SECOND EDITION

DATA CENTER HANDBOOK

PLAN, DESIGN, BUILD, AND OPERATIONS OF A SMART DATA CENTER

HWAIYU GENG P.E

WILEY

DATA CENTER HANDBOOK

DATA CENTER HANDBOOK

Plan, Design, Build, and Operations of a Smart Data Center

Second Edition

HWAIYU GENG, P.E.
Amica Research
Palo Alto, California, United States of America

WILEY

This second edition first published 2021
© 2021 by John Wiley & Sons, Inc.

Edition History
John Wiley & Sons, Inc. (1e, 2015)

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at http://www.wiley.com/go/permissions.

The right of Hwaiyu Geng, P.E. to be identified as the editor of the editorial material in this work has been asserted in accordance with law.

Registered Office
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office
111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty
While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data

Names: Geng, Hwaiyu, editor.
Title: Data center handbook : plan, design, build, and operations of a smart data center / edited by Hwaiyu Geng.
Description: 2nd edition. | Hoboken, NJ : Wiley, 2020. | Includes index.
Identifiers: LCCN 2020028785 (print) | LCCN 2020028786 (ebook) | ISBN 9781119597506 (hardback) | ISBN 9781119597544 (adobe pdf) | ISBN 9781119597551 (epub)
Subjects: LCSH: Electronic data processing departments–Design and construction–Handbooks, manuals, etc. | Electronic data processing departments–Security measures–Handbooks, manuals, etc.
Classification: LCC TH4311 .D368 2020 (print) | LCC TH4311 (ebook) | DDC 004.068/4–dc23
LC record available at https://lccn.loc.gov/2020028785
LC ebook record available at https://lccn.loc.gov/2020028786

Cover Design: Wiley
Cover Image: Particle earth with technology network over Chicago Cityscape © Photographer is my life. / Getty Images, front cover icons © Macrovector / Shutterstock except farming icon © bioraven / Shutterstock

Set in 10/12pt Times by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

To "Our Mothers Who Cradle the World" and To "Our Earth Who Gives Us Life."

BRIEF CONTENTS

ABOUT THE EDITOR/AUTHOR ix

TAB MEMBERS xi

CONTRIBUTORS xiii

FOREWORDS xv

PREFACES xxi

ACKNOWLEDGEMENTS xxv

PART I DATA CENTER OVERVIEW AND STRATEGIC PLANNING

(Chapter 1–7, Pages 1–127)

PART II DATA CENTER TECHNOLOGIES

(Chapter 8–21, Pages 143–359)

PART III DATA CENTER DESIGN & CONSTRUCTION

(Chapter 22–31, Pages 367–611)

PART IV DATA CENTER OPERATIONS MANAGEMENT

(Chapter 32–37, Pages 617–675)

ABOUT THE EDITOR/AUTHOR

Hwaiyu Geng, CMfgE, P.E., is a principal at Amica Research (Palo Alto, California, USA) promoting green technological and manufacturing programs. He has had over 40 years of diversified technological and management experience, worked with Westinghouse, Applied Materials, Hewlett-Packard, Intel, and Juniper Network on international hightech projects. He is a frequent speaker in international conferences, universities, and has presented many technical papers. A patent holder, Mr. Geng is also the editor/author of the *Data Center Handbook (2ed), Manufacturing Engineering Handbook (2ed), Semiconductor Manufacturing Handbook (2ed), and the IoT and Data Analytics Handbook*.

TECHNICAL ADVISORY BOARD

Amy Geng, M.D., Institute for Education, Washington, District of Columbia, United States of America

Bill Kosik P.E., CEM, LEED AP, BEMP, DNV GL Energy Services USA, Oak Park, Illinois, United States of America

David Fong, Ph.D., CITS Group, Santa Clara, California, United States of America

Dongmei Huang, Ph.D., Rainspur Technology, Beijing, China

Hwaiyu Geng, P.E., Amica Research, Palo Alto, California, United States of America

Jay Park, P.E., Facebook, Inc., Fremont, California, United States of America

Jonathan Jew, Co-Chair TIA TR, BICSI, ISO Standard, J&W Consultants, San Francisco, California, United States of America

Jonathan Koomey, Ph.D., President, Koomey Analytics, Burlingame, California, United States of America

Malik Megdiche, Ph.D., Schneider Electric, Eybens, France

Robert E. McFarlane, ASHRAE TC9.9 Corresponding member, ASHRAE SSPC 90.4 Voting Member, Marist College Adjunct Professor, Shen Milsom & Wilke LLC, New York City, New York, United States of America

Robert Tozer, Ph.D., MBA, CEng, MCIBSE, MASHRAE, Operational Intelligence, Ltd., London, United Kingdom

Roger R. Schmidt, Ph.D., P.E. National Academy of Engineering Member, Traugott Distinguished Professor, Syracuse University, IBM Fellow Emeritus (Retired), Syracuse, New York, United States of America

Yihlin Chan, Ph.D., Occupational Safety and Health Administration (Retired), Salt Lake City, Utah, United States of America

LIST OF CONTRIBUTORS

Ken Baudry, K.J. Baudry, Inc., Atlanta, Georgia, United States of America

Sergio Bermudez, IBM TJ Watson Research Center, Yorktown Heights, New York, United States of America

David Bonneville, Degenkolb Engineers, San Francisco, California, United States of America

David Cameron, Operational Intelligence Ltd, London, United Kingdom

Ronghui Cao, College of Information Science and Engineering, Hunan University, Changsha, China

Nicholas H. Des Champs, Munters Corporation, Buena Vista, Virginia, United States of America

Christopher Chen, Jensen Hughes, College Park, Maryland, United States of America

Chris Crosby, Compass Datacenters, Dallas, Texas, United States of America

Chris Curtis, Compass Datacenters, Dallas, Texas, United States of America

Sean S. Donohue, Jensen Hughes, Colorado Springs, Colorado, United States of America

Keith Dunnavant, Munters Corporation, Buena Vista, Virginia, United States of America

Mark Fisher, Munters Corporation, Buena Vista, Virginia, United States of America

Sophia Flucker, Operational Intelligence Ltd, London, United Kingdom

Hubertus Franke, IBM, Yorktown Heights, New York, United States of America

Ajay Garg, Intel Corporation, Hillsboro, Oregon, United States of America

Chang-Hsin Geng, Supermicro Computer, Inc., San Jose, California, United States of America

Hwaiyu Geng, Amica Research, Palo Alto, California, United States of America

Hendrik Hamann, IBM TJ Watson Research Center, Yorktown Heights, New York, United States of America

Sarah Hanna, Facebook, Fremont, California, United States of America

Skyler Holloway, Facebook, Menlo Park, California, United States of America

Ching-I Hsu, Raritan, Inc., Somerset, New Jersey, United States of America

Dongmei Huang, Beijing Rainspur Technology, Beijing, China

Robert Hunter, AlphaGuardian, San Ramon, California, United States of America

Phil Isaak, Isaak Technologies Inc., Minneapolis, Minnesota, United States of America

Alexander Jew, J&M Consultants, Inc., San Francisco, California, United States of America

Masatoshi Kajimoto, ISACA, Tokyo, Japan

Levente Klein, IBM TJ Watson Research Center, Yorktown Heights, New York, United States of America

Bill Kosik, DNV Energy Services USA Inc., Chicago, Illinois, United States of America

Nuoa Lei, Northwestern University, Evanston, Illinois, United States of America

Bang Li, Eco Atlas (Shenzhen) Co., Ltd, Shenzhen, China

Chung-Sheng Li, PricewaterhouseCoopers, San Jose, California, United States of America

Kenli Li, College of Information Science and Engineering, Hunan University, Changsha, China

Keqin Li, Department of Computer Science, State University of New York, New Paltz, New York, United States of America

Weiwei Lin, School of Computer Science and Engineering, South China University of Technology, Guangzhou, China

Chris Loeffler, Eaton, Raleigh, North Carolina, United States of America

Fernando Marianno, IBM TJ Watson Research Center, Yorktown Heights, New York, United States of America

Eric R. Masanet, Northwestern University, Evanston, Illinois, United States of America

Robert E. Mcfarlane, Shen Milsom & Wilke LLC, New York, New York, United States of America Marist College, Poughkeepsie, New York, United States of America ASHRAE TC 9.9, Atlanta, Georgia, United States of America ASHRAE SSPC 90.4 Standard Committee, Atlanta, Georgia, United States of America

Malik Megdiche, Schneider Electric, Eybens, France

Christopher O. Muller, Muller Consulting, Lawrenceville, Georgia, United States of America

Liam Newcombe, Romonet, London, United Kingdom

Jay Park, Facebook, Fremont, California, United States of America

Robert Pekelnicky, Degenkolb Engineers, San Francisco, California, United States of America

Robert Reid, Panduit Corporation, Tinley Park, Illinois, United States of America

Mark Seymour, Future Facilities, London, United Kingdom

Dror Shenkar, Intel Corporation, Israel

Ed Spears, Eaton, Raleigh, North Carolina, United States of America

Richard T. Stuebi, Institute for Sustainable Energy, Boston University, Boston, Massachusetts, United States of America

Mark Suski, Jensen Hughes, Schaumburg, Illinois, United States of America

Zhuo Tang, College of Information Science and Engineering, Hunan University, Changsha, China

Robert Tozer, Operational Intelligence Ltd, London, United Kingdom

John Weale, The Integral Group, Oakland, California, United States of America

Joseph Weiss, Applied Control Solutions, Cupertino, California, United States of America

Beth Whitehead, Operational Intelligence Ltd, London, United Kingdom

Jan Wiersma, EVO Venture Partners, Seattle, Washington, United States of America

Wentai Wu, Department of Computer Science, University of Warwick, Coventry, United Kingdom

Chao Yang, Chongqing University, Chongqing, China

Ligong Zhou, Raritan, Inc., Beijing, China

FOREWORD (1)

The digitalization of our economy requires data centers to continue to innovate to meet the new needs for connectivity, growth, security, innovation, and respect for the environment demanded by organizations. Every phase of life is putting increased pressure on data centers to innovate at a rapid pace. Explosive growth of data driven by 5G, Internet of Things (IoT), and Artificial Intelligence (AI) is changing the way data is stored, managed, and transferred. As this volume grows, data and applications are pulled together, requiring more and more computing and storage resources. The question facing data center designers and operators is how to plan for the future that accomplishes the *security, flexibility, scalability, adaptability, and sustainability needed* to support business requirements.

With this explosion of data, companies need to think more carefully and strategically about how and where their data is stored, and the security risks involved in moving data. The sheer volume of data creates additional challenges in protecting it from intrusions. This is probably one of the most important concerns of the industry – how to protect data from being hacked and being compromised in a way that would be extremely damaging to their core business and the trust of their clients.

Traditional data centers must deliver a degree of *scalability* to accommodate usage needs. With newer technologies and applications coming out daily, it is important to be able to morph the data center into the needs of the business. It is equally important to be able to integrate these technologies in a timely manner that does not compromise the strategic plans of the business. With server racks getting denser every few years, the rest of the facility must be prepared to support an ever increasing power draw. A data center built over the next decade must be expandable to accommodate for future technologies, or risk running out of room for support infrastructure. Server rooms might have more computing power in the same area, but they will also need more power and cooling to match. Institutions are also moving to install advanced applications and workloads related to AI, which requires high-performance computing. To date, these racks represent a very small percentage of total racks, but they nevertheless can present unfamiliar power and cooling challenges that must be addressed. The increasing interest in direct liquid cooling is in response to high-performance computing demands.

5G enables a new kind of network that is designed to connect virtually everyone and everything together including machines, objects, and devices. It will require more bandwidth, faster speeds, and lower latency, and the data center infrastructure must be *flexible and adaptable* in order to accommodate these demands. With the need to bring computing power closer to the point of connectivity, the end user is driving demand for edge data centers. Analyzing the data where it is created rather than sending it across various networks and data centers helps to reduce response latency, thereby removing a bottleneck from the decision-making process. In most cases, these data centers will be, remotely managed and unstaffed data centers. Machine learning will enable real-time adjustments to be made to the infrastructure without the need for human interaction.

With data growing exponentially, data centers may be impacted by significant increases in energy usage and carbon footprint. Hyperscalers have realized this and have increasingly used more and more *sustainable* technologies. This trend will cause others to follow and adopt some of the building technologies and use of renewables for their own data centers. The growing mandate for corporations to shift to a greener energy footprint lays the groundwork for new approaches to data center power.

The rapid innovations that are occurring inside (edge computing, liquid cooling, etc.) and outside (5G, IoT, etc.) of data centers will require careful and thoughtful analysis to design and operate a data center for the future that will serve the strategic imperatives of the business it supports. To help address the complex environment with competing forces, this second edition of the Data Center Handbook has assembled by leaders in the industry and academia to share their latest thinking on these issues. This handbook is the most comprehensive guide available to data center practitioners as well as academia.

Roger R. Schmidt, Ph.D.
Member, National Academy of Engineering
Traugott Distinguished Professor, Syracuse University
IBM Fellow Emeritus (Retired)

FOREWORD (2)

A key driver of innovation in modern industrial societies in the past two centuries is the application of what researchers call "general purpose technologies," which have far-ranging effects on the way the economy produces value. Some important examples include the steam engine, the telegraph, the electric power grid, the internal combustion engine, and most recently, computers and related information and communications technologies (ICTs).

ICTs represent the most powerful general-purpose technologies humanity has ever created. The pace of innovation across virtually all industries is accelerating, which is a direct result of the application of ICTs to increase efficiency, enhance organizational effectiveness, and reduce costs of manufacturing products. Services provided by data centers enable virtually all ICTs to function better.

This volume presents a comprehensive look at the current state of the data center industry. It is an essential resource for those working in the industry, and for those who want to understand where it is headed.

The importance of the data center industry has led to many misconceptions, the most common of which involves inflated estimates of how much electricity data centers use. The latest credible estimates for global electricity use of data centers are for 2018, from our article in *Science Magazine* in February 2020 (Masanet et al. 2020).

According to this analysis, data centers used about 0.9% of the world's electricity consumption in 2018 (down from 1.1% in 2010). Electricity use grew only 6% even as the number of compute instances, data transfers, and total data storage capacity grew to be 6.5 times, 11 times, and 26 times as large in 2018 as each was in 2010, respectively.

The industry was able to keep data center electricity use almost flat in absolute terms from 2010 to 2018 because of the adoption of best practices outlined in more detail in this volume. The most consequential of these best practices was the rapid adoption of hyperscale data centers, known colloquially as cloud computing. Computing output and data transfers increased rapidly, but efficiency also increased rapidly, almost completely offsetting growth in demand for computing services.

For those new to the world of data centers and information technology, this lesson is surprising. Even though data centers are increasingly important to the global economy, they don't use a lot of electricity in total, because innovation has rapidly increased their efficiency over time. If the industry aggressively adopts the advanced technologies and practices described in this volume, they needn't use a lot of electricity in the future, either.

I hope analysts and practitioners around the world find this volume useful. I surely will!

Jonathan Koomey, Ph.D.,
President, Koomey Analytics
Bay Area, California

FOREWORD (3)

The data center industry changes faster than any publication can keep up with. So why the "Data Center Handbook"? There are many reasons, but three stand out. First, fundamentals have not changed. Computing equipment may have dramatically transformed in processing power and form factor since the first mainframes appeared, but it is still housed in secure rooms, it still uses electricity, it still produces heat, it must still be cooled, it must still be protected from fire, it must still be connected to its users, and it must still be managed by humans who possess an unusual range of knowledge and an incredible ability to adapt to fast changing requirements and conditions. Second, new people are constantly entering what, to them, is this brave new world. They benefit from having grown up with a computer (i.e., "smart phone") in their hands, but are missing the contextual background behind how it came to be and what is needed to keep it working. Whether they are engineers designing their first enterprise, edge computing, hyperscale or liquid cooled facility, or IT professionals given their first facility or system management assignment within it, or are students trying to grasp the enormity of this industry, having a single reference book is far more efficient than plowing through the hundreds of articles published in multiple places every month. Third, and perhaps even more valuable in an industry that changes so rapidly, is having a volume that also directs you to the best industry resources when more or newer information is needed.

The world can no longer function without the computing industry. It's not regulated like gas and electric, but it's as critical as any utility, making it even more important for the IT industry to maintain itself reliably. When IT services fail, we are even more lost than in a power outage. We can use candles to see, and perhaps light a fireplace to stay warm. We can even make our own entertainment! But if we can't get critical news, can't pay a bill on time, or can't even make a critical phone call, the world as we now know it comes to a standstill. And that's just the personal side. Reliable, flexible, and highly adaptable computing facilities are now necessary to our very existence. Businesses have gone bankrupt after computing failures. In health care and public safety, the availability of those systems can literally spell life or death.

In this book you will find chapters on virtually every topic you could encounter in designing and operating a data center – each chapter written by a recognized expert in the field, highly experienced in the challenges, complexities, and eccentricities of data center systems and their supporting infrastructures. Each section has been brought up-to-date from the previous edition of this book as of the time of publication. But as this book was being assembled, the COVID 19 pandemic occurred, putting unprecedented demands on computing systems overnight. The industry reacted, proving beyond question its ability to respond to a crisis, adapt its operating practices to unusual conditions, and meet the inordinate demands that quickly appeared from every industry, government, and individual. A version of the famous Niels Bohr quote goes, "An expert is one who, through his own painful experience, has learned all the mistakes in a given narrow field." Adherence to the principles and practices set down by the authors of this book, in most cases gained over decades through their own personal and often painful experiences, enabled the computing industry to respond to that crisis. It will be the continued adherence to those principles, honed as the industry continues to change and mature, that will empower it to respond to the next critical situation. The industry should be grateful that the knowledge of so many experts has been assembled into one volume from which everyone in this industry can gain new knowledge.

Robert E. McFarlane
Principal, Shen Milsom & Wilke, LLC
Adjunct Faculty – Marist College, Poughkeepsie, NY

PREFACE DATA CENTER HANDBOOK (SECOND EDITION, 2021)

As Internet of Things, data analytics, artificial intelligence, 5G, and other emerging technologies revolutionize the services and products companies, the demand for computing power grows along the value chain between edge and cloud. Data centers need to improve and advance continuously to fulfill this demand.

To meet the megatrends of globalization, urbanization, demographic changes, technology advancements, and sustainability concerns, C-suite executives and technologists must work together in preparing strategic plans for deploying data centers around the world. Workforce developments and the redundancy of infrastructures required between edge and cloud need to be considered in building and positioning data centers globally.

Whether as a data center designer, user, manager, researcher, professor, or student, we all face increasing challenges in a cross-functional environment. For each data center project, we should ask, what are the goals, and work out "How to Solve It."[1] To do this, we can employ a 5W1H[2] approach applying data analytics and nurture the creativity that is needed for invention and innovation. Additionally, a good understanding of the anatomy, ecosystem, and taxonomy, of a data center will help us master and solve this complex problem.

The goal of this Data Center *Handbook* is to provide readers with the essential knowledge that is needed to plan, build, and operate a data center. This handbook embraces both emerging technologies and best practices. The handbook is divided into four parts:

Part I: *Data Center Overview and Strategic Planning* that provides an overview of data center strategic planning, while considering the impact of emerging technologies. This section also addresses energy demands, sustainability, edge to cloud computing, financial analysis, and managing data center risks.

Part II: *Data Center Technologies* that covers technologies applicable to data centers. These include software-defined applications, infrastructure, resource management, ASHRAE[3] thermal guidelines, design of energy-efficient IT equipment, wireless sensor network, telecommunication, rack level and server level cooling, data center corrosion and contamination control, cabling, cybersecurity, and data center microgrids.

Part III: *Data Center Design and Construction* that discusses plan, design, and construction of a data center that includes site selection, facility layout and rack floor plan, mechanical design, electrical design, structural design, fire protection, computational fluid dynamics, and project management for construction.

Part IV: *Data Center Operations* that covers data center benchmarking, data center infrastructure management (DCIM), energy efficiency assessment, and AI applications for data centers. This section also reviews lessons imparted from disasters, and includes mitigation strategies to ensure business continuity.

[1] Polya, G. *How to Solve It*. Princeton: Princeton University Press; 1973.
[2] The 5W1H are "Who, What, When, Where, Why, and How."
[3] ASHRAE is the American Society of Heating, Refrigerating, and Air-Conditioning Engineers.

Containing 453 figures, 101 tables and 17 pages in the index section, this second edition of *Data Center Handbook* is a single-volume, comprehensive guide to this field. The handbook covers the breadth and depth of data center technologies, and includes the latest updates from this fast-changing field. It is meant to be a relevant, practical, and enlightening resource for global data center practitioners, and will be a useful reference book for anyone whose work requires data centers.

Hwaiyu Geng, CMfgE, P.E.
Palo Alto, California, United States of America

PREFACE DATA CENTER HANDBOOK (FIRST EDITION, 2015)

Designing and operating a sustainable data center (DC) requires technical knowledge and skills from strategic planning, complex technologies, available best practices, optimum operating efficiency, disaster recovery, and more.

Engineers and managers all face challenges operating across functionalities, for example, facilities, IT, engineering, and business departments. For a mission-critical, sustainable DC project, we must consider the following:

- What are the goals?
- What are the givens?
- What are the constraints?
- What are the unknowns?
- Which are the feasible solutions?
- How is the solution validated?

How does one apply technical and business knowledge to develop an optimum solution plan that considers emerging technologies, availability, scalability, sustainability, agility, resilience, best practices, and rapid time to value? The list can go on and on. Our challenges may be as follows:

- To prepare a strategic location plan
- To design and build a mission-critical DC with energy-efficient infrastructure
- To apply best practices thus consuming less energy
- To apply IT technologies such as cloud and virtualization and
- To manage DC operations thus reducing costs and carbon footprint

A good understanding of DC components, IT technologies, and DC operations will enable one to plan, design, and implement mission-critical DC projects successfully. The goal of this handbook is to provide DC practitioners with essential knowledge needed to implement DC design and construction, apply IT technologies, and continually improve DC operations. This handbook embraces both conventional and emerging technologies, as well as best practices that are being used in the DC industry. By applying the information contained in the handbook, we can accelerate the pace of innovations to reduce energy consumption and carbon emissions and to "Save Our Earth Who Gives Us Life."

The handbook covers the following topics:

- DC strategic planning
- Hosting, colocation, site selection, and economic justifications
- Plan, design, and implement a mission-critical facility
- IT technologies including virtualization, cloud, SDN, and SDDC
- DC rack layout and MEP design
- Proven and emerging energy efficiency technologies
- DC project management and commissioning
- DC operations
- Disaster recovery and business continuity

Each chapter includes essential principles, design, and operations considerations, best practices, future trends, and further readings. The principles cover fundamentals of a technology and its applications. Design and operational considerations include system design, operations, safety, security, environment issues, maintenance, economy, and best practices. There are useful tips for planning, implementing, and controlling operational processes. The future trends and further reading sections provide visionary views

and lists of relevant books, technical papers, and websites for additional reading.

This *Data Center Handbook* is specifically designed to provide technical knowledge for those who are responsible for the design, construction, and operation of DCs. It is also useful for DC decision makers who are responsible for strategic decisions regarding capacity planning and technology investments. The following professionals and managers will find this handbook to be a useful and enlightening resource:

- C-level Executives (Chief Information Officer, Chief Technology Officer, Chief Operating Officer, Chief Financial Officer)
- Data Center Managers and Directors
- Data Center Project Managers
- Data Center Consultants
- Information Technology and Infrastructure Managers
- Network Operations Center and Security Operations Center Managers
- Network, Cabling, and Communication Engineers
- Server, Storage, and Application Managers
- IT Project Managers
- IT Consultants
- Architects and MEP Consultants
- Facilities Managers and Engineers
- Real Estate Portfolio Managers
- Finance Managers

This *Data Center Handbook* is prepared by more than 50 world-class professionals from eight countries around the world. It covers the breadth and depth of DC planning, designing, construction, and operating enterprise, government, telecommunication, or R&D Data Centers. This *Data Center Handbook* is sure to be the most comprehensive single-source guide ever published in its field.

Hwaiyu Geng, CMfgE, P.E.
Palo Alto, California, United States of America

ACKNOWLEDGEMENTS
DATA CENTER HANDBOOK (SECOND EDITION, 2021)

The Data Center Handbook is a collective representation of an international community with scientists and professionals comprising 58 experts from six countries around the world.

I am very grateful to the members of the Technical Advisory Board for their diligent reviews of this handbook, confirming technical accuracy while contributing their unique perspectives. Their guidance has been invaluable to ensure that the handbook can meet the needs of a broad audience.

I gratefully acknowledge to the contributors who share their wisdom and valuable experiences in spite of their busy schedules and personal lives.

Without the trust and support from our team members, this handbook could not have been completed. Their collective effort has resulted in a work that adds tremendous value to the data center community.

Thanks must go to the following individuals for their advice, support, and contribution:

- Nicholas H. Des Champs, Munters Corporation
- Mark Gaydos, Nlyte Software
- Dongmei Huang, Rainspur Technology
- Phil Isaak, Isaak Technologies
- Jonathan Jew, J&M Consultants
- Levente Klein, IBM
- Bill Kosik, DNV Energy Services USA Inc.
- Chung-Sheng Li, PricewaterhouseCoopers
- Robert McFarlane, Shen Milsom & Wilke
- Malik Megdiche, Schneider Electric
- Christopher Muller, Muller Consulting
- Liam Newcombe, Romonet Ltd.
- Roger Schmidt, National Academy of Engineering Member
- Mark Seymour, Future Facilities
- Robert Tozer, Operational Intelligence
- John Weale, the Integral Group.

This book benefited from the following organizations and institutes and more:

- 7×24 Exchange International
- ASHRAE (American Society of Heating, Refrigerating, and Air Conditioning Engineers)
- Asetek
- BICSI (Building Industry Consulting Service International)
- Data Center Knowledge
- Data Center Dynamics
- ENERGY STAR (the U.S. Energy Protection Agency)
- European Commission Code of Conduct
- Federal Energy Management Program (the U.S. Dept. of Energy)
- Gartner
- Green Grid, The
- IDC (International Data Corporation)
- Japan Data Center Council
- LBNL (the U.S. Dept. of Energy, Lawrence Berkeley National Laboratory)
- LEED (the U.S. Green Building Council, Leadership in Energy and Environmental Design)
- McKinsey Global Institute
- Mission Critical Magazine
- NIST (the U.S. Dept. of Commerce, National Institute of Standards and Technology)

- NOAA (the U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration)
- NASA (the U.S. Dept. of Interior, National Aeronautics and Space Administration)
- Open Compute Project
- SPEC (Standard Performance Evaluation Corporation)
- TIA (Telecommunications Industry Association)
- Uptime Institute/451 Research

Thanks are also due to Brett Kurzman and staff at Wiley for their support and guidance.

My special thanks to my wife, Limei, my daughters, Amy and Julie, and my grandchildren, Abby, Katy, Alex, Diana, and David, for their support and encouragement while I was preparing this book.

Hwaiyu Geng, CMfgE, P.E.
Palo Alto, California, United States of America

ACKNOWLEDGMENTS
DATA CENTER HANDBOOK (FIRST EDITION, 2015)

The Data Center Handbook is a collective representation of an international community with scientists and professionals from eight countries around the world. Fifty-one authors, from data center industry, R&D, and academia, plus fifteen members at Technical Advisory Board have contributed to this book. Many suggestions and advice were received while I prepared and organized the book.

I gratefully acknowledge the contributors who dedicated their time in spite of their busy schedule and personal lives to share their wisdom and valuable experience.

I would also like to thank the members at Technical Advisory Board for their constructive recommendations on the structure of this handbook and thorough peer review of book chapters.

My thanks also go to Brett Kurzman, Alex Castro, Katrina Maceda at Wiley and F. Pascal Raj at SPi Global whose can do spirit and teamwork were instrumental in producing this book.

Thanks and appreciation must go to the following individuals for their advice, support, and contributions:

Sam Gelpi, Hewlett-Packard Company

Dongmei Huang, Ph.D., Rainspur Technology, China

Madhu Iyengar, Ph.D., Facebook, Inc.

Jonathan Jew, J&M Consultants

Jonathan Koomey, Ph.D., Stanford University

Tomoo Misaki, Nomura Research Institute, Ltd., Japan

Veerendra Mulay, Ph.D., Facebook, Inc.

Jay Park, P.E., Facebook, Inc.

Roger Schmidt, Ph.D., IBM Corporation

Hajime Takagi GIT Associates, Ltd., Japan

William Tschudi, P.E., Lawrence Berkeley National Laboratory

Kari Capone, John Wiley & Sons, Inc.

This book benefited from the following organizations and institutes:

7 × 24 Exchange International

American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE)

Building Industry Consulting Service International (BICSI)

Datacenter Dynamics

European Commission Code of Conduct

The Green Grid

Japan Data Center Council

Open Compute Project

Silicon Valley Leadership Group

Telecommunications Industry Association (TIA)

Uptime Institute/451 Research

U.S. Department of Commerce, National Institute of Standards and Technology

U.S. Department of Energy, Lawrence Berkeley National Laboratory

U.S. Department of Energy, Oak Ridge National Laboratory

U.S. Department of Energy, Office of Energy Efficiency & Renewable Energy

U.S. Department of Homeland Security, Federal Emergency Management Administration

U.S. Environmental Protection Agency, ENERGY STAR Program

U.S. Green Building Council, Leadership in Energy & Environmental Design

My special thanks to my wife, Limei, my daughters, Amy and Julie, and grandchildren for their understanding, support, and encouragement when I was preparing this book.

TABLE OF CONTENTS

PART I DATA CENTER OVERVIEW AND STRATEGIC PLANNING

1 Sustainable Data Center: Strategic Planning, Design, Construction, and Operations with Emerging Technologies 1
Hwaiyu Geng

1.1 Introduction 1
1.2 Advanced Technologies 2
1.3 Data Center System and Infrastructure Architecture 6
1.4 Strategic Planning 6
1.5 Design and Construction Considerations 8
1.6 Operations Technology and Management 9
1.7 Business Continuity and Disaster Recovery 10
1.8 Workforce Development and Certification 11
1.9 Global Warming and Sustainability 11
1.10 Conclusions 12
References 12
Further Reading 13

2 Global Data Center Energy Demand and Strategies to Conserve Energy 15
Nuoa Lei and Eric R. Masanet

2.1 Introduction 15
2.2 Approaches for Modeling Data Center Energy Use 16
2.3 Global Data Center Energy Use: Past and Present 17
2.4 Global Data Center Energy Use: Forward-Looking Analysis 19
2.5 Data Centers and Climate Change 21
2.6 Opportunities for Reducing Energy Use 21
2.7 Conclusions 24
References 24
Further Reading 26

3 Energy and Sustainability in Data Centers 27
Bill Kosik

3.1 Introduction 27
3.2 Modularity in Data Centers 32

3.3 Cooling a Flexible Facility 33
3.4 Proper Operating Temperature and Humidity 35
3.5 Avoiding Common Planning Errors 37
3.6 Design Concepts for Data Center Cooling Systems 40
3.7 Building Envelope and Energy Use 42
3.8 Air Management and Containment Strategies 44
3.9 Electrical System Efficiency 46
3.10 Energy Use of IT Equipment 48
3.11 Server Virtualization 50
3.12 Interdependency of Supply Air Temperature and ITE Energy Use 51
3.13 IT and Facilities Working Together to Reduce Energy Use 52
3.14 Data Center Facilities Must Be Dynamic and Adaptable 53
3.15 Server Technology and Steady Increase of Efficiency 53
3.16 Data Collection and Analysis for Assessments 54
3.17 Private Industry and Government Energy Efficiency Programs 55
3.18 Strategies for Operations Optimization 59
3.19 Utility Customer-Funded Programs 60
References 62
Further Reading 62

4 Hosting or Colocation Data Centers 65
Chris Crosby and Chris Curtis

4.1 Introduction 65
4.2 Hosting 65
4.3 Colocation (Wholesale) 66
4.4 Types of Data Centers 66
4.5 Scaling Data Centers 72
4.6 Selecting and Evaluating DC Hosting and Wholesale Providers 72
4.7 Build Versus Buy 72
4.8 Future Trends 74
4.9 Conclusion 74
References 75
Further Reading 75

5 Cloud and Edge Computing 77
Jan Wiersma

5.1 Introduction to Cloud and Edge Computing 77
5.2 IT Stack 78
5.3 Cloud Computing 79
5.4 Edge Computing 84
5.5 Future Trends 86
References 87
Further Reading 87

6 Data Center Financial Analysis, ROI, and TCO 89
Liam Newcombe

6.1 Introduction to Financial Analysis, Return on Investment, and Total Cost of Ownership 89
6.2 Financial Measures of Cost and Return 97
6.3 Complications and Common Problems 104

6.4 A Realistic Example 114
6.5 Choosing to Build, Reinvest, Lease, or Rent 124
Further Reading 126

7 Managing Data Center Risk 127
Beth Whitehead, Robert Tozer, David Cameron and Sophia Flucker

7.1 Introduction 127
7.2 Background 127
7.3 Reflection: The Business Case 129
7.4 Knowledge Transfer 1 131
7.5 Theory: The Design Phase 131
7.6 Knowledge Transfer 2 136
7.7 Practice: The Build Phase 136
7.8 Knowledge Transfer 3: Practical Completion 137
7.9 Experience: Operation 138
7.10 Knowledge Transfer 4 140
7.11 Conclusions 140
References 141

PART II DATA CENTER TECHNOLOGIES

8 Software-Defined Environments 143
Chung-Sheng Li and Hubertus Franke

8.1 Introduction 143
8.2 Software-Defined Environments Architecture 144
8.3 Software-Defined Environments Framework 145
8.4 Continuous Assurance on Resiliency 149
8.5 Composable/Disaggregated Datacenter Architecture 150
8.6 Summary 151
References 152

9 Computing, Storage, and Networking Resource Management in Data Centers 155
Ronghui Cao, Zhuo Tang, Kenli Li and Keqin Li

9.1 Introduction 155
9.2 Resource Virtualization and Resource Management 155
9.3 Cloud Platform 157
9.4 Progress from Single-Cloud to Multi-Cloud 159
9.5 Resource Management Architecture in Large-Scale Clusters 160
9.6 Conclusions 162
References 162

10 Wireless Sensor Networks to Improve Energy Efficiency in Data Centers 163
Levente Klein, Sergio Bermudez, Fernando Marianno and Hendrik Hamann

10.1 Introduction 163
10.2 Wireless Sensor Networks 164
10.3 Sensors and Actuators 165
10.4 Sensor Analytics 166
10.5 Energy Savings 169

10.6 Control Systems 170
10.7 Quantifiable Energy Savings Potential 172
10.8 Conclusions 174
References 174

11 ASHRAE Standards and Practices for Data Centers 175
Robert E. Mcfarlane

11.1 Introduction: ASHRAE and Technical Committee TC 9.9 175
11.2 The Groundbreaking ASHRAE "Thermal Guidelines" 175
11.3 The *Thermal Guidelines* Change in Humidity Control 177
11.4 A New Understanding of Humidity and Static Discharge 178
11.5 High Humidity and Pollution 178
11.6 The ASHRAE "Datacom Series" 179
11.7 The ASHRAE Handbook and TC 9.9 Website 187
11.8 ASHRAE Standards and Codes 187
11.9 ANSI/ASHRAE Standard 90.1-2010 and Its Concerns 188
11.10 The Development of ANSI/ASHRAE Standard 90.4 188
11.11 Summary of ANSI/ASHRAE Standard 90.4 189
11.12 ASHRAE Breadth and The *ASHRAE Journal* 190
References 190
Further Reading 191

12 Data Center Telecommunications Cabling and TIA Standards 193
Alexander Jew

12.1 Why Use Data Center Telecommunications Cabling Standards 193
12.2 Telecommunications Cabling Standards Organizations 194
12.3 Data Center Telecommunications Cabling Infrastructure Standards 195
12.4 Telecommunications Spaces and Requirements 196
12.5 Structured Cabling Topology 200
12.6 Cable Types and Maximum Cable Lengths 201
12.7 Cabinet and Rack Placement (Hot Aisles and Cold Aisles) 205
12.8 Cabling and Energy Efficiency 206
12.9 Cable Pathways 208
12.10 Cabinets and Racks 208
12.11 Patch Panels and Cable Management 208
12.12 Reliability Ratings and Cabling 209
12.13 Conclusion and Trends 209
Further Reading 210

13 Air-Side Economizer Technologies 211
Nicholas H. Des Champs, Keith Dunnavant and Mark Fisher

13.1 Introduction 211
13.2 Using Properties of Ambient Air to Cool a Data Center 212
13.3 Economizer Thermodynamic Process and Schematic of Equipment Layout 213
13.4 Comparative Potential Energy Savings and Required Trim Mechanical Refrigeration 221
13.5 Conventional Means for Cooling Datacom Facilities 224
13.6 A Note on Legionnaires' Disease 224
References 225
Further Reading 225

14 Rack-Level Cooling and Server-Level Cooling 227
Dongmei Huang, Chao Yang and Bang Li

14.1 Introduction 227
14.2 Rack-Level Cooling 228
14.3 Server-Level Cooling 234
14.4 Conclusions and Future Trends 236
Acknowledgement 237
Further Reading 237

15 Corrosion and Contamination Control for Mission Critical Facilities 239
Christopher O. Muller

15.1 Introduction 239
15.2 Data Center Environmental Assessment 240
15.3 Guidelines and Limits for Gaseous Contaminants 241
15.4 Air Cleaning Technologies 242
15.5 Contamination Control for Data Centers 243
15.6 Testing for Filtration Effectiveness and Filter Life 248
15.7 Design/Application of Data Center Air Cleaning 249
15.8 Summary and Conclusion 252
15.9 Appendix 1: Additional Data Center Services 252
15.10 Appendix 2: Data Center History 253
15.11 Appendix 3: Reactivity Monitoring Data Examples: Sample Corrosion Monitoring Report 256
15.12 Appendix 4: Data Center Case Study 260
Further Reading 261

16 Rack PDU for Green Data Centers 263
Ching-I Hsu and Ligong Zhou

16.1 Introduction 263
16.2 Fundamentals and Principles 264
16.3 Elements of the System 271
16.4 Considerations for Planning and Selecting Rack PDUs 280
16.5 Future Trends for Rack PDUs 287
Further Reading 289

17 Fiber Cabling Fundamentals, Installation, and Maintenance 291
Robert Reid

17.1 Historical Perspective and The "Structured Cabling Model" for Fiber Cabling 291
17.2 Development of Fiber Transport Services (FTS) by IBM 292
17.3 Architecture Standards 294
17.4 Definition of Channel vs. Link 298
17.5 Network/Cabling Elements 300
17.6 Planning for Fiber-Optic Networks 304
17.7 Link Power Budgets and Application Standards 309
17.8 Link Commissioning 312
17.9 Troubleshooting, Remediation, and Operational Considerations for the Fiber Cable Plant 316
17.10 Conclusion 321
Reference 321
Further Reading 321

18 Design of Energy-Efficient IT Equipment **323**
Chang-Hsin Geng

18.1 Introduction 323
18.2 Energy-Efficient Equipment 324
18.3 High-Efficient Compute Server Cluster 324
18.4 Process to Design Energy-Efficient Servers 331
18.5 Conclusion 335
Acknowledgement 336
References 336
Further Reading 336

19 Energy-Saving Technologies of Servers in Data Centers **337**
Weiwei Lin, Wentai Wu and Keqin Li

19.1 Introduction 337
19.2 Energy Consumption Modeling of Servers in Data Centers 338
19.3 Energy-Saving Technologies of Servers 341
19.4 Conclusions 347
Acknowledgments 347
References 347

20 Cybersecurity and Data Centers **349**
Robert Hunter and Joseph Weiss

20.1 Introduction 349
20.2 Background of OT Connectivity in Data Centers 349
20.3 Vulnerabilities and Threats to OT Systems 350
20.4 Legislation Covering OT System Security 352
20.5 Cyber Incidents Involving Data Center OT Systems 353
20.6 Cyberattacks Targeting OT Systems 354
20.7 Protecting OT Systems from Cyber Compromise 355
20.8 Conclusion 357
References 358

21 Consideration of Microgrids for Data Centers **359**
Richard T. Stuebi

21.1 Introduction 359
21.2 Description of Microgrids 360
21.3 Considering Microgrids for Data Centers 362
21.4 U.S. Microgrid Market 364
21.5 Concluding Remarks 365
References 365
Further Reading 365

PART III DATA CENTER DESIGN & CONSTRUCTION

22 Data Center Site Search and Selection **367**
Ken Baudry

22.1 Introduction 367
22.2 Site Searches Versus Facility Searches 367
22.3 Globalization and the Speed of Light 368
22.4 The Site Selection Process 370
22.5 Industry Trends Affecting Site Selection 379

Acknowledgment 380
Reference 380
Further Reading 380

23 Architecture: Data Center Rack Floor Plan and Facility Layout Design 381
Phil Isaak

23.1 Introduction 381
23.2 Fiber Optic Network Design 381
23.3 Overview of Rack and Cabinet Design 386
23.4 Space and Power Design Criteria 389
23.5 Pathways 390
23.6 Coordination with Other Systems 392
23.7 Computer Room Design 395
23.8 Scalable Design 398
23.9 CFD Modeling 400
23.10 Data Center Space Planning 400
23.11 Conclusion 402
Further Reading 402

24 Mechanical Design in Data Centers 403
Robert Mcfarlane and John Weale

24.1 Introduction 403
24.2 Key Design Criteria 403
24.3 Mechanical Design Process 407
24.4 Data Center Considerations in Selecting Key Components 424
24.5 Primary Design Options 429
24.6 Current Best Practices 436
24.7 Future Trends 438
Acknowledgment 440
Reference 440
Further Reading 440

25 Data Center Electrical Design 441
Malik Megdiche, Jay Park and Sarah Hanna

25.1 Introduction 441
25.2 Design Inputs 441
25.3 Architecture Resilience 443
25.4 Electrical Design Challenges 450
25.5 Facebook, Inc. Electrical Design 477
Further Reading 481

26 Electrical: Uninterruptible Power Supply System 483
Chris Loeffler and Ed Spears

26.1 Introduction 483
26.2 Principal of UPS and Application 484
26.3 Considerations in Selecting UPS 498
26.4 Reliability and Redundancy 502
26.5 Alternate Energy Sources: AC and DC 513
26.6 UPS Preventive Maintenance Requirements 515
26.7 UPS Management and Control 517
26.8 Conclusion and Trends 520
Further Reading 520

27 Structural Design in Data Centers: Natural Disaster Resilience 521
David Bonneville and Robert Pekelnicky

27.1 Introduction 521
27.2 Building Design Considerations 523
27.3 Earthquakes 524
27.4 Hurricanes, Tornadoes, and Other Windstorms 527
27.5 Snow and Rain 528
27.6 Flood and Tsunami 529
27.7 Comprehensive Resiliency Strategies 530
References 532

28 Fire Protection and Life Safety Design in Data Centers 533
Sean S. Donohue, Mark Suski and Christopher Chen

28.1 Fire Protection Fundamentals 533
28.2 AHJs, Codes, and Standards 534
28.3 Local Authorities, National Codes, and Standards 534
28.4 Life Safety 535
28.5 Passive Fire Protection 537
28.6 Active Fire Protection and Suppression 537
28.7 Detection, Alarm, and Signaling 546
28.8 Fire Protection Design & Conclusion 549
References 549

29 Reliability Engineering for Data Center Infrastructures 551
Malik Megdiche

29.1 Introduction 551
29.2 Dependability Theory 552
29.3 System Dysfunctional Analysis 558
29.4 Application To Data Center Dependability 569
Further Reading 578

30 Computational Fluid Dynamics for Data Centers 579
Mark Seymour

30.1 Introduction 579
30.2 Fundamentals of CFD 580
30.3 Applications of CFD for Data Centers 588
30.4 Modeling the Data Center 592
30.5 Potential Additional Benefits of a CFD-Based Digital Twin 607
30.6 The Future of CFD-Based Digital Twins 608
References 609

31 Data Center Project Management 611
Skyler Holloway

31.1 Introduction 611
31.2 Project Kickoff Planning 611
31.3 Prepare Project Scope of Work 611
31.4 Organize Project Team 612
31.5 Project Schedule 613
31.6 Project Costs 615
31.7 Project Monitoring and Reporting 616
31.8 Project Closeout 616
31.9 Conclusion 616
Further Reading 616

PART IV DATA CENTER OPERATIONS MANAGEMENT

32 Data Center Benchmark Metrics **617**
Bill Kosik

32.1 Introduction 617
32.2 The Green Grid's PUE: A Useful Metric 617
32.3 Metrics for Expressing Partial Energy Use 618
32.4 Applying PUE in the Real World 619
32.5 Metrics Used in Data Center Assessments 620
32.6 The Green Grids XUE Metrics 620
32.7 RCI and RTI 621
32.8 Additional Industry Metrics and Standards 621
32.9 European Commission Code of Conduct 624
32.10 Conclusion 624
Further Reading 624

33 Data Center Infrastructure Management **627**
Dongmei Huang

33.1 What Is Data Center Infrastructure Management 627
33.2 Triggers for DCIM Acquisition and Deployment 629
33.3 What Are Modules of a DCIM Solution 631
33.4 The DCIM System Itself: What to Expect and Plan for 636
33.5 Critical Success Factors When Implementing a DCIM System 639
33.6 DCIM and Digital Twin 641
33.7 Future Trends in DCIM 642
33.8 Conclusion 643
Acknowledgment 643
Further Reading 643

34 Data Center Air Management **645**
Robert Tozer and Sophia Flucker

34.1 Introduction 645
34.2 Cooling Delivery 645
34.3 Metrics 648
34.4 Air Containment and Its Impact on Air Performance 651
34.5 Improving Air Performance 652
34.6 Conclusion 656
References 656

35 Energy Efficiency Assessment of Data Centers Using Measurement and Management Technology **657**
Hendrik Hamann, Fernando Marianno and Levente Klein

35.1 Introduction 657
35.2 Energy Consumption Trends in Data Centers 657
35.3 Cooling Infrastructure in a Data Center 658
35.4 Cooling Energy Efficiency Improvements 659
35.5 Measurement and Management Technology (MMT) 660
35.6 MMT-Based Best Practices 661
35.7 Measurement and Metrics 662
35.8 Conclusions 667
References 668

36 Drive Data Center Management and Build Better AI with IT Devices As Sensors **669**
Ajay Garg and Dror Shenkar

36.1 Introduction 669
36.2 Current Situation of Data Center Management 669
36.3 AI Introduced in Data Center Management 670
36.4 Capabilities of IT Devices Used for Data Center Management 670
36.5 Usage Models 670
36.6 Summary and Future Perspectives 673
Further Reading 673

37 Preparing Data Centers for Natural Disasters and Pandemics **675**
Hwaiyu Geng and Masatoshi Kajimoto

37.1 Introduction 675
37.2 Design for Business Continuity and Disaster Recovery 675
37.3 Natural Disasters 676
37.4 The 2011 Great East Japan Earthquake 676
37.5 The 2012 Eastern U.S. Coast Superstorm Sandy 679
37.6 The 2019 Coronavirus Disease (COVID-19) Pandemic 683
37.7 Conclusions 683
References 684
Further Reading 684

INDEX **687**

PART I

DATA CENTER OVERVIEW AND STRATEGIC PLANNING

1

SUSTAINABLE DATA CENTER: STRATEGIC PLANNING, DESIGN, CONSTRUCTION, AND OPERATIONS WITH EMERGING TECHNOLOGIES

Hwaiyu Geng

Amica Research, Palo Alto, California, United States of America

1.1 INTRODUCTION

The earliest known use of the term "megatrend" was in 1980s published in the Christian Science Monitor (Boston). Oxford dictionary defines megatrend as "An important shift in the progress of a society." Internet searches reveal many megatrend reports that were published by major consulting firms including Accenture, Frost, KPMG, McKinsey Global Institute, PwC, etc. as well as organizations such as UN (United Nations)* and OECD (Organization for Economic Co-operation and Development [1]). One can quickly summarize key megatrends reported that include globalization, urbanization, demographic trend, technological breakthroughs, and climate changes.

Globalization: From Asia to Africa, multinational corporations are expanding their manufacturing and R&D at a faster pace and on a larger scale than ever before. Globalization widely spreads knowledge, technologies, and modern business practices at a faster space that facilitate international cooperation. Goods and services inputs are increasingly made of countries from emerging economies who join key global players. Global value chains focus on national innovation capacities and enhance national industrial specialization. Standardization, compatibility, and harmonization are even more important in a global interlaced environment.

Urbanization: Today, more than half of the world's population live in urban areas, and more people are moving to the urban areas every day. The impacts from urbanization are enormous. Demands for infrastructure, jobs, and services must be met. Problems of human health, crime, and pollution of the environment must be solved.

Demographic trend: Longer life expectancy and lower fertility rate are leading to rapidly aging populations. We must deal with increasing population, food and water shortages, and preserving natural resources. At the same time, sex discrimination, race and wealth inequalities in every part of the world must be dealt with.

Technological changes: New technologies create both challenges and opportunities. Technological breakthroughs include Internet of Things (IoT), cyber–physical systems (CPS), data analytics, artificial intelligence (AI), robotics, autonomous vehicles (AVs) (robots, drones), cloud and edge computing, and many other emerging technologies that fuel more innovative applications. These technologies fundamentally change our lifestyle and its ecosystem. Industries may be disrupted, but more inventions and innovations are nurturing.

Climate change and sustainability: Unusual patterns of droughts, floods, and hurricanes are already happening. The world is experiencing the impacts of climate change, from melting glaciers to rising sea level to extreme weather patterns. In the April 17, 2020, *Science* magazine issue, researchers examine tree rings and report that the drought from 2000 to 2018 in the southwestern of North America is among the worst "megadroughts" that have stricken the region in the last 1,200 years. The United Nation's IPCC (Intergovernmental Panel on Climate Change) reports have described increasing dangers of climate change. At the current rising rate of

* https://www.un.org/development/desa/publications/wp-content/uploads/sites/10/2020/09/20-124-UNEN-75Report-2-1.pdf

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

greenhouse gas emissions, the global average temperature will rise by more than 3°C in the twenty-first century. Rising temperatures must be kept below 2°C before year 2050 or potential irreversible environmental changes will occur. It is imperative to find sustainable solutions and delay climate change.

This chapter will start with megatrends and emerging technologies that provide insightful roadmap of future data centers and essential elements to be included when designing and implementing a data center project.

1.1.1 Data Center Definition

Data centers are being used to orchestrate every aspect of our life that covers food, clothing, shelter, transportation, healthcare, social activities, etc. The U.S. Environmental Protection Agency defines a data center as:

- "Primarily electronic equipment used for data processing (servers), data storage (storage equipment), and communications (network equipment). Collectively, this equipment processes, stores, and transmits digital information."
- "Specialized power conversion and backup equipment to maintain reliable, high-quality power, as well as environmental control equipment to maintain the proper temperature and humidity for the ICT (information and communication technologies) equipment."

A data center could also be called data hall, data farm, data warehouse, AI lab, R&D software lab, high-performance computing lab, hosting facility, colocation, computer room, server room, etc.

An exascale data center has computing systems that perform calculation over a petaflop (a million trillion floating-point) operations. Exascale data centers are elastically configured and deployed that can meet specific workloads and be optimized for future developments in power and cooling technology.[1]

The size of a data center could range from a small closet to a hyperscale data center. The term hyperscale refers to a resilient and robust computer architecture that has the ability to increase computing ability in memory, networking, and storage resources.

Regardless of size and what it is called, all data centers perform one thing, that is, to process and deliver information.

1.1.2 Data Center Energy Consumption Trends

The energy consumption trend depends on a combination of factors including data traffic, emerging technologies, ICT equipment, and energy demand by infrastructure in data centers. The trend is a complicated and dynamic model. According to "United States Data Center Energy Usage Report, Lawrence Berkeley National Laboratory" (2016) by Arman Shehabi, Jonathan Koomey, et al. [2], U.S. data center electricity used by servers, storage, network equipment, and infrastructure in 2014 consumed an estimated of 70 billion kWh. That represents about 1.8% of total U.S. electricity consumption. The U.S. electricity used by data centers in 2016 was 2% of global electricity. For 70 billion kWh, it is equivalent to 8 nuclear reactors with 1,000 MW baseload each. 70 billion kWh provides enough energy for use by 5.9 million homes in 1 year.[2] It is equivalent to 50 million ton of carbon dioxide emission to the atmosphere. It is expected that electricity consumption will continue to increase and data centers must be valiantly controlled to conserve energy use.

1.2 ADVANCED TECHNOLOGIES

The United Nations predicts that the world's population of 7.8 billion people in 2020 will reach 8.5 billion in 2030 and 9.7 billion in 2050.[3] Over 50% of the world's population are Internet users that demand more uses of data centers. This section will discuss some of the important emerging technologies illustrated by its anatomy, ecosystem, and taxonomy. Anatomy defines components of a technology. Ecosystem describes who uses the technology. Taxonomy is to classify the components of a technology and their providers in different groups. With a good understanding of what is anatomy, ecosystem, and taxonomy of a technology, one can effectively apply and master the technology.

1.2.1 Internet of Things

The first industrial revolution (IR) started with the invention of mechanical powers. The second IR happened with the invention of assembly line and electrical power. The third IR came about with computers and automation. The fourth IR took place around 2014 as a result of the invention of IoT. IDC (International Data Corporation) forecasts an expected IoT market size of $1.1 trillion in 2023. By 2025, there will be 41.6 billion IoT connected devices that will generate 79.4 zettabytes (ZB) of data.

[1] http://www.hp.com/hpinfo/newsroom/press_kits/2008/cloudresearch/fs_exascaledatacenter.pdf

[2] https://eta.lbl.gov/publications/united-states-data-center-energy

[3] https://population.un.org/wpp/Graphs/1_Demographic%20Profiles/World.pdf

The IoT is a series of hardware coupling with software and protocols to collect, analyze, and distribute information. Using the human body as an analogy, humans have five basic senses or sensors that collect information. Nervous system acts as a network that distributes information. And the brain is accountable for storing, analyzing, and giving direction through the nervous system to five senses to execute decision. The IoT works similar to the combination of five senses, the nervous system and the brain.

1.2.1.1 Anatomy

Anatomy of the IoT comprises of all components in the following formula:

$$\begin{aligned}\text{Internet of Things} = {} & \text{Things}\left(\text{sensors/cameras/actuators}\right)\\ & + \text{edge / fog computing and AI}\\ & + \text{Wi-Fi / gateway / 5G / Internet}\\ & + \text{cloud computing / data analytics / AI}\\ & + \text{insight presentations / actions}\end{aligned}$$

Each "Thing" has a unique IPv4 or IPv6 address. A "Thing" could be a person, an animal, an AV, or alike that is interconnected at many other "Things." With increasing miniaturization and built-in AI logics, sensors are performing more computing at "edge" as well as other components in the IoT's value chain before arriving at data centers for "cloud computing." AI is embedded in every component and becomes an integral part of the IoT. This handbook considers Artificial Intelligence of Things (AIoT) the same as the IoT.

1.2.1.2 Ecosystem

There are consumer-, government-, and enterprise-facing customers within an IoT's ecosystem (Fig. 1.1). Each IoT platform contains applications that are protected by a cybersecurity system. Consumer-facing customers be composed of smart home, smart entertainment, smart health, etc. Government-facing customers are composed of smart cities, smart transportation, smart grid, etc. Enterprise-facing customers include smart retail, smart manufacturing, smart finance, etc.

1.2.1.3 Taxonomy

Using taxonomy in a hospital as an analogy, a hospital has an admission office, medical record office, internal medicine, cardiology, neurology, radiology, medical laboratory, therapeutic services, pharmacy, nursing, dietary, etc. IoT's taxonomy encompasses suppliers who provide products, equipment, or services that cover sensors (microprocessor unit, system on chip, etc.), 5G, servers, storage, network, security, data analytics, AI services, industry solutions, etc.

The Industrial IoT (IIoT) and CPS connect with many smaller IoTs. They are far more complicated in design and applications than consumer-facing IoTs.

1.2.2 Big Data Analytics and Artificial Intelligence

Data analytics is one of the most important components in IoT's value chain. Big data in size and complexity, structured, semi-structured, and unstructured, outstrips the abilities to be processed by traditional data management systems.

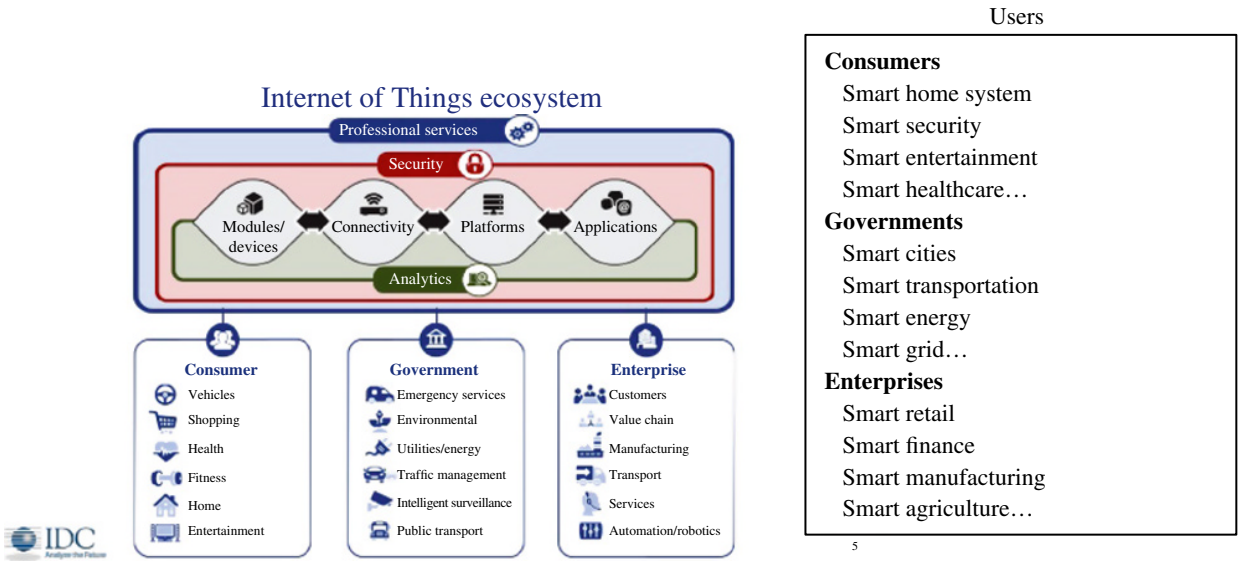


FIGURE 1.1 Internet of Things ecosystem. Source: IDC, Amica Research

1.2.2.1 Big Data Characteristics

Big data has five main characteristics that are called five V's or volume, velocity, variety, veracity, and value.

Big data signifies a huge amount of data that is produced in a short period of time. A unit of measurement (UM) is entailed to define "big." The U.S. Library of Congress (LoC) is the largest library in the world that contains 167 million items occupying 838 miles (1,340 km) of bookshelves. This quantity of information is equivalent to 15 terabytes (TB) or 15×10^6 MB of digital data.[4] Using the contents of the Library of Congress as a UM is a good way to visualize the amount of information in 15TB of digital data.

Vast stream of data is being captured by AVs for navigation and analytics consequently to develop a safe and fully automated driving experience. AV collects data from cameras, lidars, sensors, and GPS that could exceed 4 TB of data per day. Tesla sold 368,000 AVs in 2019, which is 537,280,000 TB of data or 35,800 LoCs. This is only for one car model collected in 1 year. Considering data collected from all car models, airplanes, and devices in the universe, IDC forecasts there will be 163 ZB (1 ZB = 10^9 TB) of data by 2025, which is 10.9 million LoCs.

Velocity refers to speed at which new data is generated, analyzed, and moved around. Imagining AV navigation, social media message exchanges, credit card transaction execution, or high-frequency buying or selling stocks in milliseconds, the demands for execution must be immediate with high speed.

Variety denotes the different types of data. Structured data can be sorted and organized in tables or relational data-bases. The most common example is a table containing sales information by product, region, and duration. Nowadays the majority of data is unstructured data, such as social media conversations, photos, videos, voice recording, and sensor information that cannot fit into a table. Novel big data technology, including "Unstructured Data Management-as-a-Service," harnesses and sorts unstructured data into a structured manner that can be examined for relationships.

Veracity implies authenticity, credibility, and trustworthiness of the data. With big data received and processed at high speed, quality and accuracy of some data are at risk. They must be controlled to ensure reliable information is provided to users.

Last "v" but not least is value. Fast-moving big data in different variety and veracity is only useful if it has the ability to add value to users. It is imperative that big data analytics extracts business intelligence and adds value to data-driven management to make the right decision.

1.2.2.2 Data Analytics Anatomy

The IoT, mobile telecom, social media, etc. generate data with complexity through new forms, at high speed in real time and at a very large scale. Once the big data is sorted and organized using big data algorithms, the data are ready for analytical process (Fig. 1.2). The process starts from less sophisticated descriptive to highly sophisticated prescriptive analytics that ultimately brings value to users.

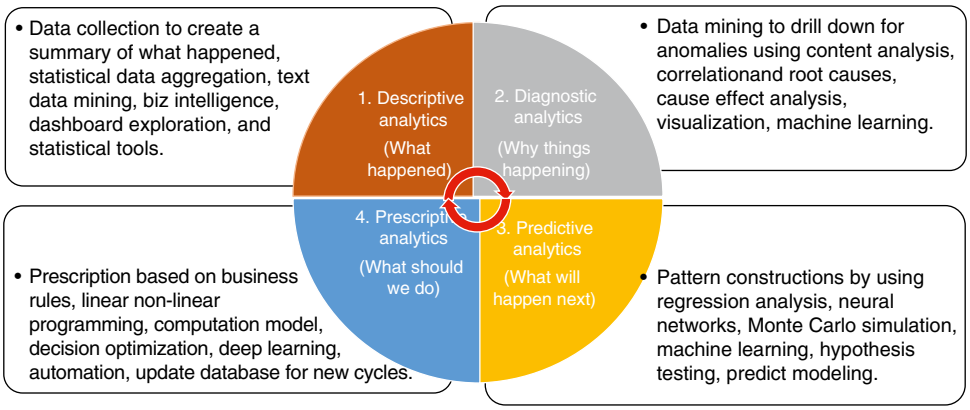


FIGURE 1.2 Virtuous Cycle of data analytics process with increasing difficulty and value. Source: © 2021 Amica Research.

[4] https://blogs.loc.gov/thesignal/2012/04/a-library-of-congress-worth-of-data-its-all-in-how-you-define-it/

Descriptive analytics does exactly what the name implies. It gathers historical data from relevant sources and cleans and transforms data into a proper format that a machine can read. Once the data is extracted, transformed, and loaded (ETL), data is summarized using data exploration, business intelligence, dashboard, and benchmark information.

Diagnostic analytics digs deeper into issues and finds in-depth root causes of a problem. It helps you understand why something happened in the past. Statistical techniques such as correlation and root cause, cause–effect analysis (Fig. 1.3), and graphic analytics visualize why the effect happened.

Predictive analytics helps businesses to forecast trends based on the current events. It predicts what is most likely to happen in the future and estimates time it will happen. Predictive analytics uses many techniques such as data mining, regression analysis, statistics, neural network, network analysis, predict modeling, Monte Carlo simulation, machine learning, etc.

Prescriptive analytics is the last and most sophisticated analytics that recommends what actions you can take to bring desired outcomes. It uses advanced tools such as decision tree, linear and nonlinear programming, deep learning, etc. to find optimal solutions and feedback to database for next analytics cycle.

Augmented analytics uses AI and machine learning to automate data preparation, discover insights, develop models and share insights among a broad range of business users. It is predicted that augmented analytics will be a dominant and destructive driver of data analytics.[5]

1.2.2.3 Artificial Intelligence

After years of fundamental research, AI is expanding and transforming every walk of life rapidly. AI has been used in IoT devices, autonomous driving, robot surgery, medical imaging and diagnosis, financial and economic modeling, weather forecasting, voice-activated digital assistance, and beyond. A well-designed AI application such as monitoring equipment failure and optimizing data center infrastructure operations and maintenance will save energy and avoid disasters.

John McCarthy, an assistant professor while at Dartmouth College, coined the term "artificial intelligence" in 1956. He defined AI as "getting a computer to do things which, when done by people, are said to involve intelligence." There is no unified definition at the time of this publication, but AI technologies consist of hardware and software and the "machines that respond to simulation consistent with traditional responses from humans, given the human capacity of contemplation, judgment and intention."[6] AI promises to drive from quality of life to the world economy. Applying both quantum computing, which stores information in 0's, 1's, or both called qubits, and parallel computing, which breaks a problem into discrete parts and solved many problems concurrently, AI can solve complicated problems faster and accurately in sophisticated ways and can conserve more energy in data centers.[7]

In data centers, AI could be used to monitor virtual machine operations and idle or running mode of servers, storages, and networking equipment to coordinate cooling loads and reduce power consumptions.

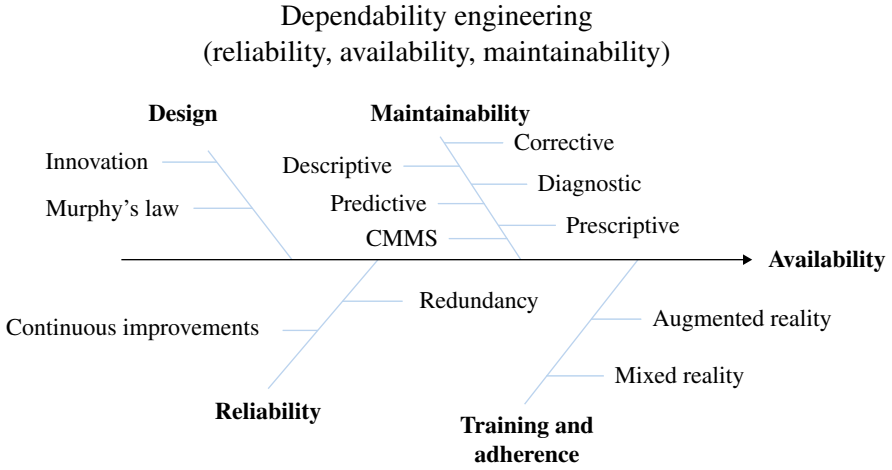


FIGURE 1.3 Cause and effect diagram. Source: © 2021 Amica Research.

[5] https://www.gartner.com/doc/reprints?id=1-1XOR8WDB&ct=191028&st=sb

[6] https://www.semanticscholar.org/paper/Applicability-of-Artificial-Intelligence-in-Fields-Shubhendu-Vijay/2480a71ef5e5a2b1f4a9217a0432c0c974c6c28c

[7] https://computing.llnl.gov/tutorials/parallel_comp/#Whatis

1.2.3 The Fifth-Generation Network

The 5G network, the fifth generation of wireless networks is changing the world and empowering how we live and work. 5G transmits median speed at 1.4 GB/s with reduced latency from 50 ms (1 ms = 0.001 s) to a few ms allowing little latency times for connected vehicles or remote surgery. There are wide spectra to provide 5G coverage. Using the high-frequency end of the spectrum, signals travel at extremely high speed, but the signals do not go as far nor through walls or obstruction. As a result, more wireless network equipment stations are required to be installed on streetlight or traffic poles. Using the lower-frequency end of the spectrum, signals travel farther but at a lower speed.

5G is one of the most important elements to power the IoT to drive smart manufacturing, smart transportation, smart healthcare, smart cities, smart entertainment, and smart everything. 5G can deliver incredibly detailed traffic, road, and hazard conditions to AV and power robotic surgery in real time. Through 5G, wearable glasses display patient's physical information and useful technical information to doctors in real time. 5G can send production instructions using wireless instead of wire at a faster speed that is critical to smart manufacturing. Virtual reality and augmented reality devices connect over 5G instead of wire that allows viewers to see the game from different angles in real time and superimpose player's statistics on the screen. By applying 5G, Airbus is piloting "Fello'fly or tandem flying" similar to migratory birds flying in a V formation to save energy.

1.3 DATA CENTER SYSTEM AND INFRASTRUCTURE ARCHITECTURE

The Oxford English dictionary defines architecture as "the art and study of designing buildings." The following are key components for architecture of a data center's system and infrastructure. They are discussed in detail in other chapters of this handbook.

- Mechanical system with sustainable cooling
- Electrical distribution and backup systems
- Rack and cabling systems
- Data center infrastructure management
- Disaster recovery and business continuity (DRBC)
- Software-defined data center
- Cloud and X-as-a-Service (X is a collective term referring to Platform, Infrastructure, AI, Software, DRBC, etc.)

1.4 STRATEGIC PLANNING

Strategic planning for data centers encompass a global location plan, site selection, design, construction, and operations. There is no one "correct way" to prepare a strategic plan. Depending on data center acquisition strategy (i.e., host, colocation, expand, lease, buy, build, or combination of above.), the level of deployments could vary from minor modifications of a server room to a complete build out of a green field project.

1.4.1 Strategic Planning Forces

The "Five Forces" described in Michael Porter's [3] "How Competitive Forces Shape Strategy" lead to a state of competition in all industries. The Five Forces are a threat of new entrants, bargaining power of customers, threat of substitute products or services, bargaining power of suppliers, and the industry jockeying for position among current competitors. Chinese strategist Sun Tzu, in the *Art of War*, stated five factors: the Moral Law, Weather, Terrain, the Commander, and Doctrine. Key ingredients in both strategic planning articulate the following:

- What are the goals
- What are fundamental factors
- What are knowns and unknowns
- What are constraints
- What are feasible solutions
- How the solutions are validated
- How to find an optimum solution

In preparing a strategic plan for a data center, Figure 1.4 shows four forces: business drivers, processes, technologies, and operations [4]. "Known" business drivers of a strategic plan include the following:

- Agility: Ability to move quickly.
- Resiliency: Ability to recover quickly from an equipment failure or natural disaster.
- Modularity and scalability: "Step and repeat" for fast and easy scaling of infrastructures.
- Reliability and availability: Ability of equipment to perform a given function and ability of an equipment to be in a state to perform a required function.
- Total cost of ownership (TCO): Total life cycle costs of CapEx (capital expenditures including land, building, design, construction, computer equipment, furniture and fixtures) and OpEx (operating expenditures including overhead, utility, maintenance, and repair costs).
- Sustainability: Apply best practices in green design, construction, and operations of data centers to reduce environmental impacts.

Additional "knowns" to each force could be expanded and added to tailor individual needs of a data center project. It is

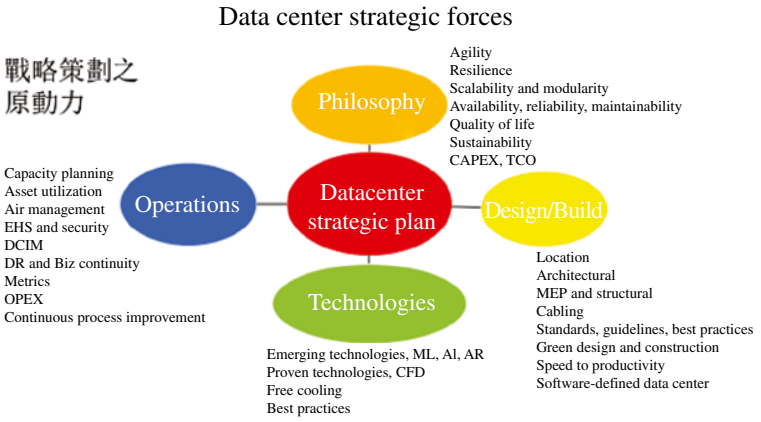


FIGURE 1.4 Data center strategic planning forces. Source: © 2021 Amica Research.

comprehensible that "known" business drivers are complicated and sometimes conflicting to each other. For example, increasing resiliency, or flexibility, of a data center will inevitably increase the costs of design and construction as well as continuous operating costs. The demand for sustainability will increase the TCO. "He can't eat his cake and have it too," so it is important to prioritize business drivers early on in the strategic planning process.

A strategic plan must anticipate the impacts of emerging technologies such as AI, blockchain, digital twin, and Generative Adversarial Networks, etc.

1.4.2 Capacity Planning

Gartner's study showed that data center facilities rarely meet the operational and capacity requirements of their initial design [5]. Microsoft's top 10 business practices estimated [6] that if a 12 Megawatt data center uses only 50% of power capacity, then every year $4–8 million in unused capital is stranded in uninterruptible power supply (UPS), generators, chillers, and other capital equipment. It is imperative to focus on capacity planning and resource utilization.

1.4.3 Strategic Location Plan

To establish data center location plan, business drivers include expanding market, emerging market, undersea fiber-optic cable, Internet exchange points, electrical power, capital investments, and many other factors. It is indispensable to have a strategic location roadmap on where to build data centers around the globe. Once the roadmap is established, a short-term data center design and implementation plan could follow. The strategic location plan starts from considering continents, countries, states, and cities down to a data center campus site. Considerations at continent and country or at macro level include:

- Political and economic stability of the country
- Impacts from political economic pacts (G20, G8+5, OPEC, APEC, RCEP, CPTPP, FTA, etc.)
- Gross domestic products or relevant indicators
- Productivity and competitiveness of the country
- Market demand and trend

Considerations at state (province) or at medium level include:

- Natural hazards (earthquake, tsunami, hurricane, tornado, volcano, etc.)
- Electricity sources with dual or multiple electrical grid services
- Electricity rate
- Fiber-optic infrastructure with multiple connectivity
- Public utilities (natural gas, water)
- Airport approaching corridor
- Labor markets (educated workforce, unemployment rate, etc.)

Considerations at city campus or at micro level include:

- Site size, shape, accessibility, expandability, zoning, and code controls
- Tax incentives from city and state
- Topography, water table, and 100-year floodplain
- Quality of life for employee retention
- Security and crime rate
- Proximity to airport and rail lines
- Proximity to chemical plant and refinery
- Proximity to electromagnetic field from high-voltage power lines
- Operational considerations

Other useful tools to formulate location plans include:

- Operations research
  - Network design and optimization
  - Regression analysis on market forecasting
- Lease vs. buy analysis or build and leaseback
- Net present value
- Break-even analysis
- Sensitivity analysis and decision tree

As a cross-check, compare your global location plan against data centers deployed by technology companies such as Amazon, Facebook, Google, Microsoft, and other international tech companies.

1.5 DESIGN AND CONSTRUCTION CONSIDERATIONS

A data center design encompasses architectural (rack layout), structural, mechanical, electrical, fire protection, and cabling system. Sustainable design is essential because a data center can consume 40–100 times more electricity compared to a similar-size office space. In this section, applicable design guidelines and considerations are discussed.

1.5.1 Design Guidelines

Since a data center involves 82–85% of initial capital investment in mechanical and electrical equipment [7], data center project is generally considered as an engineer-led project. Areas to consider for sustainable design include site selection, architectural/engineering design, energy efficiency best practices, redundancy, phased deployment, etc. There are many best practices covering site selection and building design in the Leadership in Energy and Environmental Design (LEED) program. The LEED program is a voluntary certification program that was developed by the U.S. Green Building Council (USGBC).[8]

Early on in the architecture design process, properly designed column spacing and floor elevation will ensure appropriate capital investments and minimize operating expenses. A floor plan with appropriate column spacing maximizes ICT rack installations and achieves power density with efficient cooling distribution. A floor-to-floor elevation must be carefully planned to include height and space for mechanical, electrical, structural, lighting, fire protection, and cabling system.

International technical societies have developed many useful design guidelines that are addressed in detail in other chapters of this handbook:

- ASHRAE TC9.9: Data Center Networking Equipment [8]
- ASHRAE TC9.9: Data Center Power Equipment Thermal Guidelines and Best Practice
- ASHRAE 90.1: Energy Standard for Buildings [9]
- ASHRAE: Gaseous and Particulate Contamination Guidelines for Data Centers [10]
- Best Practices Guide for Energy-Efficient Data Center Design [11]
- EU Code of Conduct on Data Centre Energy Efficiency [12]
- BICSI 002: Data Center Design and Implementation Best Practices [13]
- FEMA P-414: "Installing Seismic Restraints for Duct and Pipe" [14]
- FEMA 413: "Installing Seismic Restraints for Electrical Equipment" [15]
- FEMA, SCE, VISCMA, "Installing Seismic Restraints for Mechanical Equipment" [16]
- GB 50174: Code for Design of Data Centers [17]
- ISO 50001: Energy Management Specification and Certification
- LEED Rating Systems [18]
- Outline of Data Center Facility Standard by Japan Data Center Council (JDCC) [19]
- TIA-942: Telecommunications Infrastructure Standard for Data Centers

Chinese standard GB 50174 "Code for Design of Data Centers" provides a holistic approach of designing data centers that cover site selection and equipment layout, environmental requirements, building and structure, air conditioning (mechanical system), electrical system, electromagnetic shielding, network and cabling system, intelligent system, water supply and drainage, and fire protection and safety [17].

1.5.2 Reliability and Redundancy

"Redundancy" ensures higher reliability, but it has profound impacts on initial investments and ongoing operating costs (Fig. 1.3).

In 2011, with fierce competition against Airbus SE, Boeing Company opted to update its single-aisle 737 rather than design a new jet that is equipped with new fuel-efficient engines. The larger engines were placed farther forward on the wing that, in certain condition, caused the plane nose to pitch up too quickly. The solution to the problem was to use MCAS (Maneuvering Characteristics Augmentation System) that is a stall prevention system. For the 737 Max, a single set of "angle-of-attack" sensors was used to determine if automatic flight control commands should be triggered when the MCAS is fed sensor data. If a second set of sensors

[8] http://www.usgbc.org/leed/rating-systems

and software or redundancy design on angle of attack had been put in place, two plane crashes, which killed 346 people 5 months apart, could have been avoided [20, 21].

Uptime Institute® pioneered a tier certification program that structured data center redundancy and fault tolerance in four tiers [22]. Telecommunication Industry Association's TIA-942 contains four tables that describe building and infrastructure redundancy in four levels. Basically, different redundancies are defined as follows:

- N: Base requirement.
- $N+1$ redundancy: Provides one additional unit, module, path, or system to the minimum requirement
- $N+2$ redundancy: Provides two additional units, modules, paths, or systems in addition to the minimum requirement
- $2N$ redundancy: Provides two complete units, modules, paths, or systems for every one required for a base system
- $2(N+1)$ redundancy: Provides two complete $(N+1)$ units, modules, paths, or systems

Accordingly, a matrix table is established using the following tier levels in relation to component redundancy:

Tier I Data Center: Basic system
Tier II Data Center: Redundant components
Tier III Data Center: Concurrently maintainable
Tier V Data Center: Fault tolerant

The China National Standard GB 50174 "Code for Design of Data Centers" defines A, B, and C tier levels with A being the most stringent.

JDCC's "Outline of Data Center Facility Standard" tabulates "Building, Security, Electric Equipment, Air Condition Equipment, Communication Equipment and Equipment Management" in relation to redundancy Tiers 1, 2, 3, and 4. It is worthwhile to note that the table also includes seismic design considerations with probable maximum loss (PML) relating to design redundancy.

Data center owners should consult and establish a balance between desired reliability, redundancy, PML, and additional costs.[9]

1.5.3 Computational Fluid Dynamics

Whereas data centers could be designed by applying best practices, the locations of systems (rack, CRAC, etc.) might not be in its optimal arrangement collectively. Computational fluid dynamics (CFD) technology has been used in semiconductor's cleanroom projects for decades to ensure uniform airflow inside a cleanroom. During the initial building and rack layout design stage, CFD offers a scientific analysis and solution to visualize airflow patterns and hot spots and validate cooling capacity, rack layout, and location of cooling units. One can visualize airflow in hot and cold aisles for optimizing room design. During the operating stage, CFD could be used to emulate and manage airflow to ensure the air path does not recirculate, bypass, or create negative pressure flow.

1.5.4 Best Practices

Although designing energy-efficient data center is still evolving, many best practices could be applied whether you are designing a small server room or a large data center. One of the best practices is to build or use ENERGY STAR servers [23] and solid-state drives. The European Commission published a comprehensive "Best Practices for the EU Code of Conduct on Data Centres." The U.S. Department of Energy's Federal Energy Management Program published "Best Practices Guide for Energy-Efficient Data Center Design." Both, and many other publications, could be referred to when preparing a data center design specification. Here is a short list of best practices and emerging technologies:

- In-rack-level liquid cooling and liquid immersion cooling
- Increase server inlet temperature and humidity adjustments (ASHRAE Spec) [24]
- Hot and cold aisle configuration and containment
- Air management (to stop bypass, hot and cold air mixing, and recirculation)
- Free cooling using air-side economizer or water-side economizer
- High efficient UPS
- Variable speed drives
- Rack-level direct liquid cooling
- Fuel cell technology
- Combined heat and power (CHP) in data centers [22]
- Direct current power distribution
- AI and data analytics applications in operations control.

It is worthwhile to note that servers can operate outside the humidity and temperature ranges recommended by ASHRAE [25].

1.6 OPERATIONS TECHNOLOGY AND MANAGEMENT

Best practices in operations technology (OT) and management include benchmark metrics, data center infrastructure management, air management, cable management,

[9] www.AmicaResearch.org

preventive and predictive maintenance, 5S, disaster management, and workforce development, etc. This section will discuss some of OTs.

1.6.1 Metrics for Sustainable Data Centers

Professors Robert Kaplan and David Norton once said that "if you can't measure it, you can't manage it." Metrics, as defined in Oxford dictionary, are "A set of figures or statistics that measure results."

Data centers require well-defined metrics to make accurate measurements and act on less efficient areas with corrective actions. Power usage effectiveness (PUE), developed by the Green Grid, is a ratio of total electrical power entering a data center to the power used by IT equipment. It is a widely accepted KPI (key performance indicator) in the data center industry. Water usage effectiveness is another KPI. Accurate and real-time data dashboard information on capacity versus usage regarding space, power, and cooling provide critical benchmark information. Other information such as cabinet temperature, humidity, hot spot location, occurrence, and duration should be tracked to monitor operational efficiency and effectiveness.[10]

1.6.2 DCIM and Digital Twins

DCIM (data center infrastructure management) consists of many useful modules to plan, manage, and automate a data center. Asset management module tracks asset inventory, space/power/cooling capacity and change process, available power and data ports, bill back reports, etc. Energy management module allows integrating information from building management systems (BMS), utility meters, UPS, etc., resulting in actionable reports. Using DCIM in conjunction with CFD, data center operators could effectively optimize energy consumption.[11] A real-time dashboard allows continuous monitoring of energy consumption so as to take necessary actions. Considering data collecting points for DCIM with required connectors early on in the design stage is crucial to avoid costly installation later on.

A digital twin (DT), a 3D virtual model of a data center, replicates physical infrastructure and IT equipment from initial design to information collected from daily operations. DT tracks equipment's historical information and enables descriptive to predictive analytics.

1.6.3 Cable Management

Cabling system is a little thing but makes big impacts, and it is long lasting, costly, and difficult to replace [26, 27]. It should be planned, structured, and installed per network topology and cable distribution requirements as specified in TIA-942 and ANSI/TIA/EIA-568 standards. The cable shall be organized so that the connections are traceable for code compliance and other regulatory requirements. Poor cable management [28] could create electromagnetic interference (EMI) due to induction between cable and equipment electrical cables. To improve maintenance and serviceability, cabling should be placed in such a way that it could be disconnected to reach a piece of equipment for adjustments or changes. Pulling, stretching, or bending radius of cables beyond specified ranges should be avoided.

1.6.4 The 6S Pillars

The 6S [29], which uses 5S pillars and adds one pillar for safety, is one of best lean methods commonly implemented in the manufacturing industry. It optimizes productivity by maintaining an orderly and safe workplace. 6S is a cyclical and continuing methodology that includes the following:

- Sort: Eliminate unnecessary items from the workplace.
- Set in order: Create a workplace so that items are easy to find and put away.
- Shine: Thoroughly clean the work area.
- Standardize: Create a consistent approach which tasks and procedures are done.
- Sustain: Make a habit to maintain the procedure.
- Safety: Make accidents less likely in an orderly and shining workplace.

Applying 6S pillars to ensure cable management discipline will avoid out of control that leads to chaos in data centers. While exercising "Sort" to clean closets that are full of decommissioned storage drives, duty of care must be taken to ensure "standardized" policy and procedure are followed to avoid mistakes.

1.7 BUSINESS CONTINUITY AND DISASTER RECOVERY

In addition to natural disasters, terrorist attacks on the Internet's physical infrastructure are vulnerable and could be devastating. Statistics show that over 70% of all data centers were brought down by human errors such as improper executing procedures during maintenance. It is imperative to have detailed business continuity (BC) and disaster recovery (DR) plans that are well prepared and executed. To sustain data center buildings, BC should consider a design beyond requirements pursuant to building codes and standards. The International Building Code (IBC) and other codes generally concern life safety of

[10] https://www.sunbirddcim.com/blog/top-10-data-center-kpis
[11] http://www.raritandcim.com/

occupants with little regard to property or functional losses. Consequently, seismic strengthening design of data center building structural and nonstructural components (see Section 1.5.1) must be exercised beyond codes and standards requirements [30].

Many lessons were learned on DR from natural disasters: Los Angeles Northridge earthquake (1994), Kobe earthquake (1995), New Orleans' Hurricane Katrina (2005), Great East Japan earthquake and tsunami (2011) [31], the Eastern U.S. Superstorm Sandy (2012), and Florida's Hurricane Irma (2017) [32]. Consider what we can learn in a worst scenario with the 2020 pandemic (COVID-19) and a natural disaster happening at the same time (see Section 37.6.2).

Key lessons learned from the above natural disasters are highlighted:

- Detailed crisis management procedure and communication command line.
- Conduct drills regularly by the emergency response team using DR procedures.
- Regularly maintain and test run standby generators and critical infrastructure in a data center.
- Have enough supplies, nonperishable food, drinking water, sleeping bags, batteries, and a safe place to do their work throughout a devastating event as well as preparedness for their family.
- Fortify company properties and rooftop HVAC (heating, ventilation and air conditioning) equipment.
- Have contracts with multiple diesel oil suppliers to ensure diesel fuel deliveries.
- Use cellular phone and jam radio and have different communication mechanisms such as social networking websites.
- Get needed equipment on-site readily accessible (flashlight, backup generators, fuel, containers, hoses, extension cords, etc.).
- Brace for the worst—preplan with your customers on communication during disaster and a controlled shutdown and DR plan.

Other lessons learned include using combined diesel fuel and natural gas generator generators, fuel cell technology, and submersed fuel pump and that "a cloud computing-like environment can be very useful." Watch out for "Too many risk response manuals will serve as a 'tranquilizer' for the organization. Instead, implement a risk management framework that can serve you well in preparing and responding to a disaster."

Finally not the least, cloud is one of the most effective plans a company is able to secure its data and operations at all times [33].

1.8 WORKFORCE DEVELOPMENT AND CERTIFICATION

The traditional Henry Ford-style workforce desired a secure job, works 40-h workweek, owns a home, raises a family, and lives in peace. Rising Gen Z and the modern workforce is very different and demanding: work to be fulfilling, work any time any place, a sense of belonging, having rewarding work, and making work fun. Workforce development plays a vital role not only in retaining talents but also in having well-trained practitioners to operate data centers.

There are numerous commercial training and certification programs available. Developed by the U.S. Department of Energy, Data Center Energy Practitioner (DCEP) Program [34] offers data center practitioners with different certification programs. The U.S. Federal Energy Management Program is accredited by the International Association for Continuing Education and Training and offers free online training [35]. "Data center owners can use Data Center Energy Profiler (DC Pro) Software [36] to learn, profile, evaluate, and identify potential areas for energy efficiency improvements.

1.9 GLOBAL WARMING AND SUSTAINABILITY

Since 1880, a systematic record keeping began, and an average global surface temperature has risen about 2°F (1°C) according to scientists at the U.S. National Aeronautics and Space Administration (NASA). Separate studies are conducted by NASA, U.S. National Oceanic and Atmospheric Administration (NOAA), and European Union's Copernicus Climate Change Service, ranking 2019 the second warmest year in the decade, and the trend continued since 2017. In 2019 the average global temperature was 1.8°F (0.98°C) above the twentieth-century average (1901–2000).

In 2018, IPCC prepared a special report titled "Global Warming of 1.5°C" that states: "A number of climate change impacts that could be avoided by limiting global warming to 1.5°C compared to 2°C, or more. For instance, by 2100, global sea level rise would be 10 cm lower with global warming of 1.5°C compared with 2°C. The likelihood of an Arctic Ocean free of sea ice in summer would be once per century with global warming of 1.5°C, compared with at least once per decade with 2°C. "Every extra bit of warming matters, especially since warming of 1.5°C or higher increases the risk associated with long-lasting or irreversible changes, such as the loss of some ecosystems," said Hans-Otto Pörtner, co-chair of IPCC Working Group II. The report also examines pathways available to limit warming to 1.5°C, what it would take to achieve them, and what the consequences could be [37]. Global warming results in dry regions becoming dryer, wet region wetter, more frequent hot days and wildfires, and fewer cool days.

Humans produce all kinds of heat—from cooking food, manufacturing goods, building houses, and moving people or goods—to perform essential activities that are orchestrated by information and communication equipment (ICE) in hyperscale data centers. ICE acts as a pervasive force in global economy that includes Internet searching, online merchant, online banking, mobile phone, social networking, medical services, and computing in exascale (10^{18}) supercomputers. It will quickly analyze big data and realistically simulate complex processes and relationships such as fundamental forces of the universe.[12]

All above activities draw power and release heat in and out of data centers. One watt (W) power drawn to process data generates 1 W of heat output to the environment. Modern lifestyle will demand more energy that gives out heat, but effectively and vigilantly designing and managing a data center can reduce heat output and spare the Earth.

1.10 CONCLUSIONS

The focal points of this chapter center on how to design and operate highly available, fortified, and energy-efficient mission critical data centers with convergence of operations and information technologies. More data centers for data processing and analysis around the world have accelerated energy usages that contribute to global warming. The world has seen weather anomalies with more flood, drought, wild fire, and other catastrophes including food shortage. To design a green data center, strategic planning by applying essential drivers was introduced. Lessons learned from natural disasters and pandemic were addressed. Workforce development plays a vital role in successful application of OT.

There are more emerging technologies and applications that are driven by the IoT. International digital currency and blockchain in various applications are foreseeable. More data and analytics will be performed in the edge and fog as well as in the cloud. All these applications lead to more data centers demanding more energy that create global warming.

Dr. Albert Einstein once said, "Creativity is seeing what everyone else sees and thinking what no-one else has thought." There are tremendous opportunities for data center practitioners to apply creativity (Fig. 1.5) and accelerate the pace of invention and innovation in future data centers. By collective effort, we can apply best practices to accelerate speed of innovation to plan, design, build, and operate data centers efficiently and sustainably.

[12] https://www.exascaleproject.org/what-is-exascale/

FIGURE 1.5 Nurture creativity for invention and innovation. Source: Courtesy of Amica Research.

REFERENCES

[1] OECD science, technology, and innovation Outlook 2018. Available at http://www.oecd.org/sti/oecd-science-technology-and-innovation-outlook-25186167.htm. Accessed on March 30, 2019.

[2] Shehabi A, et al. 2016. United States Data Center Energy Usage Report, Lawrence Berkeley National Laboratory, LBNL-1005775, June 2016. Available at https://eta.lbl.gov/publications/united-states-data-center-energy. Accessed on April 1, 2019.

[3] Porter M. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press, Harvard University; 1980.

[4] Geng H. Data centers plan, design, construction and operations. Datacenter Dynamics Conference, Shanghai; September 2013.

[5] Bell MA. *Use Best Practices to Design Data Center Facilities*. Gartner Publication; April 22, 2005.

[6] Microsoft's top 10 business practices for environmentally sustainable data centers. Microsoft. Available at http://environment-ecology.com/environmentnews/122-microsofts-top-10-business-practices-for-environmentally-sustainable-data-centers-.html. Accessed on February 17, 2020.

[7] Belady C, Balakrishnan G. 2008. Incenting the right behaviors in the data center. Avaiable at https://www.uschamber.com/sites/default/files/ctec_datacenterrpt_lowres.pdf. Accessed on February 22, 2020.

[8] Data center networking equipment-issues and best practices. ASHRAE. Available at https://tc0909.ashraetcs.org/documents/ASHRAE%20Networking%20Thermal%20Guidelines.pdf. Accessed on September 3, 2020.

[9] ANSI/ASHRAE/IES Standard 90.1-2019 -- Energy Standard for Buildings Except Low-Rise Residential Buildings, https://www.ashrae.org/technical-resources/bookstore/standard-90-1. Accessed on September 3, 2020.

[10] 2011 Gaseous and particulate contamination guidelines for data centers. ASHRAE. Available at https://www.ashrae.org/

File%20Library/Technical%20Resources/Publication%20 Errata%20and%20Updates/2011-Gaseous-and-Particulate-Guidelines.pdf. Accessed on September 3, 2020.

[11] Best practices guide for energy-efficient data center design. Federal Energy Management Program. Available at https://www.energy.gov/eere/femp/downloads/best-practices-guide-energy-efficient-data-center-design. Accessed September 3,2020.

[12] 2020 Best Practice Guidelines for the EU Code of Conduct on Data Centre Energy Efficiency. Available at https://e3p.jrc.ec.europa.eu/publications/2020-best-practice-guidelines-eu-code-conduct-data-centre-energy-efficiency. Accessed on September 3,2020.

[13] BICSI data center design and implementation best practices. Available at https://www.bicsi.org/standards/available-standards-store/single-purchase/ansi-bicsi-002-2019-data-center-design. Accessed on February 22, 2020.

[14] Installing Seismic restraints for duct and pipe. FEMA P414; January 2004. Available at https://www.fema.gov/media-library-data/20130726-1445-20490-3498/fema_p_414_web.pdf. Accessed on February 22, 2020.

[15] FEMA. Installing Seismic restraints for electrical equipment. FEMA; January 2004. Available at https://www.fema.gov/media-library-data/20130726-1444-20490-4230/FEMA-413.pdf. Accessed on February 22, 2020.

[16] Installing Seismic restraints for mechanical equipment. FEMA, Society of Civil Engineers, and the Vibration Isolation and Seismic Control Manufacturers Association. Available at https://kineticsnoise.com/seismic/pdf/412.pdf. Accessed on February 22, 2020.

[17] China National Standards, Code for design of data centers: table of contents section. Available at www.AmicaResearch.org. Accessed on February 22, 2020.

[18] Rasmussen N, Torell W. Data center projects: establishing a floor plan. APC White Paper #144; 2007. Available at https://apcdistributors.com/white-papers/Architecture/WP-144%20Data%20Center%20Projects%20-%20Establishing%20a%20Floor%20Plan.pdf. Accessed September 3, 2020.

[19] Outline of data center facility standard. Japan Data Council. Available at https://www.jdcc.or.jp/english/files/facility-standard-by-jdcc.pdf. Accessed on February 22, 2020.

[20] Sider A, Tangel A. Boeing omitted MAX safeguards. The Wall Street Journal, September 30, 2019.

[21] Sherman M, Wall R. Four fixed needed before the 737 MAX is back in the air. The Wall Street Journal, August 20, 2019.

[22] Darrow K, Hedman B. *Opportunities for Combined Heat and Power in Data Centers. Arlington:* ICF International, Oak Ridge National Laboratory; 2009. Available at https://www.energy.gov/sites/prod/files/2013/11/f4/chp_data_centers.pdf. Accessed on February 22, 2020.

[23] EPA Energy Efficient Products. Available at https://www.energystar.gov/products/spec/enterprise_servers_specification_version_3_0_pd. Accessed on May 12, 2020.

[24] Server inlet temperature and humidity adjustments. Available at http://www.energystar.gov/index.cfm?c=power_mgt.datacenter_efficiency_inlet_temp. Accessed February 22, 2020.

[25] Server inlet temperature and humidity adjustments. Available at https://www.energystar.gov/products/low_carbon_it_campaign/12_ways_save_energy_data_center/server_inlet_temperature_humidity_adjustments. Accessed on February 28, 2020.

[26] 7 Best practices for simplifying data center cable management with DCIM software. Available at https://www.sunbirddcim.com/blog/7-best-practices-simplifying-data-center-cable-management-dcim-software. Accessed on February 22, 2020.

[27] Best Practices Guides: Cabling the Data Center. Brocade; 2007.

[28] Apply proper cable management in IT Racks—a guide for planning, deployment and growth. Emerson Network Power; 2012.

[29] Lean and environment training modules. Available at https://www.epa.gov/sites/production/files/2015-06/documents/module_5_6s.pdf. Accessed on February 22, 2020.

[30] Braguet OS, Duggan DC. Eliminating the confusion from seismic codes and standards plus design and installation instruction. 2019 BICSI Fall Conference, 2019. Available at https://www.bicsi.org/uploadedfiles/PDFs/conference/2019/fall/PRECON_3C.pdf. Accessed September 3, 2020.

[31] Yamanaka A, Kishimoto Z. The realities of disaster recovery: how the Japan Data Center Council is successfully operating in the aftermath of the earthquake. JDCC, Alta Terra Research; June 2011.

[32] Hurricane Irma: a case study in readiness, CoreSite. Available at https://www.coresite.com/blog/hurricane-irma-a-case-study-in-readiness. Accessed on February 22, 2020.

[33] Kajimoto M. One year later: lessons learned from the Japanese tsunami. ISACA; March 2012.

[34] Data Center Energy Practitioner (DCEP) Program. Available at https://datacenters.lbl.gov/dcep. Accessed on February 22, 2020.

[35] Federal Energy Management Program. Available at https://www.energy.gov/eere/femp/federal-energy-management-program-training. Accessed on February 22, 2020.

[36] Data center profiler tools. Available at https://datacenters.lbl.gov/dcpro. Accessed on February 22, 2020.

[37] IPCC. Global warming of 1.5°C. WMO, UNEP; October 2018. Available at http://report.ipcc.ch/sr15/pdf/sr15_spm_final.pdf. Accessed on November 10, 2018.

FURTHER READING

Huang, R., et al. Data Center IT efficiency Measures Evaluation Protocol, 2017, the National Renewable Energy Laboratory, US Dept. of Energy.

Koomey J. *Growth in Data Center Electricity Use 2005 to 2010*. Analytics Press; August 2011.

Planning guide: getting started with big data. Intel; 2013.

Voas J, Networks of 'Things', NIST Special Publication SP 800-183, July 2016.

Turn Down the Heat: Why a 4°C Warmer World Must Be Avoid. Washington, DC: The World Bank; November 18, 2012.

2

GLOBAL DATA CENTER ENERGY DEMAND AND STRATEGIES TO CONSERVE ENERGY

Nuoa Lei and Eric R. Masanet
Northwestern University, Evanston, Illinois, United States of America

2.1 INTRODUCTION

2.1.1 Importance of Data Center Energy Use

Growth in global digitalization has led to a proliferation of digital services touching nearly every aspect of modern life. Data centers provide the digital backbone of our increasingly interconnected world, and demand for the data processing, storage, and communication services that data centers provide is increasing rapidly. Emerging data-intensive applications such as artificial intelligence, the Internet of Things, and digital manufacturing—to name but a few—promise to accelerate the rate of demand growth even further. Because data centers are highly energy-intensive enterprises, there is rising concern regarding the global energy use implications of this ever-increasing demand for data. Therefore, understanding, monitoring, and managing data center energy use have become a key sustainability concern in the twenty-first century.

2.1.2 Data Center Service Demand Trends

While demand for data center services can be quantified in myriad ways, from a practical perspective, analysts must rely on macro-level indicators that capture broad industry trends at regional and national levels and that can be derived from statistics that are compiled on a consistent basis. From such indicators, it is possible to get a directional view of where demand for data center services has been and where it may be headed in the near term.

The most common macro-level indicator is annual global data center IP traffic, expressed in units of zettabytes per year (ZB/year), which is estimated by network systems company Cisco. According to Cisco [1, 2],

- Annual global data center IP traffic will reach 20.6 ZB/year by the end of 2021, up from 6.8 ZB/year in 2016 and from only 1.1 ZB/year in 2010. These projections imply that data center IP traffic will grow at a compound annual growth rate (CAGR) of 25% from 2016 to 2021, which is a CAGR much faster than societal demand in other rapidly growing sectors of the energy system. For example, demand for aviation (expressed as passenger-kilometers) and freight (expressed as ton-kilometers) rose by 6.1 and 4.6% in 2018 [3], respectively.
- Big data, defined as data deployed in a distributed processing and storage environment, is a key driver of overall data center traffic. By 2021, big data will account for 20% of all traffic within the data center, up from 12% in 2016.

While historically the relationship between data center energy use and IP traffic has been highly elastic due to substantial efficiency gains in data center technologies and operations [4], Cisco's IP traffic projections indicate that global demand for data services will continue to grow rapidly.

The number of global server workloads and compute instances provides another indicator of data center service demand. Cisco defines a server workload and compute instance as "a set of virtual or physical computer resources that is assigned to run a specific application or provide

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

computing services for one or many users" [2]. As such, this number provides a basic means of monitoring demand for data center computational services. According to Cisco [1, 2],

- The number of global server workloads and compute instances has increased from 57.5 million in 2010 to 371.7 million in 2018, a sixfold increase in only 8 years. This number is projected to grow to 566.6 million by 2021, at a CAGR of 15%.
- The nature of global server workloads and compute instances is changing rapidly. In 2010, 79% were processed in traditional data centers, whereas in 2018, 89% were processed in cloud- and hyperscale-class data centers. Furthermore, by 2021, only 6% will be processed in traditional data centers, signaling the terminus of a massive shift in global data center market structure.

Both the increase in overall demand for workloads and compute instances and the shift away from traditional data centers have implications for data center energy use. The former drives demand for energy use by servers, storage, and network communication devices, whereas the latter has profound implications for overall energy efficiency, given that cloud data centers are generally managed with greater energy efficiency than smaller traditional data centers.

Lastly, given growing demand for data storage, total storage capacity in global data centers has recently emerged as another macro-level proxy for data center service demand. According to estimates by storage company Seagate and market analysis firm IDC, in 2018, around 20 ZB of data were stored in enterprise and cloud data center environments, and this number will rise to around 150 ZB (a 7.5× increase) by 2025 [5]. Similarly, Cisco has estimated a 31% CAGR in global data center installed storage capacity through 2021 [2].

Therefore, it is clear that demand for data center services expressed as data center IP traffic, server workloads and compute instances, and storage capacity is rising rapidly and will continue to grow in the near term. Understanding the relationship between service demand growth captured by these macro-level indicators and overall energy use growth requires models of data center energy use, which are discussed in the next section.

2.2 APPROACHES FOR MODELING DATA CENTER ENERGY USE

Historically, two primary methods have been used for modeling data center energy use at the global level: (i) bottom-up methods and (ii) extrapolation-based methods based on macro-level indicators. Bottom-up methods are generally considered the most robust and accurate, because they are based on detailed accounting of installed IT device equipment stocks and their operational and energy use characteristics in different data center types.

However, bottom-up methods are data intensive and can often be costly due to reliance on nonpublic market intelligence data. As a result, bottom-up studies have been conducted only sporadically. In contrast, extrapolation-based methods are much simpler but are also subject to significant modeling uncertainties. Furthermore, extrapolations typically rely on bottom-up estimates as a baseline and are therefore not truly an independent analysis method. Each approach is discussed further in the sections that follow.

2.2.1 The Bottom-Up Approach

In the bottom-up method [4, 6–9], the model used to estimate data center energy use is typically an additive model including the energy use of servers, external storage devices, network devices, and infrastructure equipment, which can be described using a general form as:

$$E^{\mathrm{DC}} = \sum_j \left(\sum_i E_{ij}^{\mathrm{server}} + \sum_i \mathrm{E}_{ij}^{\mathrm{storage}} + \sum_i \mathrm{E}_{ij}^{\mathrm{network}} \right) \mathrm{PUE}_j \quad (2.1)$$

where

E^{DC} = data center electricity demand (kWh/year),
E_{ij}^{server} = electricity used by servers of class i in space type j (kWh/year),
$E_{ij}^{\mathrm{storage}}$ = electricity used by external storage devices of class i in space type j (kWh/year),
$E_{ij}^{\mathrm{network}}$ = electricity used by network devices of class i in space type j (kWh/year),
PUE_j = power usage effectiveness of data center in space type j (kWh/kWh).

As expressed by the equation (2.1), the total electricity use of IT devices within a given space type is calculated through the summation of the electricity used by servers, external storage devices, and network devices. The total electricity use of IT devices is then multiplied by the power usage effectiveness (PUE) of that specific space type to arrive at total data center electricity demand. The PUE, defined as the ratio of total data center energy use to total IT device energy use, is a widely used metric to quantify the electricity used by data center infrastructure systems, which include cooling, lighting, and power provisioning systems.

In a bottom-up model, careful selection of IT device categories and data center space types is needed for robust and accurate data center energy use estimates. Some typical selections are summarized in Table 2.1.

TABLE 2.1 Typical data center space type and IT devices categories

	Traditional data center				Cloud data center	
Data center space type (typical floor area)	Server closet (<100 ft²)	Server room (100 – 1,000 ft²)	Localized data center (500 – 2,000 ft²)	Mid-tier data center (2,000 – 20,000 ft²)	High-end data center (20,000 – 40,000 ft²)	Hyperscale data center (>40,000 ft²)
Server class (average sales value)	Volume server (<$25,000)		Midrange server ($25,000 – 250,000)		High-end server (>$250,000)	
Storage device types	Hard disk drive		Solid-state drive		Archival tape drives	
Network switch port speed	100 Mbps	1,000 Mbps	10 Gbps	≥40 Gbps		

Source: [7] and [9].

2.2.2 Extrapolation-Based Approaches

In the extrapolation-based method, models typically utilize a base-year value of global data center energy use derived from previous bottom-up studies. This base-year value is then extrapolated, either using a projected annual growth rate [10, 11] (Equation 2.2) or, when normalized to a unit of service (typically IP traffic), on the basis of a service demand indicator [12, 13] (Equation 2.3). Extrapolation-based methods have been applied to estimate both historical and future energy use:

$$E_{i+n}^{\mathrm{DC}} = E_i^{\mathrm{DC}} \times \left(1 + \mathrm{CAGR}\right)^n \tag{2.2}$$

$$E_{i+n}^{\mathrm{DC}} = E_i^{\mathrm{DC}} \times \left(1 + \mathrm{GR_{IP}}\right)^n \times \left(1 - \mathrm{GR_{eff}}\right)^n \tag{2.3}$$

where

E_i^{DC} = data center electricity demand in baseline year i (kWh/year),
E_{i+n}^{DC} = data center electricity demand n years from baseline year (kWh/year),
CAGR = compound annual growth rate of data center energy demand,
$\mathrm{GR_{IP}}$=annual growth rate of global data center IP traffic,
$\mathrm{GR_{eff}}$ = efficiency growth factor.

Extrapolation-based methods are simpler and rely on far fewer data than bottom-up methods. However, they can only capture high-level relationships between service demand and energy use over time and are prone to large uncertainties given their reliance on a few key parameters (see Section 2.4). Because they lack the technology-richness of bottom-up approaches, extrapolation-based methods also lack the explanatory power of relating changes in energy use to changes in various underlying technological, operations, and data center market factors over time. This limited explanatory power also reduces the utility of extrapolation-based methods for data center energy policy design [4].

2.3 GLOBAL DATA CENTER ENERGY USE: PAST AND PRESENT

Studies employing bottom-up methods have produced several estimates of global data center energy use in the past two decades [4, 8, 14]. When taken together, these estimates shed light on the overall scale of global data center energy use, its evolution over time, and its key technological, operations, and structural drivers.

The first published global bottom-up study appeared in 2008 [14]. It focused on the period 2000–2005, which coincided with a rapid growth period in the history of the Internet. Over these 5 years, the worldwide energy use of data centers was estimated to have doubled from 70.8 to 152.5 TWh/year, with the latter value representing 1% of global electricity consumption. A subsequent bottom-up study [8], appearing in 2011, estimated that growth in global data center electricity use slowed from 2005 to 2010 due to steady technological and operational efficiency gains over the same period. According to this study, global data center energy use rose to between 203 and 272 TWh/year by 2010, representing a 30–80% increase compared with 2005.

The latest global bottom-up estimates [4] produced a revised, lower 2010 estimate of 194 TWh/year, with only modest growth to around 205 TWh/year in 2018, or around 1% of global electricity consumption. The 2010–2018

flattening of global data center energy use has been attributed to substantial efficiency improvements in servers, storage devices, and network switches and shifts away from traditional data centers toward cloud- and hyperscale-class data centers with higher levels of server virtualization and lower PUEs [4].

In the following section, the composition of global data center energy use—which has been illuminated by the technology richness of the bottom-up literature—is discussed in more detail.

2.3.1 Global Data Center Energy Use Characteristics

Figure 2.1 compiles bottom-up estimates for 2000, 2005, 2010, and 2018 to highlight how energy is used in global data centers, how energy use is distributed across major data center types and geographical regions, and how these characteristics have changed over time.

Between 2000 and 2005, the energy use of global data centers more than doubled, and this growth was mainly attributable to the increased electricity use of a rapidly

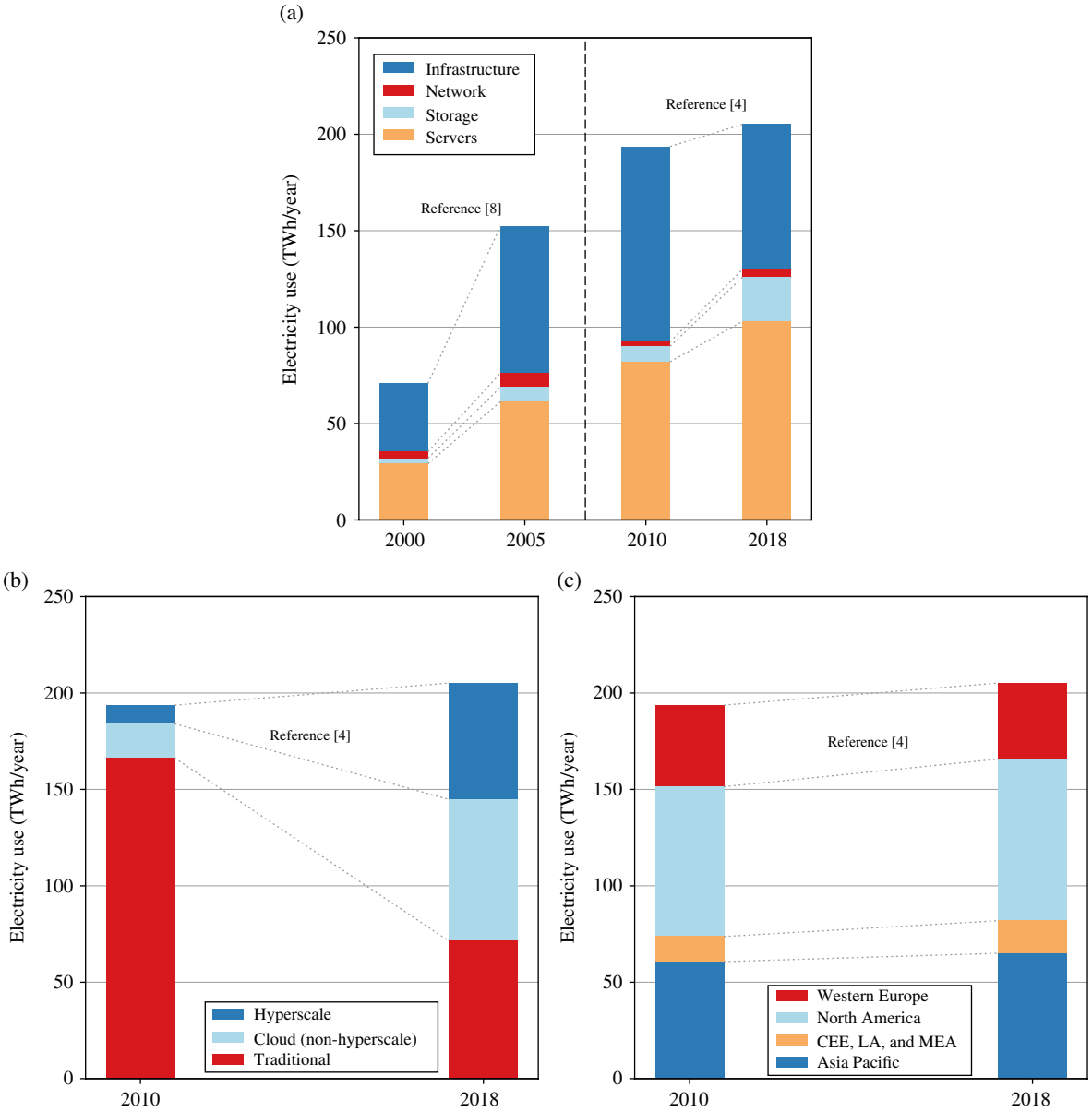


FIGURE 2.1 Global data center energy consumption by end use, data center type, and region. (a) Data center energy use by end use, (b) data center energy use by data center type, and (c) data center energy use by region. Source: © Nuoa Lei.

expanding global stock of installed servers (Fig. 2.1a). Over the same time period, minimal improvements to global average PUE were expected, leading to a similar doubling of electricity used for data center infrastructure systems [8]. By 2010, however, growth in the electricity use of servers had slowed, due to a combination of improved server power efficiencies and increasing levels of server virtualization, which also reduced growth in the number of installed servers [4].

By 2018, the energy use of IT devices accounted for the largest share of data center energy use, due to substantial increases in the energy use of servers and storage devices driven by rising demand for data center computational and data storage services. The energy use of network switches is comparatively much smaller, accounting for only a small fraction over IT device energy use. In contrast, the energy use associated with data center infrastructure systems dropped significantly between 2010 and 2018, thanks to steady improvements in global average PUE values in parallel [15, 16]. As a result of these counteracting effects, global data center energy use rose by only around 6% between 2010 and 2018, despite 11×, 6×, and 26× increases in data center IP traffic, data center compute instances, and installed storage capacity, respectively, over the same time period [4].

Figure 2.1b summarizes data center energy use by major space type category, according to space type definitions in Table 2.1. These data are presented for 2010 and 2018 only, because earlier bottom-up estimates did not consider explicit space types. Between 2010 and 2018, a massive shift away from smaller and less efficient traditional data centers occurred toward much larger and more efficient cloud data centers and toward hyperscale data centers (a subset of cloud) in particular. Over this time period, the energy use of hyperscale data centers increased by about 4.5 times, while the energy use of cloud data centers (non-hyperscale) increased by about 2.7 times. However, the energy use of traditional data centers decreased by about 56%, leading to only modest overall growth in global data center energy use.

As evident in Figure 2.1b, the structural shift away from traditional data centers has brought about significant energy efficiency benefits. Cloud and hyperscale data centers have much lower PUE values compared with traditional data centers, leading to substantially reduced infrastructure energy use (see Fig. 2.1a). Moreover, cloud and hyperscale servers are often operated at much higher utilization levels (thanks to greater server virtualization and workload management strategies), which leads to far fewer required servers compared with traditional data centers.

From a regional perspective, energy use is dominated by North America and Asia Pacific, which together accounted for around three-quarters of global data center energy use in 2018. The next largest energy consuming region is Western Europe, which represented around 20% of global energy use in 2018. It follows that data center energy management practices pursued in North America, Asia Pacific, and Western Europe will have the greatest influence on global data center energy use in the near term.

2.4 GLOBAL DATA CENTER ENERGY USE: FORWARD-LOOKING ANALYSIS

Given the importance of data centers to the global economy, the scale of their current energy use, and the possibility of significant service demand growth, there is increasing interest in forward-looking analyses that assess future data center energy use. However, such analyses are fraught with uncertainties, given the fast pace of technological change associated with IT devices and the unpredictable nature of societal demand for data center services. For these reasons, many IT industry and data center market analysts offer technology and service demand projections only for 3–5 year outlook periods.

Nonetheless, both bottom-up and extrapolation-based methods have been used in forward-looking analyses, and each method comes with important caveats and drawbacks. Extrapolation-based approaches are particularly prone to large variations and errors in forward-looking projections, given their reliance on a few macro-level modeling parameters that ignore the complex technological and structural factors driving data center energy use. In one classic example, extrapolation-based methods based on the early rapid growth phase of the Internet projected that the Internet would account for 50% of US electricity use by 2010, a forecast that was later proven wildly inaccurate when subject to bottom-up scrutiny [17].

To illustrate the sensitive nature of extrapolation-based methods, Figure 2.2b demonstrates how extrapolation-based methods would have predicted 2010–2018 global data center energy use had they been applied to project the bottom-up estimates from 2010 using growth in data center IP traffic (Fig. 2.2a) as a service demand indicator. In fact, several published extrapolation-based estimates have done exactly that [12, 13]. Four different extrapolation-based methods are considered, representing the approaches used in the published studies: (i) extrapolation based on data center electricity CAGR of 10% [11], (ii) extrapolation based on data center electricity CAGR of 12% [18], (iii) extrapolation based on CAGR of data center IP traffic (31%) with a 10% annual electricity efficiency improvement [13], and (iv) extrapolation based on CAGR of data center IP traffic (31%) with a 15% annual electricity efficiency improvement [12].

Compared with the more rigorous bottom-up estimates from 2010 to 2018, which were based on retrospective analysis of existing technology stocks, it is clear that that extrapolation-based methods would have overestimated historical growth in global data center energy use by a factor of 2–3 in 2018. Furthermore, all extrapolation-based methods based on

rising IP traffic demand result in a strong upward trajectory in data center energy use over the 2010–2018 period, implying that as service demands rise in the future, so too much global data center energy use. In contrast, by taking detailed technological stock, energy efficiency, operational, and structural factors into account, the bottom-up approach suggested that global data center energy use grew much more modestly from 2010 to 2018 due to large efficiency gains.

Because bottom-up methods rely on many different technologies, operations, and market data, they are most accurately applied to retrospective analyses for which sufficient historical data exist. When applied to forward-looking analyses, bottom-up methods are typically only employed to consider "what-if" scenarios that explore combinations of different system conditions that could lead to different policy objectives. This approach is in contrast to making explicit energy demand forecasts, given that outlooks for all variables required in a bottom-up framework might not be available. Figure 2.2b plots the only available forward-looking global scenario using bottom-up methods in the literature, which extended 2010–2018 energy efficiency trends alongside projected compute instance demand growth [4]. This scenario found that historical efficiency trends could absorb another doubling of global compute instance demand with negligible growth in data center energy use, but only if strong policy actions were taken to ensure continued uptake of energy-efficient IT devices and data center operational practices.

Also shown in Figure 2.2b are extensions of the four extrapolation-based approaches, which paint a drastically different picture of future data center energy use possibilities, ranging from 3 to 7 times what the bottom-up efficiency scenario implies by around 2023. Forward-looking extrapolations of this type have also appeared in the literature [19], often receiving lots of media attention given the alarming messages they convey about future data energy use growth. However, the historical comparison between bottom-up and extrapolation-based results in Figure 2.2b exposes the inherent risks of applying the latter to forward-looking analyses. Namely, reliance on a few macro-level indicators ignores the many technological, operational, and structural factors that govern global data center energy use, which can lead to large error propagation over time. Therefore, while extrapolation-based projections are easy to construct, their results can be unreliable and lack the explanatory power of bottom-up approaches necessary for managing global data center energy use moving forward.

In summary, bottom-up methods

- are robust and reliable for retrospective analysis, given they are based on historical technology, operations, and market data,
- illuminate key drivers of global data center energy use, and
- require many different data inputs, which can lead to costly, time-intensive, and sporadic analyses,

while extrapolation-based methods

- are simple and easy to implement, relying on only a few macro-level parameters,
- can provide high-level insights or bounding scenarios based on a few assumptions, and
- are subject to large uncertainties, since they tend to ignore important technological, operational, and market structure factor that drive data center energy use.

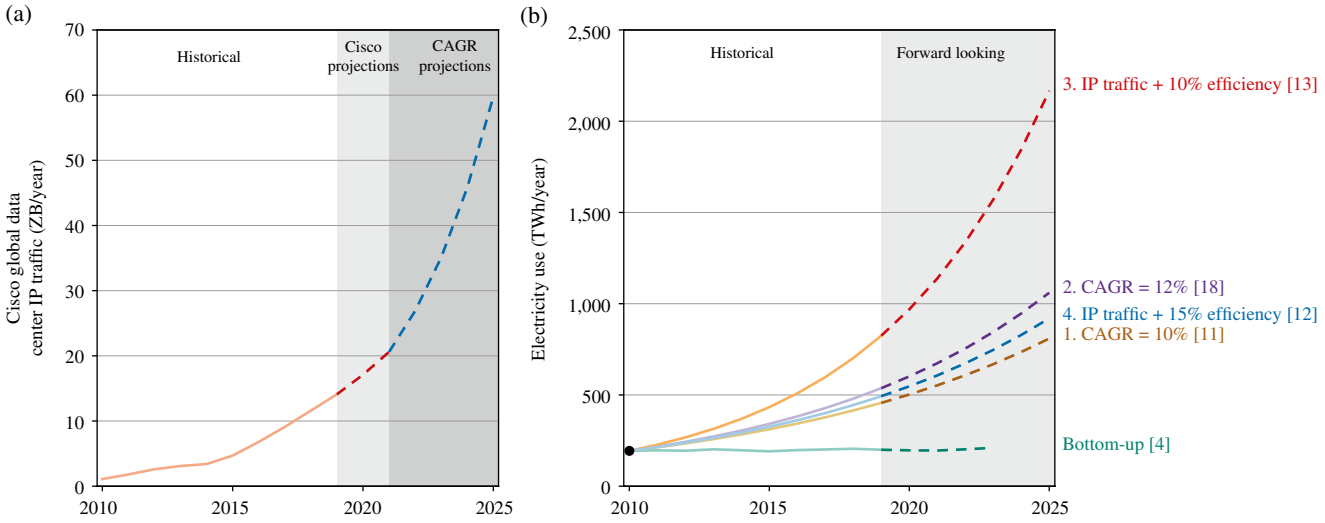


FIGURE 2.2 Comparison of forward-looking analysis methods. (a) Global data center IP traffic. (b) Comparison of forward-looking analysis methods. Source: © Nuoa Lei.

2.5 DATA CENTERS AND CLIMATE CHANGE

The electric power sector is the largest source of energy-related carbon dioxide (CO_2) emissions globally and is still highly dependent upon fossil fuels in many countries [20, 21]. Given their significant electricity use, data center operators have come under scrutiny for their potential contributions to climate change and in particular for their chosen electric power providers and electricity generation sources [22]. As demand for data center services rises in the future, scrutiny regarding the climate change impacts of data centers will likely continue.

To estimate the total CO_2 emissions associated with global data center electricity use, it is first necessary to have data at the country or regional level on data center power use, alongside information on the local electricity generating sources used to provide that power. While a few large data center operators such as Google, Facebook, Amazon, and Microsoft publish some information on their data center locations and electric power sources, the vast majority of data center operators do not. Therefore, it is presently not possible to develop credible estimates of the total CO_2 emissions of the global data center industry in light of such massive data gaps.

However, a number of data center operators are pursuing renewable electricity as part of corporate sustainability initiatives and climate commitments, alongside long-standing energy efficiency initiatives to manage ongoing power requirements. These companies are demonstrating that renewable power can be a viable option for the data center industry, paving the way for other data center operators to consider renewables as a climate change mitigation strategy.

When considering renewable power sources, data centers generally face three key challenges. First, many data center locations may not have direct access to renewable electricity via local grids, either because local renewable resources are limited or because local grids have not added renewable generation capacity. Second, even in areas with adequate renewable resources, most data centers do not have sufficient land or rooftop area for on-site self-generation, given the high-power requirements of the typical data center. Third, due to the intermittent nature of some renewable power sources (particularly solar and wind power), data centers must at least partially rely on local grids for a reliable source of power and/or turn to expensive on-site forms of energy storage to avoid power interruptions.

Therefore, some large data center operators that have adopted renewable power to date have entered into purchase power agreements (PPAs), which provide off-site renewable power to partially or fully offset on-site power drawn from the local grid. For example, Google has utilized PPAs to achieve a milestone of purchasing 100% renewable energy to match the annual electricity consumption of their global data center operations, making it the world's largest corporate buyer of renewables [23]. Google has also located data centers where local grids provide renewable electricity, for example, its North Carolina data center, where solar and wind power contribute to the grid mix [24]. Facebook has also committed to providing all of their data centers with 100% renewable energy, working with local utility partners so that their funded renewable power projects feed energy into the same grids that supply power to their data centers. To date, Facebook's investments have resulted in over 1,000 MW of wind and solar capacity additions to the US power grid [25].

Similar renewable energy initiatives are also being pursued by Apple, Amazon Web Services (AWS), and Microsoft. The global facilities of Apple (including data centers, retail stores, offices, etc.) have been powered by 100% renewable energy since the year of 2018 [26], with a total of 1.4 GW in renewable energy projects across 11 countries to date. AWS exceeded 50% renewable energy usage in 2018 and has committed to 100% renewable energy, with 13 announced renewable energy projects expected to generate more than 2.9 TWh renewable energy annually [27]. Microsoft has committed to being carbon negative by 2030 and, by 2050, to remove all the carbon it has emitted since its founding in 1975 [28]. In 2019, 50% of the power used by Microsoft's data centers had already come from renewable energy, and this percentage is expected to rise to more than 70% by 2023. Meanwhile, Microsoft is planning 100% renewable energy powered new data centers in Arizona, an ideal location for solar power generation [29]. The efforts of these large data center operators have made the ICT industry one of the world's leaders in corporate renewable energy procurement and renewable energy project investments [30].

Despite the impressive efforts of these large data center operators, there is still a long road ahead for the majority of the world's data center to break away from reliance on fossil-fuel-based electricity [31].

2.6 OPPORTUNITIES FOR REDUCING ENERGY USE

Many data centers have ample room to improve energy efficiency, which is an increasingly important strategy for mitigating growth in energy use as demand for data center services continues to rise. Additionally, optimizing energy efficiency makes good business sense, given that electricity purchases are a major component of data center operating costs. Data center energy efficiency opportunities are numerous but generally fall into two major categories: (i) improved IT hardware efficiency and (ii) improved infrastructure systems efficiency. Key strategies within each category are summarized below [32].

2.6.1 IT Hardware

2.6.1.1 Server Virtualization

The operational energy use of servers is generally a function of their processor utilization level, maximum power (i.e. power draw at 100% utilization), and idle power (i.e. power draw at 0% utilization, which can typically represent 10–70% of maximum power [9]). Servers operating at high levels of processor utilization are more efficient on an energy-per-computation basis, because constant idle power losses are spread out over more computations. Many data centers operate servers at low average processor utilization levels, especially when following the conventional practice of hosting one application per server, and sometimes for reasons of redundancy.

Server virtualization is a software-based solution that enables running multiple "virtual machines" on a single server, thereby increasing average server utilization levels and reducing the number of physical servers required to meet a given service demand. The net effect is reduced electricity use. Server virtualization is recognized as one of the single most important strategies for improving data center energy efficiency [7]. While many data centers have already adopted server virtualization, especially in cloud- and hyperscale-class data centers, there is considerable room for greater server virtualization in many data centers and particularly within traditional data centers [2].

2.6.1.2 Remove Comatose Servers

Many data centers may be operating servers whose applications are no longer in use. These "comatose" servers may represent up to 30% of all servers [33] and are still drawing large amounts of idle power for no useful computational output. Therefore, identifying and removing comatose servers can be an important energy saving strategy. While this strategy may seem obvious, in practice, there are several reasons that comatose server operations persist. For example, IT staff may not wish to remove unused servers due to service-level agreements, uncertainty about future demand for installed applications, or lack of clear server life cycle and decommissioning policies within the organization. Therefore, this strategy typically requires a corresponding change in institutional policies and corporate culture oriented around energy efficiency.

2.6.1.3 Energy-Efficient Servers

The most efficient servers typically employ the most efficient power supplies, better DC voltage regulators, more efficient electronic components, a large dynamic range (for example, through dynamic voltage and frequency scaling of processors), purpose-built designs, and the most efficient cooling configurations. In the United States, the ENERGY STAR program certifies energy-efficient servers, which are offered by many different server manufacturers [34]. According to ENERGY STAR, the typical certified server will consume 30% less energy than a conventional server in a similar application. Therefore, specifying energy-efficient servers (such as those with the ENERGY STAR rating) in data center procurement programs can lead to substantial energy savings. In addition to less electricity use by servers, this strategy also reduces cooling system loads (and hence, costs) within the data center.

2.6.1.4 Energy-Efficient Storage Devices

Historically, the energy efficiency of enterprise storage drives has been improving steadily, thanks to continuous improvements in storage density per drive and reductions in average power required per drive [4]. These trends have been realized for both hard disk drive (HDD) and solid-state drive (SSD) storage technologies. Similar to servers, an ENERGY STAR specification for data center storage has been developed [35], which should enable data center operators to identify and procure the most energy-efficient storage equipment in the future.

While SSDs consume less power than HDDs on a per-drive basis [9], the storage capacities of individual SSDs have historically been smaller than those of individual HDDs, giving HDDs an efficiency advantage from an energy per unit capacity (e.g. kilowatt-hour per terabyte (kWh/TB)) perspective. However, continued improvements to SSDs may lead to lower kWh/TB than HDDs in the future [36].

For HDDs, power use is proportional to the cube of rotational velocity. Therefore, an important efficiency strategy is to select the slowest spindle speed that provides a sufficient read/write speed for a given set of applications [37].

SSDs is becoming more popular because it is an energy-efficient alternative to HDDs. With no spinning disks, SDDs consume much less power than HDDs. The only disadvantage of SDDs is that it cost much higher than HDDs for per gigabyte of data storage.

2.6.1.5 Energy-Efficient Storage Management

While it is important to utilize the most energy-efficient storage drives, strategic management of those drives can lead to substantial additional energy savings. One key strategy involves minimizing the number of drives required by maximizing utilization of storage capacity, for example, through storage consolidation and virtualization, automated storage provisioning, or thin provisioning [38].

Another key management strategy is to reduce the overall quantities of data that must be stored, thereby leading to less required storage capacity. Some examples of this strategy include data deduplication (eliminating duplicate copies of the same data), data compression (reducing the number of

bits required to represent data), and use of delta snapshot techniques (storing only changes to existing data) [37].

Lastly, another strategy is use of tiered storage so that certain drives (i.e. those with infrequent data access) can be powered down when not in use. For example, MAID technology saves power by shutting down idle disks, thereby leading to energy savings [39].

2.6.2 Infrastructure Systems

2.6.2.1 Airflow Management

The goal of improved airflow management is to ensure that flows of cold airflows reach IT equipment racks and flows of hot air return from cooling equipment intakes in the most efficient manner possible and with minimal mixing of cold and hot air streams. Such an arrangement helps reduce the amount of energy required for air movement (e.g. via fans or blowers) and enables better optimization of supply air temperatures, leading to less electricity use by data center cooling systems. Common approaches include uses of "hot aisle/cold aisle" layouts, flexible strip curtains, and IT equipment containment enclosures, the latter of which can also reduce the required volume of air cooled [37]. The use of airflow simulation software can also help data center operators identify hot zones and areas with inefficient airflow, leading to system adjustments that improve cooling efficiency [40].

2.6.2.2 Energy-Efficient Equipment

Uninterruptible power supply (UPS) systems are a major mission-critical component within the data center. Operational energy losses are inherent in all UPS systems, but these losses can vary widely based on the efficiency and loading of the system. The UPS efficiency is expressed as power delivered from the UPS system to the data center divided by power delivered to the UPS system.

The conversion technology employed by a UPS has a major effect on its efficiency. UPS systems using double conversion technology typically have efficiencies in the low 90% range, whereas UPS systems using a delta conversion technology could achieve efficiencies as high as 97% [41]. Furthermore, the UPS efficiency increases with increasing power loading and peaks when 100% of system load capacity is reached, which suggests that proper UPS system sizing is an important energy efficiency strategy [42].

Because data center IT loads fluctuate continuously, so does the demand for cooling. The use of variable-speed drives (VSDs) on cooling system fans allows for speed to be adjusted based on airflow requirements, leading to energy savings. According to data from the ENERGY STAR program [37], the use of VSDs in data center air handling systems is also an economical investment, with simple payback times from energy savings reported from 0.5 to 1.7 years.

Another important opportunity relates to data center humidification, which can be necessary to prevent electrostatic discharge (ESD). Inefficient infrared or steam-based systems, which can raise air temperatures and place additional loads on cooling systems, can sometimes be replaced with much more energy-efficient adiabatic humidification technologies. Adiabatic humidifiers typically utilize water spraying, wetted media, or ultrasonic approaches to introduce water into the air without raising air temperatures [43].

2.6.2.3 Economizer Use

The use of so-called free cooling is one of the most common and effective means of reducing infrastructure energy use in data centers by partially or fully replacing cooling from mechanical chillers. However, the extent to which free cooling can be employed depends heavily on a data center's location and indoor thermal environment specifications [44–46]. The two most common methods of free cooling are air-side economizers and water-side economizers.

When outside air exhibits favorable temperature and humidity characteristics, an air-side economizer can be used to bring outside air into the data center for cooling IT equipment. Air-side economizers provide an economical way of cooling not only in cold climates but also in warmer climates where it can make use of cool evening and wintertime air temperatures. According to [47], using an air-side economizer may lower cooling costs by more than 50% compared with conventional chiller-based systems.

Air-side economizers can also be combined with evaporative cooling by passing outside air through a wetted media or misting device. For example, the Facebook data center in Prineville, Oregon, achieved a low PUE of 1.07 by using 100% outside air with an air-side economizer with evaporative cooling [48].

When the wet-bulb temperature of outside air (or the temperature of the water produced by cooling towers) is low enough or if local water sources with favorable temperatures are available (such as from lakes, bays, or other surface water sources), a water-side economizer can be used. In such systems, cold water produced by the water-side economizer passes through cooling coils to cool indoor air provided to the IT equipment. According to [37], the operation of water-side economizers can reduce the costs of a chilled water plant by up to 70%. In addition to energy savings, water-side economizer can also offer cooling redundancy by producing chilled water when a mechanical chiller goes offline, which reduces the risk of data center down time.

2.6.2.4 Data Center Indoor Thermal Environment

Traditionally, many data centers set their supply air dry-bulb temperature as low as 55°F. However, such a low temperature is generally unnecessary because typical servers can be safely

operated within a temperature range of 50–99 °F [37]. For example, Google found that computing hardware can be reliably run at temperatures above 90 °F; the peak operating temperature of their Belgium data center could reach 95 °F [49]. Intel investigated using only outdoor air to cool a data center; the observed temperature was between 64 and 92 °F with no corresponding server failures [50]. Therefore, many data centers can save energy simply by raising their supply air temperature set point. According to [37], every 1 °F increase in temperature can lead to 4–5% savings in cooling energy costs.

Similarly, many data centers may have an opportunity to save energy by revisiting their humidification standards. Sufficient humidity is necessary to avoid ESD failures, whereas avoiding high humidity is necessary to avoid condensation that can cause rust and corrosion. However, there is growing understanding that ASHRAE's 2008 recommended humidity ranges, by which many data centers abide, may be too restrictive [44].

For example, the risk of ESD from low humidity can be avoided by applying grounding strategies for IT equipment, while some studies have found that condensation from high humidity is rarely a concern in practice [37, 44]. Most IT equipment is rated for operating at relative humidity levels of up to 80%, while some Facebook data centers condition outdoor air up to a relative humidity of 90% to make extensive use of adiabatic cooling [51].

Therefore, relaxing previously strict humidity standards can lead to energy savings by reducing the need for humidification and dehumidification, which reduces overall cooling system energy use. In light of evolving understanding of temperature and humidity effects on IT equipment, ASHRAE has evolved its thermal envelope standards over time, as shown in Table 2.2.

2.7 CONCLUSIONS

Demand for data center services is projected to grow substantially. Understanding the energy use implications of this demand, designing data center energy management policies, and monitoring the effectiveness of those policies over time require modeling of data center energy use. Historically, two different modeling approaches have been used in national and global analyses: extrapolation-based and bottom-up approaches, with the latter generally providing the most robust insights into the myriad technology, operations, and market drivers of data center energy use. Improving data collection, data sharing, model development, and modeling best practices is a key priority for monitoring and managing data center energy use in the big data era.

Quantifying the total global CO_2 emissions of data centers remains challenging due to lack of sufficient data on many data center locations and their local energy mixes, which are only reported by a small number of major data center operators. It is these same major operators who are also leading the way to greater adoption of renewable power sources, illuminating an important pathway for reducing the data center industry's CO_2 footprint. Lastly, there are numerous proven energy efficiency improvements applicable to IT devices and infrastructure systems that can still be employed in many data centers, which can also mitigate growth in overall energy use as service demands rise in the future.

TABLE 2.2 ASHRAE recommended envelopes comparisons

Year	Dry-bulb temperature (°C)	Humidity range
2004	20 – 25	Relative humidity 40–55%
2008	18 – 27	Low end: 5.5 °C dew point High end: 60% relative humidity and 15°C dew point
2015	18 – 27	Low end: −9 °C dew point High end: 60% relative humidity and 15°C dew point

Source: [44] and [45].

REFERENCES

[1] Cisco Global Cloud Index. Forecast and methodology, 2010–2015 White Paper; 2011.

[2] Cisco Global Cloud Index. Forecast and methodology, 2016–2021 White Paper; 2018.

[3] IEA. Aviation—tracking transport—analysis, Paris; 2019.

[4] Masanet ER, Shehabi A, Lei N, Smith S, Koomey J. Recalibrating global data center energy use estimates. Science 2020;367(6481):984–986.

[5] Reinsel D, Gantz J, Rydning J. The digitization of the world from edge to core. IDC White Paper; 2018.

[6] Brown RE, et al. *Report to Congress on Server and Data Center Energy Efficiency: Public Law 109-431*. Berkeley, CA: Ernest Orlando Lawrence Berkeley National Laboratory; 2007.

[7] Masanet ER, Brown RE, Shehabi A, Koomey JG, Nordman B. Estimating the energy use and efficiency potential of US data centers. Proc IEEE 2011;99(8):1440–1453.

[8] Koomey J. Growth in data center electricity use 2005 to 2010. *A report by Analytical Press, completed at the request of The New York Times*, vol. 9, p. 161; 2011.

[9] Shehabi A, et al. United States Data Center Energy Usage Report. Lawrence Berkeley National Lab (LBNL), Berkeley, CA, LBNL-1005775; June 2016.

[10] Pickavet M, et al. Worldwide energy needs for ICT: the rise of power-aware networking. Proceedings of the 2008 2nd International Symposium on Advanced Networks and Telecommunication Systems; December 2008. p 1–3.

[11] Belkhir L, Elmeligi A. Assessing ICT global emissions footprint: trends to 2040 and recommendations. J Clean Prod 2018;177:448–463.

[12] Andrae ASG. Total consumer power consumption forecast. Presented at the Nordic Digital Business Summit; October 2017.

[13] Andrae A, Edler T. On global electricity usage of communication technology: trends to 2030. Challenges 2015;6(1):117–157.

[14] Koomey JG. Worldwide electricity used in data centers. Environ Res Lett 2008;3(3):034008.

[15] International Energy Agency (IEA). *Digitalization and Energy*. Paris: IEA; 2017.

[16] Uptime Institute. Uptime Institute Global Data Center Survey; 2018.

[17] Koomey J, Holdren JP. *Turning Numbers into Knowledge: Mastering the Art of Problem Solving*. Oakland, CA: Analytics Press; 2008.

[18] Corcoran P, Andrae A. Emerging Trends in Electricity Consumption for Consumer ICT. National University of Ireland, Galway, Connacht, Ireland, Technical Report; 2013.

[19] Jones N. How to stop data centres from gobbling up the world's electricity. Nature 2018;561:163–166.

[20] IEA. CO_2 emissions from fuel combustion 2019. IEA Webstore. Available at https://webstore.iea.org/co2-emissions-from-fuel-combustion-2019. Accessed on February 13, 2020.

[21] IEA. Key World Energy Statistics 2019. IEA Webstore. Available at https://webstore.iea.org/key-world-energy-statistics-2019. Accessed on February 13 2020.

[22] Greenpeace. Greenpeace #ClickClean. Available at http://www.clickclean.org. Accessed on February 13, 2020.

[23] Google. 100% Renewable. Google Sustainability. Available at https://sustainability.google/projects/announcement-100. Accessed on February 10, 2020.

[24] Google. The Internet is 24×7—carbon-free energy should be too. Google Sustainability. Available at https://sustainability.google/projects/24x7. Accessed on February 10, 2020.

[25] Facebook. Sustainable data centers. Facebook Sustainability. Available at https://sustainability.fb.com/innovation-for-our-world/sustainable-data-centers. Accessed on February 10, 2020.

[26] Apple. Apple now globally powered by 100 percent renewable energy. Apple Newsroom. Available at https://www.apple.com/newsroom/2018/04/apple-now-globally-powered-by-100-percent-renewable-energy. Accessed on February 13, 2020.

[27] AWS. AWS and sustainability. Amazon Web Services Inc. Available at https://aws.amazon.com/about-aws/sustainability. Accessed on February 13, 2020.

[28] Microsoft. Carbon neutral and sustainable operations. Microsoft CSR. Available at https://www.microsoft.com/en-us/corporate-responsibility/sustainability/operations. Accessed on February 13, 2020.

[29] Microsoft. Building world-class sustainable datacenters and investing in solar power in Arizona. Microsoft on the Issues July 30, 2019. Available at https://blogs.microsoft.com/on-the-issues/2019/07/30/building-world-class-sustainable-datacenters-and-investing-in-solar-power-in-arizona. Accessed on February 13, 2020.

[30] IEA. Data centres and energy from global headlines to local headaches? Analysis. IEA. Available at https://www.iea.org/commentaries/data-centres-and-energy-from-global-headlines-to-local-headaches. Accessed on February 13, 2020.

[31] Greenpeace. Greenpeace releases first-ever clean energy scorecard for China's tech industry. Greenpeace East Asia. Available at https://www.greenpeace.org/eastasia/press/2846/greenpeace-releases-first-ever-clean-energy-scorecard-for-chinas-tech-industry. Accessed on February 10, 2020.

[32] Huang R, Masanet E. Data center IT efficiency measures evaluation protocol; 2017.

[33] Koomey J, Taylor J. Zombie/comatose servers redux; 2017.

[34] ENERGY STAR. Energy efficient enterprise servers. Available at https://www.energystar.gov/products/data_center_equipment/enterprise_servers. Accessed on February 13, 2020.

[35] ENERGY STAR. Data center storage specification version 1.0. Available at https://www.energystar.gov/products/spec/data_center_storage_specification_version_1_0_pd. Accessed on February 13, 2020.

[36] Dell. Dell 2020 energy intensity goal: mid-term report. Dell. Available at https://www.dell.com/learn/al/en/alcorp1/corporate~corp-comm~en/documents~energy-white-paper.pdf. Accessed on February 13, 2020.

[37] ENERGY STAR. 12 Ways to save energy in data centers and server rooms. Available at https://www.energystar.gov/products/low_carbon_it_campaign/12_ways_save_energy_data_center. Accessed on February 14, 2020.

[38] Berwald A, et al. *Ecodesign Preparatory Study on Enterprise Servers and Data Equipment*. Luxembourg: Publications Office; 2014.

[39] SearchStorage. What is MAID (massive array of idle disks)? SearchStorage. Available at https://searchstorage.techtarget.com/definition/MAID. Accessed on February 14, 2020.

[40] Ni J, Bai X. A review of air conditioning energy performance in data centers. Renew Sustain Energy Rev 2017;67:625–640.

[41] Facilitiesnet. The role of a UPS in efficient data centers. Facilitiesnet. Available at https://www.facilitiesnet.com/datacenters/article/The-Role-of-a-UPS-in-Efficient-Data-Centers--11277. Accessed on February 13, 2020.

[42] Q. P. S. Team. When an energy efficient UPS isn't as efficient as you think. www.qpsolutions.net September 24, 2014. Available at https://www.qpsolutions.net/2014/09/when-an-energy-efficient-ups-isnt-as-efficient-as-you-think/. Accessed on September 3, 2020.

[43] STULZ. Adiabatic/evaporative vs isothermal/steam. Available at https://www.stulz-usa.com/en/ultrasonic-humidification/adiabatic-vs-isothermalsteam/. Accessed on February 13, 2020.

[44] American Society of Heating Refrigerating and Air-Conditioning Engineers. *Thermal Guidelines for Data Processing Environments*. 4th ed. Atlanta, GA: ASHRAE; 2015.

[45] American Society of Heating, Refrigerating and Air-Conditioning Engineers. *Thermal Guidelines for Data Processing Environments*. Atlanta, GA: ASHRAE; 2011.

[46] Lei, Nuoa, and Eric Masanet. "Statistical analysis for predicting location-specific data center PUE and its improvement potential." Energy 2020: 117556.

[47] Facilitiesnet. Airside economizers: free cooling and data centers. Facilitiesnet. Available at https://www.facilitiesnet.com/datacenters/article/Airside-Economizers-Free-Cooling-and-Data-Centers--11276. Accessed on February 14, 2020.

[48] Park J. Designing a very efficient data center. Facebook April 14, 2011. Available at https://www.facebook.com/notes/facebook-engineering/designing-a-very-efficient-data-center/10150148003778920. Accessed on August 11, 2018.

[49] Humphries M. Google's most efficient data center runs at 95 degrees. Geek.com March 27, 2012. Available at https://www.geek.com/chips/googles-most-efficient-data-center-runs-at-95-degrees-1478473. Accessed on September 23, 2019.

[50] Miller R. Intel: servers do fine with outside air. Data Center Knowledge. Available at https://www.datacenterknowledge.com/archives/2008/09/18/intel-servers-do-fine-with-outside-air. Accessed on September 4, 2019.

[51] Miller R. Facebook servers get hotter but run fine in the South. Data Center Knowledge. Available at https://www.datacenterknowledge.com/archives/2012/11/14/facebook-servers-get-hotter-but-stay-cool-in-the-south. Accessed on September 4, 2019.

[52] Barroso LA, Hölzle U, Ranganathan P. The datacenter as a computer: designing warehouse-scale machines. Synth Lectures Comput Archit 2018;13(3):i–189.

FURTHER READING

IEA Digitalization and Energy [15].

Recalibrating Global Data Center Energy-use Estimates [4]

The Datacenter as a Computer: Designing Warehouse-Scale Machines [52]

United States Data Center Energy Usage Report [9]

3

ENERGY AND SUSTAINABILITY IN DATA CENTERS

BILL KOSIK
DNV Energy Services USA Inc., Chicago, Illinois, United States of America

3.1 INTRODUCTION

In 1999, *Forbes* published a seminal article co-authored by Peter Huber and Mark Mills. It had a wonderful tongue-in-cheek title: "Dig More Coal—the PCs Are Coming." The premise of the article was to challenge the idea that the Internet would actually reduce overall energy use in the United States, especially in sectors such as transportation, banking, and healthcare where electronic data storage, retrieval, and transaction processing were becoming integral to business operations. The opening paragraph, somewhat prophetic, reads:

> SOUTHERN CALIFORNIA EDISON, meet Amazon.com. Somewhere in America, a lump of coal is burned every time a book is ordered on-line. The current fuel-economy rating: about 1 pound of coal to create, package, store and move 2 megabytes of data. The digital age, it turns out, is very energy-intensive. The Internet may someday save us bricks, mortar and catalog paper, but it is burning up an awful lot of fossil fuel in the process.

These words, although written more than two decades ago, are still meaningful today. Clearly Mills was trying to demonstrate that a great deal of electricity is used by servers, networking gear, and storage devices residing in large data centers that also consume energy for cooling and powering ITE (information technology equipment) systems. As the data center industry matures with respect to becoming more conversant and knowledgeable on energy efficiency, and environmental responsibility-related issues. For example, data center owners and end users are expecting better server efficiency and airflow optimization and using detailed building performance simulation techniques comparing "before and after" energy usage to justify higher initial spending to reduce ongoing operational costs.

3.1.1 Industry Accomplishments in Reducing Energy Use in Data Centers

Since the last writing of this chapter in the first edition of *The Data Center Handbook* (2015), there have been significant changes in the data center industry's approach to reducing energy usage of cooling, power, and ITE systems. But some things haven't changed: energy efficiency, optimization, usage, and cost are still some of the primary drivers when analyzing the financial performance and environmental impact of a data center. Some of these approaches have been driven by ITE manufacturers; power requirements for servers, storage, and networking gear have dropped considerably. Servers have increased in performance over the same period, and in some cases the servers will draw the same power as the legacy equipment, but the performance is much better, increasing performance-per-watt. In fact, the actual energy use of data centers is much lower than initial predictions (Fig. 3.1).

Another substantial change comes from the prevalence of cloud data centers, along with the downsizing of enterprise data centers. Applications running on the cloud have technical advantages and can result in cost savings compared to local managed servers. Elimination of barriers and reduced cost from launching Web services using the cloud offers easier start-up, scalability, and flexibility. On-demand computing is one of the prime advantages of the cloud, allowing users to start applications with minimal cost.

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

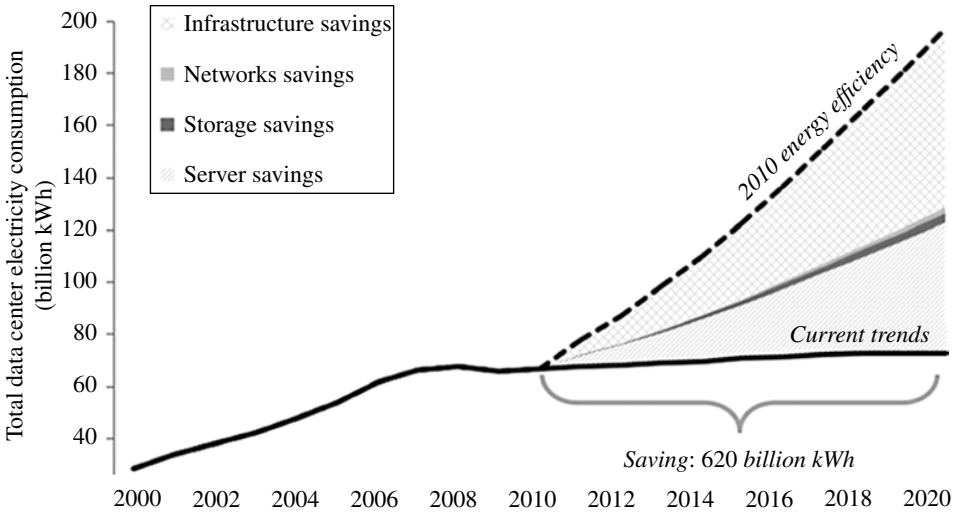


FIGURE 3.1 Actual energy use of data centers is lower than initial predictions. Source: [1].

Responding to a request from Congress as stated in Public Law 109-431, the U.S. Environmental Protection Agency (EPA) developed a report in 2007 that assessed trends in energy use, energy costs of data centers, and energy usage of ITE systems (server, storage, and networking). The report also contains existing and emerging opportunities for improved energy efficiency. This report eventually became the de facto source for projections on energy use attributable to data centers. One of the more commonly referred-to charts that was issued with the 2007 EPA report (Fig. 3.2) presents several different energy usage outcomes based on different consumption models.

3.1.2 Chapter Overview

The primary purpose of this chapter is to provide an appropriate amount of data on the drivers of energy use in data centers. It is a complex topic—the variables involved in the optimization of energy use and the minimization of environmental impacts are cross-disciplinary and include information technology (IT) professionals, power and cooling engineers, builders, architects, finance and accounting professionals, and energy procurement teams. Adding to the complexity, a data center must run 8,760 h/year, nonstop, including all scheduled maintenance (unscheduled breakdowns), and ensure that ultracritical business operations are

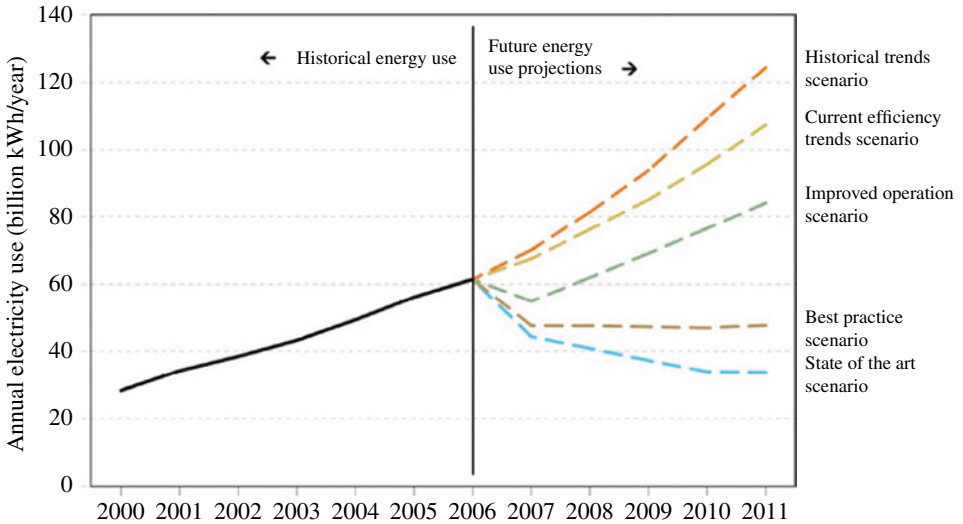


FIGURE 3.2 2007 EPA report on energy use and costs of data centers. Source: [2].

not interrupted, keeping the enterprise running. In summary, planning, design, implementation, and operations of a data center take a considerable amount of effort and attention to detail. And after the data center is built and operating, the energy cost of running the facility, if not optimized during the planning and design phases, will provide a legacy of inefficient operation and high electricity costs. Although this is a complex issue, this chapter will not be complex; it will provide concise, valuable information, tips, and further reading resources.

The good news is the industry is far more knowledgeable and interested in developing highly energy-efficient data centers. This is being done for several reasons, including (arguably the most important reason) the reduction of energy use, which leads directly to reduced operating costs. With this said, looking to the future, will there be a new technological paradigm emerging that eclipses all of the energy savings that we have achieved? Only time will tell, but it is clear that we need to continue to push hard for nonstop innovation, or as another one of my favorite authors, Tom Peters, puts it, "Unless you walk out into the unknown, the odds of making a profound difference. . .are pretty low."

3.1.3 Energy-Efficient and Environmentally Responsible Data Centers

When investigating the possible advantages of an efficient data center, questions will arise such as "Is there a business case for doing this (immediate energy savings, future energy savings, increased productivity, better disaster preparation, etc.?) Or should the focus be on the environmental advantages, such as reduction in energy and water use and reduction of greenhouse gas (GHG) emissions?" Keep in mind that these two questions are not mutually exclusive. Data centers can show a solid ROI and be considered sustainable. In fact, some of the characteristics that make a data center environmentally responsible are the same characteristics that make it financially viable. This is where the term *sustainable* can really be applied—sustainable from an environmental perspective but also from a business perspective. And the business perspective could include tactical upgrades to optimize energy use or it could include increasing market share by taking an aggressive stance on minimizing the impact on the environment—and letting the world know about it.

When planning a renovation of an existing facility, there are different degrees of efficiency upgrades that need to be considered. When looking at specific efficiency measures for a data center, there are typically some "quick wins" related to the power and cooling systems that will have paybacks of 1 or 2 years. Some have very short paybacks because there are little or no capital expenditures involved. Examples of these are adjusting set points for temperature and humidity, minimizing raised floor leakage, optimizing control and sequencing of cooling equipment, and optimizing air management on the raised floor to eliminate hot spots, which may allow for a small increase in supply air temperature, reducing energy consumption of compressorized cooling equipment. Other upgrades such as replacing cooling equipment have a larger scope of work and greater first cost. These projects are typically result in a simple payback of 5–10 years. But there are benefits beyond energy efficiency; they also will lower maintenance costs and improve reliability. These types of upgrades typically include replacement of central cooling plant components (chillers, pumps, cooling towers) as well as electrical distribution (UPS, power distribution units). These are more invasive and will require shutdowns unless the facility has been designed for concurrent operation during maintenance and upgrades. A thorough analysis, including first cost, energy cost, operational costs, and GHG emissions, is the only way to really judge the viability of different projects.

Another important aspect of planning an energy efficiency upgrade project is taking a holistic approach in the many different aspects of the project, but especially planning. For example, including information on the future plans for the ITE systems may result in an idea that wouldn't have come up if the ITE plans were not known. Newer ITE gear will reduce the cooling load and, depending on the data center layout, with improve airflow and reduce air management headaches. Working together, the facilities and ITE organizations can certainly make an impact in reducing energy use in the data center that would not be realized if the groups worked independently (see Fig. 3.3).

3.1.4 Environmental Impact

Bear in mind that a typical enterprise data center consumes *40 times*, *or more*, as much energy as a similarly sized office building. Cloud facilities and supercomputing data centers will be an order of magnitude greater than that. A company that has a large real estate portfolio including data centers will undoubtedly be at the top of list in ranking energy consumption. The data center operations have a major impact on the company's overall energy use, operational costs, and carbon footprint. As a further complication, not all IT and facilities leaders are in a position to adequately ensure optimal energy efficiency, given their level of sophistication, experience, and budget availability for energy efficiency programs. So where is the best place to begin?

3.1.5 The Role of the U.S. Federal Government and the Executive Order

Much of what the public sees coming out of the U.S. federal government is a manifestation of the political and moral will of lawmakers, lobbyists, and the President. Stretching back to George Washington's time in office, U.S. presidents have used Executive Orders (EO) to

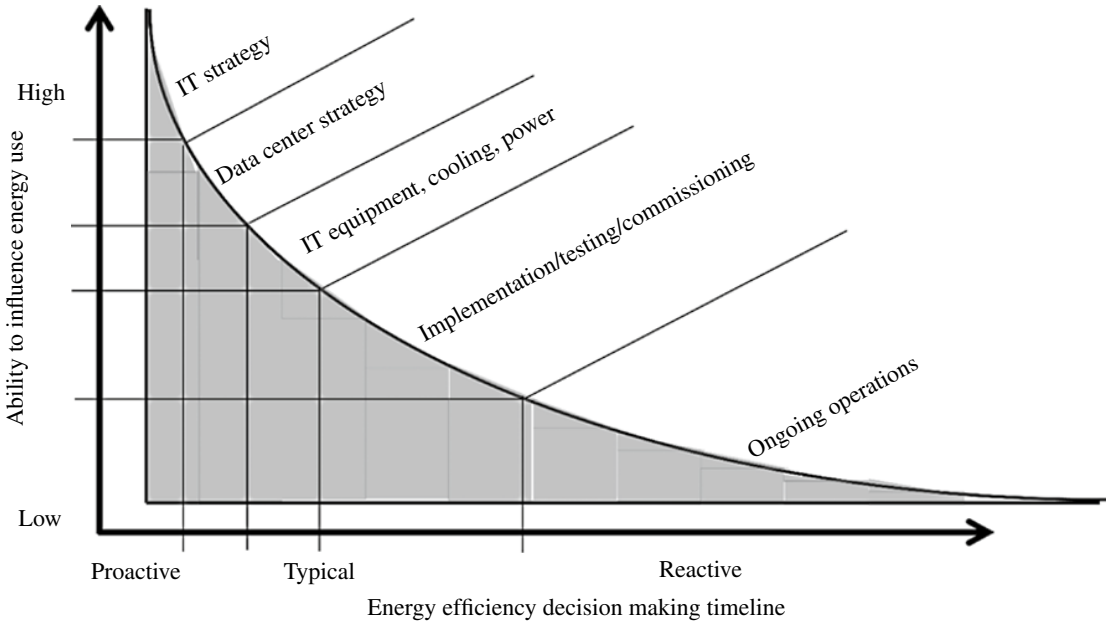


FIGURE 3.3 Data center planning timeline. Source: ©2020, Bill Kosik.

effectuate change in our country's governance. Safeguarding our country during war, providing emergency assistance to areas hit by natural disasters, encouraging/discouraging regulation by federal agencies, and avoiding financial crises are all good examples where presidents signed EO to expedite a favorable outcome.

One example, EO 13514, Federal Leadership in Environmental, Energy, and Economic Performance, signed by President Obama on October 5, 2009, outlines a mandate for reducing energy consumption, water use, and GHG emissions in U.S. federal facilities. Although the EO is written specifically for U.S. federal agencies, the broader data center industry is also entering the next era of energy and resource efficiency. The basic tenets in the EO can be applied to any type of enterprise. While the EO presents requirements for reductions for items other than buildings (vehicles, electricity generation, etc.), the majority of the EO is geared toward the built environment. Related to data centers specifically, and the impact that technology use has on the environment, there is a dedicated section on electronics and data processing facilities. An excerpt from this section states, ". . . [agencies should] promote electronics stewardship, in particular by implementing best management practices for energy-efficient management of servers and Federal data centers." Unfortunately although the EO has been revoked by EO 13834, many of the goals outlined in the EO have been put into operation by several federal, state, and local governmental bodies. Moreover, the EO raised awareness within the federal government not only on issues related to energy efficiency but also recycling, fuel efficiency, and GHG emissions; it is my hope that this awareness will endure for government employees and administrators that are dedicated to improving the outlook for our planet.

There are many other examples of where EO have been used to implement plans related to energy, sustainability, and environmental protection. The acceptance of these EO by lawmakers and the public depends on one's political leaning, personal principles, and scope of the EO. Setting aside principles of government expansion/contraction and strengthening/loosening regulation on private sector enterprises, the following are just some of the EO that have been put into effect by past presidents:

- Creation of the EPA and setting forth the components of the National Oceanic and Atmospheric Administration (NOAA), the basis for forming a "strong, independent agency," establishing and enforcing federal environmental protection laws.
- Expansion of the Federal Sustainability Agenda and the Office of the Federal Environmental Executive.
- Focusing on eliminating waste and expanding the use of recycled materials, increased sustainable building practices, renewable energy, environmental management systems, and electronic waste recycling.
- Creation of the Presidential Awards for agency achievement in meeting the President's sustainability goals.

- Directing EPA, DOE, DOT, and the USDA to take the first steps cutting gasoline consumption and GHG emissions from motor vehicles by 20%.
- Using sound science, analysis of benefits and costs, public safety, and economic growth, coordinate agency efforts on regulatory actions on GHG emissions from motor vehicles, nonroad vehicles, and nonroad engines.
- Require a 30% reduction in vehicle fleet petroleum use; a 26% improvement in water efficiency; 50% recycling and waste diversion; and ensuring 95% of all applicable contracts meet sustainability requirements.

The EO is a powerful tool to get things done quickly, and there are numerous success stories where an EO created new laws promoting environmental excellence. However, EO are fragile—they can be overturned by future administrations. Creating effective and lasting laws for energy efficiency and environmental laws must go through the legislative process, where champions from within Congress actively nurture and promote the bill; they work to gain support within the legislative branch, with the goal of passing the bill and get it on the President's desk for signing.

3.1.6 Greenhouse Gas and CO_2 Emissions Reporting

When using a certain GHG accounting and reporting protocol for analyzing the carbon footprint of an operation, the entire electrical power production chain must be considered. This chain starts at the utility-owned power plant and then all the way to the building. The utility that supplies energy in the form of electricity and natural gas impacts the operating cost of the facility and drives the amount of CO_{2eq} that is released into the atmosphere. When evaluating a comprehensive energy and sustainability plan, it is critical to understand the source of energy (fossil fuel, coal, nuclear, oil, natural gas, wind, solar, hydropower, etc.) and the efficiency of the electricity generation to develop an all-inclusive view of how the facility impacts the environment.

As an example, Scope 2 emissions, as they are known, are attributable to the generation of purchased electricity consumed by the company. And for many companies, purchased electricity represents one of the largest sources of GHG emissions (and the most significant opportunity to reduce these emissions). Every type of cooling and power system consumes different types and amounts of fuel, and each power producer uses varying types of renewable power generation technology such as wind and solar. The cost of electricity and the quantity of CO_2 emissions from the power utility have to be considered.

To help through this maze of issues, contemporary GHG accounting and reporting protocols have clear guidance on how to organize the thinking behind reporting and reducing CO_2 emissions by using the following framework:

- Accountability and transparency: Develop a clear strategic plan, governance, and a rating protocol.
- Strategic sustainability performance planning: Outline goals, identify policies, and procedures.
- Greenhouse gas management: Reduce energy use in buildings and use on-site energy sources using renewables.
- Sustainable buildings and communities: Implement strategies for developing high-performance buildings, looking at new construction, operation, and retrofits.
- Water efficiency: Analyze cooling system alternatives to determine direct water use (direct use by the heat rejection equipment at the facility) and indirect water consumption (used for cooling thermomechanical processes at the power generation facility). The results of the water use analysis, in conjunction with building energy use estimation (derived from energy modeling), are necessary to determine the optimal balancing point between energy and water use.

3.1.7 Why Report Emissions?

It is important to understand that worldwide, thousands of companies report their GHG footprint. Depending on the country the company is located, some are required to report their GHG emissions. Organizations such as the Carbon Disclosure Project (CDP) assist corporations in gathering data and reporting the GHG footprint. (This is a vast oversimplification of the actual process, and companies spend a great deal of time and money in going through this procedure.) This is especially true for companies that report GHG emissions, even though it is not compulsory. There are business-related advantages for these companies that come about as a direct result of their GHG disclosure. Some examples of these collateral benefits result in:

- Suppliers that self-report and have customers dedicated to environmental issues; the customers have actively helped the suppliers improve their environmental performance and assist in managing risks and identifying future opportunities.
- Many of the companies that publicly disclosed their GHG footprint did so at the request of their investors and major purchasing organizations. The GHG data reported by the companies is crucial to help investors in their decision making, engaging with the companies, and to reduce risks and identify opportunities.
- Some of the world's largest companies that reported their GHG emissions were analyzed against a diverse range of metrics including transparency, target-setting, and awareness of risks and opportunities. Only the very best rose to the top, setting them apart from their competitors.

3.2 MODULARITY IN DATA CENTERS

Modular design, the construction of an object by joining together standardized units to form larger compositions, plays an essential role in the planning, design, and construction of data centers. Typically, as a new data center goes live, the ITE remains in a state of minimal computing power for a period of time. After all compute, storage, and networking gear is installed, utilization starts to increase, which drives up the rate of energy consumption and intensifies heat dissipation of the IT gear well beyond the previous state of minimal compute power. The duration leading up to full power draw varies on a case-by-case basis and is oftentimes is difficult to predict in a meaningful way. And in most enterprise data centers, the equipment, by design, will never hit the theoretical maximum compute power. There are many reasons this is done In fact, most data centers contain ITE that, by design, will never hit 100% computing ability. (This is done for a number of reasons including capacity and redundancy considerations.) This example is a demonstration of how data center energy efficiency can increase using modular design with malleability and the capability to react to shifts, expansions, and contractions in power use as the business needs of the organization drive the ITE requirements.

3.2.1 What Does a Modular Data Center Look Like?

Scalability is a key strategic advantage gained when using modular data centers, accommodating compute growth as the need arises. Once a module is fully deployed and running at maximum workload, another modular data center can be deployed to handle further growth.

The needs of the end user will drive the specific type of design approach, but all approaches will have similar characteristics that will help in achieving the optimization goals of the user. Modular data centers (also see Chapter 4 in the first edition of the *Data Center Handbook*) come in many sizes and form factors, typically based around the customer's needs:

1. Container: This is typically what one might think of when discussing modular data centers. Containerized data centers were first introduced using standard 20- and 40-ft shipping containers. Newer designs now use custom-built containers with insulated walls and other features that are better suited for housing computing equipment. Since the containers will need central power and cooling systems, the containers will typically be grouped and fed from a central source. Expansion is accomplished by installing additional containers along with the required additional sources of power and cooling.
2. Industrialized data center: This type of data center is a hybrid model of a traditional brick-and-mortar data center and the containerized data center. The data center is built in increments like the container, but the process allows for a degree of customization of power and cooling system choices and building layout. The modules are connected to a central spine containing "people spaces," while the power and cooling equipment is located adjacent to the data center modules. Expansion is accomplished by placing additional modules like building blocks, including the required power and cooling sources.
3. Traditional data center: Design philosophies integrating modularity can also be applied to traditional brick-and-mortar facilities. However, to achieve effective modularity, tactics are required that diverge from the traditional design procedures of the last three decades. The entire shell of the building must accommodate space for future data center growth. The infrastructure area needs to be carefully planned to ensure sufficient space for future installation of power and cooling equipment. Also, the central plant will need to continue to operate and support the IT loads during expansion. If it is not desirable to expand within the confines of a live data center, another method is to leave space on the site for future expansion of a new data center module. This allows for an isolated construction process with tie-ins to the existing data center kept to a minimum.

3.2.2 Optimizing the Design of Modular Facilities

While we think of modular design as a solution for providing additional power and cooling equipment as the IT load increases, there might also be a power decrease or relocation that needs to be accommodated. This is where modular design provides additional benefit: an increase in energy efficiency. Using a conventional monolithic approach in the design of power and cooling systems for power in data centers will result in greater energy consumption. Looking at a modular design, the power and cooling load is spread across multiple pieces of equipment; this results in smaller equipment that can be taken on- and off-line as needed to match the IT load. This design also increases reliability because there will be redundant power and cooling modules as a part of the design. Data centers with multiple data halls, each having different reliability and functional requirements, will benefit from the use of a modular design. In this example, a monolithic approach would have difficulties in optimizing the reliability, scalability, and efficiency of the data center.

To demonstrate this idea, consider a data center that is designed to be expanded from the day-one build of one data hall to a total of three data halls. To achieve concurrent maintainability, the power and cooling systems will be designed to an $N+2$ topology. To optimize the system design and equipment selection, the operating efficiencies of the

electrical distribution system and the chiller equipment are required to determine accurate power demand at four points: 25, 50, 75, and 100% of total operating capacity. The following parameters are to be used in the analysis:

1. Electrical/UPS system: For the purposes of the analysis, a double conversion UPS was used. The unloading curves were generated using a three-parameter analysis model and capacities defined in accordance with the European Commission "Code of Conduct on Energy Efficiency and Quality of AC Uninterruptible Power Systems (UPS)." The system was analyzed at 25, 50, 75, and 100% of total IT load.
2. Chillers: Water-cooled chillers were modeled using the ASHRAE minimum energy requirements (for kilowatt per ton) and a biquadratic-in-ratio-and-DT equation for modeling the compressor power consumption. The system was analyzed at 25, 50, 75, and 100% of total IT load.

3.2.3 Analysis Approach

The goal of the analysis is to build a mathematical model defining the relationship between the electrical losses at the four loading points, comparing two system types. This same approach is used to determine the chiller energy consumption. The following two system types are the basis for the analysis:

1. Monolithic design: The approach used in this design assumes that 100% of the IT electrical requirements are covered by one monolithic system. Also, it is assumed that the monolithic system has the ability to modulate (power output or cooling capacity) to match the four loading points.
2. Modular design: This approach consists of providing four equal-sized units that correspond to the four loading points.

It is important to understand that this analysis demonstrates how to go about developing a numerical relationship between energy efficiency of a monolithic and a modular system type. There are other variables, not considered in this analysis, that will change the output and may have a significant effect on the comparison of the two system types (also see Chapter 4 "Hosting or Colocation Data Centers" in the second edition of the *Data Center Handbook*).

For the electrical system (Fig. 3.4a), the efficiency losses of a monolithic system were calculated at the four loading points. The resulting data points were then compared to the efficiency losses of four modular systems, each loaded to one-quarter of the IT load (mimicking how the power requirements increase over time). Using the modular system efficiency loss as the denominator and the efficiency losses of the monolithic system as the numerator, a multiplier was developed.

For the chillers (Fig. 3.4b), the same approach is taken, with the exception of using chiller compressor power as the indicator. A monolithic chiller system was modeled at the four loading points in order to determine the peak power at each point. Then four modular chiller systems were modeled, each at one-quarter of the IT load. Using the modular system efficiency loss as the denominator and the efficiency losses of the monolithic system as the numerator, a multiplier was developed. The electrical and chiller system multipliers can be used as an indicator during the process of optimizing energy use, expandability, first cost, and reliability.

3.3 COOLING A FLEXIBLE FACILITY

The air-conditioning system for a modular data center will typically have more equipment designed to accommodate the incremental growth of the ITE. So smaller, less capital-intensive equipment can be added over time with no disruption to the current operations. Analysis has shown that the air-conditioning systems will generally keep the upper end

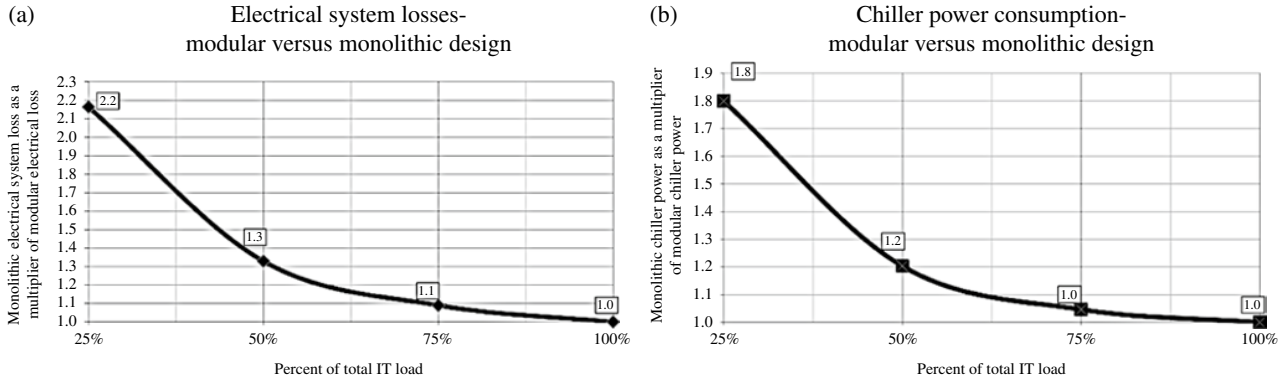


FIGURE 3.4 (a) and (b) IT load has a significant effect on electrical and cooling system between modular vs. monolithic design. Source: ©2020, Bill Kosik.

of the humidity in a reasonable range; the lower end becomes problematic, especially in mild, dry climates where there is great potential in minimizing the amount of hours that mechanical cooling is required. (When expressing moisture-level information, it is recommended to use humidity ratio or dew point temperature since these do not change relative to the dry-bulb temperature. Relative humidity (RH) will change as the dry-bulb temperature changes.)

Energy consumption in data centers is affected by many factors such as cooling system type, UPS equipment, and IT load. Air handling units designed using a modular approach can also improve introduction of outside air into the data halls in a more controlled, incremental fashion. Determining the impact on energy use from the climate is a nontrivial exercise requiring a more granular analysis technique. Using sophisticated energy modeling tools linked with multivariate analysis techniques provides the required information for geo-visualizing data center energy consumption. This is extremely useful in early concept development of a new data center giving the user powerful tools to predict approximate energy use simply by geographic siting.

As the data center design and construction industry continues to evolve and new equipment and techniques that take advantage of local climatic conditions are developed, the divergence in the PUE (power usage effectiveness) values will widen. It will be important to take this into consideration when assessing energy efficiency of data centers across a large geographic region so that facilities in less forgiving climates are not directly compared with facilities that are in climates more conducive to using energy reduction strategies. Conversely, facilities that are in the cooler climate regions should be held to a higher standard in attempting to reduce annual energy consumption and should demonstrate superior PUE values compared to the non-modular design approach.

3.3.1 Water Use in Data Centers

Studies (Pan et al. [3]) show that approximately 40% of the global population suffer from water scarcity, so managing our water resources is of utmost importance. Also, water use and energy production are inextricably linked: the water required for the thermoelectric process that generates electricity 40% to 50% of all fresh water withdrawals, even greater than the water used for irrigation. While it is outside the scope of this chapter to discuss ways of reducing water use in power plants, it is in the scope to present ideas on reducing data center energy use, which reduces the power generation requirements, ultimately reducing freshwater withdrawals. There is much more work needed on connecting the dots between cooling computers and depleting freshwater supplies; only recently the topic of data center operation impacting water consumption has become a high-priority. Unlike a commercial building, such as a corporate office or school, the greatest amount of water consumed in a data center is not the potable water used for drinking, irrigation, cleaning, or toilet flushing; it is the cooling system, namely, evaporative cooling towers and other evaporative equipment. The water gets consumed by direct evaporation into the atmosphere, by unintended water "drift" that occurs from wind carryover, and from replacing the water used for evaporation to maintain proper cleanliness levels in the water.

In addition to the water consumption that occurs at the data center (site water use), a much greater amount of water is used at the electricity generation facility (source water use) in the thermoelectrical process of making power. When analyzing locations for a facility, data center decision makers need to be well informed on this topic and understand the magnitude of how much water power plants consume, the same power plants that ultimately will provide electricity to their data center. The water use of a thermal power plant is analogous to CO_2 emissions; i.e., it is not possible for the data center owner to change or even influence the efficiency of a power plant. The environmental footprint of a data center, like any building, extends far beyond the legal boundaries of the site the data center sits on. It is vital that decisions are made with the proper data on the different types of electrical generation processes (e.g., nuclear, coal, oil, natural gas, hydroelectric) and how the cooling water is handled (recirculated or run once through). These facts, in conjunction with the power required by the ITE, will determine how much water is needed, both site and source, to support the data center. As an example, a 15-MW data center will consume between 80 and 130 million gallons annually, assuming the water consumption rate is 0.46 gallons/kWh of total data center energy use.

For the purposes of the examples shown here, averages are used to calculate the water use in gallons per megawatt-hour. (Water use discussed in this writing refers to the water used in the operation of cooling and humidification systems only.) Data comes from NREL (National Renewable Energy Laboratory) report NREL/TP-550-33905, *Consumptive Water Use for U.S. Power Production*, and *Estimating Total Power Consumption by Servers in the U.S. and the World*. It is advisable to also conduct analyses on potable water consumption for drinking, toilet/urinal flushing, irrigation, etc.

For a data center that is air-cooled (DX or direct expansion condensing units, dry coolers, air-cooled chillers), water consumption is limited to humidification. If indirect economization that uses evaporative cooling, water use consists of water that is sprayed on the heat exchanger to lower the dry-bulb temperature of the air passing through the coil. Evaporative cooling can also be used by spraying water directly into the airstream of the air handling unit (direct evaporative cooling). If the data center has water-cooled HVAC (heating, ventilation, and air conditioning) equipment, most likely some type of evaporative heat rejection (e.g., cooling tower) is being used. The operating principle of a cooling tower is fairly straightforward: Water from the facility is returned

back to the cooling tower (condenser water return or CWR) and it flows across the heat transfer surfaces, reducing the temperature of the water from evaporation. The cooler water is supplied back to the facility (condenser water supply or CWS) where it cools compressors in the main cooling equipment. It then is returned to the cooling tower.

How can we decide on where and what to build? The process includes mathematical analysis to determine preferred options:

- climate;
- HVAC system type;
- power plant water consumption rate;
- power plant GHG emissions;
- reliability;
- maintainability;
- first cost; and
- ongoing energy costs.

There are other less complex methods such as eliminating or fixing some of the variables. As an example, Table 3.1 demonstrates a parametric analysis of different HVAC system types using a diverse mix of economization techniques. When evaluated, each option includes water consumption at the site and the source. Using this type of evaluation method influences some early concepts: in some cases, when the water use at the site increases, the water used at the source (power plant) is decreased significantly This is an illustration of fixing variables as mentioned above, including climate, which has large influence on the amount of energy consumed, and the amount of water consumed. This analysis will be used as a high-level comparison, and conducting further analysis is necessary to generate thorough options to understand the trade-off between energy and water consumptions.

One more aspect is the local municipality's restrictions on water delivery and use and limitations on the amount of off-site water treatment. These are critical factors in the overall planning process, and (clearly) these need to be resolved very early in the process.

3.4 PROPER OPERATING TEMPERATURE AND HUMIDITY

Using the metaphor of water flowing through a pipe, the power and cooling distribution systems in a data center facility are located at the "end of the pipe," meaning there is little influence the HVAC systems can have on the "upstream" systems. In this metaphor, the ITE is located at the beginning of the pipe and influences everything that is "downstream." One of the design criteria for the ITE that exemplifies these ideas is the required environmental conditions (temperature and humidity) for the technology equipment located in the data center. The environmental requirements have a large impact on the overall energy use of the cooling system. If a data center is maintained at a colder temperature, the cooling equipment must work harder to maintain the required temperature. Conversely, warmer temperatures in the data center translate into less energy consumption. Analyzing the probable energy consumption of a new data center usually starts with an assessment of the thermal requirements and power demand of the ITE in the technology areas. Design dry-bulb and dew point temperatures, outside air requisites, and the supply and return temperatures will provide the data necessary for developing the first iteration of an energy analysis and subsequent recommendations to lessen the energy consumption of the cooling systems.

Most computer servers, storage devices, networking gear, etc. will come with an operating manual stating environmental conditions of 20–80% non-condensing RH and a recommended operation range of 40–55% RH. What is the difference between maximum and recommended? It has to do with prolonging the life of the equipment and avoiding failures due to electrostatic discharge (ESD) and corrosion failure that can come from out-of-range humidity levels in the facility. However, there is little, if any, industry-accepted data on what the projected service life reduction would be based on varying humidity levels. (ASHRAE's document on the subject, *2011 Thermal Guidelines for Data Processing Environments—Expanded Data Center Classes and Usage Guidance*, contains very useful information related to failure

TABLE 3.1 Different data center cooling systems that will have different electricity and water consumption

Cooling system	Economization technique	Site/source annual HVAC energy (kWh)	Site/source annual HVAC water use (gal)
Air-cooled DX	None	11,975,000	5,624,000
Air-cooled DX	Indirect evaporative cooling	7,548,000	4,566,000
Air-cooled DX	Indirect outside air	7,669,323	3,602,000
Water-cooled chillers	Water economizer	8,673,000	29,128,000
Water-cooled chillers	Direct outside air	5,532,000	2,598,000
Air-cooled chillers	Direct outside air	6,145,000	2,886,000

rates as a function of ambient temperature, but they are meant to be used as generalized guidelines only.) In conjunction with this, using outside air for cooling will reduce the power consumption of the cooling system, but with outside air come dust, dirt, and wide swings in moisture content during the course of a year. These particles can accumulate on electronic components, resulting in electrical short circuits. Also accumulation of particulate matter can alter airflow paths inside the ITE and adversely affect thermal performance. But there are data center owners/operators that can justify the cost of more frequent server failures and subsequent equipment replacement based on the reduction in energy use that comes from the use of outside air for cooling. So if a company has a planned obsolescence window for ITE of 3 years, and it is projected that maintaining higher temperatures and using outdoor air in the data center reduces the serviceable life of the ITE from 10 to 7 years, it makes sense to consider elevating the temperatures.

In order to use this type of approach, the interdependency of factors related to thermomechanical, EMC (electromagnetic compatibility), vibration, humidity, and temperature will need to be better understood. The rates of change of each of these factors, not just the steady-state conditions, will also have an impact on the failure mode. Finally, most failures occur at "interface points" and not necessarily of a component itself. Translated, this means contact points such as soldering often cause failures. So, it becomes quite the difficult task for a computer manufacturer to accurately predict distinct failure mechanisms since the computer itself is made up of many subsystems developed and tested by other manufacturers.

3.4.1 Cooling IT Equipment

When data center temperature and RH are stated in design guides, these conditions must be at the *inlet* to the computer. There are a number of legacy data centers (and many still in design) that produce air much colder than what is required by the computers. Also, the air will most often be saturated (cooled to the same value as the dew point of the air) and will require the addition of moisture in the form of humidification in order to get it back to the required conditions. This cycle is very energy intensive and does nothing to improve operation of the computers. (In defense of legacy data centers, due to the age and generational differences between ITE, airflow to the ITE is often inadequate, which causes hot spots that need to be overcome with the extra-cold air).

The use of RH as a metric in data center design is ineffective. RH changes as the dry-bulb temperature of the air changes. Wet-bulb temperature, dew point temperature, and humidity ratio are the technically correct values when performing psychrometric analysis.

What impact does all of these have on the operations of a data center? The main impact comes in the form of increased energy use, equipment cycling, and quite often simultaneous cooling/dehumidification and reheating/humidification. Discharging air at 55°F from the coils in an air handling unit is common practice in HVAC industry, especially in legacy data centers. Why? The answer is because typical room conditions for comfort cooling during the summer months are generally around 75°F and 50% RH. The dew point at these conditions is 55°F, so the air will be delivered to the conditioned space at 55°F. The air warms up (typically 20°F) due to the sensible heat load in the conditioned space and is returned to the air handling unit. It will then be mixed with warmer, more humid outside air, and then it is sent back to flow over the cooling coil. The air is then cooled and dried to a comfortable level for human occupants and supplied back to the conditioned space. While this works pretty well for office buildings, this design tactic does not transfer to data center design.

Using this same process description for an efficient data center cooling application, it would be modified as follows: Since the air being supplied to the computer equipment needs to be (as an example) 78°F and 40% RH, the air being delivered to the conditioned space would be able to range from 73 to 75°F, accounting for safety margins due to unexpected mixing of air resulting from improper air management techniques. (The air temperature could be higher with strict airflow management using enclosed cold aisles or cabinets that have provisions for internal thermal management.) The air warms up (typically 20–40°F) due to the sensible heat load in the conditioned space and is returned to the air handling unit. (Although the discharge temperature of the computer is not of concern to the computer's performance, high discharge temperatures need to be carefully analyzed to prevent thermal runaway during a loss of cooling as well as the effects of the high temperatures on the data center operators when working behind the equipment.) It will then be mixed with warmer, more humid outside air, and then it is sent back to flow over the cooling coil (or there is a separate air handling unit for supplying outside air). The air is then cooled down and returned to the conditioned space.

What is the difference in these two examples? All else being equal, the total air-conditioning load in the two examples will be the same. However, the power used by the central cooling equipment in the first case will be close to 50% greater than that of the second. This is due to the fact that much more energy needed to produce 55°F air versus 75°F air (see Section 3.5.3). Also, if higher supply air temperatures are used, the hours for using outdoor air for either air economizer or water economizer can be extended significantly. This includes the use of more humid air that would normally be below the dew point of the coil using 55°F discharge air. Similarly, if the RH or humidity ratio requirements were lowered, in cool and dry climates that are ideal for using outside air for cooling, more hours of the year could be used to reduce the load on the central cooling system without

having to add moisture to the airstream. Careful analysis and implementation of the temperature and humidity levels in the data center are critical to minimize energy consumption of the cooling systems.

3.5 AVOIDING COMMON PLANNING ERRORS

When constructing or retrofitting a data center facility, there is a small window of opportunity at the beginning of the project to make decisions that can impact long-term energy use, either positively or negatively. To gain an understanding of the best optimization strategies, there are some highly effective analysis techniques available, ensuring you're leaving a legacy of energy efficiency. Since the goal is to achieve an optimal solution, when the design concepts for cooling equipment and systems are not yet finalized, this is the perfect time to analyze, challenge, and refine system design requirements to minimize energy consumption attributable to cooling. (It is most effective if this is accomplished in the early design phases of a data center build or upgrade.)

Energy is not the only criterion that will influence the final design scheme, and other conditions will affect energy usage in the data center: location, reliability level, system topology, and equipment type, among others. There is danger in being myopic when considering design alternatives. Remember cooling systems by design are dynamic and, based on the state of other systems, will continuously adjust and course-correct to maintain the proper indoor environment. Having a full understanding of the interplay that exists between seemingly unrelated factors will enable a decision-making process that is accurate and defendable. As an example, there are a number of scenarios that, if not properly analyzed and understood, could create inefficiencies, possibly significant. These are:

- Scenario #1: Location of Facility Encumbers Energy Use.
- Scenario #2: Cooling System Mismatched with Location.
- Scenario #3: Data Center Is Way Too Cold.
- Scenario #4: Low IT Loads Not Considered in Cooling System Efficiency.
- Scenario #5: Lack of Understanding of How IT Equipment Energy Is Impacted by the Cooling System.

3.5.1 Scenario #1: Impacts of Climate on Energy Use

Climate use is just one of dozens of parameters that impacts energy use in the data center. Also considering the cost of electricity and types of the local power generation source fuel, a thorough analysis will provide a much more granular view of both environmental impacts and long-term energy costs. Without this analysis there is a risk of mismatching the cooling strategy to the local climate. True there are certain cooling systems that show little sensitivity in energy use to different climates; these are primarily ones that don't use an economization cycle. The good news is that there are several cooling strategies that will perform much better in some climates than others and there are some that perform well in many climates. A good demonstration of how climate impacts energy use comes by estimating data center energy use for the same hypothetical data center with the same power and efficiency parameters located in quite different climates (see Figs. 3.5 and 3.6).

In this analysis, where the only difference between the two alternates is the location of the data center, there are

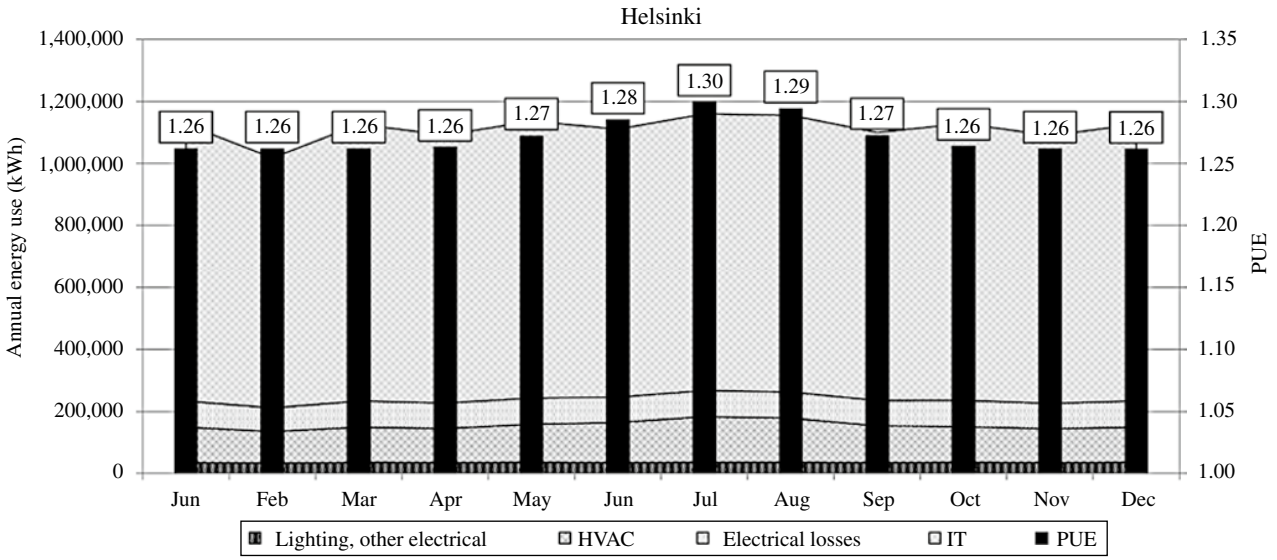


FIGURE 3.5 Monthly data center energy use and PUE for Helsinki, Finland. Source: ©2020, Bill Kosik.

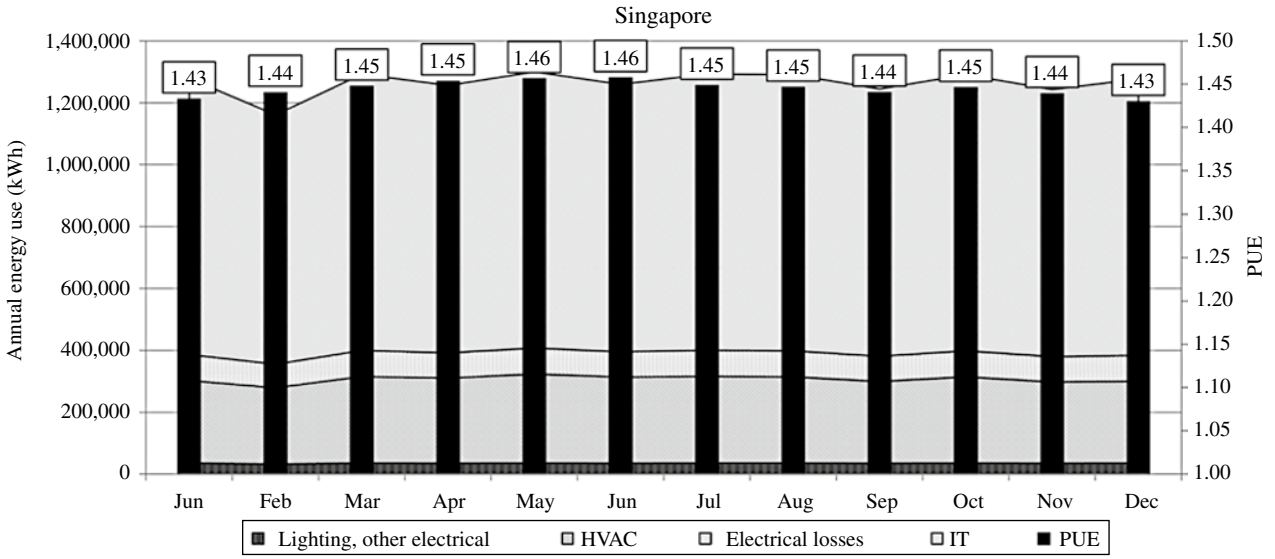


FIGURE 3.6 Monthly data center energy use and PUE for Singapore. Source: ©2020, Bill Kosik.

marked differences in annual energy consumption and PUE. It is clear that climate plays a huge role in in the energy consumption of HVAC equipment and making a good decision on the location of the data center will have long-term positive impacts.

3.5.2 Scenario #2: Establishing Preliminary PUE Without Considering Electrical System Losses

It is not unusual that data center electrical system losses attributable to the transformation and distribution of electricity could be equal to the energy consumed by the cooling system fans and pumps. Obviously losses of that magnitude will have a considerable effect on the overall energy costs and PUE. That is why it is equally important to pay close attention to the design direction of the electrical system along with the other systems. Reliability of the electrical system has a direct impact on energy use. As reliability increases, generally energy use also increases. Why does this happen? One part of increasing reliability in electrical systems is the use of redundant equipment [switchgear, UPS, PDU (power distribution unit), etc.] (see Fig. 3.7a and b). Depending on the system architecture, the redundant equipment will be online but operating at very low loads. For facilities requiring very high uptime, it is possible reliability will outweigh energy efficiency—but it will come at a high cost. This is why in the last 10–15 years manufacturers of power and cooling equipment have really transformed the market by developing products specifically for data centers. One example is new UPS technology that has very high efficiencies even at low loads.

3.5.3 Scenario #3: Data Center Is Too Cold

The second law of thermodynamics tells us that heat cannot spontaneously flow from a colder area to a hotter one; work is required to achieve this. It also holds true that the colder the area is, the more work is required to keep it cold. So, the colder the data center is, the cooling system uses more energy to do its job (Fig. 3.8). Conversely, the warmer the data center, the less energy is consumed. But this is just the half of it—the warmer the set point in the data center, the greater the amount of time the economizer will run. This means the energy-hungry compressorized cooling equipment will run at reduced capacity or not at all during times of economization.

3.5.4 Scenario #4: Impact of Reduced IT Workloads Not Anticipated

A PUE of a well-designed facility humming along at 100% load can look really great. But this operating state will rarely occur. At move-in or when IT loads fluctuates, things suddenly don't look so good. PUE states how efficiently a given IT load is supported by the facility's cooling and power systems. The facility will always have base level energy consumption (people, lighting, other power, etc.) even if the ITE is running at very low levels. Plug these conditions into the formula for PUE and what do you get? A metrics nightmare. PUEs will easily exceed 10.0 at extremely low IT loads and will still be 5.0 or more at 10%. Not until 20–30% will the PUE start resembling a number we can be proud of. So the lesson here is to be

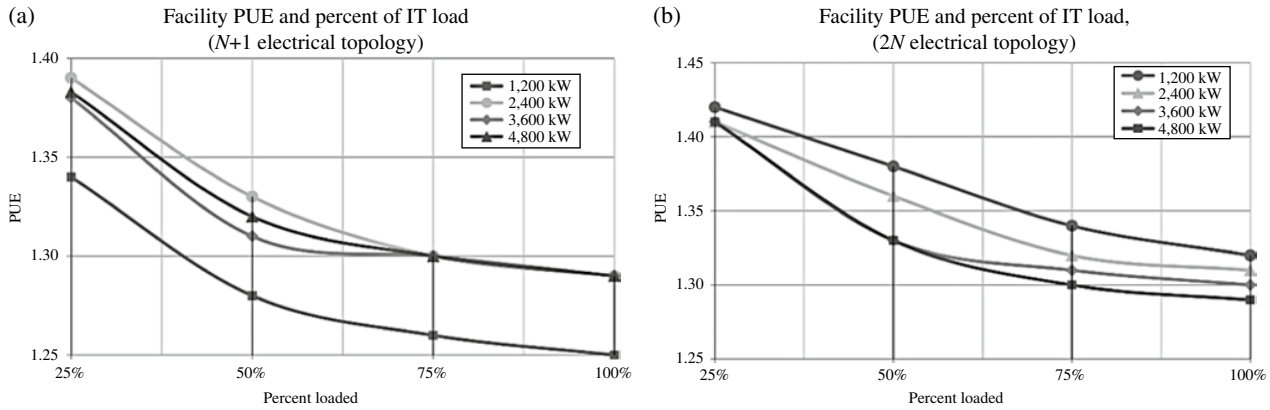


FIGURE 3.7 (a) and (b) Electrical system topology and percent of total IT load will impact overall data center PUE. In this example a scalable electrical system starting at 1,200 kW and growing to 4,800 kW is analyzed. The efficiencies vary by total electrical load as well as percent of installed IT load. Source: ©2020, Bill Kosik.

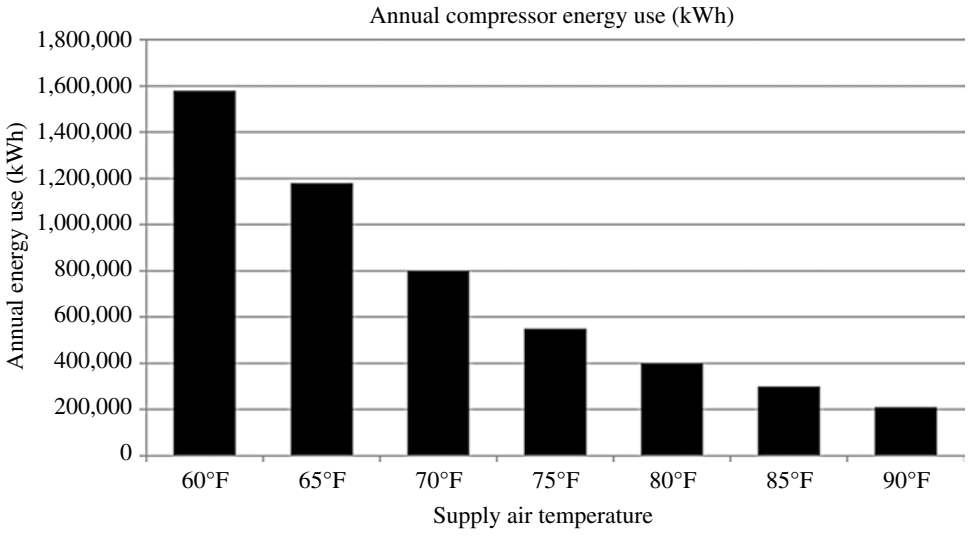


FIGURE 3.8 As supply air temperature increases, power for air-conditioning compressors decreases. Source: ©2020, Bill Kosik.

careful when predicting PUE values and considering the time frame when the estimated PUE can be achieved is presented (see Fig. 3.9).

3.5.5 Scenario #5: Not Calculating Cooling System Effects on ITE Energy

The ASHRAE TC 9.9 thermal guidelines for data centers presents expanded environmental criteria depending on the server class that is being considered for the data center. Since there are many different types of IT servers, storage, and networking equipment, the details are important here. With regard to ITE energy use, there is a point at the lower end of the range (typically 65°F) at which the energy use of a server will level out and use the same amount of energy no matter how cold the ambient temperature gets. Then there is a wide band where the temperature can fluctuate with little impact on server energy use (but a big impact on cooling system energy use—see Section 3.5.4). This band is typically 65–80°F, where most data centers currently operate. Above 80°F things start to get interesting. Depending on the age and type, server fan energy consumption will start to increase beyond 80°F and will start to become a significant part of the overall IT power consumption (as compared to the server's minimum energy consumption). The good news is that ITE manufacturers have responded to this by designing servers that can tolerate higher temperatures, no longer inhibiting high temperature data center design (Fig. 3.10).

Planning, designing, building, and operating a data center requires a lot of cooperation among the various constituents on the project team. Data centers have lots of moving parts and pieces, both literally and figuratively. This requires a dynamic decision-making process that is fed with the best

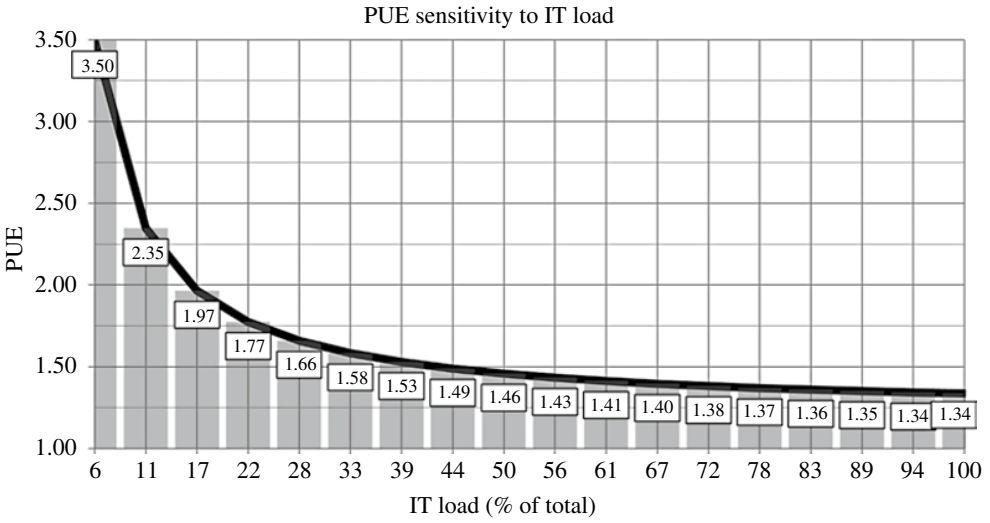


FIGURE 3.9 At very low IT loads, PUE can be very high. This is common when the facility first opens, and the IT equipment is not fully installed. Source: ©2020, Bill Kosik.

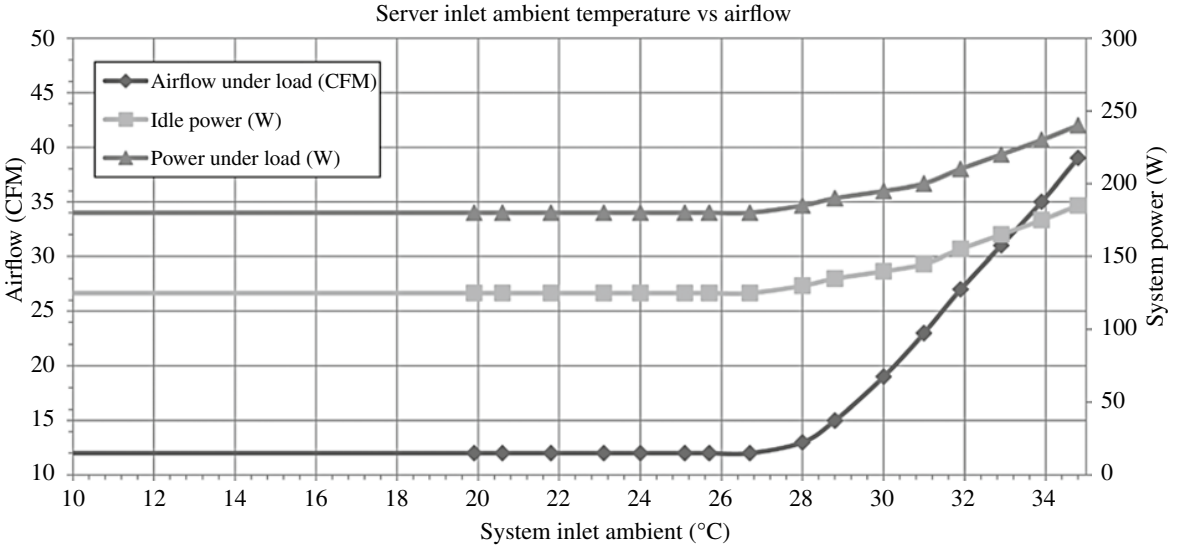


FIGURE 3.10 As server inlet temperatures increase, the overall server power will increase. Source: ©2020, Bill Kosik.

information available, so the project can continue to move forward. The key element is linking the IT and power and cooling domains, so there is an ongoing dialog about optimizing not one domain or the other, but both simultaneously. This is another area that has significantly improved.

3.6 DESIGN CONCEPTS FOR DATA CENTER COOLING SYSTEMS

In a data center, the energy consumption of the HVAC system is dependent on three main factors: outdoor conditions (temperature and humidity), the use of economization strategies, and the primary type of cooling.

3.6.1 Energy Consumption Considerations for Data Center Cooling Systems

While there a several variables that drive cooling system energy efficiency in data centers, there are factors that should be analyzed early in the design process to validate that the design is moving in the right direction:

1. The HVAC energy consumption is closely related to the outdoor temperature and humidity levels. In simple terms the HVAC equipment takes the heat from the data center and transfers it outdoors. At higher temperature and humidity levels, more work is required of the compressors to cool the

air temperature to the required levels in the data center.

2. Economization for HVAC systems is a process in which the outdoor conditions allow for reduced compressor power (or even allowing for complete shutdown of the compressors). This is achieved by supplying cool air directly to the data center (direct air economizer) or, as in water-cooled systems, cooling the water and then using the cool water in place of chilled water that would normally be created using compressors.
3. Different HVAC system types have different levels of energy consumption. And the different types of systems will perform differently in different climates. As an example, in hot and dry climates, water-cooled equipment generally consumes less energy than air-cooled systems. Conversely in cooled climates that have higher moisture levels, air-cool equipment will use less energy. The maintenance and operation of the systems will also impact energy. Ultimately, the supply air temperature and allowable humidity levels in the data center will have an influence on the annual energy consumption.

3.6.2 Transforming Data Center Cooling Concepts

To the casual observer, cooling systems for data centers have not changed a whole lot in the last 20 years. What is not obvious, however, is the foundational transformation in data center cooling resulting in innovative solutions and new ways of thinking. Another aspect is consensus-driven industry guidelines on data center temperature and moisture content. These guidelines gave data center owners, computer manufacturers, and engineers a clear path forward on the way data centers are cooled; the formal adoption of these guidelines gave the green light to many new innovative equipment and design ideas. It must be recognized that during this time, some data center owners were ahead of the game, installing never-before-used cooling systems; these companies are the vanguards in the transformation and remain valuable sources for case studies and technical information, keeping the industry moving forward and developing energy-efficient cooling systems.

3.6.3 Aspects of Central Cooling Plants for Data Centers

Generally, a central plant consists of primary equipment such as chillers and cooling towers, piping, pumps, heat exchangers, and water treatment systems. Facility size, growth plans, efficiency, reliability, and redundancy are used to determine if a central energy plant makes sense. Broadly speaking, central plants consist of centrally located equipment, generating chilled water or condenser water that is distributed to remote air handling units or CRAHs. The decision to use a central plant can be made for many different reasons, but generally central plants are best suited for large data centers and have the capability for future expansion.

3.6.4 Examples of Central Cooling Plants

Another facet to be considered is the location of the data center. Central plant equipment will normally have integrated economization controls and equipment, automatically operating based on certain operational aspects of the HVAC system and outside temperature and moisture. For a central plant that includes evaporative cooling, locations that have many hours where the outdoor wet-bulb temperature is lower than the water being cooled will reduce energy use of the central plant equipment. Economization strategies can't be examined in isolation; they need to be included in the overall discussion of central plant design.

3.6.4.1 Water-Cooled Plant Equipment

Chilled water plants include chillers (either air- or water-cooled) and cooling towers (when using water cooled chillers). These types of cooling plants are complex in design and operation but can yield superior energy efficiency. Some of the current highly efficient water-cooled chillers offer power usage that can be 50% less than legacy models.

3.6.4.2 Air-Cooled Plant Equipment

Like the water-cooled chiller plant, the air-cooled chiller plant can be complex, yet efficient. Depending on the climate, the chiller may use more energy annually than a comparably sized water-cooled chiller. To minimize this, manufacturers offer economizer modules built into the chiller that use the cold outside air to extract heat from the chilled water without using compressors. Dry coolers or evaporative coolers are also used to precool the return water back to the chiller.

3.6.4.3 Direct Expansion (DX) Equipment

DX systems have the least amount of moving parts since both the condenser and evaporator use air as the heat transfer medium, not water. This reduces the complexity, but it also can reduce the efficiency. A variation on this system is to water cool the condenser that improves the efficiency. Water-cooled computer room air-conditioning (CRAC) units fall into this category. There have been many significant developments in DX efficiency.

3.6.4.4 Evaporative Cooling Systems

When air is exposed to water spray, the dry-bulb temperature of the air will be reduced close to the wet-bulb temperature

of the air. This is the principle behind evaporative cooling. The difference between the dry bulb and wet bulb of the air is known as the wet-bulb depression. In climates that are dry, evaporative cooling works well, because the wet-bulb depression is large, enabling the evaporative process to lower the dry-bulb temperature significantly. Evaporative cooling can be used in conjunction with any of the cooling techniques outlined above.

3.6.4.5 Water Economization

Water can be used for many purposes in cooling a data center. It can be chilled via a vapor compression cycle and sent out to the terminal cooling equipment. It can also be cooled using an atmospheric cooling tower using the same principles of evaporation and used to cool compressors, or, if it is cold enough, it can be sent directly to the terminal cooling devices. The goal of a water economization, similar to direct air economization, is to use mechanical cooling as little as possible and rely on the outdoor air conditions to cool the water to the required temperature. When the system is in economizer mode, air handling unit fans, chilled water pumps, and condenser water pumps still need to operate. The energy required to run these pieces of equipment should be examined carefully to ensure that the savings that stem from the use of water economizer will not be negated by excessively high fan and pump motor energy consumption.

3.6.4.6 Direct Economization

A cooling system using direct economization (sometimes called "free" cooling) takes outside air directly to condition the data center without the use of heat exchangers. There is no intermediate heat transfer process, so the temperature outdoors is essentially the same as what is supplied to the data center. As the need lessens for the outdoor air based on indoor temperatures, the economization controls will begin to mix the outdoor air with the return air from the data center to maintain the required supply air temperature. When the outdoor temperature is no longer able to cool the data center, the economizer will completely close off the outdoor air, except for ventilation and pressurization requirements. During certain times, partial economization is achievable, where some of the outdoor air is being used for cooling, but supplemental mechanical cooling is necessary. For many climates, it is possible to run direct air economization year-round with little or no supplemental cooling. There are climates where the outdoor dry-bulb temperature is suitable for economization, but the outdoor moisture level is too high. In this case a control strategy must be in place to take advantage of the acceptable dry-bulb temperature without risking condensation or unintentionally incurring higher energy costs.

3.6.4.7 Indirect Economization

Indirect economization is used when it is not possible to use air directly from the outdoors for free cooling. Indirect economization uses the same principles as the direct outdoor air systems, but there are considerable differences in the system design and air handling equipment: in direct systems, the outdoor air is used to cool the return air by physically mixing the two airstreams. When indirect economization is used, the outdoor air is used to cool down a heat exchanger that indirectly cools the return air with no contact of the two airstreams. In indirect evaporative systems, water is sprayed on a portion of the outdoor air heat exchanger. The evaporation lowers the temperature of the heat exchanger, thereby reducing the temperature of the outdoor air. These systems are highly effective in a many climates worldwide, even humid climates. The power budget must take into consideration that indirect evaporative systems rely on a fan that draws the outside air across the heat exchanger. (This is referred to as a scavenger fan.) The scavenger fan motor power is not trivial and needs to be accounted for in estimating energy use.

3.6.4.8 Heat Exchanger Options

There are several different approaches and technology available when designing an economization system. For indirect economizer systems, heat exchanger technology varies widely:

- A rotary heat exchanger, also known as a heat wheel, uses thermal mass to cool down return air as it passes over the surface of a slowly rotating wheel. At the same time, outside air passes over the opposite side of the wheel. These two processes are separated in airtight compartments within an air handling unit to avoid cross contamination of the two airstreams.
- In a fixed crossflow heat exchanger, the two airstreams are separated and flow through two sides of the heat exchanger. Thee crossflow configuration maximizes heat transfer between the two airstreams.
- Heat pipe technology uses a continuous cycle of evaporation and condensation as the two airstreams flow across the heat pipe coil. Outside air flows across the condenser and return air at the evaporator.

Within these options there are several sub-options that will be driven by the specific application, which will ultimately inform the design strategy for the entire cooling system.

3.7 BUILDING ENVELOPE AND ENERGY USE

Buildings leak air. This leakage can have a significant impact on indoor temperature and humidity and must be

TABLE 3.2 Example of how building envelope cooling changes as a percent of total cooling load

Percent of computer equipment running (%)	Envelope losses as a percent of total cooling requirements (%)
20	8.2
40	4.1
60	2.8
80	2.1
100	1.7

Source: ©2020, Bill Kosik.

accounted for in the design process. Engineers who design HVAC systems for data centers generally understand that computers require an environment where temperature and humidity are maintained in accordance with the ASHRAE guidelines, computer manufacturers' recommendations, and the owner's requirements.

Maintaining temperature and humidity for 8,760 h/year is very energy intensive. This is one of the factors that continues to drive research on HVAC system energy efficiency. However, it seems data center industry has done little research on the *building* that houses the ITE and how it affects the temperature, humidity, and energy in the data center. There are fundamental questions that need to answered in order to gain a better understanding of the building:

1. Does the amount of leakage across the building envelope correlate to indoor humidity levels and energy use?
2. How does the climate where the data center is located affect the indoor temperature and humidity levels?
3. Are certain climates more favorable for using outside air economizer without using humidification to add moisture to the air during the times of the year when outdoor air is dry?
4. Will widening the humidity tolerances required by the computers produce worthwhile energy savings?

3.7.1 Building Envelope Effects

The building envelope is made up of the roof, exterior walls, floors, and underground walls in contact with the earth, windows, and doors. Many data center facilities have minimal amounts of windows and doors, so the remaining components are the roof, walls, and floors which need to analyzed for heat transfer and infiltration. Each of these systems have different performance characteristics; using energy modeling will help in assessing how these characteristics impact energy use. Thermal resistance (insulation), thermal mass (heavy construction such as concrete versus lightweight steel), airtightness, and moisture permeability are some of the properties that are important to understand.

3.7.1.1 Building Envelope and Energy Use

When a large data center is running at full capacity, the effects of a well-constructed building envelope on energy use (as a percent of the total) are negligible. However, when a data center is running at exceptionally low loads, the energy impact of the envelope (on a percentage basis) is much more considerable. Generally, the envelope losses start out as a significant component of the overall cooling load but decrease over time as the computer load becomes a greater portion of the total load (Table 3.2).

The ASHRAE Energy Standard 90.1 has specific information on different building envelope alternatives that can be used to meet the minimum energy performance requirements. Additionally, the ASHRAE publication *Advanced Energy Design Guide for Small Office Buildings* provides valuable details on the most effective strategies for building envelopes, categorized by climatic zone. Finally, another good source of engineering data is the CIBSE Guide A on Environmental Design. There is one thing to take into consideration specific to data centers: based on the reliability and survivability criteria, exterior systems such as exterior walls, roof, windows, louvers, etc. will be constructed to very strict standards that will survive through extreme weather events such as tornados, hurricanes, floods, etc.

3.7.2 Building Envelope Leakage

Building leakage will impact the internal temperature and RH by outside air infiltration and moisture migration. Depending on the climate, building leakage can negatively impact both the energy use of the facility and the indoor moisture content of the air. Based on several studies from the National Institute of Standards and Technology (NIST), Chartered Institution of Building Services Engineers (CIBSE), and American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) investigating leakage in building envelope components, it is clear that often building leakage is underestimated by a significant amount. Also, there is not a consistent standard on which to base building air leakage. For example:

- CIBSE TM-23, *Testing Buildings for Air Leakage*, and the Air Tightness Testing and Measurement Association (ATTMA) TS1 recommend building air leakage rates from 0.11 to 0.33 CFM/ft^2.
- Data from Chapter 27, "Ventilation and Air Infiltration" from ASHRAE Fundamentals show rates of 0.10, 0.30, and 0.60 CFM/ft^2 for tight, average, and leaky building envelopes.
- The NIST report of over 300 existing U.S., Canadian, and U.K. buildings showed leakage rates ranging from 0.47 to 2.7 CFM/ft^2 of above-grade building envelope area.

- The ASHRAE Humidity Control Design Guide indicates that typical commercial buildings have leakage rates of 0.33–2 air changes per hour and buildings constructed in the 1980s and 1990s are not significantly tighter than those constructed in the 1950s, 1960s, and 1970s.

To what extent should the design engineer be concerned about building leakage? Using hourly simulation of a data center facility and varying the parameter of envelope leakage, it is possible to develop profiles of indoor RH and air change rate.

3.7.3 Energy Modeling to Estimate Energy Impact of Envelope

Typical analysis techniques look at peak demands or steady-state conditions that are just representative "snapshots" of data center performance. These analysis techniques, while particularly important for certain aspects of data center design such as equipment sizing or estimating energy consumption in the conceptual design phase, require more granularity to generate useful analytics on the dynamics of indoor temperature and humidity—some of the most crucial elements of successful data center operation. However, using an hourly (and sub-hourly) energy use simulation tool will yield results that provide the engineer rich detail informing solutions to optimize energy use. As an example of this, the output of the building performance simulation shows marked differences in indoor RH and air change rates when comparing different building envelope leakage rates (see Fig. 3.11). Since it is not possible to develop full-scale mock-ups to test the integrity of the building envelope, the simulation process is an invaluable tool to analyze the impact to indoor moisture content based on envelope leakage. Based on research done by the author, the following conclusions can be drawn:

- There is a high correlation between leakage rates and fluctuations in indoor RH—the greater the leakage rates, the greater the fluctuations in RH.
- There is a high correlation between leakage rates and indoor RH in the winter months—the greater the leakage rates, the lower the indoor RH.
- There is low correlation between leakage rates and indoor RH in the summer months—the indoor RH levels remain relatively unchanged even at greater leakage rates.
- There is a high correlation between building leakage rates and air change rate—the greater the leakage rates, the greater the number of air changes due to infiltration.

3.8 AIR MANAGEMENT AND CONTAINMENT STRATEGIES

Proper airflow management improves efficiency that cascades through other systems in the data center. Plus, proper airflow management will significantly reduce problems

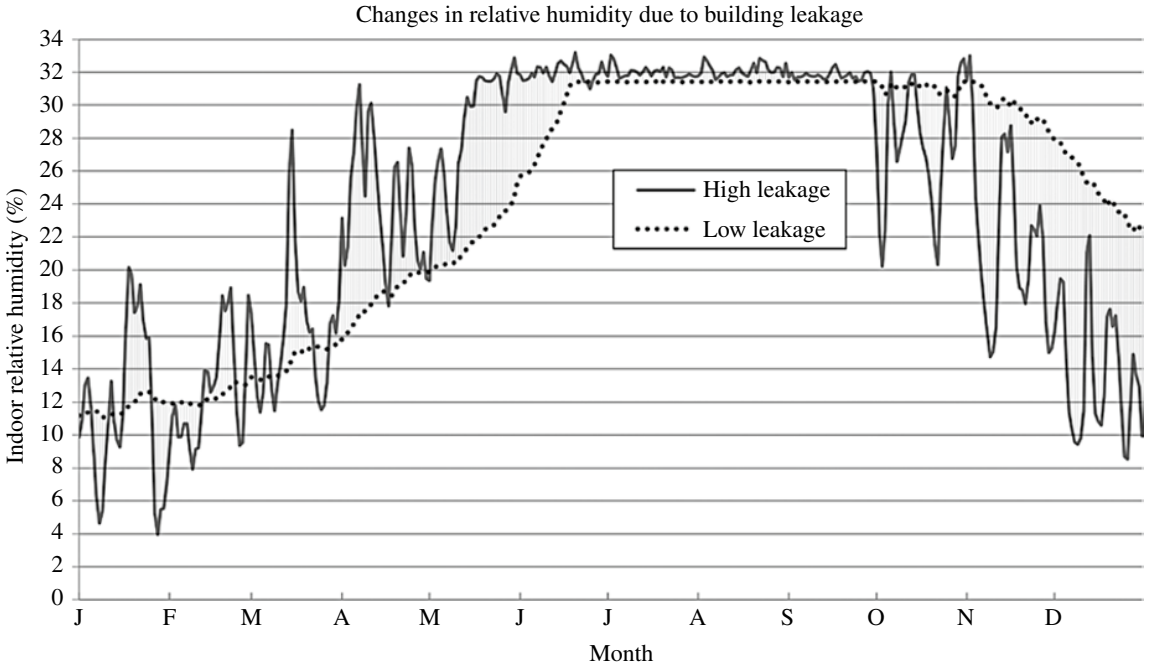


FIGURE 3.11 Internal humidity levels will correspond to outdoor moisture levels based on the amount of building leakage. Source: ©2020, Bill Kosik.

related to re-entrainment or re-circulation of hot air into the cold aisle which can lead to IT equipment shutdown due to thermal overload. Air containment creates a micro-environment with uniform temperature gradients enabling predictable conditions at the air inlets to the servers. These conditions ultimately allow for the use of increased air temperatures, which reduces the energy needed to cool the air. It also allows for an expanded window of operation for economizer use.

There are many effective remedial approaches to improve cooling effectiveness and air distribution in existing data centers. These include rearrangement of solid and perforated floor tiles, sealing openings in the raised floor, installing air dam baffles in IT cabinets to prevent air bypassing the IT gear, and other more extensive retrofits that result in pressurizing the raised floor more uniformly to ensure the air gets to where it is needed.

But arguably the most effective air management technique is the use of physical barriers to contain the air where it will be most effective. There are several approaches that give the end user options to choose from that meet the project requirements.

3.8.1 Passive Chimneys Mounted on IT Cabinets

These devices are the simplest and lowest cost of the options and have no moving parts. Depending on the IT cabinet configuration, the chimney is mounted on the top and discharges into the ceiling plenum. There are specific requirements for the cabinet, and it may not be possible to retrofit on all cabinets. Also, the chimney diameter will limit the amount of airflow from the servers, so it might be problematic to install them on higher-density cabinets.

3.8.2 Fan-Powered Chimneys Mounted on IT Cabinets

These use the same concept as the passive chimneys, but the air movement is assisted by a fan. The fan ensures a positive discharge into the ceiling plenum, but can be a point of failure and increases costs related to installation and energy use. UPS power is required if continuous operation is needed during a power failure. Though the fan-assist allows for more airflow through the chimney, it still will have limits on the amount of air that can flow through it.

3.8.3 Hot Aisle Containment

The hot aisle/cold aisle arrangement is very common and generally successful to compartmentalize the hot and cold air. Certainly, it provides benefits compared to layouts where ITE discharged hot air right into the air inlet of adjacent equipment. (Unfortunately, this circumstance still exists in many data centers with legacy equipment.) Hot aisle containment takes the hot aisle/cold aisle strategy and builds upon it. The air in the hot aisle is contained using a physical barrier that can range from the installation of a heavy plastic curtain system that is mounted at the ceiling level and terminates at the top of the IT cabinets. Other more expensive techniques used solid walls and doors that create a hot chamber that completely contains the hot air. This system is generally more applicable for new installations. The hot air is discharged into the ceiling plenum from the contained hot aisle. Since the hot air is now concentrated into a small space, worker safety needs to be considered since the temperatures can get quite high.

3.8.4 Cold Aisle Containment

While the cold aisle containment may appear to be simply a reverse of the hot aisle containment, it is more complicated in its operation. The cold aisle containment system can also be constructed from a curtain system or solid walls and doors. The difference between this and the hot aisle containment comes from the ability to manage airflow to the computers in a more granular way. When constructed out of solid components, the room can act as a pressurization chamber that will maintain the proper amount of air that is required to cool the servers by monitoring the pressure. By varying the airflow into the chamber, air handing units serving the data center are given instructions to increase or decrease air volume in order to keep the pressure in the cold aisle at a preset level. As the server fans speed up, more air is delivered; when they slow down, less is delivered. This type of containment has several benefits beyond traditional airflow management; however the design and operation are more complex.

3.8.5 Self-Contained In-Row Cooling

To tackle air management problems that are occurring in only one part of a data center, self-contained in-row cooling units are a good solution. These come in many varieties such as chilled water-cooled, air-cooled DX, low-pressure pumped refrigerant, and even CO_2-cooled. These are best applied when there is a small grouping of high-density, high-heat-generating servers that are creating difficulties for the balance of the data center. However there are many examples where entire data centers use this approach.

3.8.6 Liquid Cooling

Once required to cool large enterprise mainframe computers, water cooling decreased when microcomputers, personal computers, and then rack-mounted servers were introduced. But as processor technology and other advancements in ITE drove up power demand and the corresponding heat output of the computers, it became apparent that close-coupled or directly coupled cooling solutions were needed to remove heat from the main heat-generating

components in the computer: the CPU, memory, and the GPU. Using liquid cooling was a proven method of accomplishing this. Even after the end of the water-cooled mainframe era, companies that manufacture supercomputers were using water and refrigerant cooling in the mid-1970s. And since then, the fastest and most powerful supercomputers use some type of liquid cooling technology—it is simply not feasible to cool these high-powered computers with traditional air systems.

While liquid cooling is not strictly an airflow management strategy, it has many of the same characteristic as all-air containment systems.

- Liquid cooled computers can be located very closely to each other, without creating hot spots or re-entraining hot air from the back of the computer into the intake of an adjacent computer.
- Like computers relying on an air containment strategy, liquid-cooled computers can use higher temperature liquid, reducing energy consumption from vapor compression cooling equipment and increasing the number of hours that economizer systems will run.
- In some cases, a hot aisle/cold aisle configuration is not needed; in this case the rows of computer cabinets can be located closer together resulting in smaller data centers.

One difference with liquid cooling, however, is that the liquid may not provide 100% the cooling required. A computer like this (sometimes called a hybrid) will require air cooling for 10–30% of the total electrical load of the computer, while the liquid cooling absorbs 70% of the heat. The power requirements supercomputer equipment housed in ITE cabinets on the data center floor will vary based on the manufacturer and the nature of the computing. The equipment cabinets can have a peak demand of 60 kW to over 100 kW. Using a range of 10–30% of the total power that is not dissipated to the liquid, the heat output that will be cooled with air for liquid-cooled computing systems will range from 6 kW to over 30 kW. These are very significant cooling loads that need to be addressed and included in the air cooling design.

3.8.7 Immersion Cooling

One type of immersion cooling submerges the servers in large containers filled with dielectric fluid. The servers require some modification, but by using this type of strategy, fans are eliminated from the computers. The fluid is circulated through the container around the servers and is typically pumped to heat exchanger that is tied to outdoor heat rejection equipment. Immersion is a highly effective method of cooling—all the heat-generating components are surrounded by the liquid. (Immersion cooling is not new—it has been used in the power transformer industry for more than a century).

3.8.8 Summary

If a data center owner is considering the use of elevated supply air temperatures, some type of containment will be necessary as the margin for error (unintentional air mixing) gets smaller as the supply air temperature increases. As the use of physical air containment becomes more practical and affordable, implementing these types of energy efficiency strategies will become more feasible.

3.9 ELECTRICAL SYSTEM EFFICIENCY

In data centers, reliability and maintainability of the electrical and cooling systems are foundational design requirements to enable successful operation of the IT system. In the past, a common belief was that reliability and energy efficiency are mutually exclusive. This is no longer the case: it is possible to achieve the reliability goals and optimize energy efficiency at the same time, but it requires close collaboration among the IT and facility teams to make it happen.

The electrical distribution system in a data center includes numerous equipment and subsystems that begin at the utility entrance and building transformers, switchgear, UPS, PDUs, RPPs (remote power panels), and power supplies, ultimately powering the fans and internal components of the ITE. All of these components will have a degree of inefficiency, resulting in a conversion of the electricity into heat ("energy loss"). Some of these components have a linear response to the percent of total load they are designed to handle; others will demonstrate a very nonlinear behavior. Response to partial load conditions is an important characteristic of the electrical components; it is a key aspect when estimating overall energy consumption in a data center with varying IT loads. Also, while multiple concurrently energized power distribution paths can increase the availability (reliability) of the IT operations, this type of topology can decrease the efficiency of the overall system, especially at partial IT loads.

In order to illustrate the impacts of electrical system efficiency, there are primary factors that influence the overall electrical system performance:

1. UPS module and overall electrical distribution system efficiency
2. Part load efficiencies
3. System modularity
4. System topology (reliability)
5. Impact on cooling load

3.9.1 UPS Efficiency Curves and ITE Loading

There are many different types of UPS technologies, where some perform better at lower loads, and others are used almost exclusively for exceptionally large IT loads. The final selection of the UPS technology is dependent on the specific case. With this said, it is important to know that different UPS sizes and circuit types have different efficiency curves—it is certainly not a one-size-fits-all proposition. Each UPS type will perform differently at part load conditions, so analysis at 100, 75, 50, 25, and 0% loading is necessary to gain a complete picture of UPS and electrical system efficiency (see Fig. 3.12). At lower part load values, the higher-reliability systems (generally) will have higher overall electrical system losses as compared with a lower-reliability system. As the percent load approaches unity, the gap narrows between the two systems. The absolute losses of the high-reliability system will be 50% greater at 25% load than the regular system, but this margin drops to 23% at 100% load. When estimating annual energy consumption of a data center, it is advisable to include a schedule for the IT load that is based on the actual operational schedule of the ITE, thus providing a more accurate estimate of energy consumption. This schedule would contain the predicted weekly or daily operation, including operational hours and percent loading at each hour, of the computers (based on historic workload data), but more importantly the long-term ramp-up of the power requirements for the computers. With this type of information, planning and analysis for the overall annual energy consumption will be more precise.

3.9.2 Modularity of Electrical Systems

In addition to the UPS equipment efficiency, the modularity of the electrical system will have a large impact on the efficiency of the overall system. UPS modules are typically designed as systems, where the systems consist of multiple modules. So, within the system, there could be redundant UPS modules or there might be redundancy in the systems themselves. The ultimate topology design is primarily driven by the owner's reliability, expandability, and cost requirements. The greater the number of UPS modules, the smaller the portion of the overall load will be handled by each module. The effects of this become pronounced in high-reliability systems at low loads where it is possible to have a single UPS module working at less than 25% of its rated capacity.

Ultimately when all the UPS modules, systems, and other electrical equipment are pieced together to create a unified electrical distribution system, efficiency values at the various loading percentages are developed for the entire system. The entire system now includes all power distribution upstream and downstream of the UPS equipment. In addition to the loss incurred by the UPS equipment, losses from transformers, generators, switchgear, power distribution units (with and without static transfer switches), and distribution wiring must be accounted for. When all these components are analyzed in different system topologies, loss curves can be generated so the efficiency levels can be compared to the reliability of the system, assisting in the decision-making process. Historically, the higher the reliability, the lower the efficiency.

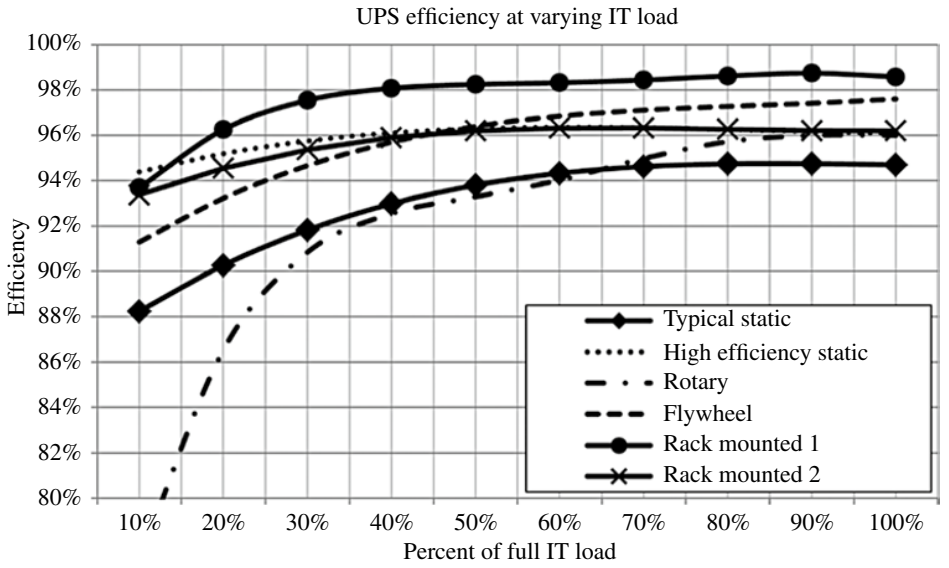


FIGURE 3.12 Example of manufacturers' data on UPS part load performance. Source: ©2020, Bill Kosik.

3.9.3 The Value of a Collaborative Design Process

Ultimately, when evaluating data center energy efficiency, it is the overall energy consumption that matters. Historically, during the conceptual design phase of a data center, it was not uncommon to develop electrical distribution and UPS system architecture separate from other systems, such as HVAC. Eventually the designs for these systems converge and were coordinated prior to the release of final construction documents. But collaboration was absent in that process, where the different disciplines would have gotten a deeper understanding of how the other discipline was approaching reliability and energy efficiency. Working as a team creates an atmosphere where the "aha" moments occur; out of this come innovative, cooperative solutions. This interactive and cooperative process produces a combined effect greater than the sum of the separate effects (synergy).

Over time, the data center design process matured, along with the fundamental understanding of how to optimize energy use and reliability. A key element of this process is working with the ITE team to gain an understanding of the anticipated IT load growth to properly design the power and cooling systems, including how the data center will grow from a modular point of view. Using energy modeling techniques, the annual energy use of the power and cooling systems is calculated based on the growth information from the ITE team. From this, the part load efficiencies of the electrical and the cooling systems (along with the ITE loading data) will determine the energy consumption that is ultimately used for powering the computers and the amount dissipated as heat.

Since the losses from the electrical systems ultimately result in heat gain (except for equipment located outdoors or in nonconditioned spaces), the mechanical engineer will need to use this data in sizing the cooling equipment and evaluating annual energy consumption. The efficiency of the cooling equipment will determine the amount of energy required to cool the electrical losses. It is essential to include cooling system energy usage resulting from electrical losses in any life cycle studies for UPS and other electrical system components. It is possible that lower-cost, lower-efficiency UPS equipment will have a higher life cycle cost from the cooling energy required, even though the capital cost may be significantly less than a high-efficiency system. In addition to the energy that is "lost," the additional cooling load resulting from the loss will negatively impact the annual energy use and PUE for the facility. The inefficiencies of the electrical system have a twofold effect on energy consumption.

3.9.4 Conclusion

Reliability and availability in the data center are of paramount importance for the center's operator. Fortunately, in recent years, the industry has responded well with myriad new products and services to help increase energy efficiency, reduce costs, and improve reliability. When planning a new data center or considering a retrofit to an existing one, the combined effect of all of the different disciplines collaborating in the overall planning and strategy for the power, cooling and IT systems result in a highly efficient and reliable plan. And using the right kind of tools and analysis techniques is an essential part of accomplishing this.

3.10 ENERGY USE OF IT EQUIPMENT

Subsequent to the release of the EPA's 2007 "EPA Report to Congress on Server and Data Center Energy Efficiency," the ongoing efforts to increase energy efficiency of servers and other ITE became urgent and more relevant. Many of the server manufacturers began to use energy efficiency as a primary platform of their marketing campaigns. Similarly, reviewing technical documentation on the server equipment, there is also greater emphasis on server energy consumption, especially at smaller workloads. Leaders in the ITE industry have been developing new transparent benchmarking criteria for ITE and data center power use. These new benchmarks are in addition to existing systems such as the US EPA's "ENERGY STAR® Program Requirements for Computer Servers" and "Standard Performance Evaluation Corporation (SPEC)." These benchmarking programs are designed to be manufacturer-agnostic, to use standardized testing and reporting criteria, and to provide clear and understandable output data for the end user.

It is clear that since 2007 when data center energy use was put in the spotlight, there have been significant improvements in energy efficiency of data centers. For example, data center energy use increased by nearly 90% from 2000 to 2005, 24% from 2005 to 2010, and 4% from 2010 to 2014. It is expected that growth rate to 2020 and beyond will hold at approximately 4%. Many of these improvements come from advances in server energy use and how software is designed to reduce energy use. And of course any reductions in energy use by the IT systems have a direct effect on energy use of the power and cooling systems.

The good news is that there is evidence, obtained through industry studies, that the energy consumption of the ITE sector is slowing significantly compared to the scenarios developed for the 2007 EPA report (Fig. 3.13). The 2016 report "United States Data Center Energy Usage Report" describes in detail the state of data center energy consumption:

1. In 2014, data centers in the United States consumed an estimated 70 billion kWh, representing about 1.8% of total U.S. electricity consumption.
2. Current study results show data center electricity consumption increased by about four percent from 2010

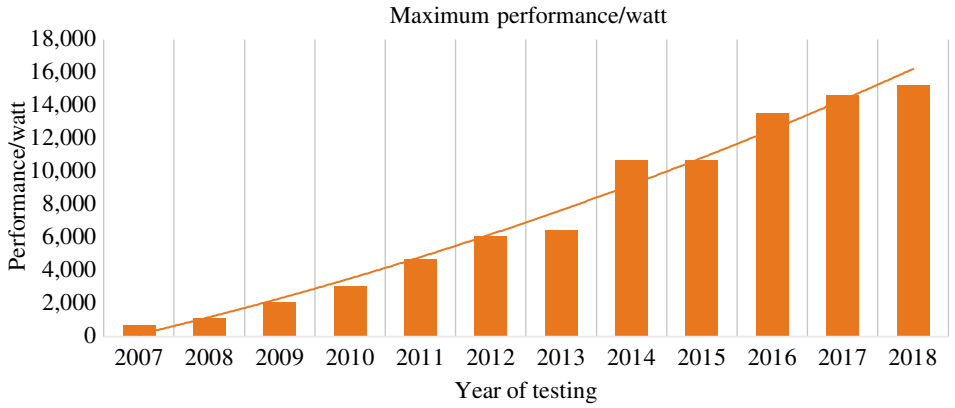


FIGURE 3.13 Since 2007, performance per watt has steadily increased. Source: ©2020, Bill Kosik.

to 2014. The initial study projected a 24% increase estimated from 2005 to 2010.

3. Servers are improving in their power scaling abilities, reducing power demand during periods of low utilization.

Although the actual energy consumption for data centers was an order of magnitude less than what was projected as a worst-case scenario by the EPA, data center energy use will continue to have strong growth. As such, it is imperative that the design of data center power and cooling systems continue to be collaborative and place emphasis on synergy and innovation.

3.10.1 U.S. Environmental Protection Agency (EPA)

The EPA has launched dozens of energy efficiency campaigns related to the built environment since the forming of the ENERGY STAR program with the U.S. Department of Energy (DOE) in 1992. The primary goal of the ENERGY STAR program is to provide unbiased information on power-consuming products and provide technical assistance in reducing energy consumption and related GHG emissions for commercial buildings and homes. Within the ENERGY STAR program, there is guidance to data center energy use. The information provided by the EPA and DOE falls into two categories:

1. Data center equipment that is ENERGY STAR certified
2. Ways to improve energy efficiency in the data center
3. Portfolio Manager

3.10.1.1 Data Center ENERGY STAR Certified Equipment

The ENERGY STAR label is one of the most recognized symbols in the United States. In addition to the hundreds of products, commercial and residential, certified by ENERGY STAR, there are products specific to data centers that have been certified by ENERGY STAR. The equipment falls into five categories:

1. Enterprise servers
2. Uninterruptible power supplies (UPS)
3. Data center storage
4. Small network equipment
5. Large network equipment

To qualify for ENERGY STAR, specific performance criteria must be met, documented, and submitted to the EPA. The EPA publishes detailed specifications on the testing methodology for the different equipment types and the overall process that must be followed to be awarded an ENERGY STAR. This procedure is a good example of how the ITE and facilities teams work in a collaborative fashion. In addition to facility-based equipment (UPS), the other products fall under the ITE umbrella. Interestingly, the servers and UPS have a similar functional test that determines energy efficiency at different loading levels. Part load efficiency is certainly a common thread running through the ITE and facilities equipment.

3.10.2 Ways to Improve Data Center Efficiency

The EPA and DOE have many "how-to" documents for reducing energy use in the built environment. The DOE's Building Technology Office (BTO) conducts regulatory activities including technology research, validation, implementation, and review, some of which are manifest in technical documents on reducing energy use in commercial buildings. Since many of these documents apply mainly to commercial buildings, the EPA has published documents specific to data centers to address systems and equipment that are only found in data centers. As an example, the EPA has a document on going after the "low-hanging fruit" (items that do not require capital funding that will reduce energy

use immediately after completion). This type of documentation is very valuable to assist data center owners in lowing their overall energy use footprint.

3.10.3 Portfolio Manager

The EPA's Portfolio Manager is a very large database containing commercial building energy consumption. But it is not a static repository of data—it is meant to be a benchmarking tool on the energy performance of similar buildings. Comparisons are made using different filters, such as building type, size, etc. As of this writing, 40% of all U.S. commercial buildings have been benchmarked in Portfolio Manager. This quantity of buildings is ideal for valid benchmarking.

3.10.4 SPECpower_ssj2008

The SPEC has designed SPECpower_ssj2008 as a benchmarking tool for server performance and a means of determining power requirements at partial workloads. Using the SPEC data, curves representing server efficiency are established at four workload levels (100, 75, 50, and 25%). When the resulting curves are analyzed, it becomes clear that the computers continue to improve their compute-power-to-electrical-power ratios, year over year.

Reviewing the data, we see that the ratio of the minimum to maximum power states has decreased from over 60% to just under 30% (Fig. 3.14). This means that at a data center level, if all the servers were in an idle state, in 2007 the running IT load would be 60% of the total IT load, while in 2013, it would be under 30%. This trickles down to the cooling and power systems consuming even more energy. Clearly this is a case for employing aggressive power management strategies in existing equipment and evaluating server equipment energy efficiency when planning an IT refresh.

The supercomputing community has developed a standardized ranking technique, since the processing ability of these types of computers is different than that of enterprise servers than run applications using greatly different amounts of processing power. The metric that is used is megaFLOPS per watt, which obtained by running a very prescriptive test using a standardized software package (HPL). This allows for a very fair head-to-head energy efficiency comparison of different computing platforms.

Since equipment manufacturers submit their server performance characteristics directly to SPEC using specific testing protocol, the SPEC database continues to grow in its wealth of performance information. Also, using the metric performance vs. power normalizes the different manufacturers' equipment by comparing power demand at different loading points and the computing performance.

3.11 SERVER VIRTUALIZATION

Studies have shown that the average enterprise server will typically have a utilization of 20% or less, with the majority being less than 10%. The principal method to reduce server energy consumption starts with using more effective equipment, which uses efficient power supplies and supports more efficient processor and memory. Second, reducing (physically or virtually) the number of servers that are required to run a given workload will reduce the overall power demand. Coupling these two approaches together with a robust power management protocol will ensure that when the servers are in operation, they are running as efficiently as possible.

It is important to understand the potential energy reduction from using virtualization and power management strategies. To demonstrate this, a 1,000-kW data center with an average of 20% utilization was modeled with 100% of the IT load attributable to compute servers. Applying power management to 20% of the servers will result in a 10% reduction

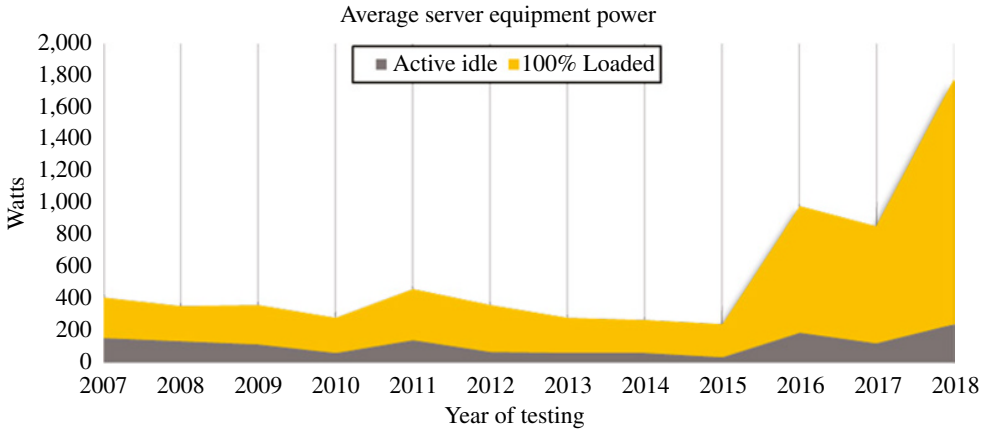


FIGURE 3.14 Servers have a much greater ratio of full load power to no load power (active idle); this equates to a lower energy consumption when the computers are idling. Source: ©2020, Bill Kosik.

TABLE 3.3 Analysis showing the impact on energy use from using power management, virtualization, and increased utilization

	Server energy (kWh)	Power and cooling energy (kWh)	Total annual energy consumption (kWh)	Reduction from base case (%)	Annual electricity expense reduction (based on $0.10/kWh)
Base case	5,452,000	1,746,523	7,198,523	Base	Base
Scenario 1: Power management	4,907,000	1,572,736	6,479,736	10%	$71,879
Scenario 2: Virtualization	3,987,000	1,278,052	5,265,052	27%	$121,468
Scenario 3: Increased Utilization	2,464,000	789,483	3,253,483	55%	$201,157

Source: ©2020, Bill Kosik.

in annual energy attributable to the servers. Virtualizing the remaining servers with a 4:1 ratio will reduce the energy another 4% to a total of 14%. Increasing the utilization of the physical servers from 20 to 40% will result in a final total annual energy reduction of 26% from the base. These might be considered modest changes in utilization and virtualization, but at 10 cents/kWh, these changes would save over $130,000/year. And this is only for the electricity for the servers, not the cooling energy and electrical system losses (see Table 3.3).

Average of all servers measured–average utilization = 7.9%.
Busiest server measured–average utilization = 16.9%.

Including the reduction in cooling energy and electrical losses in scenario 1, the consumption is reduced from 1,711,000 to 1,598,000 kWh, or 113,000 kWh/year. Further reduction for scenario 2 brings the total down to 1,528,000 kWh, which is an additional 70,000 kWh annually. Finally, for the scenario 3, the total annual energy for the power and cooling systems is further reduced to 1,483,000 kWh, or 45,000 kWh less than scenario 2 (see Figs. 3.15, 3.16, and 3.17).

3.12 INTERDEPENDENCY OF SUPPLY AIR TEMPERATURE AND ITE ENERGY USE

One aspect that demonstrates the interdependency between the ITE and the power and cooling systems is the temperature of the air delivered to the computers for cooling. A basic design tenet is to design for the highest internal air temperature allowable that will still safely cool the computer equipment and not cause the computers' internal fans to run at excessive speeds. The ASHRAE temperature and humidity guidelines for data centers recommend an upper dry-bulb limit of 80°F for the air used to cool the computers. If this

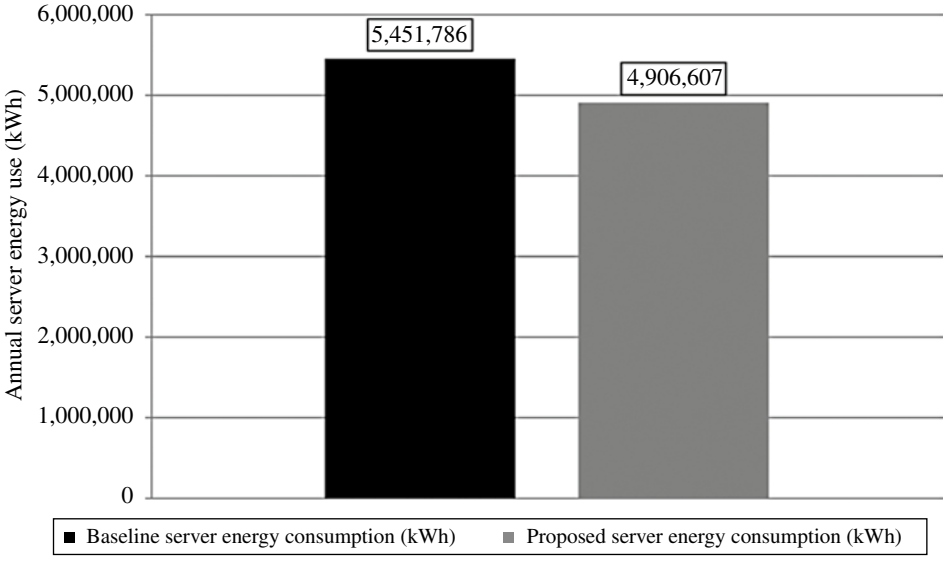


FIGURE 3.15 Energy use reduction by implementing server power management strategies. Source: ©2020, Bill Kosik.

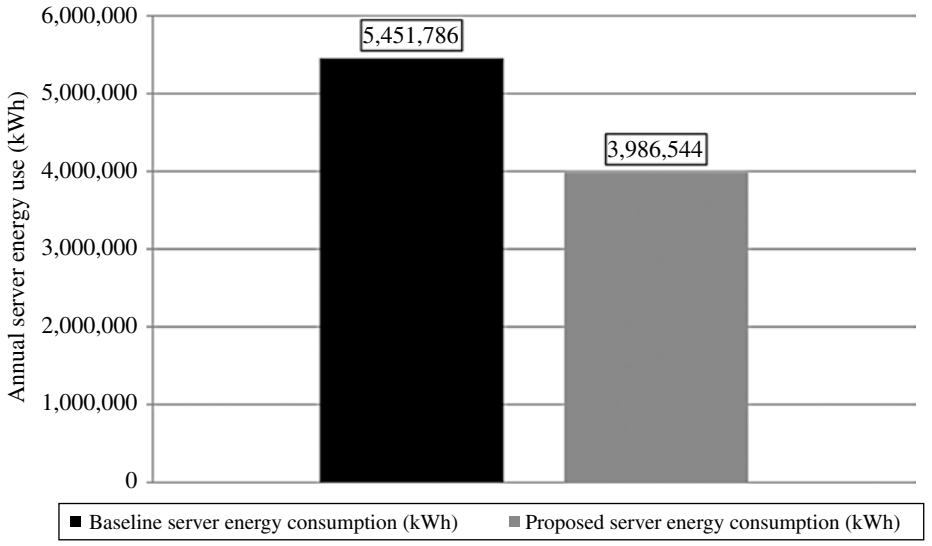


FIGURE 3.16 Energy use reduction by implementing server power management strategies and virtualizing servers. Source: ©2020, Bill Kosik.

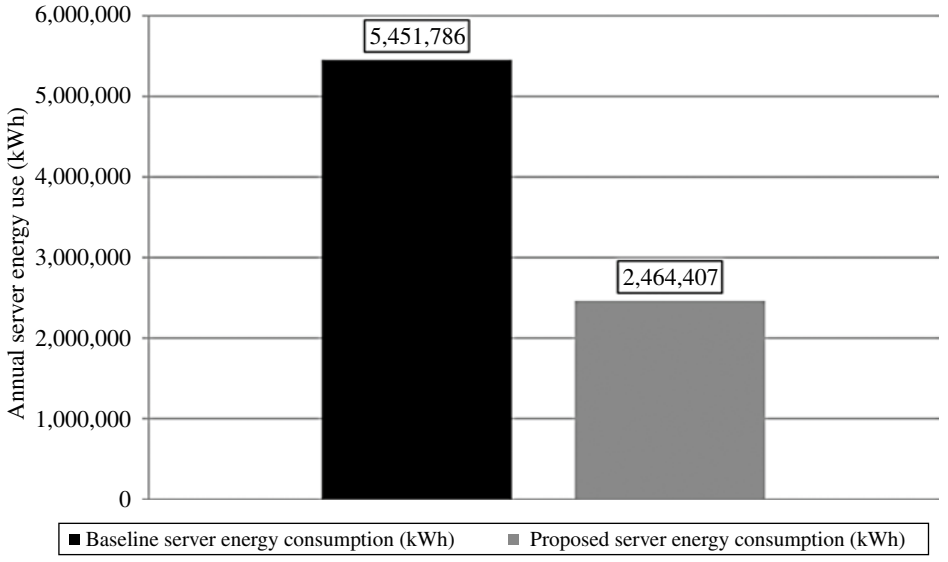


FIGURE 3.17 Energy use reduction by implementing server power management strategies, virtualizing servers, and increasing utilization. Source: ©2020, Bill Kosik.

temperature (or even higher) is used, the hours for economization will be increased. And when vapor compression (mechanical) cooling is used, the elevated temperatures will result in lower compressor power. (The climate in which the data center is located will drive the number of hours that are useful for using air economizer.)

3.13 IT AND FACILITIES WORKING TOGETHER TO REDUCE ENERGY USE

Given the multifaceted interdependencies between IT and facilities, it is imperative that close communication and coordination start early in any project. When this happens, the organization gains an opportunity to investigate how the facility, power, and cooling systems will affect the servers and other ITE from a reliability and energy use standpoint. Energy efficiency demands a holistic approach, and incorporating energy use as one of the metrics when developing the overall IT strategy will result in a significant positive impact in the subsequent planning phases of any IT enterprise project.

If we imagine the data center as the landlord and the ITE as the primary tenant, it is essential that there is an ongoing dialog to understand the requirements of the tenant and the capabilities of the landlord. This interface arguably presents

the greatest opportunities for overall energy use optimization in the data center. From a thermal standpoint, the computer's main mission is to keep its internal components at a prescribed maximum temperature to minimize risk of thermal shutdown, reduce electrical leakage, and, in extreme cases, mitigate any chances of physical damage to the equipment. The good news is that thermal engineers for the ITE have more fully embraced designing servers around the use of higher internal temperatures, wide temperature swings, and elimination of humidification equipment. From a data center cooling perspective, it is essential to understand how the ambient temperature affects the power use of the computers. Based on the inlet temperature of the computer, the overall system power will change; assuming a constant workload, the server fan power will increase as the inlet temperature increases. The data center cooling strategy must account for the operation of the computers to avoid an unintentional increase in energy use by raising the inlet temperature too high.

3.13.1 Leveraging IT and Facilities

Based on current market conditions, there is a confluence of events that can enable energy optimization of the IT enterprise. It just takes some good planning and a thorough understanding of all the elements that affect energy use. Meeting these multiple objectives—service enhancement, reliability, and reduction of operational costs—once thought to be mutually exclusive, must now be thought of as key success factors that must occur simultaneously. Current developments in ITE and operations can be leveraged to reduce/optimize a data center's energy spend: since these developments are in the domains of ITE and facilities, both must be considered to create leverage to reduce energy consumption.

3.13.2 Technology Refresh

Continual progress in increasing computational speed and the ability to handle multiple complex, simultaneous applications, ITE, systems, and software continues to be an important enabler for new businesses and expanding existing ones. As new technology matures and is released into the market, enterprises in the targeted industry sector generally embrace the new technology and use it as a transformative event, looking for a competitive edge. This transformation will require capital to acquire ITE and software. Eventually, the ITE reaches its limit in terms of computation performance and expandability. This "tail wagging the dog" phenomenon is driving new capital expenditures on technology and data centers to record high levels. This appears to be an unending cycle: faster computers enabling new software applications that, in turn, drive the need for newer, more memory and speed-intensive software applications requiring new computers! Without a holistic view of how ITE impacts overall data center energy use, this cycle will surely increase energy if use. But if the plan is to upgrade into new ITE, an opportunity arises to leverage the new equipment upgrade by looking at energy use optimization.

3.13.3 Reducing IT and Operational Costs

For companies to maintain a completive edge in pricing products and services, reducing ongoing operational costs related to IT infrastructure, architecture, applications, real estate, facility operational costs, and energy is critical given the magnitude of the electricity costs. Using a multifaceted approach starting at the overall IT strategy (infrastructure and architecture) and ending at the actual facility where the technology is housed will reap benefits in terms of reduction of annual costs. Avoiding the myopic, singular approach is of paramount importance. The best time to incorporate thinking on energy use optimization is at the very beginning of a new IT planning effort. This process is becoming more common as businesses and their IT and facilities groups become more sophisticated and aware of the value of widening out the view portal and proactively discussing energy use.

3.14 DATA CENTER FACILITIES MUST BE DYNAMIC AND ADAPTABLE

Among the primary design goals of a data center facility are future flexibility and scalability, knowing that IT systems evolve on a life cycle of under 3 years. This however can lead to short-term over-provisioning of power and cooling systems until the IT systems are fully built out. But even when fully built out, the computers, storage, and networking equipment will experience hourly, daily, weekly, and monthly variations depending what the data center is used for. This "double learning curve" of both increasing power usage over time and ongoing fluctuations of power use makes the design and operation of these types of facilities difficult to optimize. Using simulation tools can help to show how these changes affect not only energy use but also indoor environmental conditions, such as dry-bulb temperature, radiant temperature, and moisture content.

3.15 SERVER TECHNOLOGY AND STEADY INCREASE OF EFFICIENCY

Inside a server, the CPU, GPU, and memory must be operating optimally to make sure the server is reliable and fast and can handle large workloads. Servers now have greater compute power and use less energy compared with the same model from the last generation. Businesses are taking

advantage of this by increasing the number of servers in their data centers, ending up with greater capability without facing a significant increase in cooling and power. This is a win-win situation, but care must be taken in the physical placement of the new ITE to ensure the equipment gets proper cooling and has power close by. Also, the workload of the new servers must be examined to assess the impact on the cooling system. Certainly, having servers that use less energy and are more powerful compared with previous models is a great thing, and care must be taken when developing a strategy for increasing the number of servers in a data center.

With every release of the next generation of servers, storage, and networking equipment, we see a marked increase in efficiency and effectiveness. This efficiency increase is manifested by a large boost in computing power, accomplished using the same power as the previous generation. While not new, benchmarking programs such as SPECpower were (and are) being used to understand not only energy consumption but also how the power is used vis-à-vis the computational power of the server. Part of the SPECpower metric is a test that manufacturers run on their equipment, the results of which are published on the SPEC website. From the perspective of a mechanical or electrical engineer designing a data center, one of the more compelling items that appears on the SPECpower summary sheet is the power demand for servers in their database at workloads 100, 75, 50, and 25%. These data give the design engineer a very good idea of the how the power demand fluctuates depending on the workload running on the server. This also will inform the design for the primary cooling and power equipment as to the part load requirements driven by the computer equipment. But the "performance per watt" has increased significantly. This metric can be misleading if taken out of context. While the "performance per watt" efficiency of the servers has shown a remarkable growth, the power demand of the servers is also steadily increasing.

Since cooling and power, and the ITE systems are rapidly advancing, idea and technology exchange between these groups is an important step to advance the synergistic aspect of data center design. Looking at processor power consumption and cooling system efficiency together as a system with interdependent components (not in isolation) will continue to expand the realm of possibilities for creating energy efficient data centers.

3.16 DATA COLLECTION AND ANALYSIS FOR ASSESSMENTS

The cliché "You can't manage what you don't measure" is especially important for data centers, given the exceptionally high energy use power intensity. For example, knowing the relationship between server wattage requirements and mechanical cooling costs can help in determining if purchasing more efficient (and possibly more expensive) power and cooling system components is a financially sound decision. But without actual measured energy consumption data, this decision becomes less scientific and more anecdotal. For example, the operating costs of double conversion UPS compared to line reactive units must be studied to determine if long term operating costs can justify higher initial cost. While much of the data collection process to optimize energy efficiency is similar to what is done in commercial office buildings, schools, or hospitals, there are nuances, which, if not understood, will render the data collection process less effective. The following points are helpful when considering an energy audit consisting of monitoring, measurement, analysis, and remediation in a data center:

1. Identifying operational or maintenance issues: In particular, to assist in diagnosing the root cause of hot spots, heat-related equipment failure, lack of overall capacity, and other common operational problems. Due to the critical nature of data center environments, such problems are often addressed in a very nonoptimal break–fix manner due to the need for an immediate solution. Benchmarking can identify those quick fixes that should be revisited in the interests of lower operating cost or long-term reliability.
2. Helping to plan future improvements: The areas that show the poorest performance relative to other data center facilities usually offer the greatest, most economical opportunity for energy cost savings. Improvements can range from simply changing set points in order to realize an immediate payback to replacing full systems in order to realize energy savings that will show payback over the course of several years.
3. Developing design standards for future facilities: Benchmarking facilities has suggested there are some best practice design approaches that result in fundamentally lower-cost and more efficient facilities. Design standards include best practices that, in certain cases, should be developed as a prototypical design. The prototypes will reduce the cost of future facilities and identify the most effective solutions.
4. Establishing a baseline performance as a diagnostic tool: Comparing trends over time to baseline performance can help predict and avoid equipment failure, improving long-term reliability. Efficiency will also benefit by this process by identifying performance decay that occurs as systems age and calibrations are lost, degrading optimal energy use performance.

The ASHRAE publication *Procedures for Commercial Building Energy Audits* is an authoritative resource on this subject. The document describes three levels of audit from broad to very specific, each with its own set of criteria. In addition to understanding and optimizing energy use in the facility, the audits also include review of operational procedures, documentation, and set points. As the audit progresses, it becomes essential that deficiencies in operational procedures that are causing excessive energy use are separated out from inefficiencies in power and cooling equipment. Without this, false assumptions might be made on equipment performance, leading to unnecessary equipment upgrades, maintenance, or replacement.

ASHRAE Guideline 14-2002, *Measurement of Energy and Demand Savings*, builds on this publication and provides more detail on the process of auditing the energy use of a building. Information is provided on the actual measurement devices, such as sensors and meters, how they are to be calibrated to ensure consistent results year after year, and the duration they are to be installed to capture the data accurately. Another ASHRAE publication, *Real-Time Energy Consumption Measurements in Data Centers*, provides data center-specific information on the best way to monitor and measure data center equipment energy use. Finally, the document *Recommendations for Measuring and Reporting Overall Data Center Efficiency* lists the specific locations in the power and cooling systems where monitoring and measurement is required (Table 3.4). This is important for end users to consistently report energy use in non-data center areas such as UPS and switchgear rooms, mechanical rooms, loading docks, administrative areas, and corridors. Securing energy use data accurately and consistently is essential to a successful audit and energy use optimization program (Table 3.5).

3.17 PRIVATE INDUSTRY AND GOVERNMENT ENERGY EFFICIENCY PROGRAMS

Building codes, industry standards, and regulations are process integrals pervasively in the design and construction industry. Until recently, there was limited availability of documents explicitly written to improve energy efficiency in data center facilities. Many that did exist were meant to be used on a limited basis, and others tended to be primarily anecdotal. All of that has changed with an international release of design guidelines from well-established organizations, covering myriad aspects of data center design, construction, and operation. Many jurisdiction, state, and country have developed custom criteria that fit the climate, weather, economics, and sophistication level of the data center and ITE community. The goal is to deliver the most applicable and helpful energy reduction information to the data center professionals that are responsible for the implementation. And as data center technology continues to advance and ITE hardware and software maintains its rapid evolution, the industry will develop new standards and guidelines to address energy efficiency strategies for these new systems.

Worldwide there are many organizations responsible for the development and maintenance of the current documents on data center energy efficiency. In the US, there are ASHRAE, U.S. Green Building Council (USGBC), US EPA, US DOE, and The Green Grid, among others. The following is an overview of some of the standards and guidelines from these organizations that have been developed specifically to improve the energy efficiency in data center facilities.

3.17.1 USGBC: LEED Adaptations for Data Centers

The new LEED data centers credit adaptation program was developed in direct response to challenges that arose when applying the LEED standards to data center projects. These challenges are related to several factors including the extremely high power density found in data centers. In response, the USGBC has developed credit adaptations that address many of the challenges in certifying data center facilities. The credit adaptations released with the LEED version 4.1 rating system, apply to both Building Design and Construction and Building Operations and Maintenance rating systems. Since the two rating systems apply to buildings in different stages of their life cycle, the credits are adapted in different ways. However, the adaptations were developed with the same goal in mind: establish LEED credits that are applicable to data centers specifically and will help developers, owners, operators, designers, and builders to enable a reduction in energy use, minimize environmental impact, and provide a positive indoor environment for the inhabitants of the data center.

3.17.2 Harmonizing Global Metrics for Data Center Energy Efficiency

In their development of data center metrics such as PUE/DCiE, CUE, and WUE, The Green Grid has sought to achieve a global acceptance to enable worldwide standardization of monitoring, measuring, and reporting data center energy use. This global harmonization has manifested itself in the United States, European Union (EU), and Japan reaching in an agreement on guiding principles for data center energy efficiency metrics. The specific organizations that participated in this effort were U.S. DOE's Save Energy Now and Federal Energy Management Programs, U.S. EPA's ENERGY STAR Program, European Commission Joint Research Centre Data Centers Code of Conduct, Japan's Ministry of Economy, Trade and Industry, Japan's Green IT Promotion Council, and The Green Grid.

TABLE 3.4 Recommended items to measure and report overall data center efficiency

System	Units	Data source	Duration
Total recirculation fan (total CRAC) usage	kW	From electrical panels	Spot
Total makeup air handler usage	kW	From electrical panels	Spot
Total IT equipment power usage	kW	From electrical panels	Spot
Chilled water plant	kW	From electrical panels	1 week
Rack power usage, 1 typical	kW	From electrical panels	1 week
Number of racks	Number	Observation	Spot
Rack power usage, average	kW	Calculated	N/A
Other power usage	kW	From electrical panels	Spot
Data center temperatures (located strategically)	°F	Temperature sensor	1 week
Humidity conditions	R.H.	Humidity sensor	1 week
Annual electricity use, 1 year	kWh/y	Utility bills	N/A
Annual fuel use, 1 year	Therm/y	Utility bills	N/A
Annual electricity use, 3 prior years	kWh/y	Utility bills	N/A
Annual fuel use, 3 prior years	Therm/y	Utility bills	N/A
Peak power	kW	Utility bills	N/A
Average power factor	%	Utility bills	N/A
Facility (total building) area	sf	Drawings	N/A
Data center area (“electrically active floor space”)	sf	Drawings	N/A
Fraction of data center in use (fullness factor)	%	Area and rack observations	Spot
Airflow	cfm	(Designed, TAB report)	N/A
Fan power	kW	3Φ True power	Spot
VFD speed	Hz	VFD	Spot
Set point temperature	°F	Control system	Spot
Return air temperature	°F	10k Thermistor	1 week
Supply air temperature	°F	10k Thermistor	1 week
RH set point	RH	Control system	Spot
Supply RH	RH	RH sensor	1 week
Return RH	RH	RH Sensor	1 week
Status	Misc.	Observation	Spot
Cooling load	Tons	Calculated	N/A
Chiller power	kW	3Φ True power	1 week
Primary chilled water pump power	kW	3Φ True power	Spot
Secondary chilled water pump power	kW	3Φ True power	1 week
Chilled water supply temperature	°F	10k Thermistor	1 week
Chilled water return temperature	°F	10k Thermistor	1 week
Chilled water flow	gpm	Ultrasonic flow	1 week
Cooling tower power	kW	3Φ True power	1 week
Condenser water pump power	kW	3Φ True power	Spot
Condenser water supply temperature	°F	10k Thermistor	1 week

TABLE 3.4 (Continued)

System	Units	Data source	Duration
Chiller cooling load	Tons	Calculated	N/A
Backup generator(s) size(s)	kVA	Label observation	N/A
Backup generator standby loss	kW	Power measurement	1 week
Backup generator ambient temp	°F	Temp sensor	1 week
Backup generator heater set point	°F	Observation	Spot
Backup generator water jacket temperature	°F	Temp sensor	1 week
UPS load	kW	UPS interface panel	Spot
UPS rating	kVA	Label observation	Spot
UPS loss	kW	UPS interface panel or measurement	Spot
PDU load	kW	PDU interface panel	Spot
PDU rating	kVA	Label observation	Spot
PDU loss	kW	PDU interface panel or measurement	Spot
Target	Units	Data source	Duration
Outside air dry-bulb temperature	°F	Temp/RH sensor	1 week
Outside air wet-bulb temperature	°F	Temp/RH sensor	1 week

Source: ©2020, Bill Kosik.

3.17.3 Industry Consortium: Recommendations for Measuring and Reporting Overall Data Center Efficiency

In 2010, a task force consisting of representatives from leading data center organizations (7×24 Exchange, ASHRAE, The Green Grid, Silicon Valley Leadership Group, U.S. Department of Energy Save Energy Now Program, U.S. EPA's ENERGY STAR Program, USGBC, and Uptime Institute) convened to discuss how to standardize the process of measuring and reporting PUE. The purpose is to encourage data center owners with limited measurement capability to participate in programs where

TABLE 3.5 Location and data of monitoring and measurement for auditing energy use and making recommendations for increasing efficiency (Courtesy of Lawrence Berkeley National Laboratory)

ID	Data	Unit
General data center data		
dG1	Data center area (electrically active)	sf
dG2	Data center location	—
dG3	Data center type	—
dG4	Year of construction (or major renovation)	—
Data center energy data		
dA1	Annual electrical energy use	kWh
dA2	Annual IT electrical energy use	kWh
dA3	Annual fuel energy use	MMBTU
dA4	Annual district steam energy use	MMBTU
dA5	Annual district chilled water energy use	MMBTU
Air management		
dB1	Supply air temperature	°F
dB2	Return air temperature	°F
dB3	Low-end IT equipment inlet air relative humidity set point	%
dB4	High-end IT equipment inlet air relative humidity set point	%
dB5	Rack inlet mean temperature	°F
dB6	Rack outlet mean temperature	°F
Cooling		
dC1	Average cooling system power consumption	kW
dC2	Average cooling load	Tons
dC3	Installed chiller capacity (w/o backup)	Tons
dC4	Peak chiller load	Tons
dC5	Air economizer hours (full cooling)	Hours
dC6	Air economizer hours (partial cooling)	Hours

(Continued)

TABLE 3.5 (Continued)

ID	Data	Unit
dC7	Water economizer hours (full cooling)	Hours
dC8	Water economizer hours (partial cooling)	Hours
dC9	Total fan power (supply and return)	W
dC10	Total fan airflow rate (supply and return)	CFM
Electrical power chain		
dE1	UPS average load	kW
dE2	UPS load capacity	kW
dE3	UPS input power	kW
dE4	UPS output power	kW
dE5	Average lighting power	kW

Source: ©2020, Bill Kosik.

power/energy measure is required while also outlining a process that allows operators to add additional measurement points to increase the accuracy of their measurement program. The goal is to develop a consistent and repeatable measurement strategy that allows data center operators to monitor and improve the energy efficiency of their facility. A consistent measurement approach will also facilitate communication of PUE among data center owners and operators. It should be noted that caution must be exercised when an organization wishes to use PUE to compare different data centers, as it is necessary to first conduct appropriate data analyses to ensure that other factors such as levels of reliability and climate are not impacting the PUE.

3.17.4 US EPA: ENERGY STAR for Data Centers

In June 2010, the US EPA released the data center model for their Portfolio Manager, an online tool for building owners to track and improve energy and water use in their buildings. This leveraged other building models that have been developed since the program started with the release of the office building model in 1999. The details of how data center facilities are ranked in the Portfolio Manager are discussed in a technical brief available on the EPA's website.

Much of the information required in attempting to obtain an ENERGY STAR rating for a data center is straightforward. A licensed professional (architect or engineer) is required to validate the information that is contained in the Data Checklist. The licensed professional should reference the 2018 Licensed Professional's Guide to the ENERGY STAR Label for Commercial Buildings for guidance in verifying a commercial building to qualify for the ENERGY STAR.

3.17.5 ASHRAE: Green Tips for Data Centers

The ASHRAE Datacom Series is a compendium of books, authored by ASHRAE Technical Committee 9.9 that provides a foundation for developing energy-efficient designs of the data center. These 14 volumes are under continuous maintenance by ASHRAE to incorporate the newest design concepts that are being introduced by the engineering community. The newest in the series, *Advancing DCIM with IT Equipment Integration*, depicts how to develop a well-built and sustainable DCIM system that optimizes efficiency of power, cooling, and ITE systems. The Datacom Series is aimed at facility operators and owners, ITE organizations, and engineers and other professional consultants.

3.17.6 The Global e-Sustainability Initiative (GeSI)

This program demonstrates the importance of aggressively reducing energy consumption of ITE, power, and cooling systems. But when analyzing Global e-Sustainability Initiative's (GeSI) research material, it becomes clear that their vision is focused on a whole new level of opportunities to reduce energy use at a global level. This is done by developing a sustainable, resource, and energy-efficient world through ICT-enabled transformation. According to GeSI, "[They] support efforts to ensure environmental and social sustainability because they are inextricably linked in how they impact society and communities around the globe." Examples of this vision:

- . . .the emissions avoided through the use of ITE are already nearly 10 times greater than the emissions generated by deploying it.
- ITE can enable a 20% reduction of global CO_{2e} emissions by 2030, holding emissions at current levels.
- ITE emissions as a percentage of global emissions will decrease over time. Research shows the ITE sector's emissions "footprint" is expected to decrease to 1.97% of global emissions by 2030, compared to 2.3% in 2020.

3.17.7 Singapore Green Data Centre Technology Roadmap

"The Singapore Green Data Centre Technology Roadmap" aims to reduce energy consumption and improve the energy efficiency of the primary energy consumers in a data center—facilities and IT. The roadmap assesses

and makes recommendations on potential directions for research, development, and demonstration (RD&D) to improve the energy efficiency of Singapore's data centers. It covers the green initiatives that span existing data centers and new data centers.

Three main areas examined in this roadmap are facility, IT systems, and an integrated approach to design and deployment of data centers.

Of facility systems, cooling has received the most attention as it is generally the single largest energy overhead. Singapore's climate, with its year-round high temperatures and humidity, makes cooling particularly energy-intensive compared to other locations. The document examines technologies to improve the energy efficiency of facility systems:

1. Direct liquid cooling
2. Close-coupled refrigerant cooling
3. Air and cooling management
4. Passive cooling
5. Free cooling (hardening of ITE)
6. Power supply efficiency

Notwithstanding the importance of improving the energy efficiency of powering and cooling data centers, the current focal point for innovation is improving the energy performance of physical IT devices and software. Physical IT devices and software provide opportunities for innovation that would greatly improve the sustainability of data centers:

1. Software power management
2. Energy-aware workload allocation
3. Dynamic provisioning
4. Energy-aware networking
5. Wireless data centers
6. Memory-type optimization

The Roadmap explores future directions in advanced DCIM to enable the integration and automation of the disparate systems of the data center. To this end, proof-of-concept demonstrations are essential if the adoption of new technologies is to be fast-tracked in Singapore.

3.17.8 FIT4Green

An early example of collaboration among EU countries, this consortium made up of private and public organizations from Finland, Germany, Italy, Netherlands, Spain, and the United Kingdom; FIT4Green "aims at contributing to ICT energy reducing efforts by creating an energy-aware layer of plug-ins for data center automation frameworks, to improve energy efficiency of existing IT solution deployment strategies so as to minimize overall power consumption, by moving computation and services around a federation of IT data centers sites."

3.17.9 EU Code of Conduct on Data Centre Energy Efficiency 2018

This best practice supplement to the Code of Conduct is provided as an education and reference document as part of the Code of Conduct to assist data center operators in identifying and implementing measures to improve the energy efficiency of their data centers. A broad group of expert reviewers from operators, vendors, consultants, academics, and professional and national bodies have contributed to and reviewed the best practices. This best practice supplement is a full list of data center energy efficiency best practices. The best practice list provides a common terminology and frame of reference for describing an energy efficiency practice to assist participants and endorsers in avoiding doubt or confusion over terminology. Customers or suppliers of IT services may also find it useful to request or provide a list of Code of Conduct Practices implemented in a data center to assist in procurement of services that meet their environmental or sustainability standard.

3.17.10 Guidelines for Environmental Sustainability Standard for the ICT Sector

The impetus for this project came from questions being asked by customers, investors, governments, and other stakeholders to report on sustainability in the data center, but there is lack of an agreed-upon standardized measurement that would simplify and streamline this reporting specifically for the ICT sector. The standard provides a set of agreed-upon sustainability requirements for ICT companies that allows for a more objective reporting of how sustainability is practiced in the ICT sector in these key areas: sustainable buildings, sustainable ICT, sustainable products, sustainable services, end of life management, general specifications, and assessment framework for environmental impacts of the ICT sector.

There are several other standards that range from firmly established to emerging not mentioned here. The landscape for the standards and guidelines for data centers is growing, and it is important that both the IT and facilities personnel become familiar with them and apply them where relevant.

3.18 STRATEGIES FOR OPERATIONS OPTIMIZATION

Many of the data center energy efficiency standards and guidelines available today tend to focus on energy conservation measures that involve improvements to the power and

cooling systems. Or if the facility is new, strategies that can be used in the design process to improve efficiency. Arguably and equally important is how to improve energy use through better operations.

Developing a new data center includes expert design engineers, specialized builders, and meticulous commissioning processes. If the operation of the facility does not incorporate requirements of the design and construction process, it is entirely possible that deficiencies will arise in the operation of the power and cooling systems. Having a robust operations optimization process in place will identify and neutralize these discrepancies and move the data center toward enhanced energy efficiency (see Table 3.6).

3.19 UTILITY CUSTOMER-FUNDED PROGRAMS

One of the more effective ways of ensuring that a customer will reduce their building portfolio energy use footprint is if the customer is involved in a utility customer-funded efficiency program. These programs typically cover both natural gas and electricity efficiency measures in all market sectors (residential, commercial, etc.). With the proper planning, engineering, and documentation, the customer will receive incentives that are designed to help offset some of the first cost of the energy reduction project. One of the key documents developed at the state level used in these programs is called the Technical Resource Manual (TRM), which provides very granular data on how to calculate energy use reduction as it applies to the program. TRMs also can include information on other efficiency measures, such as energy conservation or demand response, water conservation, and utility customer-sited storage and distributed generation projects and renewable resources.

The primary building block of this process is called a measure. The measure is the part of the overall energy reduction strategy that outlines the process of one discreet way of energy efficiency. More than one measure is typically submitted for review and approval; ideally the measures have a synergistic effect on the other measures. The structure of a measure, while being straightforward, is rich with technical guidance. A measure is comprised of the following.

TABLE 3.6 Example of analysis and recommendations for increasing data center efficiency and improving operational performance

Title	Description
Supply air temperatures to computer equipment if too cold	Further guidance can be found in "Design Considerations for Datacom Equipment Centers" by ASHRAE and other updated recommendations. Guideline recommended range is 64.5–80°F. However, the closer the temperatures to 80°F, the more energy efficient the data center becomes
Relocate high-density equipment to within area of influence of CRACs	High-density racks should be as close as possible to CRAC/H units unless other means of supplemental cooling or chilled water cabinets are used
Distribute high-density racks	High-density IT hardware racks are distributed to avoid undue localized loading on cooling resources
Provide high-density heat containment system for the high density load area	For high-density loads there are a number of design concepts whose basic intent is to contain and separate the cold air from the heated return air on the data floor: hot aisle containment; cold aisle containment; contained rack supply, room return; room supply, contained rack return; contained rack supply, contained rack return
Install strip curtains to segregate airflows	While this will reduce recirculation, access to cabinets needs to be carefully considered
Correct situation to eliminate air leakage through the blanking panels	Although blanking panels are installed, it was observed that they are not in snug-fit "properly fit" position, and some air appears to be passing through openings up and below the blanking panels
Increase CRAH air discharge temperature and chilled water supply set points by 2°C (~4°F)	Increasing the set point by 0.6°C (1°F) reduces chiller power consumption 0.75–1.25% of fixed speed chiller kilowatt per ton and 1.5–3% for VSD chiller. Increasing the set point also widens the range of economizers operation if used; hence more saving should be expected
Widen %RH range of CRAC/H units	Humidity range is too tight. Humidifiers will come on more often. ASHRAE recommended range for servers' intake is 30–80 %RH. Widening the %RH control range (within ASHRAE guidelines) will enable less humidification ON time and hence less energy utilization. In addition, this will help to eliminate any control fighting

Source: ©2020, Bill Kosik.

3.19.1 Components of TRM Measure Characterizations

Each measure characterization uses a standardized format that includes at least the following components. Measures that have a higher level of complexity may have additional components, but also follow the same format, flow, and function.

3.19.2 Description

Brief description of measure stating how it saves energy, the markets it serves, and any limitations to its applicability.

3.19.3 Definition of Efficient Equipment

Clear definition of the criteria for the efficient equipment used to determine delta savings. Including any standards or ratings if appropriate.

3.19.4 Definition of Baseline Equipment

Clear definition of the efficiency level of the baseline equipment used to determine delta savings including any standards or ratings if appropriate. If a time of sale measure, the baseline will be new base level equipment (to replace existing equipment at the end of its useful life or for a new building). For early replacement or early retirement measures, the baseline is the existing working piece of equipment that is being removed.

3.19.5 Deemed Lifetime of Efficient Equipment

The expected duration in years (or hours) of the savings. If an early replacement measure, the assumed life of the existing unit is also provided.

3.19.6 Deemed Measure Cost

For time of sale measures, incremental cost from baseline to efficient is provided. Installation costs should only be included if there is a difference between each efficiency level. For early replacement, the full equipment and install cost of the efficient installation is provided in addition to the full deferred hypothetical baseline replacement cost.

3.19.7 Load Shape

The appropriate load shape to apply to electric savings is provided.

3.19.8 Coincidence Factor

The summer coincidence factor is provided to estimate the impact of the measure on the utility's system peak—defined as 1 p.m. to hour ending 5 p.m. on non-holiday weekdays, June through August.

3.19.9 Algorithms and Calculation of Energy Savings

Algorithms are provided followed by list of assumptions with their definition.

If there are no input variables, there will be a finite number of output values. These will be identified and listed in a table. Where there are custom inputs, an example calculation is often provided to illustrate the algorithm and provide context. The calculations with determine the following:

- Electric energy savings
- Summer coincident peak demand savings
- Natural gas savings
- Water impact descriptions and calculation
- Deemed O&M cost adjustment calculation

3.19.10 Determining Data Center Energy Use Effectiveness

When analyzing and interpreting energy use in a data center, it is essential that industry-accepted methods are used to develop the data collection forms, analysis techniques, and reporting mechanisms. This will ensure a high confidence level that the results are valid and not perceived as a non-standard process that might have built-in bias. These industry standards include ASHRAE 90.1; AHRI Standards 340, 365, and 550-590; and others. (The information contained in the ASHRAE Standard 14 is paraphrased throughout this writing.)

There are several methods available to collect, analyze, and present data to demonstrate both baseline energy consumption and projected savings resulting from the implementation of ECMs. A process called a calibrated simulation analysis incorporates a wide array of stages that range from planning though implementation. The steps listed in ASHRAE 14 are summarized below:

1. Produce a calibrated simulation plan. Before a calibrated simulation analysis may begin, several questions must be answered. Some of these questions include: Which software package will be applied? Will models be calibrated to monthly or hourly measured data, or both? What are to be the tolerances for the statistical indices? The answers to these questions are documented in a simulation plan.
2. Collect data. Data may be collected from the building during the baseline period, the retrofit period, or both. Data collected during this step include dimensions and properties of building surfaces, monthly and hourly whole-building utility data, nameplate

data from HVAC and other building system components, operating schedules, spot measurements of selected HVAC and other building system components, and weather data.

3. Input data into simulation software and run model. Over the course of this step, the data collected in the previous step are processed to produce a simulation-input file. Modelers are advised to take care with zoning, schedules, HVACs stems, model debugging (searching for and eliminating any malfunctioning or erroneous code), and weather data.
4. Compare simulation model output to measured data. The approach for this comparison varies depending on the resolution of the measured data. At a minimum, the energy flows projected by the simulation model are compared to monthly utility bills and spot measurements. At best, the two data sets are compared on an hourly basis. Both graphical and statistical means may be used to make this comparison.
5. Refine model until an acceptable calibration is achieved. Typically, the initial comparison does not yield a match within the desired tolerance. In such a case, the modeler studies the anomalies between the two data sets and makes logical changes to the model to better match the measured data. The user should calibrate to both pre- and post-retrofit data wherever possible and should only calibrate to post-retrofit data alone when both data sets are unavailable. While the graphical methods are useful to assist in this process, the ultimate determination of acceptable calibration will be the statistical method.
6. Produce baseline and post-retrofit models. The baseline model represents the building as it would have existed in the absence of the energy conservation measures. The retrofit model represents the building after the energy conservation measures are installed. How these models are developed from the calibrated model depends on whether a simulation model was calibrated to data collected before the conservation measures were installed, after the conservation measures were installed, or both times. Furthermore, the only differences between the baseline and post-retrofit models must be limited to the measures only. All other factors, including weather and occupancy, must be uniform between the two models unless a specific difference has been observed.
7. Estimate savings. Savings are determined by calculating the difference in energy flows and intensities of the baseline and post-retrofit models using the appropriate weather file.
8. Report observations and savings. Savings estimates and observations are documented in a reviewable format. Additionally, enough model development and calibration documentation shall be provided to allow for accurate recreation of the baseline and post-retrofit models by informed parties, including input and weather files.
9. Tolerances for statistical calibration indices. Graphical calibration parameters as well as two main statistical calibration indices [mean bias error and coefficient of variation (root mean square error)] are required evaluation. Document the acceptable limits for these indices on a monthly and annual basis.
10. Statistical comparison techniques. Although graphical methods are useful for determining where simulated data differ from metered data, and some quantification can be applied, more definitive quantitative methods are required to determine compliance. Two statistical indices are used for this purpose: hourly mean bias error (MBE) and coefficient of variation of the root mean squared error (CV (RMSE)).

Using this method will result in a defendable process with results that have been developed in accordance with industry standards and best practices.

REFERENCES

[1] Shehabi A, Smith S, Sartor D, Brown R, Herrlin M, Koomey J, Masanet E, Horner N, Azevedo I, Lintner W. U.S. data center energy usage report; June 2016. Available at https://www.osti.gov/servlets/purl/1372902/ (Accessed 9/9/2020)

[2] U.S. Environmental Protection Agency. Report to congress on server and data center energy efficiency, public law 109-431. U.S. Environmental Protection Agency ENERGY STAR Program; August 2, 2007.

[3] Pan SY, et al. Cooling water use in thermoelectric power generation and its associated challenges for addressing water energy nexus; 2018. p 26–41. Available at https://www.sciencedirect.com/science/article/pii/S2588912517300085. (Accessed 9/9/2020)

FURTHER READING

AHRI Standard 1060 (I-P)-2013. Performance rating of air-to-air heat exchangers for energy recovery ventilation equipment.

ANSI/AHRI 365 (I-P)-2009. Commercial and industrial unitary air-conditioning condensing units.

ANSI/AHRI 540-2004. Performance rating of positive displacement refrigerant compressors and compressor units.

ANSI/AHRI 1360 (I-P)-2013. Performance rating of computer and data processing room air conditioners.

ASHRAE Standard 90.1-2013 (I-P Edition). Energy standard for buildings except low-rise residential buildings.

ASHRAE. *Thermal Guidelines for Data Processing Environments*. 3rd ed.

ASHRAE. Liquid Cooling Guidelines for Datacom Equipment Centers.

ASHRAE. Real-Time Energy Consumption Measurements in Data Centers.

ASHRAE. *Procedures for Commercial Building Energy Audits*. 2nd ed.

ASHRAE Guideline 14-2002. Measurement of Energy and Demand Savings.

Building Research Establishment's Environmental Assessment Method (BREEAM) Data Centres 2010.

Carbon Usage Effectiveness (CUE): A Green Grid Data Center Sustainability Metric, the Green Grid.

CarbonTrust.org

Cisco Global Cloud Index: Forecast and Methodology, 2016–2021 White Paper, Updated: February 1, 2018.

ERE: A Metric for Measuring the Benefit of Reuse Energy from a Data Center, the Green Grid.

Global e-Sustainability Initiative (GeSI) c/o Scotland House Rond Point Schuman 6 B-1040 Brussels Belgium

Green Grid Data Center Power Efficiency Metrics: PUE and DCIE, the Green Grid.

Green Grid Metrics: Describing Datacenter Power Efficiency, the Green Grid.

Guidelines and Programs Affecting Data Center and IT Energy Efficiency, the Green Grid.

Guidelines for Energy-Efficient Datacenters, the Green Grid.

Harmonizing Global Metrics for Data Center Energy Efficiency Global Taskforce Reaches Agreement on Measurement Protocols for GEC, ERF, and CUE—Continues Discussion of Additional Energy Efficiency Metrics, the Green Grid.

https://www.businesswire.com/news/home/20190916005592/en/North-America-All-in-one-Modular-Data-Center-Market

Information Technology & Libraries, Cloud Computing: Case Studies and Total Costs of Ownership, Yan Han, 2011

Koomey JG, Ph.D. Estimating Total Power Consumption by Servers in the U.S. and the World.

Koomey JG, Ph.D. Growth in Data Center Electricity Use 2005 to 2010.

Lawrence Berkeley Lab High-Performance Buildings for High-Tech Industries, Data Centers.

Proxy Proposals for Measuring Data Center Productivity, the Green Grid.

PUE™: A Comprehensive Examination of the Metric, the Green Grid.

Qualitative Analysis of Power Distribution Configurations for Data Centers, the Green Grid.

Recommendations for Measuring and Reporting Overall Data Center Efficiency Version 2—Measuring PUE for Data Centers, the Green Grid.

Report to Congress on Server and Data Center Energy Efficiency Public Law 109-431 U.S. Environmental Protection Agency. ENERGY STAR Program.

Singapore Standard SS 564: 2010 Green Data Centres.

Top 12 Ways to Decrease the Energy Consumption of Your Data Center, EPA ENERGY STAR Program, US EPA

United States Public Law 109–431—December 20, 2006.

US Green Building Council—LEED Rating System.

Usage and Public Reporting Guidelines for the Green Grid's Infrastructure Metrics (PUE/DCIE) the Green Grid.

Water Usage Effectiveness (WUE™): A Green Grid Data Center Sustainability Metric, the Green Grid.

4

HOSTING OR COLOCATION DATA CENTERS

Chris Crosby and Chris Curtis
Compass Datacenters, Dallas, Texas, United States of America

4.1 INTRODUCTION

"Every day Google answers more than one billion questions from people around the globe in 181 countries and 146 languages."[1] Google does not share their search volume data. But a 2019 report estimated 70,000 search queries every second that is 5.8 billion search per day. The vast majority of this information is not only transmitted but also stored for repeated access, which means that organizations must continually expand the number of servers and storage devices to process this increasing volume of information. All of those servers and storage devices need a data center to call home, and every organization needs to have a data center strategy that will meet their computing needs both now and in the future. Not all data centers are the same, though, and taking the wrong approach can be disastrous both technically and financially. Organizations must therefore choose wisely, and this chapter provides valuable information to help organizations make an informed choice and avoid the most common mistakes.

Historically, the vast majority of corporate computing was performed within data center space that was built, owned, and operated by the organization itself. In some cases, it was merely a back room in the headquarters that was full of servers and patch panels. In other cases, it was a stand-alone, purpose-built data center facility that the organization's IT team commissioned. Whether it was a humble back room devoted to a few servers or a large facility built with a significant budget, what they had in common was that the organization was taking on full responsibility for every aspect of data center planning, development, and operations.

[1] http://www.google.com/competition/howgooglesearchworks.html.

In recent years, this strategy has proven to be cumbersome, inefficient, and costly as data processing needs have rapidly outstripped the ability of a large number of businesses to keep up with them. The size, cost, and complexity of today's data centers have prompted organizations that previously handled all their data center operations "in-house" to come to the conclusion that data centers are not their core competency. Data centers were proving to be a distraction for the organization's internal IT teams, and the capital and costs involved in these projects were becoming an increasingly larger burden on the organization's IT budget that created a market opportunity for data center providers who could relieve organizations of this technical and financial burden, and a variety of new vendors emerged to offer data center solutions that meet those needs.

Although these new businesses use a variety of business models, they may be categorized under two generalized headings:

1. Hosting
2. Colocation (wholesale data centers)

4.2 HOSTING

In their simplest form, hosting companies lease the actual servers (or space on the servers) as well as storage capacity to companies. The equipment and the data center it resides in are owned and operated by the hosting provider. Underneath this basic structure, customers are typically presented with a variety of options. These product options tend to fall within three categories:

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

1. Computing capacity
2. Storage
3. Managed services

4.2.1 Computing Capacity

Computing capacity offerings can vary widely in a hosted environment from space on a provider-owned server all the way up to one or more racks within the facility. For medium to enterprise-sized companies, the most commonly used hosting offering is typically referred to as colocation. These offerings provide customers with a range of alternatives from leasing space in a single provider-supported rack all the way up to leasing multiple racks in the facility. In all of these offerings, the customer's own server and storage equipment are housed in the leased rack space. Typically, in multirack environments, providers also offer the customer the ability to locate all their equipment in a locked cage to protect against unauthorized access to the physical space.

Customer leases in colocated environments cover the physical space and the maintenance for the data center itself. Although some providers may charge the customer for the bandwidth they use, this is not common as most companies operating in this type of environment make their own connectivity arrangements with a fiber provider that is supported in the facility. Providers typically offer facility access to multiple fiber providers to offer their customers with a choice in selecting their connectivity company. The most important lease element is for the actual power delivered to the customer. The rates charged to the customer may vary from "pass through," in which the power charge from the utility is billed directly to the customer with no markup, to a rate that includes a markup added by the data center provider.

4.2.2 Storage

Although most firms elect to use their own storage hardware, many providers do offer storage capacity to smaller customers. Typically, these offerings are based on a per-gigabyte basis with the charge applied monthly.

4.2.3 Managed Services

"Managed services" is the umbrella term used to describe the on-site support functions that the site's provider performs on behalf of their customers. Referred to as "remote" or "warm" hands, these capabilities are often packaged in escalating degrees of functions performed. At the most basic level, managed service offerings can be expected to include actions such as restarting servers and performing software upgrades. Higher-level services can include activities like hardware monitoring; performing moves, adds, and changes; administering Internet security; and the availability of customer monitoring and tracking portals. These services are typically billed to the customer on a monthly basis.

4.3 COLOCATION (WHOLESALE)

The term "colocation" as used to describe the providers who lease only data center space to their customers has been replaced by the term "wholesale" data centers. Wholesale data center providers lease physical space within their facilities to one or more customers. Wholesale customers tend to be larger, enterprise-level organizations with data center requirements of 1 MW power capacity. In the wholesale model, the provider delivers the space and power to the customer and also operates the facility. The customer maintains operational control over all of their equipment that is used within their contracted space.

Traditionally, wholesale facilities have been located in major geographic markets. This structure enables providers to purchase and build out large-capacity facilities ranging from as little as 20,000 ft^2 to those featuring a million square feet of capacity or more. Customers then lease the physical space and their required power from the provider. Within these models, multiple customers operate in a single facility in their own private data centers while sharing the common areas of the building such as security, the loading dock, and office space.

4.4 TYPES OF DATA CENTERS

Within the past 5 years, wholesale providers have found that it is more cost efficient and energy efficient to build out these facilities in an incremental fashion. As a result, many providers have developed what they refer to as "modular" data centers. This terminology has been widely adopted, but no true definition for what constitutes a modular data center has been universally embraced. At the present time, there are five categories of data centers that are generally considered to be "modular" within the marketplace.

4.4.1 Traditional Design

Traditional modular data centers (Fig. 4.1) are building-based solutions that use *shared* internal and external backplanes or plant (e.g., chilled water plant and parallel generator plant). Traditional data centers are either built all at once or, as more recent builds have been done, are expanded through adding new data halls within the building. The challenge with shared backplanes is the introduction of risk due to an entire system shutdown because of cascading failures across the backplane. For "phased builds" in which additional data halls are added over time, the key drawback

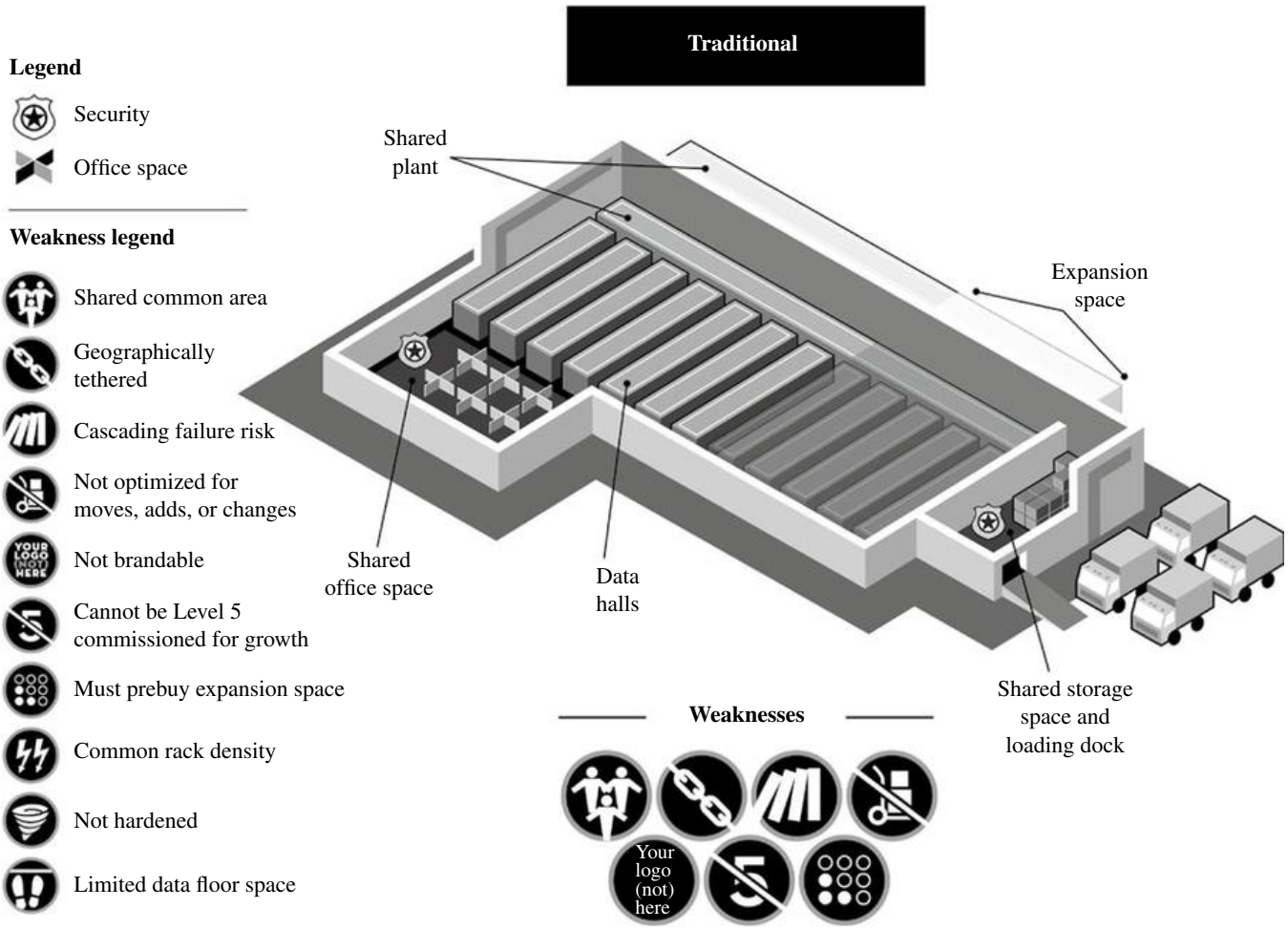


FIGURE 4.1 Traditional wholesale data centers are good solutions for IT loads above 5 MW. Source: Courtesy of Compass Datacenters.

to this new approach is the use of a shared backplane. In this scenario, future "phases" cannot be commissioned to Level 5 Integrated System Level [1] since other parts of the data center are already live.

In Level 5 Commissioning, all of the systems of the data center are tested under full load to ensure that they work both individually and in combination so that the data center is ready for use on day one.

Strengths:
- Well suited for single users
- Good for large IT loads, 5 MW+ day-one load

Weaknesses:
- Cascading failure potential on shared backplanes
- Cannot be Level 5 commissioned (in phased implementations)
- Geographically tethered (this can be a bad bet if the projected large IT load never materializes)
- Shared common areas with multiple companies or divisions (the environment is not dedicated to a single customer)
- Very large facilities that are not optimized for moves/adds/changes

4.4.2 Monolithic Modular (Data Halls)

As the name would imply, monolithic modular data centers (Fig. 4.2) are large building-based solutions. Like traditional facilities, they are usually found in large buildings and provide 5 MW+ of IT power day one with the average site featuring 5–20 MW of capacity. Monolithic modular facilities use segmentable backplanes to support their data halls so they do not expose customers to single points of failure and each data hall can be independently Level 5 commissioned prior to customer occupancy. Often, the only shared component of the mechanical and electrical plant is the medium-voltage utility gear. Because these solutions are housed

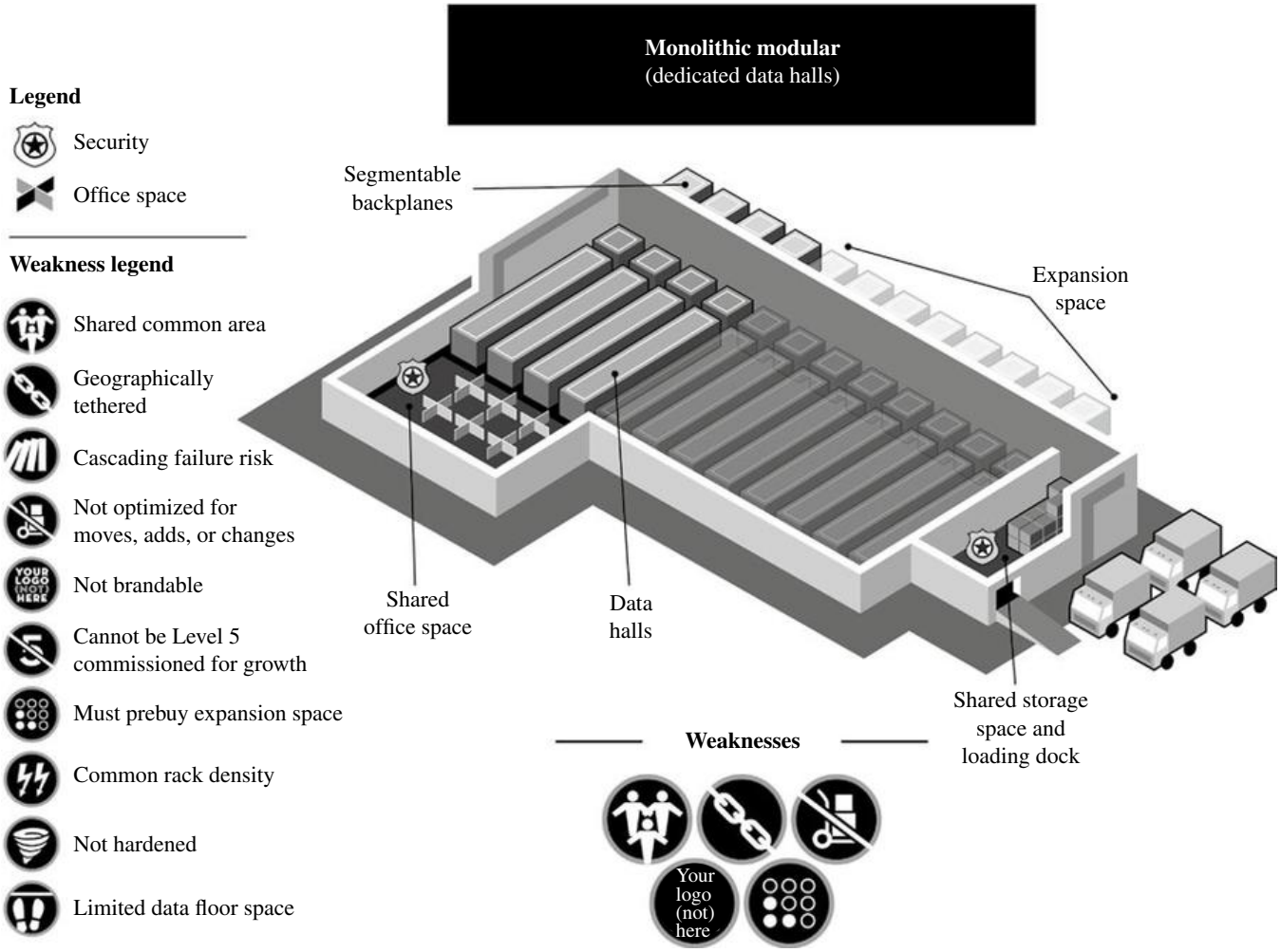


FIGURE 4.2 Monolithic modular data centers with data halls feature segmental backplanes that avoid the possibility of cascading failure found with traditional designs. Source: Courtesy of Compass Datacenters.

within large buildings, the customer may sacrifice a large degree of facility control and capacity planning flexibility if the site houses multiple customers. Additionally, security and common areas (offices, storage, staging, and the loading dock) are shared with the other occupants within the building. The capacity planning limit is a particularly important consideration as customers must prelease (and pay for) shell space within the facility to ensure that it is available when they choose to expand.

Strengths:

- Good for users with known fixed IT capacity, for example, 4 MW day one, growing to 7 MW by year 4, with fixed takedowns of 1 MW/year
- Optimal for users with limited moves/adds/changes
- Well suited for users that don't mind sharing common areas
- Good for users that don't mind outsourcing security

Weaknesses:

- Must pay for unused expansion space.
- Geographically tethered large buildings often require large upfront investment.
- Outsourced security.
- Shared common areas with multiple companies or divisions (the environment is not dedicated to a single customer).
- Very large facilities that are not optimized for moves/adds/changes.

4.4.3 Containerized

Commonly referred to as "containers" (Fig. 4.3), prefabricated data halls are standardized units contained in ISO shipping containers that can be delivered to a site to fill an immediate need. Although advertised as quick to deliver, customers are often required to provide the elements of the

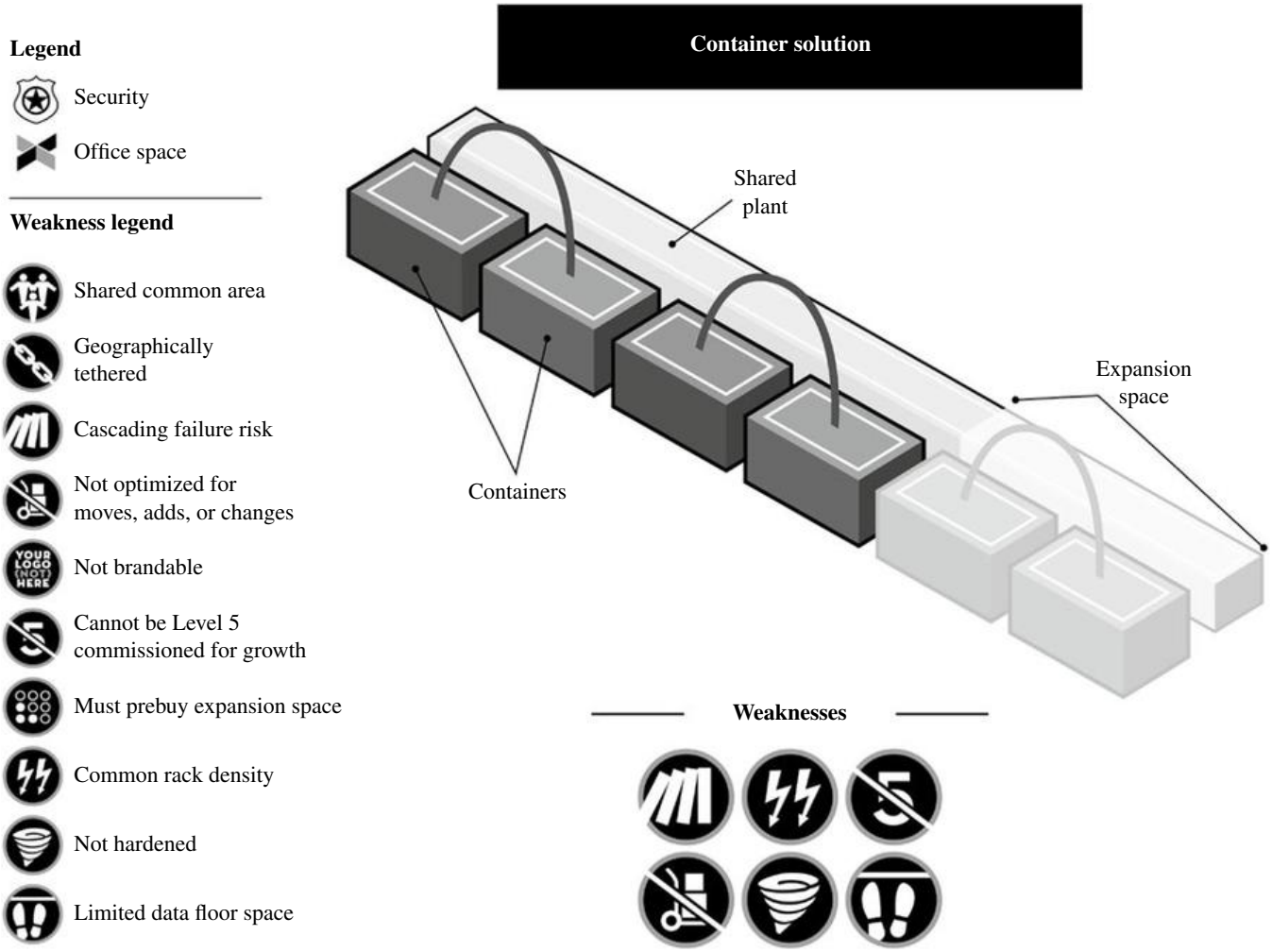


FIGURE 4.3 Container solutions are best suited for temporary applications. Source: Courtesy of Compass Datacenters.

shared outside plant including generators, switch gear, and, sometimes, chilled water. These backplane elements, if not in place, can take upward of 8 months to implement, often negating the benefit of speed of implementation. As long-term solutions, prefabricated containers may be hindered by their nonhardened designs that make them susceptible to environmental factors like wind, rust, and water penetration and their space constraints that limit the amount of IT gear that can be installed inside them. Additionally, they do not include support space like a loading dock, a storage/staging area, or security stations, thereby making the customer responsible for their provision.

Strengths:
- Optimized for temporary data center requirements
- Good for applications that work in a few hundred of KW load groups
- Support batch processing or supercomputing applications
- Suitable for remote, harsh locations (such as military locales)
- Designed for limited move/add/change requirements
- Homogeneous rack requirement applications

Weaknesses:
- Lack of security
- Nonhardened design
- Limited space
- Cascading failure potential
- Cannot be Level 5 commissioned when expanded
- Cannot support heterogeneous rack requirements
- No support space

4.4.4 Monolithic Modular (Prefabricated)

These building-based solutions are similar to their data hall counterparts with the exception that they are populated with

the provider's prefabricated data halls. The prefabricated data hall (Fig. 4.4) necessitates having tight control over the applications of the user. Each application set should drive the limited rack space to its designed load limit to avoid stranding IT capacity. For example, low-load-level groups go in one type of prefabricated data hall, and high-density-load groups go into another. These sites can use shared or segmented backplane architectures to eliminate single points of failure and to enable each unit to be Level 5 commissioned. Like other monolithic solutions, these repositories for containerized data halls require customers to prelease and pay for space in the building to ensure that it is available when needed to support their expanded requirements.

Strengths:
- Optimal for sets of applications in homogeneous load groups
- Designed to support applications that work in kW load groups of a few hundred kW in total IT load
- Good for batch and supercomputing applications
- Optimal for users with limited moves/adds/changes
- Good for users that don't mind sharing common areas

Weaknesses:
- Outsourced security.
- Expansion space must be preleased.
- Shared common areas with multiple companies or divisions (the environment is not dedicated to a single customer).
- Since it still requires a large building upfront, may be geographically tethered.
- Very large facilities that are not optimized for moves/adds/changes.

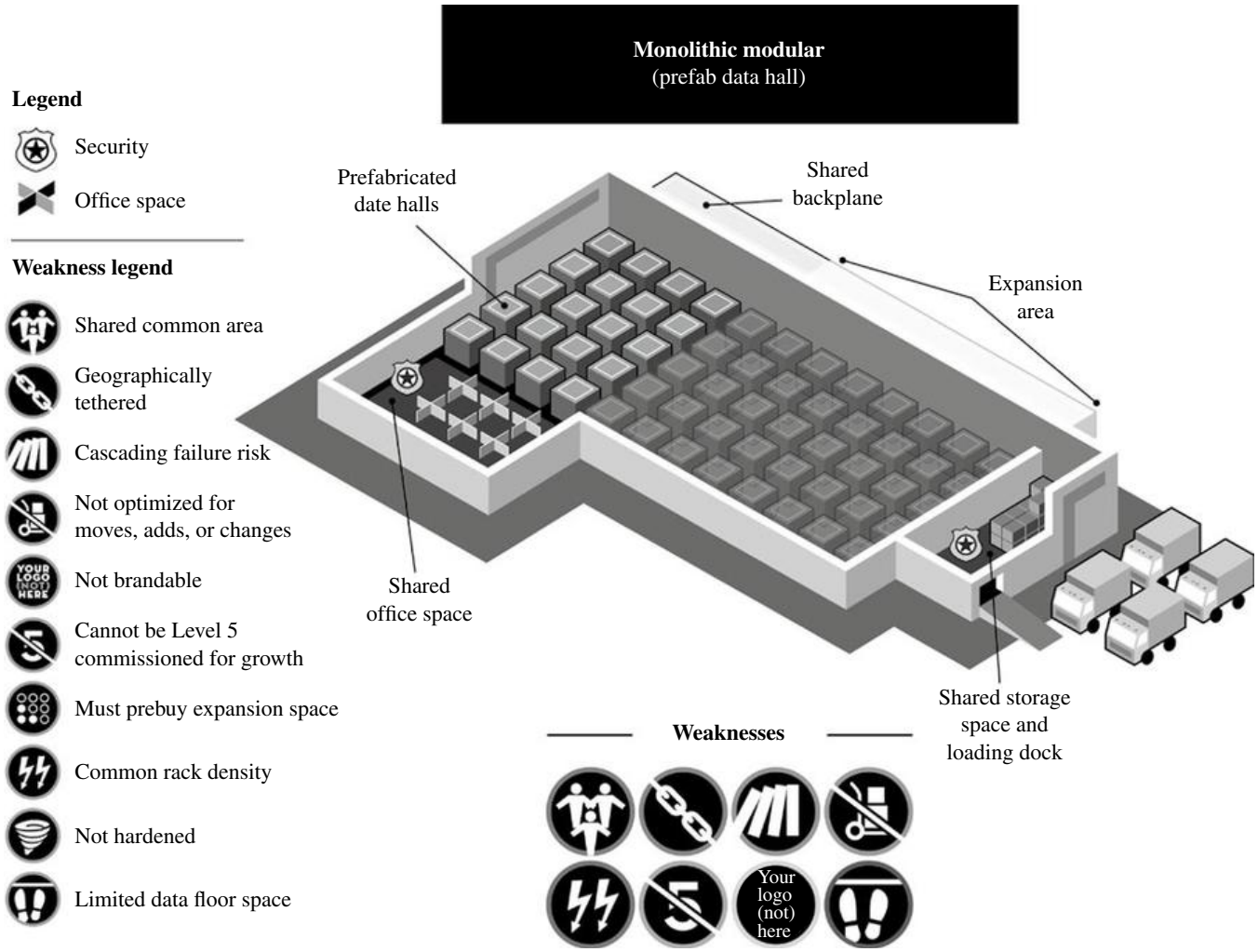


FIGURE 4.4 Monolithic modular data centers with prefabricated data halls use a shared backplane architecture that raises the risk of cascading failure in the event of an attached unit. Source: Courtesy of Compass Datacenters.

4.4.5 Stand-Alone Data Centers

Stand-alone data centers use modular architectures in which the main components of a data center have been incorporated into a hardened shell that is easily expandable in standard-sized increments. Stand-alone facilities are designed to be complete solutions that meet the certification standards for reliability and building efficiency. Stand-alone data centers have been developed to provide geographically independent alternatives for customers who want a data center dedicated to their own use, physically located where it is needed.

By housing the data center area in a hardened shell that can withstand extreme environmental conditions, stand-alone solutions differ from prefabricated or container-based data centers that require the customer or provider to erect a building if they are to be used as a permanent solution. By using standard power and raised floor configurations, stand-alone data centers simplify customers' capacity planning capability by enabling them to add capacity as it is needed rather than having to prelease space within a facility as in the case of monolithic modular solutions, for example.

Because they provide customers with their own dedicated facility, stand-alone data centers use their modular architectures to provide customers with all the site's operational components (office space, loading dock, storage and staging areas, break room, and security area) without the need to share them as in other modular solutions (Fig. 4.5).

Strengths:

- Optimized for security-conscious users
- Good for users who do not like to share any mission-critical components
- Optimal for geographically diverse locations
- Good for applications with 1–4 MW of load and growing over time
- Designed for primary and disaster recovery data centers
- Suitable for provider data centers
- Meet heterogeneous rack and load group requirements

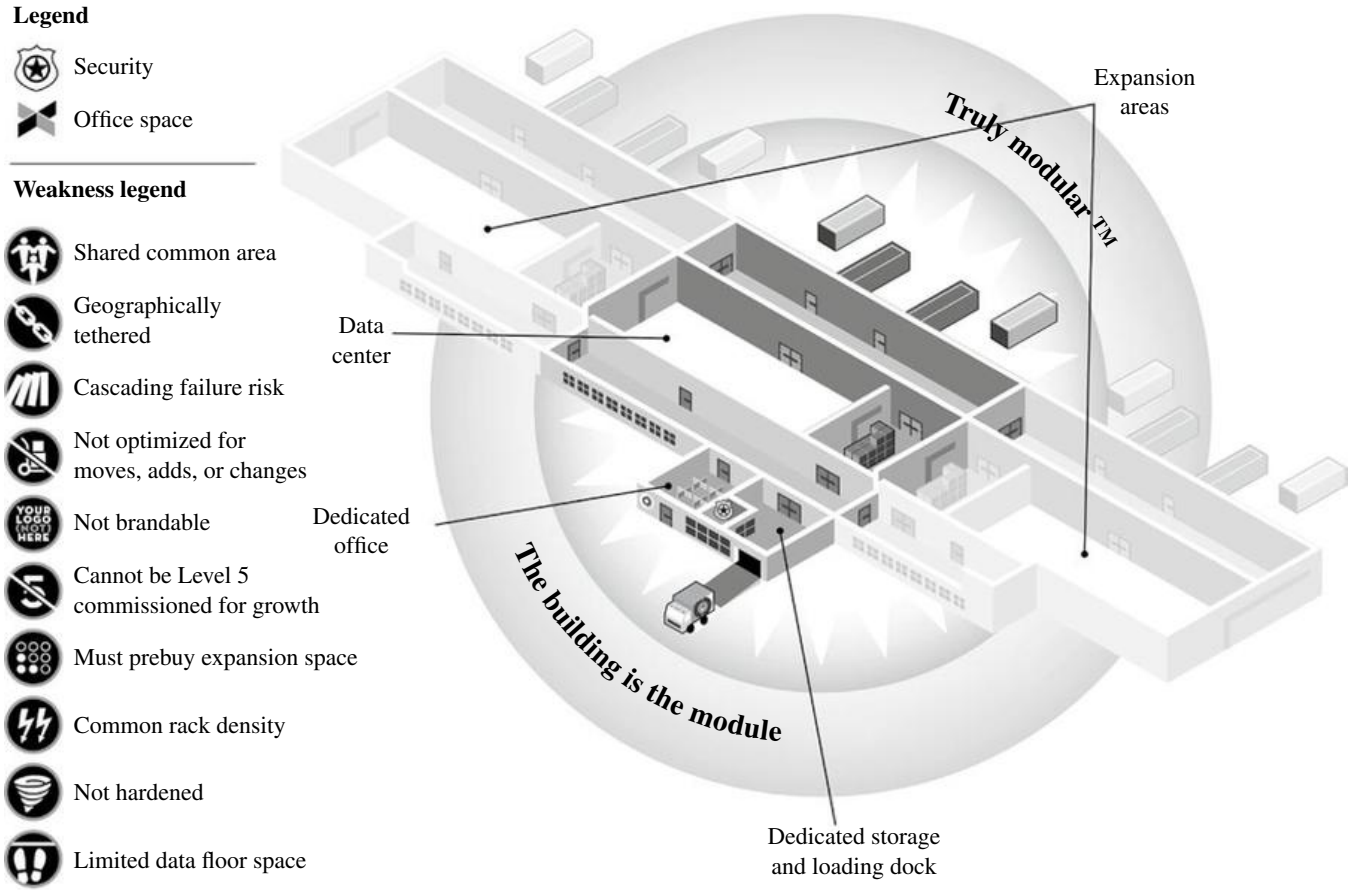


FIGURE 4.5 Stand-alone data centers combine all of the strengths of the other data center types while eliminating their weaknesses. Source: Courtesy of Compass Datacenters.

Weaknesses:
- Initial IT load over 4 MW
- Non-mission-critical data center applications

4.5 SCALING DATA CENTERS

Scaling, or adding new data centers, is possible using either a hosting or wholesale approach. A third method, build to suit, where the customer pays to have their data centers custom built where they want them, may also be used, but this approach is quite costly. The ability to add new data centers across a country or internationally is largely a function of geographic coverage of the provider and the location(s) that the customer desires for their new data centers.

For hosting customers, the ability to use the same provider in all locations limits the potential options available to them. There are a few hosting-oriented providers (e.g., Equinix and Savvis) that have locations in all of the major international regions (North America, Europe, and Asia Pacific). Therefore, the need to add hosting-provided services across international borders may require a customer to use different providers based on the region desired.

The ability to scale in a hosted environment may also require a further degree of flexibility on the part of the customer regarding the actual physical location of the site. No provider has facilities in every major country. Typically, hosted locations are found in the major metropolitan areas in the largest countries in each region. Customers seeking U.S. locations will typically find the major hosting providers located in cities such as New York, San Francisco, and Dallas, for example, while London, Paris, Frankfurt, Singapore, and Sydney tend to be common sites for European and Asia Pacific international locations.

Like their hosting counterparts, wholesale data center providers also tend to be located in major metropolitan locations. In fact, this distinction tends to be more pronounced as the majority of these firms' business models require them to operate facilities of 100,000 ft^2 or more to achieve the economies of scale necessary to offer capacity to their customers at a competitive price point. Thus, the typical wholesale customer that is looking to add data center capacity across domestic regions, or internationally, may find that their options tend to be focused in the same locations as for hosting providers.

4.6 SELECTING AND EVALUATING DC HOSTING AND WHOLESALE PROVIDERS

In evaluating potential hosting or wholesale providers from the perspective of their ability to scale, the most important element for customers to consider is the consistency of their operations. Operational consistency is the best assurance that customers can have (aside from actual Uptime Institute Tier III or IV certification[2]) that their providers' data centers will deliver the degree of reliability or uptime that their critical applications require. In assessing this capability, customers should examine each potential provider based on the following capabilities:

- Equipment providers: The use of common vendors for critical components such as UPS or generators enables a provider to standardize operations based on the vendors' maintenance standards to ensure that maintenance procedures are standard across all of the provider's facilities.
- Documented processes and procedures: A potential provider should be able to show prospective customers its written processes and procedures for all maintenance and support activities. These procedures should be used for the operation of each of the data centers in their portfolio.
- Training of personnel: All of the operational personnel who will be responsible for supporting the provider's data centers should be vendor certified on the equipment they are to maintain. This training ensures that they understand the proper operation of the equipment, its maintenance needs, and troubleshooting requirements.

The ability for a provider to demonstrate the consistency of their procedures along with their ability to address these three important criteria is essential to assure their customers that all of their sites will operate with the highest degree of reliability possible.

4.7 BUILD VERSUS BUY

Build versus buy (or lease in this case) is an age-old business question. It can be driven by a variety of factors such as the philosophy of the organization itself or a company's financial considerations. It can also be affected by issues like the cost and availability of capital or the time frames necessary for the delivery of the facility. The decision can also differ based on whether or not the customer is considering a wholesale data center or a hosting solution.

4.7.1 Build

Regardless of the type of customer, designing, building, and operating a data center are unlike any other type of building.

[2] The Uptime Institute's Tier system establishes the requirements that must be used to provide specified levels of uptime. The most common of the system's four tiers is Tier III (99.995% uptime) that requires redundant configurations on major system components. Although many providers will claim that their facilities meet these requirements, only a facility that has been certified as meeting these conditions by the Institute are actually certified as meeting these standards.

They require a specialized set of skills and expertise. Due to the unique requirements of a data center, the final decision to lease space from a provider or to build their own data center requires every business to perform a deep level of analysis of their own internal capabilities and requirements and those of the providers they may be considering.

Building a data center requires an organization to use professionals and contractors from outside of their organization to complete the project. These individuals should have demonstrable experience with data centers. This also means that they should be aware of the latest technological developments in data design and construction, and the evaluation process for these individuals and firms should focus extensively on these attributes.

4.7.2 Leasing

Buying a data center offers many customers a more expedient solution than building their own data center, but the evaluation process for potential providers should be no less rigorous. While experience with data centers probably isn't an issue in these situations, prospective customers should closely examine the provider's product offerings, their existing facilities, their operational records, and, perhaps most importantly, their financial strength as signing a lease typically means at least a 5-year commitment with the chosen provider.

4.7.3 Location

Among the most important build-versus-buy factors is the first—where to locate it. Not just any location is suitable for a data center. Among the factors that come into play in evaluating a potential data center site are the cost and availability of power (and potentially water). The site must also offer easy access to one or more fiber network carriers. Since data centers support a company's mission-critical applications, the proposed site should be far from potentially hazardous surroundings. Among the risk factors that must be eliminated are the potential for floods, seismic activity, as well as "man-made" obstacles like airplane flight paths or chemical facilities.

Due to the critical nature of the applications that a data center supports, companies must ensure that the design of their facility (if they wish to build), or that of potential providers if leasing is a consideration, is up to the challenge of meeting their reliability requirements. As we have previously discussed, the tier system of the Uptime Institute can serve as a valuable guide in developing a data center design, or evaluating a providers', that meets an organization's uptime requirements.

4.7.4 Redundancy

The concept of "uptime" was pioneered by the Uptime Institute and codified in its Tier Classification System. In this system, there are four levels (I, II, III, and IV). Within this system, the terms "N, $N+1$, and $2N$" typically refer to the number of power and cooling components that comprise the entire data center infrastructure systems. "N" is the minimum rating of any component (such as a UPS or cooling unit) required to support the site's critical load. An "N" system is nonredundant, and the failure of any component will cause an outage. "N" systems are categorized as Tier I. $N+1$ and $2N$ represent increasing levels of component redundancies and power paths that map to Tiers II–IV. It is important to note, however, that the redundancy of components does not ensure compliance with the Uptime Institute's Tier level [2].

4.7.5 Operations

Besides redundancy, the ability to do planned maintenance or emergency repairs on systems may involve the necessity to take them off-line. This requires that the data center supports the concept of "concurrent maintainability." Concurrent maintainability permits the systems to be bypassed without impacting the availability of the existing computing equipment. This is one of the key criteria necessary for a data center to receive Tier III or IV certification from the Uptime Institute.

4.7.6 Build Versus Buy Using Financial Considerations

The choice to build or lease should include a thorough analysis of the data center's compliance with these Tier requirements to ensure that it is capable of providing the reliable operation necessary to support mission-critical applications.

Another major consideration for businesses in making a build-versus-lease decision is the customer's financial requirements and plans. Oftentimes, these considerations are driven by the businesses' financial organizations. Building a data center is a capital-intensive venture. Companies considering this option must answer a number of questions including:

- Do they have the capital available?
- What is the internal cost of money within the organization?
- How long do they intend to operate the facility?
- What depreciation schedules do they intend to use?

Oftentimes, the internal process of obtaining capital can be long and arduous. The duration of this allocation and approval process must be weighed against the estimated time that the data center is required. Very often, there is also no guarantee that the funds requested will be approved, thereby stopping the project before it starts.

The cost of money (analogous to interest) is also an important element in the decision-making process to build a data

center. The accumulated costs of capital for a data center project must be viewed in comparison with other potential allocations of the same level of funding. In other words, based on our internal interest rate, are we better-off investing the same amount of capital in another project or instrument that will deliver a higher return on the company's investment?

The return on investment question must address a number of factors, not the least of which is the length of time the customer intends to operate the facility and how they will write down this investment over time. If the projected life span for the data center is relatively short, less than 10 years, for example, but the company knows it will continue to have to carry the asset on its books beyond that, building a facility may not be the most advantageous choice.

Due to the complexity of building a data center and obtaining the required capital, many businesses have come to view the ability to lease their required capacity from either a wholesale provider or hosting firm as an easier way to obtain the space they need. By leasing their data center space, companies avoid the need to use their own capital and are able to use their operational (OpEx) budgets to fund their data center requirements. By using this OpEx approach, the customer is able to budget for the expenses spelled out within their lease in the annual operation budget.

The other major consideration that customers must take into account in making their build-versus-lease decision is the timetable for the delivery of the data center. Building a data center can typically take 18–24 months (and often longer) to complete, while most wholesale providers or hosting companies can have their space ready for occupancy in 6 months or less.

4.7.7 The Challenges of Build or Buy

The decision to lease or own a data center has long-term consequences that customers should consider. In a leased environment, a number of costs that would normally be associated with owning a data center are included in the monthly lease rate. For example, in a leased environment, the customer does not incur the expense of the facility's operational or security personnel. The maintenance, both interior and exterior, of the site is also included in the lease rate. Perhaps most importantly, the customer is not responsible for the costs associated with the need to replace expensive items like generators or UPS systems. In short, in a leased environment, the customer is relieved of the responsibility for the operation and maintenance of the facility itself. They are only responsible for the support of the applications that they are running within their leased space.

While the cost and operational benefits of leasing a data center space are attractive, many customers still choose to own their own facilities for a variety of reasons that may best be categorized under the term "flexibility."

For all of the benefits found within a leased offering, some companies find that the very attributes that make these cost-effective solutions are too restrictive for their needs. In many instances, businesses, based on their experiences or corporate policies, find that their requirements cannot be addressed by prospective wholesale or hosting companies. In order to successfully implement their business models, wholesale or hosting providers cannot vary their offerings to use customer-specified vendors, customize their data center designs, or change their operational procedures. This vendor-imposed "inflexibility" therefore can be an insurmountable obstacle to businesses with very specific requirements.

4.8 FUTURE TRENDS

The need for data centers shows no signs of abating in the next 5–10 years. The amount of data generated on a daily basis and the user's desire to have instantaneous access to it will continue to drive requirements for more computing hardware for the data centers to store it in. With the proliferation of new technologies like cloud computing and big data, combined with a recognized lack of space, it is obvious that demand will continue to outpace supply.

This supply and demand imbalance has fostered the continuing entry of new firms into both the wholesale and hosting provider marketplace to offer customers a variety of options to address their data center requirements. Through the use of standardized designs and advanced building technologies, the industry can expect to see continued downward cost pressure on the providers themselves if they are to continue to offer competitive solutions for end users. Another result of the combined effects of innovations in design and technology will be an increasing desire on the part of end customers to have their data centers located where they need them. This will reflect a movement away from large data centers being built only in major metropolitan areas to meet the needs of provider's business models to a more customer-centric approach in which new data centers are designed, built, and delivered to customer-specified locations with factory-like precision. As a result, we shall see not only a proliferation of new data centers over the next decade but also their location in historically nontraditional locations.

This proliferation of options, coupled with continually more aggressive cost reduction, will also precipitate a continued decline in the number of organizations electing to build their own data centers. Building a new facility will simply become too complex and expensive an option for businesses to pursue.

4.9 CONCLUSION

The data center industry is young and in the process of an extended growth phase. This period of continued innovation and competition will provide end customers with significant

benefits in terms of cost, flexibility, and control. What will not change during this period, however, is the need for potential customers to continue to use the fundamental concepts outlined in this chapter during their evaluation processes and in making their final decisions. Stability in terms of a provider's ability to deliver reliable long-term solutions will continue to be the primary criteria for vendor evaluation and selection.

REFERENCES

[1] Building Commissioning Association. Available at http://www.bcxa.org/. Accessed on July, 2020.

[2] Data Center Knowledge. Executive Guide Series, *Build versus Buy*, p. 4.

FURTHER READING

Crosby C. The Ergonomic Data Center: Save Us from Ourselves in Data Center Knowledge. Available at https://www.datacenterknowledge.com/archives/2014/03/05/ergonomic-data-center-save-us. Accessed on September 3, 2020.

Crosby C. Data Centers Are Among the Most Essential Services A glimpse into a post-COVID world. https://www.missioncriticalmagazine.com/topics/2719-unconventional-wisdom. Accessed September 3, 2020.

Crosby C. Questions to Ask in Your RFP in Mission Critical Magazine. Available at http://www.missioncriticalmagazine.com/articles/86060-questions-to-ask-in-your-rfp. Accessed on September 3, 2020.

Crosby C, Godrich K. Data Center Commissioning and the Myth of the Phased Build in Data Center Journal. Available at http://cp.revolio.com/i/148754. Accessed on September 3, 2020.

SOURCES FOR DATA CENTER INDUSTRY NEWS AND TRENDS

Data Center Knowledge. Available at www.datacenterknowledge.com. Accessed on September 3, 2020.

Mission Critical Magazine. Available at www.missioncriticalmagazine.com. Accessed on September 3, 2020.

Web Host Talk. Available at https://www.webhostingtalk.com/. Accessed on September 3, 2020.

5

CLOUD AND EDGE COMPUTING

JAN WIERSMA
EVO Venture Partners, Seattle, Washington, United States of America

5.1 INTRODUCTION TO CLOUD AND EDGE COMPUTING

The terms "cloud" and "cloud computing" have become essential part of the information technology (IT) vocabulary in recent years, after gaining its first popularity in 2009. Cloud computing generally refers to the delivery of computing services like servers, storage, databases, networking, applications, analytics, and more over the Internet, with the aim to offer flexible resources, economies of scale, and more business agility.

5.1.1 History

The concept of delivering compute resources using a global network has its roots in the "Intergalactic Computer Network" concept created by J.C.R. Licklider in the 1960s. Licklider was the first director of the Information Processing Techniques Office (IPTO) at the US Pentagon's ARPA, and his concept inspired the creation of ARPANET, which later became the Internet. The concept of delivering computing as a public utility business model (like water or electricity) can be traced back to computer scientist John McCarthy who proposed the idea in 1961 during a speech given to celebrate MIT's (Massachusetts Institute of Technology) centennial.

As IT evolved, the technical elements needed for today's cloud computing evolved, but the required Internet bandwidth to provide these services reliably only emerged in the 1990s.

The first milestone for cloud computing was the 1999 launch of Salesforce.com providing the first concept of enterprise application delivery using the Internet and a web browser. In the years that followed, many more companies released their browser-based enterprise applications including Google with Google Apps and Microsoft launching Office 365.

Besides application delivery, IT infrastructure concepts also made their way into cloud computing concepts with Amazon Web Services (AWS) launching Simple Storage Service (S3) and the Elastic Compute Cloud (EC2) in 2006. These services enabled companies and individuals to rent storage space and compute on which to run their applications.

Easy access to cheap computer chips, memory, storage, and sensors, as enabled by the rapidly developing smartphone market, allowed companies to extend the collection and processing of data into the edges of their network. The development was assisted by the availability of cheaper and more reliable mobile bandwidth. Examples include industrial applications like sensors in factories, commercial applications like vending machines and delivery truck tracking, and consumer applications like kitchen appliances with remote monitoring, all connected using mobile Internet access. This extensive set of applications is also known as the Internet of Things (IoT), providing the extension of Internet connectivity into physical devices and everyday objects.

As these physical devices started to collect more data using various sensors and these devices started to interact more with the physical world using various forms of output, they also needed to be able to perform analytics and information creation at this edge of the network. The delivery of computing capability at the edges of the network helps to improve performance, cost, and reliability and is known as edge computing.

By virtue of both cloud and edge computing being a metaphor, they are and continue to be open to different

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

interpretations. As a lot of marketing hype has surrounded both cloud and edge computing in recent years, both terms are many times incorrectly applied. It is therefore important to use independently created, non-biased definitions when trying to describe these two important IT concepts.

5.1.2 Definition of Cloud and Edge Computing

The most common definition of cloud computing has been created by the US National Institute of Standards and Technology (NIST) in their Special Publication 800-145 released in September 2011 [1]:

> Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

The NIST definition is intended to provide a baseline for discussion on what cloud computing is and how it can be used, describing essential characteristics.

As edge computing is still evolving, the boundaries of the definition are yet to be defined. The Linux Foundation started in June 2018 to create an Open Glossary of Edge Computing containing the most commonly used definition [2]:

> The delivery of computing capabilities to the logical extremes of a network in order to improve the performance, operating cost and reliability of applications and services. By shortening the distance between devices and the cloud resources that serve them, and also reducing network hops, edge computing mitigates the latency and bandwidth constraints of today's Internet, ushering in new classes of applications. In practical terms, this means distributing new resources and software stacks along the path between today's centralized datacenters and the increasingly large number of devices in the field, concentrated, in particular, but not exclusively, in close proximity to the last mile network, on both the infrastructure and device sides.

5.1.3 Fog and Mist Computing

As the need for computing capabilities near the edge of the network started to emerge, different IT vendors began to move away from the popular term cloud computing and introduced different variations on "cloud." These include, among others, "fog" and "mist" computing. While these terms cover different computing models, they are mostly niche focused, and some vendor-specific and covered mainly by either cloud or edge computing lexicons of today.

As new technology trends emerge and new hypes get created in the IT landscape, new terms arise to describe them and attempt to point out the difference between current and future technology. The creation of a common language, terminology and the standards that go with them, will always be behind the hype that it is trying to describe.

5.2 IT STACK

To understand any computing model in the modern IT space, it is essential first to understand what is needed to provide a desired set of features to an end user. What is required to provide an end user with an app on a mobile phone or web-based email on their desktop? What are all the different components that are required to come together to deliver that service and those features? While there are many different models to explain what goes into a modern IT stack (Fig. 5.1), most of them come down to:

Facility: The physical data center location, including real estate, power, cooling, and rack space required to run IT hardware.

Network: The connection between the facility and the outside world (e.g. Internet), as well as the connectivity within the facility, all allowing the remote end user to access the system functions.

Compute and storage: The IT hardware, consisting of servers with processors, memory, and storage devices.

Virtualization: Using a hypervisor program that allows multiple operating systems (OS) and applications to share single hardware components like processor, memory, and storage.

OS: Software that supports the computer's basic functions, such as scheduling tasks, executing applications, and controlling peripherals. Examples are Microsoft Windows, Linux, and Unix.

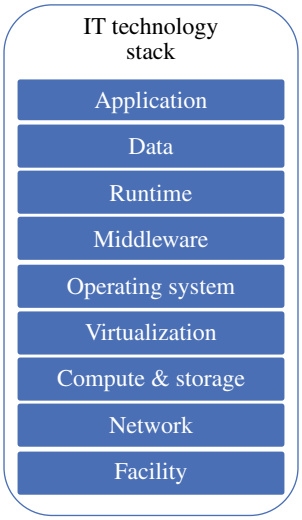


FIGURE 5.1 Layers of an IT technology stack diagram.

Middleware: Software that acts as a bridge between the OS, databases, and applications, within one system or across multiple systems.

Runtime: The runtime environment is the execution environment provided to an application by OS.

Data: Computer data is information stored or processed by a computer.

Application: The application is a program or a set of programs that allow end users to perform a set of particular functions. This is where the end user interacts with the system and the business value of the whole stack is generated.

All of these layers together are needed to provide the functionality and business value to the end user. Within a modern IT stack, these different layers can live at various locations and can be operated by different vendors.

5.3 CLOUD COMPUTING

There are a few typical characteristics of cloud computing that are important to understand:

Available over the network: Cloud computing capabilities are available over the network by a wide range of devices including mobile phones, tablets, and PC workstations. While this seems obvious, it is an often overlooked characteristic of cloud computing.

Rapid elasticity: Cloud computing capabilities can scale rapidly outward and inward with demand (elastically), sometimes providing the customer with a sense of unlimited capacity. The elasticity is needed to enable the system to provision and clear resources for shared use, including components like memory, processing, and storage. Elasticity requires the pooling of resources.

Resource pooling: In a cloud computing model, computing resources are pooled to serve multiple customers in a multi-tenant model. Virtual and physical resources get dynamically assigned based on customer demand. The multi-tenant model creates a sense of location independence, as the customer does not influence the exact location of the provided resources other than some higher-level specification like a data center or geographical area.

Measured service: Cloud systems use metering capabilities to provide usage reporting and transparency to both user and provider of the service. The metering is needed for the cloud provider to analyze consumption and optimize usage of the resources. As elasticity and resource pooling only work if cloud users are incentivized to release resources to the pool, metering by the concept of billing acts as a financial motivator, creating a resource return response.

On-demand self-service: The consumer is able to provision the needed capabilities without requiring human interaction with the cloud provider. This can typically be done by a user interface (UI), using a web browser, enabling the customer to control the needed provisioning or by an application programming interface (API). APIs allow software components to interact with one another without any human involvement, enabling easier sharing of services. Without the ability to consume cloud computing over the network, using rapid elasticity and resource pooling, on-demand self-service would not be possible.

5.3.1 Cloud Computing Service and Deployment Models

Cloud computing helps companies focus on what matters most to them, with the ability to avoid non-differentiating work such as procurement, maintenance, and infrastructure capacity planning. As cloud computing evolved, different service and deployment models emerged to meet the needs of different types of end users. Each model provides different levels of control, flexibility, and management to the customer, allowing the customer to choose the right solution for a given business problem (Fig. 5.2).

5.3.1.1 Service Models

Infrastructure as a Service: (IaaS) allows the customer to rent basic IT infrastructure including storage, network, OS, and computers (virtual or dedicated hardware), on a pay-as-you-go basis. The customer is able to deploy and run its own software on the provided infrastructure and has control over OS, storage, and limited control of select networking components. In this model, the cloud provider manages the facility to the virtualization layers of the IT stack, while the customer is responsible for the management of all layers above virtualization.

Platform as a service: (PaaS) provides the customer with an on-demand environment for developing, testing, and managing software applications, without the need to set up and manage the underlying infrastructure of servers, storage, and network. In this model, the cloud provider operates the facility to the runtime layers of the IT stack, while the customer is responsible for the management of all layers above the runtime.

Software as a service: (SaaS) refers to the capability to provide software applications over the Internet, managed by the cloud provider. The provider is responsible for the setup, management, and upgrades of the application, including all the supporting infrastructure. The application is typically accessible using

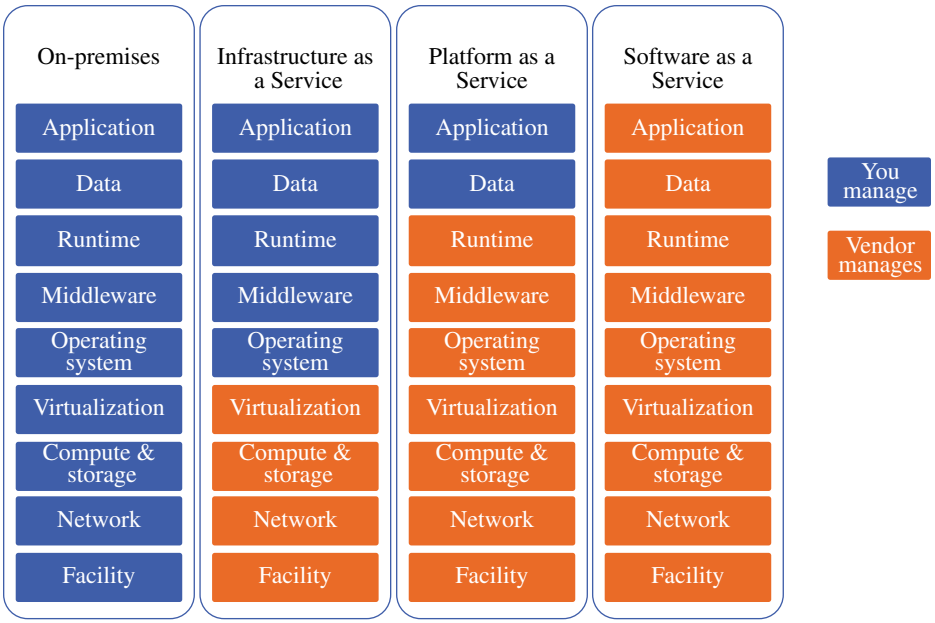


FIGURE 5.2 Diagram of ownership levels in the IT stack.

a web browser or other thin client interface (e.g. smartphone apps). The customer only has control over a limited set of application-specific configuration settings.

In this model, the cloud provider manages all layers of the IT stack.

5.3.1.2 Deployment Models

Public Cloud: Public cloud is owned and operated by a cloud service provider. In this model, all hardware, software, and supporting infrastructure are owned and managed by the cloud provider, and it is operated out of the provider's data center(s). The resources provided are made available to anyone for use, based on pay-as-you-go or for free.

Examples of public cloud providers include AWS, Microsoft Azure, Google Cloud, and Salesforce.com

Private Cloud: Private cloud refers to cloud computing resources provisioned exclusively by a single business or organization. It can be operated and managed by the organization, by a third party, or a combination of them. The deployment can be located at the customer's own data center (on premises) or in a third-party data center.

The deployment of computing resources on premises, using virtualization and resource management tools, is sometimes called "private cloud." This type of deployment provides dedicated resources, but it does not provide all of the typical cloud characteristics. While traditional IT infrastructure can benefit from modern virtualization and application management technologies to optimize utilization and increase flexibility, there is a very thin line between this type of deployment and true private cloud.

Hybrid Cloud: Hybrid cloud is a combination of public and private cloud deployments using technology that allows infrastructure and application sharing between them. The most common hybrid cloud use case is the extension of on-premises infrastructure into the cloud for growth, allowing it to utilize the benefits of cloud while optimizing existing on-premises infrastructure.

Most enterprise companies today are using a form of hybrid cloud. Typically they will use a collection of public SaaS-based applications like Salesforce, Office 365, and Google Apps, combined with public or private IaaS deployments for their other business applications.

Multi-cloud: As more cloud providers entered the market within the same cloud service model, companies started to deploy their workloads across these different provider offerings. A company may have compute workloads running on AWS and Google Cloud at the same time to ensure a best of breed solution for their different workloads. Companies are also using a multi-cloud approach to continually evaluate various providers in the market or hedge their workload risk across multiple providers.

Multi-cloud, therefore, is the deployment of workloads across different cloud providers within the same service model (IaaS/PaaS/SaaS).

5.3.2 Business View

Cloud computing is for many companies a significant change in the way they think about and consume IT resources. It has had a substantial impact on the way IT resources are used to create a competitive advantage, impacting business agility, speed of execution, and cost.

5.3.2.1 Cloud Computing Benefits

By moving to pay-as-you-go models, companies have moved their spend profile from massive upfront data center and IT infrastructure investments to only paying for what they use when they use it. This limits the need for significant long-term investments with no direct return. As in a traditional model the company would need to commit to the purchased equipment capacity and type for its entire life span, a pay-as-you-go model eliminates the associated risk in a fast-moving IT world. It has also lowered the barrier for innovation in many industries that rely on compute or storage-intensive applications. Where in the past many of these applications were only available to companies that could spend millions of dollars upfront on data centers and IT equipment, the same capacity and features are now available to a small business using a credit card and paying only for the capacity when they use it.

The cost is for the actual consumption of these cloud resources lowered by the massive economies of scale that cloud providers can achieve. As cloud providers aggregate thousands of customers, they can purchase their underlying IT infrastructure at a lower cost, which translates into lower pay-as-you-go prices.

The pay-as-you-go utility model also allows companies to worry less about capacity planning, especially in the area of idle resources. As a company is developing new business supporting applications, it is often hard to judge the needed IT infrastructure capacity for this new application, leading to either overprovisioning or underprovisioning of resources. In the traditional IT model, these resources would sit idle in the company's data center, or it would take weeks to months to add needed capacity. With cloud computing, companies can consume as much or as little capacity as they need, scaling up or down in minutes.

The ability to easily provision resources on demand typically extends across the globe. Most cloud providers offer their services in multiple regions, allowing their customers to go global within minutes. This eliminates substantial investment, procurement, and build cycles for data centers and IT infrastructure to unlock a new region in the world.

As in cloud computing new IT resources are on demand, only a few clicks away, it means companies can make these resources available quicker to their employees, going from weeks of deployment time to minutes. This has increased agility and speed for many organizations, as the time to experiment and develop is significantly lowered.

The different cloud service models have also allowed businesses to spend less time on IT infrastructure-related work, like racking and stacking of servers, allowing more focus to solve higher-level business problems.

5.3.2.2 Cloud Computing Challenges

While the business benefits for cloud computing are many, it requires handing over control of IT resources to the cloud provider. The loss of control, compared with the traditional IT model, has sparked a lot of debate around the security and privacy of the data stored and handled by cloud providers.

As both the traditional model and the cloud computing model are very different in their architecture, technologies used, and the way they are managed, it is hard to compare the two models based on security truthfully. A comparison is further complicated by the high visibility of cloud providers' security failures, compared with companies running their own traditional IT on premises. Many IT security issues originate in human error, showing that technology is only a small part of running a secure IT environment. It is therefore not possible to state that either traditional on-premises IT is more or less secure than cloud computing and vice versa. It is known that cloud providers typically have more IT security- and privacy-related certifications than companies running their own traditional IT, which means cloud providers have been audited more and are under higher scrutiny by lawmakers and government agencies. As cloud computing is based on the concept of resource pooling and multi-tendency, it means that all customers benefit from the broad set of security and privacy policies, technologies, and controls that are used by cloud providers across the different businesses they serve.

5.3.2.3 Transformation to Cloud Computing

With cloud computing having very different design philosophies compared with traditional IT, not all functionality can just be lifted and shifted from on-premises data centers and traditional IT infrastructure, expecting to work reliably in the cloud. This means companies need to evaluate their individual applications to asses if they comply with the reference architectures provided by the cloud providers to ensure the application will continue to run reliably and cost effectively. Companies also need to evaluate what cloud service model (IaaS, PaaS, SaaS) they would like to adopt for a given set of functionality. This should be done based on the strategic importance of the functionality to the business and the value of data it contains. Failure in cloud computing adoption is typically the result of not understanding how IT

designs need to evolve to work in and with the cloud, what cloud service model is applicable for the desired business features, and how to select the right cloud provider that fits with the business.

5.3.3 Technology View

As cloud computing is a delivery model for a broad range of IT functionality, the technology that powers cloud computing is very broad and spans from IT infrastructure to platform services like databases and artificial intelligence (AI), all powered by different technologies.

There are many supporting technologies underpinning and enabling cloud computing, and examples include virtualization, APIs, software-defined networking (SDN), microservices, and big data storage models. Supporting technologies also extend to new hardware designs with custom field-programmable gate array (FPGA) computer chips and to new ways of power distribution in the data center.

5.3.3.1 Cloud Computing: Architectural Principles

With so many different layers of the IT stack involved and so many innovative technologies powering those layers, it is interesting to look at IT architectural principles commonly used for designing cloud computing environments:

Simplicity: Be it either the design of a physical data center, its hardware, or the software build on top of it to power the cloud; they all benefit from starting simple, as successful complex systems always evolve from simple systems. Focus on basic functions, test, fix, and learn.

Loose coupling: Design the system in a way that reduces interdependencies between components. This design philosophy helps to avoid changes or failure in one component to affect others and can only be done by having well-defined interfaces between them. It should be possible to modify underlying system operations without affecting other components. If a component failure does happen, the system should be able to handle this gracefully, helping to reduce impact. Examples are queueing systems that can manage queue buildup after a system failure or component interactions that understand how to handle error messages.

Small units of work: If systems are built in small units of work, each focused on a specific function, then each can be deployed and redeployed without impacting the overall system function. The work unit should focus on a highly defined, discrete task, and it should be possible to deploy, rebuild, manage, and fail the unit without impacting the system. Building these small units helps to focus on *simplicity*, but can only be successful when they are *loosely coupled*. A popular way of achieving this in software architecture is the microservices design philosophy.

Compute resources are disposable: Compute resources should be treated as disposable resources while always being consistent and tested. This is typically done by implementing immutable infrastructure patterns, where components are replaced rather than changed. When a component is deployed, it never gets modified, but instead gets redeployed when needed due to, for example, failure or a new configuration.

Design for failure: Things will fail all the time: software will have bugs, hardware will fail, and people will make mistakes. In the past IT systems, design would focus on avoidance of service failure by pushing as much redundancy as (financially) possible into designs, resulting in very complicated and hard-to-manage services. Running reliable IT services at massive scale is notoriously hard, forcing an IT design rethink in the last few years. Risk acceptance and focusing on the ability to restore the service quickly have shown to be a better IT design approach. *Simple, small, disposable* components that are *loosely coupled* help to design for failure.

Automate everything: As both cloud providers and their customers are starting to deal with systems at scale, they are no longer able to manage these systems manually. Cloud computing infrastructure enables users to deploy and modify using on-demand self-service. As these self-service points are exposed using APIs, it allows components to interact without human intervention. Using monitoring systems to pick up signals and orchestration systems for coordination, automation is used anywhere from auto-recovery to auto-scaling and lifecycle management.

Many of these architectural principles have been captured in the Reactive Manifesto, released in 2014 [3], and the Twelve-Factor App [4], first published in 2011.

5.3.4 Data Center View

Technological development in the field of data center and infrastructure relating to cloud computing is split between two areas: on-premises deployments and public cloud provider deployments.

The on-premises deployments, often referred to as private cloud, have either been moving to standard rackmount server and storage hardware combined with new software technology like OpenStack or Microsoft Azure Stack or more packaged solutions. As traditional hardware deployments have not always provided customers with the cloud benefits needed due to management overhead, converged infrastructure solutions have been getting traction in the market. A converged infrastructure solution packages networking,

servers, storage, and virtualization tools on a turnkey appliance for easy deployment and management.

As more and more compute consumption moved into public cloud computing, a lot of technical innovation in the data center and IT infrastructure has been driven by the larger public cloud providers in recent years. Due to the unprecedented scale at which these providers have to operate, their large data centers are very different from traditional hosting facilities. Individual "pizza box" servers or single server applications no longer work in these warehouses full of computers. By treating these extensive collections of systems as one massive warehouse-scale computer (WSC) [5], these providers can provide the levels of reliability and service performance business and customers nowadays expect. In order to support thousands of physical servers, in these hyperscale data centers, cloud providers had to develop new ways to deploy and maintain their infrastructure, maximizing the compute density while minimizing the cost of power, cooling, and human labor.

If one were running a cluster of 10,000 physical servers, that would have stellar reliability for the hardware components used; it would still mean that in a given year on average, one server would fail every day. In order to manage hardware failure in WSCs, cloud providers started with different rack server designs to enable more straightforward swap out of failed servers and general lower operational cost. As part of a larger interconnected system, WSC servers are low-end server based, built in tray or blade enclosure format. Racks that hold together tens of servers and supporting infrastructure like power conversion and delivery cluster these servers into a single rack compute unit. The physical racks can be a completely custom design by the cloud provider enabling specific applications for compute, storage, or machine learning (ML). Some cloud providers cluster these racks into 20–40-ft shipping containers using the container as a deployment unit within the WSC.

5.3.4.1 Open-Source Hardware and Data Center Designs

The 2011 launched Open Compute Project [6] contains detailed specifications of the used racks and hardware components used by companies like Facebook, Google, and Microsoft to build their WSCs. As these hardware designs, as well as many software component designs for hyperscale computing, have been made open source, it has enabled broader adoption and quicker innovation. Anyone can use, modify, collaborate, and contribute back to these custom designs using the open-source principles.

Examples of contributions include Facebook's custom designs for servers, power supplies, and UPS units and Microsoft's Project Olympus for new rack-level designs. LinkedIn launched its open data center effort with its launch of the Open19 Foundation in 2016 [7]. Networking has seen similar open-source and collaboration initiatives in, for example, the OpenFlow initiative [8].

5.3.4.2 Cloud Computing with Custom Hardware

As the need for more compute power started to rise due to the increase in cloud computing consumption, cloud providers also began to invest in custom hardware chips and components. General-purpose CPUs have been replaced or supported by FPGAs or application-specific integrated circuits (ASICs) in these designs. These alternative architectures and specialized hardware like FPGAs and ASICs can provide cloud providers with cutting-edge performance to keep up with the rapid pace of innovation.

One of the innovation areas that cloud providers have responded to is the wide adoption of deep learning models and real-time AI, requiring specialized computing accelerators for deep learning algorithms. While this type of computing started with widely deployed graphical processing units (GPUs), several cloud providers have now build their own custom chips. Examples include the Google tensor processing unit (TPU) and Microsoft's Project Catapult for FPGA usage.

5.3.4.3 Cloud Computing: Regions and Zones

As cloud computing is typically delivered across the world and cloud vendors need to mitigate the risk of one WSC (data center) going offline due to local failures, they usually split their offerings across regions and zones. While cloud vendor-specific implementations may differ, typically regions are independent geographic areas that consist of multiple zones. The zone is seen as a single failure domain, usually composed of one single data center location (one WSC), within a region. This enables deployment of fault-tolerant applications across different zones (data centers), providing higher availability. To protect against natural disasters impacting a specific geographic area, applications can be deployed across multiple regions.

Cloud providers may also provide managed services that are distributed by default across these zones and regions, providing redundancy without the customer needing to manage the associated complexity. As a result, these services have constraints and trade-offs on latency and consistency, as data is synchronized across multiple data centers spread across large distances.

To be able to achieve a reliable service, cloud providers have not only built their own networking hardware and software but also invested in worldwide high-speed network links, including submarine cables across continents.

The scale at which the largest cloud providers operate has forced them to rethink the IT hardware and infrastructure they use to provide reliable services. At hyperscale these providers have encountered unique challenges in networking

and computing while trying to manage cost, sparking innovation across the industry.

5.4 EDGE COMPUTING

Workloads in IT have been changing over the years, moving from mainframe systems to client/server models, on to the cloud and in recent years expanding to edge computing recently. The emergence of the IoT has meant devices started to interact more with the physical world, collecting more data and requiring faster analysis and bandwidth to operate successfully.

The model for analyzing and acting on data from these devices by edge computing technologies typically involves:

- Capturing, processing, and analyzing time-sensitive data at the network edge, close to the source
- Acting on data in milliseconds
- Using cloud computing to receive select data for historical analysis, long-term storage, and training ML models

The device is a downstream compute resource that can be anything from a laptop, tablet to cars, environmental sensors, or traffic lights. These edge devices can be single function focused or fully programmable compute nodes, which live in what is called the "last mile network" that delivers the actual service to the end consumer.

A model where edge devices are not dependent on a constant high bandwidth connection to a cloud computing backend does not only eliminate network latency problems and lowers cost but also improves reliability as local functionality is less impacted by disconnection from the cloud.

The independence of functions living in a distant cloud data center needs to be balanced with the fact that edge devices, in general, are limited in memory size, battery life, and heat dissipation (limiting processing power). More advanced functions, therefore, need to offload energy-consuming compute to the edge of the network.

The location of this offload processing entirely depends on what business problem it needs to solve, leading to a few different edge computing concepts.

5.4.1 Edge Computing Initiatives

Network edge-type processing and storage concepts date back to the 1990s' concept of content delivery networks (CDNs) that aimed to resolve Internet congestion by caching website content at the edges of the network. Cisco recognized the growing amount of Internet-enabled devices on networks and launched in 2012 the concept of fog computing. It assumes a distributed architecture positioning compute and storage at its most optimal place between the IoT device and the centralized cloud computing resources. The effort was consolidated in the OpenFog Consortium in 2015 [9]. The European Telecommunications Standards Institute (ETSI) launched the idea for multi-access edge computing (MEC) in 2014 [10], aiming to deliver standard MEC and APIs supporting third-party applications. MEC is driven by the new generation of mobile networks (4G and 5G) requiring the deployment of applications at the edge due to low latency. The Open Edge Computing (OEC) Initiative launched in 2015 [11] has taken the approach of cloudlets, just-in-time provisioning of applications to the edge compute nodes, and dynamic handoff of virtual machines from one node to the next depending on the proximity of the consumer. Cloud providers like AWS, Google, and Microsoft have also entered the IoT and edge computing space using hubs to easily connect for device–cloud and device–device communication with a centralized data collection and analysis capabilities. Overall edge computing has received interest from infrastructure, network, and cloud operators alike, all looking to unlock its potential in their own way.

Computing at the network edge is currently approached in different architectural ways, and while these four (OpenFog, ETSI MEC, OEC, and cloud providers) highlight some of the most significant initiatives, they are not the only concepts in the connecting circle to Edge device block evolving edge computing field. Overall one can view all these as initiatives and architectural deployment options as a part of edge computing and its lexicon.

In general, these concepts either push computing to the edge device side of the last mile, in network layer near the device, or bridge the gap from the operator side of the network and a central location (Fig. 5.3). The similarity

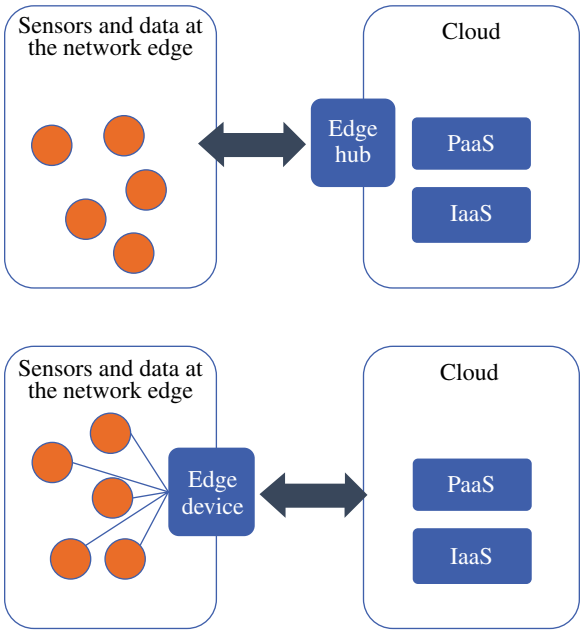


FIGURE 5.3 Edge device/compute concepts.

between the different approaches is enabling third parties to deploy applications and services on the edge computing infrastructure using standard interfaces, as well as the openness of the projects themselves allowing collaboration and contribution from the larger (vendor) community. The ability to successfully connect different devices, protocols, and technologies in one seamless edge computing experience is all about interoperability that will only emerge from open collaboration.

Given the relative complexity of the edge computing field and the ongoing research, selection of the appropriate technology, model, or architecture should be done based on specific business requirements for the application, as there is no one size fits all.

5.4.2 Business View

Edge computing can be a strategic benefit to a wide range of industries, as it covers industrial, commercial, and consumer application of the IoT and extends to advanced technologies like autonomous cars, augmented reality (AR), and smart cities. Across these, edge computing will transform businesses as it lowers cost, enables faster response times, provides more dependable operation, and allows for interoperability between devices. By allowing processing of data closer to the device, it reduces data transfer cost and latency between the cloud and the edge of the network. The availability of local data also allows for faster processing and gaining actionable insights by reducing round-trips between edge and the cloud. Having instantaneous data analysis has allowed autonomous vehicles to avoid collisions and has prevented factory equipment from failure. Smart devices that need to operate with a very limited or unreliable Internet connection depend on edge computing to operate without disruption. This unlocks deployments in remote locations such as ships, airplanes, and rural areas. The wide field of IoT has also required interoperability between many different devices and protocols, both legacy and new, to make sure historical investment is protected and adoption can be accelerated.

5.4.2.1 Edge Computing: Adoption Challenges

Most companies will utilize both edge computing and cloud computing environments for their business requirements, as edge computing should be seen as complementary to cloud computing—real-time processing data locally at the device while sending select data to the cloud for analysis and storage.

The adoption of edge computing still is not a smooth path. IoT device adoption has shown to have longer implementation durations with higher cost than expected. Especially the integration into legacy infrastructure has seen significant challenges, requiring heavy customizations. The initial lack of vendor collaboration has also slowed down IoT adoption. This has been recognized by the vendor community, resulting in the different "open" consortiums like OpenFog and OEC for edge computing in 2015. The collaboration need also extends into the standards and interoperability space with customers pushing vendors to work on common standards. IoT security, and the lack thereof, has also slowed down adoption requiring more IoT and edge computing-specific solutions than expected.

While these challenges have provided a slower than expected adoption of IoT and edge computing, it is clear that the potential remains huge and the technology is experiencing growing pains. Compelling new joint solutions have started to emerge like the Kinetic Edge Alliance [12] that combines wireless carrier networks, edge colocation, hosting, architecture, and orchestration tools in one unified experience for the end user across different vendors. Given the current collaboration between vendors, with involvement from academia, combined with the massive investments made, these challenges will be overcome in the next few years.

5.4.3 Technology View

From a technology usage perspective, many elements make up the edge computing stack. In a way, the stack resembles a traditional IT stack, with the exception that devices and infrastructure can be anywhere and anything.

Edge computing utilizes a lot of cloud native technology as its prime enabler, as well as deployment and management philosophies that have made cloud computing successful. An example is edge computing orchestration that is required to determine what workloads to run where in a highly distributed infrastructure. Compared with cloud computing architectures, edge computing provides cloud-like capabilities but at a vast number of local points of presence and not as infinitely scalable as cloud. Portability of workloads and API usage are other examples. All this allows to extend the control plane to edge devices in the field and process workloads at the best place for execution, depending on many different criteria and policies set by the end user. This also means the end user defines the edge actual boundaries, depending on business requirements.

To serve these business purposes, there are different approaches to choose from, including running fog-type architectures, or standardized software platforms like the MEC initiative, all on top of edge computing infrastructure.

One of the major difference between architecting a solution for the cloud and one for edge computing is handling of program state. Within cloud computing the application model is stateless, as required by the abstractions of the underlying technologies used. Cloud computing also uses the stateless model to enable application operation at scale, allowing many servers to execute the same service

simultaneously. Processing data collected from the physical world with the need to process instantly is all about stateful processing. For applications deployed on edge computing infrastructure, this requires careful consideration of design choices and technology selection.

5.4.3.1 Edge Computing: Hardware

Another essential technology trend empowering edge computing is advancements made in IT hardware. Examples include smart network interface cards (NICs) that help to offload work from the central edge device's processor, Tesla's full self-driving (FSD) computer optimized to run neural networks to read the road, to Google's Edge TPU that provides a custom chip to optimize high-quality ML for AI.

Powered by these new hardware innovations, network concepts like network functions virtualization (NFV) enabling virtualization at the edge and federated ML models allowing the data to stay at the edge for near-real-time analysis are quickly advancing the field of edge computing.

5.4.4 Data Center View

Data centers for edge computing are the midpoint between the edge device and the central cloud. They are deployed as close as possible to the edge of the network to provide low latency. While edge data centers perform the same functions as a centralized data center, they are smaller in size and distributed across many physical locations. Sizes typically vary between 50 and 150 kW of power consumption. Due to their remote nature, they are "lights out"—operating autonomously with local resilience. The deployment locations are nontraditional like at the base of cellular network towers. Multiple edge data centers may be interconnected into a mesh network to provide shared capacity and failover, operating as one virtual data center.

Deployment examples include racks inside larger cellular network towers, 20- or 40-ft shipping containers, and other nontraditional locations that provide opportunities for nonstandard data center technologies like liquid cooling and fuel cells.

5.5 FUTURE TRENDS

5.5.1 Supporting Technology Trends: 5G, AI, and Big Data

Several technology trends are supporting cloud and edge computing:

Fifth generation wireless or 5G is the latest iteration of cellular technology, capable of supporting higher network speeds, bandwidth, and more devices per square kilometer. The first significant deployments are launched in April 2019. With 5G providing the bandwidth needed for more innovative edge device usage, edge computing in the last mile of the network needs to deliver low latency to localized compute resources. 5G technology also allows the cellular network providers to treat their radio access network (RAN) as "intelligent." This will enable mobile network providers, for example, to provide multiple third-party tenants to use their base stations, enabling new business and commercial models.

Big data is the field of extracting information from large or complex data sets and analysis. Solutions in the big data space aim to make capturing, storing analysis, search, transfer, and visualizing easier while managing cost. Supporting technologies include Apache Hadoop, Apache Spark, MapReduce, and Apache HBase. Several cloud computing providers have launched either storage or platform services around big data, allowing customers to focus on generating business value from data, without the burden of managing the supporting technology. As capturing and storing large amounts of data has become relatively easy, the focus of the field has shifted to data science and ML.

AI is a popular term for intelligence demonstrated by machines. For many in the current IT field, the most focus and investment is given to a specific branch of the AI technology stack—ML. ML allows building computer algorithms that will enable computer programs to improve through experience automatically and is often mislabeled as AI due to the "magic" of its outcomes. Examples of successful ML deployment include speech recognition, machine language translation, and intelligent recommendation engines. ML is currently a field of significant research and investment, with massive potential in many industries.

5.5.2 Future Outlook

The IT landscape has significantly changed in the last 10 years (2009–2019). Cloud-based delivery and consumption models have gone mainstream, powering a new wave of innovation across industries and domains. The cloud-based delivery of infrastructure, platform and software (IaaS, PaaS, SaaS), has enabled companies to focus on solving higher-level business problems by eliminating the need for large upfront investments and enabling business agility.

The consumption of cloud computing has also been accelerated by an economical paradox called the Jevons effect; when technology progress increases the efficiency of a resource, the rate of consumption of that resource rises due to increasing demand. The relative low cost and low barrier of entry for cloud usage have fueled a massive consumption growth.

This growth has seen the emergence of new companies like AWS and Google Cloud (GCP), the reboot of large companies like Microsoft (with Azure), and the decline of other large IT vendors that could not join the race to cloud on time.

Modern-day start-ups start their business with cloud consumption from the company inception, and most of them never move to their own data centers or co-lo data centers during their company growth. Most of them do end up consuming cloud services across multiple providers, ending up in a multi-cloud situation. Examples include combinations of SaaS services like Microsoft Office 365, combined with IaaS services from AWS and GCP.

Many larger enterprises still have sunken capital in their own data centers or in co-lo data center setups. Migrations from these data centers into the cloud can encounter architectural and/or financial challenges, limiting the ability to quickly eliminate these data centers from the IT portfolio. For most enterprises this means they end up managing a mixed portfolio of own/co-lo data centers combined with multiple cloud providers for their new application deployments. The complexity of managing IT deployments across these different environments is one of the new challenge's IT leaders will face in the next few years.

One attempt to address some of these challenges can be found in the emergence of serverless and container-type architectures and technologies to help companies with easier migration to and between cloud platforms.

The introduction of AI delivered as a service, and more specific ML, has allowed companies of all sizes to experiment at scale with these emerging technologies without significant investment. The domain of storage of large data sets, combined with ML, will accelerate cloud consumption in the next few years.

Edge computing will also continue to see significant growth, especially with the emergence of 5G wireless technology, with the management of these new large highly distributed edge environments being one of the next challenges. Early adaptors of edge computing maybe inclined to deploy their own edge locations, but the field of edge computing has already seen the emergence of service providers offering cloud delivery-type models for edge computing including pay-as-you-go setups.

Overall data center usage and growth will continue to rise, while the type of actual data center tenants will change, as well as some of the technical requirements. Tenant types will change from enterprise usage to service providers, as enterprises move their focus to cloud consumption. Examples of changes in technical requirements include specialized hardware for cloud service delivery and ML applications.

REFERENCES

[1] NIST. NIST 800-145 publication. Available at https://csrc.nist.gov/publications/detail/sp/800-145/final. Accessed on October 1, 2019.

[2] LF Edge Glossary. Available at https://github.com/lf-edge/glossary/blob/master/edge-glossary.md. Accessed on October 1, 2019.

[3] Reactive Manifesto. Available at https://www.reactivemanifesto.org. Accessed on October 1, 2019.

[4] 12factor app. Available at https://12factor.net. Accessed on October 1, 2019.

[5] Barraso LA, Clidaras J, Holzle U. The datacenter as a computer. Available at https://www.morganclaypool.com/doi/10.2200/S00874ED3V01Y201809CAC046. Accessed on October 1, 2019.

[6] Open Compute Project. Available at https://www.opencompute.org/about. Accessed on October 1, 2019.

[7] Open19 Foundation. Available at https://www.open19.org/. Accessed on October 1, 2019.

[8] OpenFlow. Available at https://www.opennetworking.org/. Accessed on October 1, 2019.

[9] Openfog. Available at https://www.openfogconsortium.org/. Accessed on October 1, 2019.

[10] ETSI MEC. Available at https://www.etsi.org/technologies/multi-access-edge-computing. Accessed on October 1, 2019.

[11] OEC. Available at http://openedgecomputing.org/about.html. Accessed on October 1, 2019.

[12] Kinetic Edge Alliance. Available at https://www.vapor.io/kinetic-edge-alliance/. Accessed on October 1, 2019.

FURTHER READING

Barraso LA, Clidaras J, Holzle U. *The Datacenter as a Computer*. 2009/2013/2018.

Building the Internet of Things: Implement New Business Models, Disrupt Competitors, Transform Your Industry ISBN-13: 978-1119285663.

Carr N. *The Big Switch: Rewiring the World, from Edison to Google*. 1st ed.

Internet of Things for Architects: Architecting IoT Solutions by Implementing Sensors, Communication Infrastructure, Edge Computing, Analytics, and Security ISBN-13: 978-1788470599.

6

DATA CENTER FINANCIAL ANALYSIS, ROI, AND TCO

LIAM NEWCOMBE
Romonet, London, United Kingdom

6.1 INTRODUCTION TO FINANCIAL ANALYSIS, RETURN ON INVESTMENT, AND TOTAL COST OF OWNERSHIP

Anywhere you work in the data center sector, from an enterprise business that operates its own data centers to support business activities, a colocation service provider whose business is to operate data centers, a cloud provider that delivers services from data centers, or for a company that delivers products or services to data center operators, any project you wish to carry out is likely to need a business justification. In the majority of cases, this business justification is going to need to be expressed in terms of the financial return the project will provide to the business if they supply the resources and funding. Your proposals will be tested and assessed as investments, and therefore, you need to be able to present them as such.

In many cases, this will require you to not only simply assess the overall financial case for the project but also deal with split organizational responsibility or contractual issues, each of which can prevent otherwise worthwhile projects from going ahead. This chapter seeks to introduce not only the common methods of Return on Investment (ROI) and Total Cost of Ownership (TCO) assessment but also how you may use these tools to prioritize your limited time, resources, and available budget toward the most valuable projects.

A common mistake made in many organizations is to approach an ROI or TCO analysis as being the justification for engineering decisions that have already been made; this frequently results in the selection of the first project option to exceed the hurdle set by the finance department. To deliver the most effective overall strategy, project analysis should consider both engineering and financial aspects to identify the most appropriate use of the financial and personnel resources available. Financial analysis is an additional set of tools and skills to supplement your engineering skill set and enable you to provide a better selection of individual projects or overall strategies for your employer or client.

It is important to remember as you perform or examine others' ROI analysis that any forecast into the future is inherently imprecise and requires us to make one or more estimations. An analysis that uses more data or more precise data is not necessarily any more accurate as it will still be subject to this forecast variability; precision should not be mistaken for accuracy. Your analysis should clearly state the inclusions, exclusions, and assumptions made in your TCO or ROI case and clearly identify what estimates of delivered value, future cost, or savings you have made; what level of variance should be expected in these factors; and how this variance may influence the overall outcome. Equally, you should look for these statements in any case prepared by somebody else, or the output is of little value to you.

This chapter provides an introduction to the common financial metrics used to assess investments in the data center and provides example calculations. Some of the common complications and problems of TCO and ROI analysis are also examined, including site and location sensitivity. Some of the reasons why a design or project optimized for data center A is not appropriate for data center B or C and why the vendor case studies probably don't apply to your data center are considered. These are then brought together in an example ROI analysis for a realistic data center reinvestment scenario where multiple options are assessed and the presented methods used to compare the project options.

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

The chapter closes with a discussion from a financial perspective of likely future trends in data centers. The changing focus from engineering to financial performance accelerated by the threat of cloud and commoditization is discussed along with the emergence of energy service and guaranteed energy performance contracts. A sample of existing chargeback models for the data center is reviewed, and their relative strengths and weaknesses compared. The impact on data centers of the current lack of effective chargeback models is examined in terms of the prevalent service monoculture problem. The prospect of using Activity-Based Costing (ABC) to break out of this trap provides effective unit costing and fosters the development of a functioning internal market for enterprise operators, and per customer margin management for service providers is examined. The development from our current, energy-centric metric, PUE, toward more useful overall financial performance metrics such as cost per delivered IT kWh is discussed, and lastly, some of the key points to consider when choosing which parts of your data center capacity should be built, leased, colocated, or deployed in the cloud are reviewed.

This chapter provides a basic introduction to the financial analysis methods and tools; for a more in-depth treatment of the subject, a good management finance text should be consulted such as Wiley's "Valuation: Measuring and Managing the Value of Companies" (ISBN 978-0470424704).

6.1.1 Market Changes and Mixed ICT Strategies

Data centers are a major investment for any business and present a series of unusual challenges due to their combination of real estate, engineering, and information technology (IT) demands. In many ways, a data center is more like a factory or assembly plant than any normal business property or operation. The high power density, high cost of failure, and the disconnect between the 20+ year investment horizons on the building and major plant and the 2–5-year technology cycle on the IT equipment all serve to make data centers a complex and expensive proposition.

The large initial capital cost, long operational cost commitments, high cost of rectifying mistakes, and complex technology all serve to make data centers a relatively specialist, high risk area for most businesses. At the same time, as data centers are becoming more expensive and more complex to own, there is a growing market of specialist providers offering everything from outsourced management for your corporate data center to complete services rented by the user hour. This combination of pressures is driving a substantial change in the views of corporate CFOs, CIOs, and CEOs on how much of their IT estate they should own and control.

There is considerable discussion in the press of IT moving to a utility model like power or water in which IT services are all delivered by specialist operators from a "cloud" and no enterprise business needs to own any servers or employ any IT staff. One of the key requirements for this utility model is that the IT services are completely homogeneous and entirely substitutable for each other, which is clearly not presently the case. The reality is likely to be a more realistic mix of commercial models and technology.

Most businesses have identified that a substantial part of their IT activity is indeed commodity and represents little more than an overhead on their cost of operating; in many cases, choosing to obtain these services from a specialist service provider is a sensible choice. On the other hand, most businesses also have something that they believe differentiates them and forms part of their competitive advantage. In a world where the Internet is the majority of media for customer relationships and more services are delivered electronically, it is increasingly common to find that ICT is an important or even a fundamental part of that unique competitive advantage. There are also substantial issues with application integration when many independent providers of individual specific service components are involved as well as security, legal, risk, and regulatory compliance concerns. Perhaps the biggest threat to cloud adoption is the same vendor lock-in problem businesses currently face with their internal applications where it is difficult or impossible to effectively move the data built up in one system to another.

In reality, most enterprise businesses are struggling to find the right balance of cost, control, compliance, security, and service integration. They will find their own mix of in-house data center capacity, owned IT equipment in colocation facilities, and IT purchased as a service from cloud providers.

Before any business can make an informed decision on whether to build a service in their own data center capacity or outsource it to a cloud provider, they must be able to assess the cost implications of each choice. A consistent and unbiased assessment of each option that includes the full costs over the life cycle is an essential basis for this decision that may then be considered along with the deployment time, financial commitment, risk, and any expected revenue increase from the project.

6.1.2 Common Decisions

For many organizations, there is a substantial, and ever growing, range of options for their data center capacity against which any option or investment may be tested by the business:

- Building a new data center
- Capacity expansion of an existing data center
- Efficiency improvement retrofit of an existing data center
- Sale and leaseback of an existing data center
- Long-term lease of private capacity in the form of wholesale colocation (8+ years)

- Short-term lease of shared capacity in the form of retail colocation
- Medium-term purchase of a customized service on dedicated IT equipment
- Medium-term purchase of a commodity service on dedicated IT equipment
- Short-term purchase of a commodity service on provider-owned equipment

For each project, the relative costs of delivery internally will increasingly need to be compared with the costs of partial or complete external delivery. Where a project requires additional capital investment to private data center capacity, it will be particularly hard to justify that investment against the individually lower capital costs of external services.

6.1.3 Cost Owners and Fragmented Responsibility

ICT and, particularly, data center cost are subject to an increasing level of scrutiny in business, largely due to the increased fraction of the total business budget that is absorbed by the data center. As this proportion of cost has increased, the way in which businesses treat IT and data center cost has also started to change. In many organizations, the IT costs were sufficiently small to be treated as part of the shared operating overhead and allocated across consuming parts of the business in the same way that the legal or tax accounts department costs would be spread out. This treatment of costs failed to recognize any difference in the cost of IT services supporting each function and allowed a range of suboptimal behaviors to develop.

A common issue is for the responsibility and budget for the data center and IT to be spread across a number of separate departments that do not communicate effectively. It is not uncommon for the building to be owned and the power bill paid by the corporate real estate (CRE) group, a facilities group to own and manage the data center mechanical and electrical infrastructure, while another owns the IT hardware, and individual business units are responsible for the line of business software. In these situations, it is very common for perverse incentives[1] to develop and for decisions to be made, which optimize that individual department's objectives or cost at the expense of the overall cost to the business.

A further pressure is that the distribution of cost in the data center is also changing, though in many organizations the financial models have not changed to reflect this. In the past, the data center infrastructure was substantially more expensive than the total power cost over the data center lifetime, while both of these costs were small compared to the IT equipment that was typically purchased from the end user department budget. In the past few years, IT equipment capital cost has fallen rapidly, while the performance yield from each piece of IT equipment has increased rapidly. Unfortunately, the power efficiency of IT equipment has not improved at the same rate that capital cost has fallen, while the cost of energy has also risen and for many may continue on its upward path. This has resulted in the major cost shifting away from the IT hardware and into the data center infrastructure and power. Many businesses have planned their strategy based on the apparently rapidly falling cost of the server, not realizing the huge hidden costs they were also driving.[2]

In response to this growth and redistribution of data center costs, many organizations are now either merging responsibility and strategy for the data center, power, and IT equipment into a single department or presenting a direct cross-charge for large items such as data center power to the IT departments. For many organizations, this, coupled with increasing granularity of cost from external providers, is the start of a more detailed and effective chargeback model for data center services.

Fragmented responsibility presents a significant hurdle for many otherwise strong ROI cases for data center investment that may need to overcome in order to obtain the budget approval for a project. It is common to find issues, both within a single organization and between organizations, where the holder of the capital budget does not suffer the operational cost responsibility and vice versa. For example:

- The IT department does not benefit from the changes to airflow management practices and environmental control ranges, which would reduce energy cost because the power cost is owned by CRE.
- A wholesale colocation provider has little incentive to invest or reinvest in mechanical and electrical equipment, which would reduce the operational cost of the data center as this is borne by the lease-holding tenant who, due to accounting restrictions, probably cannot invest in capital infrastructure owned by a supplier.

To resolve these cases of fragmented responsibility, it is first necessary to make realistic and high confidence assessments of the cost and other impacts of proposed changes to provide the basis for a negotiation between the parties. This may be a matter of internal budget holders taking a joint case to the CFO, which is deemed to be in the business overall interests, or it may be a complex customer–supplier contract and service level agreement (SLA) issue that requires commercial negotiations. This aspect will be explored in more detail under the Section 6.4.8.6.

[1] A perverse incentive occurs when a target or reward program, instead of having the desired effect on behavior, produces unintended and undesirable results contrary to the goals of those establishing the target or reward.

[2] C. Belady, "In the data center, power and cooling costs more than the it equipment it supports," Electronic Cooling, February 2007. http://www.electronics-cooling.com/2007/02/in-the-data-center-power-and-cooling-costs-more-than-the-it-equipment-it-supports/

6.1.4 What Is TCO?

TCO is a management accounting concept that seeks to include as many of the costs involved in a device, product, service, or system as possible to provide the best available decision-making information. TCO is frequently used to select one from a range of similar products or services, each of which would meet the business needs, and in order to minimize the overall cost. For example, the 3-year TCO of a server may be used as the basis for a service provider pricing a managed server or for cross-charge to consuming business units within the same organization.

As a simple example, we may consider a choice between two different models of server that we wish to compare for our data center, one is more expensive but requires less power and cooling than the other; the sample costs are shown in Table 6.1.

On the basis of this simplistic TCO analysis, it would appear that the more expensive server A is actually cheaper to own than the initially cheaper server B. There are, however, other factors to consider when we look at the time value of money and Net Present Value (NPV), which are likely to change this outcome.

When considering TCO, it is normal to include at least the first capital cost of purchase and some element of the operational costs, but there is no standard definition of which costs you should include in a TCO analysis. This lack of definition is one of the reasons to be careful with TCO and ROI analyses provided by other parties; the choices made regarding the inclusion or exclusion of specific items can have a substantial effect on the outcome, and it is as important to understand the motivation of the creator as their method.

6.1.5 What Is ROI?

In contrast to TCO, an ROI analysis looks at both costs and incomes and is commonly used to inform the decision whether to make a purchase at all, for example, whether it makes sense to upgrade an existing device with a newer, more efficient device.

TABLE 6.1 Simple TCO example, not including time

Costs	Server A	Server B
Capital purchase	$2,000	$1,500
3-year maintenance contract	$900	$700
Installation and cabling	$300	$300
3-year data center power and cooling capacity	$1,500	$2,000
3-year data center energy consumption	$1,700	$2,200
3-year monitoring, patches, and backup	$1,500	$1,500
TCO	$7,900	$8,200

In the case of an ROI analysis, the goal is, as for TCO, to attempt to include all of the relevant costs, but there are some substantial differences:

- The output of TCO analysis is frequently used as an input to an ROI analysis.
- ROI analysis is typically focused on the difference between the costs of alternative actions, generally "what is the difference in my financial position if I make or do not make this investment?"
- Where a specific cost is the same over time between all assessed options, omission of this cost has little impact and may simplify the ROI analysis, for example, a hard-to-determine staff cost for support and maintenance of the device.
- Incomes due to the investment are a key part of ROI analysis; for example, if the purchased server is to be used to deliver charged services to customers, then differences in capacity that result in differences in the per server income are important.

We may consider an example of whether to replace an existing old uninterruptible power supply (UPS) system with a newer device, which will both reduce the operational cost and address a constraint on data center capacity, allowing a potential increase in customer revenue, as shown in Table 6.2.

In this case, we can see that the balance is tipped by the estimate of the potential increase in customer revenue available after the upgrade. Note that both the trade-in rebate of the new UPS from the vendor and the estimate of increased customer revenue are of the opposite sign to the costs. In this case, we have shown the costs as negative and the income as positive. This is a common feature of ROI analysis; we treat all costs and income as cash flows in or out of our analysis; whether costs are signed positive or negative only makes a difference to how we explain and present our output, but they should be of the opposite sign to incomes. In this case, we present the answer as follows: "The ROI of the $100,000 new UPS upgrade is $60,000 over 10 years."

As for the simple TCO analysis, this answer is by no means complete as we have yet to consider how the values change over time and is thus unlikely to earn us much credit with the CFO.

6.1.6 Time Value of Money

While it may initially seem sensible to do what is presented earlier in the simple TCO and ROI tables and simply add up the total cost of a project and then subtract the total cost

TABLE 6.2 Simple ROI example, not including time

Income received or cost incurred	Existing UPS upgrade	New UPS	Difference
New UPS purchase	$0	−$100,000	−$100,000
New UPS installation	$0	−$10,000	−$10,000
Competitive trade-in rebate for old UPS	$0	$10,000	$10,000
UPS battery costs (old UPS also requires replacement batteries)	−$75,000	−$75,000	$0
10-year UPS service and maintenance contract	−$10,000	−$5,000	$5,000
Cost of power lost in UPS inefficiency	−$125,000	−$50,000	$75,000
Additional customer revenue estimate	$0	$80,000	$80,000
Total	−$210,000	−$150,000	$60,000

saving or additional revenue growth, this approach does not take into account what economists and business finance people call the "time value of money."

At a simple level, it is relatively easy to see that the value of a certain amount of money, say $100, depends on when you have it; if you had $100 in 1900, this would be considerably more valuable than $100 now. There are a number of factors to consider when we need to think about money over a time frame.

The first factor is inflation; in the earlier example, the $100 had greater purchasing power in 1900 than now due to inflation, the rise in costs of materials, energy, goods, and services between then and now. In the context of a data center evaluation, we are concerned with how much more expensive a physical device or energy may become over the lifetime of our investment.

The second factor is the interest rate that could be earned on the money; the $100 placed in a deposit account with 5% annual interest would become $105 at the end of year 1, $110.25 in year 2, $115.76 in year 3, and so on. If $100 was invested in a fixed interest account with 5% annual interest in 1912, when RMS Titanic departed from Southampton, the account would have increased to $13,150 by 2012 and in a further 100 years in 2112 would have become $1,729,258 (not including taxes or banking fees). This nonlinear impact of compound interest is frequently the key factor in ROI analysis.

The third factor, one that is harder to obtain a defined number, or even the method agreed for, is risk. If we invest the $100 in April on children's toys that we expect to sell from a toy shop in December, we may get lucky and be selling the must-have toy; alternatively, we may find ourselves selling most of them off at half price in January. In a data center project, the risk could be an uncertain engineering outcome affecting operational cost savings, uncertainty in the future cost of energy, or potential variations in the customer revenue received as an outcome of the investment.

6.1.7 Cost of Capital

When we calculate the Present Value (PV) of an investment option, the key number we will need for our calculation is the discount rate. In simple examples, the current interest rate is used as the discount rate, but many organizations use other methods to determine their discount rate, and these are commonly based on their cost of capital; you may see this referred to as the Weighted Average Cost of Capital (WACC).

The cost of capital is generally given in the same form as an interest rate and expresses the rate of return that the organization must achieve from any investment in order to satisfy its investors and creditors. This may be based on the interest rate the organization will pay on loans or on the expected return on other investments for the organization. It is common for the rate of ROIs in the normal line of business to be used for this expected return value. For example, an investment in a data center for a pharmaceuticals company might well be evaluated against the return on investing in new drug development.

There are various approaches to the calculation of cost of capital for an organization, all of which are outside the scope of this book. You should ask the finance department of the organization to whom you are providing the analysis what discount rate or cost of capital to use.

6.1.8 ROI Period

Given that the analysis of an investment is sensitive to the time frame over which it is evaluated, we must consider this time frame. When we are evaluating a year one capital cost against the total savings over a number of years, both the number of years' savings we can include and the discount rate have a significant impact on the outcome. The ROI period will depend on both the type of project and the accounting practices in use by the organization whose investment you are assessing.

The first aspect to consider is what realistic lifetime the investment has. In the case of a reinvestment in a data center that is due to be decommissioned in 5 years, we have a fairly clear outer limit over which it is reasonable to evaluate savings. Where the data center has a longer or undefined lifetime, we can consider the effective working life of the devices affected by our investment. For major elements of data center infrastructure such as transformers, generators, or chillers, this can be 20 years or longer, while for other elements such as computer room air conditioning/computer room air handling (CRAC/CRAH) units, the service lifetime may be shorter, perhaps 10–15 years. Where the devices have substantial periodic maintenance costs such as UPS battery refresh, these should be included in your analysis if they occur within the time horizon.

One key consideration in the assessment of device lifetime is proximity to the IT equipment. There are a range of devices such as rear door and in-row coolers that are installed very close to the IT equipment, in comparison with traditional devices such as perimeter CRAC units or air handling units (AHUs). A major limiting factor on the service lifetime of data center infrastructure is the rate of change in the demands of the IT equipment. Many data centers today face cooling problems due to the increase in IT power density. The closer coupled an infrastructure device is to the IT equipment, the more susceptible it is likely to be to changes in IT equipment power density or other demands. You may choose to adjust estimates of device lifetimes to account for this known factor.

In the case of reinvestment, particularly those designed to reduce operational costs by improving energy efficiency, the allowed time frame for a return is likely to be substantially shorter; NPV analysis durations as short as 3 years are not uncommon, while others may calculate their Internal Rate of Return (IRR) with savings "to infinity."

Whatever your assessment of the service lifetime of an investment, you will need to determine the management accounting practices in place for the organization and whether there are defined ROI evaluation periods, and if so, which of these is applicable for the investment you are assessing. These defined ROI assessment periods are frequently shorter than the device working lifetimes and are set based on business, not technical, criteria.

6.1.9 Components of TCO and ROI

When we are considering the TCO or ROI of some planned project in our data center, there are a range of both costs and incomes that we are likely to need to take into account. While TCO focuses on costs, this does not necessarily exclude certain types of income; in an ROI analysis, we are likely to include a broader range of incomes as we are looking for the overall financial outcome of the decision.

It is useful when identifying these costs to determine which costs are capital and which are operational, as these two types of cost are likely to be treated quite differently by the finance group. Capital costs not only include purchase costs but also frequently include capitalized costs occurring at the time of purchase of other actions related to the acquisition of a capital asset.

6.1.9.1 Initial Capital Investment

The initial capital investment is likely to be the first value in an analysis. This cost will include not only the capital costs of equipment purchased but also frequently some capitalized costs associated with the purchase. These might include the cost of preparing the site, installation of the new device(s), and the removal and disposal of any existing devices being replaced. Supporting items such as software licenses for the devices and any cost of integration to existing systems are also sometimes capitalized.

You should consult the finance department to determine the policies in place within the organization for which you are performing the analysis, but there are some general guidelines for which costs should be capitalized.

Costs are capitalized where they are incurred on an asset that has a useful life of more than one accounting period; this is usually one financial year. For assets that last more than one period, the costs are amortized or depreciated over what is considered to be the useful life of the asset. Again, it is important to note that the accounting lifetime and therefore depreciation period of an asset may well be shorter than the actual working life you expect to achieve based on accounting practice or tax law.

The rules on capitalization and depreciation vary with local law and accounting standards; but as a conceptual guide, the European Financial Reporting Standard guidance indicates that the costs of fixed assets should initially be "directly attributable to bringing the asset into working condition for its intended use."

Initial capitalized investment costs for a UPS replacement project might include the following:

- Preparation of the room
- Purchase and delivery
- Physical installation
- Wiring and safety testing
- Commissioning and load testing
- Installation and configuration of monitoring software
- Training of staff to operate the new UPS and software
- Decommissioning of the existing UPS devices
- Removal and disposal of the existing UPS devices

Note that disposal does not always cost money; there may be a scrap value or rebate payment; this is addressed in the additional incomes section that follows.

6.1.9.2 Reinvestment and Upgrade Costs

There are two circumstances in which you would need to consider this second category of capital cost.

The first is where your project does not purchase completely new equipment but instead carries out remedial work or an upgrade to existing equipment to reduce the operating cost, increase the working capacity, or extend the lifetime of the device, the goal being "enhances the economic benefits of the asset in excess of its previously assessed standard of performance." An example of this might be reconditioning a cooling tower by replacing corroded components and replacing the old fixed speed fan assembly with a new variable frequency drive (VFD) controlled motor and fan. This both extends the service life and reduces the operating cost and, therefore, is likely to qualify as a capitalized cost.

The second is where your project will require additional capital purchases within the lifetime of the device such as a UPS system that is expected to require one or more complete replacements of the batteries within the working life in order to maintain design performance. These would be represented in your assessment at the time the cost occurs. In financial terminology, these costs "relate to a major inspection or overhaul that restores the economic benefits of the asset that have been consumed by the entity."

6.1.9.3 Operating Costs

The next major group of costs relates to the operation of the equipment. When considering the operational cost of the equipment, you may include any cost attributable to the ownership and operation of that equipment including staffing, service and maintenance contracts, consumables such as fuel or chemical supplies, operating licenses, and water and energy consumption.

Operating costs for a cooling tower might include the following:

- Annual maintenance contract including inspection and cleaning.
- Cost of metered potable water.
- Cost of electrical energy for fan operation.
- Cost of electrical energy for basin heaters in cold weather.
- Cost of the doping chemicals for tower water.

All operating costs should be represented in the accounting period in which they occur.

6.1.9.4 Additional Income

It is possible that your project may yield additional income, which could be recognized in the TCO or ROI analysis. These incomes may be in the form of rebates, trade-in programs, salvage values for old equipment, or additional revenue enabled by the project. If you are performing a TCO analysis to determine the cost at which a product or service may be delivered, then the revenue would generally be excluded from this analysis. Note that these additional incomes should be recognized in your assessment in the accounting period in which they occur.

Additional income from a UPS replacement project might include the following:

- Salvage value of the existing UPS and cabling.
- Trade-in value of the existing UPS from the vendor of the new UPS devices.
- Utility, state, or government energy efficiency rebate programs where project produces an energy saving that can realistically be shown to meet the rebate program criteria.

6.1.9.5 Taxes and Other Costs

One element that varies greatly with both location and the precise nature of the project is taxation. The tax impact of a project should be at least scoped to determine if there may be a significant risk or saving. Additional taxes may apply when increasing capacity in the form of emissions permits for diesel generators or carbon allowances if your site is in an area where a cap-and-trade scheme is in force, particularly if the upgrade takes the site through a threshold. There may also be substantial tax savings available for a project due to tax rebates, for example, rebates on corporate tax for investing or creating employment in a specific area. In many cases, corporate tax may be reduced through the accounting depreciation of any capital assets purchased. This is discussed further in the Section 6.3.3.

6.1.9.6 End-of-Life Costs

In the case of some equipment, there may be end-of-life decommissioning and disposal costs that are expected and predictable. These costs should be included in the TCO or ROI analysis at the point at which they occur. In a replacement project, there may be disposal costs for the existing equipment that you would include in the first capital cost as it occurs in the same period as the initial investment. Disposal costs for the new or modified equipment at the end of service life should be included and valued as at the expected end of life.

6.1.9.7 Environmental, Brand Value, and Reputational Costs

Costs in this category for a data center project will vary substantially depending on the organization and legislation in the operating region but may also include the following:

- Taxation or allowances for water use.
- Taxation or allowances for electricity use.
- Taxation or allowances for other fuels such as gas or oil.
- Additional energy costs from "green tariffs."
- Renewable energy certificates or offset credits.
- Internal cost of carbon (or equivalent).

There is a demonstrated link between greenhouse gases and the potential impacts of global warming. The operators of data centers come under a number of pressures to control and minimize their greenhouse gas and other environmental impacts.

Popular recognition of the scale of energy use in data centers has led to a substantial public relations and brand value issue for some operators. Governments have recognized the concern; in 2007 the US Environmental Protection Agency presented a report to Congress on the energy use of data centers[3]; in Europe in 2008 the EC launched the Code of Conduct for Data Centre Energy Efficiency.[4]

6.1.10 Green Taxes

Governmental concerns relating to both the environmental impact of CO_2 and the cost impacts of energy security have led to market manipulations that seek to represent the environmental or security cost of energy from certain sources. These generally take the form of taxation, which seeks to capture the externality[5] through increasing the effective cost of energy. In some areas carbon taxes are proposed or have been implemented; at the time of writing only the UK Carbon Reduction Commitment[6] and Tokyo Cap-and-Trade scheme are operating. At a higher level schemes such as the EU Emissions Trading Scheme[7] generally affect data center operators indirectly as electricity generators must acquire allowances and this cost is passed on in the unit cost of electricity. There are few data center operators who consume or generate electricity on a sufficient scale to acquire allowances directly.

6.1.11 Environmental Pressures

Some Non-Governmental Organizations (NGO) have succeeded in applying substantial public pressure to data center operators perceived as either consuming too much energy or energy from the wrong source. This pressure is frequently away from perceived "dirty" sources of electricity such as coal and oil and toward "clean" or renewable sources such as solar and hydroelectric; whether nuclear is "clean" depends upon the political objectives of the pressure group.

In addition to this direct pressure to reduce the carbon intensity of data center energy, there are also efforts to create a market pressure through "scope 3"[8] accounting of the greenhouse gas emissions associated with a data center or a service delivered from that data center. The purpose of this is to create market pressure on data center operators to disclose their greenhouse gas emissions to customers, thereby allowing customers to select services based on their environmental qualities. The major NGO in this area is the Greenhouse Gas Protocol.[9]

In many cases operators have selected alternate locations for data centers based on the type of local power-generating capacity or invested in additional renewable energy generation close to the data center in order to demonstrate their environmental commitment. As these choices directly affect construction and operating costs (in many cases the "dirty" power is cheaper), there needs to be a commercial justification for the additional expense. This justification commonly takes the form of lost trade and damage to the organization's brand value (name, logos, etc.). In these cases, an estimate is made of the loss of business due to adverse publicity or to the reduction in brand value. For many large organizations, the brand has an identifiable and substantial value as it represents the organization and its values to customers; this is sometimes referred to as "goodwill." Damage to this brand through being associated with negative environmental outcomes reduces the value of the company.

6.1.12 Renewable or Green Energy

Some data center operators choose to purchase "renewable" or "zero carbon" energy for their data center and publish this fact. This may be accomplished in a number of ways dependent upon the operating region and source of energy. Those who become subject to a "scope 3" type emissions disclosure may find it easier to reduce their disclosable emissions to zero than to account them to delivered services or customers.

While some operators choose to colocate with a source of renewable energy generation (or a source that meets the local regulations for renewable certification such as combined heat and power), this is not necessary to obtain recognized renewable energy for the data center.

In some cases a "green tariff" is available from the local utility provider. These can take a number of forms but are

[3] http://www.energystar.gov/index.cfm?c=prod_development.server_efficiency_study.

[4] http://iet.jrc.ec.europa.eu/energyefficiency/ict-codes-conduct/data-centres-energy-efficiency.

[5] In this case an externality is a cost that is not borne by the energy consumer but other parties; taxes are applied to externalities to allow companies to modify behavior to address the overall cost of an activity, including those which they do not directly bear without the taxation.

[6] http://www.decc.gov.uk/en/content/cms/emissions/crc_efficiency/crc_efficiency.aspx.

[7] http://ec.europa.eu/clima/policies/ets/index_en.htm.

[8] http://www.ghgprotocol.org/standards/scope-3-standard.

[9] http://www.ghgprotocol.org/.

generally based on the purchase of renewable energy or certificates to equal the consumed kWh on the tariff. Care should be taken with these tariffs as many include allowances or certificates that would have been purchased anyway in order to meet local government regulation and fail to meet the "additionality test" meaning that they do not require additional renewable energy generation to be constructed or to take place. Those that meet the additionality test are likely to be more expensive than the normal tariff.

An alternative approach is to purchase "offset" for the carbon associated with electricity. In most regions, a scheme is in place to allow organizations that generate electricity from "renewable" energy sources or take other actions recognized as reducing carbon to obtain certificates representing the amount of carbon saved through the action. These certificates may then be sold to another organization that "retires" the certificate and may then claim to have used renewable or zero carbon energy. If the data center operator has invested in renewable energy generation at another site, then they may be able to sell the electricity to the local grid as regular "dirty" electricity and use the certificates obtained through generation against the electricity used by their data center. As with green tariffs, care should be taken with offsets as the qualification criteria vary greatly between different regions and offsets purchased may be perceived by NGO pressure groups as being "hostage offsets" or otherwise invalid. Further, the general rule is that offsets should only be used once all methods of reducing energy consumption and environmental impact have already been exhausted. Organizations that are deemed to have used offsets instead of minimizing emissions are likely to gain little, if any, value from the purchase.

6.1.13 Cost of Carbon

In order to simplify the process of making a financial case for a project that reduces carbon or other greenhouse gas emissions, many organizations now specify an internal financial cost for CO_2. Providing a direct cost for CO_2 allows for a direct comparison between the savings from emission reduction due to energy efficiency improvements or alternate sources of energy and the cost of achieving the reductions.

The cost of CO_2 within the organization can vary substantially but is typically based upon one of the following:

- The cost of an emission allowance per kg of CO_2 based on the local taxation or cap-and-trade scheme, this is the direct cost of the carbon to the organization
- The cost of carbon offsets or renewable certificates purchased to cover the energy used by the data center
- The expected loss of business or impact to brand value from a negative environmental image or assessment

Some organizations will assign a substantially higher value to each unit of CO_2 than the current cost of an allowance or offset as a form of investment. This depends upon the view that in the future their customers will be sensitive to the environmental history of the company. Therefore an investment now in reducing environmental impact will repay over a number of future years.

CO_2 is by no means the only recognized greenhouse gas. Other gases are generally converted to CO_2 through the use of a published equivalency table although the quantities of these gases released by a data center are likely to be small in comparison with the CO_2.

6.2 FINANCIAL MEASURES OF COST AND RETURN

When the changing value over time is included in our assessment of project costs and returns, it can substantially affect the outcome and viability of projects. This section provides an introduction and examples for the basic measures of PV and IRR, followed by a short discussion of the relative strengths and weaknesses.

6.2.1 Common Business Metrics and Project Approval Tests

There are a variety of relatively standard financial methods used and specified by management accountants to analyze investments and determine their suitability. It is likely that the finance department in your organization has a preferred metric that you will be expected to use—in many larger enterprises, a template spreadsheet or document is provided that must be completed as part of the submission. It is not unusual for there to be a standard "hurdle" for any investment expressed in terms of this standard calculation or metric such as "all projects must exceed a 30% IRR."

The measures you are most likely to encounter are as follows:

- TCO: Total Cost of Ownership
- NPV: The Net Present Value of an option
- IRR: The Internal Rate of Return of an investment

Both the NPV and IRR are forms of ROI analysis and are described later.

While the essence of these economic hurdles may easily be misread as "We should do any project that exceeds the hurdle" or "We should find the project with the highest ROI metric and do that," there is, unsurprisingly, more to consider than which project scores best on one specific metric. Each has its own strengths and weaknesses, and making good decisions is as much about understanding the relative strengths of the metric as how to calculate them.

6.2.1.1 Formulae and Spreadsheet Functions

In this section, there are several formulae presented; in most cases where you are calculating PV or IRR, there are spreadsheet functions for these calculations that you can use directly without needing to know the formula. In each case, the relevant Microsoft Office Excel function will be described in addition to the formula for the calculation.

6.2.2 Present Value

The first step in calculating the PV of all the costs and savings of an investment is to determine the PV of a single cost or payment. As discussed under time value of money, we need to discount any savings or costs that occur in the future to obtain an equivalent value in the present. The basic formula for the PV of a single payment a at time n accounting periods into the future at discount rate i per period is given by the following relation:

$$\mathrm{PV}_n = \frac{a}{(1+i)^n}$$

In Microsoft Office Excel, you would use the PV function:

$$= \mathrm{PV}(\mathrm{rate}, \mathrm{nper}, \mathrm{pmt}, \mathrm{fv}) = \mathrm{PV}(i, n, 0, -a)$$

We can use this formula or spreadsheet function to calculate the PV (i.e. the value today) of a single income we receive or cost we incur in the future. Taking an income of $1,000 and an interest rate of 10%/annum, we obtain the following:

$$\text{End of year 1 PV} = 1{,}000 \times \frac{1}{(1+0.1)^1} = 1{,}000 \times \frac{1}{(1.1)^1} = 1{,}000 \times \frac{1}{1.1} = 909.09$$

$$\text{End of year 2 PV} = 1{,}000 \times \frac{1}{(1+0.1)^2} = 1{,}000 \times \frac{1}{(1.1)^2} = 1{,}000 \times \frac{1}{1.21} = 826.45$$

$$\text{End of year 3 PV} = 1{,}000 \times \frac{1}{(1+0.1)^3} = 1{,}000 \times \frac{1}{(1.1)^3} = 1{,}000 \times \frac{1}{1.331} = 751.31$$

If we consider an annual income of $1,000 over 10 years, with the first payment at the end of this year, then we obtain the series of PVs shown in Table 6.3 for our $1,000/year income stream.

The values of this series of individual $1,000 incomes over a 20-year period are shown in Figure 6.1.

Figure 6.1 shows that the PVs of the incomes reduce rapidly at our 10% discount rate toward a negligible value. If we plot the total of the annual income PVs over a 50-year period, we see that the total tends toward $10,000 as shown in Figure 6.2.

TABLE 6.3 PV of $1,000 over 10 years at 10% discount rate

Year	1	2	3	4	5	6	7	8	9	10
Income	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000
Scalar	0.91	0.83	0.75	0.68	0.62	0.56	0.51	0.47	0.42	0.39
Present value at 10%	$909.09	$826.45	$751.31	$683.01	$620.92	$564.47	$513.16	$466.51	$424.10	$385.54

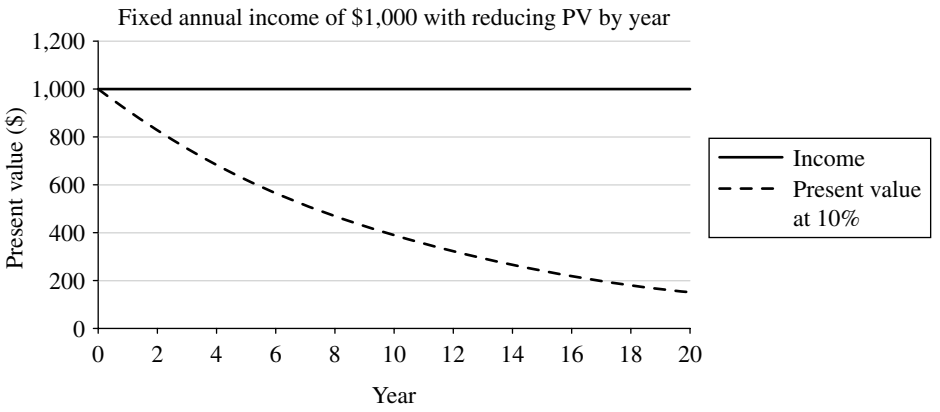


FIGURE 6.1 PV of $1,000 annual incomes at 10% interest rate.

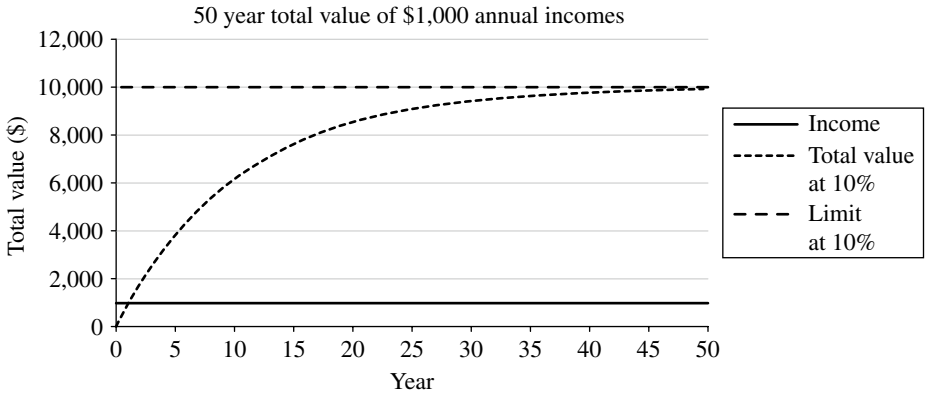


FIGURE 6.2 Total value of $1,000 incomes at 10% interest rate over varying periods.

This characteristic of the PV is important when assessing the total value of savings against an initial capital investment; at higher discount rates, increasing the number of years considered for return on the investment has little impact. How varying the interest rate impacts the PVs of the income stream is shown in Figure 6.3.

As the PV of a series of payments of the same value is a geometric series, it is easy to use the standard formulae for the sum to n terms and to infinity to determine the total value of the number of payments PV_A or the sum of a perpetual series of payments PV_P that never stops:

$$PV_A = \frac{a}{i}\left(1 - \frac{1}{(1+i)^n}\right) = PV(\text{rate}, \text{nper}, \text{pmt}) = PV(i, n, -a)$$

$$PV_P = \frac{a}{i}$$

Note: In Excel, the PV function uses payments, not incomes; to obtain a positive value from the PV function, we must enter incomes as negative payments.

TABLE 6.4 Value of $1,000 incomes over varying periods and discount rates

Discount rate	5 years	10 years	20 years	Perpetual
1%	$4,853	$9,471	$18,046	$100,000
5%	$4,329	$7,722	$12,462	$20,000
10%	$3,791	$6,145	$8,514	$10,000
15%	$3,352	$5,019	$6,259	$6,667
20%	$2,991	$4,192	$4,870	$5,000
30%	$2,436	$3,092	$3,316	$3,333

Using these formulae, we can determine the value of a perpetual series of $1,000 income over any period for any interest rate as shown in Table 6.4.

These values may be easily calculated using the financial functions in most spreadsheets; in Microsoft Office Excel, the PV function takes the argument PV (Interest Rate, Number of Periods, Payment Amount).

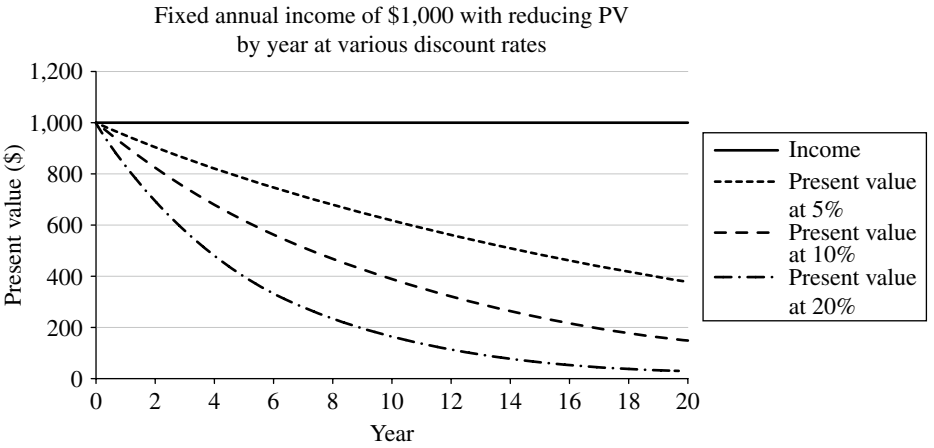


FIGURE 6.3 PV of $1,000 annual incomes at varied interest rates.

To calculate the value to 10 years of the $1,000 annual payments at 5% discount rate in a spreadsheet, we can use

$$= \mathbf{PV}\left(\mathbf{0.05}, \mathbf{10}, \mathbf{-1,000}\right) \quad \text{which returns } \$7,721.73$$

6.2.3 Net Present Value

To calculate the NPV of an investment, we need to consider more than just a single, fixed value, saving over the period; we must include the costs and savings, in whichever accounting period they occur, to obtain the overall value of the investment.

6.2.3.1 Simple Investment NPV Example

As an example, if an energy saving project has a $7,000 implementation cost, yields $1,000 savings/year, and is to be assessed over 10 years, we can calculate the income and resulting PV in each year as shown in Table 6.5.

The table shows one way to assess this investment. Our initial investment of $7,000 is shown in year zero as this money is spent up front, and therefore, the PV is—$7,000. We then have a $1,000 accrued saving at the end of each year for which we calculate the PV based on the 5% annual discount rate. Totaling these PVs gives the overall NPV of the investment as $722.

Alternatively, we can calculate the PV of each element and then combine the individual PVs to obtain our NPV as shown in Table 6.6; this is an equivalent method, and choice depends on which is easier in your particular case.

The general formula for NPV is as follows:

$$\mathrm{NPV}\left(i, N\right) = \sum_{n=0}^{N} \frac{R_n}{\left(1+i\right)^n} \begin{array}{l} = \mathrm{NPV}\left(\text{rate, value 1, value 2}, \ldots\right) \\ = \mathrm{NPV}\left(i, R_1, R_2, \ldots\right) \end{array}$$

where

R_t = the cost incurred or income received in period t,
i = the discount rate (interest rate),
N = the number of costs or income periods,
n = the time period over which to evaluate NPV.

In the Excel formula, R_1, R_2, etc. are the individual costs or incomes. Note that in the Excel, the first cost or income is R_1 and not R_0, and therefore one period's discount rate is applied to the first value; we must handle the year zero capital cost separately.

6.2.3.2 Calculating Break-Even Time

Another common request when forecasting ROI is to find the time (if any) at which the project investment is equaled by the incomes or savings of the project to determine the break-even time of the project. If we simply use the cash flows, then the break-even point is at 7 years where the total income of $7,000 matches the initial cost. The calculation becomes more complex when we include the PV of the project incomes as shown in Figure 6.4.

Including the impact of discount rate, our break-even points are shown in Table 6.7.

As shown in the graph and table, the break-even point for a project depends heavily on the discount rate applied to the analysis. Due to the impact of discount rate on the total PV of the savings, it is not uncommon to find that a project fails to achieve breakeven over any time frame despite providing ongoing returns that appear to substantially exceed the implementation cost.

As for the NPV, spreadsheets have functions to help us calculate the break-even point; in Microsoft Office Excel, we can use the NPER (discount rate, payment, PV) function but only for constant incomes. Once you consider any aspect of a project that changes over time, such as the energy tariff or planned changes in IT load, you are more likely to have to calculate the annual values and look for the break-even point manually.

6.2.4 Profitability Index

One of the weaknesses of NPV as an evaluation tool is that it gives no direct indication of the scale of return compared with the initial investment. To address this, some organizations use a simple variation of the NPV called profitability index, which simply divides the PV of the incomes by the initial investment.

TABLE 6.5 Simple investment example as NPV

Year	0	1	2	3	4	5	6	7	8	9	10
Cost	$7,000										
Savings		$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000
Annual cost or savings	−$7,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000
PV at 5%	−$7,000	$952	$907	$864	$823	$784	$746	$711	$677	$645	$614
NPV at year 0											$722

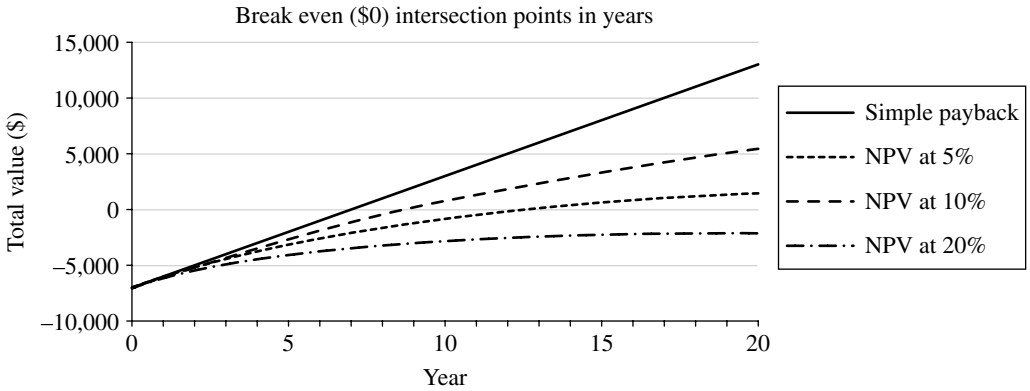


FIGURE 6.4 Breakeven of simple investment example.

TABLE 6.6 Calculate combined NPV of cost and saving

	Amount	Periods	Discount rate	Present value
Cost	$7,000			–$7,000
Saving	–$1,000	10	5%	$7,722
NPV				$722

TABLE 6.7 Break-even point of simple investment example under varying discount rates

Case	Break-even years	Formula
Simple payback	7.0	=NPER(0, –1,000, 7,000)
NPV = 0 at 5%	8.8	=NPER(0.05, –1,000, 7,000)
NPV = 0 at 10%	12.6	=NPER(0.1, –1,000, 7,000)
NPV = 0 at 20%	#NUM!	=NPER(0.2, –1,000, 7,000)

The general formula for profitability index is as follows:

$$\text{Profitability index} = \frac{\text{PV}(\text{future incomes})}{\text{Intial investment}} \begin{array}{l} = \text{NPV}(\text{rate, value 1, value 2,}\ldots)/\text{investiment} \\ = \text{NPV}(i, N_1, N_2, \ldots)/\text{investiment} \end{array}$$

where i is the discount rate (interest rate) and N_1 and N_2 are the individual costs or incomes.

For our simple investment example presented earlier, the Profitability Indexes would be as shown in Table 6.8.

6.2.5 NPV of the Simple ROI Case

Returning to the simple ROI case used previously of a UPS replacement, we can now recalculate the ROI including the discount rate and assess whether our project actually provides an overall return and, if so, how much. In our simple addition previously, the project outcome was a saving of $60,000; for this analysis we will assume that the finance department has requested the NPV over 10 years with a 10% discount rate as shown in Table 6.9.

TABLE 6.8 Profitability index of simple investment example

Discount rate	Profitability index	Formula (PV/initial investment)
0% simple payback	2.86	=$20,000/$7,000
5%	1.78	=$12,462/$7,000
10%	1.22	=$8,514/$7,000
20%	0.70	=$4,869/$7,000

With the impact of our discount rate reducing the PV of our future savings at 10%/annum, our UPS upgrade project now evaluates as showing a small loss over the 10-year period.

The total NPV may be calculated either by summing the individual PVs for each year or by using the annual total costs or incomes to calculate the NPV. In Microsoft Office Excel, we can use the NPV worksheet function that takes the arguments: NPV (Discount Rate, Future Income 1, Future Income 2, etc.). It is important to treat each cost or income in the correct period. Our first cost occurs at the beginning of the first year, but our payments occur at the end of the year; this must be separately added to the output of the NPV function. The other note is that the NPV function takes incomes rather than payments, so the signs are reversed as compared with the PV function.

To calculate our total NPV in the cells already mentioned, we would use the formula = **B9 + NPV(0.1, C9:L9)**, which takes the initial cost and adds the PV of the savings over the 10-year period.

6.2.6 Internal Rate of Return

The IRR is closely linked to the NPV calculation. In the NPV calculation, we use a discount rate to reduce the PV of

TABLE 6.9 Calculation of the NPV of the simple ROI example

	A	B	C	D	E	F	G	H	I	J	K	L
1	Year	0	1	2	3	4	5	6	7	8	9	10
2	New UPS purchase	–$100,000										
3	New UPS installation	–$10,000										
4	Competitive trade-in rebate	$10,000										
5	UPS battery costs	$0										
6	UPS maintenance contract		$500	$500	$500	$500	$500	$500	$500	$500	$500	$500
7	UPS power costs		$7,500	$7,500	$7,500	$7,500	$7,500	$7,500	$7,500	$7,500	$7,500	$7,500
8	Additional revenue		$8,000	$8,000	$8,000	$8,000	$8,000	$8,000	$8,000	$8,000	$8,000	$8,000
9	Annual total	–$100,000	$16,000	$16,000	$16,000	$16,000	$16,000	$16,000	$16,000	$16,000	$16,000	$16,000
10	PV	–$100,000	$14,545	$13,223	$12,021	$10,928	$9,935	$9,032	$8,211	$7,464	$6,786	$6,169
11	NPV											–$1,687

costs or incomes in the future to determine the overall net value of an investment. To obtain the IRR of an investment, we simply reverse this process to find the discount rate at which the NPV of the investment is zero.

To find the IRR in Microsoft Office Excel, you can use the IRR function:

$$= \mathrm{IRR}\left(\mathrm{values}, \left[\mathrm{guess}\right]\right)$$

6.2.6.1 Simple Investment IRR Example

We will find the IRR of the simple investment example from NPV given earlier of a $7,000 investment that produced $1,000/annum operating cost savings. We tested this project to yield an NPV of $722 at a 5% discount rate over 10 years. The IRR calculation is shown in Table 6.10.

The IRR was calculated using the formula = IRR(B4:L4), which uses the values in the "annual cost" row from the initial –$7,000 to the last $1,000.

In this case, we see that the IRR is just over 7%; if we use this as the discount rate in the NPV calculation, then our NPV evaluates to zero as shown in Table 6.11.

6.2.6.2 IRR Over Time

As observed with the PV, incomes later in the project lifetime have progressively less impact on the IRR of a project; in this case, Figure 6.5 shows the IRR of the simple example given earlier up to 30 years project lifetime. The IRR value initially increases rapidly with project lifetime but can be seen to be tending toward approximately 14.3%.

TABLE 6.10 Calculation of IRR for the simple investment example

Year	0	1	2	3	4	5	6	7	8	9	10
Cost	$7,000										
Saving		$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000
Annual cost	–$7,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000
IRR											7.07%

TABLE 6.11 NPV of the simple investment example with a discount rate equal to the IRR

Year	0	1	2	3	4	5	6	7	8	9	10
Cost	$7,000										
Saving		$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000
Annual cost	−$7,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000	$1,000
PV	−$7,000	$934	$872	$815	$761	$711	$664	$620	$579	$541	$505
NPV											$0

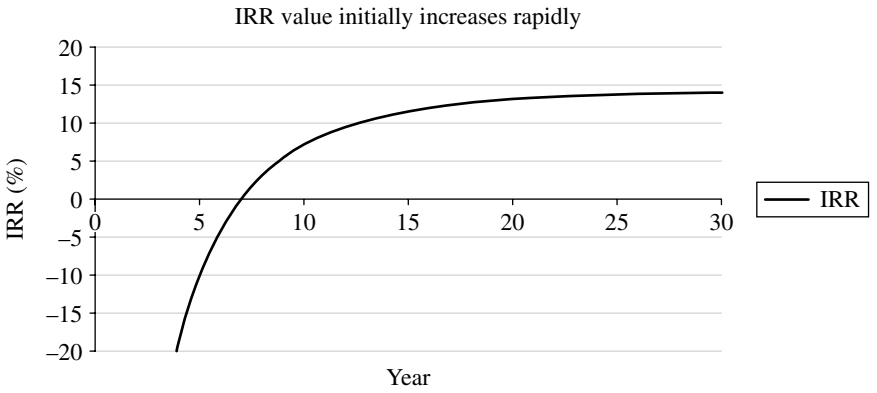


FIGURE 6.5 IRR of simple investment example.

6.2.7 Choosing NPV or IRR

In many cases, you will be required to present either an NPV or an IRR case, based on corporate policy and sometimes within a standard form, without which finance will not consider your proposal. In other cases, you may need to choose whether to use an IRR or NPV analysis to best present the investment case. In either case, it is worth understanding what the relative strengths and weaknesses of NPV and IRR analysis are to select the appropriate tool and to properly manage the weaknesses of the selected analysis method.

At a high level, the difference is that NPV provides a total money value without indication of how large the return is in comparison with the first investment, while IRR provides a rate of return with no indication of the scale. There are, of course, methods of dealing with both of these issues, but perhaps the simplest is to lay out the key numbers for investment, NPV, and IRR to allow the reader to compare the projects in their own context.

To illustrate some of the potential issues with NPV and IRR, we have four simple example projects in Table 6.12, each of which has a constant annual return over 5 years, evaluated at a discount rate of 15%.

6.2.7.1 Ranking Projects

The first issue is how to rank these projects. If we use NPV to rank the projects, then we would select project D with the highest NPV when, despite requiring twice the initial investment of project C, the return is less than 1% larger. If we rank the projects using only the profitability index or IRR, then projects A and C would appear to be the same despite C being five times larger in both investment and return than A. If we are seeking maximum total return, then C would be preferable; conversely, if there is substantial risk in the projects, we may choose to take project A rather than C.

TABLE 6.12 NPV and IRR of four simple projects

Project	Capital cost	Annual return	NPV	Profitability index	IRR
A	−$100,000	$50,000	$67,608	1.68	41%
B	−$500,000	$200,000	$170,431	1.34	29%
C	−$500,000	$250,000	$338,039	1.68	41%
D	−$1,000,000	$400,000	$340,862	1.34	29%

A further complication in data center projects is that in many cases the project options are mutually exclusive, either because there is limited total budget available or because the projects cannot both be implemented such as options to upgrade or replace the same piece of equipment. If we had $1 million to invest and these four projects to choose from, we might well choose B and C; however, if these two projects are an either-or option, then A and C would be our selection, and we would not invest $400k of our available budget.

Clearly, neither NPV nor IRR alone is suitable for ranking projects; this can be a particular issue in organizations where the finance group sets a minimum IRR for any project, and it may be appropriate to present options that are near to the minimum IRR but have larger available returns than those that exceed the target IRR.

6.2.7.2 Other Issues

IRR should not be used to compare projects of different durations; your finance department will typically have a standard number of years over which an IRR calculation is to be performed.

IRR requires both costs and savings; you can't use IRR to compare purchasing or leasing a piece of equipment.

In a project with costs at more than one time, such as modular build of capacity, there may be more than one IRR at different times in the project.

6.3 COMPLICATIONS AND COMMON PROBLEMS

All of the examples so far have been relatively simple with clear predictions of the impact of the changes to allow us to clearly assess the NPV or IRR of the project. In the real world, things are rarely this easy, and there will be many factors that are unknown, variable, or simply complicated, which will make the ROI analysis less easy. This section will discuss some of the complications as well as some common misunderstandings in data center financial analysis.

6.3.1 ROI Analysis Is About Optimization, Not Just Meeting a Target Value

When assessing the financial viability of data center projects, there will generally be a range of options in how the projects are delivered and which will affect the overall cost and overall return. The art of an effective financial analysis is to break down the components of each project and understand how each of these contributes to the overall ROI outcome. Once you have this breakdown of benefit elements, these may be weighed against the other constraints that you must work within. In any organization with more than one data center, it will also be necessary to balance the available resources across the different sites.

A good ROI analysis will find an effective overall balance considering the following:

- Available internal resource to evaluate, plan, and implement or manage projects.
- Projects that are mutually exclusive for engineering or practical reasons.
- The total available budget and how it is distributed between projects.

6.3.2 Sensitivity Analysis

As already stated, analysis of a project requires that we make a number of assumptions and estimations of future events. These assumptions may be the performance of devices once installed or upgraded, the changing cost of electricity over the next 5 years, or the increase in customer revenue due to a capacity expansion. While the estimated ROI of a project is important, it is just as vital to understand and communicate the sensitivity of this outcome to the various assumptions and estimations.

At a simple level, this may be achieved by providing the base analysis, accompanied by an identification of the impact on ROI of each variable. To do this, you can state the estimate and the minimum and maximum you would reasonably expect for each variable and then show the resulting ROI under each change.

As a simple example, a project may have an estimated ROI of $100,000 at a power cost of $0.10/kWh, but your estimate of power cost ranges from $0.08 to $0.12/kWh, which result in ROIs of $50,000 and $150,000, respectively. It is clearly important for the decision maker to understand the impact of this variability, particularly if the company has other investments that are subject to variation in energy cost.

There are, of course, more complex methods of assessing the impact of variability on a project; one of the more popular, Monte Carlo analysis, is introduced later in this chapter.

6.3.2.1 Project Benefits Are Generally not Cumulative

One very common mistake is to independently assess more than one data center project and then to assume that the results may be added together to give a total capacity release or energy savings for the combined projects if implemented together.

The issue with combining multiple projects is that the data center infrastructure is a system and not a set of individual components. In some cases, the combined savings of two projects can exceed the sum of the individual savings; for example, the implementation of airflow containment with VFD fan upgrades to the CRAC units coupled with the addition of a water side economizer. Either project would

save energy, but the airflow containment allows the chilled water system temperature to be raised, which will allow the economizer to further decrease the compressor cooling requirement.

More frequently, some or all of the savings of two projects rely on reducing the same overheads in the data center. The same overhead can't be eliminated twice, and therefore, the total savings will not be the sum of the individual projects. A simple example might be the implementation of raised supply temperature set points and adiabatic intake air cooling in a data center with direct outside air economizing AHUs. These two projects would probably be complementary, but the increase in set points seeks to reduce the same compressor cooling energy as the adiabatic cooling, and therefore, the total will almost certainly not be the sum of the parts.

6.3.3 Accounting for Taxes

In many organizations, there may be an additional potential income stream to take account of in your ROI analysis in the form of reduced tax liabilities. In most cases, when a capital asset is purchased by a company, the cost of the asset is not dealt for tax purposes as one lump at the time of purchase. Normal practice is to depreciate the asset over some time frame at a given rate; this is normally set by local tax laws. This means that, for tax purposes, some or all of the capitalized cost of the project will be spread out over a number of years; this depreciation cost may then be used to reduce tax liability in each year. This reduced tax liability may then be included in each year of the project ROI analysis and counted toward the overall NPV or IRR. Note that for the ROI analysis, you should still show the actual capital costs occurring in the accounting periods in which they occur; it is only the tax calculation that uses the depreciation logic.

The discussion of regional tax laws and accounting practices related to asset depreciation and taxation is clearly outside of the scope of this book, but you should consult the finance department in the organization for whom you are producing the analysis to determine whether and how they wish you to include tax impacts.

6.3.4 Costs Change over Time: Real and Nominal Discount Rates

As already discussed, the value of money changes over time; however, the cost of goods, energy, and services also changes over time, and this is generally indicated for an economy by an annual percentage inflation or deflation. When performing financial analysis of data center investments, it may be necessary to consider how costs or incomes may change independently of a common inflation rate.

The simpler method of NPV analysis uses the real cash flows. These are cash flows that have been adjusted to the current value or, more frequently, simply estimated at their current value. This method then applies what is called the real discount rate that includes both the nominal interest rate and a reduction to account for the inflation rate. The relationship between the real and nominal rates is shown as follows:

$$\text{Real} = \left(\frac{1+\text{nominal}}{1+\text{inflation}}\right) - 1$$

The second method of NPV analysis allows you to make appropriate estimates for the changes in both costs and revenues over time. This is important where you expect changes in goods or energy costs that are not well aligned with inflation or each other. In this case, the actual (nominal) cash flows are used, and the full nominal discount rate is applied.

As an example, consider a project with a $100,000 initial capital investment, which we expect to produce a $50,000 income in today's money across each of 3 years. For this project, the nominal discount rate is 10%, but we expect inflation over the period to be 2.5%, which gives a real discount rate of 7.3%.

We can perform an NPV analysis using real cash flows and the real discount rate as in Table 6.13.

Alternatively, we can include the effect of our expected inflation in the cash flows and then discount them at the nominal discount rate as in Table 6.14.

The important thing to note here is that both NPV calculations return the same result. Where the future costs and

TABLE 6.13 NPV of real cash flows at the real discount rate

Capital	1	2	3	NPV	Notes
£100,000	£50,000	£50,000	£50,000		Real cash flows
	£46,591	£43,414	£40,454	£30,459	Real discount rate at 7.3%

TABLE 6.14 NPV of nominal cash flows at the nominal discount rate

Capital	1	2	3	NPV	Notes
£100,000	£51,250	£52,531	£53,845		Nominal cash flows
	£46,591	£43,414	£40,454	£30,459	Nominal discount rate at 10.0%

revenues all increase at the same rate as our inflation factor, the two calculations are equivalent. Where we expect any of the future cash flows to increase or decrease at any rate other than in line with inflation, it is better to use the nominal cash flows and nominal discount rate to allow us to account for these changes. Expected changes in the future cost of energy are the most likely example in a data center NPV analysis. This latter approach is illustrated in both the Monte Carlo and main realistic example analysis later in this chapter.

6.3.5 Multiple Solutions for IRR

One of the issues in using IRR is that there is no simple formula to give an IRR; instead, you or the spreadsheet you are using must seek a value of discount rate for which the NPV evaluates to zero. When you use the IRR function in a spreadsheet such as Microsoft Office Excel, there is an option in the formula to allow you to provide a guess to assist the spreadsheet in determining the IRR you seek:

$$= \mathrm{IRR}\left(\mathrm{values}, \left[\mathrm{guess}\right]\right)$$

This is not because the spreadsheet has trouble iterating through different values of discount rate; but because there is not always a single unique solution to the IRR for a series of cash flows. If we consider the series of cash flows in Table 6.15, we can see that our cash flows change sign more than once; that is, they start with a capital investment, negative, then change between incomes, positive, and further costs, negative.

The chart in Figure 6.6 plots the NPV over the 4 years against the applied discount rate. It is evident that the NPV is zero twice due to the shape of the curve; in fact, the IRR solves to both 11 and 60% for this series of cash flows.

There are a number of methods for dealing with this issue, from supplying an appropriate guess to the spreadsheet IRR function to assisting it in converging on the value you are looking for to using alternative methods such as the Modified Internal Rate of Return (MIRR), which is provided in most spreadsheet packages but is outside the scope of this chapter.

6.3.6 Broken and Misused Rules of Thumb

In the data center industry, there are many standard practices and rules of thumb; some of these have been developed over many years of operational experience, while others have taken root on thin evidence due to a lack of available information to disprove them. It is generally best to make an individual assessment; where only a rule of thumb is available, this is unlikely to be an effective assumption in the ROI case.

Some of the most persistent of these are related to the cooling system and environmental controls in the data center. Some common examples are as follows:

- It is best to operate required capacity +1 of the installed CRAC/AHU; this stems from systems operating constant speed fans with flow dampers where energy was relatively linear with airflow and operating hours meant wear-out maintenance costs. In modern VFD controlled systems, the large savings of fan speed reduction dictate that, subject to minimum speed requirements, more units should operate in parallel and at the same speed.
- We achieve *X%* saving in cooling energy for every degree increase in supply air or water temperature. This may have been a good rule of thumb for entirely

TABLE 6.15 Example cash flow with multiple IRR solutions

Year	0	1	2	3	4
Income	–$10,000	$27,000	–$15,000	–$7,000	$4,500

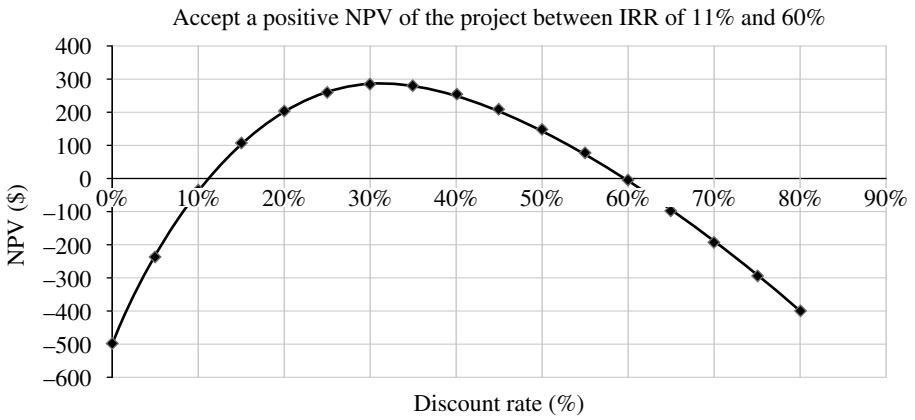


FIGURE 6.6 Varying NPV with discount rate.

compressor-cooled systems; but in any system with free cooling, the response is very nonlinear.

- The "optimum" IT equipment supply temperature is 25°C; above this IT equipment fan energy increases faster than cooling system energy. The minimum overall power point does, of course, depend upon not only the changing fan power profile of the IT equipment but also the response of the cooling system and, therefore, varies for each data center as well as between data centers.
- Applying a VFD to a fan or pump will allow the energy to reduce as the cube of flow; this is close to the truth for a system with no fixed head and the ability to turn down to any speed. But in the case of pumps that are controlled to a constant pressure such as secondary distribution water pumps, the behavior is very different.

6.3.7 Standardized Upgrade Programs

In many end user and consulting organizations, there is a strong tendency to implement data center projects based on a single strategy that is believed to be tested and proven. This approach is generally flawed for two major reasons.

First, each data center has a set of opportunities and constraints defined by its physical building, design, and history. You should not expect a data center with split direct expansion (DX) CRAC units to respond in the same way to an airflow management upgrade as a data center with central AHUs and overhead distribution ducts.

Second, where the data centers are distributed across different climates or power tariffs, the same investment that delivered excellent ROI in Manhattan may well be a waste of money in St. Louis even when applied to a building identical in cooling design and operation.

There may well be standard elements, commonly those recognized as best practice by programs such as the EU Code of Conduct, which should be on a list of standard options to be applied to your estate of data centers. These standard elements should then be evaluated on a per-opportunity basis in the context of each site to determine the selection of which projects to apply based on a tailored ROI analysis rather than habit.

6.3.7.1 Climate Data

Climate data is available in a range of formats, each of which is more or less useful for specific types of analysis. There are a range of sources for climate data, many of which are regional and have more detailed data for their region of operation.

While the majority of the climate data available to you will be taken from quite detailed observations of the actual climate over a substantial time period, this is generally processed before publication, and the data you receive will be some sort of summary. The common formats you are likely to come across are as follows.

6.3.7.2 Design Conditions

The design conditions for a site are generally given as the minimum and maximum temperature expected over a specified number of years. These values are useful only for ensuring the design is able to operate at the climate extremes it will encounter.

6.3.7.3 Heating/Cooling Hours

It is common to find heating and cooling hours in the same data sets as design conditions; these are of no realistic use for data center analysis.

6.3.7.4 Temperature Binned Hours

It is common to see analysis of traditional cooling components such as chillers carried out using data that sorts the hours of the year into temperature "bins," for example, "2316 annual hours between 10 and 15°C Dry Bulb." The size of the temperature bin varies with the data source. A major issue with this type of data is that the correlation between temperature and humidity is destroyed in the binning process. This data may be useful if no less processed data is available, but only where the data center cooling load does not vary with the time of day, humidity control is not considered (i.e. no direct air economizer systems), and the utility energy tariff does not have off-peak/peak periods or peak demand charges.

6.3.7.5 Hourly Average Conditions

Another common processed form of data is the hourly average; in this format, there are 24 hourly records for each month of the year, each of which contains an average value for dry bulb temperature, humidity, and frequently other aspects such as solar radiation or wind speed and direction. This format can be more useful than binned hours where the energy tariff has peak/off-peak hours but is of limited use for humidity sensitive designs and may give false indications of performance for economized cooling systems with sharp transitions.

6.3.7.6 Typical Meteorological Year

The preferred data type for cooling system analysis is Typical Meteorological Year (TMY). This data contains a set of values for each hour of the year, generally including dry bulb temperature, dew point, humidity, atmospheric pressure, solar radiation, precipitation, wind speed, and direction. This data is generally drawn from recorded observations but is carefully processed to represent a "typical" year.

6.3.7.7 Recorded Data

You may have actual recorded data from a Building Management System for the site you are analyzing or another nearby site in the same climate region. This data can be useful for historical analysis, but in most cases, correctly processed TMY data is preferred for predictive analysis.

6.3.7.8 Sources of Climate Data

Some good sources of climate data are the following:

- ASHRAE[10] and equivalent organizations outside the United States such as ISHRAE.[11]
- The US National Renewable Energy Laboratory of the Department of Energy (DOE) publish an excellent set of TMY climate data for use in energy simulations and converter tools between common file formats on the DOE Website.
- Weather Underground[12] where many contributors upload data recorded from weather stations that is then made freely available.

6.3.8 Location Sensitivity

It is easy to see how even the same data center design may have a different cooling overhead in Finland than in Arizona and also how utility electricity may be cheaper in North Carolina than in Manhattan or Singapore. As an example, we may consider a relatively common 1 MW water-cooled data center design. The data center uses water-cooled chillers and cooling towers to supply chilled water to the CRAC units in the IT and plant areas. The data center has plate heat exchangers between the condenser water and chilled water circuits to provide free cooling when the external climate allows.

For the first part of the analysis, the data center was modeled[13] in four configurations, representing four different chilled water supply (CHWS) temperatures; all of the major variables in the cooling system are captured. The purpose of the evaluation is to determine the available savings from the cooling plant if the chilled water temperature is increased. Once these savings are known, it can be determined whether the associated work in airflow management or increase in IT equipment air supply temperature is worthwhile.

The analysis will be broken into two parts, first the PUE response to the local climate and then the impact of the local power tariff.

[10] American Society of Heating Refrigeration and Air Conditioning Engineers.
[11] Indian Society of Heating Refrigerating and Air Conditioning Engineers.
[12] www.weatherundergound.com.
[13] Using Romonet Software Suite to perform analysis of the entire data center mechanical and electrical infrastructure with full typical meteorological year climate data.

6.3.8.1 Climate Sensitivity

The first part of the analysis is to determine the impact on the annual PUE for the four set points:

- 7°C (45°F) CHWS with cooling towers set to 5°C (41°F) in free cooling mode.
- 11°C (52°F) CHWS with cooling towers set to 9°C (48°F) in free cooling mode and chiller Coefficient of Performance (CoP) increased based on higher evaporator temperature.
- 15°C (59°F) CHWS with cooling towers set to 13°C (55°F) in free cooling mode and chiller CoP increased based on higher evaporator temperature.
- 19°C (66°F) CHWS with cooling towers set to 17°C (63°F) in free cooling mode, chiller CoP as per the 15°C (59°F) variant and summer mode cooling tower return set point increased by 5°C (9°F).

The output of the analysis is shown in Figure 6.7 for four different TMY climates selected to show how the response of even this simple change depends on the location and does not follow a rule of thumb for savings. The PUE improvement for Singapore is less than 0.1 as the economizer is never active in this climate and the only benefit is improved mechanical chiller efficiency. St. Louis, Missouri, shows a slightly stronger response, but still only 0.15, as the climate is strongly modal between summer and winter with few hours in the analyzed economizer transition region. Sao Paulo shows a stronger response above 15°C, where the site transitions from mostly mechanical cooling to mostly partial or full economizer. The largest saving is shown in San Jose, California, with a 0.24 reduction in PUE, which is substantially larger than the 0.1 for Singapore.

6.3.8.2 Energy Cost

Both the cost and the charge structure for energy vary greatly across the world. It is common to think of electricity as having a unit kWh cost, but when purchased at data center scale, the costs are frequently more complex; this is particularly true in the US market, where byzantine tariffs with multiple consumption bands and demand charges are common.

To demonstrate the impact of these variations in both energy cost and type of tariff, the earlier analysis for climate sensitivity also includes power tariff data every hour for the climate year:

- Singapore has a relatively high cost of power with peak/off-peak bands and a contracted capacity charge that is unaffected by the economizer implementation as no reduction in peak draw is achieved.
- Sao Paulo also has a relatively high cost of power but in this instance on a negotiated flat kWh tariff.

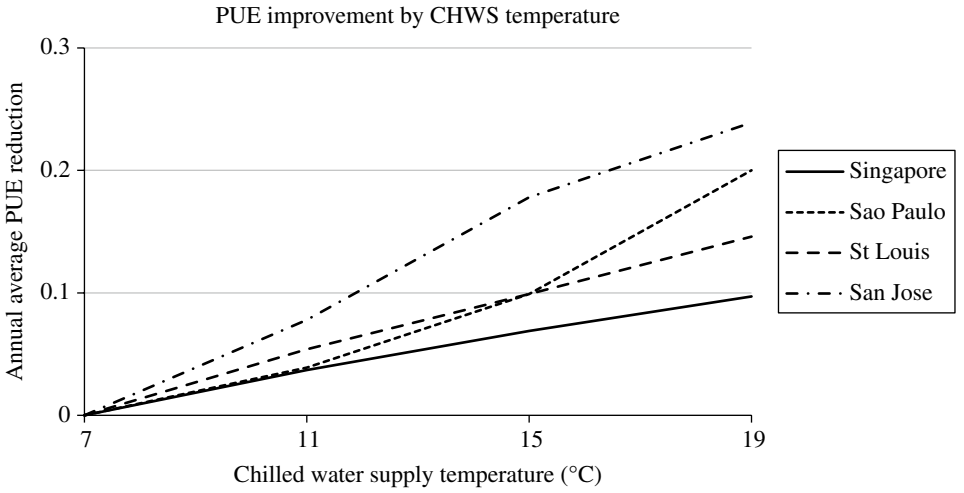


FIGURE 6.7 Climate sensitivity analysis—PUE variation with chilled water supply temperature.

- St. Louis, Missouri, has a very low kWh charge as it is in the "coal belt" with an additional small capacity charge.
- San Jose, California, has a unit kWh charge twice that of St. Louis.

Note that the free cooling energy savings will tend to be larger during off-peak tariff hours and so, to be accurate, the evaluation must evaluate power cost for each hour and not as an average over the period.

The impact of these charge structures is shown in the graph in Figure 6.8. Singapore, despite having only two-third of the PUE improvement of St. Louis, achieves more than twice the energy cost saving due to the high cost of power, particularly in peak demand periods. Sao Paolo and San Jose both show large savings but are again in inverse order of their PUE savings.

The cost outcomes shown here show us that we should consider the chilled water system upgrade very differently in St. Louis than in San Jose or Sao Paulo.

As with any part of our ROI analysis, these regional energy cost and tariff structure differences are based on the current situation and may well change over time.

No Chiller Data Centers

In recent years, the concept of a data center with no compressor-based cooling at all has been popularized with a number of operators building such facilities and claiming financial or environmental benefits due to this elimination of chillers.

While there are some benefits to eliminate the chillers from data centers, the financial benefit is primarily first capital cost, as neither energy efficiency nor energy cost is improved significantly. Depending on the climate the data center operates in, these benefits may come at the cost of the

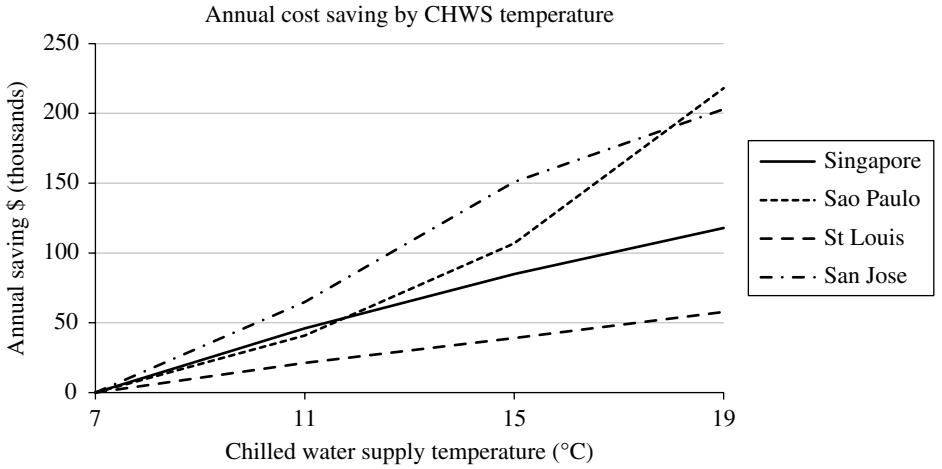


FIGURE 6.8 Energy cost sensitivity analysis—annual cost saving by chilled water supply (CHWS) temperature.

requirement of a substantial expansion of the working environmental range of the IT equipment.

As discussed in the section on free cooling that follows, the additional operational energy efficiency and energy cost benefits of reducing chiller use from a few months per year to never are minimal. There may be substantial first capital cost benefits, however, not only in the purchase and installation cost of the cooling plant but also in the elimination of upstream electrical equipment capacity otherwise required to meet compressor load. Additional operational cost benefits may be accrued through the reduction of peak demand or power availability charges as these peaks will no longer include compressor power.

The balancing factor against the cost benefits of no-chiller designs is the expansion in environmental conditions the IT equipment must operate in. This may be in the form of increased temperature, humidity range, or both. Commonly direct outside air systems will use adiabatic humidifiers to maintain temperature at the expense of high humidity. Other economizer designs are more likely to subject the IT equipment to high temperature peaks during extreme external conditions. The additional concern with no-chiller direct outside air systems is that they cannot revert to air recirculation in the event of an external air pollution event such as dust, smoke, or pollen, which may necessitate an unplanned shutdown of the data center.

Free Cooling, Economizer Hours, and Energy Cost

Where a free cooling system is in use, it is quite common to see the performance of the free cooling expressed in terms of "economizer hours," usually meaning the number of hours during which the system requires mechanical compressor cooling. While the type of economizer may vary, from direct external air to plate heat exchangers for the chilled water loop, the objective of cooling economizers is to reduce the energy consumed to reject the heat from the IT equipment.

As the cooling system design and set points are improved, it is usual to expect some energy saving. As described earlier in the section on climate sensitivity, the level of energy saving is not linear with the changes in air or water set point temperature; this is not only due to the number of hours in each temperature band in the climate profile but also due to the behavior of the free cooling system.

Figure 6.9 shows a simplified overview of the relationship between mechanical cooling energy, economizer hours, and chiller elimination.

At the far left (A) is a system that relies entirely on mechanical cooling with zero economizer hours—the mechanical cooling energy is highest at this point. Moving to the right (B), the cooling set points are increased, and this allows for some of the cooling to be performed by the economizer system. Initially, the economizer is only able to reduce the mechanical cooling load, and the mechanical cooling must still run for the full year. As the set points increase further (C), the number of hours per year that the mechanical cooling is required for reduces, and the system moves to primarily economized cooling. When the system reaches zero hours of mechanical cooling (D) in a typical year, it may still require mechanical cooling to deal with peak hot or humid conditions,[14] even though these do not regularly occur. Beyond this point (E), it is common to install mechanical cooling of reduced capacity to supplement the free cooling

[14] Commonly referred to as the design conditions.

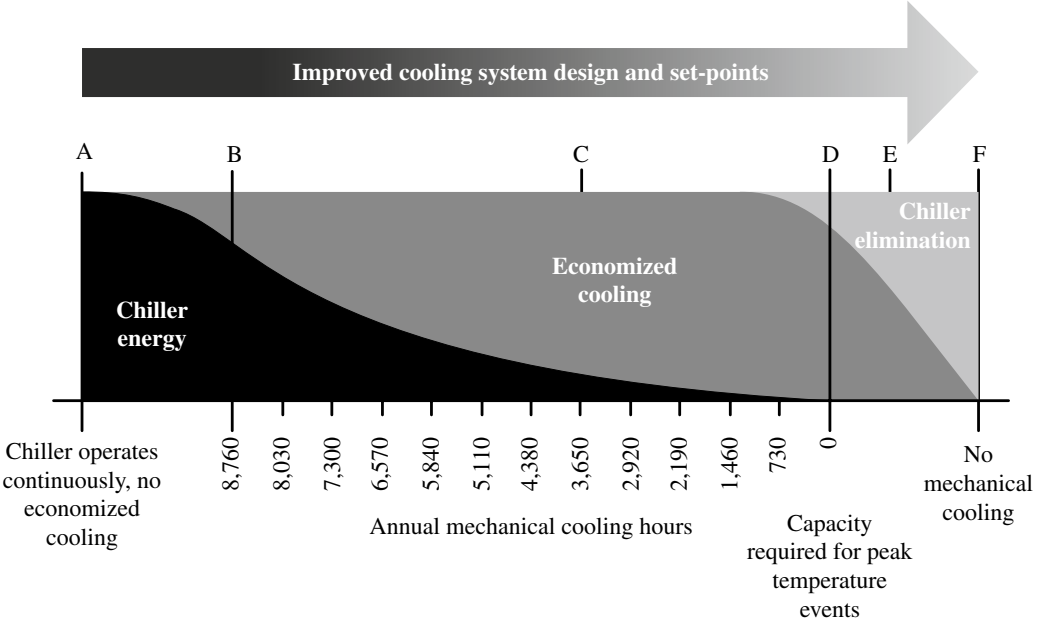


FIGURE 6.9 Chiller energy by economizer hours.

system. At the far right (F) is a system that is able to meet all of the heat rejection needs even at peak conditions without installing any mechanical cooling at all.

The area marked "chiller energy" in the chart indicates (approximately, dependent on the system design and detailed climate profile) the amount of energy consumed in mechanical cooling over the year. This initially falls sharply and then tails off, as the mechanical cooling energy is a function of several variables. As the economized cooling capacity increases,

- The mechanical cooling is run for fewer hours, thus directly using less energy;
- The mechanical cooling operates at part load for many of the hours it is run, as the free cooling system takes part of the load, thus using less energy;
- The mechanical cooling system is likely to work across a smaller temperature differential, thus allowing a reduction in compressor energy, either directly or through the selection of a unit designed to work at a lower temperature differential.

These three factors combine to present a sharp reduction in energy and cost initially as the economizer hours start to increase; this allows for quite substantial cost savings even where only one or two thousand economizer hours are achieved and substantial additional savings for small increases in set points. As the economized cooling takes over, by point (C), there is very little mechanical cooling energy consumption left to be saved, and the operational cost benefits of further increases in set point are minimal. Once the system is close to zero mechanical cooling hours (D), additional benefit in capital cost may be obtained by reducing or completely eliminating the mechanical cooling capacity installed.

Why the Vendor Case Study Probably Doesn't Apply to You
It is normal for vendor case studies to compare the best reasonably credible outcome for their product, service, or technology with a "base case" that is carefully chosen to present the value of their offering in the most positive light possible. In many cases, it is easy to establish that the claimed savings are in fact larger than the energy losses of those parts of your data center that are to be improved and, therefore, quite impossible for you to achieve.

Your data center will have a different climate, energy tariff, existing set of constraints, and opportunities to the site selected for the case study. You can probably also achieve some proportion of the savings with lower investment and disruption; to do so, break down the elements of the savings promised and how else they may be achieved to determine how much of the claimed benefit is actually down to the product or service being sold.

The major elements to consider when determining how representative a case study may be of your situation are as follows:

- Do the climate or IT environmental conditions impact the case study? If so, are these stated and how close to your data center are the values?
- Are there physical constraints of the building or regulatory constraints such as noise that would restrict the applicability?
- What energy tariff was used in the analysis? Does this usefully represent your tariff including peak/off-peak, seasonal, peak demand, and availability charge elements?
- How much better than the "before" condition of the case study is your data center already?
- What other cheaper, faster, or simpler measures could you take in your existing environment to produce some or all of the savings in the case study?
- Was there any discount rate included in the financial analysis of the case study? If not, are the full implementation cost and savings shown for you to estimate an NPV or IRR using your internal procedures?

The process shown in the Section 6.4 is a good example of examining how much of the available savings are due to the proposed project and how much may be achieved for less disruption or cost.

6.3.9 IT Power Savings and Multiplying by PUE

If the project you are assessing contains an element of IT power draw reduction, it is common to include the energy cost savings of this in the project analysis. Assuming that your data center is not perfectly efficient and has a PUE greater than 1.0, you may expect some infrastructure overhead energy savings in addition to the direct IT energy savings.

It is common to see justifications for programs such as IT virtualization or server refresh using the predicted IT energy saving and multiplying these by the PUE to estimate the total energy savings. This is fundamentally misconceived; it is well recognized that PUE varies with IT load and will generally increase as the IT load decreases. This is particularly severe in older data centers where the infrastructure overhead is largely fixed and, therefore, responds very little to IT load.

IT power draw multiplied by PUE is not suitable for estimating savings or for charge-back of data center cost. Unless you are able to effectively predict the response of the data center to the expected change in IT load, the predicted change in utility load should be no greater than the IT load reduction.

6.3.10 Converting Other Factors into Cost

When building an ROI case, one of the more difficult elements to deal with is probability and risk. While there is a risk element in creating any forecast into the future, there are some revenues or costs that are more obviously at risk and should be handled more carefully. For example, an upgrade reinvestment business case may improve reliability at the same time as reducing operational costs requiring us to put a value on the reliability improvement. Alternatively, for a service provider, an investment to create additional capacity may rely on additional customer revenue for business justification; there can be no guarantee of the amount or timing of this additional revenue, so some estimate must be used.

6.3.10.1 Attempt to Quantify Costs and Risks

For each of the external factors that could affect the outcome of your analysis, make a reasonable attempt to quantify the variables so that you may include them in your assessment. In reality, there are many bad things that may happen to a data center that could cost a lot of money, but it is not always worth investing money to reduce those risks. There are some relatively obvious examples; the cost of adding armor to withstand explosives is unlikely to be an effective investment for a civilian data center but may be considered worthwhile for a military facility.

The evaluation of risk cost can be quite complex and is outside the scope of this chapter. For example, where the cost of an event may vary dependent on the severity of the event, modeling the resultant cost of the risk requires some statistical analysis.

At a simplistic level, if a reasonable cost estimate can be assigned to an event, the simplest way to include the risk in your ROI analysis is to multiply the estimated cost of the event by the probability of it occurring. For example, your project may replace end-of-life equipment with the goal of reducing the risk of a power outage from 5 to 0.1%/year. If the expected cost of the power outage is $500,000 in service credit and lost revenue, then the risk cost would be:

- Without the project, $0.05 \times \$500{,}000 = \$25{,}000$/annum
- With the project, $0.001 \times \$500{,}000 = \$500$/annum

Thus, you could include $24,500/annum cost saving in your project ROI analysis for this mitigated risk. Again, this is a very simplistic analysis, and many organizations will use more effective tools for risk quantification and management, from which you may be able to obtain more effective values.

6.3.10.2 Create a Parameterized Model

Where your investment is subject to external variations such as the cost of power over the evaluation time frame, it may be necessary to evaluate how your proposed project performs under a range of values for each external factor. In these cases, it is common to construct a model of the investment in a spreadsheet that responds to the variable external factors and so allows you to evaluate the range of outcomes and sensitivity of the project to changes in these input values.

The complexity of the model may vary from a control cell in a spreadsheet to allow you to test the ROI outcome at $0.08, $0.10, and $0.12/kWh power cost through to a complex model with many external variables and driven by a Monte Carlo analysis[15] package.

6.3.10.3 A Project that Increases Revenue Example

It is not uncommon to carry out a data center project to increase (or release) capacity. The outcome of this is that there is more data center power and cooling capacity to be sold to customers or cross-charged to internal users. It is common in capacity upgrade projects to actually increase the operational costs of the data center by investing capital to allow more power to be drawn and the operational cost to increase. In this case, the NPV or IRR will be negative unless we consider the additional business value or revenue available.

As an example of this approach, a simple example model will be shown that evaluates the ROI of a capacity release project. This project includes both the possible variance in how long it takes to utilize the additional capacity and the power cost over the project evaluation time frame.

For this project we have the following:

- $100,000 capital cost in year 0.
- 75 kW increase in usable IT capacity.
- Discount rate of 5%.
- Customer power multiplier of 2.0 (customer pays metered kWh × power cost × 2.0).
- Customer kW capacity charge of $500/annum.
- Customer power utilization approximately 70% of contracted.
- Estimated PUE of 1.5 (but we expect PUE to fall from this value with increasing load).
- Starting power cost of $0.12/kWh.

From these parameters, we can calculate in any year of the project the additional cost and additional revenue for each extra 1 kW of the released capacity we sell to customers.

We construct our simple spreadsheet model such that we can vary the number of years it takes to sell the additional capacity and the annual change in power cost.

[15] A numerical analysis method developed in the 1940s during the Manhattan Project that is useful for modeling phenomena with significant uncertainty in inputs that may be modeled as random variables.

We calculate the NPV as before, at the beginning of our project, year zero, we have the capital cost of the upgrade, $100,000. Then, in each year, we determine the average additional customer kW contracted and drawn based on the number of years it takes to sell the full capacity. In Table 6.16 is a worked example where it takes 4 years to sell the additional capacity.

The spreadsheet uses a mean and variance parameter to estimate the increase in power cost each year; in this case, the average increase is 3% with a standard deviation of ±1.5%.

From the values derived for power cost contracted and drawn kW, we are able to determine the annual additional revenue and additional cost. Subtracting the cost from the revenue and applying the formula for PV, we can obtain the PV for each year. Summing these provides the total PV across the lifetime—in this case, $119,933, as shown in Table 6.16.

We can use this model in a spreadsheet for a simple Monte Carlo analysis by using some simple statistical functions to generate for each trial:

- The annual power cost increase based on the specified mean and standard deviation of the increase (In this example, I used the NORM.INV[RAND(), mean, standard deviation] function in Microsoft Office Excel to provide the annual increase assuming a normal distribution).
- The number of years before the additional capacity is fully sold (In this example the NORM.INV[RAND(), expected fill out years, standard deviation] function is used, again assuming a normal distribution).

By setting up a reasonably large number of these trials in a spreadsheet, it is possible to evaluate the likely range of financial outcomes and the sensitivity to changes in the external parameters. The outcome of this for 500 trials is shown in Figure 6.10; the dots are the individual trials plotted as years to fill capacity versus achieved NPV; the horizontal lines show the average project NPV across all trials and the boundaries of ±1 standard deviation.

TABLE 6.16 Calculation of the NPV for a single trial

Parameter	0	1	2	3	4	5	6
Annual power cost		$0.120	$0.124	$0.126	$0.131	$0.132	$0.139
Additional kW sold		9	28	47	66	75	75
Additional kW draw		7	20	33	46	53	53
Additional revenue	$0	$18,485	$56,992	$96,115	$138,192	$159,155	$165,548
Additional cost	$100,000	$10,348	$32,197	$54,508	$79,035	$91,241	$96,036
Annual present value	–$100,000	$7,749	$22,490	$35,942	$48,669	$53,212	$51,871
Total present value	–$100,000	–$92,251	–$69,761	–$33,819	$14,850	$68,062	$119,933

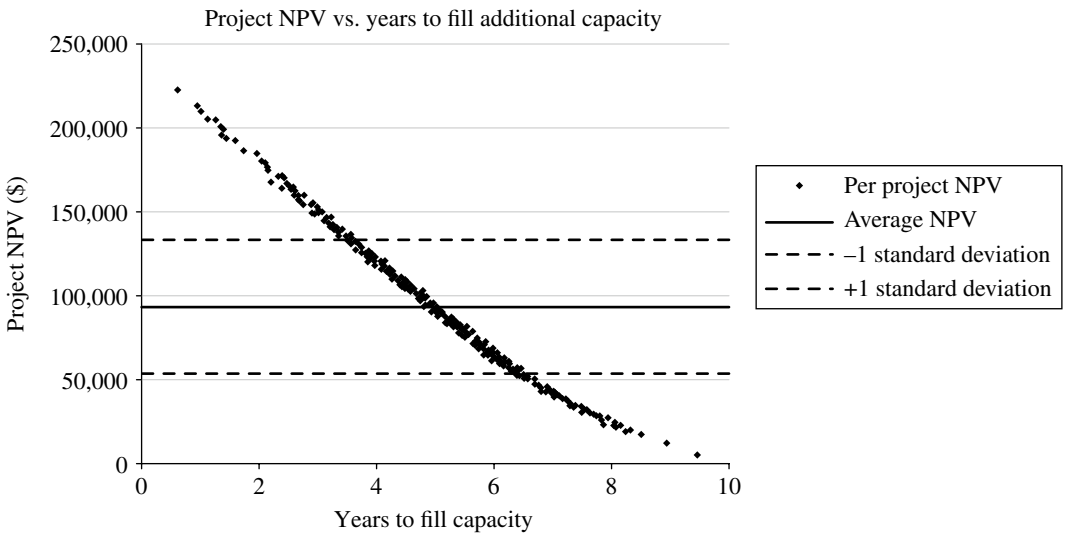


FIGURE 6.10 Simple Monte Carlo analysis of capacity upgrade project.

There are a number of things apparent from the chart:

- Even in the unlikely case of it taking 10 years to sell all of the additional capacity, the overall outcome is still likely to be a small positive return.
- The average NPV is just under $100,000, which against an investment of $100,000 for the capacity release is a reasonable return over the 6-year project assessment time frame.

An alternative way to present the output of the analysis is to perform more trials and then count the achieved NPV of each trial into a bin to determine the estimated probability of an NPV in each range. To illustrate this, 5,000 trials of the earlier example are binned into NPV bands of $25,000 and plotted in Figure 6.11.

6.3.10.4 Your Own Analysis

The earlier example is a single simplistic example of how you might assess the ROI of a project that is subject to one or more external factors. There are likely to be other plots and analyses of the output data that provide insight for your situation; those shown are merely examples. Most spreadsheet packages are capable of Monte Carlo analysis, and there are many worked examples available in the application help and online. If you come to use this sort of analysis regularly, then it may be worth investing in one of the commercial software packages[16] that provide additional tools and capability in this sort of analysis.

6.4 A REALISTIC EXAMPLE

To bring together some of the elements presented in this chapter, an example ROI analysis will be performed for a common reinvestment project. The suggested project is to implement cooling improvements in an existing data center. The example data center:

- Has a 1 MW design total IT load,
- Uses chilled water CRAC units supplied by a water-cooled chiller with cooling towers,
- Has a plate heat exchanger for free cooling when external conditions permit with a CHWS temperature of 9°C/48°F,
- Is located in Atlanta, Georgia, USA.

The ROI analysis is to be carried out over 6 years using a discount rate of 8% at the request of the finance group.

6.4.1 Airflow Upgrade Project

There are two proposals provided for the site:

- In-row cooling upgrade with full Hot Aisle Containment (HAC).
- Airflow management and sensor network improvements and upgrade of the existing CRAC units with electronically commutated (EC) variable speed fans combined with a distributed temperature sensor network that optimizes CRAC behavior based on measured temperatures

6.4.2 Break Down the Options

While one choice is to simply compare the two options presented with the existing state of the data center, this is unlikely to locate the most effective investment option for our site. In order to choose the best option, we need to break down which changes are responsible for the project savings and in what proportion.

[16] Such as Palisade @Risk or Oracle Crystal Ball.

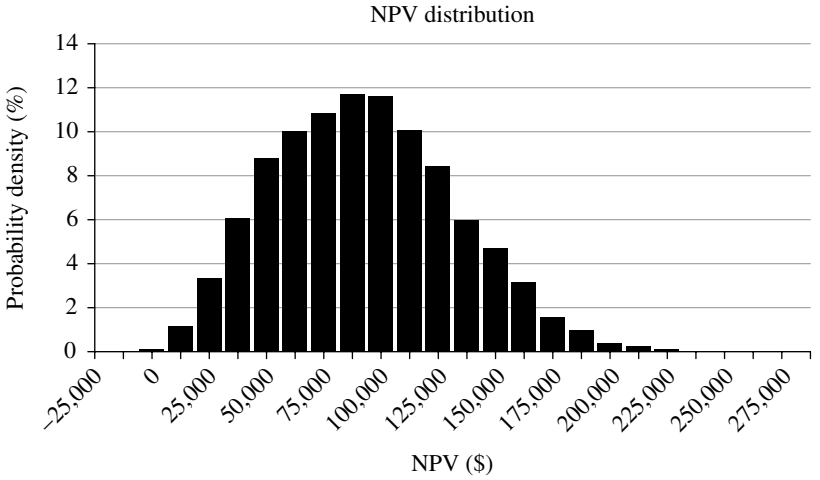


FIGURE 6.11 Probability density plot of simple Monte Carlo analysis.

In this example, the proposed cost savings are due to improved energy efficiency in the cooling system. In both options, the energy savings come from the following:

- A reduction in CRAC fan motor power through the use of variable speed drives enabled by reducing or eliminating the mixing of hot return air from the IT equipment with cold supply air from the CRAC unit. This airflow management improvement reduces the volume required to maintain the required environmental conditions at the IT equipment intake.
- A reduction in chilled water system energy consumption through an increase in supply water temperature also enabled by reducing or eliminating the mixing of hot and cold air. This allows for a small increase in compressor efficiency but more significantly an increase in the free cooling available to the system.

To evaluate our project ROI, the following upgrade options will be considered.

6.4.2.1 Existing State

We will assume that the site does not have existing issues that are not related to the upgrade such as humidity over control or conflicting set points. If there are any such issues, they should be remediated independently and not confused with the project savings as this would present a false and misleading impression of the project ROI.

6.4.2.2 Proposed Option One: In-Row Cooling

The in-row cooling upgrade eliminates 13 of the 15 current perimeter CRAC units and replaces the majority of the data hall cooling with 48 in-row cooling units. The in-row CRAC units use EC variable speed fans operated on differential pressure to reduce CRAC fan power consumption. The HAC allows for an increase in supply air and, therefore, chilled water loop temperature to 15°C/59°F. The increased CHWS temperature allows for an increase in achieved free cooling hours as well as a small improvement in operating chiller efficiency. The remaining two perimeter CRAC units are upgraded with a VFD and set to 80% minimum airflow.

6.4.2.3 Proposed Option Two: Airflow Management and Sensor Network

The more complex proposal is to implement a basic airflow management program that stops short of airflow containment and is an upgrade of the existing fixed speed fans in the CRAC units to EC variable speed fans. This is coupled with a distributed sensor network that monitors the supply temperature to the IT equipment. There is no direct saving from the sensor network, but it offers the ability to reduce CRAC fan power and increase in the CHWS temperature to allow for more free cooling hours. This option is also evaluated at 15°C/59°F CHWS temperature.

6.4.2.4 Airflow Management and VFD Upgrade

Given that much of the saving is from reduced CRAC fan power, we should also evaluate a lower capital cost and complexity option. In this case, the same basic airflow management retrofit as in the sensor network option will be deployed but without the sensor network; a less aggressive improvement in fan speed and chilled water temperature will be achieved. In this case, a less expensive VFD upgrade to the existing CRAC fans will be implemented with a minimum airflow of 80% and fan speed controlled on return air temperature. The site has $N+20\%$ CRAC units, so the 80% airflow will be sufficient even without major reductions in hot/cold remix. The chilled water loop temperature will only be increased to 12°C/54°F.

6.4.2.5 EC Fan Upgrade with Cold Aisle Containment

As the in-row upgrade requires the rack layout to be adjusted to allow for HAC, it is worth evaluating a similar option. As the existing CRAC units feed supply air under the raised floor, in this case, Cold Aisle Containment (CAC) will be evaluated with the same EC fan upgrade to the existing CRAC units as in the sensor network option. But in this case controlled on differential pressure to meet IT air demand. The contained airflow allows for the same increase in CHWS temperature to 15°C (59°F).

6.4.3 Capital Costs

The first step in evaluation is to determine the capitalized costs of the implementation options. This will include capital purchases, installation costs, and other costs directly related to the upgrade project. The costs provided in this analysis are, of course, only examples, and as for any case study, the outcome may or may not apply to your data center:

- The airflow management and HAC/CAC include costs for both airflow management equipment and installation labor.
- The In-Row CRAC unit costs are estimated to cost 48 units × $10,000 each.
- The In-Row system also requires four coolant distribution units and pipework at a total of $80,000.
- The 15 CRAC units require $7,000 upgrades of fans and motors for the two EC fan options.
- The distributed temperature sensor network equipment, installation, and software license are $100,000.

TABLE 6.17 Capitalized costs of project options

	Existing state	Airflow management and VFD fan	In-row cooling	EC fan upgrade and CAC	AFM, EC fan, and sensor network
Airflow management		$100,000			$100,000
HAC/CAC			$250,000	$250,000	
In-row CRAC			$480,000		
CDU and pipework			$80,000		
EC fan upgrade				$105,000	$105,000
VFD fan upgrade		$60,000	$8,000		
Sensor network					$100,000
CFD analysis		$20,000	$20,000	$20,000	$20,000
Total capital	$0	$180,000	$838,000	$375,000	$325,000

- Each of the options requires a $20,000 Computational Fluid Dynamic (CFD) analysis; prior to implementation, this cost is also capitalized.

The total capitalized costs of the options are shown in Table 6.17.

6.4.4 Operational Costs

The other part of the ROI assessment is the operational cost impact of each option. The costs of all options are affected by both the local climate and the power cost. The local climate is represented by a TMY climate data set in this analysis.

The energy tariff for the site varies peak and off-peak as well as summer to winter, averaging $0.078 in the first year. This is then subject to a 3% annual growth rate to represent an expected increase in European energy costs.

6.4.4.1 Efficiency Improvements

Analysis[17] of the data center under the existing state and upgrade conditions yields the achieved annual PUE results shown in Table 6.18.

These efficiency improvements do not translate directly to energy cost savings as there is an interaction between the peak/off-peak, summer/winter variability in the energy tariff, and the external temperature, which means that more free cooling hours occur at lower energy tariff rates. The annual total energy costs of each option are shown in Table 6.19.

[17] The analysis was performed using Romonet Software Suite simulating the complete mechanical and electrical infrastructure of the data center using full typical meteorological year climate data.

TABLE 6.18 Analyzed annual PUE of the upgrade options

Option	PUE
Existing state	1.92
Airflow management and VFD fan	1.72
In-row cooling	1.65
EC fan upgrade and CAC	1.63
AFM, EC fan, and sensor network	1.64

6.4.4.2 Other Operational Costs

As an example of other cost changes due to a project, the cost of quarterly CFD airflow analysis has been included in the operational costs. The use of CFD analysis to adjust airflow may continue under the non-contained airflow options, but CFD becomes unnecessary once either HAC or CAC is implemented, and this cost becomes a saving of the contained airflow options. The 6-year operational costs are shown in Table 6.19.

6.4.5 NPV Analysis

To determine the NPV of each option, we first need to determine the PV of the future operational costs at the specified discount rate of 8%. This is shown in Table 6.20.

The capitalized costs do not need adjusting as they occur at the beginning of the project. Adding together the capitalized costs and the total of the operational PVs provides a total PV for each option. The NPV of each upgrade option is the difference between the total PV for the existing state and the total PV for that option as shown in Table 6.21.

TABLE 6.19 Annual operational costs of project options

	Existing state	Airflow management and VFD fan	In-row cooling	EC fan upgrade and CAC	AFM, EC fan, and sensor network
Annual CFD analysis	$40,000	$40,000	$0	$0	$40,000
Year 1 energy	$1,065,158	$957,020	$915,394	$906,647	$912,898
Year 2 energy	$1,094,501	$983,437	$940,682	$931,691	$938,117
Year 3 energy	$1,127,336	$1,012,940	$968,903	$959,642	$966,260
Year 4 energy	$1,161,157	$1,043,328	$997,970	$988,432	$995,248
Year 5 energy	$1,198,845	$1,077,134	$1,030,284	$1,020,439	$1,027,474
Year 6 energy	$1,231,871	$1,106,866	$1,058,746	$1,048,627	$1,055,858

TABLE 6.20 NPV analysis of project options at 8% discount rate

	Existing state	Airflow management and VFD fan	In-row cooling	EC fan upgrade and CAC	AFM, EC fan, and sensor network
6-year CFD analysis PV	$184,915	$184,915	$0	$0	$184,915
Year 1 energy PV	$986,258	$886,129	$847,587	$839,488	$845,276
Year 2 energy PV	$938,359	$843,138	$806,483	$798,775	$804,284
Year 3 energy PV	$894,916	$804,104	$769,146	$761,795	$767,048
Year 4 energy PV	$853,485	$766,877	$733,537	$726,527	$731,537
Year 5 energy PV	$815,914	$733,079	$701,194	$694,493	$699,282
Year 6 energy PV	$776,288	$697,514	$667,190	$660,813	$665,370

TABLE 6.21 NPV of upgrade options

	Existing state	Airflow management and VFD fan	In-row cooling	EC fan upgrade and CAC	AFM, EC fan, and sensor network
Capital	$0	$180,000	$838,000	$375,000	$325,000
PV Opex	$5,450,134	$4,915,757	$4,525,136	$4,481,891	$4,697,712
Total PV	$5,450,134	$5,095,757	$5,363,136	$4,856,891	$5,022,712
NPV	$0	$354,377	$86,997	$593,243	$427,422

6.4.6 IRR Analysis

The IRR analysis is performed with the same capitalized and operational costs but without the application of the discount rate. To set out the costs so that they are easy to supply to the IRR function in a spreadsheet package, we will subtract the annual operational costs of each upgrade option from the baseline costs to give the annual saving as shown in Table 6.22.

From this list of the first capital cost shown as a negative number and the annual incomes (savings) shown as positive numbers, we can use the IRR function in the spreadsheet to determine the IRR for each upgrade option.

6.4.7 Return Analysis

We now have the expected change in PUE, the NPV, and the IRR for each of the upgrade options. The NPV and IRR of the existing state are zero, as this is the baseline against which the other options are measured. The analysis summary is shown in Table 6.23.

It is perhaps counterintuitive that there is little connection between the PUE improvement and the ROI for the upgrade options.

The airflow management and VFD fan upgrade option has the highest IRR and the highest ratio of NPV to invested capital. The additional $145,000 capital investment for the

TABLE 6.22 IRR analysis of project options

Option	Existing state	Airflow management and VFD fan	In-row cooling	EC fan upgrade and CAC	AFM, EC fan, and sensor network
Capital cost	$0	–$180,000	–$838,000	–$375,000	–$325,000
Year 1 savings	$0	$108,139	$189,765	$198,512	$152,261
Year 2 savings	$0	$111,065	$193,820	$202,810	$156,385
Year 3 savings	$0	$114,397	$198,434	$207,694	$161,076
Year 4 savings	$0	$117,829	$203,187	$212,725	$165,909
Year 5 savings	$0	$121,711	$208,561	$218,406	$171,371
Year 6 savings	$0	$125,005	$213,125	$223,244	$176,013

TABLE 6.23 Overall return analysis of project options

	Existing state	Airflow management and VFD fan	In-row cooling	EC fan upgrade and CAC	AFM, EC fan, and sensor network
Capital	$0	$180,000	$838,000	$375,000	$325,000
PUE	1.92	1.72	1.65	1.63	1.64
NPV	$0	$354,377	$86,997	$593,243	$427,422
IRR	0%	58%	11%	50%	43%
Profitability index		2.97	1.10	2.58	2.32

EC fans and distributed sensor network yields only a $73,000 increase in the PV, thus the lower IRR of only 43% for this option. The base airflow management has already provided a substantial part of the savings, and the incremental improvement of the EC fan and sensor network is small. If we have other projects with a similar return to the base airflow management and VFD fan upgrade on which we could spend the additional capital of the EC fans and sensor network, these would be better investments. The IRR of the sensor network in addition to the airflow management is only 23%, which would be unlikely to meet approval as an individual project.

The two airflow containment options have very similar achieved PUE and operational costs; they are both quite efficient and neither requires CFD or movement of floor tiles. There is, however, a substantial difference in the implementation cost; so despite the large energy saving, the in-row cooling option has the lowest return of all the options, while the EC fan upgrade and CAC has the highest NPV.

It is interesting to note that there is no one "best" option here as the airflow management and VFD fan have the highest IRR and highest NPV per unit capital, while the EC fan upgrade and CAC have the highest overall NPV.

6.4.8 Break-Even Point

We are also likely to be asked to identify the break-even point for our selected investments; we can do this by taking the PV in each year and summing these over time. We start with a negative value for the year 0 capitalized costs and then add the PV of each year's operational cost saving over the 6-year period. The results are shown in Figure 6.12.

The break-even point is where the cumulative NPV of each option crosses zero. Three of the options have a break-even point of between 1.5 and 2.5 years, while the in-row cooling requires 5.5 years to breakeven.

6.4.8.1 Future Trends

This section examines the impact of the technological and financial changes on the data center market and how these may impact the way you run your data center or even dispose of it entirely. Most of the future trends affecting data centers revolve around the commoditization of data center capacity and the change in focus from technical performance criteria to business financial criteria. Within this is the impact of cloud, consumerization of ICT, and the move toward post-PUE financial metrics of data center performance.

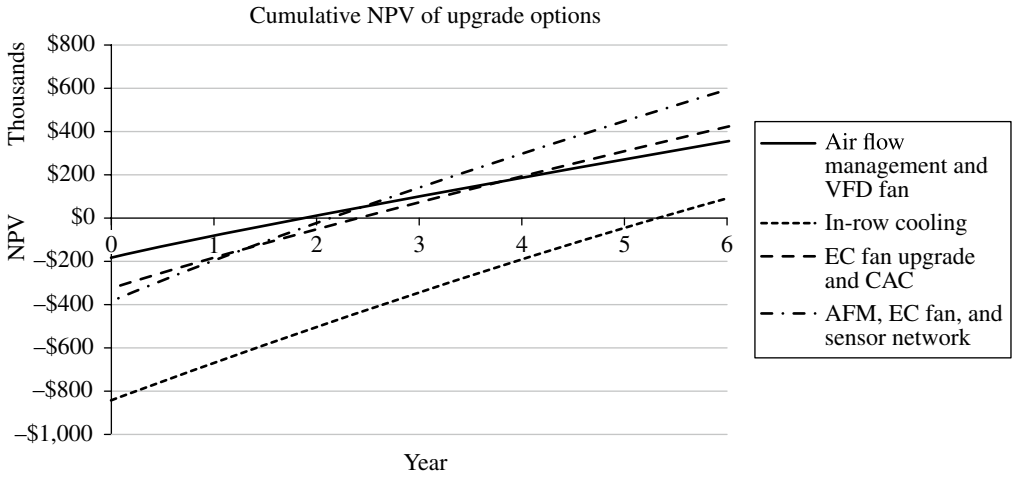


FIGURE 6.12 Break-even points of upgrade options.

6.4.8.2 The Threat of Cloud and Commoditization

At the time of writing, there is a great deal of hype about cloud computing and how it will turn IT services into utilities such as water or gas. This is a significant claim that the changes of cloud will erase all distinctions between IT services and that any IT service may be transparently substituted with any other IT service. If this were to come true, then IT would be subject to competition on price alone with no other differentiation between services or providers.

Underneath the hype, there is little real definition of what actually constitutes "cloud" computing, with everything from free webmail to colocation services, branding itself as cloud. The clear trend underneath the hype, however, is the commoditization of data center and IT resources. This is facilitated by a number of technology changes including the following:

- Server, storage, and network virtualization at the IT layer have substantially reduced the time, risk, effort, and cost of moving services from one data center to another. The physical location and ownership of IT equipment are of rapidly decreasing importance.
- High-speed Internet access is allowing the large-scale deployment of network-dependent end user computing devices; these devices tend to be served by centralized platform vendors such as Apple, Microsoft, or Amazon rather than corporate data centers.
- Web-based application technology is replacing many of the applications or service components that were previously run by enterprise users. Many organizations now select externally operated platforms such as Salesforce because of their integration with other Web-based applications instead of requiring integration with internal enterprise systems.

6.4.8.3 Data Center Commoditization

Data centers are commonly called the factories of IT; unfortunately, they are not generally treated with the same financial rigor as factories. While the PUE of new data centers may be going down (at least in marketing materials), the data center market is still quite inefficient. Evidence of this can be seen in the large gross margins made by some operators and the large differences in price for comparable products and services at both M&E device and data center levels.

The process of commoditization will make the market more efficient, to quote one head of data center strategy "this is a race to the bottom, and the first one there wins." This recognition that data centers are a commodity will have significant impacts not only on the design and construction of data centers but also on the component suppliers who will find it increasingly hard to justify premium prices for heavily marketed but nonetheless commodity products.

In general, commoditization of a product is the process of distinguishing factors becoming less relevant to the purchaser and thereby becoming simple commodities. In the data center case, commoditization comes about through several areas of change:

- Increased portability: It is becoming faster, cheaper, and easier for customers of data center capacity or services delivered from data centers to change supplier and move to another location or provider. This prevents "lock-in" and so increases the impact of price competition among suppliers.
- Reductions in differentiating value: Well-presented facilities with high levels of power and cooling resilience or availability certifications are of little value in a world where customers neither know nor care which data center their services are physically located in, and

service availability is handled at the network and software level.

- Broadening availability of the specific knowledge and skills required to build and operate a financially efficient data center; while this used to be the domain of a few very well-informed experts, resources such as the EU Code of Conduct on data centers and effective predictive financial and operational modeling of the data center are making these capabilities generally available.
- Factory assembly of components through to entire data centers being delivered as modules, so reducing the capital cost of delivering new data center capacity compared with traditional on-site construction.
- Business focus on financial over technical performance metrics.

While there are many barriers obstructing IT services or data centers from becoming truly undifferentiated utility commodities, such as we see with water or oil, much of the differentiation, segmentation, and price premium that the market have so far enjoyed are disappearing. There will remain some users for whom there are important factors such as physical proximity to, or distance from, other locations, but even in these cases it is likely that only the minimum possible amount of expensive capacity will be deployed to meet the specific business issue and the remainder of the requirement will be deployed across suitable commodity facilities or providers.

6.4.8.4 Driving Down Cost in the Data Center Market

Despite the issues that are likely to prevent IT from ever becoming a completely undifferentiated commodity such as electricity or gas, it is clear that the current market inefficiencies will be eroded and the cost of everything from M&E (mechanical and electrical) equipment to managed application services will fall. As this occurs, both enterprise and service provider data centers will have to substantially reduce cost in order to stay competitive.

Enterprise data centers may:

- Improve both their cost and flexibility closer to that offered by cloud providers to reduce the erosion of internal capacity and investment by low capital and short commitment external services.
- Target their limited financial resource and data center capacity to services with differentiating business value or high business impact of failure while exporting commodity services that may be cheaply and effectively delivered by other providers.
- Deliver multiple grades of data center at multiple cost levels to meet business demands and facilitate a functioning internal market.

Cloud providers are likely to be even more vulnerable than enterprise data centers as their applications are, almost by definition, commodity, fast and easy to replace with a cheaper service. It is already evident that user data is now the portability issue and that some service providers resist competition by making data portability for use in competitive services as difficult as possible.

6.4.8.5 Time Sensitivity

One of the key issues in the market for electricity is our present inability to economically store any large quantity of it once generated. The first impact of this is that sufficient generating capacity to meet peak demand must be constructed at high capital cost but not necessarily full utilization. The second is the substantial price fluctuation over short time frames with high prices at demand peaks and low prices when there is insufficient demand to meet the available generating capacity.

For many data centers, the same issue exists, the workload varies due to external factors, and the data center must be sized to meet peak demand. Some organizations are able to schedule some part of their data center workload to take place during low load periods, for example, Web crawling and construction of the search index when not serving search results. For both operators purchasing capacity and cloud providers selling it through markets and brokers, price fluctuation and methods of modifying demand schedules are likely to be an important issue.

6.4.8.6 Energy Service Contracts

Many data center operators are subject to a combination of capital budget reductions and pressure to reduce operational cost or improve energy efficiency. While these two pressures may seem to be contradictory, there is a financial mechanism that is increasingly used to address this problem.

In the case where there are demonstrable operational cost savings available from a capital upgrade to a data center, it is possible to fund the capital reinvestment now from the later operational savings. While energy service contracts take many forms, they are in concept relatively simple:

1. The expected energy cost savings over the period are assessed.
2. The capitalized cost of the energy saving actions including equipment and implementation is assessed.
3. A contract is agreed, and a loan is provided or obtained for the capitalized costs of the implementation; this loan funds some or all of the project implementation costs and deals with the capital investment hurdle.
4. The project is implemented, and the repayments for the loan are serviced from some or all of the energy cost savings over the repayment period.

Energy service contracts are a popular tool for data center facilities management outsourcing companies. While the arrangement provides a mechanism to reduce the up-front cost of an energy performance improvement for the operator, there are a number of issues to consider:

- The service contract tends to commit the customer to the provider for an extended period; this may be good for the provider and reduces direct price competition for their services.
- There is an inherent risk in the process for both the provider and customer; the cost savings on which the loan repayments rely may either not be delivered or it may not be possible to prove that they have been delivered due to other changes, in which case responsibility for servicing the loan will still fall to one of the parties.
- There may be a perverse incentive for outsource facilities management operators to "sandbag" on operational changes, which would reduce energy in order to use these easy savings in energy service contract-funded projects.

6.4.8.7 Guaranteed Performance and Cost

The change in focus from technical to financial criteria for data centers coupled with the increasing brand value importance of being seen to be energy efficient is driving a potentially significant change in data center procurement. It is now increasingly common for data center customers to require their design or build provider to state the achieved PUE or total energy consumption of their design under a set of IT load fill out conditions. This allows the customer to make a more effective TCO optimization when considering different design strategies, locations, or vendors.

The logical extension of this practice is to make the energy and PUE performance of the delivered data center part of the contractual terms. In these cases, if the data center fails to meet the stated PUE or energy consumption, then the provider is required to pay a penalty. Contracts are now appearing, which provide a guarantee that if the data center fails to meet a set of PUE and IT load conditions, the supplier will cover the additional energy cost of the site.

This form of guarantee varies from a relative simple PUE, that is above a certain kW load, to a more complex definition of performance at various IT load points or climate conditions.

A significant issue for some purchasers of data centers is the split incentive inherent in many of the build or lease contracts currently popular. It is common for the provider of the data center to pay the capital costs of construction but to have no financial interest in the operational cost or efficiency. In these cases, it is not unusual for capital cost savings to be made directly at the expense of the ongoing operational cost of the data center, which results in a substantial increase in the total TCO and poor overall performance. When purchasing or leasing a data center, it is essential to ensure that the provider constructing the data center has a financial interest in the operational performance and cost to mitigate these incentives. This is increasingly taking the form of energy performance guarantees that share the impact of poor performance with the supplier.

6.4.8.8 Charging for the Data Center: Activity-Based Costing

With data centers representing an increasing proportion of the total business operating cost and more business activity becoming critically reliant upon those data centers, a change is being forced in the way in which finance departments treat data centers. It is becoming increasingly unacceptable for the cost of the data center to be treated as a centralized operating overhead or to be distributed across business units with a fixed finance "allocation formula" that is often out of date and has little basis in reality. Many businesses are attempting to institute some level of chargeback model to apply the costs of their data center resources to the (hopefully value-generating) business units that demand and consume them.

These chargeback models vary a great deal in their complexity and accuracy all the way from square feet to detailed and realistic ABC models. For many enterprises, this is further complicated by a mix of data center capacity that is likely to be made up of the following:

- One or more of their own data centers, possibly in different regions with different utility power tariffs and at different points in their capital amortization and depreciation.
- One or more areas of colocation capacity, possibly with different charging models as well as different prices, dependent upon the type and location of facility.
- One or more suppliers of cloud compute capacity, again with varying charging mechanisms, length of commitment, and price.

Given this mix of supply, it is inevitable that there will be tension and price competition between the various sources of data center capacity to any organization. Where an external colo or cloud provider is perceived to be cheaper, there will be a pressure to outsource capacity requirements. A failure to accurately and effectively cost internal resources for useful comparison with outsourced capacity may lead to the majority of services being outsourced, irrespective of whether it makes financial or business sense to do so.

6.4.8.9 The Service Monoculture

Perhaps the most significant issue facing data center owners and operators is the service monoculture that has been allowed to develop and remains persistent by a failure to properly understand and manage data center cost. The symptoms of this issue are visible across most types of organization, from large enterprise operators with legacy estates through colocation to new build cloud data centers. The major symptoms are a single level of data center availability, security, and cost with the only real variation being due to local property and energy costs. It is common to see significant data center capacity built to meet the availability, environmental, and security demands of a small subset of the services to be supported within it.

This service monoculture leads to a series of problems that, if not addressed, will cause substantial financial stress for all types of operator as the data center market commoditizes, margins reduce, and price pressure takes effect.

As an example of this issue, we may consider a fictional financial services organization that owns a data center housing a mainframe that processes customer transactions in real time. A common position for this type of operator when challenged on data center cost efficiency is that they don't really care what the data center housing the mainframe costs, as any disruption to the service would cost millions of dollars per minute and the risk cost massively outweighs any possible cost efficiencies. This position fails to address the reality that the operator is likely to be spending too much money on the data center for no defined business benefit while simultaneously underinvesting in the critical business activity. Although the mainframe is indeed business critical, the other 90% plus of the IT equipment in the data center is likely to range from internal applications to development servers with little or no real impact of downtime. The problem for the operator is that the data center design, planning, and operations staff are unlikely to have any idea which servers in which racks could destroy the business and which have not been used for a year and are expensive fan heaters.

This approach to owning and managing data center resources may usefully be compared to Soviet Union era planned economies. A central planning group determines the amount of capacity that is expected to be required, provides investment for, and orders the delivery of this capacity. Business units then consume the capacity for any requirement they can justify and, if charged at all, pay a single fixed internal rate. Attempts to offer multiple grades and costs of capacity are likely to fail as there is no incentive for business units to choose anything but the highest grade of capacity unless there is a direct impact on their budget. The outcomes in the data center or the planned economy commonly include insufficient provision of key resources, surplus of others, suboptimal allocation, slow reaction of the planning cycle to demand changes, and centrally dictated resource pricing.

6.4.8.10 Internal Markets: Moving Away from the Planned Economy

The increasing use of data center service charge-back within organizations is a key step toward addressing the service monoculture problem. To develop a functioning market within the organization, a mixture of internal and external services, each of which has a cost associated with acquisition and use, is required. Part of the current momentum toward use of cloud services is arguably not due to any inherent efficiency advantages of cloud but simply due to the ineffective internal market and high apparent cost of capacity within the organization, allowing external providers to undercut the internal resources.

As organizations increasingly distribute their data center spend across internal, colocation, and cloud resources and the cost of service is compared with the availability, security, and cost of each consumed resource, there is a direct opportunity for the organization to better match the real business needs by operating different levels and costs of internal capacity.

6.4.8.11 Chargeback Models and Cross Subsidies

The requirement to account or charge for data center resources within both enterprise and service provider organizations has led to the development of a number of approaches to determining the cost of capacity and utilization. In many cases, the early mechanisms have focused on data gathering and measurement precision at the expense of the accuracy of the cost allocation method itself.

Each of the popular chargeback models, some of which are introduced in the following, has its own balance of strengths and weaknesses and creates specific perverse incentives. Many of these weaknesses stem from the difficulty in dealing with the mixture of fixed and variable costs in the data center. There are some data center costs that are clearly fixed, that is, they do not vary with the IT energy consumption, such as the capital cost of construction, staffing, rent, and property taxes. Others, such as the energy consumption at the IT equipment, are obviously variable cost elements.

6.4.8.12 Metered IT Power

Within the enterprise, it is common to see metering of the IT equipment power consumption used as the basis for chargeback. This metered IT equipment energy is then multiplied by a measured PUE and the nominal energy tariff to arrive at an estimate of total energy cost for the IT loads. This frequently requires expensive installation of metering equipment coupled with significant data gathering and maintenance requirements to identify which power cords are related to which delivered service. The increasing use of virtualization

and the portability of virtual machines across the physical infrastructure present even more difficulties for this approach.

Metered IT power × PUE × tariff is a common element of the cost in colocation services where it is seen by both the operator and client as being a reasonably fair mechanism for determining a variable element of cost. The metering and data overheads are also lower as it is generally easier to identify the metering boundaries of colo customer areas than IT services. In the case of colocation, however, the metered power is generally only part of the contract cost.

The major weakness of metered IT power is that it fails to capture the fixed costs of the data center capacity occupied by each platform or customer. Platforms or customers with a significant amount of allocated capacity but relatively low draw are effectively subsidized by others that use a larger part of their allocated capacity.

6.4.8.13 Space

Historically, data center capacity was expressed in terms of square feet or square meters, and therefore, costs and pricing models were based on the use of space, while the power and cooling capacity were generally given in kW per square meter or foot. Since that time, the power density of the IT equipment has risen, transferring the dominant constraint to the power and cooling capacity. Most operators charging for space were forced to apply power density limits, effectively changing their charging proxy to kW capacity. This charging mechanism captures the fixed costs of the data center very effectively but is forced to allocate the variable costs as if they were fixed and not in relation to energy consumption.

Given that the majority of the capital and operational costs for most modern data centers are related to the kW capacity and applied kW load, the use of space as a weak proxy for cost is rapidly dying out.

6.4.8.14 Kilowatt Capacity or Per Circuit

In this case, the cost is applied per kilowatt capacity or per defined capacity circuit provided. This charge mechanism is largely being replaced by a combination of metered IT power and capacity charge for colocation providers, as the market becomes more efficient and customers better understand what they are purchasing. This charging mechanism is still popular in parts of North America and some European countries where local law makes it difficult to resell energy.

This mechanism has a similar weakness and, therefore, exploitation opportunity to metered IT power. As occupiers pay for the capacity allocated irrespective of whether they use it, those who consume the most power from each provided circuit are effectively subsidized by those who consume a lower percentage of their allocated capacity.

6.4.8.15 Mixed kW Capacity and Metered IT Power

Of the top–down charge models, this is perhaps the best representation of the fixed and variable costs. The operator raises a fixed contract charge for the kilowatt capacity (or circuits, or space as a proxy for kilowatt capacity) and a variable charge based on the metered IT power consumption. In the case of colocation providers, the charge for metered power is increasingly "open book" in that the utility power cost is disclosed and the PUE multiplier stated in the contract allowing the customer to understand some of the provider margin. The charge for allocated kW power and cooling capacity is based on the cost of the facility and amortizing this over the period over which this cost is required to be recovered. In the case of colocation providers, these costs are frequently subject to significant market pressures, and there is limited flexibility for the provider.

This method is by no means perfect; there is no real method of separating fixed from variable energy costs, and it is also difficult to deal with any variation in the class and, therefore, cost of service delivered within a single data center facility.

6.4.8.16 Activity-Based Costing

As already described, two of the most difficult challenges for chargeback models are separating the fixed from variable costs of delivery and differentially costing grades of service within a single facility or campus. None of the top–down cost approaches discussed so far are able to properly meet these two criteria, except in the extreme case of completely homogenous environments with equal utilization of all equipment.

An approach popular in other industries such as manufacturing is to cost the output product as a supply chain, considering all of the resources used in the production of the product including raw materials, energy, labor, and licensing. This methodology, called activity-based costing, may be applied to the data center quite effectively not only to produce effective costing of resources but also to allow for the simultaneous delivery of multiple service levels with properly understood differences in cost. Instead of using fixed allocation percentages for different elements, ABC works by identifying relationships in the supply chain to objectively assign costs.

By taking an ABC approach to the data center, the costs of each identifiable element, from the land and building, through mechanical and electrical infrastructure to staffing and power costs, are identified and allocated to the IT resources that they support. This process starts at the initial resources, the incoming energy feed, and the building and passes costs down a supply chain until they arrive at the IT devices, platforms, or customers supported by the data center.

Examples of how ABC may result in differential costs are as follows:

- If one group of servers in a data hall has single-corded feed from a single $N+1$ UPS room, while another is dual corded and fed from two UPS rooms giving $2(N+1)$ power, the additional capital and operational cost of the second UPS room would only be borne by the servers using dual-corded power.
- If two data halls sharing the same power infrastructure operate at different temperature and humidity control ranges to achieve different free cooling performance and cost, this is applied effectively to IT equipment in the two halls.

For the data center operator, the most important outcomes of ABC are as follows:

- The ability to have a functioning internal and external market for data center capacity and thereby invest in and consume the appropriate resources.
- The ability to understand whether existing or new business activities are good investments. Specifically, where business activities require data center resources, the true cost of these resources should be reflected in the cost of the business activity.

For service providers, this takes the form of per customer margin assessment and management. It is not unusual to find that through cross subsidy between customers, frequently, the largest customers (usually perceived as the most valuable) are in fact among the lowest margin and being subsidized by others, to whom less effort is devoted to retaining their business.

6.4.8.17 Unit Cost of Delivery: $/kWh

The change in focus from technical to financial performance metrics for the data center is also likely to change focus from the current engineering-focused metrics such as PUE to more financial metrics for the data center. PUE has gained mind share through being both simple to understand and being an indicator of cost efficiency. The use of ABC to determine the true cost of delivery of data center loads provides the opportunity to develop metrics that capture the financial equivalent of the PUE, the unit cost of each IT kWh, or $/kWh.

This metric is able to capture a much broader range of factors for each data center, such as a hall within a data center or individual load, than PUE can ever do. The capital or lease cost of the data center, staffing, local taxes, energy tariff, and all other costs may be included to understand the fully loaded unit cost. This may then be used to understand how different data centers within the estate compare with each other and how internal capacity compares for cost with outsourced colocation or cloud capacity.

When investment decisions are being considered, the use of full-unit cost metrics frequently produces what are initially counterintuitive results. As an example, consider an old data center for which the major capital cost is considered to be amortized, operating in an area where utility power is relatively cheap, but with a poor PUE; we may determine the unit delivery cost to be 0.20 $/kWh, including staffing and utility energy. It is not uncommon to find that the cost of a planned replacement data center, despite having a very good PUE, once the burden of the amortizing capital cost is applied, cannot compete with the old data center. Frequently, relatively minor reinvestments in existing capacity are able to produce lower unit costs of delivery than even a PUE = 1 new build.

An enterprise operator may use the unit cost of delivery to compare multiple data centers owned by the organization and to establish which services should be delivered from internal versus external resources, including allocating the appropriate resilience, cost, and location of resource to services.

A service provider may use unit cost to meet customer price negotiation by delivering more than one quality of service at different price points while properly understanding the per deal margin.

6.5 CHOOSING TO BUILD, REINVEST, LEASE, OR RENT

A major decision for many organizations is whether to invest building new data center capacity, reinvest in existing, lease capacity, colocate, or use cloud services. There is, of course, no one answer to this; the correct answer for many organizations is neither to own all of their own capacity nor to dispose all of it and trust blindly in the cloud. At the simplest level, colocation providers and cloud service providers need to make a profit and, therefore, must achieve improvements in delivery cost over that which you can achieve, which are at least equal to the required profit to even achieve price parity.

The choice of how and where to host each of your internal or customer-facing business services depends on a range of factors, and each option has strengths and weaknesses. For many operators, the outcome is likely to be a mixture of the following:

- High-failure impact services, high security requirement services, or real differentiating business value operated in owned or leased data centers that are run close to capacity to achieve low unit cost.

- Other services that warrant ownership and control of the IT equipment or significant network connectivity operated in colocation data centers.
- Specific niche and commodity services such as email that are easily outsourced, supplied by low-cost cloud providers.
- Short-term capacity demands and development platforms delivered via cloud broker platforms that auction for the current lowest cost provider.

As a guide, some of the major benefits and risks of each type of capacity are described in the following. This list is clearly neither exhaustive nor complete but should be considered a guide as to the questions to ask.

6.5.1 Owned Data Center Capacity

Data center capacity owned by the organization may be known to be located in the required legal jurisdiction, operated at the correct level of security, maintained to the required availability level, and operated to a high level of efficiency. It is no longer difficult to build and operate a data center with a good PUE. Many facilities management companies provide the technical skills to maintain the data center at competitive rates, eliminating another claimed economy of scale by the larger operators. In the event of an availability incident, the most business-critical platforms may be preferentially maintained or restored to service. In short, the owner controls the data center.

The main downside of owning capacity is the substantial capital and ongoing operational cost commitment of building a data center although this risk is reduced if the ability to migrate out of the data center and sell it is included in the assessment.

The two most common mistakes are the service monoculture, building data center capacity at a single level of service, quality, and cost, and failing to run those data centers at full capacity. The high fixed cost commitments of the data center require that high utilization be achieved to operate at an effective unit cost, while migrating services out of a data center you own into colo or cloud simply makes the remainder more expensive unless you can migrate completely and dispose of the asset.

6.5.2 Leased Data Center Capacity

Providers of wholesale or leased data center capacity claim that their experience, scale, and vendor price negotiation leverage allow them to build a workable design for a lower capital cost than the customer would achieve.

Leased data center capacity may be perceived as reducing the capital cost commitment and risk. However, in reality, the capital cost has still been financed, and a loan is being serviced. Furthermore, it is frequently as costly and difficult to get out of a lease as it is to sell a data center you own.

The risk defined in Section 6.4.8.6 may be mitigated by ensuring contractual commitments by the supplier to the ongoing operational cost and energy efficiency of the data center.

As for the owned capacity, once capacity is leased, it should generally be operated at high levels of utilization to keep the unit cost acceptable.

6.5.3 Colocation Capacity

Colocation capacity is frequently used in order to leverage the connectivity available at the carrier neutral data center operators. This is frequently of higher capacity and lower cost than may be obtained for your own data center; where your services require high speed and reliable Internet connectivity, this is a strong argument in favor of colocation. There may also be other bandwidth-intensive services available within the colocation data center made available at lower network transit costs within the building than would be incurred if those services were to be used externally.

It is common for larger customers to carry out physical and process inspections of the power, cooling, and security at colocation facilities and to physically visit them reasonably frequently to attend to the IT equipment. This may provide the customer with a reasonable assurance of competent operation.

A common perception is that colocation is a much shorter financial commitment than owning or leasing data center capacity. In reality, many of the contracts for colocation are of quite long duration, and when coupled with the time taken to establish a presence in the colo facility, install and connect network equipment, and then install the servers, storage, and service platforms, the overall financial commitment is of a similar length.

Many colocation facilities suffer from the service monoculture issue and are of high capital cost to meet the expectations of "enterprise colo" customers as well as being located in areas of high real estate or energy cost for customer convenience. These issues tend to cause the cost base of colocation to be high when compared with many cloud service providers.

6.5.4 Cloud Capacity

The major advantages of cloud capacity are the short commitment capability, sometimes as short as a few hours, relatively low unit cost, and the frequent integration of cloud services with other cloud services. Smart cloud operators build their data centers to minimal capital cost in cheap locations and negotiate for cheap energy. This allows them to operate at a very low basic unit cost, sometimes delivering complete managed services for a cost comparable to colocating your own equipment in traditional colo.

One of the most commonly discussed downsides of cloud is the issue of which jurisdiction your data is in and whether you are meeting legal requirements for data retention or privacy laws.

The less obvious downside of cloud is that, due to the price pressures, cloud facilities are built to low cost, and availability is generally provided at the software or network layer rather than spending money on a resilient data center infrastructure. While this concept is valid, the practical reality is that cloud platforms also fail, and when they do, thanks to the high levels of complexity, it tends to be due to human error, possibly combined with an external or hardware event. Failures due to operator misconfiguration or software problems are common and well reported.

The issue for the organization relying on the cloud when their provider has an incident is that they have absolutely no input to or control over the order in which services are restored.

FURTHER READING

Cooling analysis white paper (prepared for the EU CoC). With supporting detailed content Liam Newcombe, IT environmental range and data centre cooling analysis, May 2011. https://www.bcs.org/media/2914/cooling_analysis_summary_v100.pdf. Accessed September 3, 2020.

Drury C. *Management and Cost Accounting*. 7th Rev ed. Hampshire: Cengage Learning; 2007.

Liam Newcombe, et al., Data Centre Fixed to Variable Energy Ratio Metric, BCS Data Centre Specialist Group. https://www.bcs.org/media/2917/dc_fver_metric_v10.pdf. Accessed September 3, 2020.

EU Code of Conduct for Energy Efficiency in Data Centres, https://ec.europa.eu/jrc/en/energy-efficiency/code-conduct/datacentres. Accessed September 3, 2020.

7

MANAGING DATA CENTER RISK

Beth Whitehead, Robert Tozer, David Cameron and Sophia Flucker
Operational Intelligence Ltd, London, United Kingdom

7.1 INTRODUCTION

The biggest barriers to risk reduction in any system are human unawareness of risk, a lack of formal channels for knowledge transfer within the life cycle of a facility and onto other facilities, and design complexity.

There is sufficient research into the causes of failure to assert that any system with a human interface will eventually fail. In their book, *Managing Risk: The Human Element*, Duffey and Saull [1] found that when looking at various industries, such as nuclear, aeronautical, space, and power, 80% of failures were due to human error or the human element. Indeed, the Uptime Institute [2] report that over 70% of data center failures are caused by human error, the majority of which are due to management decisions and the remaining to operators and their lack of experience and knowledge, and complacency.

It is not, therefore, a case of eliminating failure but rather reducing the risk of failure by learning about the system and sharing that knowledge among those who are actively involved in its operation through continuous site-specific facility-based training. To enable this, a learning environment that addresses human unawareness at the individual and organizational level must be provided. This ensures all operators understand how the various systems work and interact and how they can be optimized. Importantly, significant risk reduction can only be achieved through active engagement of all facility teams and through each disparate stage of the data center's life cycle. Although risk management may be the responsibility of a few individuals, it can only be achieved if there is commitment from all stakeholders.

The identification of risks is also important. By identifying risks and increasing stakeholder awareness of them, it is possible to better manage and minimize their impact—after all it is hard to manage something that you are unaware of. Many sites undertake risk analyses, but without a way to transfer this knowledge, the findings are often not shared with the operators, and much of their value is lost. Finally, limiting human interfaces in the design and overall design complexity is imperative for a resilient data center. Each business model requires a certain resilience that can be achieved through different designs of varying complexity. The more complex a system, the more important training and knowledge sharing becomes, particularly where systems are beyond the existing knowledge base of the individual operator. Conversely the less complex a system is, the less training that is required.

7.2 BACKGROUND

To better understand risk and how it can be managed, it is essential to first consider how people and organizations learn, what causes human unawareness, and how knowledge is transferred during data center projects.

7.2.1 Duffey and Saull: Learning

Duffey and Saull [1] used a 3D cube to describe their universal learning curve and how risk and experience interact in a learning space (Fig. 7.1). They expressed failure rate in terms of two variables: accumulated learning experience of the organization (organizational) and the depth of experience of an individual (operator). Figure 7.1 shows that when

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

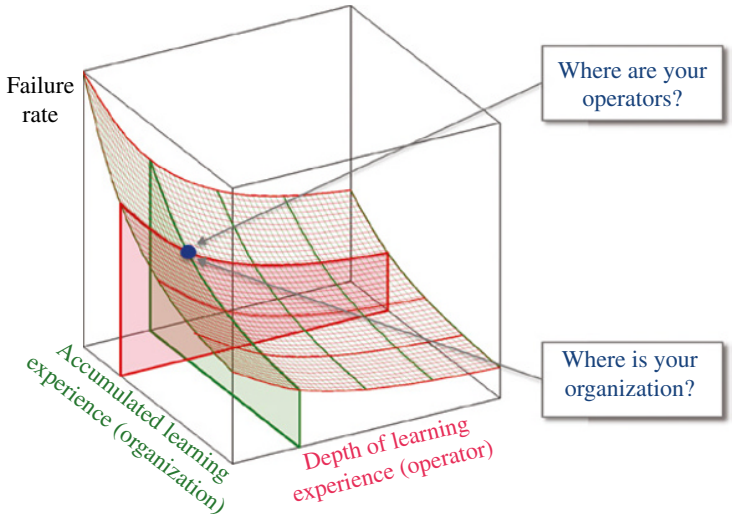


FIGURE 7.1 The universal learning curve. Source: Courtesy of Operational Intelligence Ltd.

experience is at a minimum, risk is at a maximum, but with learning and increased experience, the failure rate drops exponentially, tending to, but never quite reaching zero. This is because there may be unknowns that cannot be managed, and complacency tends to increase with time. However, if it is a learning environment, the failure rate will reduce with time.

From the authors' experience of numerous failure analyses, eight different areas in which organizational vulnerabilities have been found and which can result in failures have been identified:

- Structure and resources
- Maintenance
- Change management
- Document management
- Commissioning
- Operability and maintainability
- Capacity
- Organization and operator learning

These vulnerabilities align with those of the Uptime Institute for operational sustainability and management and operation [3]. To minimize risk, these areas should be focused on and adequate training provided for each vulnerability. Likewise, the authors have also classified three key elements relating to individual operator vulnerabilities:

- General and site-specific knowledge
- Experience from other sites
- Attitude toward people and learning

A detailed analysis of these vulnerabilities should be completed once a site is in operation. However, some very high-level thinking into some of the areas is useful at the start of the project. In particular, the timing and extent of commissioning should be considered to ensure there are adequate resources (both financial and manpower) made available during the build phase. This includes appointment of a commissioning manager and other subcontractors in good time before the commissioning starts to ensure a smooth handover with better knowledge transfer from the build to operations teams. Furthermore, ensuring there is provision for future training sets the foundations for a learning environment in which organizations and operators can operate their facilities efficiently and safely.

7.2.2 Human Unawareness

The impact of human unawareness on risk and failure was discussed in Section 7.1. Traditionally in the facilities sector, people may work in silos based on their discipline, experience, and management position. If a blame culture is adopted, these silos can become fortresses with information retained within them, meaning operators are often unaware of the impact their actions have on other parts of the facility or of mistakes made by others that might also be putting their area at risk. For example, if IT are unaware that negative pressure can be induced into floor grilles placed too close to a CRAC unit, they might place their most heavily loaded cabinet here, thus starving it of air. Had IT understood more about how airflow works in a data hall, they could have made a better-informed design for their layout.

If risk is to be reduced, knowledge and awareness must be increased at all levels of the business, and it must be accepted that failure and "near misses" are inevitable. There must be opportunity to learn from these failures and near misses and to gain knowledge on how the facility works as a whole. It is important that the management create an environment where

staff feel they have a voice and are recognized for their role in delivering a high-performing environment. In a learning environment, it can be acknowledged that failures are often due to mistakes by the operator or poor management decisions and ensures lessons can be learned not only from an individual's mistakes but also from the mistakes of others. This ensures knowledge is transferred easily and free of blame.

7.2.3 Knowledge Transfer, Active Learning, and the Kolb Cycle

At its simplest, active learning is learning by doing. When learning is active, we make discoveries and experiment with knowledge firsthand, rather than reading or hearing about the experiences of others. Research shows that active learning approaches result in better recall, understanding, and enjoyment.

The educational theorist David Kolb said that learning is optimized when we move through the four quadrants of the experiential learning cycle [4]. These are concrete experience, reflective observation, abstract conceptualization, and active experimentation. The cycle demonstrates how we make connections between what we already know and new content to which we are exposed. For the purpose of this chapter, we refer to these quadrants as experience, reflection, theory, and practice, as shown in Figure 7.2.

When you compare the Kolb cycle with the data center construction industry, it is clear that each quadrant is inhabited by different teams with contractual boundaries between adjacent quadrants. The transfer of technical information and knowledge is therefore rarely, if ever, perfect. Figure 7.3 shows these teams and, with reference to the construction and operation of a data center, the knowledge transfer that is required at each boundary and the specific activities carried out to address risk in each quadrant. To minimize risk, learning needs to be optimized in each quadrant, and rather than staying in each quadrant, like a silo, knowledge needs freedom to pass through the contractual boundaries. In the following sections, the content of this adapted Kolb cycle will be explained to show how risk in the data center can be better managed.

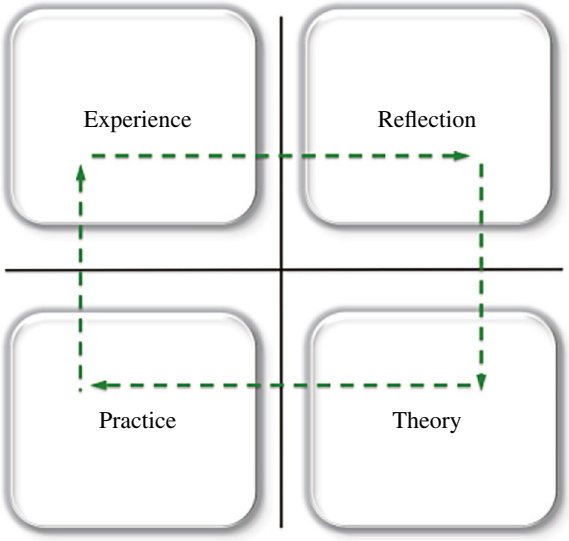


FIGURE 7.2 The Kolb cycle. Source: Courtesy of Operational Intelligence Ltd.

7.3 REFLECTION: THE BUSINESS CASE

The first quadrant refers to the business aspect of the data center. This is where a client should set their brief (Owner's Project Requirements [OPR]) and lay out the design requirements for their facility. Note that in the United Kingdom this phase overlaps RIBA (Royal Institute of British Architects) Stages 0 (Strategic Definition) and 1 (Preparation and Brief). The design should match the business requirements by understanding what the acceptable level of risk is to the business and the cost. For example, a small engineering design consultancy can cope with website downtime of 2 days and would expect to see little impact on their business, whereas a large online trader could not.

7.3.1 Quantifying the Cost of Failure

The cost of failure can be quantified using the following equation, where risk is the cost per year, likelihood is the number of failures per year, and severity is the cost per failure:

$$\text{Risk} = \text{likelihood} \times \text{severity}$$

This cost of failure can then be used to compare different design options that could mitigate this risk. For example, a facility could experience one failure every 2 years. Each failure might cost the business \$10,000,000; therefore the cost to the business of this risk would be

$$\text{Risk} = 1\,\text{failure} / 2\,\text{years} \times \$10{,}000{,}000$$

$$\text{Risk} = \$5{,}000{,}000 / \text{year}$$

If this failure were to occur every 2 years for 10 years, the total cost to the business would be \$50 million over that period of time. The cost of different design options and their impact on the likelihood of failure and risk could then be examined. For example, a design option costing \$2 million extra could be considered. If these works could reduce the likelihood of failure to 1 failure in the whole 10-year period, the risk to the business would become

$$\text{Risk} = 1/10 \times \$10{,}000{,}000 = \$1{,}000{,}000 / \text{year}$$

For a \$2 million investment, the risk of failure has dropped from \$5 to \$1 million/year, and the total cost is now \$12 million, \$38 million less than the original \$50 million

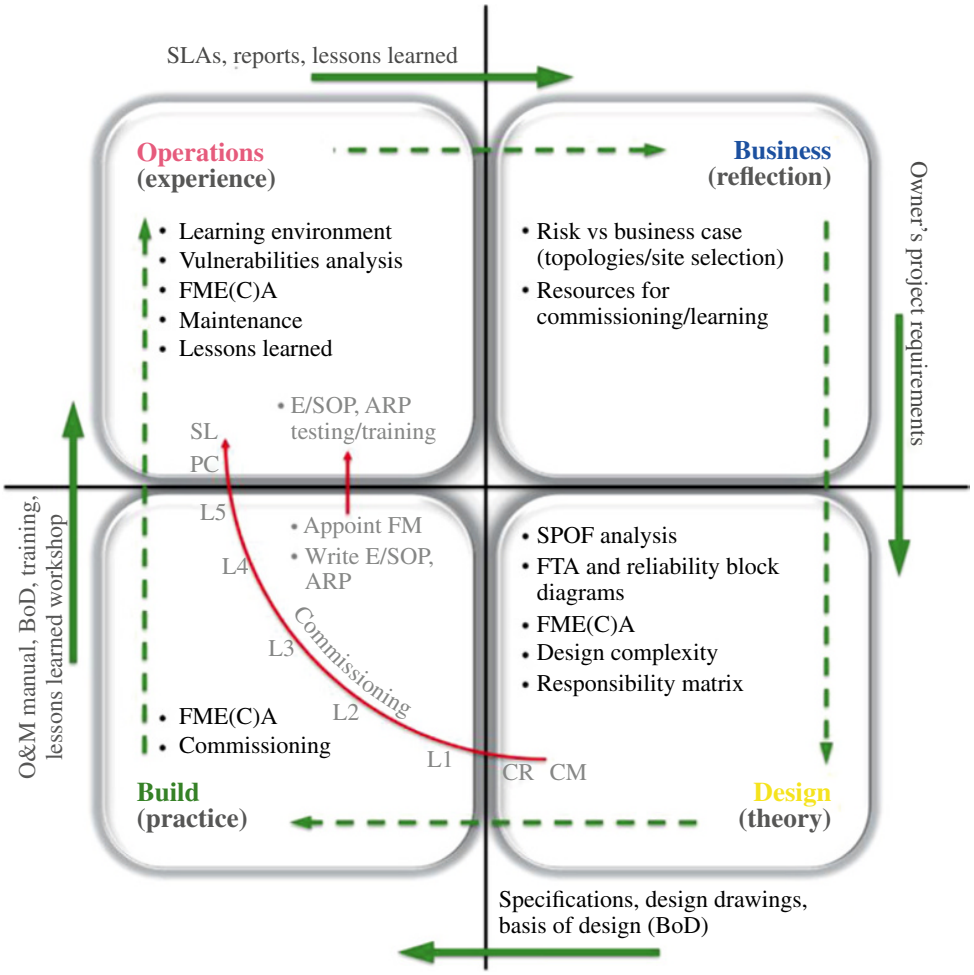


FIGURE 7.3 The Kolb cycle and the data center. (Key: SPOF, single point of failure; FTA, fault tree analysis; FME(C)A, failure mode and effect (criticality) analysis; CM, commissioning manager; CR, commissioning review; L1-5, commissioning levels 1–5; PC, practical completion; SL, soft landings; E/SOP, emergency/standard operating procedures; ARP, alarm response procedures; FM, facilities management; O&M, operation and maintenance; SLA, service-level agreement). Source: Courtesy of Operational Intelligence Ltd.

had the additional investment not been made. Finally, a payback period can be calculated:

$$\text{Payback}\left(\text{years}\right) = \text{cost of compensating provision}\left(\$\right) / \text{risk reduction}\left(\$ / \text{year}\right)$$

$$\text{Payback} = \$2{,}000{,}000 / \left(\$5{,}000{,}000 - \$1{,}000{,}000\right) = 0.5\,\text{years}$$

7.3.2 Topology

The topology of the various data center systems, be it the mechanical and electrical systems or networking systems, can be classified according to the typical arrangement of components contained within them, as shown in Table 7.1. At this stage a client will define a topology based on a desired level of reliability. It is important that there is a business need for this chosen topology—the higher the level, the more expensive the system is and the more complex the system can become. Different IT services may have different availability needs; this can be addressed by providing different topology options within the same facility or even outside of the facility. For example, resilience may be achieved by having multiple data centers.

7.3.3 Site Selection

Any potential data center site will have risks inherent to its location. These need to be identified, and their risk to the business needs to be analyzed along with ways to mitigate it. Locations that could pose a risk include those with terrorist/security threats, on floodplains, in areas of extreme weather such as tornados or typhoons, with other environmental/climate concerns, with poor accessibility (particularly for disaster recovery), in earthquake zones, under a flight path, next to a motorway or railway, with poor connection to the

TABLE 7.1 Different topologies

Tier/ level/ class	Description
1	No plant redundancy (N)
2	Plant redundancy ($N+1$) No system redundancy
3	Concurrently maintainable: System redundancy (active + passive paths) to allow for concurrent maintenance
4	Fault tolerant: System redundancy (active paths) to permit fault tolerance. No single points of failure of any single event (plant/system/control/power failure, or flood, or fire, or explosion, or any other single event)

grid and other utilities, or next to a fireworks (or other highly flammable products) factory [5]. Other risks to consider are those within the facility such as [5] space configuration, impact of plant on the building, ability for future expansion, and emergency provisions, as well as any planning risks such as ecology, noise, and carbon tax/renewable contribution. For each case, the severity and likelihood should be established and compiled in a risk schedule, and the resulting risk weighed up against other business requirements.

Another factor that impacts site selection is latency. Some businesses will locate multiple facilities in close proximity to reduce latency between them. However, facilities located too close can be exposed to the same risks. Another option is to scatter facilities in different locations. These can be live, and performing different workloads, but can also provide mirroring and act as redundancy with the capacity to take on the workload of the other, were the other facility to experience downtime (planned or otherwise). For instance, some companies will have a totally redundant facility ready to come online should their main facility fail. This would be at great cost to the business and would be unlikely to fit the business profile in the majority of cases. However, the cost to the business of a potential failure may outweigh the cost of providing the additional facility.

7.3.4 Establishing a Learning Environment, Knowledge Transfer, and the Skills Shortage

It has already been described how risk stems from the processes and people that interact with a facility and how it can be addressed by organizational (the processes) and operator (the people) learning. In this quadrant, it is important for the business to financially plan for a learning environment once the facility is live. For many businesses, training is considered once a facility is live and funds may not be available. If the link between a lack of learning and risk is understood, then the business case is clear from the start and funds allocated.

Learning is particularly important as the data center industry has a skills shortage. Operatives who are unaware are more likely to contribute toward a failure in the facility. The business needs to decide whether it will hire and fire or spend money and time to address this shortfall in knowledge through training. The skills shortage also means there is high demand and operatives may move to a facility offering bigger financial benefits. This high turnover can pose constant risk to a data center. However, if a learning environment is well established, then the risk associated with a new operative is more likely to be managed over time. If the business does not compare the cost of this training with failure, it can be easy to think there is little point in it when the turnover is so high, and yet this is the very reason why training is so important.

Furthermore, the skill sets of the staff with the most relevant knowledge may not, for example, include the ability to write a 2000-word technical incident report; instead, there should be the forum to facilitate that transfer of knowledge to someone who can. This can only occur in an open environment where the operative feels comfortable discussing the incident.

7.4 KNOWLEDGE TRANSFER 1

If there is no way to transfer knowledge in the data center life cycle, the quadrants of the Kolb cycle become silos with expertise and experience remaining within them. The first contractual boundary in the data center life cycle comes between the business and design phases. At this point the client's brief needs to move into the design quadrant via the OPR, which documents the expected function, use, and operation of the facility [6]. This will include the outcome of considering risk in relation to the topology (reliability) and site selection.

7.5 THEORY: THE DESIGN PHASE

During this phase, the OPR is taken and turned into the Basis of Design (BoD) document that forms the foundations of the developed and technical design that comes later in this quadrant. Note in the United Kingdom this quadrant corresponds with RIBA Stages 2 (Concept Design), 3 (Developed Design), and 4 (Technical Design). The BoD "clearly conveys the assumptions made in developing a design solution that fulfils the intent and criteria in the OPR" [6] and should be updated throughout the design phase (with the OPR as an appendix).

It is important to note here the value the BoD [7] can have throughout the project and beyond. If passed through each future boundary (onto the build and operation phases), it can (if written simply) provide a short, easily accessible overview of the philosophy behind the site design that is updated as the design intent evolves. Later in the Kolb cycle, the information in the BoD provides access to the design intent from which the design and technical specifications are created. These technical specifications contain a lot of information that is not so easily digested. However, by reading the BoD, new operators to site can gain quick access to the basic information on how the systems work and are configured, not something that is so instantly possible from technical specifications. It can also be used to check for any misalignments or inconsistencies in the specifications. For example, the BoD might specify the site be concurrently maintainable, but something in the specifications undermines this. Without the BoD, this discrepancy might go unnoticed and is also true when any future upgrades are completed on-site.

It is important to note that although it would be best practice for this document to pass over each boundary, this rarely happens. Traditionally the information is transferred into the design and the document remains within the design phase.

Bearing in mind reliability and complexity (less complex designs are inherently lower risk, meaning that the requirement on training is reduced), the first step in this phase is to define different M&E and IT designs that fulfill the brief. To minimize risk and ensure the design is robust while fulfilling the business case, the topologies of these different solutions should be analyzed and compared (and a final design chosen) using various methods:

- Single point of failure (SPOF) analysis
- Fault tree analysis (FTA) (reliability block diagrams)
- Failure mode and effect analysis (FMEA) and failure mode and effect criticality analysis (FMECA)

The eventual design must consider the time available for planned maintenance, the acceptable level of unplanned downtime, and its impact on the business while minimizing risk and complexity.

7.5.1 Theoretical Concepts: Availability/Reliability

Availability is the percentage of time a system or piece of equipment is available or ready to use. In Figure 7.4 the solid line denotes a system that is available and working. This is the mean time between failures (MTBF) and is often referred to as uptime. The dashed line denotes an unavailable system that is in failure mode or down for planned maintenance. This is the mean time to repair (MTTR) and is often referred to as downtime.

The availability of the system can be calculated as the ratio of the MTBF to total time:

$$\text{Availability} = \text{MTBF} / (\text{MTBF} + \text{MTTR})$$

If the IT equipment in a facility were unavailable due to failure for 9 hours in a 2-year period, availability would be

$$\text{Availability} = ((2 \times 365 \times 24) - 9) / (2 \times 365 \times 24) = 0.9995$$

The availability is often referred to by the number of 9s, so, for example, this is three 9s. Six 9s (99.9999%) would be better, and two 9s (99%) would be worse. An availability of 99.95% looks deceptively high, but there is no indication of the impact of the failure. This single failure could have cost the business $100,000 or it could have cost $10,000,000. Furthermore, the same availability could be achieved from 9 separate events of 1-hour duration each, and yet each failure could have cost the same as the single event. For example, 9 failures each costing $10,000,000 would result in a total cost of $90,000,000, 9 times that of the single failure event ($10,000,000).

Reliability is therefore used in the design process as it provides a clearer picture. Reliability is the probability that a system will work over time given its MTBF. If, for example, a UPS system has a MTBF of 100 years, it will work, on average (it could fail at any point before or after the MTBF), for 100 years without failure. Reliability is therefore time dependent and can be calculated using the following equation. Note MTBF (which includes the repair time) is almost equal in value to mean time to fail (MTTF), which is used in the case of non-repairable items (such as bearings) [8]. MTTF is the inverse of failure rate (failures/year), and so here the same is assumed for MTBF. It should also be noted that different authors differ in their use of these terms [8–11]:

$$\text{Reliability} = e^{-\text{time}/\text{MTBF}}$$

where e is the base of the natural logarithm that is a mathematical constant approximately equal to 2.71828.

In Figure 7.5 it can be seen that when time is zero, reliability is 100% and as time elapses, reliability goes down.

FIGURE 7.4 Availability. Source: Courtesy of Operational Intelligence Ltd.

The equation can be used to compare different topologies. If redundancy is added ($N+1$), a parallel system is created, and the reliability equation (where R denotes reliability and R_1 and R_2 denote the reliability of systems 1 and 2) becomes [9]

$$\text{Reliability} = 1-(1-R_1)(1-R_2)$$

As the units (equipment) are the same, $R_1 = R_2 = R$, therefore

$$\text{Reliability} = 1-(1-R)^2$$

When plotted in Figure 7.5d, it gives a much higher reliability. Note that Figure 7.5a–c shows the same relationship for MTBFs of 10, 20, and 50 years. Adding redundancy to a system therefore increases the reliability while still reducing over time. However, as time goes by, there will eventually be a failure even though the failure rate or MTBF remains constant, because of the human interface with the system. A facility's ability to restore to its original state (resilience) after a failure is therefore not only related to the reliability of the systems it contains but also related to the people operating it. Although this theoretical modeling can be used to compare different topologies and design options, it cannot model the impact of this human element and is one of the reasons why effective training and knowledge transfer is so important in managing data center risk.

7.5.2 SPOF Analysis

The removal of all SPOFs means that a failure can only occur in the event of two or more simultaneous events. Therefore, a SPOF analysis is used for high-reliability designs where a SPOF-free design is essential in achieving the desired reliability. In other designs, it may be possible to remove certain SPOFs, increasing the reliability without significant additional cost. Many designs will accept SPOFs, but awareness of their existence helps to mitigate the associated risk, for example, it may inform the maintenance strategy. This analysis may also be repeated at the end of the design phase to ensure SPOFs have not been introduced due to design complexities.

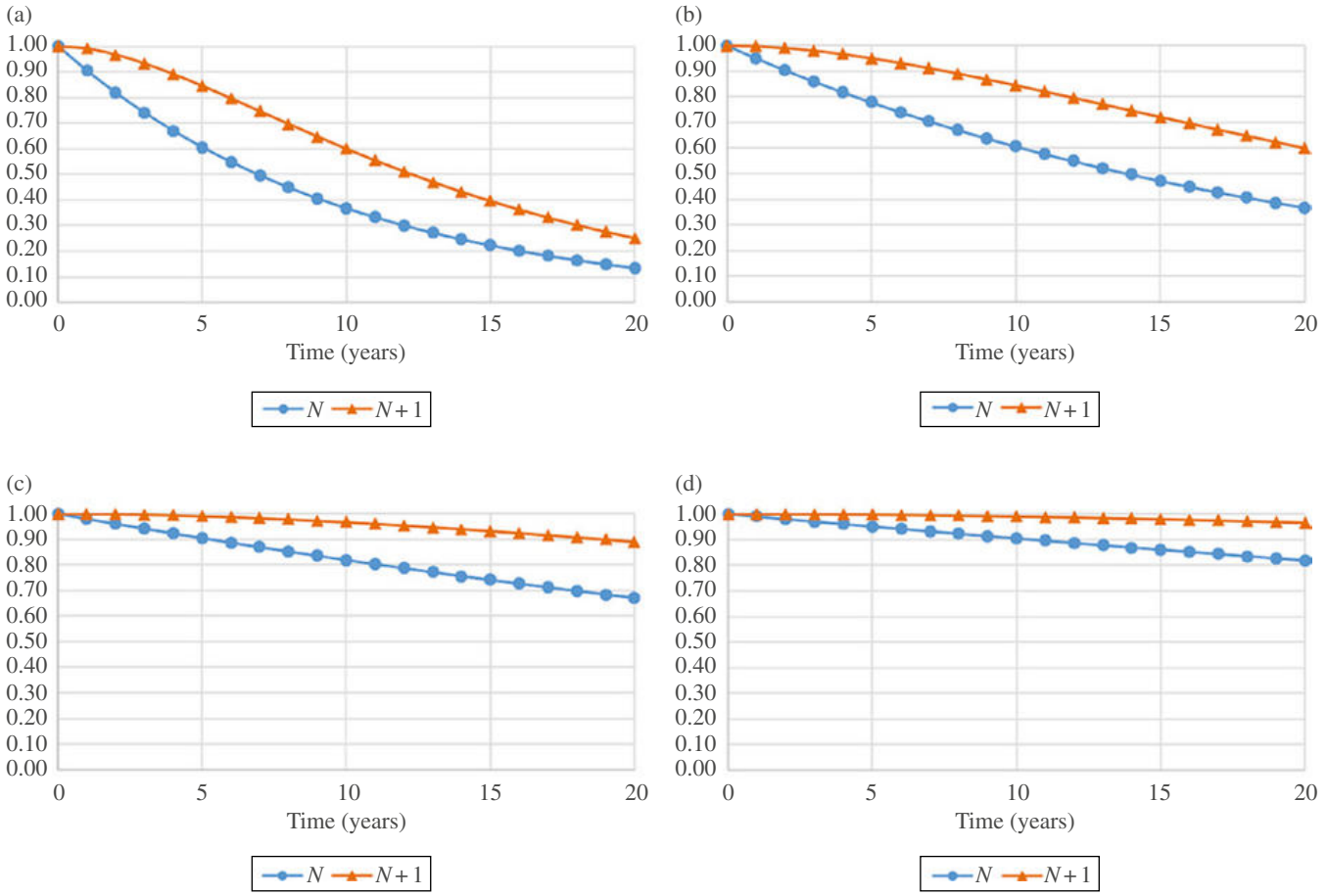


FIGURE 7.5 Reliability vs. time for a UPS system with MTBF of 10, 20, 50, and 100 years. (a) Reliability for MTBF = 10 years, (b) reliability for MTBF = 20 years, (c) reliability for MTBF = 50 years, (d) reliability for MTBF = 100 years. Source: Courtesy of Operational Intelligence Ltd.

7.5.3 Fault Tree Analysis (FTA) and Reliability Block Diagrams

The reliability of a system depends on the reliability of the elements contained within it. Consideration of data center reliability is essential, and the earlier it is considered in the design process, the more opportunity there is to influence the design [9] by minimizing design weaknesses and system vulnerabilities. It also ensures that the desired level of reliability is met and that it is appropriate to the business need, while minimizing costs to the project, and should be considered through all stages of the design.

FTA is a "top-down" method used to analyze complex systems and understand ways in which systems fail and subsystems interact [9]. A component can fail in a number of different ways, resulting in different outcomes or failure modes. In turn these failure modes can impact on other parts of the system or systems. In an FTA a logic diagram is constructed with a failure event at the top. Boolean arguments are used to trace the fault back to a number of potential initial causes via AND and OR gates and various sub-causes. These initial causes can then be removed or managed, and the probabilities combined to determine an overall probability [9, 10]:

$$\text{Probability of } A \text{ AND } B = P_A \times P_B$$

$$\text{Probability of } A \text{ OR } B = P_A + P_B$$

$$\text{Probability of } A \text{ OR } B \text{ or } A \text{ AND } B = P_A + P_B - P_A \times P_B$$

Reliability block diagrams can be used to represent pictorially much of the information in an FTA. An FTA, however, represents the probability of a system failing, whereas reliability block diagrams represent the reliability of a system, or rather the probability of a system not failing or surviving [9]. If the elements of a system are in series, then each element must survive in order for the system not to fail. The probability that the system survives is therefore the product of each element reliability [10]:

$$R_{\text{series}} = R_1 R_2 \ldots R_i \ldots R_m$$

Assuming a constant failure rate (which is adequate for a large number of systems and results in an exponential reliability time distribution [10]), then

$$R_i = \mathrm{e}^{-\lambda_i t}$$

$$R_{\text{series}} = \mathrm{e}^{-\lambda_1 t}\mathrm{e}^{-\lambda_2 t} \ldots \mathrm{e}^{-\lambda_i t} \ldots \mathrm{e}^{-\lambda_m t}$$

$$R_{\text{series}} = \mathrm{e}^{-\lambda_{\text{system}} t} = \mathrm{e}^{-(\lambda_1 + \lambda_2 + \cdots + \lambda_i + \cdots + \lambda_m)t}$$

where

e = 2.71828
λ_i = failure rate (failures/year), and
t = time

If the elements of a system are in parallel (as would be the case in a system with redundancy), then all elements must fail for the system to fail, and reliability (for one redundant unit) would become [8–11]

$$R_{\text{parallel}} = 1 - (1 - R_1)(1 - R_2)$$

As the redundant units will be identical, $R_1 = R_2 = R$; therefore

$$R_{\text{parallel}} = 1 - (1 - R)^2$$

7.5.4 FMEA/FMECA

FMEA is a "bottom-up" design tool used to establish potential failure modes and the effects they have for any given system within the data center. It is used to minimize risk and achieve target hazard rates by designing out vulnerabilities and is used to compare design options. In an FMEA the smallest parts (or elements) of a component (within subassemblies/assemblies/subsystems/systems) are listed, and the system failures that result from their potential failure modes are determined. The effect on each step of the system (subsystem, assembly, subassembly) is listed alongside the likelihood of occurrence [8–11].

An FMECA takes the output of an FMEA and rates each vulnerability according to how critical it is to the continued running of the data center. Vulnerabilities can then be accepted or designed out according to the potential impact they have and the level of risk that is acceptable to the business. A simple example of how an FMECA can be used to compare two (centralized vs. decentralized) cooling options is shown in Table 7.2. Note there are three data halls each with three cooling units and one redundant cooling unit. The risk is calculated by severity/MTTF.

7.5.5 Design Complexity

Although added redundancy improves reliability, a more complex system can undermine this. An FTA will highlight combinations of events that result in failure; however, it is very difficult to model complex designs and the human element as the data used in this modeling will always be subjective and the variables infinite. Reducing complexity therefore helps to manage this aspect of risk. The simpler a system, the more reliable it can be and in turn the less learning that is required to understand and operate it. In short, complex designs require more training to operate them. Therefore, less complex systems can help to manage risk. Before considering system complexity, it is necessary to understand that for a resilient system with no SPOFs, a failure event must be, by definition, the result of two or more simultaneous events. These can be component failures or incorrect human intervention as previously noted.

A 2*N* system could be considered the minimum requirement to achieve a SPOF-free installation. For simplicity, this could contain A and B electrical and A and B

TABLE 7.2 Example of FMECA

Option	Failure event	MTTF (years/ failure)	Impact	Severity[a] (£m/ failure)	Risk (£m/year)
CRACs/DX	Any two of four grouped CRACs	5	1/3 of a data hall	1	0.2
CRAHs/chilled water set	Chilled water system	18	3 data halls	9	0.5

[a] £ = 1.3 USD.

mechanical systems. If the systems are diverse throughout and physically separated in this 2*N* system, then any action on one system should have no impact on the other. However, it is not uncommon for "improvements" to be introduced that take the simple 2*N* system and add disaster recovery links (see Figure 7.6) or common storage vessels, for example, providing an interconnection between the A and B systems. The system is no longer SPOF-free. On large-scale projects, this might be using automatic control systems such as SCADA and BMS, as opposed to simple mechanical interlocks. The basic principles of 2*N* have therefore been compromised, and the complexity of the system has risen exponentially, along with the skills required by the operations teams.

A desktop review would show that a 2*N* design had been achieved; however, the resulting complexity and challenges of operability undermine the fundamental requirement of a high-availability design. Furthermore, the particular sequence of events that leads to a failure is often unforeseen, and until it has occurred, there is no knowledge that it would do so. In other words, these event sequences are unknown until they become known and would not therefore form part of an FTA. The more complex the system, the more of these unknown sequences there are, and the more reliant the system is on comprehensive training.

7.5.6 Commissioning

Commissioning is an important phase that is often rushed and essential to proper management of facility risk. It allows the design to be tested prior to the site going live and ensures that when the installation contractor hands over the facility to the operations teams, the systems work as they were designed to and that knowledge on these systems is transferred. Commissioning therefore reduces the risk of the facility failing once the IT becomes live and runs from the beginning of the next (build) quadrant. Although it does not start until the next phase, initiating planning at this stage can help to manage risk. In particular, the commissioning responsibility matrix [5] should be considered. Among other information, this sets out the key deliverables of the commissioning process and who is responsible for it. This ensures that contractual responsibilities for commissioning are understood as early as possible mitigating risks from arising later where responsibilities are unknown. As the project moves through the design phase, more detail should be added.

Traditionally, a commissioning review will begin at the start of the installation phase. However, it can start earlier, toward the end of the design phase. It is also important during this phase to appoint a commissioning manager. This can minimize the problems associated with different teams

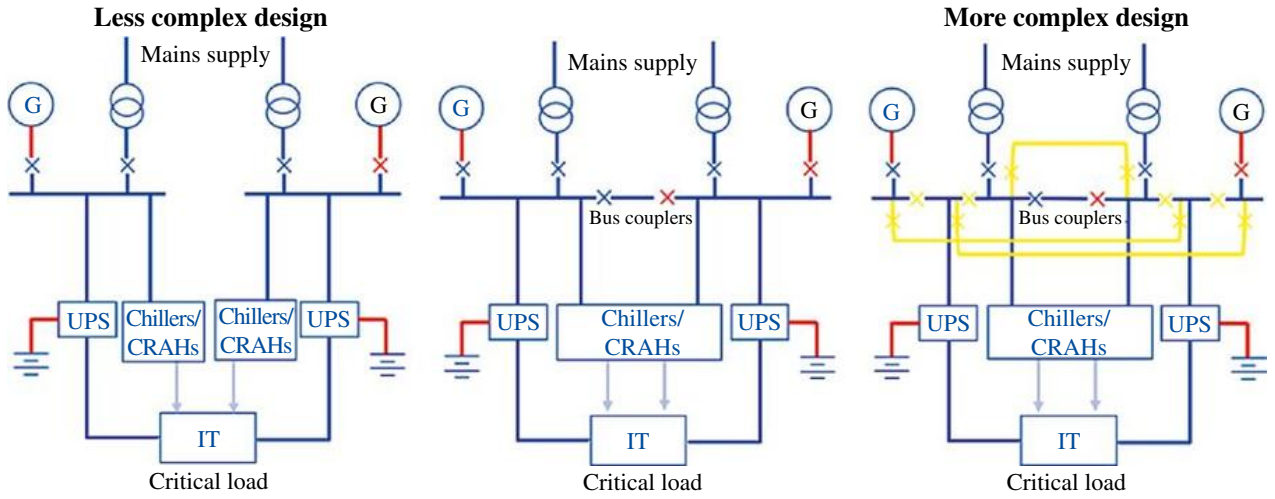


FIGURE 7.6 Design complexity. Source: Courtesy of Operational Intelligence Ltd.

inhabiting each quadrant of the Kolb cycle and facilitate improved knowledge transfer over the boundaries.

7.6 KNOWLEDGE TRANSFER 2

The second contractual boundary occurs between the design and build phases. During the design phase, the content of the BoD is transferred into the data center design, the information of which is passed into the build quadrant via design documents, technical specifications, and drawings. The commissioning specifications should include the details agreed in the commissioning responsibility matrix. It is the most mature of the boundaries, and for this reason it undergoes less scrutiny. Therefore, it is important at this stage that the needs set out in the BoD have been met by the design and that any discrepancies between the design and the brief can be identified. The BoD, with the OPR in an appendix (though not commonplace), should therefore be transferred at this boundary.

7.7 PRACTICE: THE BUILD PHASE

During this phase (RIBA Stage 5—Construction), it is essential that the systems and plant are installed correctly and optimized to work in the way they were designed to. This optimization should consider risk (as well as energy) and is achieved via commissioning.

7.7.1 Commissioning

Commissioning runs alongside installation and is not a single event. The commissioning plan should document this process (key design decisions and logistics) and should evolve with the project. Commissioning, shown in Figure 7.7, starts with a commissioning review—during which the commissioning plan will be started—and follows through the following five levels, the end of which is practical completion (PC) [5]:

- Level 1 (L1): Factory acceptance testing (FAT)/factory witness testing (FWT) of critical infrastructure equipment
- Level 2 (L2): Supplier/subcontractor installation testing of critical infrastructure components
- Level 3 (L3): Full witness and demonstration testing of installation/equipment to client/consultants (plant commissioning/site acceptance testing)
- Level 4 (L4): Testing of interfaces between different systems (i.e. UPS/generators/BMS) to demonstrate functionality of systems and prove design (systems testing)
- Level 5 (L5): Integrated systems testing (IST)

The commissioning plan is an iterative process and should be reviewed and updated on a regular basis as the installation progresses. Some problems will be identified and remedied during this process, meaning some testing might no longer be required, while some additional testing might be required. The commissioning responsibility matrix must also be reviewed to ensure all contractual obligations are met and any late additional requirements are addressed.

L5 or IST is now common on data center projects, but it is still very much the domain of the project delivery team, often with only limited involvement of the operations team. The testing is used to satisfy a contractual requirement and misses the opportunity to impart knowledge from the construction

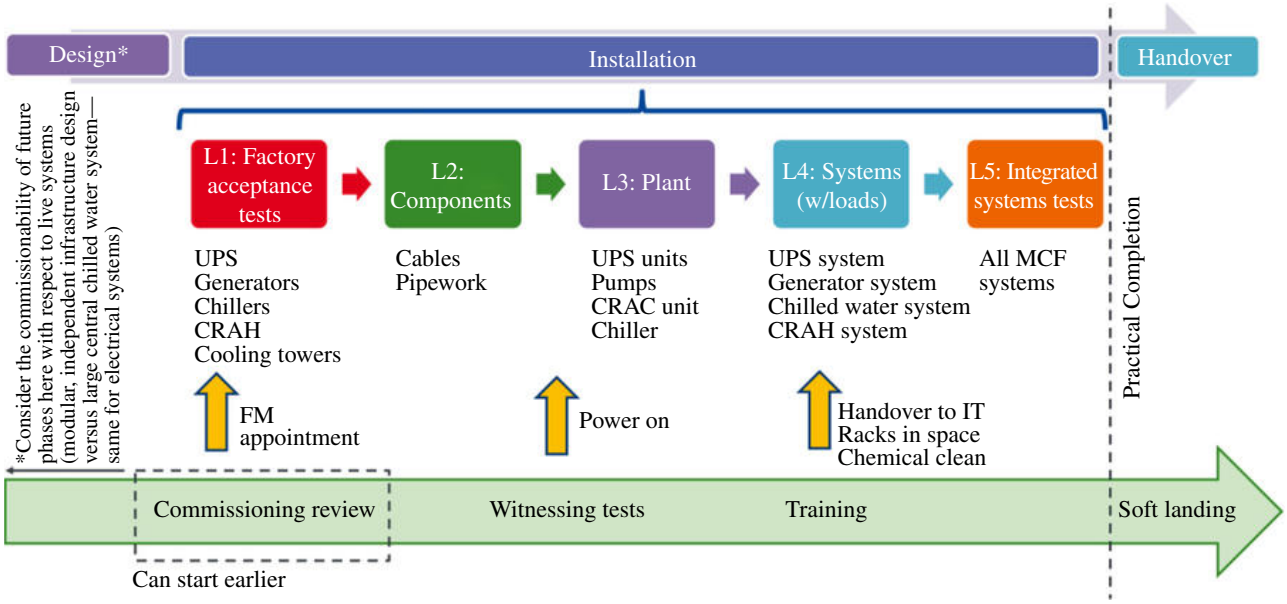


FIGURE 7.7 The commissioning plan. Source: Courtesy of Operational Intelligence Ltd.

phase into the operation phase. In many cases, particularly with legacy data centers, the operations team instead has little or no access to the designer or installation contractor, resulting in a shortfall in the transfer of knowledge to the people who will actually operate the facility. However, risk could be reduced if members of the facilities management (FM) team were appointed and involved in this stage of the commissioning. Instead, operators often take control of a live site feeling insufficiently informed, and over time they can become less engaged, introducing risks due to unawareness.

7.7.2 Additional Testing/Operating Procedures

Operating and response procedures ensure operators understand the systems that have been built and how they operate in emergencies (emergency operating procedures [EOP]) and under normal conditions (standard operating procedures [SOP]) and what steps should be followed in response to alarms (alarm response procedures [ARP]). These procedures are essential to the smooth running of a facility and help to minimize the risk of failure due to incorrect operation. They need to be tested on-site and operators trained in their use.

Relevant test scripts from the commissioning process can form the basis of some of these procedures. The testing of which would therefore be completed by the commissioning engineer if included in their scope. The remaining procedures will be written by the FM team. Traditionally appointment of the FM team would be at the start of the operation phase, and so procedures would be written then. However, appointment of members of the FM team during this phase can ensure continuity across the next contractual boundary and allows for collaboration between the FM and commissioning teams when writing the procedures. At this stage (and the next), FMEA/FMECA can be used to inform the testing.

7.7.3 Maintenance

Once the facility is in operation, regular maintenance is essential to allow continuous operation of the systems with desired performance. Without maintenance, problems that will end in failure go unnoticed. Maintenance information should form the basis of the maintenance manual contained within the O&M manual and should include [5, 12] equipment/system descriptions, description of function, recommended procedures and frequency, recommended spare parts/numbers and location, selection sheets (including vendor and warranty information), and installation and repair information. This information should then be used by the FM team to prepare the maintenance management program once the facility is in operation. As with the commissioning, if members of the FM team are appointed during the build phase, this program can be established in collaboration with the commissioning engineers.

The philosophy adopted for maintenance management is of particular importance for managing risk. This philosophy can be (among others) planned preventative maintenance (PPM), reliability-centered maintenance (RCM), or predictive centered maintenance (PCM). PPM is the bare minimum. It is the cheapest to set up and therefore the most widely adopted. In this approach components (i.e. a filter) are replaced on a regular basis regardless of whether it is needed or not. This approach, however, tends to increase overall total cost of ownership (TCO) because some components will be replaced before they require it and some will fail before replacement, which can result in additional costs beyond the failed component (due to system failure, for example).

In an RCM approach, the reliability of each component is considered, and maintenance provided based on its criticality. For example, a lightbulb in a noncritical area could be left until it blows to be changed; however, a lightbulb over a switchboard would be critical in the event of a failure and therefore checked on a more regular basis than in PPM. PCM could then be applied to these critical components. PCM is the specific monitoring of critical components to highlight problems prior to failure. For example, if the pressure drop across a CRAC unit filter is monitored, the filter can be changed when the pressure exceeds the value displayed by a dirty filter. Or the noise in a critical set of bearings may be monitored via sensors enabling their replacement when a change in noise (associated with a failing bearing) is heard. This type of maintenance is more fine-tuned to what is actually happening, ensuring components are only replaced when needed. It is expensive to set up but reduces overall TCO. Because the RCM and PCM approaches monitor components more closely, they are also likely to reduce the risk of componentry failures.

Interestingly, these latter maintenance philosophies could be considered examples of applying Internet of Things (IoT) and data analytics within the data center. However, it must be remembered that limiting complexity is crucial in managing risk in the data center and adding sensors could undermine this approach.

7.8 KNOWLEDGE TRANSFER 3: PRACTICAL COMPLETION

This boundary coincides with RIBA Stage 6—Handover and Close Out. The handover from installation to operations teams can be the most critical of the boundaries and is the point of PC. If knowledge embedded in the project is not transferred here, the operations teams are left to manage a live critical facility with limited site-specific training and only a set of record documents to support them. Risk at this point can be reduced if there has been an overlap between the commissioning and FM teams so that the transfer is not

solely by documents. The documents that should occur at this boundary include:

- O&M manual [5, 12]: This includes (among other things) information on the installed systems and plant, the commissioning file including commissioning results (levels 1–5) and a close out report for levels 4 and 5, as-commissioned drawings, procedures, and maintenance documents.
- BoD: This ensures the philosophy behind the facility is not lost and is easily accessed by new operatives and during future maintenance, retrofits, and upgrades. This should be contained within the O&M manual.

Knowledge transferring activities that should occur at this boundary include:

- Site/system/plant-specific training: Written material is often provided to allow self-directed learning on the plant, but group training can improve the level of understanding of the operators and provide an environment to share knowledge/expertise and ask questions. The written documentation should be contained within the O&M manual.
- Lessons learned workshop: To manage risk once the site is live, it is imperative that lessons learned during the installation and commissioning are transferred to the next phase and inform the design of future facilities.

7.9 EXPERIENCE: OPERATION

In the final quadrant, the site will now be live. In the United Kingdom this refers to RIBA Stage 7. Post-PC, a soft landings period during which commissioning engineers are available to provide support and troubleshooting helps to minimize risk. The term "soft landings" [13] refers to a mindset in which the risk and responsibility of a project is shared by all the teams involved in the life cycle of a building (from inception through design, build, and operation) and aligns with the content discussed in this chapter. The soft landings period in this quadrant bridges the building performance gap and should be a contractual obligation with a defined duration of time. During this phase, the site is optimized, and any snags (latent defects and incipient faults) that remain after commissioning are rectified. Providing continuity of experience and knowledge transfer beyond the last boundary can help to minimize the risk of failure that can occur once the site is handed over.

7.9.1 Vulnerability Analysis, the Risk Reduction Plan, and Human Error

With the site live, it is now important that the organization and operator vulnerabilities discussed in Section 7.2.1 are identified and a risk reduction plan created. Examples of vulnerabilities and their contribution to failure for each area are shown in Tables 7.3 and 7.4.

TABLE 7.3 Organizational vulnerabilities and their potential contribution to failure

Area	Vulnerability	Contribution to failure
Structure/resources	Technical resources	Unaware of how to deal with a failure
	Insufficient team members	Unable to get to the failure/increased stress
	Management strategy: unclear roles and responsibilities	Unaware of how to deal with a failure, and team actions overlap rather than support
Maintenance	No operating procedures	Plant not maintained
	No predictive techniques (infrared)	Plant fails before planned maintenance
	No client to Facilities Management (FM) Service-Level Agreement (unclear objectives)	Unaware of failed plant criticality
Change management	No tracking of activity progress	Steps are missed, for example, after returning from a break
	Deviations from official procedures	Increased risk
	No timeline/timestamps for tasks	Human error goes undetected
Document management	Drawings not indexed or displayed in M&E rooms	Unable to find information in time
	No SOP/EOP/ARP or not displayed	Misinterpretation of procedures trips the plant
	Reliance on undocumented knowledge of individuals	SPOF—absence leaves those left unsure of what to do

TABLE 7.3 (Continued)

Area	Vulnerability	Contribution to failure
Commissioning (incipient faults and latent defaults)	No mission-critical plant testing and systems commissioning documentation	Accidental system trip
	No IST documented	Unaware the action would trip the system
	Snagging not managed/documented	Failure due to unfinished works
Operability and Maintainability	No emergency backup lights in M&E rooms	Poor visibility to rectify the fault
	No alarm to BMS auto-paging	Unaware of failed plant/system
	Disparity between design intent and operation	Operation in unplanned-for modes
Capacity	Load greater than the redundant capacity	Load not supported in event of downtime
	Growth in load	Overcapacity and/or overredundant capacity
	System upgrade without considering capacity	Overcapacity and/or overredundant capacity
Organization and Operator Learning	No plant training	Unaware of how to deal with a failure
	No systems training	Unaware of how to deal with a failure
	No SOP/EOP/ARP training	Misinterpretation of procedures trips the MCF (mission-critical facilities)

TABLE 7.4 Operator vulnerabilities analysis

Area	Vulnerability	Contribution to failure
Knowledge	No involvement in commissioning	Unaware of how systems work and failure
	Lack of learning environment/training	Unaware of how systems work and failure
	No access to procedures	Unaware of how systems work and failure
Experience	No prior involvement in commissioning	Unaware of how systems work and failure
	No prior experience of failures	Unaware of how to react to a failure
Attitude	Blind repetition of a process	Complacency leading to failure
	Poor communication	Reduced motivation and lack of engagement leading to failure
	Unopen to learning	Unawareness and failure

Traditional risk analyses are not applicable to human error in which data is subjective and variables are infinite. One option (beyond the vulnerabilities analysis above) for human error analysis is TESEO (tecnica empirica stima errori operatori) (empirical technique to estimate operator failure). In TESEO [8] five factors are considered: activity factor, time stress factor, operator qualities, activity anxiety factor, and activity ergonomic (i.e. plant interface) factor. The user determines a level for each factor, and a numerical value (as defined within the method) is assigned. The probability of failure of the activity is determined by the product of these factors. While the method is simplistic, it is coarse and subjective (one person's definition of a "highly trained" operator could be very different to that of another), and so it is difficult to replicate the results between users. Nonetheless it can help operators look at their risk.

7.9.2 Organization and Operator Learning

It has already been established that a learning environment is crucial in the management of data center risk. It recognizes the human contribution toward operational continuity of any

critical environment and the reliance on teams to avoid unplanned downtime and respond effectively to incidents.

Training should not stop after any initial training on the installed plant/systems (through the last boundary); rather, it should be continuous throughout the operational life of the facility and specific to the site and systems installed. It should not consider only the plant operation, but how the mechanical and electrical systems work and their various interfaces. It should also be facility-based and cross-disciplinary, involving all levels of the team from management to operators. This approach helps each team to operate the facility holistically, understanding how each system works and interacts, and promotes communication between the different teams and members. This improved communication can empower individuals and improve operator engagement and staff retention. In this environment, where continuous improvement is respected, knowledge sharing on failures and near misses also becomes smoother, enabling lessons to be learned and risk to be better managed.

Training provides awareness of the unique requirements of the data center environment and should include site maintenance, SOP and EOP and ARP, site policies and optimization, inductions, and information on system upgrades.

7.9.3 Further Risk Analyses

Further risk analyses might be completed at this stage. Data centers undergo upgrades, expansion, maintenance, and changes, and in particular on sites where data halls have been added to existing buildings, the operations team might lose clarity on how the site is working, and complexities may have crept in. At this point it is important to run additional FMEA/FME(C)A to ensure risk continues to be managed in the new environment. It is also important that any changes made to the facility as a result are documented in the O&M manual and (where required) additional training is provided to the operators.

In the event of a failure, root cause analysis (RCA) may be used to learn from the event. In an RCA, three categories of vulnerabilities are considered—the physical design (material failure), the human element (something was/was not done), and the processes (system, process, or policy shortcomings)—and the combination of these factors that led to the failure determined. Note that with complex systems there are usually a number of root causes. It can then be used to improve any hidden flaws and contributing factors. RCA can be a very powerful tool, and when used in an environment that is open to learning from failures (rather than apportioning blame), it can provide clear information on the primary drivers of the failure, which can be shared throughout the business ensuring the same incident does not happen again. It also enables appropriate improvements to the design, training, or processes that contributed to the event and supports a culture of continuous improvement of the facility and operators.

7.10 KNOWLEDGE TRANSFER 4

This is the final contractual boundary, where knowledge and information are fed back to the client. This includes service-level agreements (SLAs), reports, and lessons learned from the project. It is rare at the end of projects for any consideration to be made to the lessons that can be learned from the delivery process and end product. However, from the experience of the authors, the overriding message that has come from the few that they have participated in is the need for better communication to ensure awareness of what each team is trying to achieve. Indeed, this reinforces the approach suggested in this chapter for managing data center risk and in particular the need for improved channels of communication and to address what lessons can be learned throughout the whole process. By improving the project communication, the lessons that can be learned from the process could move on beyond this topic and provide valuable technical feedback (good and bad) to better inform future projects. This boundary also needs to support continuous transformation of the facility and its operation in response to changing business needs.

7.11 CONCLUSIONS

To manage risk in the data center, attention must be paid to identifying risks, and reducing design complexity and human unawareness through knowledge transfer and training.

In such a complex process, it is almost impossible to guarantee every procedure addresses all eventualities. In the event of an incident, it is imperative that the team has the best chance of responding effectively. It is well established that the human interface is the biggest risk in the data center environment and relying on an individual to do the right thing at the right time without any investment in site-specific training is likely to result in more failures and increased downtime. As an industry, more effort should be made to improve the process of knowledge sharing throughout the project lifetime and in particular at project handover on completion of a facility to ensure lessons can be learned from the experience. What is more, this should extend beyond the confines of the business to the industry as a whole—the more near misses and failures that are shared and learned from, the more the industry has to gain. They represent an opportunity to learn and should be embraced rather than dealt with and then brushed aside.

Once a facility is in operation, continuous site-specific training of staff will increase knowledge and identify unknown failure combinations, both of which can reduce the number of unknown failure combinations and resulting downtime. Finally, reducing complexity not only reduces the number of unknown sequences of events that cause a failure but also reduces the amount of training required.

REFERENCES

[1] Duffey RB, Saull JW. *Managing Risk: The Human Element*. Wiley; 2008.

[2] Onag G. 2016. Uptime institute: 70% of DC outages due to human error. Computer World HK. Available at https://www.cw.com.hk/it-hk/uptime-institute-70-dc-outages-due-to-human-error. Accessed on October 18, 2018.

[3] Uptime Institute. Data center site infrastructure. Tier standard: operational sustainability; 2014.

[4] Kolb DA. *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice Hall; 1984.

[5] CIBSE. *Data Centres: An Introduction to Concepts and Design*. CIBSE Knowledge Series. London: CIBSE; 2012.

[6] ASHRAE. *ASHRAE Guideline 0-2013. The Commissioning Process*. ASHRAE; 2013.

[7] Briones V, McFarlane D. Technical vs. process commissioning. Basis of design. ASHRAE J 2013;55:76–81.

[8] Smith D. *Reliability, Maintainability and Risk. Practical Methods for Engineers*. 8th ed. Boston: Butterworth Heinemann; 2011.

[9] Leitch RD. *Reliability Analysis for Engineers. An Introduction*. 1st ed. Oxford: Oxford University Press; 1995.

[10] Bentley JP. *Reliability and Quality Engineering*. 2nd ed. Boston: Addison Wesley; 1998. Available at https://www.amazon.co.uk/Introduction-Reliability-Quality-Engineering-Publisher/dp/B00SLTZUTI.

[11] Davidson J. *The Reliability of Mechanical Systems*. Oxford: Wiley-Blackwell; 1994.

[12] ASHRAE. *ASHRAE Guideline 4-2008. Preparation of Operating and Maintenance Documentation for Building Systems*. Atlanta, GA: ASHRAE; 2008.

[13] BSRIA. *BG 54/2018. Soft Landings Framework 2018. Six Phases for Better Buildings*. Bracknell: BSRIA; 2018.

PART II

DATA CENTER TECHNOLOGIES

8

SOFTWARE-DEFINED ENVIRONMENTS

Chung-Sheng Li[1] and Hubertus Franke[2]
[1] *PwC, San Jose, California, United States of America*
[2] *IBM, Yorktown Heights, New York, United States of America*

8.1 INTRODUCTION

The worldwide public cloud services market, which includes business process as a service, software as a service, platform as a service, and infrastructure as a service, is projected to grow 17.5% in 2019 to total $214.3 billion, up from $182.4 billion in 2018, and is projected to grow to $331.2 billion by 2022.[1] The hybrid cloud market, which often includes simultaneous deployment of on premise and public cloud services, is expected to grow from $44.60 billion in 2018 to $97.64 billion by 2023, at a compound annual growth rate (CAGR) of 17.0% during the forecast period.[2] Most enterprises are taking *cloud first* or *cloud only* strategy and are migrating both of their mission-critical and performance-sensitive workloads to either public or hybrid cloud deployment models. Furthermore, the convergence of mobile, social, analytics, and artificial intelligence workloads on the cloud definitely gave strong indications shift of the value proposition of cloud computing from cost reduction to simultaneous efficiency, agility, and resilience.

Simultaneous requirements on agility, efficiency, and resilience impose potentially conflicting design objectives for the computing infrastructures. While cost reduction largely focused on the virtualization of infrastructure (IaaS, or infrastructure as a service), agility focuses on the ability to rapidly react to changes in the cloud environment and workload requirements. Resilience focuses on minimizing the risk of failure in an unpredictable environment and providing maximal availability. This requires a high degree of automation and programmability of the infrastructure itself. Hence, this shift led to the recent disruptive trend of software-defined computing for which the entire system infrastructure—compute, storage, and network—is becoming software defined and dynamically programmable. As a result, software-defined computing receives considerable focus across academia [1, 2] and every major infrastructure company in the computing industry [3–12].

Software-defined computing originated from the compute environment in which the computing resources are virtualized and managed as virtual machines [13–16]. This enabled mobility and higher resource utilization as several virtual machines are colocated on the same server, and variable resource requirements can be mitigated by being shared among the virtual machines. Software-defined networks (SDNs) move the network control and management planes (functions) away from the hardware packet switches and routers to the server for improved programmability, efficiency, extensibility, and security [17–21]. Software-defined storage (SDS), similarly, separates the control and management planes from the data plane of a storage system and dynamically leverages heterogeneous storage to respond to changing workload demands [22, 23]. Software-defined environments (SDEs) bring together software-defined compute, network, and storage and unify the control and management planes from each individual software-defined component.

The SDE concept was first coined at IBM Research during 2012 [24] and was cited in the 2012 IBM Annual report [25] at the beginning of 2013. In SDE, the unified control planes are assembled from programmable resource

[1] https://www.gartner.com/en/newsroom/press-releases/2019-04-02-gartner-forecasts-worldwide-public-cloud-revenue-to-g
[2] https://www.marketwatch.com/press-release/hybrid-cloud-market-2019-global-size-applications-industry-share-development-status-and-regional-trends-by-forecast-to-2023-2019-07-12

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

abstractions of the compute, network, and storage resources of a system (also known as *fit-for-purpose* systems or workload-optimized systems) that meet the specific requirements of individual workloads and enable dynamic optimization in response to changing business requirements. For example, a workload can specify the abstracted compute and storage resources of its various workload components and their operational requirements (e.g. I/O [input/output] operations per second) and how these components are interconnected via an abstract wiring that will have to be realized using the programmable network. The decoupling of the control/management plane from the data/compute plane and virtualization of available compute, storage, and networking resources also lead to the possibility of resource pooling at the physical layer known as disaggregated or composable systems and datacenter [26–28].

In this chapter, we provide an overview of the vision, architecture, and current incarnation of SDEs within industry, as shown in Figure 8.1. At the top, workload abstractions and related tools provide the means to construct workloads and services based on preexisting patterns and to capture the functional and nonfunctional requirements of the workloads. At the bottom, heterogeneous compute, storage, and networking resources are pooled based on their capabilities, potentially using composable system concept. The workloads and their contexts are then mapped to the best-suited resources. The unified control plane dynamically constructs, configures, continuously optimizes, and proactively orchestrates the mapping between the workload and the resources based on the desired outcome specified by the workload and the operational conditions of the cloud environment. We also demonstrate at a high level how this architecture achieves agility, efficiency, and continuous outcome optimized infrastructure with proactive resiliency and security.

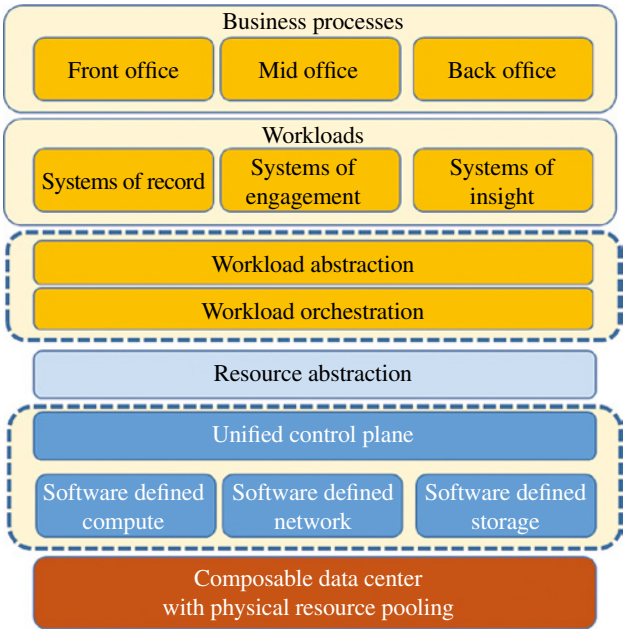


FIGURE 8.1 Architecture of software-defined environments. Workloads are complex wirings of components and are represented through abstractions. Given a set of abstract resources the workloads are continuously mapped (orchestrated) into the environment through the unified control plane. The individual resource controllers program the underlying virtual resources (compute, network, and storage). Source: © 2020 Chung-Sheng Li.

8.2 SOFTWARE-DEFINED ENVIRONMENTS ARCHITECTURE

Traditional virtualization and cloud solutions only allow basic abstraction of the computing, storage, and network resources in terms of their capacity [29]. These approaches often call for standardization of the underlying system architecture to simplify the abstraction of these resources. The convenience offered by the elasticity for scaling the provisioned resources based on the workload requirements, however, is often achieved at the expense of overlooking capability differences inherent in these resources. Capability differences in the computing domain could be:

- Differences in the instruction set architecture (ISA), e.g. Intel x86 versus ARM versus IBM POWER architectures
- Different implementations of the same ISA, e.g. Xeon by Intel versus EPYC by AMD
- Different generations of the same ISA by the same vendor, e.g. POWER7 versus POWER8 versus POWER9 from IBM and Nehalem versus Westmere versus Sandy Bridge versus Ivy Bridge versus Coffee Lake from Intel.
- Availability of various on-chip or off-chip accelerators including graphics processing units (GPUs) such as those from Nvidia, Tensor Processing Unit (TPU) from Google, and other accelerators such as those based on FPGA or ASIC for encryption, compression, extensible markup language (XML) acceleration, machine learning, deep learning, or other scalar/vector functions.

The workload-optimized system approaches often call for tight integration of the workload with the tuning of the underlying system architecture for the specific workload. The *fit-for-purpose* approaches tightly couple the special capabilities offered by each micro-architecture and by the system level capabilities at the expense of potentially labor-intensive tuning required. These workload-optimized approaches are not sustainable in an environment where the workload might be unpredictable or evolve rapidly as a

result of growth of the user population or the continuous changing usage patterns.

The conundrum created by these conflicting requirements in terms of standardized infrastructure vs. workload-optimized infrastructure is further exacerbated by the increasing demand for agility and efficiency as more enterprise applications from systems of record, systems of engagement, and systems of insight require fast deployment while continuously being optimized based on the available resources and unpredictable usage patterns. *Systems of record* usually refer to enterprise resource planning (ERP) or operational database systems that conduct online transaction processing (OLTP). *Systems of engagement* usually focused on engagement with large set of end users, including those applications supporting collaboration, mobile, and social computing. *Systems of insight* often refer to online analytic processing (OLAP), data warehouse, business intelligence, predictive/prescriptive analytics, and artificial intelligence solutions and applications. Emerging applications including chatbot, natural language processing, knowledge representation and reasoning, speech recognition/synthesis, computer vision and machine learning/deep learning all fall into this category.

Systems of records, engagement, and insight can be mapped to one of the enterprise applications areas:

- *Front office*: Including most of the customer facing functions such as corporate web portal, sales and marketing, trading desk, and customer and employee support,
- *Mid office:* Including most of the risk management, and compliance areas,
- *Back office:* The engine room of the corporation and often includes corporate finance, legal, HR, procurement, and supply chain.

Systems of engagement, insight, and records are deployed into front, mid, and back office application areas, respectively. Emerging applications such as chatbot based on artificial intelligence and KYC (know your customer) banking solutions based on advanced analytics, however, blurred the line among front, mid, and back offices. Chatbot, whether it is based on Google DialogFlow, IBM Watson Conversation, Amazon Lex, or Microsoft Azure Luis, is now widely deployed for customer support in the front office and HR & procurement in the back office area. KYC solutions, primarily deployed in front office, often leverage customer data from back office to develop comprehensive customer profiling and are also connected to most of the major compliance areas including anti-money laundering (AML) and Foreign Account Tax Compliance Act (FATCA) in the mid office area.

It was observed in [30] that a fundamental change in the axis of IT innovation happened around 2000. Prior to 2000, new systems were introduced at the very high end of the economic spectrum (large public agencies and Fortune 500 companies). These innovations trickled down to smaller businesses, then to home office applications, and finally to consumers, students, and even children. This innovation flow reversed after 2000 and often started with the consumers, students, and children leading the way, especially due to the proliferation of mobile devices. These innovations are then adopted by nimble small-to-medium-size businesses. Larger institutions are often the last to embrace these innovations. The author of [30] coined the term systems of engagement for the new kinds of systems that are more focused on engagement with the large set of end users in the consumer space. Many systems of engagement such as Facebook, Twitter, Netflix, Instagram, Snap, and many others are born on the cloud using public cloud services from Amazon Web Services (AWS), Google Cloud Platform (GCP), Microsoft Azure, etc. These systems of engagement often follow the agility trajectory. On the other hand, the workload-optimized system concept is introduced to the systems of record environment, which occurred with the rise of client–server ERP systems on top of the Internet. Here the entire system, from top to bottom, is tuned for the database or data warehouse environment.

SDEs intended to address the challenge created from the desire for simultaneous agility, efficiency, and resilience. SDEs decouple the abstraction of resources from the real resources and only focus on the salient capabilities of the resources that really matter for the desired performance of the workload. SDEs also establish the workload definition and decouple this definition from the actual workloads so that the matching between the workload characteristics and the capabilities of the resources can be done efficiently and continuously. Simultaneous abstraction of both resources and workloads to enable late binding and flexible coupling among workload definitions, workload runtime, and available resources is fundamental to addressing the challenge created by the desire for both agility and optimization in deploying workloads while maintaining nearly maximal utilization of available resources.

8.3 SOFTWARE-DEFINED ENVIRONMENTS FRAMEWORK

8.3.1 Policy-Based and Goal-Based Workload Abstraction

Workloads are generated by the execution of business processes and activities involving systems of record, systems of engagement, and systems of insight applications and solutions within an enterprise. Using the order-to-cash (OTC) process—a common corporate finance function as an example—the business process involves (i) generating quote

after receiving the RFQ/RFP or after receiving a sales order, (ii) recording trade agreement or contract, (iii) receiving purchase order from client, (iv) preparing and shipping the order, (v) invoicing the client, (vi) recording invoice on account receivable within general ledger, (vii) receiving and allocating customer payment against account receivable, (viii) processing customer return as needed, and (ix) conducting collection on those delinquent invoices. Most of these steps within the OTC process can be automated through, for example, robotic process automation (RPA) [31]. The business process or workflow is often captured by an automation script within the RPA environment, where the script is executed by the orchestration engine of the RPA environment. This script will either invoke through direct API call or perform screen scraping of a VDI client (such as Citrix client) of those systems of records (the ERP system) that store and track the sales order, trade agreement, and purchase order, invoice, and account receivable; systems of engagement (email or SMS) for sending invoice and payment reminders; and systems of insight such as prediction of which invoices are likely to encounter challenges in collection. The execution of these applications in turn contributes to the workloads that need to be orchestrated within the infrastructure.

Executing workloads involves mapping and scheduling the tasks that need to be performed, as specified by the workload definition, to the available compute, storage, and networking resources. In order to optimize the mapping and scheduling, workload modeling is often used to achieve evenly distributed manageable workloads, to avoid overload, and to satisfy service level objectives.

The workload definition has been previously and extensively studied in the context of the Object Management Group (OMG) Model-Driven Architecture (MDA) initiative during the late 1990s as an approach to system specification and interoperability based on the use of formal models. In MDA, platform-independent models are described in a platform-independent modeling language such as Unified Modeling Language (UML). The platform-independent model is then translated into a platform-specific model by mapping the platform independent models to implementation languages such as Java, XML, SOAP (Simple Object Access Protocol), or various dynamic scripting languages such as Python using formal rules.

Workload concepts were heavily used in the grid computing era, for example, IBM Spectrum Symphony, for defining and specifying tasks and resources and predicting and optimizing the resources for the tasks in order to achieve optimal performance. IBM Enterprise Workload Manager (eWLM) allows the user to monitor application-level transactions and operating system processes, allows the user to define specific performance goals with respect to specific work, and allows adjusting the processing power among partitions in a partition workload group to ensure that performance goals are met. More recently, workload automation and development for deployment have received considerable interests as the development and operations (DevOps) concept becomes widely deployed. These workload automation environments often include programmable infrastructures that describe the available resources and characterization of the workloads (topology, service-level agreements, and various functional and nonfunctional requirements). Examples of such environments include Amazon Cloud Formation, Oracle Virtual Assembly Builder, and VMware vFabric.

A workload, in the context of SDEs, is often composed of a complex wiring of services, applications, middleware components, management agents, and distributed data stores. Correct execution of a workload requires that these elements be wired and mapped to appropriate logical infrastructure according to workload-specific policies and goals. Workload experts create workload definitions for specific workloads, which codify the best practices for deploying and managing the workloads. The workload abstraction specifies all of the workload components including services, applications, middleware components, management agents, and data. It also specifies the relationships among components and policies/goals defining how the workload should be managed and orchestrated. These policies represent examples of workload context embedded in a workload definition. They are derived based on expert knowledge of a specific workload or are learned in the course of running the workload in SDE. These policies may include requirements on continuous availability, minimum throughput, maximum latency, automatic load balancing, automatic migration, and auto-scaling in order to satisfy the service-level objectives. These contexts for the execution of the workload need to be incorporated during the translation from workload definition to an optimal infrastructure pattern that satisfies as many of the policies, constraints, and goals that are pertinent to this workload as possible.

In the OTC business process example, the ERP system (which serves as the systems of record) will need to have very high availability and low latency to be able to sustain high transaction throughput needed to support mission-critical functions such as sales order capturing, shipping, invoicing, account receivable, and general ledger. In contrast, the email server (which is part of the systems of engagement) still need high availability but can tolerate lower throughput and higher latency. The analytic engine (which is part of the systems of insight) might not need to have high availability nor high throughput.

8.3.2 Capability-Based Resource Abstraction and Software-Defined Infrastructure

The abstraction of resources is based on the capabilities of these resources. Capability-based pooling of heterogeneous resources requires classification of these resources based on

workload characteristics. Using compute as an example, server design is often based on the thread speed, thread count, and effective cache/thread. The fitness of the compute resources (servers in this case) for the workload can then be measured by the serial fitness (in terms of thread speed), the parallel fitness (in terms of thread count), and the data fitness (in terms of cache/thread).

Capability-based resource abstraction is an important step toward decoupling heterogeneous resources provisioning from the workload specification. Traditional resource provisioning is mostly based on capacity, and hence the differences in characteristics of the resource are often ignored. The Pfister framework [32] has been used to describe workload characteristics [1] in a two-dimensional space where one axis describes the amount of thread contention and the other axis describes the amount of data contention. We can categorize the workload into four categories based on the Pfister framework: Type 1 (mixed workload updating shared data or queues), Type 2 (highly threaded applications, including WebSphere* applications), Type 3 (parallel data structures with analytics, including Big Data, Hadoop, etc.), and Type 4 (small discrete applications, such as Web 2.0 apps).

Servers are usually optimized to one of the corners of this two-dimensional space, but not all four corners. For instance, the IBM System z [33] is best known for its single-thread performance, while IBM Blue Gene [34] is best known for its ability to carry many parallel threads. Some of the systems (IBM System x3950 [Intel based] and IBM POWER 575) were designed to have better I/O capabilities. Eventually there is not a single server that can fit all workloads described above while delivering required performance by the workloads.

This leads to a very important observation: the majority of workloads (whether they are systems of record or systems of engagement or systems of insight) always consist of multiple workload types and are best addressed by a combination of heterogeneous servers rather than homogeneous servers.

We envision resource abstractions based on different computing capabilities that are pertinent to the subsequent workload deployments. These capabilities could include high memory bandwidth resources, high single thread performance resources, high I/O throughout resources, high cache/thread resources, and resources with strong graphics capabilities. Capability-based resource abstraction eliminates the dependency on specific instruction-set architectures (e.g. Intel x86 versus IBM POWER versus ARM) while focusing on the true capability differences (AMD Epyc versus Intel Xeon, and IBM POWER8 versus POWER9 may be represented as different capabilities).

Previously, it was reported [35] that up to 70% throughput improvement can be achieved through careful selection of the resources (AMD Opteron versus Intel Xeon) to run Google's workloads (content analytics, Big Table, and web search) in its heterogeneous warehouse scale computer center. Likewise, storage resources can be abstracted beyond the capacity and block versus file versus objects. Additional characteristics of storage such as high I/O throughput, high resiliency, and low latency can all be brought to the surface as part of storage abstraction. Networking resources can be abstracted beyond the basic connectivity and bandwidth. Additional characteristics of networking such as latency, resiliency, and support for remote direct memory access (RDMA) can be brought to the surface as part of the networking abstraction.

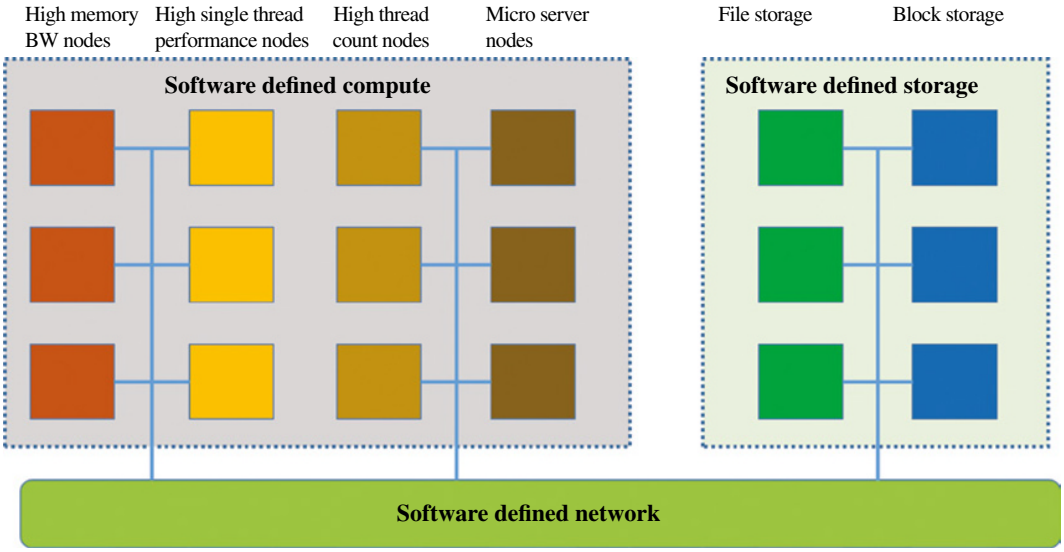


FIGURE 8.2 Capability-based resource abstraction. Source: © 2020 Chung-Sheng Li.

The combination of capability-based resource abstraction for software-defined compute, storage, and networking forms the software-defined infrastructure, as shown in Figure 8.2. This is essentially an abstract view of the available compute and storage resources interconnected by the networking resources. This abstract view of the resources includes the pooling of resources with similar capabilities (for compute and storage), connectivity among these resources (within one hop or multiple hops), and additional functional or nonfunctional capabilities attached to the connectivity (load balancing, firewall, security, etc.). Additional physical characteristics of the datacenter are often captured in the resource abstraction model as well. These characteristics include clustering (for nodes and storage sharing the same top-of-the-rack switches and that can be reached within one hop), point of delivery (POD) (for nodes and storage area network (SAN)-attached storage sharing the same aggregation switch and can be reached within four hops), availability zones (for nodes sharing the same uninterrupted power supply (UPS) and A/C), and physical data center (for nodes that might be subject to the same natural or man-made disasters). These characteristics are often needed during the process of matching workload requirements to available resources in order to address various performance, throughput, and resiliency requirements.

8.3.3 Continuous Optimization

As a business increasingly relies on the availability and efficiency of its IT infrastructure, linking the business operations to the agility and performance of the deployment and continuous operation of IT becomes crucial for the overall business optimization. SDEs provide an overall framework for directly linking the business operation to the underlying IT as described below. Each business operation can be decomposed into multiple tasks, each of which has a priority. Each task has a set of key performance indicators (KPIs), which could include confidentiality, integrity, availability, correctness/precision, quality of service (QoS) (latency, throughput, etc.), and potentially other KPIs.

As an example, a procure-to-pay (PTP) business operation might include the following tasks: (i) send out request for quote (RFQ) or request for proposal (RFP); (ii) evaluate and select one of the proposals or bids to issue purchase order based on the past performance, company financial health, and competitiveness of the product in the marketplace; (iii) take delivery of the product (or services); (iv) receive the invoice for the goods or services rendered; (v) perform three-way matching among purchase order, invoice, and goods received; (vi) issue payment based on the payment policy and deadline of the invoice. Each of these tasks may be measured by different KPIs: the KPI for the task of sending out RFP/RFQ or PO might focus on availability, while the KPI for the task of performing three-way matching and issue payment might focus on integrity. The specification of the task decomposition of a business operation, the priority of each task, and KPIs for each task allow trade-offs being made among these tasks when necessary. Using RFP/RFQ as an example, availability might have to be reduced when there is insufficient capacity until the capacity is increased or the load is reduced.

The KPIs for the task often are translated to the architecture and KPIs for the infrastructure. Confidentiality usually translates to required isolation for the infrastructure. Availability potentially translates into redundant instantiation of the runtime for each task using active–active or active–passive configurations—and may need to take advantage of the underlying availability zones provided by all major cloud service providers. Integrity of transactions, data, processes, and policies is managed at the application level, while the integrity of the executables and virtual machine images is managed at the infrastructure level. Correctness and precision need to be managed at the application level, and QoS (latency, throughput, etc.) usually translates directly to the implications for infrastructures. Continuous optimization of the business operation is performed to ensure optimal business operation during both normal time (best utilization of the available resources) and abnormal time (ensures the business operation continues in spite of potential system outages). This potentially requires trade-offs among KPIs in order to ensure the overall business performance does not drop to zero due to outages. The overall closed-loop framework for continuous optimizing is as follows:

- The KPIs of the service are continuously monitored and evaluated at each layer (the application layer and the infrastructure layer) so that the overall utility function (value of the business operation, cost of resource, and risk to potential failures) can be continuously evaluated based on the probabilities of success and failure. Deep introspection, i.e. a detailed understanding of resource usage and resource interactions, within each layer is used to facilitate the monitoring. The data is fed into the behavior models for the SDE (which includes the workload, the data (usage patterns), the infrastructure, and the people and processes).
- When triggering events occur, what-if scenarios for deploying different amount of resources against each task will be evaluated to determine whether KPIs can be potentially improved.
- The scenario that maximizes the overall utility function is selected, and the orchestration engine will orchestrate the SDE through the following: (i) adjustment to resource provisioning (scale up or down), (ii) quarantine of the resources (in various resiliency and security scenarios), (iii) task/workload migration, and (iv) server rejuvenation.

8.4 CONTINUOUS ASSURANCE ON RESILIENCY

The resiliency of a service is often measured by the availability of this service in spite of hardware failures, software defects, human errors, and malicious cybersecurity threats. The overall framework on continuous assurance of resiliency is directly related to the continual optimization of the services performed within the SDEs, taking into account the value created by the delivery of service, subtracting the cost for delivering the service and the cost associated with a potential failure due to unavailability of the service (weighted by the probability of such failure). This framework enables proper calibration of the value at risk for any given service so that the overall metric will be *risk-adjusted cost performance*. Continuous assurance on resiliency, as shown in Figure 8.3, ensures that the value at risk (VAR) is always optimal while maintaining the risk of service unavailability due to service failures and cybersecurity threats below the threshold defined by the service level agreement (SLA).

Increased virtualization, agility and resource heterogeneity within SDE on one hand improves the flexibility for providing resilience assurance and on the other hand also introduces new challenges, especially in the security area:

- *Increased virtualization obfuscates monitoring*: Traditional security architectures are often physically based, as IT security relies on the identities of the machine and the network. This model is less effective when there are multiple layers of virtualization and abstractions, which could result in many virtual systems being created within the same physical system or multiple physical systems virtualized into a single virtual system. This challenge is further compounded by the use of dedicated or virtual appliances in the computing environment.
- *Dynamic binding complicates accountability*: SDEs enable standing up and tearing down computing, storage, and networking resources quickly as the entire computing environment becomes programmable and breaks the long-term association between security policies and the underlying hardware and software environment. The SDE environment requires the ability to quickly set up and continuously evolve security policies directly related to users, workloads, and the software-defined infrastructure. There are no permanent associations (or bindings) between the logical resources and physical resources as software-defined systems can be continuously created from scratch and can be continuously evolved and destroyed at the end. As a result, the challenge will be to provide a low-overhead approach for capturing the provenance (who has done what, at what time, to whom, in what context), to identify the suspicious events in a rapidly changing virtual topology.
- *Resource abstraction mask vulnerability*: In order to accommodate heterogeneous compute, storage, and network resources in an SDE, resources are abstracted in terms of capability and capacity. This normalization of the capability across multiple types of resources masks the potential differences in various nonfunctional aspects such as the vulnerabilities to outages and security risk.

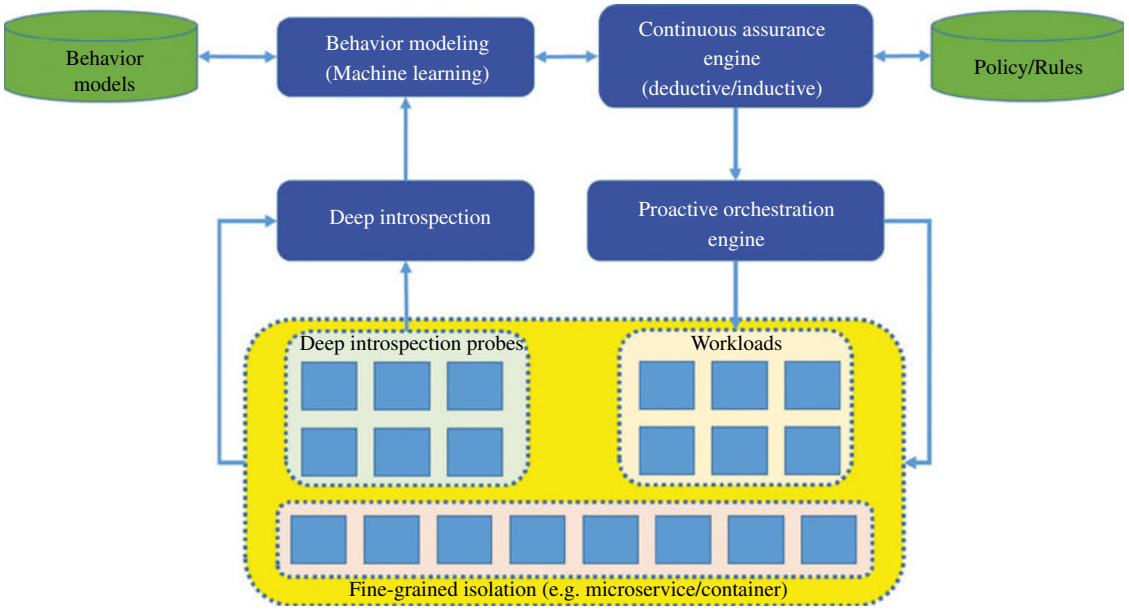


FIGURE 8.3 Continuous assurance for resiliency and security helps enable continuous deep introspection, advanced early warning, and proactive quarantine and orchestration for software-defined environments. Source: © 2020 Chung-Sheng Li.

To ensure continuous assurance and address the challenges mentioned above, the continuous assurance framework within SDEs includes the following design considerations:

- *Fine-grained isolation*: By leveraging the fine-grained virtualization environments such as those provided by the microservice and Docker container framework, it is possible to minimize the potential interference between microservices within different containers so that the failure of one microservice in a Docker container will not propagate to the other containers. Meanwhile, fine-grained isolation is feasible to contain the cybersecurity breach or penetration within a container while maintaining the continuous availability of other containers and maximize the resilience of the services.
- *Deep introspection*: Works with probes (often in the form of agents) inserted into the governed system to collect additional information that cannot be easily obtained simply by observing network traffic. These probes could be inserted into the hardware, hypervisors, guest virtual machines, middleware, or applications. Additional approaches include micro-checkpoints and periodic snapshots of the virtual machine or container images when they are active. The key challenge is to avoid introducing unnecessary overhead while providing comprehensive capabilities for monitoring and rollback when abnormal behaviors are found.
- *Behavior modeling*: The data collected from deep introspection are assimilated with user, system, workload, threat, and business behavior models. Known causalities among these behavior models allow early detection of unusual behaviors. Being able to provide early warning of these abnormal behaviors from users, systems, and workloads, as well as various cybersecurity threats, is crucial for taking proactive actions against these threats and ensuring continuous business operations.
- *Proactive failure discovery*: Complementing deep introspection and behavior modeling is active fault (or chaos) injection. Introduced originally as chaos monkey for Netflix [36] and subsequently generalized into chaos engineering [37], pseudo-random failures can be injected into an SDE environment to discover potential failure modes proactively and ensure that the SDE can survive the type of failures that are being tested. Coupling with the containment structures such as Docker container for microservices defined within SDE, the "blast radius" of the failure injection can be controlled without impacting the availability of the services.
- *Policy-based Adjudication*: The behavior model assimilated from the workloads and their environments—including the network traffic—can be adjudicated based on the policies derived from the obligations extracted from pertinent regulations to ensure continuous assurance with respect to these regulations.
- *Self-healing with automatic Investigation and remediation*: A case for subsequent follow-up is created whenever an anomaly (such as microservice failure or network traffic anomaly) is detected from behavior modeling or an exception (such as SSAE 16 violation) is determined from the policy-based adjudication. Automatic mechanisms can be used to collect the evidence, formulate multiple hypothesis, and evaluate the likelihood of each hypothesis based on the available evidence. The most likely hypothesis will then be used to generate recommendation and remediation. A properly designed microservice architecture within an SDE enables fault isolation so that crashed microservices can be detected and restarted automatically without human intervention to ensure continuous availability of the application.
- *Intelligent orchestration*: The assurance engine will continuously evaluate the predicted trajectory of the user, system, workload, and threats and compare against the business objectives and policies to determine whether proactive actions need to be taken by the orchestration engine. The orchestration engine receives instructions from the assurance engine and orchestrates defensive or offensive actions including taking evasive maneuvers as necessary. Examples of these defensive or offensive actions includes fast workload migration from infected areas, fine-grained isolation and quarantine of infected areas of the system, server rejuvenation of those server images when the risk of server image contamination due to malware is found to be unacceptable, and Internet Protocol (IP) address randomization of the workload, making it much more difficult to accurately pinpoint an exact target for attacks.

8.5 COMPOSABLE/DISAGGREGATED DATACENTER ARCHITECTURE

Capability-based resource abstraction within SDE not only decouples the resource requirements of workloads from the details of the computing architecture but also drives the resource pooling at the physical layers for optimal resource utilization within cloud datacenters. Systems in a cloud computing environment often have to be configured according to workload specifications. Nodes within a traditional datacenter are interconnected in a spine–leaf model—first by top-of-rack (TOR) switches within the same racks, then interconnected through the spine switches among racks. There is a conundrum between performance and resource utilization (and hence the cost of computation) when statically configuring these nodes across a wide spectrum of big data &

AI workloads, as nodes optimally configured for CPU-intensive workloads could leave CPUs underutilized for I/O intensive workloads. Traditional systems also impose identical life cycle for every hardware component inside the system. As a result, all of the components within a system (whether it is a server, storage, or switches) are replaced or upgraded at the same time. The "synchronous" nature of replacing the whole system at the same time prevents earlier adoption of newer technology at the component level, whether it is memory, SSD, GPU, or FPGA.

Composable/disaggregated datacenters achieve resource pooling at the physical layer through constructing each system at a coarser granularity so that individual resources such as CPU, memory, HDD, SSD, and GPU can be pooled together and dynamically composed into workload execution units on demand. A composable datacenter architecture is ideal for SDE with heterogeneous and fast evolving workloads as SDEs often have dynamic resource requirements and can benefit from the improved elasticity of the physical resource pooling offered by the composable architecture. From the simulations reported in [28], it was shown that the composable system sustains nearly up to 1.6 times stronger workload intensity than that of traditional systems, and it is insensitive to the distribution of workload demands.

Composable resources can be exposed through hardware-based, hypervisor/operating system based, and middleware-/application-based approaches. Directly expose resource composability through capability-based resource abstraction methodology within SDE to policy-based workload abstractions that allow applications to manage the resources using application-level knowledge is likely to achieve the best flexibility and performance gain.

Using Cassandra (a distributed NoSQL database) as an example, it is shown in [26] that accessing data from across multiple disks connected via Ethernet poses less of a bandwidth restriction than SATA and thus improves throughput and latency of data access and obviates the need for data locality. Overall composable storage systems are cheaper to build and manage and incrementally scalable and offer superior performance than traditional setups.

The primary concern for the composable architecture is the potential performance impacts arising from accessing resources such as memory, GPU, and I/O from nonlocal shared resource pools. Retaining sufficient local DRAM serving as the cache for the pooled memory as opposed to full disaggregation of memory resources and retain no local memory for the CPU is always recommended to minimize the performance impact due to latency incurred from accessing remote memory. Higher SMT levels and/or explicit management by applications that maximize thread level parallelism are also essential to further minimize the performance impact. It was shown in [26] that there is negligible latency and throughput penalty incurred in the Memcached experiments for the read/update operations if these operations are 75% local and the data size is 64 KB. Smaller data sizes result in larger latency penalty, while larger data sizes result in larger throughput penalty when the ratio of nonlocal operations is increased to 50 and 75%.

Frequent underutilization of memory is observed, while CPU is more fully utilized across the cluster in the Giraph experiments. However, introducing composable system architecture in this environment is not straightforward as sharing memory resources among nodes within a cluster through configuring RamDisk presents very high overhead. Consequently, it is stipulated that sharing unused memory across the entire compute cluster instead of through a swap device to a remote memory location is likely to be more promising in minimizing the overhead. In this case, rapid allocation and deallocation of remote memory is imperative to be effective.

It is reported in [38] that there is the notion of *effective memory resource requirements* for most of the big data analytic applications running inside JVMs in distributed Spark environments. Provisioning memory less than the effective memory requirement may result in rapid deterioration of the application execution in terms of its total execution time. A machine learning-based prediction model proposed in [38] forecasts the effective memory requirement of an application in an SDE like environment given its SLA. This model captures the memory consumption behavior of big data applications and the dynamics of memory utilization in a distributed cluster environment. With an accurate prediction of the effective memory requirement, it is shown in [38] that up to 60% savings of the memory resource is feasible if an execution time penalty of 10% is acceptable.

8.6 SUMMARY

As the industry is quickly moving toward converged systems of record and systems of engagement, enterprises are increasingly aggressive in moving mission-critical and performance-sensitive applications to the cloud. Meanwhile, many new mobile, social, and analytics applications are directly developed and operated on the cloud. These converged systems of records and systems of engagement will demand simultaneous agility and optimization and will inevitably require SDEs for which the entire system infrastructure—compute, storage, and network—is becoming software defined and dynamically programmable and composable.

In this chapter, we described an SDE framework that includes capability-based resource abstraction, goal-/policy-based workload definition, and continuous optimization of the mapping of the workload to the available resources. These elements enable SDEs to achieve agility, efficiency, continuously optimized provisioning and management, and continuous assurance for resiliency and security.

REFERENCES

[1] Temple J, Lebsack R. Fit for purpose: workload based platform selection. Journal of Computing Resource Management 2011;129:20–43.

[2] Prodan R, Ostermann S. BA survey and taxonomy of infrastructure as a service and web hosting cloud providers. Proceedings of the 10th IEEE/ACM International Conference on Grid Computing, Banff, Alberta, Canada; 2009. p 17–25.

[3] Data Center and Virtualization. Available at http://www.cisco.com/en/US/netsol/ns340/ns394/ns224/index.html. Accessed on June 24, 2020.

[4] RackSpace. Available at http://www.rackspace.com/. Accessed on June 24, 2020.

[5] Wipro. Available at https://www.wipro.com/en-US/themes/software-defined-everything--sdx-/software-defined-compute--sdc-/. Accessed on June 24, 2020.

[6] Intel. Available at https://www.intel.com/content/www/us/en/data-center/software-defined-infrastructure-101-video.html. Accessed on June 24, 2020.

[7] HP. Available at https://www.hpe.com/us/en/solutions/software-defined.html. Accessed on June 24, 2020.

[8] Dell. Available at https://www.dellemc.com/en-us/solutions/software-defined/index.htm. Accessed on June 24, 2020.

[9] VMware. Available at https://www.vmware.com/solutions/software-defined-datacenter.html. Accessed on June 24, 2020.

[10] Amazon Web Services. Available at http://aws.amazon.com/. Accessed on June 24, 2020.

[11] IBM Corporation. IBM Cloud Computing Overview, Armonk, NY, USA. Available at http://www.ibm.com/cloud-computing/us/en/. Accessed on June 24, 2020.

[12] Cloud Computing with VMWare Virtualization and Cloud Technology. Available at http://www.vmware.com/cloud-computing.html. Accessed on June 24, 2020.

[13] Madnick SE. Time-sharing systems: virtual machine concept vs. conventional approach. Mod Data 1969;2(3):34–36.

[14] Popek GJ, Goldberg RP. Formal requirements for virtualizable third generation architectures. Commun ACM 1974;17(7):412–421.

[15] Barman P, Dragovic B, Fraser K, Hand S, Harris T, Ho A, Neugebauer R, Pratt I, Warfield A. Xen and the art of virtualization. Proceedings of ACM Symposium on Operating Systems Principles, Farmington, PA; October 2013. p 164–177.

[16] Bugnion E, Devine S, Rosenblum M, Sugerman J, Wang EY. Bringing virtualization to the x86 architecture with the original VMware workstation. ACM Trans Comput Syst 2012;30(4):12:1–12:51.

[17] Li CS, Liao W. Software defined networks [guest editorial]. IEEE Commun Mag 2013;51(2):113.

[18] Casado M, Freedman MJ, Pettit J, Luo J, Gude N, McKeown N, Shenker S. Rethinking enterprise network control. IEEE/ACM Trans Netw 2009;17(4):1270–1283.

[19] Kreutz D, Ramos F, Verissimo P. Towards secure and dependable software-defined networks. Proceedings of 2nd ACM SIGCOMM Workshop Hot Topics in Software Design Networking; August 2013. p 55–60.

[20] Stallings W. Software-defined networks and openflow. Internet Protocol J 2013;16(1). Available at https://wxcafe.net/pub/IPJ/ipj16-1.pdf. Accessed on June 24, 2020.

[21] Security Requirements in the Software Defined Networking Model. Available at https://tools.ietf.org/html/draft-hartman-sdnsec-requirements-00. Accessed on June 24, 2020.

[22] ViPR: Software Defined Storage. Available at http://www.emc.com/data-center-management/vipr/index.htm. Accessed on June 24, 2020.

[23] Singh A, Korupolu M, Mohapatra D. BServer-storage virtualization: integration and load balancing in data centers. Proceedings of the 2008 ACM/IEEE Conferenceon Supercomputing, Austin, TX; November 15–21, 2008; Piscataway, NJ, USA: IEEE Press. p 53:1–53:12.

[24] Li CS, Brech BL, Crowder S, Dias DM, Franke H, Hogstrom M, Lindquist D, Pacifici G, Pappe S, Rajaraman B, Rao J. Software defined environments: an introduction. IBM J Res Dev 2014;58(2/3):1–11.

[25] 2012 IBM Annual Report. p 25. Available at https://www.ibm.com/annualreport/2012/bin/assets/2012_ibm_annual.pdf. Accessed on June 24, 2020.

[26] Li CS, Franke H, Parris C, Abali B, Kesavan M, Chang V. Composable architecture for rack scale big data computing. Future Gener Comput Syst 2017;67:180–193.

[27] Abali B, Eickemeyer RJ, Franke H, Li CS, Taubenblatt MA. 2015. Disaggregated and optically interconnected memory: when will it be cost effective?. *arXiv preprint arXiv:1503.01416*.

[28] Lin AD, Li CS, Liao W, Franke H. Capacity optimization for resource pooling in virtualized data centers with composable systems. IEEE Trans Parallel Distrib Syst 2017;29(2):324–337.

[29] Armbrust M, Fox A, Griffith R, Joseph AD, Katz RH, Konwinski A, Lee G, Patterson DA, Rabkin A, Stoica I, Zaharia M. BAbove the Clouds: A Berkeley View of Cloud Computing. University of California, Berkeley, CA, USA. Technical Report No. UCB/EECS-2009-28; February 10, 2009. Available at http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html. Accessed on June 24, 2020.

[30] Moore J. System of engagement and the future of enterprise IT: a sea change in enterprise IT. AIIM White Paper; 2012. Available at http://www.aiim.org/~/media/Files/AIIM%20White%20Papers/Systems-of-Engagement-Future-of-Enterprise-IT.ashx. Accessed on June 24, 2020.

[31] IBM blue gene. IBM J Res Dev 2005;49(2/3).

[32] Mars J, Tang L, Hundt R. Heterogeneity in homogeneous warehouse-scale computers: a performance opportunity. Comput Archit Lett 2011;10(2):29–32.

[33] Basiri A, Behnam N, De Rooij R, Hochstein L, Kosewski L, Reynolds J, Rosenthal C. Chaos engineering. IEEE Softw 2016;33(3):35–41.

[34] Bennett C, Tseitlin A. Chaos monkey released into the wild. Netflix Tech Blog, 2012. p. 30.

[35] Darema-Rogers F, Pfister G, So K. Memory access patterns of parallel scientific programs. Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer System, Banff, Alberta, Canada; May 11–14, 1987. p 46–58.

[36] van der Aalst WMP, Bichler M, Heinzl A. Bus Inf Syst Eng 2018;60:269. doi: https://doi.org/10.1007/s12599-018-0542-4.

[37] Haas J, Wallner R. IBM zEnterprise systems and technology. IBM J Res Dev 2012;56(1/2):1–6.

[38] Tsai L, Franke H, Li CS, Liao W. Learning-based memory allocation optimization for delay-sensitive big data processing. IEEE Trans Parallel Distrib Syst 2018;29(6):1332–1341.

9

COMPUTING, STORAGE, AND NETWORKING RESOURCE MANAGEMENT IN DATA CENTERS

Ronghui Cao[1], Zhuo Tang[1], Kenli Li[1] and Keqin Li[2]

[1] *College of Information Science and Engineering, Hunan University, Changsha, China*

[2] *Department of Computer Science, State University of New York, New Paltz, New York, United States of America*

9.1 INTRODUCTION

Current data centers can contain hundreds of thousands of servers [1]. It is no doubt that the performance and stability of data centers have been significantly impacted by resource management. Moreover, in the course of data center construction, the creation of dynamic resource pools is essential. Some technology companies have built their own data centers for various applications, such as the deep learning cloud service run by Google. Resource service providers usually rent computation and storage resources to users at a very low cost.

Cloud computing platform, which rent various virtual resources to tenants, is becoming more and more popular for resource service websites or data applications. However, with the increasing of virtualization technologies and various clouds continue to expand their server clusters, resource management is becoming more and more complex. Obviously, adding more hardware devices to extend the cluster scale of the data center easily causes unprecedented resource management pressures in data centers.

Resource management in cloud platforms refers to how to efficiently utilize and schedule the virtual resources, such as computing resources. With the development of various open-source approaches and expansion of open-source communities, multiple resource management technologies have been widely used in the date centers. OpenStack [2], KVM [3], and Ceph [4] are some typical examples developed over the past years. It is clear that these resource management methods are considered critical factors for data center creation.

However, some resource management challenges are still impacting the modern data centers [7]. The first challenge is how to integrate various resources (hardware resource and virtual resource) into a unified platform. The second challenge is how to easily manage various resources in the data centers. The third challenge is resource services, especially network services. Choosing an appropriate resource management method among different resource management platforms and virtualization techniques is hence difficult and complex. Therefore, the following criteria should be taken into account: ease of resource management, provisional storage pool, and flexibility in performing the network architectures (such as resource transmission across different instances).

In this chapter, we will first explain the resource virtualization and resource management in data centers. We will then elaborate on the cloud platform demands for data centers and the related open-source cloud offerings focusing mostly on cloud platforms. Next, we will elaborate on the single-cloud bottlenecks and the multi-cloud demands in data centers. Finally, we will highlight the different large-scale cluster resource management architectures based on the OpenStack cloud platform.

9.2 RESOURCE VIRTUALIZATION AND RESOURCE MANAGEMENT

9.2.1 Resource Virtualization

In computing, virtualization refers to the act of creating a virtual (rather than actual) version of something, including

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

virtual computer hardware platforms, storage devices, and computer network resources.

Hardware virtualization refers to the creation of virtual resources acts like the real computer with a full operating system. Software executed on these virtual resources is not directly running on the underlying hardware resources. For example, a computer that is running Microsoft Windows may host a virtual machine (VM) that looks like a computer with the Ubuntu Linux operating system; Ubuntu-based software can be run on the VM.

According to the different deployment patterns and operating mechanism, resource virtualization can be divided into two types: full virtualization and paravirtualization (Fig. 9.1). Full virtualization is also called primitive virtualization technology. It uses the VM to coordinate the guest operating systems and the original hardware devices. Some protected instructions must be captured and processed by the hypervisor. Paravirtualization is another technology that similar to the full virtualization. It uses hypervisor to share the underlying hardware devices, but its guest operating systems integrate the resource virtualization codes. In the past 5 years, the full virtualization technologies gained polarity with the rise of KVM, Xen, etc. KVM is open-source software, and the kernel component of KVM is included in mainline Linux, as of 2.6.20. The first version of KVM was developed at a small Israeli company, Qumranet, which has been acquired by Red Hat in 2008.

For the resource management, data centers must not only comprehensively consider various factors such as manufacturers, equipment, applications, users and technology, etc. but also consider the integration with operation maintenance process of data centers. Obviously, building an open, standardized, easy-to-expand, and interoperable unified intelligent resource management platform is not easy. The scale of data centers is getting larger and more complex, and the types of applications are becoming more and more complex, which makes the difficulty of resource management even more difficult:

- *Multitenant support:* Management of multiple tenants and their applied resources, applications, and operating systems in large-scale data centers with different contracts and agreements.
- *Multi-data center support:* Management of multiple data centers with different security levels, hardware devices, resource management approaches, and resource virtualization technologies.
- *Resource monitor:* Monitoring of various resources with different tenant requests, hardware devices, management platforms, and cluster nodes up to date.
- *Budget control:* Manage the cost of data centers and reduce budget as much as possible, where resources are procured based on "best cost"—regardless if it is deployed at the hardware devices or used for resource virtualization. Additionally, energy and cooling costs are also the principal aspects of budget reducing.
- *Application deploying:* Deploy new applications and services faster with limited understanding of resource availability as well as inconsistent policies and structure.

Data centers with heterogeneous architecture make the above problems particularly difficult since the resource management solutions with high scalability and performance are emergency needed. By tackling these problems, data services can be made more efficient and reliable, notably reducing the internal server costs and increasing the utilization of energy and resource in data centers.

As a result, various virtualization technologies and architectures have been used in data centers to simplify resource management. Without question, the wide use of virtualization brings many benefits for data centers, but it also incurs some costs caused by the virtual machine monitor (VMM) or called hypervisor. These costs usually come from various activities within the virtualization layer such as code rewriting, OS memory operations, and, most commonly, resource scheduling overhead. The hypervisor is the kernel of virtual resource

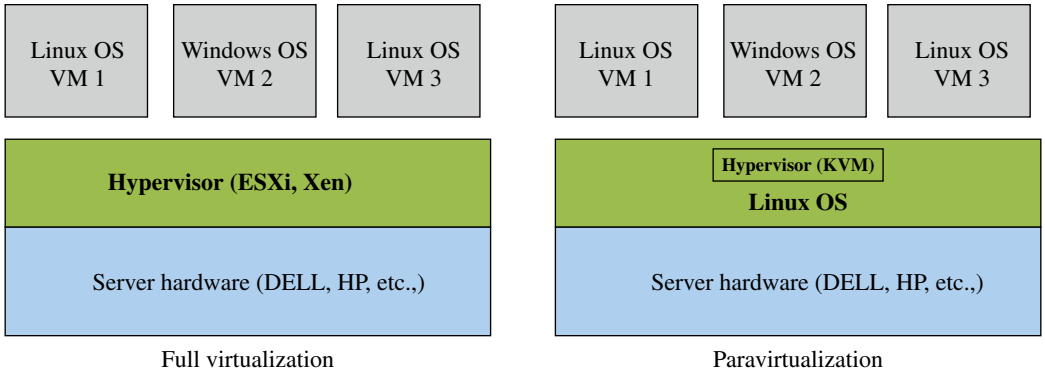


FIGURE 9.1 Two resource virtualization methods.

management, especially for VM. It can be software, firmware, or hardware used to build and execute VM.

Actually, resource virtualization is not a new technology for the large-scale server cluster. It was largely used in the 1960s for mainframe and been widely used in early 2000 for resource pool creation and cloud platforms [5]. In a traditional concept of virtual servers, multiple virtual servers or VMs can be simultaneously operated on one traditional single physical server. As a result, the data centers can operate using VM to improve utilization of server resource capacity and consequently reduce the hardware device cost in data centers. With advances in virtualization technology, we are able to run over 100 VMs on one physical server node.

9.2.2 Resource Management

The actual overhead of resource management and scheduling in data centers vary depending on the virtualization technologies and cloud platforms being used. With greater resource multiplexing, hardware costs can be decreased by resource virtualization. While many data centers would like to move various applications to VMs to lower energy and hardware costs, this kind of transition should be ensured that will not be disrupted by correctly estimating the resource requirements. Fortunately, the disrupt problem can be solved by monitoring the workload of applications and attempt to configure the VMs.

Several earlier researches describe various implementations of hypervisor. The performance results showed that hypervisor can measure the overhead impact of resource virtualization on microbenchmark or macrobenchmark. Some commercial tools use trace-based methods to support server load balancing, resource management, and simulating placement of VMs to improve server resource utilization and cluster performance. Other commercial tools use the trace-based resource management solution that scales the resource usage traces by a given CPU multiplier. In addition, cluster system activities and application operations can incur additional overheads of CPUs.

9.2.2.1 VM Deployment

With the increasing task scale in data centers, breaking down a large serial task into several small tasks and assigning them to different VMs to complete the task in parallel is the main method to reduce the task completion time. Therefore, in modern data centers, how to deploy VMs has become one of the important factors that determine the task completion time and improve resource utilization.

When the VM deployment, the utilization of computation resource, and I/O resource are considered together, it may help to find a multi-objective optimization VM deployment model. Moreover, some VM-optimized deployment mechanisms based on the resource matching bottleneck can also reduce data transmission response time in the data centers. Unfortunately, the excessive time complexity of these VM deployment algorithms will seriously affect the overall operation of data centers.

9.2.2.2 VM Migration

In order to meet the real-time changing requirements of the task, the VM migration technology is introduced in modern data centers. The primary application scenario is using VM migration to integrate resources and decrease energy consumption by monitoring the state of VMs. Green VM Migration Controller (GVMC) combines the resource utilization of the physical servers with the destination nodes of VM migration to minimize the cluster size of data centers. Classical genetic algorithm is often improved and optimized for VM migration to solve the energy consumption problem in data centers.

The VM migration duration is another interesting resource management issue for data centers. It is determined by many factors, including the image size of VM, the memory size, the choice of the migration node, etc. How to reduce the migration duration by optimizing these factors has always been one of the hot topics in data center resource management. Some researchers formalize the joint routing and VM placement problem and leverage the Markov approximation technique to solve the online resource joint optimization problem, with the goal of optimizing the long-term averaged performance under changing workloads.

Obviously, the traditional resource virtualization technologies or resource management mechanisms in data centers are both cannot meet the needs of the new generation of high-density servers and storage devices. On the other hand, the capacity growth of information technology (IT) infrastructure in data centers is severely constrained by floor space. The cloud platform deployed in data centers emerges as the resource management infrastructure to solve these problems.

9.3 CLOUD PLATFORM

The landscape of IT has been evolving ever since the first rudimentary computers were introduced at the turn of the twentieth century. With the introduction of the cloud computing model, the design and deployment of modern data centers have been transformed in the last two decades. Essentially, the difference between cloud service and traditional data service is that in the cloud platform, users can access their resources and data through the Internet. The cloud provider performs ongoing maintenance and updates for resources and services, often owning multiple data centers in several geographic locations to safeguard user data during outages and other failures. The resource management

in the cloud platform is a departure from traditional data center strategies since it provides a resource pool that can be consumed by users as services as opposed to dedicating infrastructure to each individual application.

9.3.1 Architecture of Cloud Computing

The introduction of the cloud platform enabled a redefinition of resource service that includes a new perspective—all virtual resources and services are available remotely. It offers three different model or technical use of resource service (Fig. 9.2): Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).

Each layer of the model has a specific role:

- IaaS layer corresponds to the hardware infrastructure of data centers. It is a service model in which IT infrastructure is provided as a service externally through the network and users are charged according to the actual use of resources.
- PaaS is a model that is "laying" on the IaaS. It provides a computing platform and solution services and allows the service providers to outsource the middleware applications, databases, and data integration layer.
- SaaS is the final layer of cloud and deploys application software on the PaaS layer. It defines a new delivery method, which also makes the software return to the essence of service. SaaS changes the way traditional software services provided, reduces the large amount of upfront investment required for local deployment, and further highlights the service attributes of information software.

9.3.2 Common Open-Source Cloud Platform

Some open-source cloud platforms take a more comprehensive approach, all of which integrate all necessary functions (including virtualization, resource management, application interfaces, and service security) in one platform. If deployed on servers and storage networks, these cloud platforms can provide a flexible cloud computing and storage infrastructure (IaaS).

9.3.2.1 OpenNebula

OpenNebula is an interesting open-source application (under the Apache license) developed at Universidad Complutense de Madrid. In addition to supporting private cloud structures, OpenNebula also supports the hybrid cloud architecture. Hybrid clouds allow the integration of private cloud infrastructure with public cloud infrastructure, such as Amazon, to provide a higher level of scalability. OpenNebula supports Xen, KVM/Linux, and VMware and relies on libvirt for resource management and introspection [8].

9.3.2.2 OpenStack

OpenStack cloud platform was released in July 2010 and quickly became the most popular open-source IaaS solution. The cloud platform is originally combined of two cloud plans, namely, Rackspace Hosting (cloud files) and Nebula platform from NASA (National Aeronautics and Space Administration). It is a cloud operating system that controls large pools of compute, storage, and networking resources throughout a data center, all managed through a dashboard that gives administrators control while empowering their users to provision resources through a web interface [9].

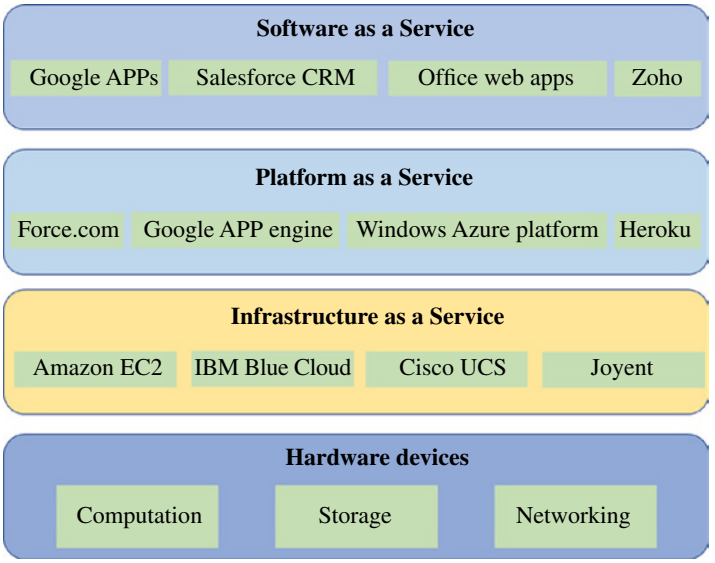


FIGURE 9.2 Architecture of cloud computing model.

9.3.2.3 Eucalyptus

Eucalyptus is one of the most popular open-source cloud solutions that used to build cloud computing infrastructure. Its full name is Elastic Utility Computing Architecture for Linking Your Programs to Useful Systems. The special of eucalyptus is that its interface is compatible with Amazon Elastic Compute Cloud (Amazon EC2—Amazon's cloud computing interface). In addition, Eucalyptus includes Walrus, a cloud storage application that is compatible with Amazon Simple Storage Service (Amazon S3—Amazon's cloud storage interface) [10].

9.3.2.4 Nimbus

Nimbus is another IaaS solution focused on scientific computing. It can borrow remote resources (such as remote storage provided by Amazon EC2) and manage them locally (resource configuration, VM deployment, status monitoring, etc.). Nimbus was evolved from the workspace service project. Since it is based on Amazon EC2, Nimbus supports Xen and KVM.

9.4 PROGRESS FROM SINGLE-CLOUD TO MULTI-CLOUD

With the ever-growing need of resource pool and the introduction of high-speed network devices, the data centers enable building scalable services through the scale-out model by utilizing the elastic pool of computing resources provided by such platforms. However, unlike native components, these extended devices typically do not provide specialized data services or multi-cloud resource management approaches. Therefore, enterprises have to consider the bottlenecks of computing performance and storage stability of single-cloud architecture. In addition, there is no doubt that traditional single-cloud platforms are more likely to suffer from single-point failures and vendor lock-in.

9.4.1 The Bottleneck of Single-Cloud Platform

Facing various resources as well as their diversity and heterogeneity, data center vendors may be confused about whether existing resource pools can completely meet the resource requirements of customer data. If not, no matter the level of competition or development, it is urgent for providers to extend hardware devices and platform infrastructures. To overcome the difficulties, data center vendors usually build a new resource pool under the acceptable bound of the risk and increase the number of resource nodes as the growing amount of data. However, when the cluster scales to 200 nodes, a request message will not respond until at least 10 seconds. David Willis, head of research and development at a UK telecom regulator, estimated that a lone OpenStack controller could manage around 500 computing nodes at most [6]. Figure 9.3 shows a general single-cloud architecture.

The bottlenecks of traditional single-cloud systems first lie in the scalability of architecture, which surely generates considerable expense of data migration. The extension of existing cloud platforms also makes customers suffer from service adjustments of cloud vendors that are not uncommon. For example, resource fluctuation in cloud platforms will affect the price of cloud services. Uncontrolled data availability further aggravates the decline in confidence of users. Some disruptions even lasted for several hours and directly destroy users' confidence. Therefore, vendors were confronted with a dilemma that they could do nothing but build a new cloud platform with a separate cloud management system.

9.4.2 Multi-cloud Architecture

Existing cloud resources exhibit great heterogeneities in terms of both performances and fault-tolerant requirements. Different cloud vendors build their respective infrastructures and keep upgrading them with newly emerging gears. Some multi-cloud architectures that rely on multiple cloud platforms for placing resource data have been used by current cloud providers (Fig. 9.4). Compared with the single-cloud storage, the multi-cloud platform can provide better service quality and more storage features. These features are extremely beneficial to the platform itself or cloud applications such as data backup, document archiving, and electronic health recording, which need to keep a large

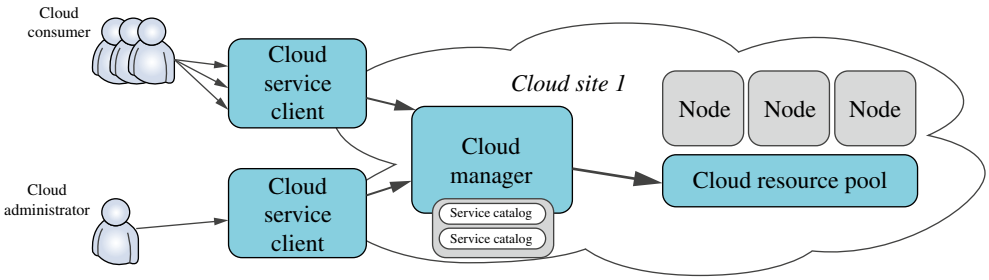


FIGURE 9.3 A general single-cloud site.

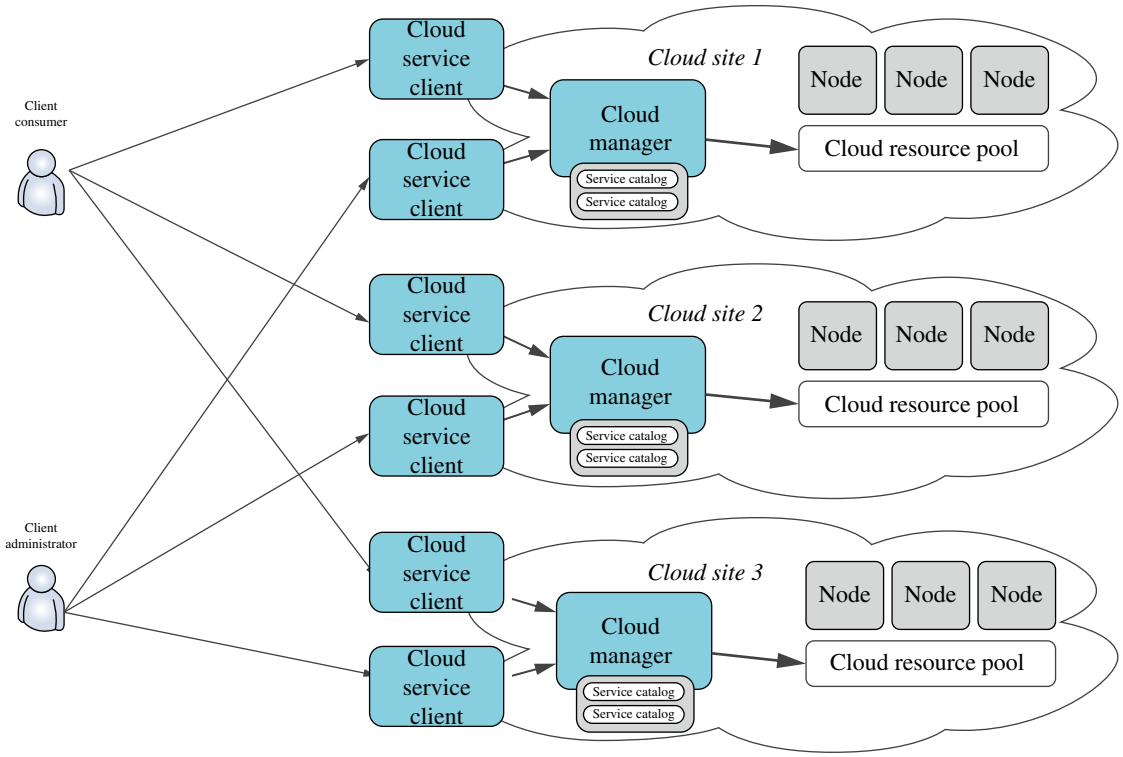


FIGURE 9.4 Multi-cloud environment.

amount of data. Although the multi-cloud platform is a better selection, both administrators and maintainers are still inconvenienced since each bottom cloud site is managed by each provider separately and the corresponding resources are also independent.

Customers have to consider which cloud site is the most appropriate one to store their data with the highest cost effectiveness. Cloud administrators need to manage various resources in different manners and should be familiar with different management clients and configurations among bottom cloud sites. It is no doubt that these problems and restrictions can bring more challenges for resource storage in the multi-cloud environment.

9.5 RESOURCE MANAGEMENT ARCHITECTURE IN LARGE-SCALE CLUSTERS

When dealing with large-scale problems, naturally divide-and-conquer strategy is the best solution. It decomposes a problem of size N into K smaller subproblems. These subproblems are independent of each other and have the same nature as the original problem. In the most popular open-source cloud community, OpenStack community, there are three kinds of divide-and-conquer strategies for resource management in large-scale clusters: multi-region, multi-cell, and resource cascading mechanism. The difference among them is the management concept.

9.5.1 Multi-region

In the OpenStack cloud platform, it supports to divide the large-scale cluster into different regions. The regions shared all the core components, and each of them is a complete OpenStack environment. When deploying in multi-region, the data center only needs to deploy a set of public authentication service of OpenStack, and other services and components can be deployed like a traditional OpenStack single-cloud platform. Users must specify a specific area/region when requesting any resources and services. Distributed resources in different regions can be managed uniformly, and different deployment architectures and even different OpenStack versions can be adopted between regions. The advantages of multi-region are simple deployment, fault domain isolation, flexibility, and freedom. It also has obvious shortcomings that every region is completely isolated from each other and the resources cannot be shared with each other. Cross-region resource migration can also not be supported. Therefore, it is particularly suitable for scenarios that the resources cross different data centers and distribute in different regions.

9.5.2 Nova Cells

The computation component of OpenStack provides nova multi-cell method for large-scale cluster environment. It is different from multi-region; it divides the large-scale clusters according to the service level, and the ultimate goal is to achieve that the single-cloud platform can support the capabilities of deployment and flexible expansion in data centers. The main strategy of nova cells (Fig. 9.5) is to divide different computing resources into cells and organize them in the form of a tree. The architecture of nova cells is shown as follows.

There are also some nova cell use cases in industry:

1. CERN (European Organization for Nuclear Research) OpenStack cluster may be the largest OpenStack deployment cluster currently disclosed. The scale of deployment as of February 2016 is as follows [11]:
   - Single region and 33 cells
   - 2 Ceph clusters
   - 5500 compute nodes, totaling 140k cores
   - More than 17,000 VMs
2. Tianhe-2 is one of the typical examples of the scale of China's thousand-level cluster, and it has been deployed and provided services in the National Supercomputer Center in Guangzhou in early 2014. The scale of deployment is as follows [12].
   - Single region and 8 cells.
   - Each cell contains 2 control nodes and 126 computing nodes.
   - The total scale includes 1152 physical nodes.

9.5.3 OpenStack Cascading

OpenStack cascading is a large-scale OpenStack cluster deployment supported by Huawei to support scenarios including 100,000 hosts, millions of VMs, and unified management across multiple data centers (Fig. 9.6). The strategy it adopts is also divide and conquer, that is, split a large OpenStack cluster into multiple small clusters and cascade the divided small clusters for unified management [13].

When users request resources, they first submit the request to the top-level OpenStack API. The top-level OpenStack will select a suitable bottom OpenStack based on a certain scheduling policy. The selected bottom OpenStack is responsible for the actual resource allocation.

This solution claims to support spanning up to 100 data centers, supports the deployment scale of 100,000 computing nodes, and can run 1 million VMs simultaneously. At present,

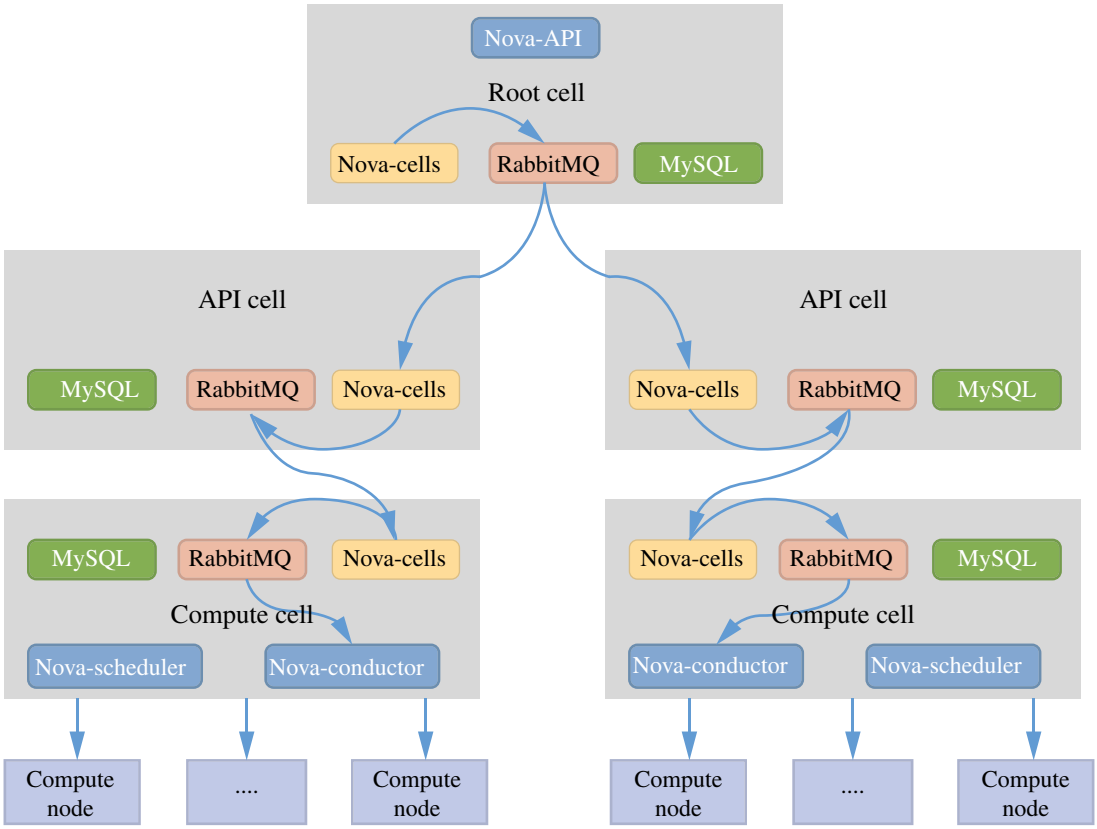


FIGURE 9.5 Nova cell architecture.

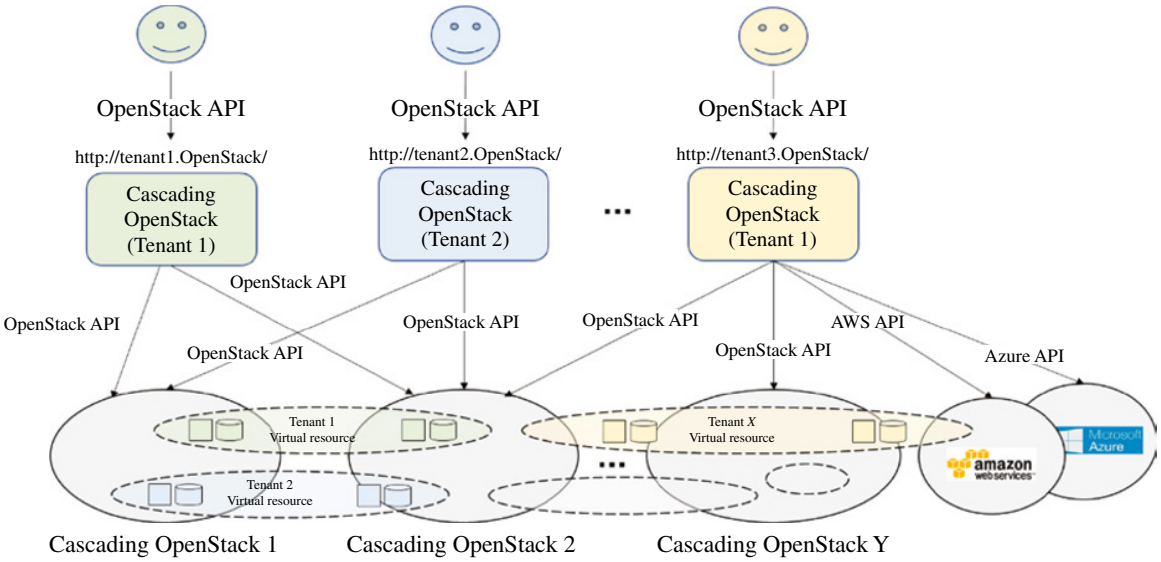


FIGURE 9.6 OpenStack cascading architecture.

the solution has separated two independent big-tent projects: one is Tricircle, which is responsible for network automation development in multi-cloud environment with networking component Neutron, and the other is Trio2o, which provides a unified API gateway for computation and storage resource management in multi-region OpenStack clusters.

9.6 CONCLUSIONS

The resource management of data centers is indispensable. The introduction of virtualization technologies and cloud platforms undoubtedly significantly increased in the resource utilization of data centers. Numerous scholars have produced a wealth of research on various types of resource management and scheduling in the data centers, but there is still further research value in many aspects. On the one hand, the resource integration limit still exists in a traditional data center and single-cloud platform. On the other hand, due to the defects of nonnative management of additional management plugins, existing multi-cloud architectures make resource management and scheduling often accompanied by high bandwidth and data transmission overhead. Therefore, the resource management of data centers based on the multi-cloud platform emerges at the historic moment under the needs of the constantly developing service applications.

REFERENCES

[1] Geng H. Chapter 1: Data Centers--Strategic Planning, Design, Construction, and Operations,*Data Center Handbook*. Wiley, 2014.

[2] Openstack. Available at http://www.openstack.org. Accessed on May 20, 2014.

[3] KVM. Available at http://www.linux-kvm.org/page/Main_Page. Accessed on May 5, 2018.

[4] Ceph. Available at https://docs.ceph.com/docs/master/. Accessed on February 25, 2018.

[5] Kizza JM. Africa can greatly benefit from virtualization technology–Part II. Int J Comput ICT Res 2012;6(2).Available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.372.8407&rep=rep1&type=pdf. Accessed on June 29, 2020.

[6] Cao R, et al. A scalable multi-cloud storage architecture for cloud-supported medical Internet of Things. IEEE Internet Things J, March 2020;7(3):1641–1654.

[7] Beloglazov A, Buyya R. Energy efficient resource management in virtualized cloud data centers. Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, Melbourne, Australia; May 17–20, 2010. IEEE. p. 826–831.

[8] Milojičić D, Llorente IM, Montero RS. Opennebula: a cloud management tool. IEEE Internet Comput 2011;15(2):11–14.

[9] Sefraoui O, Aissaoui M, Eleuldj M. OpenStack: toward an open-source solution for cloud computing. Int J Comput Appl 2012;55(3):38–42.

[10] Boland DJ, Brooker MIH, Turnbull JW. Eucalyptus Seed; 1980. Available at https://www.worldcat.org/title/eucalyptus-seed/oclc/924891653?referer=di&ht=edition. Accessed on June 29, 2020.

[11] Herran N. Spreading nucleonics: The Isotope School at the Atomic Energy Research Establishment, 1951–67. Br J Hist Sci 2006;39(4):569–586.

[12] Xue W, et al. Enabling and scaling a global shallow-water atmospheric model on Tianhe-2. Proceedings of the 2014 IEEE 28th International Parallel and Distributed Processing Symposium, Phoenix, AZ; May 19–23, 2014. IEEE. p. 745–754.

[13] Mayoral A, et al. Cascading of tenant SDN and cloud controllers for 5G network slicing using Transport API and Openstack API. Proceedings of the Optical Fiber Communication Conference. Optical Society of America, Los Angeles, CA; March 19–23, 2017. M2H. 3.

10

WIRELESS SENSOR NETWORKS TO IMPROVE ENERGY EFFICIENCY IN DATA CENTERS

Levente Klein, Sergio Bermudez, Fernando Marianno and Hendrik Hamann
IBM TJ Watson Research Center, Yorktown Heights, New York, United States of America

10.1 INTRODUCTION

Data center (DC) environments play a critical role in maintaining the reliability of computer systems. Typically, manually controlled air-cooling strategies are implemented to mitigate temperature increase through usage of computer room air conditioning (CRAC) units and to eliminate overheating of information technology (IT) equipment. Most DCs are provisioned to have at least the minimum required *N* CRAC units to maintain safe operating conditions, with an additional unit, total N+1 provisioned to ensure redundancy. Depending on the criticality of DC operations, the CRAC units can be doubled to 2*N* to increase DC uptime and avoid accidental shutdown [1].

The main goal of control systems for CRAC units is to avoid overheating and/or condensation of moisture on IT equipment. The CRAC units are driven to provide the necessary cooling and maintain server's manufacturer-recommended environmental operating parameters. Many DCs recirculate indoor air to avoid accidental introduction of moisture or contamination, even when the outdoor air temperature is lower than the operating point of the DC. Most DCs operate based on the strategy of maintaining low temperature in the whole DC, and their local (in-unit) control loops are based on recirculating and cooling indoor air based on a few (in-unit) temperature and relative humidity sensors. While such control loops are simple to implement and result in facility-wide cooling, the overall system performance is inefficient from the energy consumption perspective—indeed the energy consumed on cooling can be comparable to the cost of operating the IT equipment. Currently, an industry-wide effort is underway to improve the overall cooling performance of DCs by minimizing cooling cost while keeping the required environmental conditions for the IT equipment. Since currently DCs lack enough environmental sensor data, the first step to improve energy efficiency in a DC is to measure and collect such data. The second step is to analyze the data to find optimal DC operating conditions. Finally, implementing automatic control of the DC attains the desired energy efficiency.

Measuring and collecting environmental data can be achieved by spatially dense and granular measurements either by using (i) a mobile measurement system (Chapter 35) or (ii) by deploying a wireless sensor network. The advantage of dense monitoring is the high temporal and spatial resolution and quick overheating detection around IT equipment, which can lead to more targeted cooling. The dynamic control of CRAC systems can reduce significantly the energy consumption in a DC by optimizing targeted cooled airflow to only those locations that show significant overheating ("hot spots"). A wireless sensor network can capture the local thermal trends and fluctuations; furthermore, these sensor readings are incorporated in analytics models that generate decision rules that govern the control loops in a DC, which ensures that local environmental conditions are maintained within safe bounds.

Here we present two strategies to reduce energy consumption in DCs: (i) dynamic control of CRACs in response to DC environment and (ii) utilization of outside air for cooling. Both strategies rely on detailed knowledge of the environmental parameters within the DCs, where the sensor networks are integral part of the control system. In this chapter, we discuss the basics of sensor network architecture and the implemented control loops—based on sensor analytics—to reduce energy usage in DCs and maintain reliable operations.

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

10.2 WIRELESS SENSOR NETWORKS

Wireless sensor networks are ubiquitous monitoring systems, here used to measure environmental conditions in DCs, as different sensors can be connected to microcontroller/radios and simultaneously acquire multiple environmental parameters [2]. A wireless sensor network is composed of untethered devices with sensors (called motes or nodes) and a gateway (also called central hub or network manager) that manages and collects data from the sensor network, as well as serving as the interface to systems outside the network. Wireless radios ensure local sensor measurements are transmitted to the central location that aggregates data from all the radios [3]. The wireless solution has distinctive advantages compared with wired monitoring solutions, like allowing full flexibility for placing the sensors in the most optimal location (Fig. 10.1). For example, sensors can be placed readily below the raised floors, at air intakes, air outlets, or around CRAC units.

An ongoing consideration when using a sensor network for DC operations is the installation cost and sensor network maintenance. Each sensor network requires both provision of power and a communication path to the central data collection point. Since DCs can be frequently rearranged, as new IT equipment is installed or removed from the facility, effortless rearranging of the sensor network can be achieved by using wireless sensor networks—since sensor nodes can be easily relocated and extended in sensing from one to multiple environmental parameters (e.g. adding a new pressure or relative humidity sensor to a wireless node that measured only temperature). Sensors are connected through a wireless path where each radio can communicate with other neighboring radios and relay their data in multi-hop fashion to the central collection point [4]. If a data packet is generated at the edge of the network, the packet will hop from the acquisition point to the next available node until it reaches the central control point (gateway). Depending on the network architecture, either all motes or only a subset of them are required to be available in order to facilitate the sensor information flow. The network manager can also facilitate the optimization of the data transmission path, data volume, and latency such that data can be acquired from all points of interest.

The communication between low-power nodes enables a data transfer rate of up to 250 kbps, depending on the wireless technology being used. The computational power of a typical microcontroller is enough to perform simple processing on the raw measurements, like averaging, other statistical operations, or unit conversions. For the energy efficiency and air quality monitoring tasks, the motes were built using four environmental sensors (temperature, airflow, corrosion, and relative humidity) as described in the next section. Each mote is powered with two AA lithium batteries and takes samples of each of those three sensors every minute. Since the rest of the time the mote is sleeping, the battery lifetime of each unit is around 5 years [5].

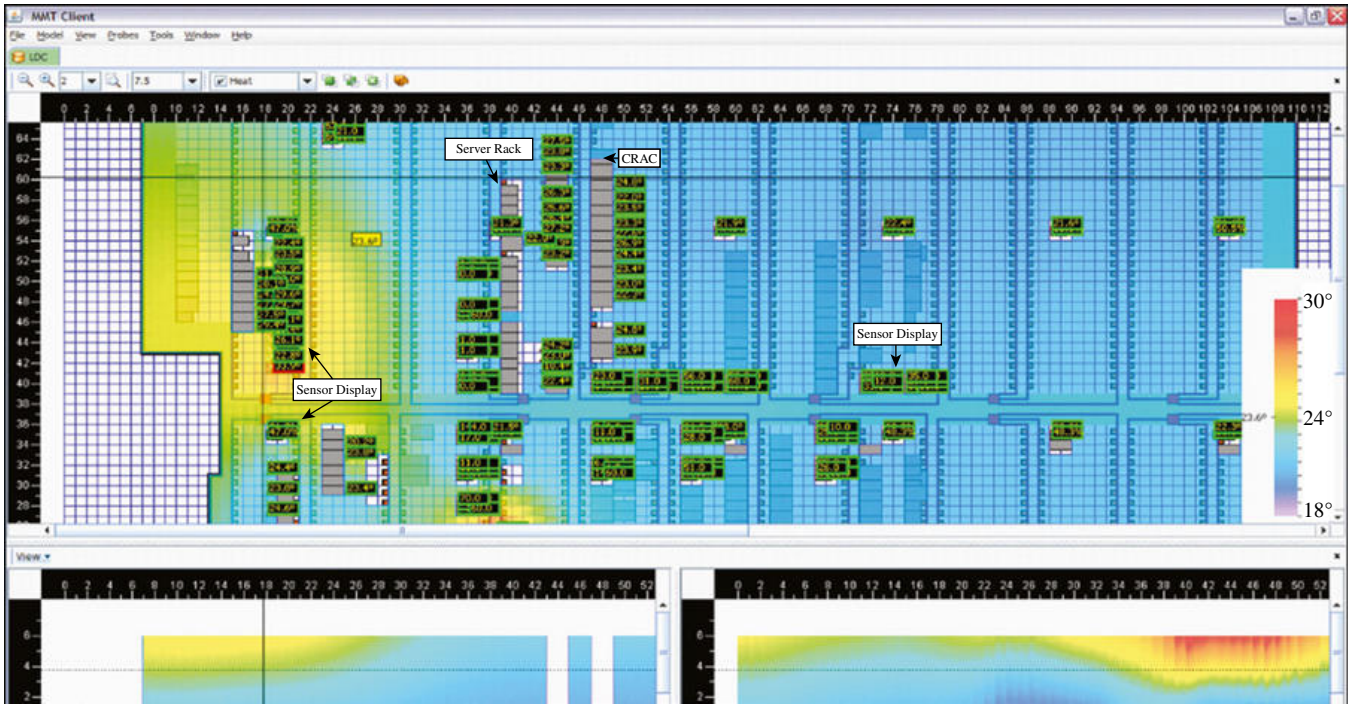


FIGURE 10.1 Data center layout with servers and wireless sensor network overlaid on a computational fluid dynamics simulation of temperature in horizontal direction, while the lower right image shows cross-sectional temperature variation in vertical direction.

The radios can be organized in different network topologies like star, mesh, or cluster tree [6]. The most common approach for network topology is a mesh network, where every radio will connect to one or more nearby radios. In a mesh network, messages hop from node to node extending the network across a large area and reporting back to the gateway, which aggregates data from each radio and timestamps each data point (note that a timestamp can also be applied by the mote that takes the measurement). Each wireless radio is a node of the mesh network, and its data propagates to the central gateway, which is the external interface of the wireless network. One gateway can support hundreds of motes simultaneously, and the mesh network can cover a lateral span of hundreds of meters. Current development and hardening of wireless networks have made them extremely robust and reliable, even for mission-critical solutions where more than 99.9% reliability can be achieved with data acquisition [7].

Multiple communication protocols can be implemented for wireless radios like Zigbee, 6LoWPAN, WirelessHART, SmartMesh IP, and ISA100.11a, all of which use a 2.4 GHz radio (unlicensed ISM band). Many of the above communication technologies will require similar hardware and firmware development. Wireless sensor networks in DCs have a few special requirements compared with other sensor networks such as (i) high reliability, (ii) low latency (close to real-time response), and (iii) enhanced security (for facilities with critical applications).

10.3 SENSORS AND ACTUATORS

The most common sensors used in DCs are temperature, relative humidity, airflow, pressure, acoustic, smoke, and air quality sensors. Temperature sensors can be placed either in front of servers where cold air enters in the server rack or at the back of the server measuring the exhaust temperature of servers. The difference between the inlet and exhaust temperature is an indicator of the IT computational load, as central processing units (CPUs) and memory units heat up during operations. Also, such temperature sensors can be placed at different heights, which enables to understand the vertical cooling provided through raised floor DCs. Additionally, pressure sensors can be placed under the raised floor to measure the pressure levels, which are good indicators of potential airflow from the CRAC units.

The accuracy and number of sensors deployed in a DC are driven by the expected granularity of the measurement and the requirements of the physical or statistical models that predict the dynamics of temperature changes within the DCs. Higher-density sensor network can more accurately and timely capture the dynamic environment within the DCs to increase potentially energy savings. In Figure 10.1 a typical DC with a wireless sensor network is shown where sensor readings are used in a computational fluid dynamics (CFD) model to assess the distribution of temperature across the whole facility. At the bottom of the image, the cross sections of temperature distributions along horizontal and vertical directions in the DC are shown as extracted from the CFD model. The CFD model can run operationally regularly, or it can be updated on demand. These dynamic maps, created from the sensor network readings, are useful to pinpoint hot spot locations in a DC (the left side of the DC in Fig. 10.1 shows 5ºC higher temperature in the gray scale or in yellow/red in the ebook) regions as indicated in the temperature heat map). The sensor readings are part of the boundary conditions used by the CFD model, and such models can be updated periodically as the IT loads on servers' changes.

Most of the sensors deployed in DCs are commercially available with digital or analog output. Each sensor can be attached to a mote that displays the local measurement at that point (Fig. 10.2a). In case that special sensors are required, they can be custom manufactured. One such sensor is the corrosion sensor that was developed to monitor air quality for DCs, which are either retrofitted or equipped to use air-side economization [8]. In addition to sensors, control relays can be mounted around CRAC that can turn them on/off (Fig. 10.2c). The relays are turned on/off based on sensor reading that are distributed around racks (Fig. 10.2d).

The corrosion sensors (Fig. 10.2b) measure the overall contamination levels in a DC and are based on thin metal films that are fully exposed to the DC environment. The metal reaction with the chemical pollutants in the air changes the film surface chemistry and provides an aggregated response of the sensor to concentration of chemical pollutants (like sulfur-bearing gases). The sensor measures the chemical change of metal films (i.e. corrosion); this change is an indicator of how electronic components (e.g. memory, CPUs, etc.) or printed circuit boards may react with the environment and get degraded over time. The metals used for corrosion sensors are copper and silver thin films. Copper is the main metal used for connecting electronic components on printed circuit boards, while silver is part of solder joints anchoring electronic components on the circuit boards. If sulfur is present in air, it gets deposited on the silver films, and, in combination with temperature, it creates a nonconductive Ag_2S thin layer on top of the Ag film—or Cu_2S on top of the Cu film for copper-based corrosion sensors. As the nonconductive film grow in thickness on the top of Ag and Cu thin films, it reduces the conductive film thickness, resulting in an increased sensor resistance. The change in resistance is monitored through an electric circuit where the sensor is an active part of a Wheatstone bridge circuit. The air quality is assessed through corrosion rate estimations, where the consumed film thickness over a certain period of time is measured, rather than the absolute change of film thickness [9]. The corrosion rate measurement

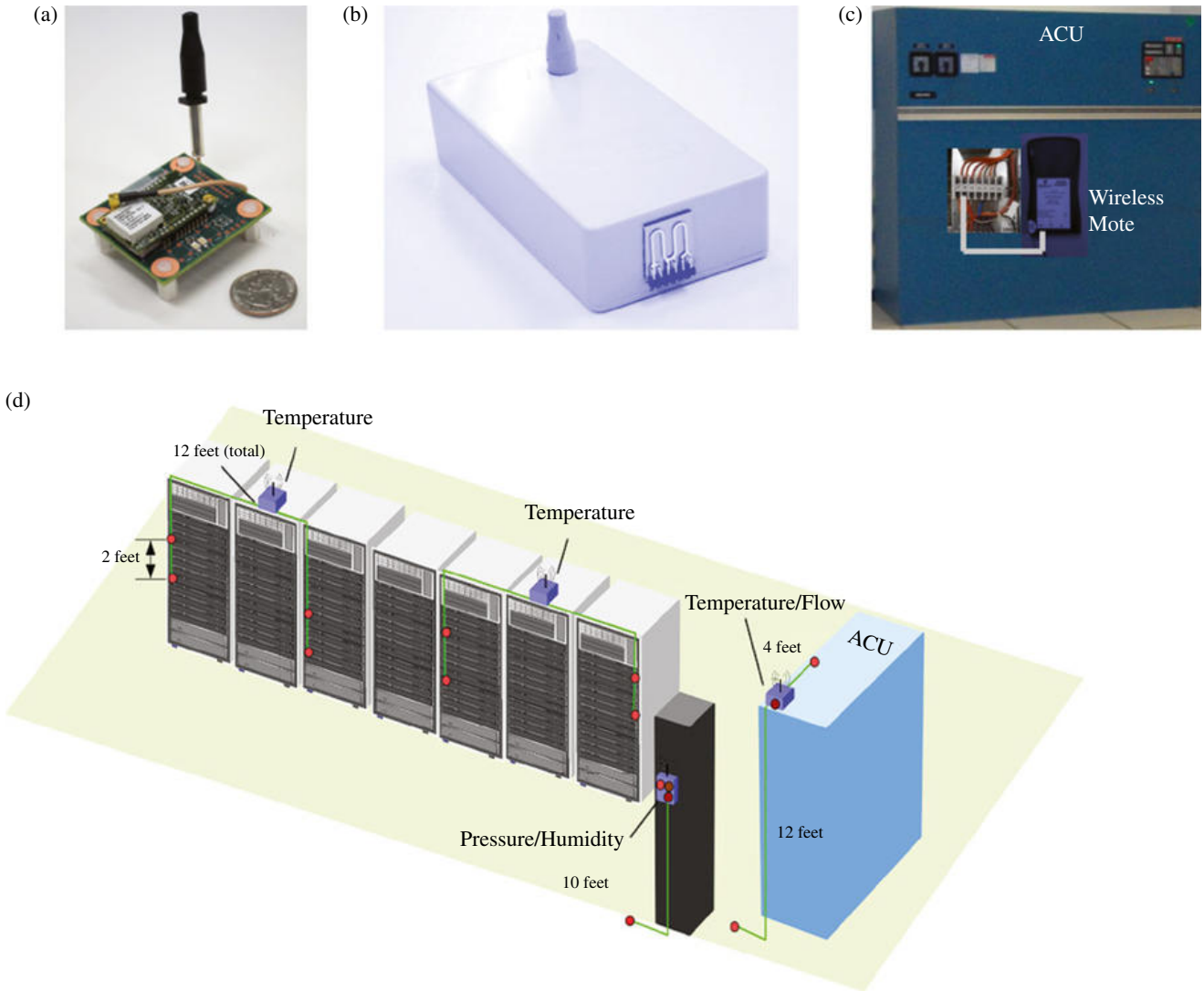


FIGURE 10.2 (a) Wireless sensor mote with integrated microcontroller and radios, (b) corrosion sensor with integrated detection circuit, (c) CRAC control relays, and (d) wireless radio mounted on a rack in a data center.

is an industry-wide agreed measure where a contamination-free DC is characterized by a corrosion rate of less than 200 Å/month for silver and 300 Å/month for copper [10], while higher corrosion rate values indicate the presence of contamination in the air. The main concern for higher concentrations of contaminating gases is the reduced reliable lifetime of electronic components, which can lead to server shutdown [11].

The corrosion sensor can be connected to a controller, which then forms a system that can automatically adjust the operation of CRAC units. These controllers react to the sensor network readings (Fig. 10.2c) and will be discussed below in details. Controllers and sensors are distributed across DCs, with multiple sensors associated with each rack (Fig. 10.3a). One schematic implementation is shown in Figure 10.3b where sensors are positioned at the inlet of servers, CRAC, under the raised floor, and air exchanger (AEx). Data aggregation and control of the cooling units are carried out in a cloud platform.

10.4 SENSOR ANALYTICS

10.4.1 Corrosion Management and Control

The corrosion sensors along with temperature, relative humidity, static pressure, differential pressure, and airflow sensors are continuously measuring the environmental conditions in a DC. Temperature sensors are located at the inlet side of servers, as well as at the air supply and air return areas of each CRAC. To monitor if CRAC units are used, airflow sensors are positioned at the inlet and outlet to

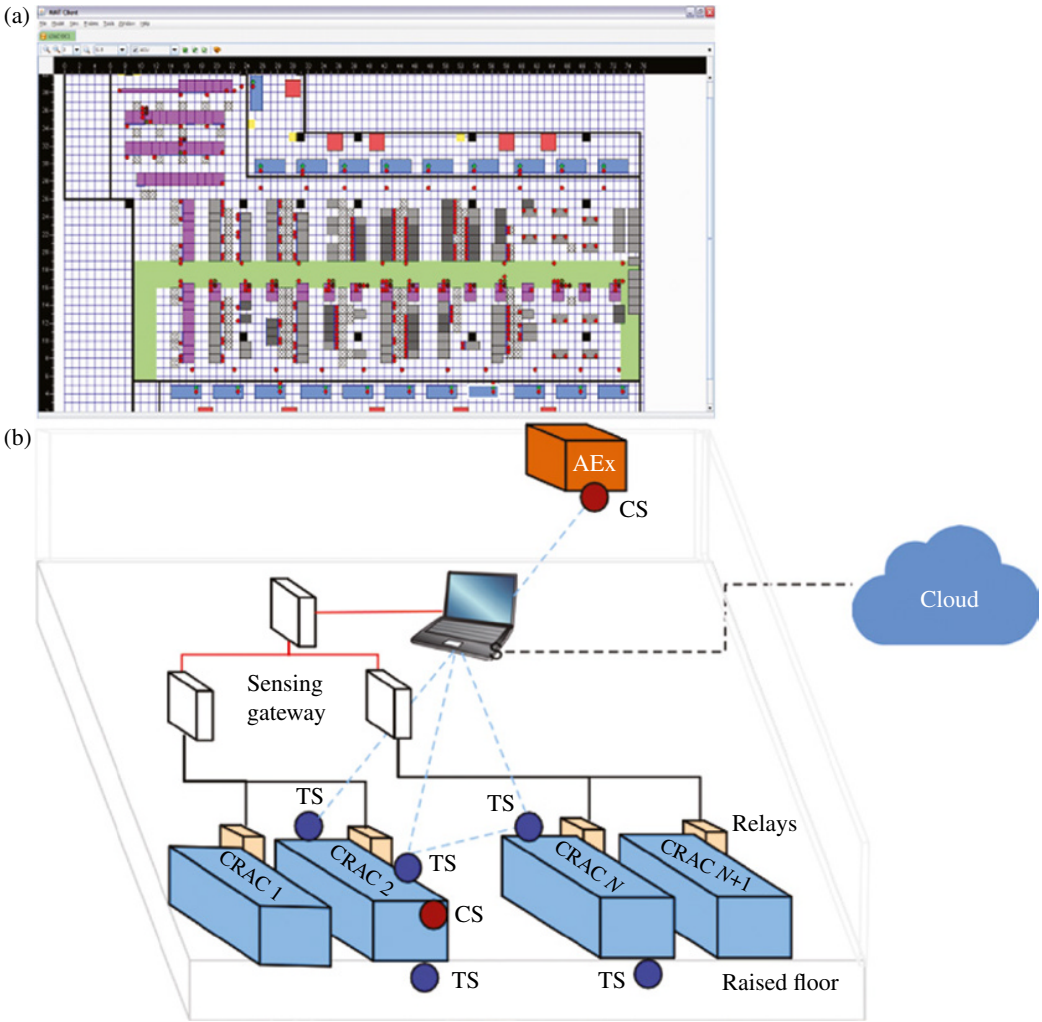


FIGURE 10.3 (a) Data center layout with CRACS (blue), racks (gray), and sensors (red dots). (colors are shown in ebook.) (b) Schematics of sensor network layout in a data center with temperature sensors (TS), corrosion sensor (CS), and relays positioned around CRACS, air exchanger (AEX), and under raised floor.

measure the air transferred. Temperature sensors mounted in the same locations can assess the intake and outlet air temperature and measure the performance of CRAC units. Additionally, corrosion sensors can be placed at air exchange (AEx) and CRAC air intake positions as most of the air moving through a DC will pass through these points (Fig. 10.3). If the corrosion sensor reading is small, then outside air may be introduced in the DC and mixed with the indoor air without any risk of IT equipment damage. The amount of outside air allowed into the DC can be controlled by the feedback from the wireless sensing network that monitors hot spot formation in the DC and the corrosion sensor reading to assure optimal air quality.

The output of the corrosion sensor (resistance change) is expressed as corrosion rate, where instantaneous resistance changes are referenced to the resistance values 24 hours in the past. The historical reference point is obtained by averaging the resistance values across a 12-hour period, centered on the time that is 24 hours in the past from the moment when the corrosion rate is calculated. Averaging the current reading across a short time interval (e.g. using the last 10 readings) and averaging the historical reference point (e.g. over 2-week period) reduce the noise in the corrosion rate calculations and provide more robust temperature compensation by minimizing inherent sensor reading fluctuations. In addition, the corrosion sensor can pass through a Kalman filter that predicts the trends of sensor reading to integrate predictive capabilities in operations [12].

The corrosion rate can vary significantly across a few months, mainly influenced by the pollution level of outside air and temperature of the DC indoor environment. When the corrosion rate measured value for outside air is below the accepted threshold (200 Å/month for silver), then the outside air can be allowed into the DC through the air

exchanger that controls the volume of outside air introduced in the facility. Examples of possible additional constraints are (i) to require the temperature of the outside air to be below the cooling set point for the DC and (ii) air humidity to be below 90%. Since the combination of temperature and relative humidity has a synergistic contribution, the environment needs to be monitored to avoid condensation, which may occur if the temperature of IT servers falls below the dew point of air and may result in water accumulation.

Figure 10.4a shows the corrosion rate for a DC where the corrosion rate exceeds for a short period of time the threshold value (300 Å/month) when the air exchanger was closed, which resulted in a gradual decrease of the corrosion rate. The corrosion rate values were validated through standard silver and copper coupon measurements (Fig. 10.4a).

10.4.2 CRAC Control

The main CRAC control systems are implemented using actuators that can change the operating state of the CRAC units. The remote controller actuator is attached to each CRAC in the DC (Fig. 10.3b). The base solution could be applied to two different types of CRACs: discrete and variable speed. The former ones only accept on or off commands, so the unit is either in standby mode or in full operating mode. The latter CRAC types (e.g. variable-frequency drive [VFD]) can have their fan speed controlled to different levels—thus increasing even more the potential energy optimization. For simplicity purposes, this chapter only considers the discrete control of CRACS, i.e. a unit that can be in only one of two states: *on* or *off*. But similar results apply to VFD CRACs.

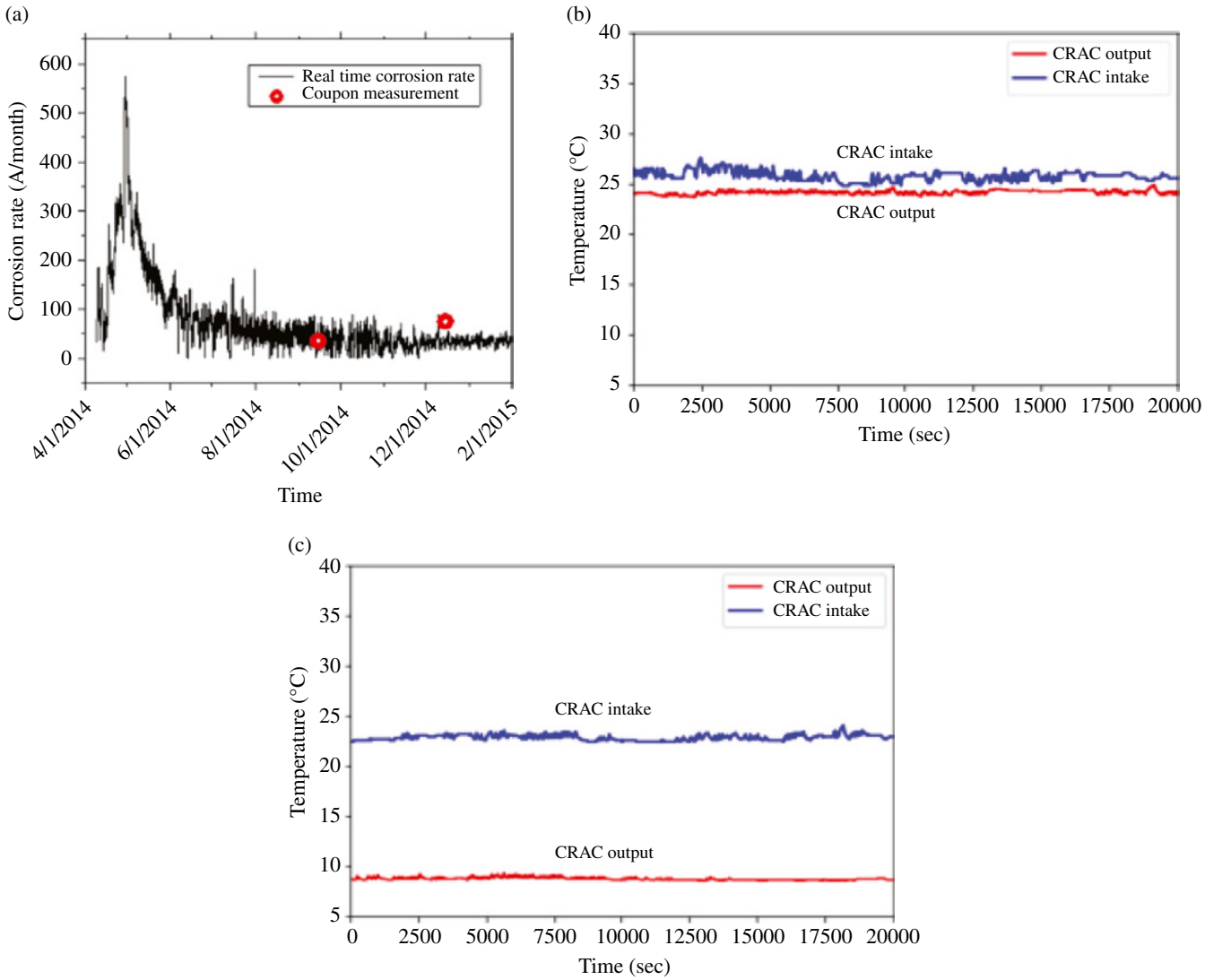


FIGURE 10.4 (a) Corrosion rate in a data center where rate exceeds the acceptable 200 Å/month level and (b) the inlet and outlet temperature of a poorly operated CRAC unit and (c) the inlet and outlet temperature of a well-utilized CRAC unit.

In the base solution, each CRAC unit is controlled independently based on the readings of the group of sensors positioned at the inlet of server racks that are in the area of influence of such CRAC. The remote controller actuators have a watchdog timer for fail-safe purposes, and there is one actuator per CRAC (Fig. 10.3b). Additionally, the inlet and outlet temperature of each CRAC unit is monitored by a sensor mounted at those locations. It is expected that the outlet temperature is lower than the inlet temperature that collects the warmed-up air in the DC. The CRAC utilization is not optimal when the difference between the inlet and outlet temperature is similar (Fig. 10.4b). For a CRAC that is being efficiently managed, this difference in temperatures can be significant (Fig. 10.4c).

Both sensors and actuators communicate with the main server that keeps track of the state of each CRAC units in the DC. This communication link can take multiple forms, e.g. a direct link via Ethernet or through another device (i.e. an intermediate or relay computer, as it is shown within a dotted box in Fig. 10.3b).

By using the real-time data stream from the environmental sensors and through DC analytics running in the software platform, it is possible to know if and which CRACs are being underutilized. With such information, the software control agents can turn off a given set of CRACs when being underutilized, or they can turn on a CRAC when a DC event occurs (e.g. a hot spot or a CRAC failure). See more details in the Section 10.6.2.

10.5 ENERGY SAVINGS

The advantage of air-side economizer can be simply summarized as the energy savings associated with turning off underutilized CRAC units and chillers. For an underutilized CRAC unit, pump and blowers are consuming power while contributing very little to cooling. Those underutilized CRACs can be turned off or replaced with outside air cooling [13–16].

The energy savings potential is estimated using the coefficient of performance (COP) metric. The energy consumed in a DC is divided in two parts: (i) energy consumed for air transport (pumps and blowers) and (ii) energy consumed to refrigerate the coolant that is used for cooling [17]. In a DC, the total cooling power can be defined as

$$P_{\text{Cool}} = P_{\text{Chill}} + P_{\text{CRAC}} \tag{10.1}$$

where P_{Chill} is the power consumed on refrigeration and P_{CRAC} is the power consumed on circulating the coolant. The energy required to move coolant from the cooling tower to the CRACs is not considered in these calculations. If the total dissipated power is P_{RF}, the COP metric is defined for chillers and for CRACs, respectively:

$$COP_{\text{Chill}} = P_{\text{RF}} / P_{\text{Chill}} \quad \text{and} \quad COP_{\text{CRAC}} = P_{\text{RF}} / P_{\text{CRAC}} \tag{10.2}$$

The cooling power can be expressed as

$$P_{\text{Cool}} = P_{\text{RF}} \left(1/COP_{\text{Chill}} + 1/COP_{\text{CRAC}} \right) \tag{10.3}$$

In the case of the cooling control system, the total power consumed for CRAC operations can be neglected, while in the case of outdoor air cooling, the calculations are detailed below.

For savings' calculation, a power baseline at moment $t = 0$ is considered, where the total power at any moment of time t is based on evaluating business as usual (BAU) or no changes to improve energy efficiency:

$$P_{\text{Cool}}^{\text{BAU}}(t) = P_{\text{RF}(t)} / COP(t=0) \tag{10.4}$$

The actual power consumption at time t is

$$P_{\text{Cool}}^{\text{Actual}}(t) = P_{\text{RF}(t)} / COP(t) \tag{10.5}$$

where energy efficiency measures are implemented.

Power savings can be calculated as the difference between actual and BAU power consumption:

$$P_{\text{Cool}}^{\text{Savings}}(t) = P_{\text{Cool}}^{\text{Actual}}(t) - P_{\text{Cool}}^{\text{BAU}}(t) \tag{10.6}$$

The cumulated energy savings can then be calculated over a certain period (t_1, t_2) as

$$E_{\text{Cool}}^{\text{Savings}}(t) = \int_{t_1}^{t_2} P_{\text{Cool}}^{\text{Savings}}(t)\,dt. \tag{10.7}$$

In the case of air-side economization, the main factors that drive energy savings are the set point control of the chilling system and chiller utilization factor.

The power consumption of the chilling system can be assumed to be composed of two parts: (i) power dissipation due to compression cycle and (ii) power consumed to pump the coolant and power consumed by the cooling tower.

A simplified formula used for estimating chiller power consumption is

$$P_{\text{Chill}} = \frac{P_{\text{RF}}\chi}{COP_{\text{Chill}}\left(1 + m_2\left(T_{\text{OS,O}} - T_{\text{OS}}\right)\right)\left(1 + m_1\left(T_{\text{S}} - T_{\text{S,O}}\right)\right)} + P_{\text{RF}} f_{\text{Chill}} \tag{10.8}$$

where

χ is chiller utilization factor,
COP_{Chill} is the chiller's COP,
$T_{\text{OS,O}}$ is outside air temperature,
T_{OS} is the air discharge temperature set point (the temperature the CRAC is discharging),
m_1 and m_2 are coefficients that describe COP_{Chill} change as function of T_{OS} and set point temperature T_{S},
f_{Chill} is on the order of 5%.

Values for m_1 and m_2 can be as large as 5%/°C [15]. The discharge set point temperature (T_S) is controlled by DC operators, and the highest possible set point can be extracted by measuring the temperature distribution in the DC using a distributed wireless sensor network.

Assuming a normal distribution with a standard deviation σ_T, the maximum allowable hot spot temperature T_{HS} can be defined as

$$T_{HS} = T_S \pm 3\sigma_T \tag{10.9}$$

where a three-sigma rule is assumed with the expectations that less than 0.25% of the servers will see temperatures at inlet higher than the chosen hot spot temperature T_{HS}.

The chiller may be fully utilized for a closed DC. Since there may be additional losses in the heat exchange system as outside air is moved into the facility, the effect of heating the air as it is moved to servers can be aggregated into a temperature value (ΔT), where its value can vary between 0.5 and 2°C. The outside temperature threshold value where the system is turned on to allow outside air in is

$$T_{FC} = T_{OS} - \Delta T$$

That will determine the chiller utilization factor:

$$\chi = \begin{cases} 1 & \text{for } T_{OS} \geq T_{FC} \\ 0 & \text{for } T_{OS} < T_{FC} \end{cases} \tag{10.10}$$

For free air cooling, the utilization factor χ is zero (chiller can be turned off) for as long as the outside air temperature is lower than T_{FC} (Fig. 10.5a). The time period can be calculated based on hourly outside weather data when the temperature is below the set point temperature (T_S).

10.6 CONTROL SYSTEMS

10.6.1 Software Platform

The main software application resides on a server, which can be a cloud instance or a local machine inside the DC. Such application, which is available for general DC usage [9], contains all the required software components for a full solution, including a graphical interface, a database, and a repository. The information contained in the main software is very comprehensive, and it includes the "data layout" of the DC, which is a model representing all the detailed physical characteristics of interest of the DC—covering from location of IT equipment and sensors to power ratings of servers and CRACs. The data layout provides specific information for the control algorithm, e.g. the power rating of CRACs (to estimate their utilization) and their location (to decide which unit to turn on or off).

In addition, the main software manages all the sensor data, which allows it to perform basic analysis, like CRAC utilization, simple threshold alarm for sensor readings, or flagging erroneous sensor readings (out-of-range or physically impossible values). The application can run more advanced analyses, like the CFD models that permit to pinpoint hot spots, estimate cooling airflow, and delineate CRAC influence zones [18].

The control algorithm is based on underutilized CRACs and events in the DC. The CRACs can be categorized as being in two states, standby or active, based on their utilization level—e.g. by setting a threshold level below which a CRAC is considered redundant. The CRAC utilization is proportional to the difference between air return and supply temperatures [19]. An underutilized CRAC is wasting energy by running the blowers to move air while providing minimum cooling to the DC. Figure 10.4b and c shows an example of the air return and supply temperatures in two different CRACs during a period of one week. In that figure it is clearly noticeable that CRAC in Fig. 10.4b is being underutilized (since there are only a couple of degrees of temperature difference between air return and supply temperatures), while the CRAC in Fig. 10.4c is not underutilized (since there are around 13°C difference between air return and supply temperatures).

Sample graphs quantifying the CRAC utilization are shown in Figure 10.5b and c. Given a $N + 1$ or $2N$ DC cooling design, some CRACs will be clearly underutilized due to overprovisioning, so those units are categorized to be on standby state. The CRAC control agents decide whenever a standby CRAC can become active (turned on) or inactive (turned off), and the software platform directly sends the commands to the CRACs via the remote controller actuators.

Note that a CRAC utilization level depends on the unit capacity, its heat exchange efficiency (supply and return air temperature), and air circulation patterns, which can be obtained through the CFD modeling as shown in [18]. Once the CRACs are categorized, the control algorithm is regulated by events within the DC as described next.

10.6.2 CRAC Control Agents

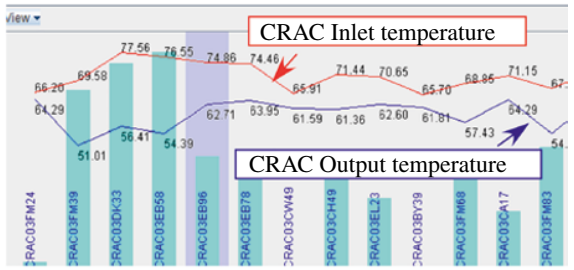
The CRAC categorization is an important grouping step of the control algorithm because, given the influence zones of a CRAC [10], the always active units provide the best trade-off between power consumption and DC cooling power (i.e. these CRACs are the least underutilized ones).

The CRAC discrete control mechanism is based on a set of events that can trigger an action (Fig. 10.6). Once having the infrastructure that provides periodic sensor data stream, an optimal method is implemented to control the CRACs in a DC. As mentioned, the first step is to identify underutilized CRACs; such CRACs are turned *off* sequentially at specified and configurable times as defined by the CRAC control agents. Given that such CRACs are underutilized, the total cooling power of the DC will remain almost the same if not slightly cooler, depending on the threshold used to categorize a CRAC as standby. If any DC event (e.g. a hot spot, as described below) occurs after a CRAC has been turned off, then such unit is turned back on.

(a)

Data centers (DC)	Heat load (kW)	Chiller efficiency	Average temp (°C)	Annual potential savings (kW)	Annual potential savings (%)
DC1	4770	3.0	16	1950	41
DC2	1603	7.3	7	590	36
DC3	2682	7.0	7	975	36
DC4	2561	3.4	6	2430	94
DC5	1407	3.5–5.9	11	675	47
DC6	2804	3.5	15	1320	47
DC7	3521	3.5–6.9	12	1550	44
DC8	1251	3.5	11	965	77

(b)



(c)

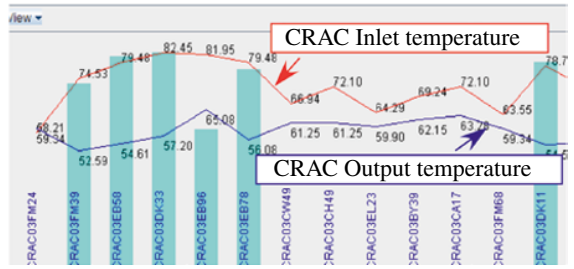


FIGURE 10.5 (a) The energy savings potential for air-side economized data centers based on data center performance and (b) CRAC utilization levels for normal operation and (c) CRAC utilization levels when DC is under the distributed control mechanism.

Regarding practical implementation concerns, the categorization of CRACs can be performed periodically or whenever there are changes in the DC, for example, an addition or removal of IT equipment, racks, etc. or rearrangements of the perforated tiles to adjust cooling.

Once the CRACs are off (standby state), the control agents monitor the data from the sensor network and check if they cross predefined threshold values. Whenever the threshold is crossed, an event is created, and an appropriate control command, e.g. turn on a CRAC, is executed. Figure 10.7 illustrates the flow diagram of the CRAC control agents.

The basic events that drive the turning *on* of an inactive CRAC are summarized in Figure 10.6a and b. The control events can be grouped in three categories:

1. Sensor measurements: Temperature, pressure, flow, corrosion, etc. For example, when a hot spot emerges (high temperature—above a threshold—in a localized area); very low pressures in plenum in DC (e.g. below the required level to push enough cool air to the top servers in the racks); very large corrosion rate measured by sensor (air intake)
2. Communication links: For example, no response from a remote controller, a relay computer (if applicable), or a sensor network gateway, or there is any type of network disruption
3. Sensor or device failure: For example, no sensor reading or out-of-bounds measurement value (e.g. physically impossible value); failure of an active CRAC (e.g. no airflow measured when the unit should be active)

The control agent can determine the location of an event within the DC layout layers in the software platform—such data layer stores the location information for all motes, servers, CRACs, and IT equipment. Thus, when activating a CRAC in order to address an event, the control agent selects the CRAC with the closest geometric distance to where the event occurred. Alternatively, the control agent could use the CRAC influence zone map (which is an outcome of the CFD capabilities of the main software [18]) to select a unit to become active. Influence zone is the area where the impact of a CRAC is most dominant, based on airflow and underfloor separating wall.

Once a CRAC is turned on, the DC status is monitored for some period by the control agent (i.e. the time required to increase the DC cooling power), and the initial alarm raised by an event will go inactive if the event is resolved. If the event continues, then the control agent will turn on another CRAC. The control agent will follow this pattern until all the CRACs are active—at which point no more redundant cooling power is available.

Furthermore, when the initial alarm becomes inactive, then, after a configurable waiting period, the control agent will turn off the CRAC that has been activated (or multiple CRACs if several units were turned on). This process will occur in a sequential approach, i.e. one unit at a time, while the control agents keep monitoring for the status of events. The sequence of turning units off can be configured, for example, units will only be turned off during business hours, or two units at a time will be turned off, or the interval between units being turned off can be adjusted, etc.

The events defined in Figure 10.6 are weighted by severity and reported accordingly, e.g. a single sensor event triggers no change, but two sensor events will turn on one CRAC. Once a standby CRAC has been turned *on*, it will remain *on* for a specified time after the event (e.g. 1 hour, a time that is dependent on the thermal mass of the DC or how fast the DC can respond to cooling) that caused it to become *on* in the first place has disappeared. These weights and waiting mechanisms provide a type of hysteresis loop in order to avoid frequent turning on and off CRACs.

In addition to the control algorithms, the CRAC control system includes a fail-safe mechanism, which is composed of watchdog timers. Such mechanism becomes active in case a control agent fails to perform periodical communication, and its purpose is to turn on standby CRACs. Also note that for

(a)

Event type	Description
Sensor (S)	Sensor measurement
Communication (C)	Communication failure
Failure (F)	Failure of control system

(b)

T-events

Number of thermal sensor readings above CRAC threshold	Event weight	Action
Lower than 1	0	
Between 2–4	1	Turn 1 of closest CRAC on
Between 4–6	2	Turn 2 of the closest CRACs on
Between 6–8	3	Turn 3 of the closest CRACs on
Above 8	4	Turn all CRACs on

P-events

Number of pressure sensor readings above CRAC threshold	Event weight	Action
Above 10	0	
Between 8–10	1	Turn 1 of closest CRAC on
Between 6–8	2	Turn 2 of the closest CRACs on
Between 4–6	3	Turn 3 of the closest CRACs on
Below 4	4	Turn all CRACs on

F-events

Number of flow sensor on active CRACs are "OFF"	Event weight	Action
16	0	Turn 1 of closest CRAC on
15	1	Turn 2 of the closest CRACs on
14	2	Turn 3 of the closest CRACs on
Below 13	3	Turn all CRACs on

FIGURE 10.6 (a) Three different events (sensor, communication, and failure) that are recorded by monitoring system and initiate a CRAC response and (b) different sensors' occurrence response and corresponding action..

manual operation, each CRAC unit is fitted with an override switch, which allows an operator to manually control a CRAC, thus bypassing the distributed control system.

10.7 QUANTIFIABLE ENERGY SAVINGS POTENTIAL

10.7.1 Validation for Free Cooling

In case partial free air cooling is used, the chiller utilization can be between 0 and 1 depending on the ratio of outside and indoor air used for cooling.

As a case study, DCs in eight locations were evaluated for the potential energy savings coming from air-side economization. Weather data were analyzed for two consecutive prior years to establish a baseline, and the numbers of hours when temperature falls below T_S, the set point temperature, were calculated for each year. The value for Ts is specified in Figure 10.5a for each DC along with the heat load and COP_{Chill}. We note that a high value of COP_{Chill} is desirable and values between 1 and 5 are considered poor metrics, while a value of 8 is a very good value.

Air quality measurements were started 6 months before the study and continue up till today. Each DC has at least one silver and one copper corrosion sensors. The copper corrosion sensor readings are in general less than 50 Å/month for the period of study and are not further discussed here. Silver sensors show periodic large changes as

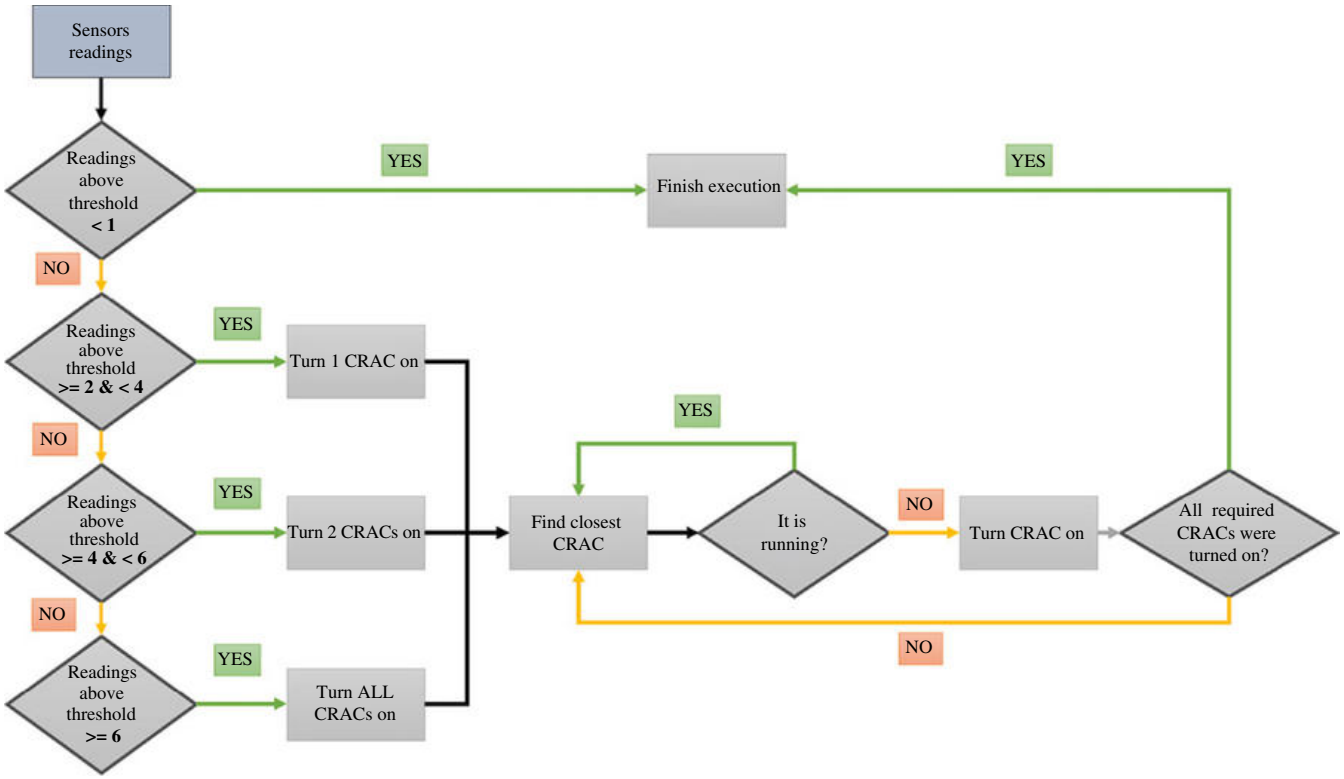


FIGURE 10.7 Flow diagram of the CRAC control agents based on sensor events.

illustrated in Figure 10.4a. The energy savings potential for the eight DCs are summarized in Figure 10.5a. These values assume that air quality is contamination-free and the only limitations are set by temperature and relative humidity (the second assumption is that when the outside air relative humidity goes above 80%, mechanical cooling will be used). The potential savings of DCs are dependent on the geographical locations of the DC; most of the DCs can reduce energy consumption by 20% or higher in a moderate climate zone.

10.7.2 Validation for CRAC Control

Figure 10.5b shows the status of the DC during normal operations—without the distributed control system enabled. In this state there are 2 CRACs off, and there are several units being largely underutilized—e.g. the leftmost bar, with 6% utilization, whose air return temperature is 19°C and air supply temperature is 18.5°C.

Once the distributed control system is enabled, as shown in Figure 10.5c, and after steady state is reached, seven CRACs are turned off—the most underutilized ones—and the utilization metric of the remaining active CRACs increases, as expected. Since, at the beginning, 2 CRACs were already normally off, a total of 5 additional CRACs were turned off by the control agent.

Given the maximum total active cooling capacity numbers and the total heat load of the DC in the previous subsection, having 8 CRACs active will provide enough cooling to the DC. The underfloor pressure slightly dropped after the additional 5 CRACs were turned off, although the resulting steady-state pressure was still within acceptable ranges for the DC. If the pressure had gone under the defined lower threshold, a standby CRAC would have been turned back on by the control agents (represented by an event as outlined in Fig. 10.6a and b).

Note that in this representative DC scenario, the optimal number of active CRACs is in the sense of keeping the same average return temperature at all the CRACs. Such metric is equivalent to maintaining a given cooling power within the DC, i.e. average inlet temperature at all the servers or racks. This optimality definition has as a constraint using the minimum number of active CRAC units along with having no DC events (as defined in the previous section, e.g. hot spots). Other optimality metric that could be used is maintaining the average under constant plenum pressure.

As a result, by turning off the 5 most underutilized CRACs in this DC, the average supply temperature decreased by 2°C. For a medium-size DC like this, the potential savings of keeping the five CRACs off are more than $60k/year calculated at a price of 10 cents/kWh.

10.8 CONCLUSIONS

Wireless sensor networks offer the advantage of both dense spatial and temporal monitoring across very large facilities with the potential to quickly identify hot spot locations and respond to those changes by adjusting CRAC operations. The wireless sensor networks enable dynamic assessment of DC's environments and are essential part of real-time sensor analytics that are integrated in control loops that can turn on/off CRACs. A dense wireless sensor network enables a more granular monitoring of DCs that can lead to substantial energy savings compared with few sensor-based facility-wide cooling strategies. Two different methods of energy savings are presented: free air cooling and discrete control of CRAC units. Turning on/off CRACs and combined with outside air cooling can be implemented to maximize energy efficiency. The sensor network and control loop analytics can also integrate information from DC equipment to improve energy efficiency while ensuring the reliable operation of IT servers.

REFERENCES

[1] Dunlap K, Rasmussen N. The advantages of row and rack-oriented cooling architectures for data centers. West Kingston: Schneider Electric ITB; 2006. APC White Paper-Schneider #130.

[2] Rajesh V, Gnanasekar J, Ponmagal R, Anbalagan P. Integration of wireless sensor network with cloud. Proceedings of the 2010 International Conference on Recent Trends in Information, Telecommunication and Computing, India; March 12–13, 2010. p. 321–323.

[3] Ilyas M, Mahgoub I. *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*. CRC Press; 2004.

[4] Jun J, Sichitiu ML. The nominal capacity of wireless mesh networks. IEEE Wirel Commun 2003;10:8–14.

[5] Hamann HF, et al. Uncovering energy-efficiency opportunities in data centers. IBM J Res Dev 2009;53:10:1–10:12.

[6] Gungor VC, Hancke GP. Industrial wireless sensor networks: challenges, design principles, and technical approaches. IEEE Trans Ind Electron 2009;56:4258–4265.

[7] Gungor VC, Lu B, Hancke GP. Opportunities and challenges of wireless sensor networks in smart grid. IEEE Trans Ind Electron 2010;57:3557–3564.

[8] Klein L, Singh P, Schappert M, Griffel M, Hamann H. Corrosion management for data centers. Proceedings of the 2011 27th Annual IEEE Semiconductor Thermal Measurement and Management Symposium, San Jose, CA; March 20–24,2011, p. 21–26.

[9] Singh P, Klein L, Agonafer D, Shah JM, Pujara KD. Effect of relative humidity, temperature and gaseous and particulate contaminations on information technology equipment reliability. Proceedings of the ASME 2015 International Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Microsystems collocated with the ASME 2015 13th International Conference on Nanochannels, Microchannels, and Minichannels, San Francisco, CA; July 6–9, 2015.

[10] ASHRAE: American Society of Heating, Refrigerating and Air-Conditioning Engineers. 9.2011 gaseous and particulate contamination guidelines for data centers. ASHRAE J 2011.

[11] Klein LJ, Bermudez SA, Marianno FJ, Hamann HF, Singh P. Energy efficiency and air quality considerations in airside economized data centers. Proceedings of the ASME 2015 International Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Microsystems collocated with the ASME 2015 13th International Conference on Nanochannels, Microchannels, and Minichannels, San Francisco, CA; July 6–9, 2015.

[12] Klein LI, Manzer DG. Real time numerical computation of corrosion rates from corrosion sensors. Google Patents; 2019.

[13] Zhang H, Shao S, Xu H, Zou H, Tian C. Free cooling of data centers: a review. Renew Sustain Energy Rev 2014;35:171–182.

[14] Meijer GI. Cooling energy-hungry data centers. Science 2010;328:318–319.

[15] Siriwardana J, Jayasekara S, Halgamuge SK. Potential of air-side economizers for data center cooling: a case study for key Australian cities. Appl Energy 2013;104:207–219.

[16] Oró E, Depoorter V, Garcia A, Salom J. Energy efficiency and renewable energy integration in data centres. Strategies and modelling review. Renew Sustain Energy Rev 2015;42:429–445.

[17] Stanford HW, III. *HVAC Water Chillers and Cooling Towers: Fundamentals, Application, and Operation*. CRC Press; 2016.

[18] Lopez V, Hamann HF. Measurement-based modeling for data centers. Proceedings of the 2010 12th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, Las Vegas, NV; June 2–5 2010. p. 1–8.

[19] Hamann HF, López V, Stepanchuk A. Thermal zones for more efficient data center energy management. Proceedings of the 2010 12th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, Las Vegas, NV; June 2–5, 2010. p. 1–6.

11

ASHRAE STANDARDS AND PRACTICES FOR DATA CENTERS

Robert E. McFarlane[1,2,3,4]

[1] *Shen Milsom & Wilke LLC, New York, New York, United States of America*
[2] *Marist College, Poughkeepsie, New York, United States of America*
[3] *ASHRAE TC 9.9, Atlanta, Georgia, United States of America*
[4] *ASHRAE SSPC 90.4 Standard Committee, Atlanta, Georgia, United States of America*

11.1 INTRODUCTION: ASHRAE AND TECHNICAL COMMITTEE TC 9.9

Many reputable organizations and institutions publish a variety of codes, standards, guidelines, and best practice documents dedicated to improving the performance, reliability, energy efficiency, and economics of data centers. Prominent among these are publications from ASHRAE—*The American Society of Heating, Refrigeration and Air-Conditioning Engineers*. ASHRAE [1], despite the nationalistic name, is actually international and publishes the most comprehensive range of information available for the heating, ventilation, and air-conditioning (HVAC) industry. Included are more than 125 ANSI standards; at least 25 guidelines; numerous white papers; the four-volume *ASHRAE Handbook*, which is considered the "bible" of the HVAC industry; and the *ASHRAE Journal*.

The documents relating to data centers have originated primarily in ASHRAE Technical Committee *TC 9.9* [2], whose formal name is *Mission-critical Facilities, Data Centers, Technology Spaces*, *and Electronic Equipment*. *TC 9.9* is the largest of the 96 ASHRAE TCs, with more than 250 active members. Its history dates back to 1998 when it was recognized that standardization of thermal management in the computing industry was needed. This evolved into an ASHRAE Technical Consortium in 2002 and became a recognized ASHRAE Technical Committee in 2003 under the leadership of Don Beaty, whose engineering firm has designed some of the best known data centers in the world, and Dr. Roger Schmidt, an IBM Distinguished Engineer and IBM's Chief Thermal Engineer, now retired, but continuing his service to the industry on the faculty of Syracuse University. Both remain highly active in the committee's activities.

11.2 THE GROUNDBREAKING ASHRAE "THERMAL GUIDELINES"

ASHRAE TC 9.9 came to prominence in 2004 when it published the *Thermal Guidelines for Data Processing Environments,* the first of the ASHRAE *Datacom Series*, which consists of 14 books at the time of this book publication. For the first time, *Thermal Guidelines* gave the industry a bona fide range of environmental temperature and humidity conditions for data center computing hardware. Heretofore, there were generally accepted numbers based on old Bellcore/Telcordia data that was commonly used for "big iron" mainframe computing rooms. Anyone familiar with those earlier days of computing knows that sweaters and jackets were de rigueur in the frigid conditions where temperatures were routinely kept at 55°F or 12.8°C and relative humidity (RH) levels were set to 50%. As the demand grew to reduce energy consumption, it became necessary to reexamine legacy practices. A major driver of this movement was the landmark 2007 US Department of Energy study on data center energy consumption in the United States and its prediction that the data processing industry would outstrip generating capacity within 5 years if its growth rate continued. The industry took note, responded and, thankfully, that dire prediction did not

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

materialize. But with the never-ending demand for more and faster digital capacity, the data processing industry cannot afford to stop evolving and innovating in both energy efficiency and processing capacity. ASHRAE continues to be one of the recognized leaders in that endeavor, *The Green Grid (TGG)* [3] being the other major force. These two industry trendsetters jointly published one of the *Datacom Series* books, described later in this chapter, detailing the other landmark step in improving data center energy efficiency—the *Power Usage Effectiveness* or *PUE*™ metric developed by *TGG* and now universally accepted.

But changing legacy practices is never easy. When the *Thermal Guidelines* [7] first appeared, its recommendations violated many existing warranties, as well as the "recommended conditions" provided by manufacturers with their expensive computing equipment. But *Thermal Guidelines* had not been developed in a vacuum. Dr. Roger Schmidt, due to his prominence in the industry, was able to assemble designers from every major computing hardware manufacturer to address this issue. Working under strict nondisclosure, and relying on the highly regarded, noncommercial ethics of ASHRAE, they all revealed their actual equipment environmental test data to each other. It became clear that modern hardware could actually operate at much higher temperatures than those generally recommended in manufacturer's data sheets, with no measurable reductions in reliability, failure rates, or computing performance. As a result, ASHRAE TC 9.9 was able to publish the new *recommended* and *allowable* ranges for Inlet Air Temperatures to computing hardware, with full assurance that their use would not violate warranties, impair performance, or reduce equipment life.

The guidelines are published for different classifications of equipment. The top of the *recommended* range is for the servers and storage equipment commonly used in data centers (Class A1) and is set at 27°C (80.6°F). This was a radical change for the industry that, for the first time, had a validated basis for cooling designs that would not only ensure reliable equipment operation but also result in enormous savings in energy use and cost. It takes a lot of energy to cool air, so large operations quickly adopted these new guidelines since energy costs comprise a major portion of their operating expenses. Many smaller users, however, initially balked at such a radical change, but slowly began to recognize both the importance and value of these guidelines provide in reducing energy consumption.

The *Thermal Guidelines* book is in its fourth edition at the time of this printing, with more equipment classifications and ranges added that are meant to challenge manufacturers to design equipment that can operate at even higher temperatures. Equipment in these higher classes could run in any climate zone on Earth with *no* need for mechanical cooling at all. Considering the rate at which this industry evolves, equipment meeting these requirements will likely be commonly available before this textbook is published and may even become "standard" before it is next revised.

Some enterprise facilities even operate above the *Recommended* temperature ranges in order to save additional energy. Successive editions of the *Thermal Guidelines* book have addressed these practices by adding detailed data, along with methods of statistically estimating potential increased failure rates, when computing hardware is consistently subjected to the higher temperatures. Operations that do this tend to cycle hardware faster than any increased incidence of failure, making the resulting energy savings worthwhile. However, when looking at the temperature ranges in each classification, it is still important to understand several things:

- The original and primary purpose of developing guidelines for increased temperature operation was to save energy. This was meant to occur partly through a reduction in refrigeration energy, but mainly to make possible more hours of "free cooling" in most climate zones each year. "Free cooling" is defined as the exclusive use of outside air for heat removal, with no mechanical refrigeration needed. This is possible when the outside ambient air temperature is lower than the maximum inlet temperature of the computing hardware.
- The upper limit of 27°C (80.6°F) for Class A1 hardware was selected because it is the temperature at which most common servers begin to significantly ramp up internal fan speeds. Fan energy essentially follows a cube-law function, meaning that doubling fan speed can result in eight times the energy use ($2 \times 2 \times 2$). Therefore, it is very possible to save cooling energy by increasing server inlet temperature above upper limit of the *Recommended* range, but to offset, or even exceed, that energy savings with increased equipment fan energy consumption.
- It is also important to recognize that the *Thermal Envelope* (the graphical representation of temperature and humidity limits in the form of what engineers call a psychrometric chart) and its high and low numerical limits are based on *inlet temperature* to the computing hardware (Fig. 11.1). Therefore, simply increasing air conditioner set points so as to deliver higher temperature air in order to save energy may not have the desired result. Cooling from a raised access floor provides the best example of oversimplifying the interpretation of the thermal envelope. Warm air rises (or, in actuality, cool air, being more dense, falls, displacing less dense warmer air and causing it to rise). Therefore, pushing cool air through a raised floor airflow panel, and expecting it to rise to the full height of a rack cabinet, is actually contrary to the laws of physics. As a result, maintaining uniform temperature from bottom to top of the rack with under-floor cooling is impossible.

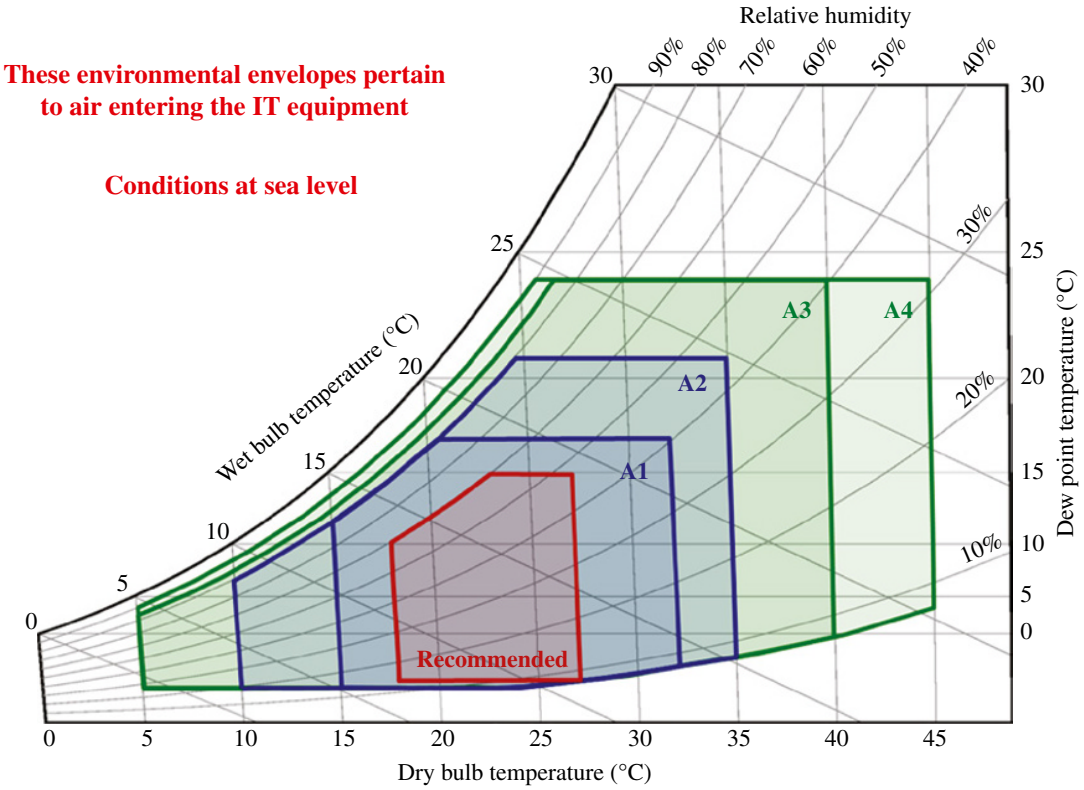


FIGURE 11.1 Environmental guidelines for air-cooled equipment. 2015 Thermal Guidelines SI Version Psychrometric Chart. Source: ©ASHRAE www.ashrae.org.

There are ways to significantly improve the situation, the more useful being "containment." But if you were to deliver 80°F (27°C) air from the floor tile, even within the best possible air containment environment, it could easily be 90°F (32°C) by the time it reached the top of the rack. Without good containment, you could see inlet temperatures at the upper level equipment of 100°F (38°C). In short, good thermal design and operation are challenging.

- The *Thermal Guidelines* also specifies "*Allowable* temperature ranges," which are higher than the *Recommended* ranges. These "allowable" ranges tell us that, in the event of a full or partial cooling failure, we need not panic. Computing hardware can still function reliably at a higher inlet temperature for several days without a significant effect on performance or long-term reliability.
- The *Thermal Guidelines* also tell us that, when using "free cooling," it is not necessary to switch to mechanical refrigeration if the outside air temperature exceeds the *Recommended* limit for only a few hours of the day. This means that "free cooling" can be used more continuously, minimizing the number of cooling transfers, each of which has the potential of introducing a cooling failure.

11.3 THE *THERMAL GUIDELINES* CHANGE IN HUMIDITY CONTROL

The other major change in *Thermal Guidelines* was the recommendation to control data center moisture content on dew point ("DP") rather than on relative humidity ("RH"), which had been the norm for decades. The reason was simple, but not necessarily intuitive for non-engineers. DP is also known as "absolute humidity." It is the amount of moisture in the air measured in grains of water vapor per unit volume. It is essentially uniform throughout a room where air is moving, which is certainly the case in the data center environment. In short, it's called "absolute humidity" for an obvious reason.

DP is the temperature, either Fahrenheit or Celsius, at which water vapor in the air condenses and becomes liquid. Very simply, if the dry-bulb temperature (the temperature measured with a normal thermometer), either in a room or on a surface, is higher than the DP temperature (also known as wet-bulb temperature measured with a special thermometer), the moisture in the air will remain in the vapor state and will not condense. However, if the dry-bulb temperature falls to where it equals the DP temperature, the water vapor turns to liquid. Outdoors, it may become dew on the cool lawn, or water on cool car windows, or it will turn to rain if the temperature falls enough in higher levels of the

atmosphere. Within the data center, it will condense on equipment, which is obviously not good. The concept is actually quite simple.

RH, on the other hand, is measured in percent and, as its name implies, is related to temperature. Therefore, even when the actual amount of moisture in the air is uniform (same DP temperatures everywhere), the RH number will be considerably different in the hot and cold aisles. It will even vary in different places within those aisles because uniform temperature throughout a space is virtually impossible to achieve. Therefore, when humidity is controlled via RH measurement, the amount of moisture either added to or removed from the air depends on where the control points are located and the air temperatures at those measurement points. These are usually in the air returns to the air conditioners, and those temperatures can be considerably different at every air conditioner in the room. The result is that one unit may be humidifying while another is dehumidifying. It also means that energy is being wasted by units trying to oppose each other and that mechanical equipment is being unnecessarily exercised, potentially reducing its service life.

When humidity is controlled on DP, however, the temperature factor is removed, and every device is seeing the same input information and working to maintain the same conditions. That's both more energy and more operationally efficient. Of course, modern air conditioners are processor controlled and can intercommunicate to avoid crossed-purpose operation. But both efficiency and accuracy of control are still much better when DP is used as the benchmark.

11.4 A NEW UNDERSTANDING OF HUMIDITY AND STATIC DISCHARGE

The next radical change to come out of the TC 9.9 Committee's work was a big revision in the humidity requirement part of the *Thermal Guidelines*. The concern with humidity has always been one of preventing static discharge, which everyone has experienced on cold winter days when the air is very dry. Static discharges can be in the millions of electron volts, which would clearly be harmful to microelectronics. But there was no real data on how humidity levels actually relate to static discharge in the data center environment and the equipment vulnerability to it. Everything is well grounded in a data center, and high static generating materials like carpeting do not exist. Therefore, TC 9.9 sponsored an ASHRAE-funded research project into this issue, which, in 2014, produced startling results. The study *The Effect of Humidity on Static Electricity Induced Reliability Issues of ICT Equipment in Data Centers* [4] was done at the Missouri University of Science and Technology under the direction of faculty experts in static phenomena. It examined a wide range of combinations of floor surfaces, footwear and even cable being pulled across a floor, under a wide range of humidity conditions. The conclusion was that, even at RH levels as low as 8%, static discharge in the data center environment was insufficient to harm rack-mounted computing hardware. This is another enormous change from the 50% RH level that was considered the norm for decades. Again, this publication caused both disbelief and concern. It also created a conflict in how data center humidity is measured, since the ASHRAE recommendation is to control DP or "absolute" humidity, and static discharge phenomena are governed by RH. But an engineer can easily make the correlation using a psychrometric chart, and the *Thermal Guidelines* book provides an easy method of relating the two. So the ASHRAE recommendation is still to control humidity on DP.

This further change in environmental considerations provides increased potential for energy reduction. The greatest opportunities to utilize "free cooling" occur when the outside air is cool, which also correlates with dryer air since cool air can retain less moisture than warm air. It requires considerable energy to evaporate moisture, so adding humidity to dry air is very wasteful. Therefore, this important ASHRAE information provides a further opportunity to save energy and reduce operating costs without unduly exposing critical computing equipment to an increased potential for failure. The only real caveat to very low humidity operation is the restriction of a particular type of plastic-soled footwear. The other caveat, which should be the "norm" anyway, is that grounding wrist straps *must* be used when working *inside* the case of any piece of computing hardware. The study was to assess the potential for damage to *mounted* equipment in a data center. Equipment is always vulnerable to static damage when the case is opened, regardless of humidity level.

11.5 HIGH HUMIDITY AND POLLUTION

But the *Thermal Guidelines* also stipulate an upper limit of 60% for RH. While much lower humidity levels have been proven acceptable, humidity can easily exceed 60% RH in the hot, humid summers experienced in many locales. That outside air should not be brought into the data center without being conditioned. The reason is a relatively new one, where humidity combines with certain contaminants to destroy connectors and circuit boards, as detailed in the next paragraph. The upper limit RH specification is to avoid that possibility.

Contamination is the subject of another one of the TC 9.9 Datacom books in the series described in more detail below. In essence, it demonstrates that above 60% RH, the high moisture level combines with various environmental contaminants to produce acids. Those acids, primarily sulfuric and hydrochloric, can eat away at tiny circuit board lands and connector contacts, particularly where they are soldered.

This concern results from the European Union's *RoHS* Directive [5] (pronounced "*Ro*Hass"). RoHS stands for the *Restriction of Hazardous Substances* in electrical and electronic equipment. It was first issued in 2002 and was recast in 2011. Lead, which was historically a major part of electrical solder, is one of the more than 100 prohibited substances. Virtually every manufacturer of electronic equipment now follows RoHS guidelines, which means that lead-silver solder can no longer be used on circuit boards. Since lead is an inert element, but silver is not, connections are now susceptible to corrosive effects that did not previously affect them, and the number of circuit board and connector failures has skyrocketed as a result. These airborne contaminants, such as sulfur dioxide compounds, are a less serious concern in most developed countries, but in some parts of the world, and anywhere in close proximity to certain chemical manufacturing plants or high traffic roadways, they can be. So it is best to observe the 60% maximum RH limit regardless. This simply means that either mechanical refrigeration or desiccant filters may be required to remove moisture when using air-side free cooling in high humidity environments. And charcoal filters may also be recommended for incoming air in environments with high levels of gaseous contaminants.

All these parameters have been combined into both the psychrometric chart format commonly used by engineers and a tabular format understandable to everyone. There is much more detail in the *Thermal Guidelines* book, but these charts provide a basic understanding of the environmental envelope ranges.

11.6 THE ASHRAE "DATACOM SERIES"

The beginning of this chapter noted that *Thermal Guidelines* was the first of the ASHRAE TC 9.9 *Datacom Series,* comprised of 14 books (see Further Reading) at the time of this book publication. The books cover a wide range of topics relevant to the data center community, and many have been updated since original publication, in some cases several times, to keep pace with this fast-changing industry. The *Datacom* series is written to provide useful information to a wide variety of users, including those new to the industry, those operating and managing data centers, and the consulting engineers who design them. Data centers are very unique and highly complex infrastructures in which many factors interact, and change is a constant as computing technology continues to advance. It is an unfortunate reality that many professionals are not aware of the complexities and significant challenges of these facilities and are not specifically schooled in the techniques of true "mission-critical" design. When considering professionals to design a new or upgraded data center, an awareness of the material in the ASHRAE publications can be useful in selecting those who are truly qualified to develop the infrastructure of a high-availability computing facility.

The detail in these books is enormous, and the earlier books in the series contain chapters providing fundamental information on topics such as contamination, structural loads, and liquid cooling that are covered in depth in later publications. A summary of each book provides guidance to the wealth of both technical and practical information available in these ASHRAE publications. All books provide vendor-neutral information that will empower data center designers, operators, and managers to better determine the impact of varying design and operating parameters, in particular encouraging innovation that maintains reliability while reducing energy use. In keeping with the energy conservation and "green" initiatives common to the book topics, the books are available in electronic format, but many of the paper versions are printed on 30% postconsumer waste using soy-based inks. Where color illustrations are utilized, the downloadable versions are preferable since the print versions are strictly in black and white. All editions listed below are as of the date of this book publication, but the rapidity with which this field changes means that the books are being constantly reviewed and later editions may become available at any time.

11.6.1 Book #1: *Thermal Guidelines for Data Processing Environments, 4th Edition* [6]

This book should be required reading for every data center designer, operator, and facility professional charged with maintaining a computing facility. The fundamentals of thermal envelope and humidity control included in this landmark book have been covered above, but there is much more information in the full publication. The ASHRAE summary states: "*Thermal Guidelines for Data Processing Environments* provides a framework for improved alignment of efforts among IT equipment (ITE) hardware manufacturers (including manufacturers of computers, servers, and storage products), HVAC equipment manufacturers, data center designers, and facility operators and managers. This guide covers five primary areas:

- Equipment operating environment guidelines for air-cooled equipment
- Environmental guidelines for liquid-cooled equipment
- Facility temperature and humidity measurement
- Equipment placement and airflow patterns
- Equipment manufacturers' heat load and airflow requirements reporting."

In short, *Thermal Guidelines* provides the foundation for all modern data center design and operation.

Equipment Environment Specifications for Air Cooling							
	Product Operation[b,c]					Product Power Off[c,d]	
Class [a]	Dry-Bulb Temperature[e,g], °C	Humidity Range, Noncondensing[h,i,k,l]	Maximum Dew Point[k], °C	Maximum Elevation[e,j,m], m	Maximum Rate of Change[f], °C/h	Dry-Bulb, Temperature, °C	Relative Humidity[k], %
Recommended (Suitable for all four classes; explore data center metrics in this book for conditions outside this range.)							
A1 to A4	18 to 27	–9°C DP to 15°C DP and 60% rh					
Allowable							
A1	15 to 32	–12°C DP and 8% rh to 17°C DP and 80% rh	17	3050	5/20	5 to 45	8 to 80
A2	10 to 35	–12°C DP and 8% rh to 21°C DP and 80% rh	21	3050	5/20	5 to 45	8 to 80
A3	5 to 40	–12°C DP and 8% rh to 24°C DP and 85% rh	24	3050	5/20	5 to 45	8 to 80
A4	5 to 45	–12°C DP and 8% rh to 24°C DP and 90% rh	24	3050	5/20	5 to 45	8 to 80
B	5 to 35	8% to 28°C DP and 80% rh	28	3050	N/A	5 to 45	8 to 80
C	5 to 40	8% to 28°C DP and 80% rh	28	3050	N/A	5 to 45	8 to 80

* For potentially greater energy savings, refer to the section "Detailed Flowchart for the Use and Application of the ASHRAE Data Center Classes" in Appendix C for the process needed to account for multiple server metrics that impact overall TCO.

a. Classes A3, A4, B, and C are identical to those included in the 2011 edition of *Thermal Guidelines for Data Processing Environments*. The 2015 version of the A1 and A2 classes have expanded RH levels compared to the 2011 version.
b. Product equipment is powered ON.
c. Tape products require a stable and more restrictive environment (similar to 2011 Class A1). Typical requirements: minimum temperature is 15°C, maximum temperature is 32°C, minimum RH is 20%, maximum RH is 80%, maximum dew point is 22°C, rate of change of temperature is less than 5°C/h, rate of change of humidity is less than 5% rh per hour, and no condensation.
d. Product equipment is removed from original shipping container and installed but not in use, e.g., during repair, maintenance, or upgrade.
e. Classes A1, A2, B, and C—Derate maximum allowable dry-bulb temperature 1°C/300 m above 900 m. Above 2400 m altitude, the derated dry-bulb temperature takes precedence over the recommended temperature. Class A3—Derate maximum allowable dry-bulb temperature 1°C/175 m above 900 m. Class A4—Derate maximum allowable dry-bulb temperature 1°C/125 m above 900 m.
f. For tape storage: 5°C in an hour. For all other ITE: 20°C in an hour and no more than 5°C in any 15 minute period of time. The temperature change of the ITE must meet the limits shown in the table and is calculated to be the maximum air inlet temperature minus the minimum air inlet temperature within the time window specified. The 5°C or 20°C temperature change is considered to be a temperature change within a specified period of time and not a rate of change. See Appendix K for additional information and examples.
g. With a diskette in the drive, the minimum temperature is 10°C (not applicable to Classes A1 or A2).
h. The minimum humidity level for Classes A1, A2, A3, and A4 is the higher (more moisture) of the –12°C dew point and the 8% rh. These intersect at approximately 25°C. Below this intersection (~25°C) the dew point (–12°C) represents the minimum moisture level, while above it, RH (8%) is the minimum.
i. Based on research funded by ASHRAE and performed at low RH, the following are the minimum requirements:
1) Data centers that have non-ESD floors and where people are allowed to wear non-ESD shoes may want to consider increasing humidity given that the risk of generating 8 kV increases slightly from 0.27% at 25% rh to 0.43% at 8% (see Appendix D for more details).
2) All mobile furnishing/equipment is to be made of conductive or static dissipative materials and bonded to ground.
3) During maintenance on any hardware, a properly functioning and grounded wrist strap must be used by any personnel who contacts ITE.
j. To accommodate rounding when converting between SI and I-P units, the maximum elevation is considered to have a variation of ±0.1%. The impact on ITE thermal performance within this variation range is negligible and enables the use of rounded values of 3050 m (10,000 ft).
k. See Appendix L for graphs that illustrate how the maximum and minimum dew-point limits restrict the stated relative humidity range for each of the classes for both product operations and product power off.
l. For the upper moisture limit, the limit is the minimum absolute humidity of the DP and RH stated. For the lower moisture limit, the limit is the maximum absolute humidity of the DP and RH stated.
m. Operation above 3050 m requires consultation with IT supplier for each specific piece of equipment.

FIGURE 11.2 Environmental guidelines for air-cooled equipment. 2015 Recommended and Allowable Envelopes for ASHRAE Classes A1, A2, A3, and A4, B and C. Source: ©ASHRAE www.ashrae.org.

11.6.2 Book #2: *IT Equipment Power Trends, 3rd Edition* [7]

Computing equipment has continued to follow Moore's law, formulated in 1964 by Roger Moore, then president of Intel. Moore predicted that the number of transistors on a chip would double every 18 months and believed this exponential growth would continue for as long as 10 years. It actually continued more than five decades, and began to slow only as nanotechnology approached a physical limit. When components cannot be packed any closer together, the lengths of microscopic connecting wires become a limiting factor in

processor speed. But each increase in chip density brings with it a commensurate increase in server power consumption and, therefore, in heat load.

While the fundamentals of energy efficient design are provided in *Thermal Guidelines*, long-term data center power and cooling solutions cannot be developed without good knowledge of both initial and future facility power requirements. Predicting the future in the IT business has always been difficult, but Dr. Schmidt was again able to assemble principal design experts from the leading ITE manufacturers to develop the ASHRAE *Power Trends* book. These people have first-hand knowledge of the technology in development, as well as what is happening with chip manufacturers and software developers. In short, they are in the best positions to know what can be expected in the coming years and were willing to share that information and insight with ASHRAE.

The book originally predicted growth rates for ITE to 2014 in multiple categories of type and form factor. At the time of this textbook publication, the *Power Trends* book and its charts have been revised twice, extending the predictions through 2025. The information can be used to predict future capacity and energy requirements with significant accuracy, enabling both power and cooling systems to be designed with minimal "first costs," as well as for logical, nondisruptive expansion, and with the minimum energy use necessary to serve actual equipment needs. The book can also help operators and facilities professionals predict when additional capacity will be needed so prudent investments can be made in preplanned capacity additions.

The third edition of this book also takes a different approach to presenting the information than was used in the previous publications. The purpose is to provide users with better insight into the power growth that can be expected in their particular computing facilities. The focus is now on the workloads and applications the hardware must run, which gives better insight into future power trends than focusing on equipment type and form factor alone. The major workloads analyzed include business processing, analytics, scientific, and cloud-based computing. Further, projections are provided for both rack power densities and annualized power growth rates and even for individual server and storage equipment components. These categories provide better insight into what is actually driving the change in ITE power consumption.

Under-designing anything is inefficient because systems will work harder than should be necessary. But under-designing cooling systems is particularly inefficient because compressors will run constantly without delivering sufficient cooling, in turn making server fans run at increased speed, all of which compounds the wasteful use of energy. Overdesign results in both cooling and UPS (uninterruptable power supply) systems operating in the low efficiency ranges of their capabilities, which wastes energy directly. This is particularly concerning with high-availability redundant configurations. Compounding the design problem is the way the power demands of IT hardware continue to change. While equipment has become significantly more efficient on a "watts per gigaflop" basis, both servers and storage equipment have still increased in both power usage and power density. This means that each cabinet of equipment has both higher power demands and greater cooling requirements. Modern UPS systems can be modular, enabling capacity to grow along with the IT systems so that capacity is matched to actual load. Cooling systems can be variable capacity as well, self-adjusting to demand when operated by the right distribution of sensors and controls.

11.6.3 Book #3: *Design Considerations for Datacom Equipment Centers, 2nd Edition* [8]

The design of computer rooms and telecommunications facilities is fundamentally different from the design of buildings and offices used primarily for human occupancy. To begin with, power densities can easily be 100 times what is common to office buildings, or even more. Further, data center loads are relatively constant day and night and all year-around, temperature and humidity requirements are much different than for "comfort cooling," and reliability usually takes precedence over every other consideration.

While the *Design Considerations* book is based on the information in *Thermal Guidelines* and *Power Trends*, it provides actual guidance in developing the design criteria and applying this information to the real world of data center design. The book begins with basic computer room cooling design practices (both air and liquid), which requires consideration of many interrelated elements. These include establishing HVAC load, selection of operating temperature, temperature rate of change, RH, DP, redundancy, systems availability air distribution, and filtration of contaminants. For those already experienced in designing and operating data centers, more advanced information is also provided on energy efficiency, structural and seismic design and testing, acoustical noise emissions, fire detection and suppression, and commissioning. But since a full data center consists of more than the actual machine room or "white space," guidance is also provided in the design of battery plants, emergency generator rooms, burn-in rooms, test labs, and spare parts storage rooms. The book does not, however, cover electrical or electronic system design and distribution.

11.6.4 Book #4: *Liquid Cooling Guidelines for Datacom Equipment Centers, 2nd Edition* [9]

This is one of the several books in the Datacom series that significantly expands information covered more generally in previous books. While power and the resulting heat loads have been increasing for decades, it is the power and heat

densities that have made equipment cooling increasingly difficult to accomplish efficiently. With more heat now concentrated in a single cabinet than existed in entire rows of racks not many years ago, keeping equipment uniformly cooled can be extremely difficult. Server cooling requirements, in particular, are based on the need to keep silicon junction temperatures within specified limits. Inefficient cooling can, therefore, result in reduced equipment life, poor computing performance, and greater demand on cooling systems to the point where they operate inefficiently as well. Simply increasing the number of cooling units, without thoroughly understanding the laws of thermodynamics and airflow, wastes precious and expensive floor space and may still not solve the cooling problem.

This situation is creating an increasing need to implement liquid cooling solutions. Moving air through modern high-performance computing devices at sufficient volumes to ensure adequate cooling becomes even more challenging as the form factors of the hardware continue to shrink. Further, smaller equipment packaging reduces the space available for air movement in each successive equipment generation. It has become axiomatic that conventional air cooling cannot sustain the continued growth of compute power. Some form of liquid cooling will be necessary to achieve the performance demands of the industry without resorting to "supercomputers," which are already liquid-cooled. It also comes as a surprise to most people, and particularly to those who are fearful of liquid cooling, that laptop computers have been liquid-cooled for several generations. They use a closed-loop liquid heat exchanger that transfers heat directly from the processor to the fan, which sends it to the outside. Failures and leaks in this system are unheard of.

Liquid is thousands of times more efficient per unit volume than air at removing heat. (Water is more than 3,500 times as efficient, and other coolants are not far behind that.) Therefore, it makes sense to directly cool the internal hardware electronics with circulating liquid that can remove large volumes of heat in small spaces and then transfer the heat to another medium such as air outside the hardware where sufficient space is available to accomplish this efficiently. But many users continue to be skeptical of liquid circulating anywhere near their hardware, much less inside it, with the fear of leakage permanently destroying the equipment. The *Design Considerations* book dispels these concerns with solid information about proven liquid cooling systems, devices such as spill-proof connectors, and examples of "best practices" liquid cooling designs.

The second edition of *Liquid Cooling Guidelines* goes beyond direct liquid cooling, also covering indirect means such as rear door heat exchangers (RDhX) and full liquid immersion systems. It also addresses design details such as approach temperatures, defines liquid and air cooling for ITE, and provides an overview of both chilled water and condenser water systems and how they interface to the liquid equipment cooling loops. Lastly, the book addresses the fundamentals of water quality conditioning, which is important to maintaining trouble-free cooling systems, and the techniques of thermal management when both liquid and air cooling systems are used together in the data center.

11.6.5 Book #5: *Structural and Vibration Guidelines for Datacom Equipment Centers, 1st Edition* [10]

This is another of the books that expands on information covered more generally in the books covering fundamentals.

As computing hardware becomes more dense, the weight of a fully loaded rack cabinet becomes problematic, putting loads on conventional office building structures that can go far beyond their design limits. Addressing the problem by spreading half-full cabinets across a floor wastes expensive real estate. Adding structural support to an existing floor, however, can be prohibitively expensive, not to mention dangerously disruptive to any ongoing computing operations.

When designing a new data center building or evaluating an existing building for the potential installation of a computing facility, it is important to understand how to estimate the likely structural loads and to be able to properly communicate that requirement to the architect and structural engineer. It is also important to be aware of the techniques that can be employed to solve load limit concerns in different types of structures. If the structural engineer doesn't have a full understanding of cabinet weights, aisle spacings, and raised floor specifications, extreme measures may be specified, which could price the project out of reality, when more realistic solutions could have been employed. *Structural and Vibration Guidelines* addresses these issues in four sections:

- The Introduction discusses "best practices" in the cabinet layout and structural design of these critical facilities, providing guidelines for both new buildings and the renovation of existing ones. It also covers the realities of modern datacom equipment weights and structural loads.
- Section 2 goes into more detail on the structural design of both new and existing buildings, covering the additional weight and support considerations when using raised access floors.
- Section 3 delves into the issues of shock and vibration testing for modern datacom equipment, and particularly for very high density hard disk drives that can be adversely affected, and even destroyed, by vibration.
- Lastly, the book addresses the challenges of seismic restraints for cabinets and overhead infrastructure when designing data centers in seismic zones.

11.6.6 Book #6: *Best Practices for Datacom Facility Energy Efficiency, 2nd Edition* [11]

This is a very practical book that integrates key elements of the previous book topics into a practical guide to the design of critical datacom facilities. With data center energy use and cost continuing to grow in importance, some locales are actually restricting their construction due to their inordinate demand for power in an era of depleting fuel reserves and the inability to generate and transmit sufficient energy.

With global warming of such concern, the primary goal of this book is to help designers and operators reduce energy use and life cycle costs through knowledgeable application of proven methods and techniques. Topics include environmental criteria, mechanical equipment and systems, economizer cycles, airflow distribution, HVAC controls and energy management, electrical distribution equipment, datacom equipment efficiency, liquid cooling, total cost of ownership, and emerging technologies. There are also appendices on such topics as facility commissioning, operations and maintenance, and actual experiences of the datacom facility operators.

11.6.7 Book #7: *High Density Data Centers—Case Studies and Best Practices, 2nd Edition* [12]

While most enterprise data centers still operate with power and heat densities not exceeding 7–10 kW per cabinet, many are seeing cabinets rise to levels of 20 kW, 30 kW, or more. Driving this density is the ever-increasing performance of datacom hardware, which rises year after year with the trade-off being higher heat releases. This has held true despite the fact that performance has generally grown without a linear increase in power draw. There are even cabinets in specialized computing operations (not including "supercomputers") with cabinet densities as high as 60 kW. When cabinet densities approach these levels, and even in operations running much lower density cabinets, the equipment becomes extremely difficult to cool. Operations facing the challenges of cooling the concentrated heat releases produced by these power densities can greatly benefit from knowledge of how others have successfully faced these challenges.

This book provides case studies of a number of actual high density data centers and describes the ventilation approaches they used. In addition to providing practical guidance from the experiences of others, these studies confirm that there is no one "right" solution to addressing high density cooling problems and that a number of different approaches can be successfully utilized.

11.6.8 Book #8: *Particulate and Gaseous Contamination in Datacom Environments, 2nd Edition* [13]

Cleanliness in data centers has always been important, although it has not always been enforced. But with smaller form factor hardware, and the commensurate restricted airflow, cleanliness has actually become a significant factor in running a "mission-critical" operation. The rate of air movement needed through high density equipment makes it mandatory to keep filters free of dirt. That is much easier if the introduction of particulates into the data center environment is minimized. Since data center cleaning is often done by specialized professionals, this also minimizes OpEx by reducing direct maintenance costs. Further, power consumption is minimized when fans aren't forced to work harder than necessary. There are many sources of particulate contamination, many of which are not readily recognized. This book addresses the entire spectrum of particulates and details ways of monitoring and reducing contamination.

While clogged filters are a significant concern, they can at least be recognized by visual inspection. That is not the case for damage caused by gaseous contaminants, which, when combined with high humidity levels, can result in acids that eat away at circuit boards and connections. As mentioned in the discussion of RoHS compliance and the changes it has made to solder composition, the result can be catastrophic equipment failures that are often unexplainable except through factory and laboratory analysis of the failed components.

The ASHRAE 60% RH limit for data center moisture content noted in the previous humidity discussion should not be a great a concern in most developed countries, where high levels of gaseous contamination are not generally prevalent. But anyplace that has high humidity should at least be aware. Unfortunately, there is no way to alleviate concerns without proper testing and evaluation. That requires copper and silver "coupons" to be placed in the environment for a period of time and then analyzed in a laboratory to determine the rate at which corrosive effects have occurred. The measurements are in angstroms (Å), which are metric units equal to 10^{-10} or one-ten-billionth of a meter.

Research referenced in the second edition of this book has shown that silver coupon corrosion at a rate of less than 200 Å/month is not likely to cause problems. Although this may sound like a very small amount of damage, when considered in terms of the thickness of circuit board lands, it can be a significant factor. But the even bigger problem is the deterioration of soldered connections, particularly from sulfur dioxide compounds. These can be present in relatively high concentrations where automobile traffic, fossil-fuel-fired power plants and boilers, and chemical plants exist. The sulfur compound gases combine with water vapor to create sulfuric acid that can rapidly eat away at silver-soldered connections and silver-plated contacts. As noted earlier in this chapter, the advent of RoHS, and its elimination of lead from solder, has made circuit boards particularly vulnerable to gaseous contaminant damage. Analysis with silver coupons has proven to be the best indicator of this type of contamination.

There is also a chapter in the *Contamination* book on strategies for contamination prevention and control, along with an update to the landmark ASHRAE survey of gaseous contamination and datacom equipment published in the first edition of the book. This book includes access to a supplemental download of *Particulate and Gaseous Contamination Guidelines for Data Centers* at no additional cost.

11.6.9 Book #9: *Real-Time Energy Consumption Measurements in Data Centers, 1st Edition* [14]

The adage "You can't manage what you can't measure" has never been more true than in data centers. The wide variety of equipment, the constant "churn" as hardware is added and replaced, and the moment-to-moment changes in workloads make any single measurement of energy consumption a poor indicator of actual conditions over time. Moreover, modern hardware, both computing systems and power and cooling infrastructures, provide thousands of monitoring points generating volumes of performance data. Control of any device, whether to modify its operational parameters or to become aware of an impending failure, requires both real-time and historical monitoring of the device, as well as of the overall systems. This is also the key to optimizing energy efficiency.

But another important issue is the need for good communication between IT and facilities. These entities typically report to different executives, and they most certainly operate on different time schedules and priorities and speak very different technical languages. Good monitoring that provides useful information to both entities (as opposed to "raw data" that few can interpret) can make a big difference in bridging the communication gap that often exists between these two groups. If each part of the organization can see the performance information important to the systems for which they have responsibility, as well as an overall picture of the data center performance and trends, there can be significant improvements in communication, operation, and long-term stability and reliability. This, however, requires the proper instrumentation and monitoring of key power and cooling systems, as well as performance monitoring of the actual computing operation. This book provides insight into the proper use of these measurements, but a later book in the Datacom Series thoroughly covers the *Data Center Infrastructure Management* or "DCIM" systems that have grown out of the need for these measurements. DCIM can play an important role in turning the massive amount of "data" into useful "information."

Another great value of this book is the plethora of examples showing how energy consumption data can be used to calculate *PUE*™ (*Power Utilization Effectiveness*). One of the most challenging aspects of the *PUE*™ metric is calculation in mixed-use facilities. Although a later book in the *Datacom Series* focuses entirely on *PUE*™, this book contains a practical method of quantifying *PUE*™ in those situations. Facilities that use combined cooling, heat, and power systems make *PUE*™ calculations even more challenging. This book provides clarifications of the issues affecting these calculations.

11.6.10 Book #10: *Green Tips for Data Centers, 1st Edition* [15]

The data center industry has been focused on improving energy efficiency for many years. Yet, despite all that has been written in books and articles and all that has been provided in seminars, many existing operations are still reluctant to adopt what can appear to be complex, expensive, and potentially disruptive cooling methods and practices. Even those who have been willing and anxious to incorporate "best practices" for efficient cooling became legitimately concerned when *ASHRAE Standard 90.1* suddenly removed the exemption for data centers from its requirements, essentially forcing this industry to adopt energy-saving approaches commonly used in office buildings. Those approaches can be problematic when applied to the critical systems used in data centers, which operate continuously and are never effectively "unoccupied" as are office buildings in off-hours when loads decrease significantly.

The ultimate solution to the concerns raised by *Std. 90.1* was *ANSI/ASHRAE Standard 90.4*, discussed in detail in the following Sections 11.8, 11.10, and 11.11. But there are many energy-saving steps that can be taken in existing data centers without subjecting them to the requirements of 90.1, and ensuring compliance with 90.4. The continually increasing energy costs associated with never-ending demands for more compute power, the capital costs of cooling systems, and the frightening disruptions when cooling capacity must be added to an existing operation require that facilities give full consideration to ways of making their operations more "green" in the easiest ways possible.

ASHRAE TC 9.9 recognizes that considerable energy can be saved in the data center without resorting to esoteric means. Savings can be realized in the actual power and cooling systems, often by simply having a better understanding of how to operate them efficiently. Savings can also accrue in the actual ITE by operating in ways that avoid unnecessary energy use. The *Green Tips* book condenses many of the more thorough and technical aspects of the previous books in order to provide simplified understandings and solutions for users. It is not intended to be a thorough treatise on the most sophisticated energy-saving designs, but it does provide data center owners and operators, in nontechnical language, with an understanding of the energy-saving opportunities that exist and practical methods of achieving them. *Green Tips* covers both mechanical cooling and electrical systems, including backup and emergency power efficiencies. The organization of the book also provides a method of conducting an energy usage assessment internally.

11.6.11 Book #11: *PUE™: A Comprehensive Examination of the Metric, 1st Edition* [16]

The *Power Usage Effectiveness* metric, or *PUE*™, has become the most widely accepted method of quantifying the efficiency of data center energy usage that has ever been developed. It was published in 2007 by *TGG*, a nonprofit consortium of industry leading data center owners and operators, policy makers, technology providers, facility architects, and utility companies, dedicated to energy-efficient data center operation and resource conservation worldwide. *PUE*™ is deceptively simple in concept; the total energy consumed by the data center is divided by the IT hardware energy to obtain a quotient. Since the IT energy doesn't include energy used for cooling, or energy losses from inefficiencies such as power delivery through a UPS, IT energy will always be less than total energy. Therefore, the *PUE*™ quotient must always be greater than 1.0, which would be perfect, but is unachievable since nothing is 100% efficient. *PUE*™ quotients as low as 1.1 have been claimed, but most facilities operate in the 1.5–2.0 range. *PUE*s of 2.5–3.0 or above indicate considerable opportunity for energy savings.

Unfortunately, for several years after its introduction, the *PUE*™ metric was grossly misused, as major data centers began advertising *PUE*™ numbers so close to 1.0 as to be unbelievable. There were even claims of *PUE*s less than 1.0, which would be laughable if they didn't so clearly indicate an egregious misunderstanding. "Advertised" *PUE*s were usually done by taking instantaneous power readings at times of the day when the numbers yielded the very best results. The race was on to publish *PUE*s as low as possible, but the *PUE*™ metric was never intended to compare the efficiencies of different data centers. So although the claims sounded good, they really meant nothing. There are too many variables involved, including climate zone and the type of computing being done, for such comparisons to be meaningful. Further, while *PUE*™ can certainly be continually monitored, and noted at different times during the day, it is only the *PUE*™ based on total energy usage over time that really matters. "Energy" requires a time component, such as kilowatt-hours (kWh). Kilowatts (kW) is only a measurement of instantaneous power at any given moment. So while a *PUE*™ based on power can be useful when looking for specific conditions that create excessive loads, it is the energy measurement that provides a true *PUE*™ number and is the most meaningful. That requires accumulating power data over time—usually a full year.

To remedy this gross misuse of the *PUE*™ metric, in 2009 TGG published a revised metric called Version 2.1, or more simply, *PUE*v2™, that provided four different levels of *PUE*™ measurement. The first, and most basic level, remains the instantaneous power readings. But when that is done, it must be identified as such with the designation "PUE_0." Each successive measurement method requires long-term cumulative energy tracking and also requires measuring the ITE usage more and more accurately. At PUE_3, IT energy use is derived directly from internal hardware data collection.

In short, the only legitimate use of the *PUE*™ metric is to monitor one's own energy usage in a particular data center over time in order to quantify relative efficiency as changes are made. But it is possible, and even likely, to make significant reductions in energy consumption, such as by consolidating servers and purchasing more energy-efficient compute hardware, and see the *PUE*™ go up rather than down. This can be disconcerting, but should not be regarded as failure, since total energy consumption has still been reduced. Data center upgrades are usually done incrementally, and replacing power and cooling equipment, just to achieve a better *PUE*™, is not as easily cost-justified as replacing obsolete IT hardware. So an increase in *PUE*™ can occur when commensurate changes are not made in the power and cooling systems. Mathematically, if the numerator of the equation is not reduced by as much as the denominator, a higher quotient will result despite the reduction in total energy use. That should still be considered a good thing.

In cooperation with TGG, ASHRAE TC 9.9 published *PUE™: An Examination of the Metric* [17] with the intent of providing the industry with a thorough explanation of *PUE*™, an in-depth understanding of what it is and is not, and a clarification of how it should and should not be used. This book consolidates all the material previously published by TGG, as well as adding new material. It begins with the concept of the *PUE*™ metric, continues with how to properly calculate and apply it, and then specifies how to report and analyze the results. This is critical for everyone involved in the operation of a data center, from facility personnel to executives in the C-suite for whom the *PUE*™ numbers, rather than their derivations, can be given more weight than they should, and become particularly misleading.

11.6.12 Book #12: *Server Efficiency—Metrics for Computer Servers and Storage, 1st Edition* [17]

Simply looking for the greatest server processing power or the fastest storage access speed on data sheets is no longer a responsible way to evaluate computing hardware. Energy awareness also requires examining the energy required to produce useful work, which means evaluating "performance per watt" along with other device data. A number of different energy benchmarks are used by manufacturers. This book examines each of these metrics in terms of its application and target market. It then provides guidance on interpreting the data, which will differ for each type of device in a range of applications. In the end, the information in this book enables users to select the best measure of performance and power for each server application.

11.6.13 Book #13: *IT Equipment Design Impact on Data Center Solutions, 1st Edition* [18]

The data center facility, the computing hardware that runs in it, and the OS and application code that runs on that hardware together form a "system." The performance of that "system" can be optimized only with a good understanding of how the ITE responds to its environment. This knowledge has become increasingly important as the Internet of Things (IoT) drives the demand for more and faster processing of data, which can quickly exceed the capabilities for which most data centers were designed. That includes both the processing capacities of the IT hardware and the environment in which it runs. Hyperscale convergence, in particular, has required much rethinking of the data center systems and environment amalgamation.

The goal of this book is to provide an understanding for all those who deal with data centers of how ITE and environmental system designs interact so that selections can be made that are flexible, scalable, and adaptable to new demands as they occur. The intended audience includes facility designers, data center operators, ITE and environmental systems manufacturers, and end users, all of whom must learn new ways of thinking in order to respond effectively to the demands that this enormous rate of change is putting on the IT industry. The book is divided into sections that address the concerns of three critical groups:

- Those who design the infrastructure, who must therefore have a full understanding of how the operating environment affects the ITE that must perform within it.
- Those who own and operate data centers, who must therefore understand how the selection of the ITE and its features can either support or impair both optimal operation and the ability to rapidly respond to changes in processing demand.
- IT professionals, who must have a holistic view of how the ITE and its environment interact, in order to operate their systems with optimal performance and flexibility.

11.6.14 Book #14: *Advancing DCIM with IT Equipment Integration, 1st Edition* [19]

One of the most important data center industry advances in recent years is the emergence, growth, and increasing sophistication of *DCIM* or *Data Center Infrastructure Management* tools. All modern data center equipment, including both IT and power/cooling hardware, generates huge amounts of data from potentially thousands of devices and sensors. (See Section 11.6.9). Unless this massive amount of data is converted to useful information, most of it is worthless to the average user. But when monitored and accumulated by a sophisticated system that reports consolidated results in meaningful and understandable ways, this data is transformed into a wealth of information that can make a significant difference in how a data center is operated.

It is critical in today's diverse data centers to effectively schedule workloads, and to manage and schedule power, cooling, networking, and space requirements, in accordance with actual needs. Providing all required assets in the right amounts, and at the right times, even as load and environmental demands dynamically change, results in a highly efficient operation—efficient in computing systems utilization as well as efficient in energy consumption. Conversely, the inability to maintain a reasonable balance can strand resources, limit capacity, impair operations, and be wasteful of energy and finances. At the extreme, poor management and planning of these resources can put the entire data center operation at risk.

DCIM might be called ERP (enterprise resource planning) for the data center. It's a software suite for managing both the data center infrastructure and its computing systems by collecting data from IT and facilities gear, consolidating it into relevant information, and reporting it in real time. This enables the intelligent management, optimization, and future planning of data center resources such as processing capacity, power, cooling, space, and assets. DCIM tools come in a wide range of flavors. Simple power monitoring is the most basic, but the most sophisticated systems provide complete visibility across both the management and operations layers. At the highest end, DCIM can track assets from order placement through delivery, installation, operation, and decommissioning. It can even suggest the best places to mount new hardware based on space, power, and cooling capacities and can track physical location, power and data connectivity, energy use, and processor and memory utilization. A robust *DCIM* can even use artificial intelligence (AI) to provide advance alerts to impending equipment failures by monitoring changes in operational data and comparing them with preset thresholds. But regardless of the level of sophistication, the goal of any *DCIM* tool is to enable operations to optimize system performance on a holistic basis, minimize cost, and report results to upper management in understandable formats. The Covid-19 pandemic also proved the value of DCIM when operators could not physically enter their data centers, and had to rely on information obtained remotely. A robust DCIM is likely to become an important part of every facility's disaster response planning.

The ASHRAE book Foreword begins with the heading "DCIM—Don't Let Data Center Gremlins Keep You Up At Night." Chapters include detailed explanations and definitions, information on industry standards, best practices, interconnectivity explanations, how to properly use measured data, and case examples relating to power, thermal, and capacity planning measurements. There are appendices to assist with proper sensor placement and use of performance metrics, and the introduction of "*CITE*" *Compliance for IT Equipment*, *CITE* defines the types of telemetry that should be incorporated into ITE designs so that *DCIM* solutions can be used to maximum advantage. In short, this is the first comprehensive treatment of one of the industry's most valuable tools in the

arsenal now available to the data center professional. But due to the number of different approaches taken by the multiple providers of *DCIM* solutions and the range of features available, *DCIM* is also potentially confusing and easy to misunderstand. The aim of this book is to remedy that situation.

11.7 THE ASHRAE HANDBOOK AND TC 9.9 WEBSITE

As noted at the beginning of this chapter, there are many resources available from ASHRAE, with the Datacom book series being the most thorough. Another worthwhile publication is the *ASHRAE Handbook*. This 4-volume set is often called the "bible" of the HVAC industry, containing chapters written by every Technical Committee in ASHRAE and covering virtually every topic an environmental design professional will encounter. The books are updated on a rotating basis so that each volume is republished every 4 years. However, with the advent of online electronic access, out-of-sequence updates are made to the online versions of the handbooks when changes are too significant to be delayed to the next book revision. Chapter 20 of the *Applications* volume (formerly Chapter 19 before the 2019 edition) is authored by *TC 9.9* and provides a good overview of data center design requirements, including summaries of each of the books in the Datacom series. In addition, the TC 9.9 website (http://tc0909.ashraetcs.org) contains white papers covering current topics of particular relevance, most of which are ultimately incorporated into the next revisions of the Datacom book series and, by reference or summary, into the Handbook as well.

11.8 ASHRAE STANDARDS AND CODES

As also previously noted, ASHRAE publishes several standards that are very important to the data center industry. Chief among these, and the newest for this industry, is *Standard 90.4*, *Energy Standard for Data Centers* [20]. *Std 90.4* was originally published in July 2016 and has been significantly updated for the 2019 Code Cycle. Other relevant standards include *Std.127*, *Testing Method for Unitary Air Conditioners,* which is mainly applicable to manufacturers of precision cooling units for data centers. *Standard 127* is an advisory standard, meaning manufacturers are encouraged to comply with it, but are not required to do so. Most manufacturers of data center cooling solutions comply with *Std. 127*, but some may not. End users looking to purchase data center cooling equipment should be certain that the equipment they are considering has been tested in accordance with this standard so that comparisons of capacities and efficiencies are made on a truly objective basis.

That having been said, a word about standards and codes is appropriate here, as a preface to understanding the history and critical importance of *Std. 90.4*, which will then be discussed in detail.

"Codes" are documents that have been adopted by local, regional, state, and national authorities for the purpose of ensuring that new construction, as well as building modifications, use materials and techniques that are safe and, in more recent years, environmentally friendly. Codes have the weight of law and are enforceable by the adopting authority, known as the Authority Having Jurisdiction, or "AHJ" for short. Among the best known of those that significantly affect the data center industry in the United States is probably the *National Electrical Code* or *NEC*. It is published by the *National Fire Protection Association* or *NFPA* and is officially known as *NFPA-70®*. Other countries have similar legal requirements for electrical, as well as for all other aspects of construction. Another important code would be *NFPA-72®*, the *National Fire Alarm and Signaling Code*. There are relatively few actual "codes" and all are modified to one degree or another by each jurisdiction, both to address the AHJ's local concerns and to conform with their own opinions of what is and is not necessary. California, for example, makes significant modifications to address seismic concerns. Even the *NEC* may be modified in each state and municipality.

Standards, on the other hand, exist by the thousands. ASHRAE alone publishes more than 125 that are recognized by ANSI (American National Standards Institute). Virtually every other professional organization, including the *NFPA and* the *IEEE (Institute of Electrical and Electronics Engineers)*, also publishes standards that are highly important to our industry, but are never adopted by the AHJ as "code." These are known as "advisory standards," which, as noted for *ASHRAE Std. 127*, means that a group of high-ranking industry professionals, usually including manufacturers, users, professional architects and engineers, and other recognized experts, strongly recommend that the methods and practices in the documents be followed. Good examples in the data center industry are *NFPA-75*, *Standard for Fire Protection of Information Technology Equipment*, and *NFPA-76*, *Standard for Fire Protection of Telecommunications Facilities*. Advisory standards can have several purposes. Most provide "best practices" for an industry, establishing recognized ways designers and owners can specify the level to which they would like facilities to be designed and constructed. But other standards are strictly to establish a uniform basis for comparing and evaluating similar types of equipment. Again, *ASHRAE Std. 127* is a good example of this. All reputable manufacturers of computer room air conditioners voluntarily test their products according to this standard, ensuring that their published specifications are all based on the same criteria and can be used for true "apples-to-apples" comparisons. There is no legal requirement for anyone to do this, but it is generally accepted that products of any kind must adhere to certain standards in order to be

recognized, accepted, and trusted by knowledgeable users in any industry.

But when a standard is considered important enough by the AHJ to be mandated. A major example of this is *ASHRAE Standard 90.1*, *Energy Standard for Buildings Except Low Rise Residential*. As the title implies, virtually every building except homes and small apartment buildings is within the purview of this standard. *ASHRAE 90.1*, as it is known for short, is adopted into code or law by virtually every local, state, and national code authority in the United States, as well as by many international entities. This makes it a very important standard. Architects and engineers are well acquainted with it, and it is strictly enforced by code officials.

11.9 ANSI/ASHRAE STANDARD 90.1-2010 AND ITS CONCERNS

For most of its existence, *ASHRAE Std. 90.1* included an exemption for data centers. Most codes and standards are revised and republished on a 3-year cycle, so the 2007 version of *Std. 90.*1 was revised and republished in 2010. In the 2010 revision, the data center exemption was simply removed, making virtually all new, expanded, and renovated data centers subject to all the requirements of the *90.1 Standard*. In other words, data centers were suddenly lumped into the same category as any other office or large apartment building. This went virtually unnoticed by most of the data center community because new editions of codes and standards are not usually adopted by AHJ's until about 3 years after publication. Some jurisdictions adopt new editions sooner, and some don't adopt them until 6 or more years later, but as a general rule, a city or state will still be using the 2016 edition of a code long after the 2019 version has been published. Some will still use the 2016 edition even after the 2022 version is available. In short, this seemingly small change would not have been recognized by most people until at least several years after it occurred.

But the removal of the data center exemption was actually enormous and did not go unnoticed by *ASHRAE TC 9.9*, which argued, lobbied, and did everything in its power to get the *Std. 90.1* committee to reverse its position. A minor "Alternative Compliance Path" was finally included, but the calculations were onerous, so it made very little difference.

This change to *Std. 90.1* raised several significant concerns, the major one being that *Std. 90.1* is prescriptive. For the most part, instead of telling you what criteria and numbers you need to achieve, it tells you what you need to include in your design to be compliant. In the case of cooling systems, that means a device known as an economizer, which is essentially a way of bypassing the chiller plant when the outside air is cool enough to maintain building air temperatures without mechanical refrigeration—in other words "free cooling." That can require a second cooling tower, which is that large box you see on building roofs, sometimes emitting a plume of water vapor that looks like steam.

There's nothing fundamentally wrong with economizers. In fact, they're a great energy saver, and *Std. 90.1* has required them on commercial buildings for years. But their operation requires careful monitoring in cold climates to ensure that they don't freeze up, and the process of changing from chiller to economizer operation and back again can result in short-term failures of the cooling systems. That's not a great concern in commercial buildings that don't have the reliability demands of high-availability data centers. But for mission-critical enterprises, those interruptions would be disastrous. In fact, in order to meet the availability criteria of a recognized benchmark like Uptime Institute Tier III or Tier IV, or a corresponding TIA Level, two economizer towers would be needed, along with the redundant piping to serve them. That simply exacerbates the second concern about mandating economizers, namely, where to put them and how to connect them on existing buildings, especially on existing high-rise structures. If one wanted to put a small data center in the Empire State Building in New York City, for example, *Standard 90.1-2010* would preclude it. You would simple not be able to meet the requirements.

11.10 THE DEVELOPMENT OF ANSI/ASHRAE STANDARD 90.4

Concern grew rapidly in the data center community as it became aware of this change. A*SHRAE TC 9.9* also continued to push hard for *Std. 90.1* addenda and revisions that would at least make the onerous requirements optional. When that did not occur, the ASHRAE Board suggested that *TC 9.9* propose the development of a new standard specific to data centers. The result was *Standard 90.4*.

Standards committees are very different than TCs. Members are carefully selected to represent a balanced cross section of the industry. In this case, that included industry leading manufacturers, data center owners and operators, consulting engineers specializing in data center design, and representatives of the power utilities. In all, 15 people were selected to develop this standard. They worked intensely for 3 years to publish in 2016 so it would be on the same 3-year Code Cycle as *Std. 90.1*. This was challenging since standards committees must operate completely in the open, following strict requirements dictated by ANSI (American National Standards Institute) to be recognized. Committee meetings must be fully available to the public, and must be run in accordance with Robert's Rules of Order, with thorough minutes kept and made accessible for public consumption. Only committee members can vote, but others can be recognized during meetings to contribute advice or make comments. The most important and time-consuming

requirement, however, is that ANSI standards must be published for public review before they can be published, with each substantive comment formally answered in writing using wording developed by and voted on by the committee. If comments are accepted, the Draft Standard is revised and then resubmitted for another public review. Comments on the revisions are reviewed in the same way until the committee has either satisfied all concerns or objections or has voted to publish the standard without resolving comments they consider inappropriate to include, even if the commenter still disagrees. In other words, it is an onerous and lengthy process, and achieving publication by a set date requires significant effort. That is what was done to publish *Std. 90.4* on time, because the committee felt it was so important to publish simultaneously with *Std. 90.1*. By prior agreement, the two standards were to cross-reference each other when published.

Unfortunately, the best laid plans don't always materialize. While *Std. 90.4* was published on time, due to an ANSI technicality, *Std. 90.1-2016* was published without the pre-agreed cross-references to *Std. 90.4*. This resulted in two conflicting ASHRAE standards, which was both confusing and embarrassing. That was remedied with publication of the 2019 versions of both *Standard 90.1* and *Standard 90.4,* which now reference each other. *Standard 90.4* applies to data centers, which are defined as having design IT loads of at least 10 kW and 20 W/ft^2 or 215 W/m^2. Smaller facilities are defined as computer rooms and are still subject to the requirements of *Standard 90.1*.

11.11 SUMMARY OF ANSI/ASHRAE STANDARD 90.4

ASHRAE/ANSI Standard 90.4 is a performance-based standard. In other words, contrary to the prescriptive approach of *Std. 90.1*, *Std. 90.4* establishes minimum efficiencies for which the mechanical and electrical systems must be designed. But it does *not* dictate what designers must do to achieve them. This is a very important distinction. The data center industry has been focused on energy reduction for a long time, which has resulted in many innovations in both power and cooling technologies, with more undoubtedly to come. None of these cooling approaches is applicable to office or apartment buildings, but each is applicable to the data center industry, depending on the requirements of the design. Under *Std. 90.4*, designers are able to select from multiple types and manufacturers of infrastructure hardware according to the specific requirements and constraints of each project. Those generally include flexibility and growth modularity, in addition to energy efficiency and the physical realities of the building and the space. Budgets, of course, also play a major role. But above all, the first consideration in any data center design is reliability. The Introduction to *Std. 90.4* makes it clear that this standard was developed with reliability and availability as overriding considerations in any mission critical design.

Standard 90.4 follows the format of *Standard 90.1* so that cross-references are easy to relate. Several sections, such as service water heating and exterior wall constructions, do not have mission-critical requirements that differ from those already established for energy efficient buildings, so *Std. 90.4* directs the user back to *Std. 90.1* for those aspects.

The central components of *Std. 90.4* are the mechanical and electrical systems. It was determined early in the development process that the *PUE*™ metric, although widely recognized, is not a "design metric" and would be highly misleading if used for this purpose since it is an operational metric that cannot be accurately calculated in the design stage of a project. Therefore, the *Std. 90.4* committee developed new, more appropriate metrics for these calculations. These are known as the mechanical load component (MLC) and the electrical loss component (ELC). The MLC is calculated from the equations in the 90.4 Standard and must be equal to or lower than the values stipulated in the standard for each climate zone. The ELC is calculated from three different segments of the electrical systems: the incoming service segment, the UPS segment, and the distribution segment. The totals of these three calculations result in the ELC. ELC calculations are based on the IT design load, and the standard assumes that IT power is virtually unaffected by climate zone, so it can be assumed to be constant throughout the year. Total IT energy, therefore, is the IT design load power times the number of hours in a year (8,760 hours). The ELC, however, is significantly affected by redundancies, numbers of transformers, and wire lengths, so the requirements differ between systems with "2*N*" or greater redundancy and "*N*" or "*N* + 1" systems. UPS systems also tend to exhibit a significant difference in efficiency below and above 100 kW loads. Therefore, charts are provided in the standard for each level of redundancy at each of these two load points. While the charts do provide numbers for each segment of the ELC, the only requirement is that the total of the ELC segments meets the total ELC requirement. In other words, "trade-offs" are allowed among the segments so that a more efficient distribution component, for example, can compensate for a less efficient UPS component, or vice versa. The final ELC number must simply be equal to or less than the numbers in the 90.4 Standard tables.

The standard also recognizes that data center electrical systems are complex, with sometimes thousands of circuit paths running to hundreds of cabinets. If the standard were to require designers to calculate and integrate every one of these paths, it would be unduly onerous without making the result any more accurate, or the facility any more efficient. So *Std. 90.4* requires only that the worst-case (greatest loss) paths be calculated. The assumption is that if the worst-case paths meet the requirements, the entire data center electrical

system will be reasonably efficient. Remember any standard establishes a *minimum* performance requirement. It is expected, and hoped, that the vast majority of installations will exceed the minimum requirements. But any standard or code is mainly intended to ensure that installations using inferior equipment and/or shortcut methods unsuitable for the applications are not allowed.

Standard 90.4 also allows trade-offs between the MLC and ELC, similar to those allowed among the ELC components. Of course, it is hoped that the MLC and ELC will each meet or exceed the standard requirements. But if they don't, and one element can be made sufficiently better than the other, the combined result will still be acceptable if together they meet the combined requirements of the 90.4 Standard tables. The main reason for allowing this trade-off, however, is for major upgrades and/or expansions of either an electrical or mechanical system where the other system is not significantly affected. It is not the intent of the standard to require unnecessary and prohibitively expensive upgrades of the second system, but neither is it the intention of the standard to give every old, inefficient installation a "free pass." The trade-off method set forth in the standard allows a somewhat inefficient electrical system, for example, to be retained, so long as the new or upgraded mechanical system can be designed with sufficiently improved efficiency to offset the electrical system losses. The reverse is also allowed.

ANSI/ASHRAE Standard 90.4 is now under continuous maintenance, which means that suggestions for improvements from any user, as well as from members of the committee, are received and reviewed for applicability. Any suggestions the committee agrees will improve the standard, either in substance or understandability, are then submitted for public review following the same exacting process as for the original document. If approved, the changes are incorporated into the revisions that occur every 3 years. The 2019 version of *Standard 90.4* includes a number of revisions that were made in the interim 3-year period. Most significant among these were tightening of both the MLC and ELC minimum values. The 2022 and subsequent versions will undoubtedly contain further revisions. The expectation is that the efficiency requirements will continue to strengthen.

Since ASHRAE *Standard 90.4-2019* is now recognized and referenced within *Standard 90.1-2019*, it is axiomatic that it will be adopted by reference wherever *Std. 90.1-2019* is adopted. This means it is very important that data center designers, contractors, owners, and operators be familiar with the requirements of *Std. 90.4*.

11.12 ASHRAE BREADTH AND THE *ASHRAE JOURNAL*

Historically, ASHRAE has been an organization relevant primarily to mechanical engineers. But the work done by, or in cooperation with, Technical Committee *TC 9.9* has become a very comprehensive resource for information relating to data center standards, best practices, and operation.

Articles specific to data center system operations and practices often also appear in the *ASHRAE Journal*, which is published monthly. Articles that appear in the journal have undergone thorough double-blind reviews, so these can be considered highly reliable references. Since these articles usually deal with very current technologies, they are important for those who need to be completely up to date in this fast-changing industry. Some of the information published in articles is ultimately incorporated into new or revised books in the Datacom Series, into Chapter 20 of the ASHRAE Handbook, and/or into the *90.4 Standard*.

In short, ASHRAE is a significant source of information for the data center industry. Although it addresses primarily the facilities side of an enterprise, knowledge and awareness of the available material can also be very important to those on the operations side of the business.

REFERENCES

[1] *The American Society of Heating, Refrigeration and Air Conditioning Engineers*. Available at https://www.ashrae.org/about. Accessed on March 1, 2020.

[2] ASHRAE. Technical Committee *TC 9.9*. Available at http://tc0909.ashraetcs.org/. Accessed on March 1, 2020.

[3] The Green Grid (TGG). Available at https://www.thegreengrid.org/. Accessed on March 1, 2020.

[4] Wan F, Swenson D, Hillstrom M, Pommerenke D, Stayer C. The Effect of Humidity on Static Electricity Induced Reliability Issues of ICT Equipment in Data Centers Source_ASHRAE_Transactions"ASHRAE Transactions, vol. 119, p. 2; January 2013. Available at https://www.esdemc.com/public/docs/Publications/Dr.%20Pommerenke%20Related/The%20Effect%20of%20Humidity%20on%20Static%20Electricity%20Induced%20Reliability%20Issues%20of%20ICT%20Equipment%20in%20Data%20Centers%20%E2%80%94Motivation%20and%20Setup%20of%20the%20Study.pdf. Accessed on June 29, 2020.

[5] European Union's. *RoHS* Directive. Available at https://ec.europa.eu/environment/waste/rohs_eee/index_en.htm. Accessed on March 1, 2020.

[6] Book 1*: Thermal Guidelines or Data Processing Environments*. 4th ed.; 2015.

[7] Book 2*: IT Equipment Power Trends*. 2nd ed.; 2009.

[8] Book 3: *Design Considerations for Datacom Equipment Centers*. 3rd ed.; 2020.

[9] Book 4: *Liquid Cooling Guidelines for Datacom Equipment Centers*. 2nd ed.; 2013.

[10] Book 5*: Structural and Vibration Guidelines for Datacom Equipment Centers*. 2008.

[11] Book 6: *Best Practices for Datacom Facility Energy Efficiency*. 2nd ed.; 2009.

[12] Book 7: *High Density Data Centers – Case Studies and Best Practices*. 2008.

[13] Book 8: *Particulate and Gaseous Contamination in Datacom Facilities*. 2nd ed.; 2014.

[14] Book 9: *Real-Time Energy Consumption Measurements in Data Centers*. 2010.

[15] Book 10: *Green Tips for Data Centers*. 2011.

[16] Book 11: *PUE™: A Comprehensive Examination of the Metric*. 2014.

[17] Book 12: *Server Efficiency – Metrics for Computer Servers and Storage*. 2015.

[18] Book 13: *IT Equipment Design Impact on Data Center Solutions*. 2016.

[19] Book 14: *Advancing DCIM with IT Equipment Integration*. 2019.

[20] (a) ANSI/ASHRAE/IES Standard 90.1-2019. *Energy Standard for Buildings Except Low-Rise Residential Buildings*. Available at https://www.techstreet.com/ashrae/subgroups/42755. Accessed on March 1, 2020.*;*
(b) ANSI/ASHRAE Standard 90.4-2019. *Energy Standard for Data Centers;*
(c) ANSI/ASHRAE Standard 127-2012. *Method of Testing for Rating Computer and Data Processing Unitary Air Conditioners*;
(d) ANSI/TIA Standard 942-B-2017. *Telecommunications Infrastructure Standard for Data Centers;*
(e) NFPA Standard 70-2020. *National Electric Code;*
(f) NFPA Standard 75-2017. *Fire Protection of Information Technology Equipment;*
(g) NFPA Standard 76-2016. *Fire Protection of Telecommunication Facilities;*
(h) McFarlane R. *Get to Know ASHRAE 90.4, the New Energy Efficiency Standard*. TechTarget. Available at https://searchdatacenter.techtarget.com/tip/Get-to-know-ASHRAE-904-the-new-energy-efficiency-standard. Accessed on March 1, 2020;
(i) McFarlane R. *Addendum Sets ASHRAE 90.4 as Energy-Efficiency Standard*. TechTarget. Available at https://searchdatacenter.techtarget.com/tip/Addendum-sets-ASHRAE-904-as-energy-efficiency-standard. Accessed on March 1, 2020.

FURTHER READING

ASHRAE. *Datacom Book Series*. Available at https://www.techstreet.com/ashrae/subgroups/42755. Accessed on March 1, 2020.

Pommerenke D., Swenson D. *The Effect of Humidity on Static Electricity Induced Reliability Issues of ICT Equipment in Data Center. ASHRAE Research Project RP-1499, Final Report;* 2014.

12

DATA CENTER TELECOMMUNICATIONS CABLING AND TIA STANDARDS

Alexander Jew
J&M Consultants, Inc., San Francisco, California, United States of America

12.1 WHY USE DATA CENTER TELECOMMUNICATIONS CABLING STANDARDS?

When mainframe and minicomputer systems were the primary computing systems, data centers used proprietary cabling that was typically installed directly between equipment. See Figure 12.1 for an example of a computer room with unstructured nonstandard cabling designed primarily for mainframe computing.

With unstructured cabling built around nonstandard cables, cables are installed directly between the two pieces of equipment that need to be connected. Once the equipment is replaced, the cable is no longer useful and should be removed. Although removal of abandoned cables is a code requirement, it is common to find abandoned cables in computer rooms.

As can be seen in the Figure 12.1, the cabling system is disorganized. Because of this lack of organization and the wide variety of nonstandard cable types, such cabling is typically difficult to troubleshoot and maintain.

Figure 12.2 shows an example of the same computer room redesigned using structured standards-based cabling.

Structured standards-based cabling saves money:

- Standards-based cabling is available from multiple sources rather than a single vendor.
- Standards-based cabling can be used to support multiple applications (for example, local area networks (LAN), storage area networks (SAN), console, wide area network (WAN) circuits), so the cabling can be left in place and reused rather than removed and replaced.
- Standards-based cabling provides an upgrade path to higher-speed protocols because they are developed in conjunction with committees that develop LAN and SAN protocols.
- Structured cabling is organized, so it is easier to administer and manage.

Structured standards-based cabling improves availability:

- Standards-based cabling is organized, so tracing connections is simpler.
- Standards-based cabling is easier to troubleshoot than nonstandard cabling.

Since structured cabling can be preinstalled in every cabinet and rack to support most common equipment configurations, new systems can be deployed quickly.

Structured cabling is also very easy to use and expand. Because of its modular design, it is easy to add redundancy by (copying) the design of a horizontal distribution area (HDA) or a backbone cable. Using structured cabling breaks the entire cabling system into smaller pieces, which makes it easier to manage, compared with having all cables in one big group.

Adoption of the standards is voluntary, but the use of standards greatly simplifies the design process, ensures compatibility with application standards, and may address unforeseen complications.

During the planning stages of a data center, the owner will want to consult architects and engineers to develop a functional facility. During this process, it is easy to become confused and perhaps overlook some crucial aspect of data center construction, leading to unexpected expenses or downtime. The data center standards try to avoid this outcome by informing the reader. If data center

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

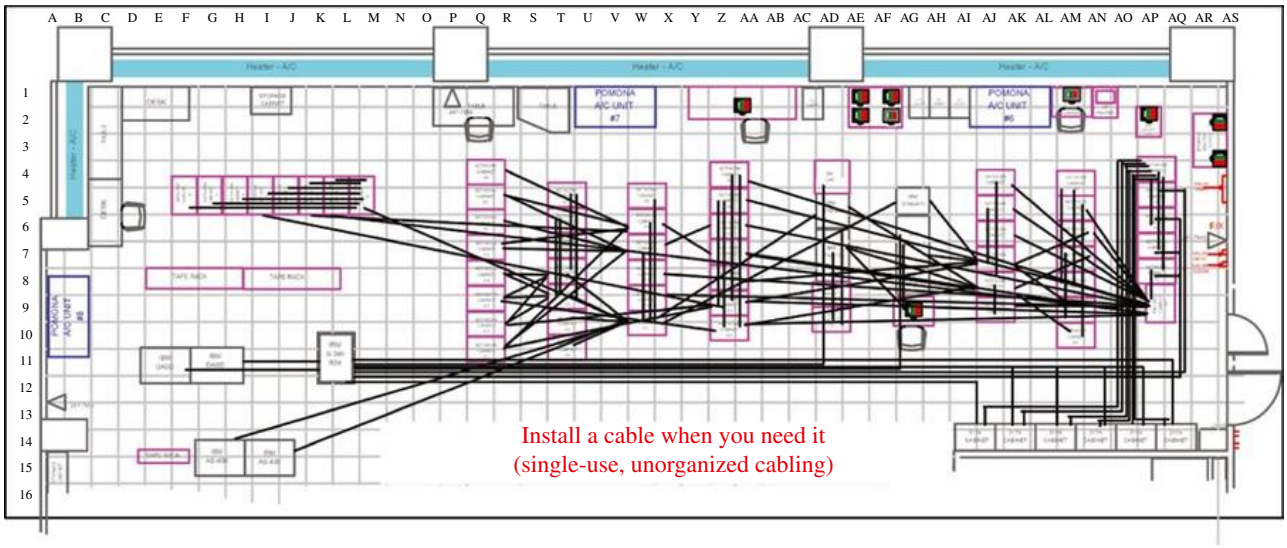


FIGURE 12.1 Example of computer room with unstructured nonstandard cabling. Source: © J&M Consultants, Inc.

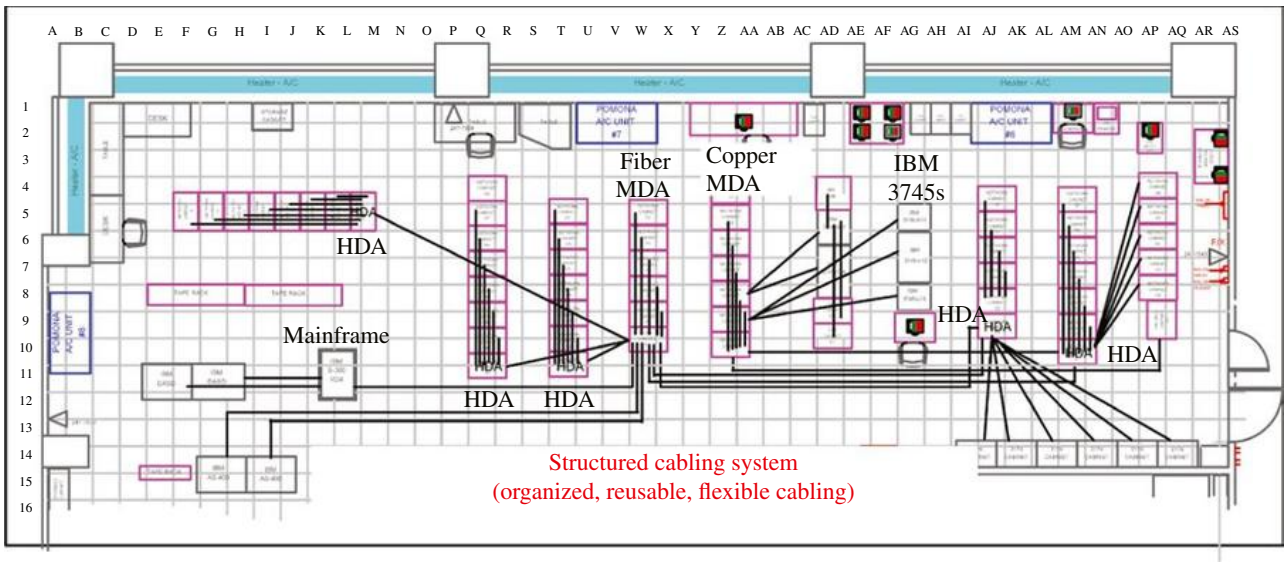


FIGURE 12.2 Example of computer room with structured standards-based cabling. Source: © J&M Consultants, Inc.

owners understand their options, they can participate during the designing process more effectively and can understand the limitations of their final designs. The standards explain the basic design requirements of a data center, allowing the reader to better understand how the designing process can affect security, cable density, and manageability. This will allow those involved with a design to better communicate the needs of the facility and participate in the completion of the project.

Common services that are typically carried using structured cabling include LAN, SAN, WAN, systems console connections, out-of-band management connections, voice, fax, modems, video, wireless access points, security cameras, distributed antenna systems (DAS), and other building signaling systems (fire, security, power controls/monitoring, HVAC controls/monitoring, etc.). There are even systems that permit LED lighting to be provisioned using structured cabling. With the development of the Internet of Things (IoT), more building systems and sensors will be using structured cabling.

12.2 TELECOMMUNICATIONS CABLING STANDARDS ORGANIZATIONS

Telecommunications cabling infrastructure standards are developed by several organizations. In the United States and Canada, the primary organization responsible for

telecommunications cabling standards is the Telecommunications Industry Association (TIA). TIA develops information and communications technology standards and is accredited by the American National Standards Institute and the Canadian Standards Association to develop telecommunications standards.

In the European Union, telecommunications cabling standards are developed by the European Committee for Electrotechnical Standardization (CENELEC). Many countries adopt the international telecommunications cabling standards developed jointly by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC).

These standards are consensus based and are developed by manufacturers, designers, and users. These standards are typically reviewed every 5 years, during which they are updated, reaffirmed, or withdrawn according to submissions by contributors. Standards organizations often publish addenda to provide new content or updates prior to publication of a complete revision to a standard.

12.3 DATA CENTER TELECOMMUNICATIONS CABLING INFRASTRUCTURE STANDARDS

Data center telecommunications cabling infrastructure standards by TIA, CENELEC, and ISO/IEC cover the following subjects:

- Types of cabling permitted
- Cable and connecting hardware specifications
- Cable lengths
- Cabling system topologies
- Cabinet and rack specifications and placement
- Telecommunications space design requirements (for example, door heights, floor loading, lighting levels, temperature, and humidity)
- Telecommunications pathways (for example, conduits, optical fiber duct, and cable trays)
- Testing of installed cabling
- Telecommunications cabling system administration and labeling

The TIA data center standard is ANSI/TIA-942-B *Telecommunications Infrastructure Standard for Data Centers*. The ANSI/TIA-942-B standard is the second revision of the ANSI/TIA-942 standard. This standard provides guidelines for the design and installation of a data center, including the facility's layout, cabling system, and supporting equipment. It also provides guidance regarding energy efficiency and provides a table with design guidelines for four ratings of data center reliability.

ANSI/TIA-942-B references other TIA standards for content that is common with other telecommunications cabling standards. See Figure 12.3 for the organization of the TIA telecommunications cabling standards.

Thus, ANSI/TIA-942-B references each of the common standards:

- ANSI/TIA-568.0-D for generic cabling requirements including cable installation and testing.
- ANSI/TIA-569-D regarding pathways, spaces, cabinets, and racks.
- ANSI/TIA-606-C regarding administration and labeling.
- ANSI/TIA-607-C regarding bonding and grounding.
- ANSI/TIA-758-B regarding campus/outside cabling and pathways.
- ANSI/TIA-862-B regarding cabling for intelligent building systems including IP cameras, security systems, and monitoring systems for the data center electrical and mechanical infrastructure.
- ANSI/TIA-5017 regarding physical network security.

Detailed specifications for the cabling are specified in the component standards ANSI/TIA-569.2-D, ANSI/TIA-568.3-D, and ANSI/TIA-568.4-D, but these standards are meant primarily for manufacturers. So the data center telecommunications cabling infrastructure designer in the United States or Canada should obtain ANSI/TIA-942-B

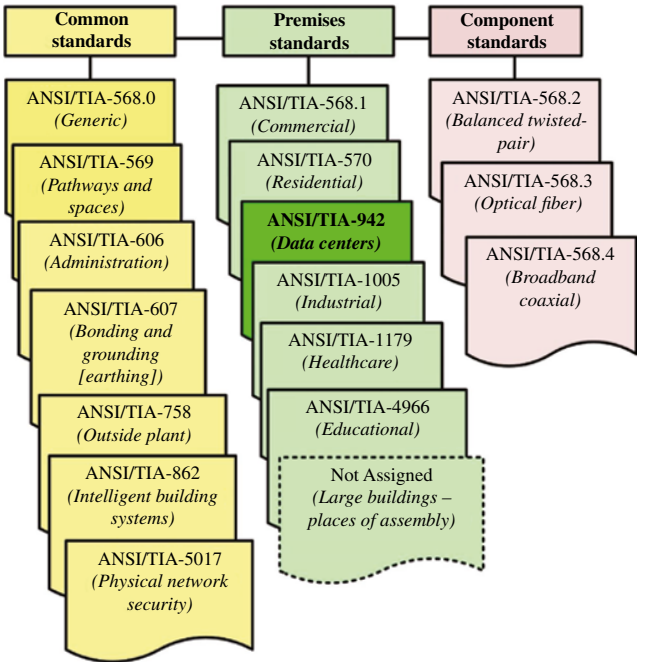


FIGURE 12.3 Organization of TIA telecommunications cabling standards. Source: © J&M Consultants, Inc.

and the common standards ANSI/TIA-568.0-D, ANSI/TIA-569-D, ANSI/TIA-606-C, ANSI/TIA-607-C, ANSI/TIA-758-B, and ANSI/TIA-862-B.

The CENELEC telecommunications standards for the European Union also have a set of common standards that apply to all types of premises and separate premises cabling standards for different types of buildings. See Figure 12.4.

A designer that intends to design telecommunications cabling for a data center in the European Union would need to obtain the CENELEC premises-specific standard for data centers (CENELEC EN 50173-5 and the common standards CENELEC EN 50173-1, EN 50174-1, EN 50174-2, EN 50174-3, EN 50310, and EN 50346.

See Figure 12.5 for the organization of the ISO/IEC telecommunications cabling standards.

A designer that intends to design telecommunications cabling for a data center using the ISO/IEC standards would need to obtain the ISO/IEC premises-specific standard for data centers—ISO/IEC 11801-5—and the common standards ISO/IEC 11801-1, ISO/IEC 14763-2, and ISO/IEC 14763-3.

The data center telecommunications cabling standards use the same topology for telecommunications cabling infrastructure but use different terminology. This handbook uses the terminology used in ANSI/TIA-942-B. See Table 12.1 for a cross-reference between the TIA, ISO, and CENELEC terminology.

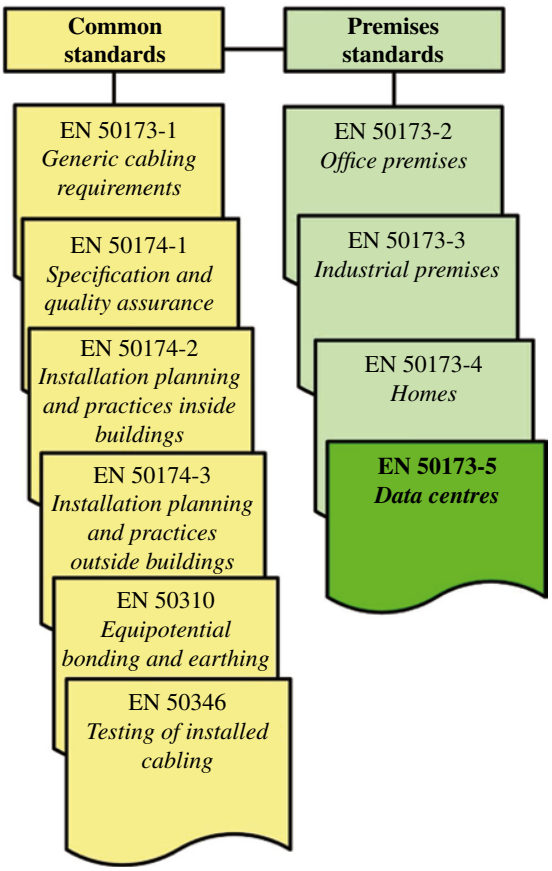


FIGURE 12.4 Organization of CENELEC telecommunications cabling standards. Source: © J&M Consultants, Inc.

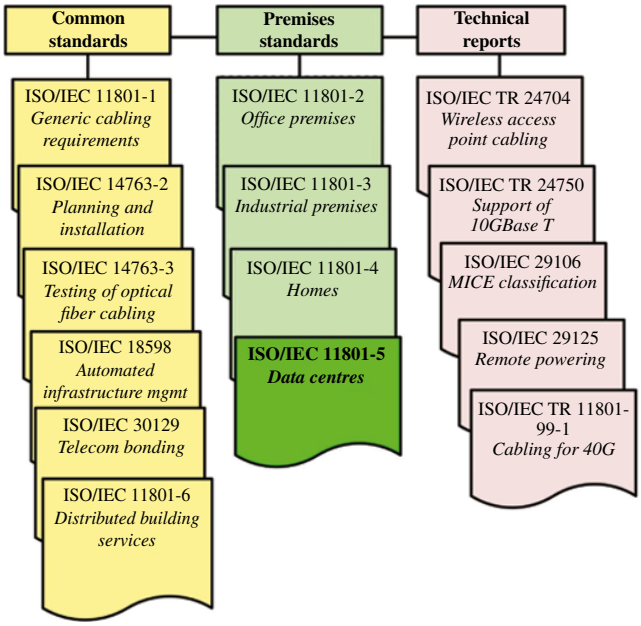


FIGURE 12.5 Organization of ISO/IEC telecommunications cabling standards. Source: © J&M Consultants, Inc.

ANSI/BICSI-002 Data Center Design and Implementation Best Practices standard is another useful reference. It is an international standard meant to supplement the telecommunications cabling standard that applies in your country—ANSI/TIA-942-B, CENELEC EN 50173-5, ISO/IEC 24764, or other—and provides best practices beyond the minimum requirements specified in these other data center telecommunications cabling standards.

12.4 TELECOMMUNICATIONS SPACES AND REQUIREMENTS

12.4.1 General Requirements

A computer room is an environmentally controlled room that serves the sole purpose of supporting equipment and cabling directly related to the computer and networking systems. The data center includes the computer room and all related support spaces dedicated to supporting the computer room such as the operations center, electrical rooms, mechanical rooms, staging area, and storage rooms.

The floor layout of the computer room should be consistent with the equipment requirements and the facility providers' requirements, including floor loading, service clearance, airflow, mounting, power, and equipment connectivity length requirements. Computer rooms should be located away from building components that would

TABLE 12.1 Cross-reference of TIA, ISO/IEC, and CENELEC terminology

ANSI/TIA-942-B	ISO/IEC 11801-5	CENELEC EN 50173-5
Telecommunications distributors		
Telecommunications entrance room (TER)	Not defined	Not defined
Main distribution area (MDA)	Not defined	Not defined
Intermediate distribution area (IDA)	Not defined	Not defined
Horizontal distribution area (HDA)	Not defined	Not defined
Zone distribution area (ZDA)	Not defined	Not defined
Equipment distribution area (EDA)	Not defined	Not defined
Cross-connects and distributors		
External network interface (ENI) in telecommunications entrance room (TER)	External network interface (ENI)	External network interface (ENI)
Main cross-connect (MC) in the main distribution area (MDA)	Main distributor (MD)	Main distributor (MD)
Intermediate cross-connect (IC) in the intermediate distribution area (IDA)	Intermediate distributor (ID)	Intermediate distributor (ID)
Horizontal cross-connect (HC) in the horizontal distribution area (HDA)	Zone distributor (ZD)	Zone distributor (ZD)
Zone outlet or consolidation point in the zone distribution area (ZDA)	Local distribution point (LDP)	Local distribution point (LDP)
Equipment outlet (EO) in the equipment distribution area (EDA)	Equipment outlet (EO)	Equipment outlet (EO)
Cabling subsystems		
Backbone cabling (from TER to MDAs, IDAs, and HDAs)	Network access cabling subsystems	Network access cabling subsystems
Backbone cabling (from MDA to IDAs and HDAs)	Main distribution cabling subsystems	Main distribution cabling subsystems
Backbone cabling (from IDAs to HDAs)	Intermediate distribution cabling subsystem	Intermediate distribution cabling subsystem
Horizontal cabling	Zone distribution cabling subsystem	Zone distribution cabling subsystem

Source: © J&M Consultants, Inc.

restrict future room expansion, such as elevators, exterior walls, building core, or immovable walls. They should also not have windows or skylights, as they allow light and heat into the computer room, making air conditioners work more and use more energy.

The rooms should be built with security doors that allow only authorized personnel to enter. It is also just as important that keys or passcodes to access the computer rooms are only accessible to authorized personnel. Preferably, the access control system should provide an audit trail.

The ceiling should be at least 2.6 m (8.5 ft) tall to accommodate cabinets up to 2.13 m (7 ft) tall. If taller cabinets are to be used, the ceiling height should be adjusted accordingly. There should also be a minimum clearance of 460 mm (18 in) between the top of cabinets and sprinklers to allow them to function effectively.

Floors within the computer room should be able to withstand at least 7.2 kPa (150 lb/ft^2), but 12 kPa (250 lb/ft^2) is recommended. Ceilings should also have a minimum hanging capacity so that loads may be suspended from them. The minimum hanging capacity should be at least 1.2 kPa (25 lb/ft^2), and a capacity of 2.4 kPa (50 lb/ft^2) is recommended.

The computer room needs to be climate controlled to minimize damage and maximize the life of computer parts.

The room should have some protection from environmental contaminants like dust. Some common methods are to use vapor barriers, positive room pressure, or absolute filtration. Computer rooms do not need a dedicated HVAC system if it can be covered by the building's and has an automatic damper; however, having a dedicated HVAC system will improve reliability and is preferable if the building's might not be on continuously. If a computer room does have a dedicated HVAC system, it should be supported by the building's backup generator or batteries, if available.

A computer room should have its own separate power supply circuits with its own electrical panel. It should have duplex convenience outlets for noncomputer use (e.g., cleaning equipment, power tools, fans, etc.). The convenience outlets should be located every 3.65 m (12 ft) unless specified otherwise by local ordinances. These should be wired on separate power distribution units/panels from those used by the computers and should be reachable by a 4.5 m (15 ft) cord. If available, the outlets should be connected to a standby generator, but the generator must be rated for electronic loads or be "computer grade."

All computer room environments including the telecommunications spaces should be compatible with $M_1I_1C_1E_1$ environmental classifications per ANSI/TIA-568.0-D. MICE classifications specify environmental requirements for M, mechanical; I, ingress; C, climatic; and E, electromagnetic. Mechanical specifications include conditions such as vibration, bumping, impact, and crush. Ingress specifications include conditions such as particulates and water immersion. Climatic includes temperature, humidity, liquid contaminants, and gaseous contaminant. Electromagnetic includes electrostatic discharge (ESD), radio-frequency emissions, magnetic fields, and surge. The CENELEC and ISO/IEC standards also have similar MICE specifications.

Temperature and humidity for computer room spaces should follow current ASHRAE TC 9.9 and manufacturer equipment guidelines.

The telecommunications spaces such as the main distribution area (MDA), intermediate distribution area (IDA), and HDA could be separate rooms within the data center but are more often a set of cabinets and racks within the computer room space.

12.4.2 Telecommunications Entrance Room (TER)

The telecommunications entrance room (TER) or entrance room refers to the location where telecommunications cabling enters the building and not the location where people enter the building. This is typically the demarcation point—the location where telecommunications access providers hand-off circuits to customers. The TER is also the location where the owner's outside plant cable (such as campus cabling) terminates inside the building.

The TER houses entrance pathways, protector blocks for twisted-pair entrance cables, termination equipment for access provider cables, access provider equipment, and termination equipment for cabling to the computer room.

The interface between the data center structured cabling system and external cabling is called the external network interface (ENI).

The telecommunications access provider's equipment is housed in this room, so the provider's technicians will need access. Because of this, it is not recommended to put the entrance room inside a computer room and that it is housed within a separate room, such that access to it does not compromise the security of any other room requiring clearance.

The room's location should also be determined so that the entire circuit length from the demarcation point does not exceed the maximum specified length. If the data center is very large:

- The TER may need to be in the computer room space.
- The data center may need multiple entrance rooms.

The location of the TER should also not interrupt airflow, piping, or cabling under floor.

The TER should be adequately bonded and grounded (for primary protectors, secondary protectors, equipment, cabinets, racks, metallic pathways, and metallic components of entrance cables).

The cable pathway system should be the same type as the one used in the computer room. Thus, if the computer room uses overhead cable tray, the TER should use overhead cable tray as well.

There may be more than one entrance room for large data centers, additional redundancy, or dedicated service feeds. If the computer rooms have redundant power and cooling, TER power and cooling should be redundant to the same degree.

There should be a means of removing water from the entrance room if there is a risk. Water pipes should also not run above equipment.

12.4.3 Main Distribution Area (MDA)

The MDA is the location of the main cross-connect (MC), the central point of distribution for the structured cabling system. Equipment such as core routers and switches may be located here. The MDA may also contain a horizontal cross-connect (HC) to support horizontal cabling for nearby cabinets. If there is no dedicated entrance room, the MDA may also function as the TER. In a small data center, the MDA may be the only telecommunications space in the data center.

The location of the MDA should be chosen such that the cable lengths do not exceed the maximum length restrictions.

If the computer room is used by more than one organization, the MDA should be in a separate secured space (for example, a secured room, cage, or locked cabinets). If it has its own room, it may have its own dedicated HVAC system and power panels connected to backup power sources.

There may be more than one MDA for redundancy.

Main distribution frame (MDF) is a common industry term for the MDA.

12.4.4 Intermediate Distribution Area (IDA)

The IDA is the location of an intermediate cross-connect (IC)—an optional intermediate-level distribution point within the structured cabling system. The IDA is not vital and may be absent in data centers that do not require three levels of distributors.

If the computer room is used by multiple organizations, it should be in a separate secure space—for example, a secured room, cage, or locked cabinets.

The IDA should be located centrally to the area that it serves to avoid exceeding the maximum cable length restrictions.

This space also typically houses switches (LAN, SAN, management, console).

The IDA may contain an HC to support horizontal cabling to cabinets near the IDA.

12.4.5 Horizontal Distribution Area (HDA)

The HDA is a space that contains an HC, the termination point for horizontal cabling to the equipment cabinets and racks (equipment distribution areas [EDAs]). This space typically also houses switches (LAN, SAN, management, console).

If the computer room is used by multiple organizations, it should be in a separate secure space—for example, a secured room, cage, or locked cabinets

There should be a minimum of one HC per floor, which may be in an HDA, IDA, or MDA.

The HDA should be located to avoid exceeding the maximum backbone length from the MDA or IDA for the medium of choice. If it is in its own room, it is possible for it to have its own dedicated HVAC or electrical panels.

To provide redundancy, equipment cabinets and racks may have horizontal cabling to two different HDAs.

Intermediate distribution frame (IDF) is a common industry term for the HDA.

12.4.6 Zone Distribution Area (ZDA)

The zone distribution area (ZDA) is the location of either a consolidation point or equipment outlets (EOs). A consolidation point is an intermediate administration point for horizontal cabling. Each ZDA should be limited to 288 coaxial cable or balanced twisted-pair cable connections to avoid cable congestion. The two ways that a ZDA can be deployed—as a consolidation point or as a multiple outlet assembly—are illustrated in Figure 12.6.

The ZDA shall contain no active equipment, nor should it be a cross-connect (i.e., have separate patch panels for cables from the HDAs and EDAs).

ZDAs may be in under-floor enclosures, overhead enclosures, cabinets, or racks.

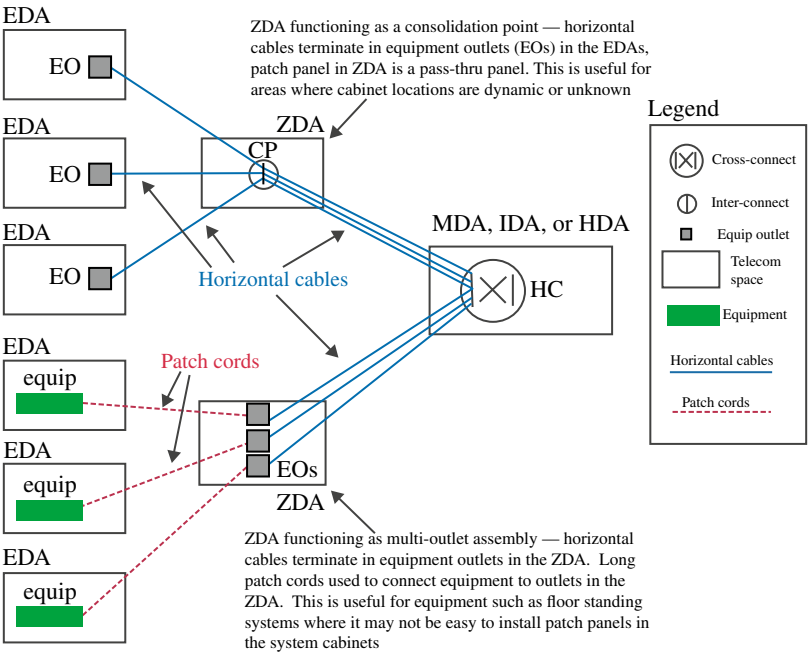


FIGURE 12.6 Two examples of ZDAs. Source: © J&M Consultants, Inc.

12.4.7 Equipment Distribution Area (EDA)

The EDA is the location of end equipment, which is composed of the computer systems, communications equipment, and their racks and cabinets. Here, the horizontal cables are terminated in EOs. Typically, an EDA has multiple EOs for terminating multiple horizontal cables. These EOs are typically located in patch panels located at the rear of the cabinet or rack (where the connections for the servers are usually located).

Point-to-point cabling (i.e., direct cabling between equipment) may be used between equipment located in EDAs. Point-to-point cabling should be limited to 7 m (23 ft) in length and should be within a row of cabinets or racks. Permanent labels should be used on either end of each cable.

12.4.8 Telecommunications Room (TR)

The telecommunications room (TR) is an area that supports cabling to areas outside of the computer room, such as operations staff support offices, security office, operations center, electrical room, mechanical room, or staging area. They are usually located outside of the computer room but may be combined with an MDA, IDA, or HDA.

12.4.9 Support Area Cabling

Cabling for support areas of the data center outside the computer room is typically supported from one or more dedicated TRs to improve security. This allows technicians working on telecommunications cabling, servers, or network hardware for these spaces to remain outside the computer room.

Operation rooms and security rooms typically require more cables than other work areas. Electrical rooms, mechanical rooms, storage rooms, equipment staging rooms, and loading docks should have at least one wall-mounted phone in each room for communication within the facility. Electrical and mechanical rooms need at least one data connection for management system access and may need more connections for equipment monitoring.

12.5 STRUCTURED CABLING TOPOLOGY

The structured cabling system topology described in data center telecommunications cabling standards is a hierarchical star. See Figure 12.7 for an example.

The horizontal cabling is the cabling from the HCs to the EDAs and ZDAs. This is the cabling that supports end equipment such as servers.

The backbone cabling is the cabling between the distributors where cross-connects are located—TERs, TRs, MDAs, IDAs, and HDAs.

Cross-connects are patch panels that allow cables to be connected to each other using patch cords. For example, the HC allows backbone cables to be patched to horizontal cables. An interconnect, such as a consolidation point in a ZDA, connects two cables directly through the patch panel. See Figure 12.8 for examples of cross-connects and interconnects used in data centers.

Note that switches can be patched to horizontal cabling (HC) using either a cross-connect or interconnect scheme. See the two diagrams on the right side of Figure 12.8. The interconnect scheme avoids another patch panel; however the cross-connect scheme may allow more compact cross-connects since the switches don't need to be located in or adjacent to the cabinets containing the HCs. Channels using Category 8, 8.1, or 8.2 for 25Gbase-T or 40GBase-T can only use the interconnect scheme as only two patch panels total are permitted from end to end.

Most of the components of the hierarchical star topology are optional. However, each cross-connect must have backbone cabling to a higher-level cross-connect:

- ENIs must have backbone cabling to an MC. They may also have backbone cabling to an IC or HC as required to ensure that WAN circuit lengths are not exceeded.
- HCs in TRs located in a data center must have backbone cabling to an MC and may optionally have backbone cabling to other distributors (ICs, HCs).
- ICs must have backbone cabling to an MC and one or more HCs. They may optionally have backbone cabling to an ENI or IC either for redundancy or to ensure that maximum cable lengths are not exceeded.
- HCs in an HDA must have backbone cabling to an MC or IC. They may optionally have backbone cabling to an HC, ENI, or IC either for redundancy or to ensure that maximum cable lengths are not exceeded.
- Because ZDAs only support horizontal cabling, they may only have cabling to an HDA or EDA.

Cross-connects such as the MC, IC, and HC should not be confused with the telecommunications spaces in which they are located: the MDA, IDA, and HDA. The cross-connects are components of the structured cabling system and are typically composed of patch panels. The spaces are dedicated rooms or more commonly dedicated cabinets, racks, or cages within the computer room.

EDAs and ZDAs may have cabling to different HCs to provide redundancy. Similarly, HCs, ICs, and ENIs may have redundant backbone cabling. The redundant backbone cabling may be to different spaces (for maximum redundancy) or between the same to spaces on both ends but follow different routes. See Figure 12.9 for degrees of redundancy in the structured cabling topology at various rating levels as defined in ANSI/TIA-942-B.

A rated 1 cabling infrastructure has no redundancy.

A rated 2 cabling infrastructure requires redundant access

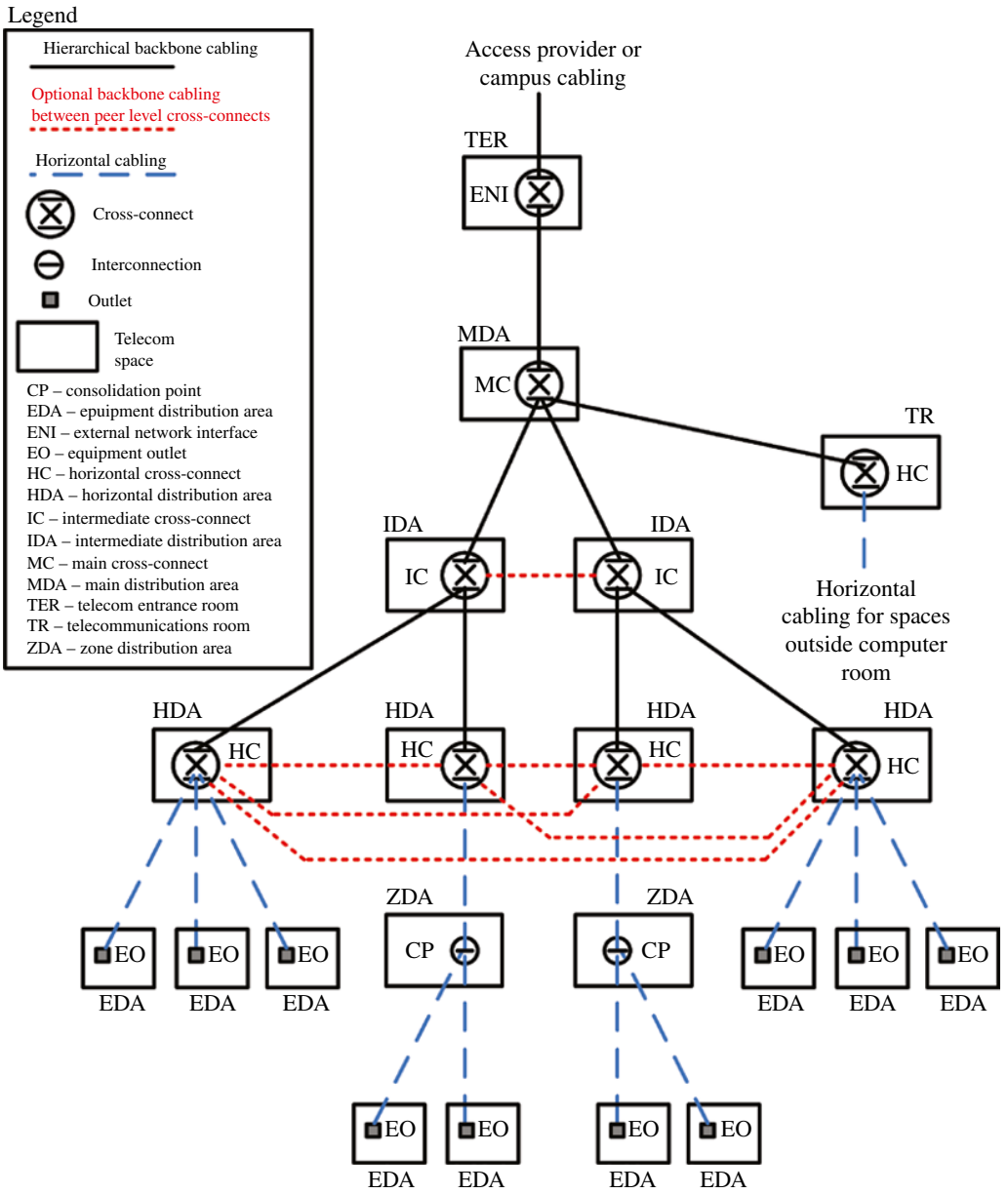


FIGURE 12.7 Hierarchical star topology. Source: © J&M Consultants, Inc.

provider (telecommunications carrier) routes into the data center. The two redundant routes must go to different carrier central offices and be separated from each other along their entire route by at least 20 m (66 ft).

A rated 3 cabling infrastructure has redundant TERs. The data center must be served by two different access providers (carriers). The redundant routes that the circuits take from the two different carrier central offices to the data center must be separated by at least 20 m (66 ft).

A rated 3 data center also requires redundant backbone cabling. The backbone cabling between any two cross-connects must use at least two separate cables, preferably following different routes within the data center.

A rated 4 data center adds redundant MDAs, IDAs, and HDAs. Equipment cabinets and racks (EDAs) must have horizontal cabling to two different HDAs. HDAs must have redundant backbone cabling to two different IDAs (if present) or MDAs. Each entrance room must have backbone cabling to two different MDAs.

12.6 CABLE TYPES AND MAXIMUM CABLE LENGTHS

There are several types of cables one can use for telecommunications cabling in data centers. Each has

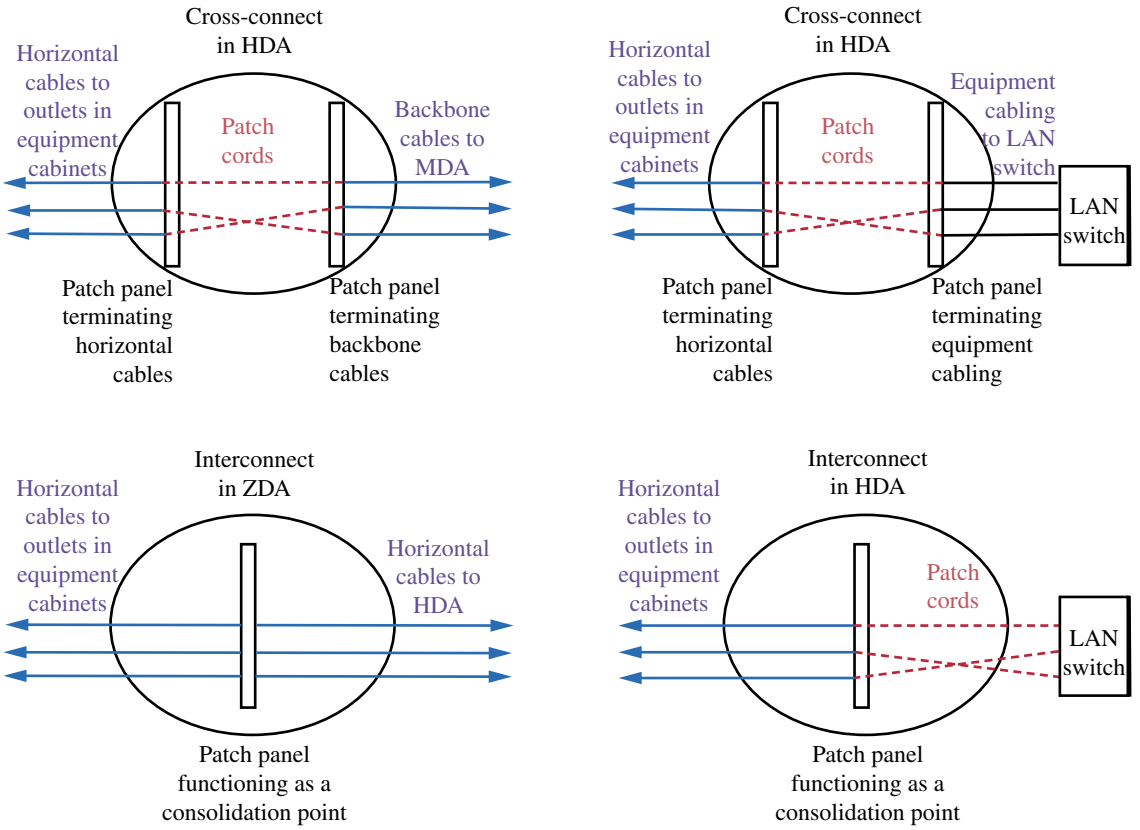


FIGURE 12.8 Cross-connects and interconnect examples. Source: © J&M Consultants, Inc.

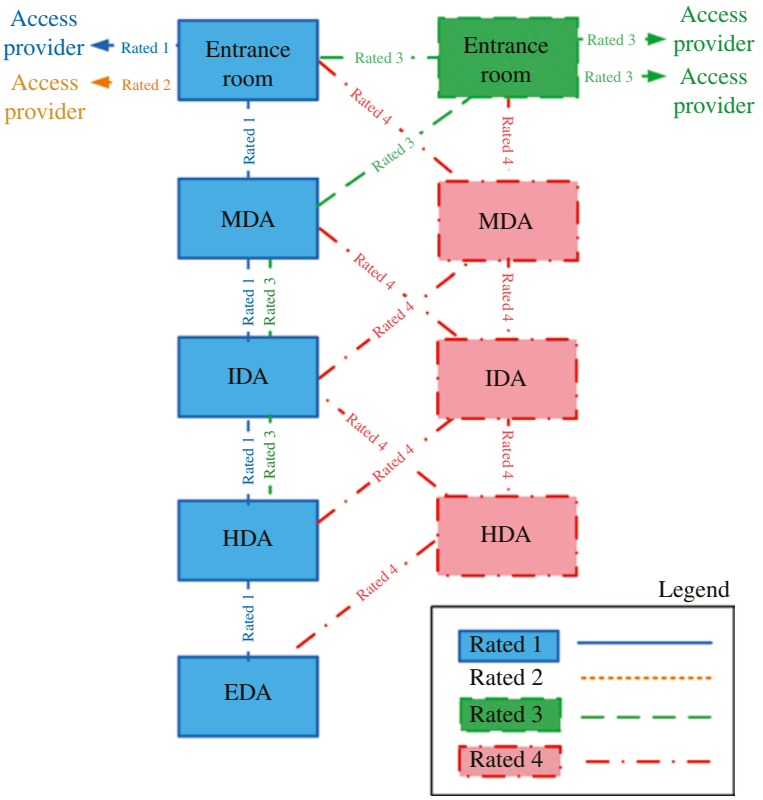


FIGURE 12.9 Structured cabling redundancy at various rating levels. Source: © J&M Consultants, Inc.

different characteristics and chosen to suit the various conditions to which they are subject. Some cables are more flexible than others. The size of the cable can affect its flexibility as well as its shield. A specific type of cable may be chosen because of space constrains or required load or because of bandwidth or channel capacity. Equipment vendors may also recommend cable for use with their equipment.

12.6.1 Coaxial Cabling

Coaxial cables are composed of a center conductor, surrounded by an insulator, surrounded by a metallic shield, and covered in a jacket. The most common types of coaxial cable used in data centers are the 75 ohm 734- and 735-type cables used to carry E-1, T-3, and E-3 wide area circuits; see Telcordia Technologies GR-139-CORE regarding specifications for 734- and 735-type cables and ANSI/ATIS-0600404.2002 for specifications regarding 75 ohm coaxial connectors.

Circuit lengths are longer for the thicker, less flexible 734 cable. These maximum cable lengths are decreased by intermediate connectors and DSX panels—see ANSI/TIA-942-B.

Broadband coaxial cable is also sometimes used in data centers to distribute television signals. The specifications of the broadband coaxial cables (Series 6 and Series 11) and connectors (F type) are specified in ANSI/TIA-568.4-D.

12.6.2 Balanced Twisted-Pair Cabling

The 100 ohm balanced twisted-pair cable is a type of cable that uses multiple pairs of copper conductors. Each pair of conductors is twisted together to protect the cables from electromagnetic interference.

- Unshielded twisted-pair (UTP) cables have no shield.
- The cable may have an overall cable screen made of either or both foil and braided shield.
- Each twisted pair may also have a foil shield.

Balanced twisted-pair cables come in different categories or classes based on the performance specifications of the cables. See Table 12.2.

Category 3, 5e, 6, and 6A cables are typically UTP cables but may have an overall screen or shield.

Category 7, 7A, and 8.2 cables have an overall shield and a shield around each of the four twisted pairs.

Category 8 and 8.1 cables have an overall shield.

Balanced twisted-pair cables used for horizontal cabling has 4 pairs. Balanced twisted-pair cables used for backbone cabling may have 4 or more pairs. The pair count above 4 pairs is typically a multiple of 25 pairs.

Types of balanced twisted-pair cables required and recommended in standards are as specified in Table 12.3.

Note that TIA-942-B recommends and ISO/IEC 11801-5 requires a minimum of Category 6A balanced twisted-pair cabling so as to be able to support 10G Ethernet. Category 6 cabling may support 10G Ethernet for shorter distances (less than 55 m), but it may require limiting the number of cables that support 10G Ethernet and other mitigation measures to function properly; see TIA TSB-155-A *Guidelines for the*

TABLE 12.2 Balanced twisted-pair categories

TIA categories	ISO/IEC and CENELEC classes/ categories	Max frequency (MHz)	Common application
Category 3	N/A	16	Voice, wide area network circuits, serial console, 10 Mbps Ethernet
Category 5e	Class D/Category 5	100	As above + 100 Mbps and 1 Gbps Ethernet
Category 6	Class E/Category 6	250	Same as above
Augmented Category 6 (Cat 6A)	Class EA/Category 6A	500	As above + 10G Ethernet
N/A	Class F/Category 7	600	Same as above
N/A	Class F_A/Category 7A	1,000	Same as above
Category 8	Class I/Category 8.1	2,000	As above + 25G and 40G Ethernet
N/A	Class II/Category 8.2	2,000	As above + 25G and 40G Ethernet

ISO/IEC and CENELEC categories refer to components such as cables and connectors. Classes refer to channels comprised of installed cabling including cables and connectors.
Note that TIA doesn't currently specify cabling categories above category 6A. However, higher performance Category 7/Class F and Category 7A/Class F_A are specified in ISO/IEC and CENELEC cabling standards.
Category 3 is no longer supported in ISO/IEC and CENELEC cabling standards.
Source: © J&M Consultants, Inc.

TABLE 12.3 Balanced twisted-pair requirements in standards

Standard	Type of cabling	Balanced twisted-pair cable categories/classes permitted
TIA-942-B	Horizontal cabling	Category 6, 6A, or 8, Category 6A or 8, recommended
TIA-942-B	Backbone cabling	Category 3, 5e, 6, or 6A, Category 6A or 8 recommended
ISO/IEC 11801-5	All cabling except network access cabling	Category 6A/E_A, 7/F, 7A/F_A, 8.1, or 8.2
ISO/IEC 11801-5	Network access cabling (to/from telecom entrance room/ENI)	Category 5/Class D, 6/E, 6A/E_A, 7/F, 7A/F_A, 8.1, or 8.2
CENELEC EN 51073-5	All cabling except network access cabling	Category 6/Class F, 6A/E_A, 7/F, 7A/F_A, 8.1, or 8.2
CENELEC EN 51073-5	Network access cabling (to/from telecom entrance room/ENI)	Category 5/Class D, 6/E, 6A/E_A, 7/F, 7A/F_A, 8.1, or 8.2

Source: © J&M Consultants, Inc.

Assessment and Mitigation of Installed Category 6 to Support 10GBase-T.

Category 8, 8.1, and 8.2 cabling are designed to support 25G and 40G Ethernet, but the end-to-end distances are limited to 30 m with only two-patch panel within the channel from switch to device.

12.6.3 Optical Fiber Cabling

Optical fiber is composed of a thin transparent filament, typically glass, surrounded by a cladding, which is used as a waveguide. Both single- and multimode fibers can be used over long distances and have high bandwidth. Single-mode fiber uses a thinner core, which allows only one mode (or path) of light to propagate. Multimode fiber uses a wider core, which allows multiple modes (or paths) of light to propagate. Multimode fiber uses less expensive transmitters and receivers but has less bandwidth than single-mode fiber. The bandwidth of multimode fiber reduces over distance, because light following different modes will arrive at the far end at different times.

There are five classifications of multimode fiber: OM1, OM2, OM3, OM4, and OM5. OM1 is 62.5/125 μm multimode optical fiber. OM2 can be either 50/125 μm or 62.5/125 μm multimode optical fiber. OM3 and OM4 are both 50/125 μm 850 nm laser-optimized multimode fiber, but OM4 optical fiber has higher bandwidth. OM5 is like OM4 but supports wave division multiplexing with four signals at slightly different wavelengths on each fiber.

A minimum of OM3 is specified in data center standards. TIA-942-B recommends the use of OM4 or OM5 multimode optical fiber cable to support longer distances for 100G and higher-speed Ethernet.

There are two classifications of single-mode fiber: OS1a and OS2. OS1a is a tight-buffered optical fiber cable used primarily indoors. OS2 is a loose-tube fiber (with the fiber sitting loose in a slightly larger tube) and is primarily for outdoor use. Both OS1a and OS2 use low water peak single-mode fiber that is processed to reduce attenuation at 1,400 nm frequencies allowing those frequencies to be used. Either type of single-mode optical fiber may be used in data centers, but OS2 is typically for outdoor use. OS1, tight-buffered single-mode optical fiber that is not a low fiber, is obsolete and no longer recognized in the standards.

12.6.4 Maximum Cable Lengths

The following Table 12.4 reflects the maximum circuit lengths over 734 and 735 type coaxial cables with only two connectors (one at each end) and no DSX panel.

Generally, the maximum length for LAN applications that are supported by balanced twisted-pair cables is 100 m (328 ft), with 90 m being the maximum length permanent link between patch panels and 10 m allocated for patch cords.

Channel lengths (lengths including permanently installed cabling and patch cords) for common data center LAN applications over multimode optical fiber are shown in Table 12.5. Channel lengths for single-mode optical fiber are several kilometers since single-mode fiber is used for long-haul communications.

TABLE 12.4 E-1, T-3, and E-3 circuits' lengths over coaxial cable

Circuit type	734 cable	735 cable
E-1	332 m (1088 ft)	148 m (487 ft)
T-3	146 m (480 ft)	75 m (246 ft)
E-3	160 m (524 ft)	82 m (268 ft)

Source: © J&M Consultants, Inc.

TABLE 12.5 Ethernet channel lengths over multimode optical fiber

Fiber type	1G Ethernet	10G Ethernet	25/40/50G Ethernet	40G Ethernet	100G Ethernet	200G Ethernet	400G Ethernet
# of fibers	2	2	2	8	4 or 8	8	8 (future) 32 (current)
OM1	275 m	26 m	Not supported	Not supported	Not supported	Not supported	Not supported
OM2	550 m	82 m	Not supported	Not supported	Not supported	Not supported	Not supported
OM3	**800 m**[a]	300 m	70 m	100 m	70 m	70 m	70 m
OM4	**1040 m**[a]	**550 m**[a]	100 m	150 m	100 m	100 m	100 m
OM5	**1040 m**[a]	**550 m**[a]	100 m	150 m	100 m	100 m	100 m (150 m with future 8 fiber version)

[a] Distances in bold are specified by manufacturers, but not in IEEE standards. Source: © J&M Consultants, Inc.

Refer to ANSI/TIA-568.0-D and ISO 11801-1 for tables that provide more details regarding maximum cable lengths for other applications.

12.7 CABINET AND RACK PLACEMENT (HOT AISLES AND COLD AISLES)

It is important to keep computers cool; computers create heat during operation, and heat decreases their functional life and processing speed, which in turn uses more energy and increases cost. The placement of computer cabinets or racks affects the effectiveness of a cooling system. Airflow blockages can prevent cool air from reaching computer parts and can allow heat to build up in poorly cooled areas.

One efficient method of placing cabinets is using hot and cold aisles, which creates convection currents that helps circulate air. See Figure 12.10. This is achieved by placing cabinets in rows with aisles between each row. Cabinets in each row are oriented such that they face one another. The hot aisles are the walkways with the rears of the cabinets on either side, and cold aisles are the walkways with the front of the cabinets on either side.

Telecommunications cables are placed under access floors and should be placed under the hot aisles so as to not restrict airflow if under-floor cooling ventilation is to be used. If power cabling is distributed under the access floors, the power cables should be placed on the floor in the cold aisles to ensure proper separation of power and telecommunications cabling. See Figure 12.10.

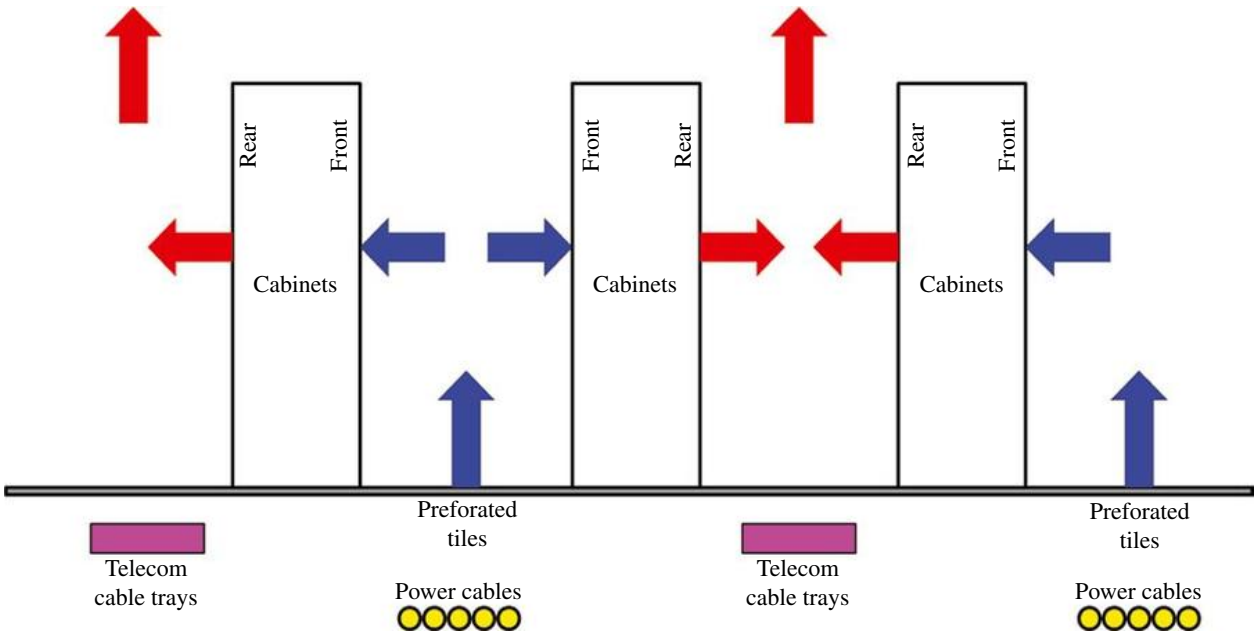


FIGURE 12.10 Hot and cold aisle example. Source: © J&M Consultants, Inc.

Lighting and telecommunications cabling shall be separated by at least 5 in.

Power and telecommunications cabling shall be separated by the distances specified in ANSI/TIA-569-D or ISO/IEC 14763-2. Generally, it is best to separate large numbers of power cables and telecommunications cabling by at least 600 mm (2 ft). This distance can be halved if the power cables are completely surrounded by a grounded metallic shield or sheath.

The minimum clearance at the front of the cabinets and racks is 1.2 m (4 ft), the equivalent of two full tiles. This ensures that there is proper clearance at the front of the cabinets to install equipment into the cabinets—equipment is typically installed in cabinets from the front. The minimum clearance at the rear of cabinets and equipment at the rear of racks is 900 mm (3 ft). This provides working clearance at the rear of the equipment for technicians to work on equipment. If cool air is provided from ventilated tiles at the front of the cabinets, more than 1.2 m (4 ft) of clearance may be specified by the mechanical engineer to provide adequate cool air.

The cabinets should be placed such that either the front or rear edges of the cabinets align with the floor tiles. This ensures that the floor tiles at both at the rear of the cabinets can be lifted to access systems below the access floor. See Figure 12.11.

If power and telecommunications cabling are under the access floor, the direction of airflow from air-conditioning equipment should be parallel to the rows of cabinets and racks to minimum interference caused by the cabling and cable trays.

Openings in the floor tiles should only be made for cooling vents or for routing cables through the tile. Openings for floor tile for cables should minimize air pressure loss by not cutting excessively large holes and by using a device that restricts airflow around cables, like brushes or flaps. The holes for cable management should not create tripping hazards; ideally, they should be located either under the cabinets or under vertical cable managers between racks.

If there are no access floors, or if they are not to be used for cable distribution, cable trays shall be routed above cabinets and racks, and not above the aisles.

Sprinklers and lighting should be located above aisles rather than above cabinets, racks, and cable trays, where their efficiency will be significantly reduced.

12.8 CABLING AND ENERGY EFFICIENCY

There should be no windows in the computer room; it allows light and heat into the environmentally controlled area, which creates an additional heat load.

TIA-942-B specifies that the 2015 ASHRAE TC 9.9 guidelines be used for the temperature and humidity in the computer room and telecommunications spaces.

ESD could be a problem at low humidity (dew point below 15°C [59°F], which corresponds approximately to

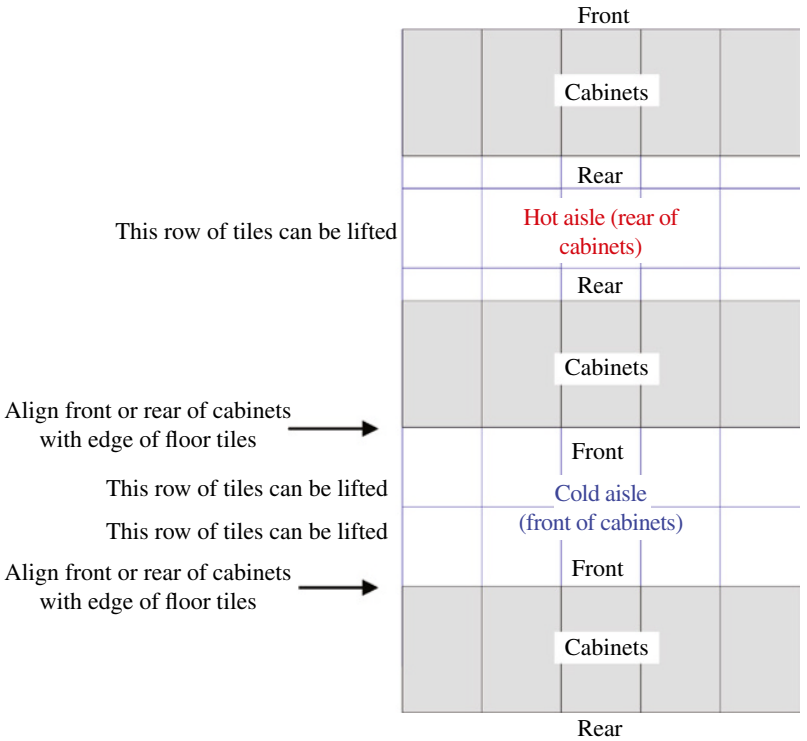


FIGURE 12.11 Cabinet placement example. Source: © J&M Consultants, Inc.

44% relative humidity at 18°C [64°F] and 25% relative humidity at 27°C [81°F]). Follow the guidelines in TIA TSB-153 *Static Discharge Between LAN Cabling and Data Terminal Equipment* for mitigation of ESD if the data center will operate in low humidity for extended periods. The guidelines include use of grounding patch cords to dissipate ESD built up on cables and use wrist straps per manufacturers' guidelines when working with equipment.

The attenuation of balanced twisted-pair telecommunications cabling will increase as temperatures increase. Since the ASHRAE guidelines permit temperatures measured at inlets to be as high as 35°C (95°F), temperatures in the hot aisles where cabling may be located can be as high as 55°C (131°F). See ISO/IEC 11801-1, CENELEC EN 50173-1, or ANSI/TIA-568.2-D for reduction in maximum cable lengths based on the average temperature along the length of the cable. Cable lengths may be further decreased if the cables are used to power equipment, since the cables themselves will also generate heat.

TIA-942-B recommends that energy-efficient lighting such as LED be used in the data center and that the data center follow a three-level lighting protocol depending on human occupancy of each space:

- Level 1: With no occupants, the lighting level should only be bright enough to meet the needs of the security cameras.
- Level 2: Detection of motion triggers higher lighting levels to provide safe passage through the space and to permit security cameras to identify persons.
- Level 3: This level is used for areas occupied for work—these areas shall be lit to 500 lux.

Cooling can be affected both positively and negatively by the telecommunications and IT infrastructure. For example, use of the hot aisle/cold aisle cabinet arrangement described above will enhance cooling efficiency. Cable pathways should be designed and located so as to minimize interference with cooling.

Generally, overhead cabling is more energy efficient than under-floor cabling if the space under the access floor is used for cooling since overhead cables will not restrict airflow or cause turbulence.

If overhead cabling is used, the ceilings should be high enough so that air can circulate freely around the hanging devices. Ladders or trays should be stacked in layers in high capacity areas so that cables are more manageable and do not block the air. If present, optical fiber patch cords should be protected from copper cables.

If under-floor cabling is used, they will be hidden from view, which will give a cleaner appearance. Installation is generally easier. Care should be made to separate telecommunications cables from the under-floor electrical wiring. Smaller cable diameters should be used. Shallower, wider cable trays are preferred as they don't obstruct under-floor airflow as much. Additionally, if under-floor air conditioning is used, cables from cabinets should run in the same direction of airflow to minimize air pressure attenuation.

Either overhead or under-floor cable trays should be no deeper than 6 in (150 mm). Cable trays used for optical fiber patch cords should have solid bottoms to prevent microbends in the optical fibers.

Enclosure or enclosure systems can also assist with air-conditioning efficiency. Consider using systems such as:

- Cabinets with isolated air returns (e.g., chimney to plenum ceiling space) or isolated air supply.
- Cabinets with in-cabinet cooling systems (e.g., door cooling systems).
- Hot aisle containment or cold aisle containment systems—note that cold aisle containment systems will generally mean that most of the space including the space occupied by overhead cable trays will be warm.
- Cabinets that minimize air bypass between the equipment rails and the side of the cabinet.

The cable pathways, cabinets, and racks should minimize the mixing of hot and cold air where not intended. Openings in cabinets, access floors, and containment systems should have brushes, grommets, and flaps at cable openings to decrease air loss around cable holes.

The equipment should match the cooling scheme—that is, equipment should generally have air intakes at the front and exhaust hot air out the rear. If the equipment does not match this scheme, the equipment may need to be installed backward (for equipment that circulates air back to front) or the cabinet may need baffles (for equipment that has air intakes and exhausts at the sides).

Data center equipment should be inventoried. Unused equipment should be removed (to avoid powering and cooling unnecessary equipment).

Cabinets and racks should have blanking panels at unused spaces to avoid mixing of hot and cold air.

Unused areas of the computer room should not be cooled. Compartmentalization and modular design should be taken into consideration when designing the floor plans; adjustable room dividers and multiple rooms with dedicated HVACs allow only the used portions of the building to be cooled and unoccupied rooms to be inactive.

Also, consider building the data center in phases. Sections of the data center that are not fully built require less capital and operating expenses. Additionally, since future needs may be difficult to predict, deferring construction of unneeded data center space reduces risk.

12.9 CABLE PATHWAYS

Adequate space must be allocated for cable pathways. In some cases either the length of the cabling (and cabling pathways) or the available space for cable pathways could limit the layout of the computer room.

Cable pathway lengths must be designed to avoid exceeding maximum cable lengths for WAN circuits, LAN connections, and SAN connections:

- Length restrictions for WAN circuits can be avoided by careful placement of the entrance rooms, demarcation equipment, and wide area networking equipment to which circuits terminate. In some cases, large data centers may require multiple entrance rooms.
- Length restrictions for LAN and SAN connections can be avoided by carefully planning the number and location of MDAs, IDAs, and HDAs where the switches are commonly located.

There must be adequate space between stacked cable trays to provide access for installation and removal of cables. TIA and BICSI standards specify a separation of 12 in (300 mm) between the top of one tray and the bottom of the tray above it. This separation requirement does not apply to cable trays run at right angles to each other.

Where there are multiple ratings of cable trays, the depth of the access floor or ceiling height could limit the number of cable trays that can be placed.

Standards in the NFPA and National Electrical Code limit the maximum depth of cable and cable fill of cable trays:

- Cabling inside cable trays must not exceed a depth of 150 mm (6 in) regardless of the depth of the tray.
- With cable trays that do not have solid bottom, the maximum fill of the cable trays is 50% by cross-sectional area of the cables.
- With cable trays that have solid bottoms, the maximum fill of the cable trays is 40%.

Cables in under-floor pathways should have a clearance of at least 50 mm (2 in) from the bottom of the floor tiles to the top of the cable trays to provide adequate space between the cable trays and the floor tiles to route cables and avoid damage to cables when floor tiles are placed.

Optical fiber patch cords should be placed in cable trays with solid bottoms to avoid attenuation of signals caused by micro-bends.

Optical fiber patch cords should be separated from other cables to prevent the weight of other cables from damaging in the fiber patch cords.

When they are located below the access floors, cable trays should be located in the cold aisles. When they are located overhead, they should be located above the cabinets and racks. Lights and sprinklers should be located above the aisles rather than the cable trays and cabinets/racks.

Cabling shall be at least 5 in (130 mm) from lighting and adequately separated from power cabling as previously specified.

12.10 CABINETS AND RACKS

Racks are frames with side mounting rails on which equipment may be fastened. Cabinets have adjustable mounting rails, panels, and doors and may have locks. Because cabinets are enclosed, they may require additional cooling if natural airflow is inadequate; this may include using fans for forced airflow, minimizing return airflow obstructions, or liquid cooling.

Empty cabinet and rack positions should be avoided. Cabinets that have been removed should be replaced, and gaps should be filled with new cabinets/racks with panels to avoid recirculation of hot air.

If doors are installed in cabinets, there should be at least 63% open space on the front and rear doors to allow for adequate airflow. Exceptions may be made for cabinets with fans or other cooling mechanisms (such as dedicated air returns or liquid cooling) that ensure that the equipment is adequately cooled.

In order to avoid difficulties with instillation and future growth, consideration should be taken when designing and installing the preliminary equipment. 480 mm (19 in) racks should be used for patch panels in the MDA, IDA, and HDA, but 585 mm (23 in) racks may be required by the service provider in the entrance room. Both racks and cabinets should not exceed 2.4 m (8 ft) in height.

Except for cable trays/ladders for patching between racks within the MDA, IDA, or HDA, it is not desirable to secure cable ladders to the top of cabinets and racks as it may limit the ability to replace the cabinets and racks in the future.

To ensure that infrastructure is adequate for unexpected growth, vertical cable management size should be calculated by the maximum projected fill plus a minimum of 50% growth.

The cabinets should be at least 150 mm (6 in) deeper than the deepest equipment to be installed.

12.11 PATCH PANELS AND CABLE MANAGEMENT

Organization becomes increasingly difficult as more interconnecting cables are added to equipment. Labeling both cables and patch panels can save time, as accidentally switching or removing the wrong cable can cause outages

that can take an indefinite amount of time to locate and correct. The simplest and most reliable method of avoiding patching errors is by clearly labeling each patch panel and each end of every cable as specified in ANSI/TIA-606-C.

However, this may be difficult if high-density patch panels are used. It is not generally considered a good practice to use patch panels that have such high density that they cannot be properly labeled.

Horizontal cable management panels should be installed above and below each patch panel; preferably, there should be a one-to-one ratio of horizontal cable management to patch panel unless angled patch panels are used. If angled patch panels are used instead of horizontal cable managers, vertical cable managers should be sized appropriately to store cable slack.

Separate vertical cable managers are typically required with racks unless they are integrated into the rack. These vertical cable managers should provide both front and rear cable management.

Patch panels should not be installed on the front and back of a rack or cabinet to save space, unless both sides can be easily accessed from the front.

12.12 RELIABILITY RATINGS AND CABLING

Data center infrastructure ratings have four categories: telecommunications (T), electrical (E), architectural (A), and mechanical (M). Each category is rated from one to four with one providing the lowest availability and four providing the highest availability. The ratings can be written as $T_N E_N A_N M_N$, with TEAM standing for the four categories and *N* being the rating of the corresponding category. Higher ratings are more resilient and reliable but more costly. Higher ratings are inclusive of the requirements for lower ratings. So, a data center with rated 3 telecommunications, rated 2 electrical, rated 4 architectural, and rated 3 mechanical infrastructure would be classified as TIA-942 Rating $T_3E_2A_4M_3$. The overall rating for the data center would be rated 2, the rating of the lowest level portion of the infrastructure (electrical rated 2).

The TIA-942 rating classifications are specified in more detail in ANSI/TIA-942-B. There are also other schemes for assessing the reliability of data centers. In general, systems that require more detailed analysis of the design and operation of a data center provide a better indicator of the expected availability of a data center.

12.13 CONCLUSION AND TRENDS

The requirements of telecommunications cabling, including maximum cable lengths, size, and location of telecommunications distributors, and requirements for cable pathways influence the configuration and layout of the data center.

The telecommunications cabling infrastructure of the data center should be planned to handle the expected near-term requirements and preferably at least one generation of system and network upgrades to avoid the disruption of removing and replacing the cabling.

For current data centers, this means that:

- Balanced twisted-pair cabling should be Category 6A or higher.
- Multimode optical fiber should be OM4 or higher.
- Either install or plan capacity for single-mode optical fiber backbone cabling within the data center.

It is likely that LAN and SAN connections for servers will be consolidated. The advantages of consolidating LAN and SAN networks include the following:

- Fewer connections permit use of smaller form factor servers that cannot support a large number of network adapters.
- Reduces the cost and administration of the network because it has fewer network connections and switches.
- It simplifies support because it avoids the need for a separate Fibre Channel network to support SANs.

Converging LAN and SAN connections requires high-speed and low-latency networks. The common server connection for converged networks will likely be 10 or 40 Gbps Ethernet. Backbone connections will likely be 100 Gbps Ethernet or higher.

The networks required for converged networks will require low latency. Additionally, cloud computing architectures typically require high-speed device-to-device communication within the data center (e.g., server-to-storage array and server to server). New data center switch fabric architectures are being developed to support these new data center networks.

There are a wide variety of implementations of data center switch fabrics. See Figure 12.12 for an example of the fat-tree or leaf-and-spine configuration, which is one common implementation.

The various implementations and the cabling to support them are described in an ANSI/TIA-942-B. Common attributes of data center switch fabrics are (i) the need for much more bandwidth than the traditional switch architecture and (ii) many more connections between switches than the traditional switch architecture.

When planning data center cabling, consider the likely future need for data center switch fabrics.

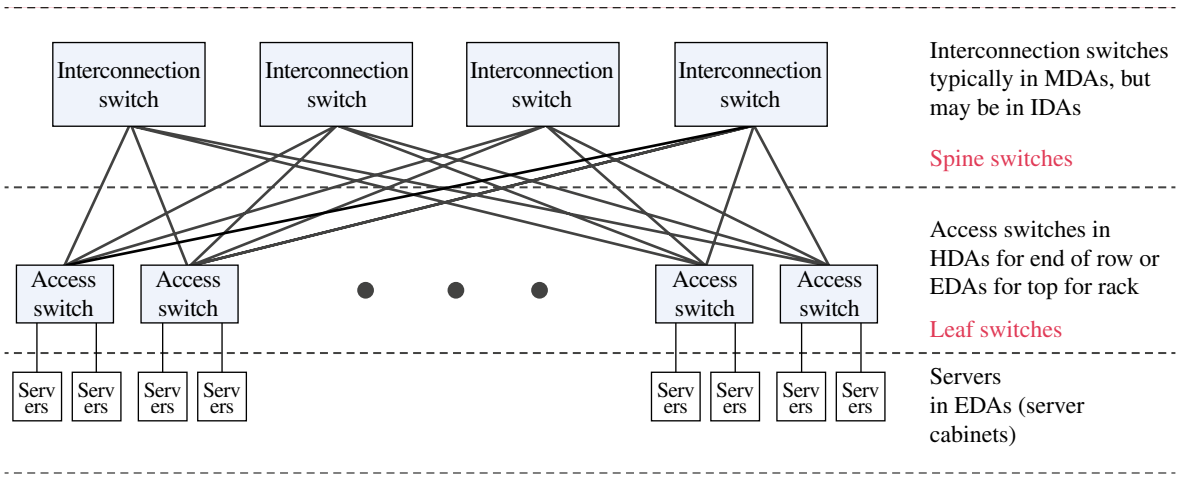


FIGURE 12.12 Data center switch fabric example. Source: © J&M Consultants, Inc.

FURTHER READING

For further reading, see the following telecommunications cabling standards

ANSI/BICSI-002. *Data Center Design and Implementation Best Practices Standard*.

ANSI/NECA/BICSI-607. *Standard for Telecommunications Bonding and Grounding Planning and Installation Methods for Commercial Buildings*.

ANSI/TIA-942-B. *Telecommunications Infrastructure Standard for Data Centers*.

ANSI/TIA-568.0-D. *Generic Telecommunications Cabling for Customer Premises*.

ANSI/TIA-569-D. *Telecommunications Pathways and Spaces*.

ANSI/TIA-606-C. *Administration Standard for Telecommunications Infrastructure*.

ANSI/TIA-607-C. *Telecommunications Bonding and Grounding (Earthing) for Customer Premises*.

ANSI/TIA-758-B. *Customer-Owned Outside Plant Telecommunications Infrastructure Standard*.

In Europe, the TIA standards may be replaced by the equivalent CENELEC standard:

CENELEC EN 50173-5. *Information Technology: Generic Cabling – Data Centres*.

CENELEC EN 50173-1. *Information Technology: Generic Cabling – General Requirements*.

CENELEC EN 50174-1. *Information Technology: Cabling Installation – Specification and Quality Assurance*.

CENELEC EN 50174-2. *Information Technology: Cabling Installation – Installation Planning and Practices Inside Buildings*.

CENELEC EN 50310. *Application of Equipotential Bonding and Earthing in Buildings with Information Technology Equipment*.

In locations outside the United States and Europe, the TIA standards may be replaced by the equivalent ISO/IEC standard.

ISO/IEC 11801-5. *Information Technology: Generic Cabling Systems for Data Centres*.

ISO/IEC 11801-1. *Information Technology: Generic Cabling for Customer Premises*.

ISO/IEC 14763-2. *Information Technology: Implementation and Operation of Customer Premises Cabling – Planning and Installation*.

Also note that standards are being continually updated; please refer to the most recent edition and all addenda to the listed standards.

13

AIR-SIDE ECONOMIZER TECHNOLOGIES

Nicholas H. Des Champs, Keith Dunnavant and Mark Fisher
Munters Corporation, Buena Vista, Virginia, United States of America

13.1 INTRODUCTION

The development and use of computers for business and science was a result of attempts to remove the drudgery of many office functions and to speed the time required to do mathematically intensive scientific computations. As computers developed from the 1950s tube-type mainframes, such as the IBM 705, through the minicomputers of the 70s and 80s, they were typically housed in a facility that was also home to many of the operation's top-level employees. And, because of the cost of these early computers and the security surrounding them, they were housed in a secure area within the main facility. It was not uncommon to have them in an area enclosed in lots of glass so that the computers and peripheral hardware could be seen by visitors and employees. It was an asset that presented the operation as one that was at the leading edge of technology.

These early systems generated considerably more heat per instruction than today's servers. Also, the electronic equipment was more sensitive to temperature, moisture, and dust. As a result, the computer room was essentially treated as a modern-day clean room. That is, high-efficiency filtration, humidity control, and temperatures comparable to operating rooms were standard. Since the computer room was an integral part of the main facility and had numerous personnel operating the computers and the many varied pieces of peripheral equipment, maintaining the environment was considered by the facilities personnel as a more precise form of "air conditioning."

Development of the single-chip microprocessor during the mid-1970s is considered to be the beginning of an era in which computers would be low enough in cost and had the power to perform office and scientific calculation, allowing individuals to have access to their own "personal" computers. The early processors and their host computers produced very little heat and were usually scattered throughout a department. For instance, an 8086 processor (refer to Table 13.1) generated less than 2 W of heat, and its host computer generated on the order of 25 W of heat (without monitor). Today's servers can generate up to 500 W of heat or more and when used in modern data centers (DCs) are loaded into a rack and can result in very high densities of heat in a very small footprint. Consider a DC with 200 racks at a density of 20,000 W/rack that results in 4 MW of heat to dissipate in a very small space.

Of course, there would be no demand for combining thousands of servers in large DCs had it not been for the development of the Internet and launching of the World Wide Web (WWW) in 1991 (at the beginning of 1993 only 50 servers were known to exist on the WWW), development of sophisticated routers, and many other ancillary hardware and software products. During the 1990s, use of the Internet and personal computers mushroomed as is illustrated by the rapid growth in routers: in 1991 Cisco had 251 employees and $70 million in sales, and by 1997 it had 11,000 employees and $7 billion in sales. Another example of this growth is shown by the increasing demand for server capacity: in 2011 there were 300 million new websites created, bringing the total to 555 million by the end of that year. The total number of Internet servers worldwide is estimated to be greater than 75 million.

As technology has evolved during the last several decades, so have the cooling requirements. No longer is a new DC "air-conditioned," but instead it is considered "process cooling"

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

TABLE 13.1 Chronology of computing processors

Processor	Clock speed	Introduction	Mfg. process	Transistors
4004	108 KHz	November 1971	10 μm	2,300
8086	10 MHz	June 1978	3 μm	29,000 1.87 W (sustained)
386	33 MHz	June 1988	1.5 μm	275,000
486	33 MHz	November 1992	0.8 μm	1.4 million
Pentium	66 MHz	March 1993	0.8 μm	3.1 million
Pentium II	233 MHz	May 1997	0.35 μm	7.5 million
Pentium III	900 MHz	March 2001	0.18 μm	28 million
Celeron	2.66 GHz	April 2008	65 nm	105 million
Xeon MP X7460	2.66 GHz	September 2008	45 nm	1.90 billion 170.25 W (sustained)

Source: Intel Corporation.

where air is delivered to a cold aisle, absorbs heat as it traverses the process, is sent to a hot aisle, and then is either discarded to ambient or returned to dedicated machines for extraction of the process heat and then sent back to the cold aisle. Today's allowable cooling temperatures reflect the conceptual change from air conditioning (AC) to process cooling. There have been four changes in ASHRAE's cooling guidelines [1] during the last nine years. In 2004, ASHRAE recommended Class 1 temperature was 68–77°F (20–25°C); in 2008 it was 64.4–80.6°F (18–27°C). In 2012, the guidelines remained the same in terms of *recommended* range but greatly expand the *allowable* range of temperatures and humidity in order to give operators more flexibility in doing compressor-less cooling (using ambient air directly or indirectly) to remove the heat from the DC with the goal of increasing the DC cooling efficiency and reducing the energy efficiency metric, power usage effectiveness (PUE). Today, the 2015 guidelines further expanded the recommended range to a lower humidity level, reducing the amount of humidification needed to stay within the range.

13.2 USING PROPERTIES OF AMBIENT AIR TO COOL A DATA CENTER

In some instances it is the ambient conditions that are the principal criteria that determine the future location of a DC, but most often the location is based on acceptance by the community, access to networks, and adequate supply and cost of utilities in addition to being near the market it serves. Ambient conditions have become a more important factor as a result of an increase in allowable cooling temperature for the information technology (IT) equipment. The cooler, and sometimes drier, the climate, the greater period of time a DC can be cooled by using ambient air. For instance, in Reno, NV, air can be supplied all year at 72°F (22°C) with no mechanical refrigeration by using evaporative cooling techniques.

Major considerations by the design engineers when selecting the cooling system for a specific site are:

(a) Cold aisle temperature and maximum temperature rise across server rack
(b) Critical nature of continuous operation for individual servers and peripheral equipment
(c) Availability of sufficient usable water for use with evaporative cooling
(d) Ambient design conditions, i.e., yearly typical design as well as extremes of dry-bulb (db) and wet-bulb (wb) temperature
(e) Site air quality, i.e., particulate and gases
(f) Utility costs

Other factors are projections of initial capital cost, full-year cooling cost, reliability, complexity of control, maintenance cost, and the effectiveness of the system in maintaining the desired space temperature, humidity, and air quality during normal operation and during a power or water supply failure.

Going forward, the two air-side economizer cooling approaches, direct and indirect, are discussed in greater detail. A *d*irect *a*ir-*s*ide *e*conomizer (DASE) takes outdoor air (OA), filters and conditions it, and delivers it directly to the space. An *i*ndirect *a*ir-*s*ide *e*conomizer (IASE) uses ambient air to indirectly cool the recirculating airstream without delivering ambient air to the space. Typically a DASE system will include a direct evaporative cooler (DEC) cooling system for cooling; ambient air traverses

the wetted media, lowering the db temperature, and is controlled to limit the amount of moisture added to keep the space within the desired RH%. An IASE system typically uses some form of air-to-air heat exchanger (AHX) that does not transfer latent energy between airstreams. Typically, plate-type, tubular, thermosiphon, or heat pipe heat exchangers are used. Please refer to Ref. [2] for information on AHXs.

13.3 ECONOMIZER THERMODYNAMIC PROCESS AND SCHEMATIC OF EQUIPMENT LAYOUT

13.3.1 Direct Air-Side Economizer (DASE)

13.3.1.1 Cooling with Ambient Dry-Bulb Temperature

The simplest form of an air-side economizer uses ambient air directly supplied to the space to remove heat generated by IT equipment. Figure 13.1 shows a schematic of a typical DASE arrangement that includes a DEC, item 1, and a cooling coil, item 2. Without item 1 this schematic would represent a DASE that uses the db temperature of the ambient air to cool the DC. For this case, ambient air can be used to perform all the cooling when its temperature is below the design cold aisle temperature and a portion of the cooling when it is below the design hot aisle temperature. When ambient temperature is above hot aisle temperature, or ambient dew point (dp) exceeds the maximum allowed by the design, then the system must resort to full recirculation and all mechanical cooling. When ambient temperature is below the design cold aisle temperature, some of the heated process air is returned to the inlet plenum to mix with the incoming OA to yield the desired delivery temperature. In almost all cases, except in extreme cold climates, some level of mechanical cooling is required to meet the space cooling requirements, and, in most cases, the mechanical supplement will be designed to handle the full cooling load. The result is that for most regions of the world, the full-year energy reduction is appreciable, but the capital equipment cost reflects the cost of having considerable mechanical refrigeration on board. Other factors to consider are costs associated with bringing high levels of OA into the building that result in higher rate of filter changes and less control of space humidity. Also, possible gaseous contaminants, not captured by standard high-efficiency filters, could pose a problem.

13.3.1.2 Cooling with Ambient Wet-Bulb Temperature

If a source of usable water is available at the site, then an economical approach to extend the annual hours of economizer cooling, as discussed in the previous paragraph, is to add a DEC, item 1, as shown in Figure 13.1. The evaporative pads in a DEC typically can achieve 90–95% efficiency in cooling the ambient air to approach wb temperature from db temperature, resulting in a db temperature being delivered to space at only a few degrees above the ambient wb temperature. The result is that the amount of trim mechanical cooling required is considerably reduced from using ambient db and in many cases may be eliminated completely. In addition, there is greater space humidity control by using the DEC to add water to the air during colder ambient conditions. The relative humidity within the space, during cooler periods, is

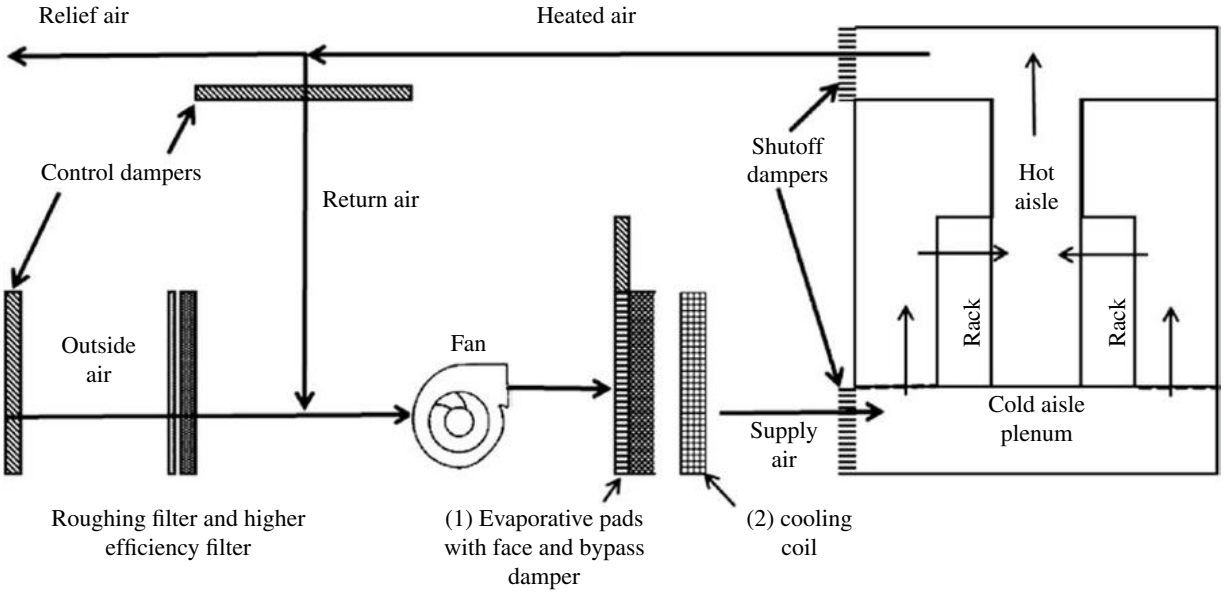


FIGURE 13.1 Schematic of a typical direct air-side economizer.

controlled with the face and bypass dampers on the DEC. It is important that the system is designed to prevent freeze conditions at the DEC or condensate formation in supply ductwork or outlet areas. There would be no humidity control however during the warmer ambient conditions. In fact, lack of humidity control is the single biggest drawback in using DASE with DEC. As with the db cooling, factors to consider are costs associated with bringing high levels of OA into the building, which results in higher rates of filter changes and less control of space humidity. Also, possible gaseous contaminants, not captured by standard high-efficiency filters, could pose a problem. Even with these operating issues, the DASE using DEC is arguably the most efficient and least costly of the many techniques for removing waste heat from DCs, except for DASE used on facilities in extreme climates where the maximum ambient db temperature never exceeds the specified maximum cold aisle temperature.

A DASE with DEC cooling process is illustrated in Figure 13.2. In this instance, the cold aisle temperature is 75°F, and the hot aisle is 95°F, which is a fairly typical 20°F temperature difference, or Delta T (ΔT) across the IT equipment. With an ambient design wb of 67.7°F and a 90% effective evaporative process, the supply air (SA) to the space can be cooled to 70°F from 91.2°F, which is lower than specified. Under this type of condition, there are several control schemes that are used to satisfy the space cooling requirements:

1. Reduce the process air flow to maintain the hot aisle temperature at 95°F, which increases the ΔT between the hot and cold aisles. Decreasing the process airflow results in considerably less fan power. This scheme is shown as the process between the two square end marks.
2. Maintain the specified 20°F ΔT by holding the process airflow at the design value, which results in a lower hot aisle temperature. This is shown in the horizontal process line starting from "Out of DEC" but only increasing up to 90°F db return temperature.
3. Use face and bypass dampers on the DEC to control the cold aisle SA temperature to 75°F as shown in the process between the two triangular end marks.

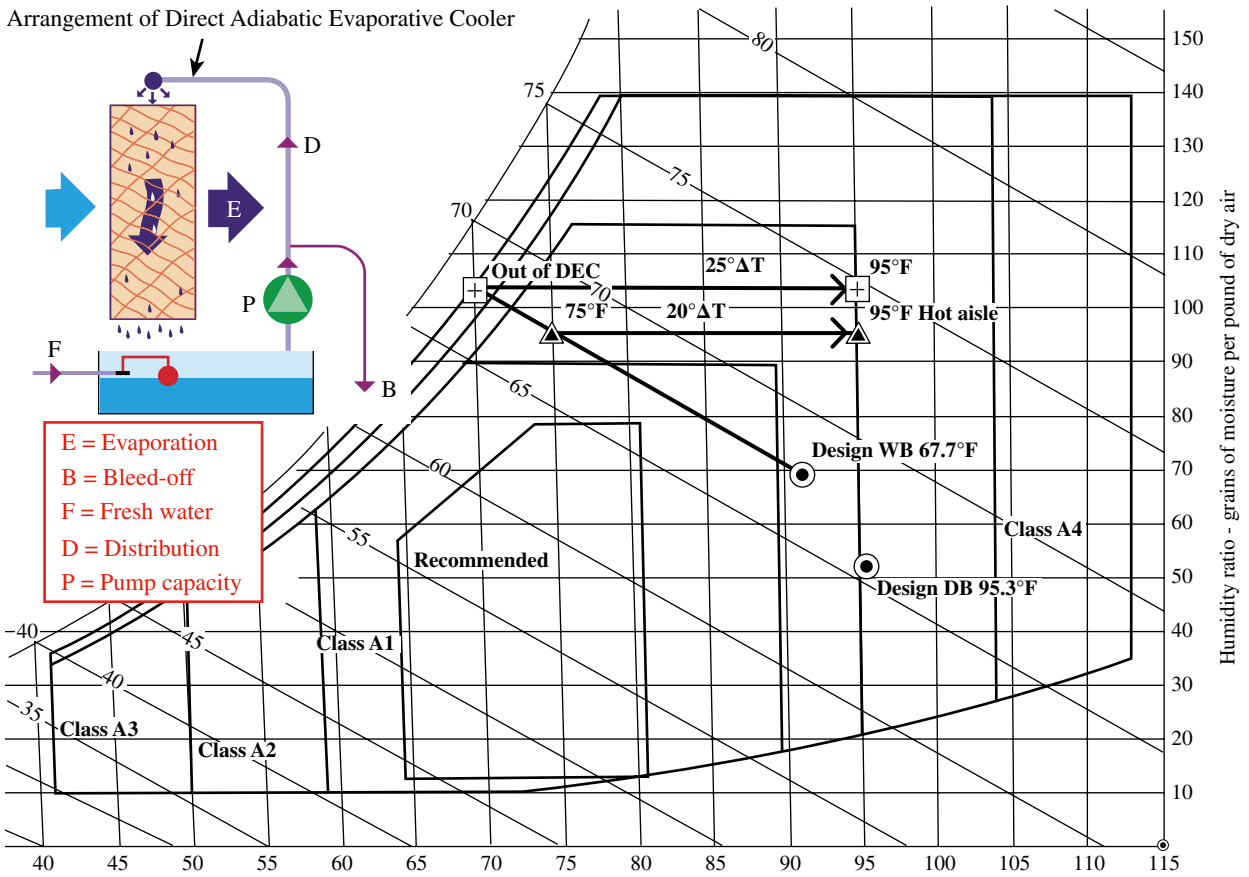


FIGURE 13.2 Direct cooling processes shown with ASHRAE recommended and allowable envelopes for DC supply temperature and moisture levels.

FIGURE 13.3 At left: cooling system using bank of DASE with DEC units at the end wall of a DC; at right: the array of evaporative cooling media. Source: Courtesy of Munters Corporation.

13.3.1.3 Arrangement of Direct Adiabatic Evaporative Cooler

A bank of multiple DASE with DEC units arranged in parallel is shown in Figure 13.3. Each of these units supplies 40,000 cubic feet per minute (CFM) of adiabatically cooled OA during warm periods and a blend of OA and recirculated air, as illustrated in Figure 13.1, during colder periods. The cooling air is supplied directly to the cold aisle, travels through the servers and other IT equipment, and is then directed to the relief dampers on the roof. Also shown in Figure 13.3 is a commonly used type of rigid, fluted direct evaporative cooling media.

13.3.2 Indirect Air-Side Economizer (IASE)

13.3.2.1 Air-to-Air Heat Exchangers

In many Datacom cooling applications, it is desirable to indirectly cool recirculated DC room air as opposed to delivering ambient OA directly into the space for cooling. This indirect technique allows for much more stable humidity control and significantly reduces the potential of airborne contaminants entering the space compared to DASE designs. When cooling recirculated air, dedicated makeup air units are added to the total cooling system to control space humidity and building pressure. An AHX serves as the intermediary that permits the use of ambient OA to cool the space without actually bringing the ambient OA into the space. The most commonly used types of AHX used for this purpose are plate and heat pipe as shown in Figure 13.4. Sensible wheel heat exchangers have also been used in IASE systems, but are no longer recommended due to concerns with air leakage, contaminant and/or humidity carryover, and higher air filtration requirements when compared with passive plate or heat pipe heat exchangers. Please refer to Ref. [2], Chapter 26, *Air-To-Air Energy Recovery Equipment*, for further information regarding performance and descriptions of AHX. Figure 13.5 illustrates the manner in which the AHX is used to transfer the heat from the hot aisle return air (RA) to the cooling air, com-

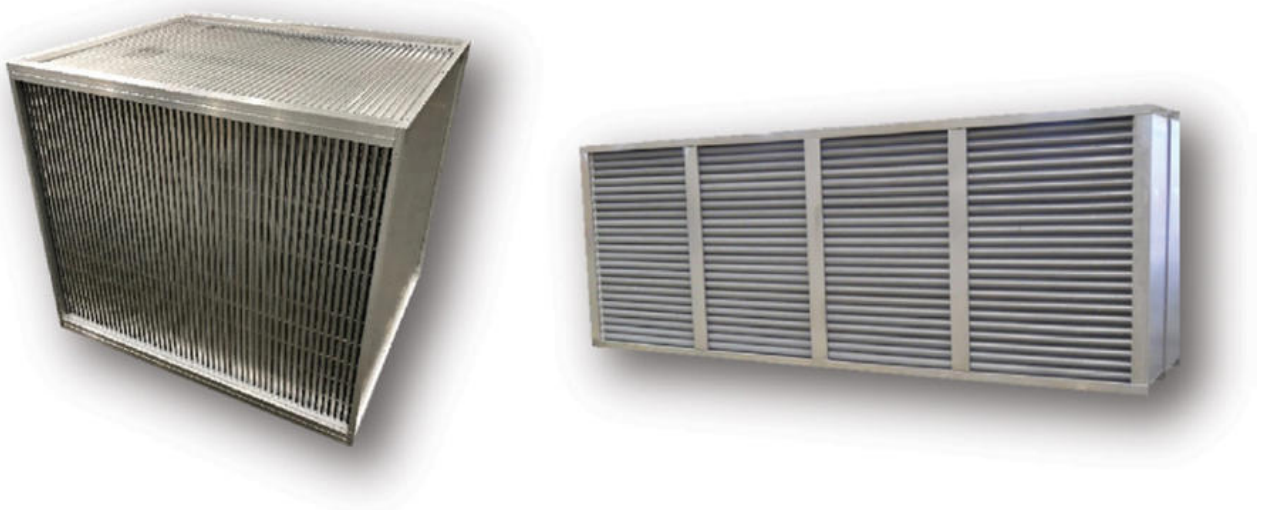


FIGURE 13.4 Plate-type (left) and heat pipe (right) heat exchangers.

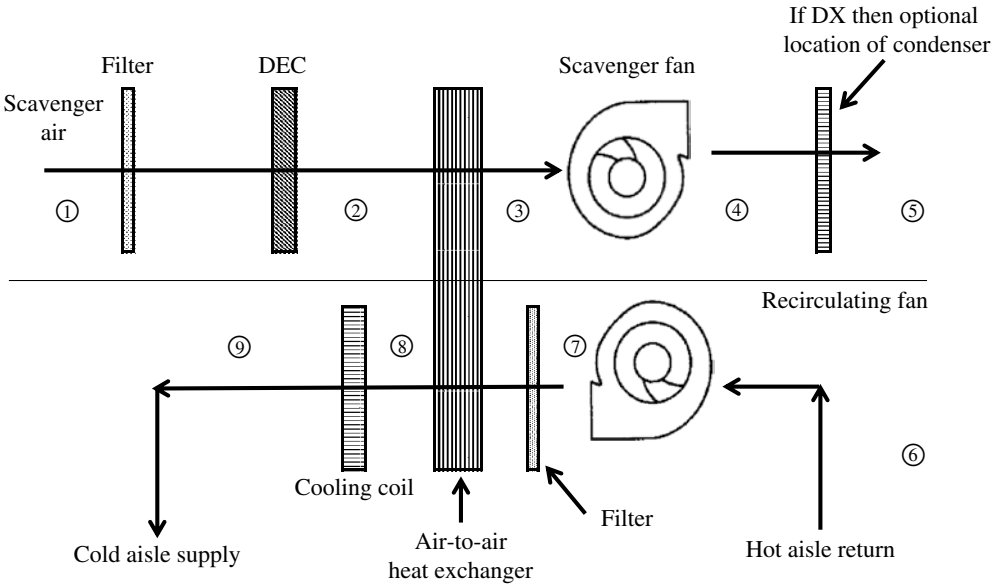


FIGURE 13.5 Schematic of typical indirect air-side economizer.

monly referred to as scavenger air (ScA) since it is discarded to ambient after it performs its intended purpose, that of absorbing heat. The effectiveness of an AHX, when taking into consideration the cost, size, and pressure drop, is usually selected to be between 65 and 75% when operating at equal airflows for the ScA and recirculating air.

Referring to the schematic shown in Figure 13.5, the ScA enters the system through a roughing filter at ① that removes materials that are contained in the OA that might hamper the operation of the components located in the scavenger airstream. If a sufficient amount of acceptable water is available at the site, then cooling the ScA with a DEC before it enters the AHX at ② should definitely be considered. Evaporatively cooling the ScA will not only extend the energy-saving capability of the IASE over a greater period of time, but it will and also reduce the amount of mechanical refrigeration required at the extreme ambient design conditions. The ambient conditions used for design of cooling equipment are generally extreme db temperature if just an AHX is used and extreme wb temperature if a form of evaporative cooling is used to precool the ScA before it enters the heat exchanger. Extreme ambient conditions are job dependent and are usually selected using either Typical Meteorological Year 3 (TMY3) data, the extreme ASHRAE data, or even the 0.4% ASHRAE annual design conditions.

When DEC is used as shown in Figure 13.5, and trim direct expansion refrigeration (DX) cooling is required, then it is advantageous to place the condenser coil in the leaving scavenger airstream since its db temperature, in almost all cases, is lower than the ambient db temperature. If no DEC is used, then there could be conditions where the ScA temperature is above the RA. Under these circumstances, there should be a means to prevent the AHX from transferring heat in the wrong direction; otherwise heat will be transferred from the ScA to the recirculating air, and the trim mechanical refrigeration will not be able to cool the recirculating air to the specified cold aisle temperature. Vertical heat pipe AHXs automatically prevent heat transfer at these extreme conditions because if the ambient OA is hotter than the RA, then no condensing of the heat pipe working fluid will occur (process ② to ③ as shown in Fig. 13.5), and therefore no liquid will be returned to the portion of the heat pipe in the recirculating airstream (process ⑦ to ⑧). With the plate heat exchanger, a face and bypass section to direct ScA around the AHX may be necessary in order to prevent heat transfer, or else the condenser will need to be in a separate section, which would allow the scavenger fans to be turned off.

As an example, when using just an AHX without DEC and assuming an effectiveness of 72.5% (again using 75°F cold aisle and 95°F hot aisle), the economizer can do all of the cooling when the ambient db temperature is below 67.4°F. At lower ambient temperatures the scavenger fans are slowed in order to remove the correct amount of heat and save on scavenger fan energy. Above 67.4°F ambient the mechanical cooling system is staged on until at an ambient of 95°F or higher the entire cooling load is borne by the mechanical cooling system.

When precooling ScA with a DEC, it is necessary to discuss the cooling performance with the aid of a psychrometric chart. The numbered points on Figure 13.6 correspond to the numbered locations shown in Figure 13.5. On a design wb day ① of 92°F db/67.7°F wb, the DEC lowers the ScA db

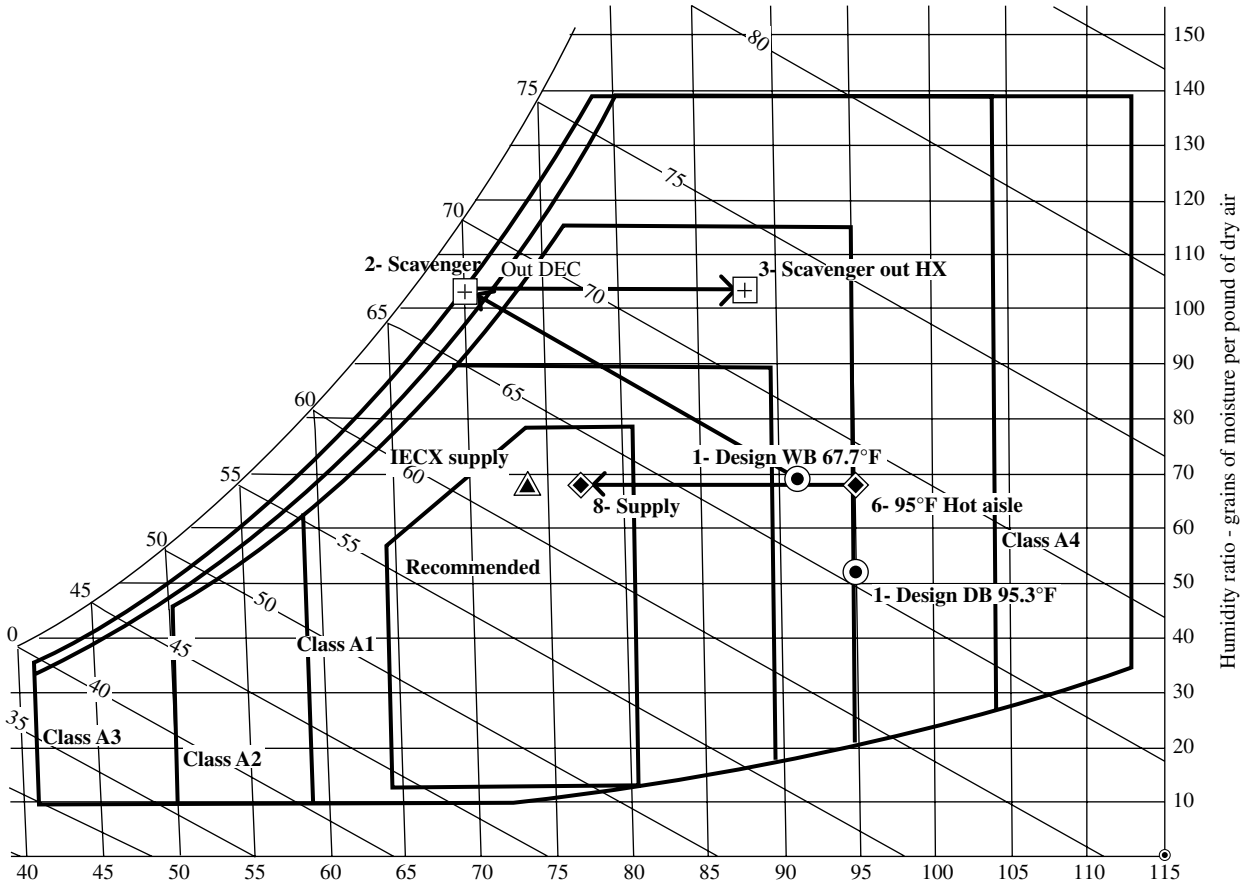


FIGURE 13.6 Psychrometric chart showing performance of IASE system with DEC precooling scavenger airstream.

temperature from 92 to 70.1°F ②. The ScA then enters the heat exchanger and heats to 88.2°F ③. During this process, air returning from the hot aisle ⑥ is cooled from 95°F (no fan heat added) to 77.2°F ⑧, or 89% of the required cooling load. Therefore, on a design day using DEC and an AHX, the amount of trim mechanical cooling required ⑨ in Figure 13.5 is only 11% of the full cooling load, and the trim would only be called into operation for a short period of time during the year.

13.3.2.2 Integral Air-to-Air Heat Exchanger/Indirect Evaporative Cooler

The previous section used a separate DEC and AHX to perform an indirect evaporative cooling (IEC) process. The two processes can be integrated into a single piece of equipment, known as an indirect evaporative cooling heat exchanger (IECX). The IECX approach, which uses wb temperature as the driving potential to cool Datacom facilities, can be more efficient than using a combination of DEC and AHX since the evaporative cooling process occurs in the same area as the heat removal process. It is important to note that with this process the evaporative cooling effect is achieved indirectly, meaning no moisture is introduced into the process airstream.

Configuration of a typical IECX is illustrated in Figure 13.7. The recirculating Datacom air returns from the hot aisle at, for example, 95°F and enters the horizontal tubes from the right side and travels through the inside of the tubes where it cools to 75°F. The recirculating air cools as a result of the indirect cooling effect of ScA evaporating water that is flowing downward over the outside of the tubes. Because of the evaporative cooling effect, the water flowing over the tubes and the tubes themselves approach ambient wb temperature. Typically, an IECX is designed to have wb depression efficiency (WBDE) in the range of 70–80% at peak ambient wb conditions. Referring to Figure 13.6, with all conditions remaining the same as the example with the dry AHX with a DEC precooler on the ScA, a 78% efficient IECX process is shown to deliver a cold aisle temperature of 73.7°F, shown on the chart as a triangle, which is below the required 75°F. Under these conditions the ScA fan speed is reduced to maintain the specified cold aisle temperature at 75 instead of 73.7°F.

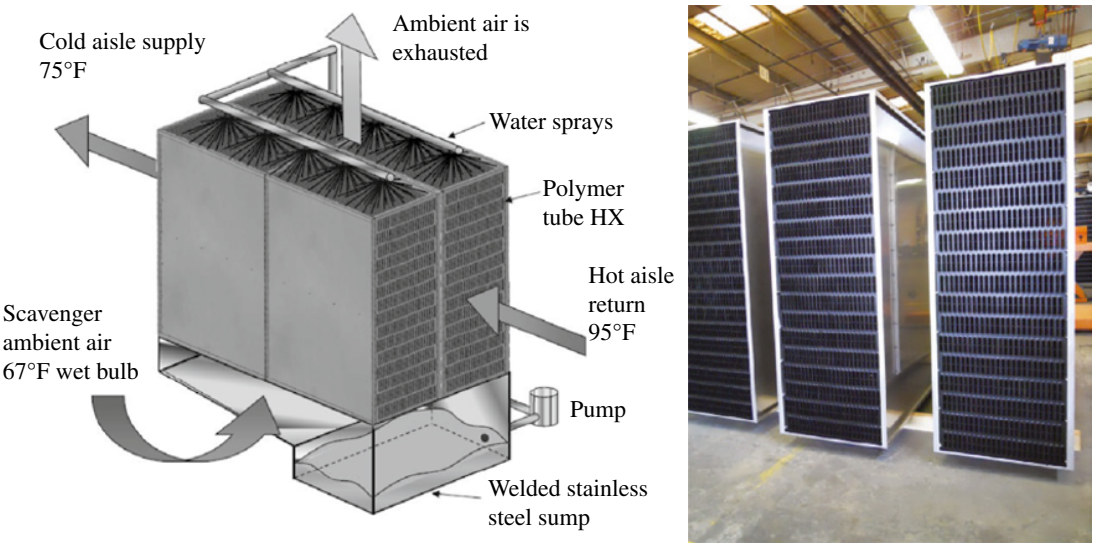
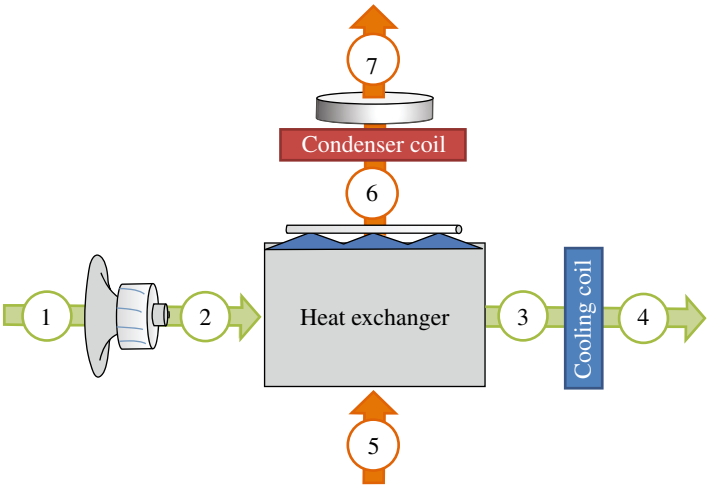


FIGURE 13.7 Indirect evaporative cooled heat exchanger (IECX). Source: Courtesy of Munters Corporation.

A unit schematic and operating conditions for a typical IECX unit design are shown in Figure 13.8. Referring to the airflow pattern in the schematic, air at 96°F comes back to the unit from the hot aisle ①, heats to 98.2°F through the fan ②, and enters the tubes of the IECX where it cools to 83.2°F ③ on a design ambient day of 109°F/75°F (db/wb). The trim DX then cools the supply to the specified cold aisle temperature of 76°F. At these extreme operating conditions, the IECX removes 67% of the heat load, and the DX removes the remaining 33% of the heat. This design condition will be a rare occurrence, but the mechanical trim cooling is sized to handle this extreme. For a facility in Quincy, WA, operating



Operating Point	Critical			Normal		
	DB (°F)	WB (°F)	ACFM	DB (°F)	WB (°F)	ACFM
T1	96.0	68.9	65,910	96.0	68.9	53,926
T2	98.2	69.5	66,171	97.6	69.4	54,081
T3	83.2	65.0	64,392	82.5	64.8	52,616
T4	76.0	62.6	63,538	76.0	62.6	51,985
T5	109.0	75.0	43,859	109.0	75.0	33,855
T6	81.0	78.5	41,940	81.0	78.5	34,286
T7	93.9	81.5	42,939	94.2	81.6	35,123
	ITE load rejected		380.7 kW	ITE load rejected		311.5 kW

FIGURE 13.8 Schematic of a typical DC cooling IECX unit. Source: Courtesy of Munters Corporation.

with these design parameters, the IEC economizer is predicted to remove 99.2% of the total annual heat load.

The period of time during the year that an economizer is performing the cooling function is extremely important because a Datacom facility utilizing economizer cooling has a lower PUE than a facility with conventional cooling using chillers and computer room air handler (CRAH) units or computer room air conditioner (CRAC) units. PUE is a metric used to determine the energy efficiency of a Datacom facility. PUE is determined by dividing the amount of power entering a DC by the power used to run the computer infrastructure within it. PUE is therefore expressed as a ratio, with overall efficiency improving as the quotient decreases toward 1. There is no firm consensus of the average PUE; from a survey of over 500 DCs conducted by the Uptime Institute in 2011 the average PUE was reported to be 1.8, but in 2012 the CTO of Digital Realty indicated that the average PUE for a DC was 2.5. Economizer PUE values typically range from as low as 1.07 for a DASE using DEC to a high of about 1.3, while IECX systems range from 1.1 to 1.2 depending upon the efficiency of the IECX and the site location. For example, if the economizer at a given location reduced the time that the mechanical refrigeration was operating by 99.7% during a year, then the cooling costs would be reduced by a factor of around 5 relative to a DC with the same server load operating at a PUE of 2.0.

Typically, in a DC facility where hot aisle containment is in place, the IECX system is able to provide cooling benefit even during the most extreme ambient design conditions. As a result, the mechanical refrigeration system, if it is required at all, is in most cases able to be sized significantly smaller than the full cooling load. This smaller amount of refrigeration is referred to as "trim DX," since it only has to remove a portion of the total heat load. A further benefit of the IECX system is that, referring again to Figure 13.8, the ScA leaving the IECX ⑥ is brought close to saturation (and thus cooler than the ambient temperature) before performing its second job, that of removing heat from the refrigeration condenser coil. This cooler temperature entering the condenser coil improves compressor performance with the resulting lower condensing temperature.

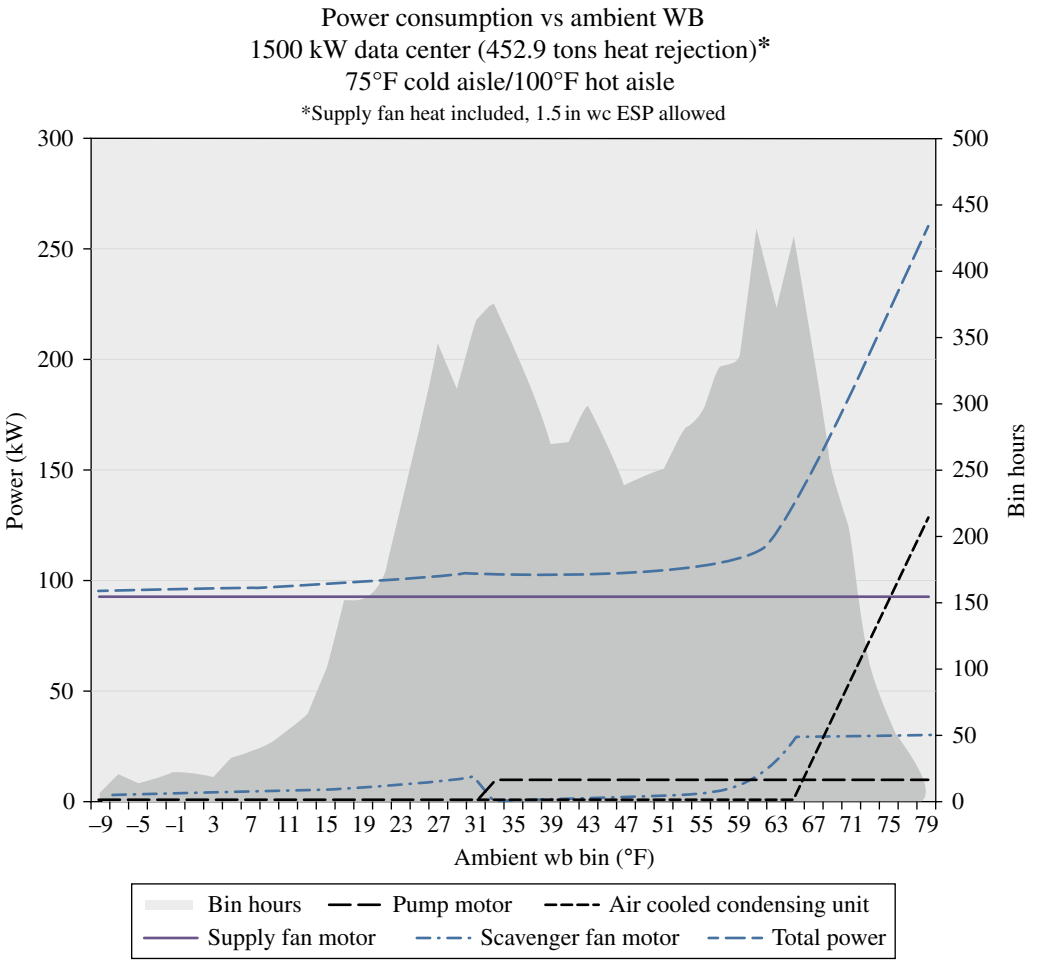


FIGURE 13.9 Power consumption for a typical IECX IASE cooling unit. Source: Courtesy of Munters Corporation.

Figure 13.9 shows a graph of operating power vs. ambient wb condition for a typical IECX unit, and the shaded area represents the number of hours in a typical year that each condition occurs in a given location, providing a set of bin hours (right ordinate) at which the IECX might operate at each wet bulb. Most of the hours are between about 11 and 75°F. The upper curve, medium dashed line, is the total operating power of the economizer cooling system. The short-dashed curve is the DX power and the dot-dash curve is the scavenger fan motor, both of which operate at full capacity at the extreme wb temperatures. The average weighted total power for the year is 117 kW. Typically, the lights and other electrical within the DC are about 3% of the IT load, so the total average load into the facility is 1500 kW × 1.03 + 117 kW or 1662 kW. This yields an average value of PUE of 1662/1500 or 1.108, an impressive value when compared with conventional cooling PUEs of 1.8–2.5. For this example, the onboard trim DX represented 24% of the 452.9 tons of heat rejection, which results in a lower connected load to be backed up with generators, as long as short-term water storage is provided.

One company that has experienced great success implementing IECX systems is colocation provider Sabey Data Centers Inc. Figure 13.10 illustrates an aerial view of one of several Sabey DC facilities at their Intergate campus located in Quincy, WA. This campus has one of the largest IECX installations in the Western United States with a reported annual PUE of 1.13. Overall, the annual PUE of the campus is less than 1.2, which is impressive considering that these colocation facilities have variable loads and are often operating at partial loads below the design capacity (and higher efficiency points) of the equipment.

In order to give even a better understanding of how the IECX performs at different climatic conditions and altitudes, Figure 13.11 shows the percentage of cooling ton-hours performed during the year: first the IECX operating wet (warm conditions using wb temperature), then second the IECX operating dry (cool conditions using db temperature), and third at extreme conditions the operation of DX. Fifteen cities are listed with elevations ranging from sea level to over 5,000 ft. The embedded chart gives a graphical representation of the energy saved during each operating mode. The last column is the percentage of time during the year that there are no compressors staged on and the IECX is handling the entire cooling load.

13.3.2.3 Trim Cooling Drain Trap Considerations

When using an indirect economizer in combination with a cooling coil used for trim cooling, there will be extended periods of time when mechanical cooling is not active. This can lead to "dry-out" of the condensate traps, resulting in air leakage in or out of the recirculating air handler. This situation can impact pressurization control within the DC, and can also increase the amount of conditioned make-up air required. It is recommended that the cooling coil drain traps be designed to prevent dry-out and the resulting airflow within the condensate drain line, such as use of an Air-Trap (which uses the fan pressure to trap air) instead of a P-Trap (which uses water to trap air).

FIGURE 13.10 Aerial view of (42) indirect air-side economizers (IASEs). Source: Courtesy of Munters Corporation.

Location	Elevation (ft)	0.4% WB design (MCDB/WB °F)	% reduction of peak mechanical cooling requirement*	% annual ton-hours IASE (wet)	% annual ton-hours IASE (dry)	% annual ton-hours mechanical cooling	% annual hours mechanical cooling is off
Ashburn, VA (IAD)	325	88.8/77.7	65.7	53.0	44.1	2.9	78.7
Atlanta, GA	1,027	88.2/77.2	67.2	73.7	22.0	4.3	70.8
Boston, MA	0	86.3/76.2	70.1	51.5	47.8	0.7	91.6
Chicago, IL	673	88.2/77.9	65.1	46.4	52.3	1.3	88.8
Dallas, TX	597	91.4/78.6	63.0	69.8	22.8	7.4	62.1
Denver, CO	5,285	81.8/64.6	100.0	51.3	48.7	0.0	100.0
Houston, TX	105	89.0/80.1	58.4	74.8	15.2	10.0	48.0
Los Angeles, CA	325	78.0/70.2	87.4	99.2	0.7	0.1	97.9
Miami, FL	0	86.8/80.2	58.1	84.1	0.3	15.6	24.5
Minneapolis, MN	837	87.5/76.9	68.1	46.4	52.3	1.3	90.3
Newark, NJ	0	88.8/77.7	65.7	54.6	43.7	1.7	84.9
Phoenix, AZ	1,106	96.4/76.1	70.3	83.1	14.5	2.4	80.7
Salt Lake City, UT	4,226	86.8/67.0	95.9	50.1	49.9	0.0	99.8
San Francisco, CA	0	78.2/65.4	100.0	70.7	29.3	0.0	100.0
Seattle, WA	433	82.2/66.5	97.5	58.6	41.4	0.0	99.8

System design parameters:

1 MW load, $n = 4$

Target supply air = 75°F, target return air = 96°F

N+1 redundancy, with redundant unit operating for annual analysis

MERV 13 filtration consolidated in only (2) units

Water sprays turned off below 50°F ambient db

1.0" ESP (supply + return)

IECX WBDE ≈ 75% and dry effectiveness ≈ 56%

Notes:

System (wet) rejects 100% of ITe load when ambient wet bulb temperature is below 67°F

System (dry) rejects 100% of ITe load when ambient wet bulb temperature is below 55°F

System does not introduce any outside air into data hall, all cooling effects are produced indirectly

*Percentage reduction in mechanical cooling equipment normally required at peak load based on N units operating

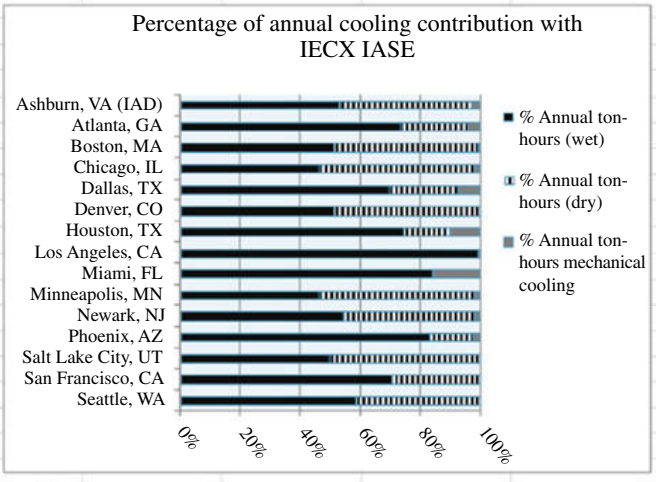


FIGURE 13.11 Analysis summary for modular DC cooling solution using IECX.

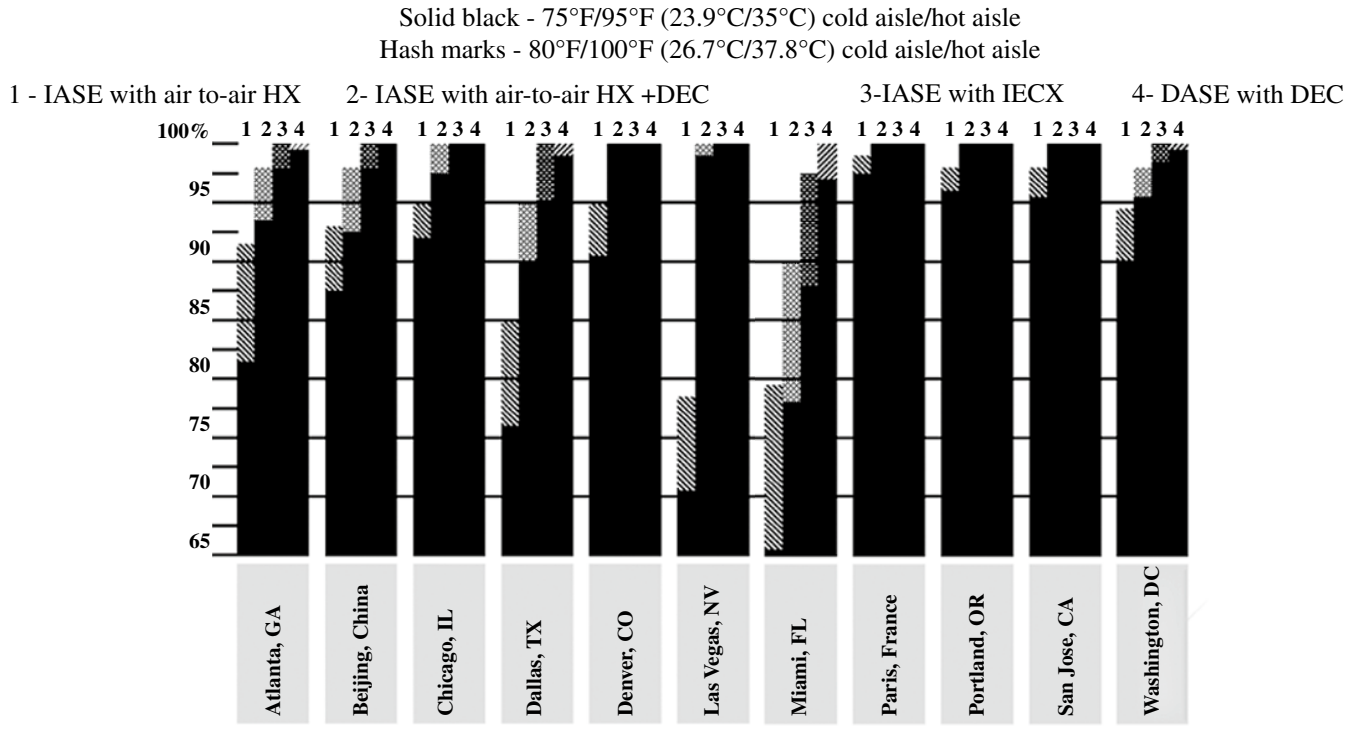
13.3.2.4 Other Indirect Economizer Types

There are other types of indirect economizers that can be considered based on the design of the DC facility and cooling systems. These are indirect water-side economizer (IWSE) and indirect refrigerant-side economizer (IRSE), and the primary difference between these and the IASEs discussed in this chapter is the working fluid from which the economizer is removing the heat energy. Where heat energy is being transported with water or glycol, an IWSE can be implemented. Similarly, in systems where a refrigerant is used to transport heat energy, an IRSE can be implemented. Either of these economizer types can be implemented with or without evaporative cooling, in much the same way an IASE can be, and similarly the overall efficiency of these economizer types depends on the efficiency of the heat exchange devices, efficiency of other system components, and facility location.

13.4 COMPARATIVE POTENTIAL ENERGY SAVINGS AND REQUIRED TRIM MECHANICAL REFRIGERATION

Numerous factors have an influence on the selection and design of a DC cooling system. Location, water availability, allowable cold aisle temperature, and extreme design conditions are four of the major factors. Figure 13.12 shows a comparison of the cooling concepts previously discussed as they relate to percentage of cooling load during the year that the economizer is capable of removing and the capacity of trim mechanical cooling that has to be on board to supplement the economizer on hot and/or humid days, the former representing full-year energy savings and the latter initial capital cost.

To aid in using Figure 13.12, take the following steps:



Trim DX using TMY maximum temperatures, tons

75/95°F(23.9/35°C) cold aisle/hot aisle temperature

	Atlanta, GA	Beijing, China	Chicago, IL	Dallas, TX	Denver, CO	Las Vegas, NV	Miami, FL	Paris, France	Portland, OR	San Jose, CA	Washington, DC
1	1.80	1.80	1.80	1.80	1.80	1.88	1.80	1.22	1.80	1.80	1.80
2	1.06	1.11	0.95	1.10	0.25	0.58	0.90	0.52	0.88	0.34	0.94
3	0.55	0.97	0.78	0.96	0.00	0.34	0.73	0.27	0.70	0.06	0.77
4	3.58°F	8.62°F	10.53°F	9.0°F	0°F	1.91°F	5.64°F	0°F	6.05°F	0°F	6.72°F

80/100°F(26.7/37°C) cold aisle/hot aisle temperature

	Atlanta, GA	Beijing, China	Chicago, IL	Dallas, TX	Denver, CO	Las Vegas, NV	Miami, FL	Paris, France	Portland, OR	San Jose, CA	Washington, DC
1	1.68	1.75	1.48	1.80	1.80	1.80	1.55	0.89	1.71	1.55	1.74
2	0.77	0.81	0.65	0.80	0.00	0.28	0.61	0.23	0.58	0.05	0.64
3	0.20	0.62	0.43	0.61	0.00	0.00	0.37	0.00	0.35	0.00	0.42
4	0°F	3.6°F	1.55°F	4.0°F	0°F	0°F	0.64°F	0°F	1.05°F	0°F	1.72°F

Trim DX using extreme 50 year maximum temperatures, tons

75/95°F(23.9/35°C) cold aisle/hot aisle temperature

	Atlanta, GA	Beijing, China	Chicago, IL	Dallas, TX	Denver, CO	Las Vegas, NV	Miami, FL	Paris, France	Portland, OR	San Jose, CA	Washington, DC
1	1.80	1.80	1.48	1.80	1.80	1.80	1.80	1.80	1.80	1.80	1.80
2	1.06	1.38	1.11	1.09	0.29	1.00	1.20	0.85	1.29	0.85	1.15
3	0.92	1.29	0.98	0.95	0.00	0.84	1.08	0.66	1.20	0.66	1.03
4	9.6°F	15.9°F	6.55°F	10.86°F	0°F	9.93°F	11.17°F	6.24°F	13.57°F	6.7°F	11.2°F

80/100°F(26.7/37.8°C) cold aisle/hot aisle temperature

	Atlanta, GA	Beijing, China	Chicago, IL	Dallas, TX	Denver, CO	Las Vegas, NV	Miami, FL	Paris, France	Portland, OR	San Jose, CA	Washington, DC
1	1.80	1.80	1.80	1.80	1.80	1.80	1.76	1.80	1.80	1.80	1.80
2	0.77	1.08	0.82	0.80	0.00	0.70	0.90	0.56	1.00	0.56	0.86
3	0.56	0.94	0.63	0.60	0.00	0.49	0.73	0.31	0.85	0.31	0.68
4	4.66°F	9.9°F	5.53°F	5.86°F	0°F	4.93°F	0.17°F	1.24°F	8.57°F	1.7°F	6.2°F

Tons of additional mechanical AC per 1000 SCFM of cooling air required to achieve desired delivery

Temperature when using air economizers - with no economizer the full AC load is 1.8 tons/1000 SCFM

FIGURE 13.12 Annualized economizer cooling capability based on TMY3 (Typical Meteorological Year) data

1. Select the city of interest and use that column to select the following parameters.
2. Select either TMY maximum or 50-year extreme section for the ambient cooling design.
3. Select the desired cold aisle/hot aisle temperature section within the section selected in step 2.
4. Compare the trim mechanical cooling required for each of the four cooling systems under the selected conditions.

Dallas, Texas, using an AHX, represented by the no. 1 at the top of the column, will be used as the first example. Operating at a cold aisle temperature of 75°F and a hot aisle of

95°F, represented by the solid black bars, 76% of the cooling ton-hours during the year will be supplied by the economizer. The other 24% will be supplied by a cooling coil. The size of the trim mechanical cooling system is shown in the lower part of the table as 1.8 tons/1000 standard cubic feet per minute (SCFM) of cooling air, which is also the specified maximum cooling load that is required to dissipate the IT heat load. Therefore, for the AHX in Dallas, the amount of trim cooling required is the same tonnage as would be required when no economizer is used. That is because the TMY3 design db temperature is 104°F, well above the RA temperature of 95°F. Even when the cold aisle/hot aisle setpoints are raised to 80°F/100°F, the full capacity of mechanical cooling is required. If a DEC (represented by no. 2 at top of column) is placed in the ScA (TMY3 maximum wb temperature is 83°F), then 90% of the yearly cooling is supplied by the economizer and the trim cooling drops to 1.1 tons/1000 scfm from 1.8 tons.

For the second example, we will examine Washington, D.C., where the engineer has determined that the design ambient conditions will be based on TMY3 data. Using 75°F/95°F cold aisle/hot aisle conditions, the IECX and DASE with DEC, columns no. 3 and no. 4, can perform 98 and 99% of the yearly cooling, respectively, leaving only 2 and 1% of the energy to be supplied by the mechanical trim cooling. The AHX (no. 1) accomplishes 90% of the yearly cooling, and if a DEC (no. 2) is added to the scavenger airstream, the combination does 96% of the cooling. The trim cooling for heat exchangers 1, 2, and 3, respectively, is 1.8, 0.94, and 0.77 tons where 1.8 is full load tonnage. Increasing the cold aisle/hot aisle to 80°F/110°F allows no. 3 and no. 4 to supply all of the cooling with the economizers, and reduces the amount of onboard trim cooling for 1 and 2.

It should be apparent from Figure 13.11 that even in hot and humid climates such as Miami, Florida, economizers can provide a significant benefit for DCs. As ASHRAE standard 90.4 is adopted, selecting the right economizer cooling system should allow a design to meet or exceed the required mechanical efficiency levels. In addition, the economizers presented in this section will become even more desirable for energy savings as engineers and owners become more familiar with the recently introduced allowable operating environments A1 through A4 as shown on the psychrometric charts of Figures 13.2 and 13.4. In fact, if the conditions of A1 and A2 were allowed for a small portion of the total operating hours per year, then for no. 2 and no. 3 all of the cooling could be accomplished with the economizers, and there would be no requirement for trim cooling when using TMY3 extremes. For no. 4, the cooling could also be fully done with the economizer, but the humidity would exceed the envelope during hot, humid periods.

There are instances when the cooling system is being selected and designed for a very critical application where the system has to hold space temperature under the worst possible ambient cooling condition. In these cases the ASHRAE 50-year extreme annual design conditions are used as referred in Chapter 14 of Ref. [3] and designated as "complete data tables" and underlined in blue in the first paragraph. These data can only be accessed by means of the disk that accompanies the ASHRAE Handbook. The extreme conditions are shown in Table 13.2, which also includes for comparison the maximum conditions from TMY3 data.

Using the 50-year extreme temperatures of Table 13.2, the amount of trim cooling, which translates to additional initial capital cost, is shown in the lower portion of Figure 13.12. All values of cooling tons are per 1000 scfm

TABLE 13.2 Design temperatures that aid in determining the amount of trim cooling

	50-year extreme				Maximum from TMY3 data			
	db		wb		db		wb	
	°F	°C	°F	°C	°F	°C	°F	°C
Atlanta	105.0	40.6	82.4	28.0	98.1	36.7	77.2	25.1
Beijing	108.8	42.7	87.8	31.0	99.3	37.4	83.2	28.4
Chicago	105.6	40.9	33.3	28.5	95.0	35.0	80.5	26.9
Dallas	112.5	44.7	82.9	28.3	104.0	40.0	83.0	28.3
Denver	104.8	40.4	69.3	20.7	104.0	40.0	68.6	20.3
Las Vegas	117.6	47.6	81.3	27.4	111.9	44.4	74.2	23.4
Miami	99.4	37.4	84.7	29.3	96.1	35.6	79.7	26.5
Paris	103.2	39.6	78.8	26.0	86.0	30.0	73.2	22.9
Portland	108.1	42.3	86.4	30.2	98.6	37.0	79.3	26.3
San Jose	107.8	42.1	78.8	26.0	96.1	35.6	70.2	21.2
Washington, D.C.	106.0	41.1	84.0	28.9	99.0	37.2	80.3	26.8

Source: ASHRAE Fundamentals 2013 and NREL

(1699 m^3/h) with a ΔT of 20°F (11.1°C). For the DASE with DEC designated as number 4, instead of showing tons, temperature rise above desired cold aisle temperature is given.

From a cost standpoint, just what does it mean when the economizer reduces or eliminates the need for mechanical cooling? This can best be illustrated by comparing the mechanical partial PUE (pPUE) of an economizer system to that of a modern conventional mechanical cooling system. Mechanical pPUE in this case is a ratio of (IT cooling load + power consumed in cooling IT load)/(IT load). The mechanical pPUE value of economizers ranges from 1.07 to about 1.3. For refrigeration systems the value ranges from 1.8 to 2.5. Taking the average of the economizer performance as being 1.13 and using the lower value of a refrigeration (better performance) system of 1.8, the economizer uses only 1/6 of the operating energy to cool the DC when all cooling is performed by the economizer.

As an example of cost savings, if a DC operated at an IT load of 5 MW for a full year and the electrical utility rate was $0.10/kW-h, then the power cost to operate the IT equipment would be $4,383,000/year. To cool with mechanical refrigeration equipment with a pPUE of 1.80, the cooling cost would be $3,506,400/year for a total electrical cost of $7,889,000. If the economizer handled the entire cooling load, the cooling cost would be reduced to $570,000/year. If the economizer could only do 95% of the full cooling load for the year, then the cooling cost would still be reduced from $3,506,400 to $717,000—a reduction worth investigating.

13.5 CONVENTIONAL MEANS FOR COOLING DATACOM FACILITIES

In this chapter we have discussed techniques for cooling that first and foremost consider economization as the principal form of cooling. There are more than 20 ways to cool a DC using mechanical refrigeration with or without some form of economizer as part of the cooling strategy. References [4] and [5] are two articles that cover these various mechanical cooling techniques. Also, Chapter 20, *Data Centers and Telecommunication Facilities*, of Ref. [6] discusses standard techniques for DC cooling.

13.6 A NOTE ON LEGIONNAIRES' DISEASE

IEC is considered to share the same operating and maintenance characteristics as conventional DEC, except that the evaporated water is not added to the process air. As a result, ASHRAE has included IEC in chapter 53, *Evaporative Cooling*, of Ref. [5]. Below is an excerpt from the handbook:

> Legionnaires' Disease. There have been no known cases of Legionnaires' disease with air washers or wetted-media evaporative air coolers. This can be attributed to the low temperature of the recirculated water, which is not conducive to Legionella bacteria growth, as well as the absence of aerosolized water carryover that could transmit the bacteria to a host. (ASHRAE Guideline 12-2000 [7])

IECs operate in a manner closely resembling DECs and not resembling cooling towers. A typical cooling tower process receives heated water at 95–100°F, sprays the water into the top of the tower fill material at the return temperature, and is evaporatively cooled to about 85°F with an ambient wb of 75°F before it flows down into the sump and is then pumped back to the process to complete the cycle. The ScA leaving the top of a cooling tower could be carrying with it water droplets at a temperature of over 100°F. On the other hand, an IEC unit sprays the coolest water within the system on top of the IECX surface, and then the cool water flows down over the tubes. It is the cooled water that totally covers the tubes that is the driving force for cooling the process air flowing within the tubes. The cooling water then drops to the sump and is pumped back up to the spray nozzles, so the water temperature leaving at the bottom of the HX is the same temperature as the water being sprayed into the top of the IECX. On hot days, at any point within the IECX, the water temperature on the tubes is lower than the temperature of either the process airstream or the wetted scavenger airstream. From an ETL, independently tested IECX, similar to the units being used in DCs and operating at DC temperatures, high ambient temperature test data show that the sump water temperature, and therefore the spray water temperature, is 78°F when the return from the hot aisle is 101.2°F and the ambient ScA is 108.3/76.1°F. Or the spray water temperature is 1.9°F above ambient wb temperature.

In addition to having the sump temperature within a few degrees of the wb temperature on hot days, thus behaving like a DEC, there is essentially, with proper design, no chance that water droplets will leave the top of the IEC unit. This is because there is a moisture eliminator over the IECX and then there is a warm condenser coil over the eliminator (on the hottest days the trim DX will be operating and releasing its heat into the scavenger airstream, which, in the unlikely event that a water droplet escapes through the eliminator, that droplet would evaporate to a gas as the air heats through the condenser coil).

So, IEC systems inherently have two of the ingredients that prevent *Legionella*: cool sump and spray temperatures and only water vapor leaving the unit. The third is to do a good housekeeping job and maintain the sump area so that it is clean and fresh. This is accomplished with a combination of sump water bleed-off, scheduled sump dumps, routine inspection and cleaning, and biocide treatment if necessary. With good sump maintenance, all three criteria to prevent *Legionella* are present.

REFERENCES

[1] ASHRAE. *Thermal Guidelines for Data Processing Environments*. 4th ed. Atlanta: ASHRAE; 2015.

[2] ASHRAE. *ASHRAE Handbook-Systems and Equipment*. Atlanta: American Society of Heating Refrigeration and Air Conditioning Engineers, Inc.; 2020.

[3] ASHRAE. *ASHRAE Handbook-Fundamentals*. Atlanta: American Society of Heating Refrigeration and Air Conditioning Engineers, Inc.; 2017.

[4] Evans T. The different technologies for cooling data centers, Revision 2. Available at http://www.apcmedia.com/salestools/VAVR-5UDTU5/VAVR-5UDTU5_R2_EN.pdf. Accessed on May 15, 2020.

[5] Kennedy D. Understanding data center cooling energy usage and reduction methods. Rittal White Paper 507; February 2009.

[6] ASHRAE. *ASHRAE Handbook-Applications*. Atlanta: American Society of Heating Refrigeration and Air Conditioning Engineers, Inc.; 2019.

[7] ASHRAE Standard. *Minimizing the risk of legionellosis associated with building water systems*, ASHRAE Guideline12-2000, ISSN 1041-2336. Atlanta, Georgia: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.

FURTHER READING

Atwood D, Miner J. *Reducing Data Center Cost with an Air Economizer*. Hillsboro: Intel; 2008.

Dunnavant K. *Data center heat rejection*. ASHRAE J 2011;53(3):44–54.

Quirk D, Sorell V. *Economizers in Datacom: risk mission vs. reward environment?* ASHRAE Trans 2010;116(2):9, para.2.

Scofield M, Weaver T. *Using wet-bulb economizers, data center cooling*. ASHRAE J 2008;50(8):52–54, 56–58.

Scofield M, Weaver T, Dunnavant K, Fisher M. Reduce data center cooling cost by 75%. Eng Syst 2009;26(4):34–41.

Yury YL. *Waterside and airside economizers, design considerations for data center facilities*. ASHRAE Trans 2010;116(1):98–108.

14

RACK-LEVEL COOLING AND SERVER-LEVEL COOLING

Dongmei Huang[1], Chao Yang[2] and Bang Li[3]

[1] *Beijing Rainspur Technology, Beijing, China*
[2] *Chongqing University, Chongqing, China*
[3] *Eco Atlas (Shenzhen) Co., Ltd, Shenzhen, China*

14.1 INTRODUCTION

This chapter provides a brief introduction to rack-level cooling and server-level cooling as applied to information technology (IT) equipment support. At rack-level cooling, the cooling unit is closer to heat source (IT equipment). And at server-level cooling, the coolant is close to heat source. This chapter will introduce various cooling types from remote to close to heat source. In each type, the principle, pros, and cons are discussed. There are a lot of types for server-level cooling; only liquid cooling is described on high-density servers. When using liquid cooling, it will reduce the data center energy costs. Liquid cooling is gaining its marketing share in the data center industry.

14.1.1 Fundamentals

A data center is typically a dedicated building used to house computer systems and associated equipment such as electronic data storage arrays and telecommunications hardware. The IT equipment generates a large amount of heat when it works. All the heat released inside the data center must be removed and released to the outside environment, often in the form of water evaporation (using a cooling tower).

The total cooling system equipment and processes are split into two groups (Fig. 14.1):

1. Located inside the data center room – Room cooling
2. Located outside the data center room – Cooling infrastructure

Rack-level cooling is primarily focused on equipment and processes associated with room cooling. Existing room cooling equipment is typically comprised of equipment such as computer room air handlers (CRAHs) or computer room air conditioners (CRACs). These devices most commonly pull warm air from the ceiling area, cool it using a heat exchanger, and force it out to an underfloor plenum using fans. The method is often referred to as raised floor cooling as shown in Figure 14.1.

The heat from the IT equipment, power distribution, and cooling systems inside the data center (including the energy required by the CRAH/CRAC units) must be transferred to the cooling infrastructure via the CRAH/CRAC units. This transfer typically takes place using a cooling water loop. The cooling infrastructure, commonly using a water-cooled chiller and cooling tower, receives the heated water from the room cooling systems and transfers the heat to the environment.

14.1.2 Data Center Cooling

14.1.2.1 Introduction

Rack-level cooling is applied to the heat energy transport inside a data center. Therefore a brief overview of a typical existing cooling equipment is provided so we can understand how rack-level cooling fits into an overall cooling system.

It is important to note that rack-level cooling depends on having a cooling water loop (mechanically chilled water or cooling tower water). Facilities without cooling water systems are unlikely to be good candidates for rack-level cooling.

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

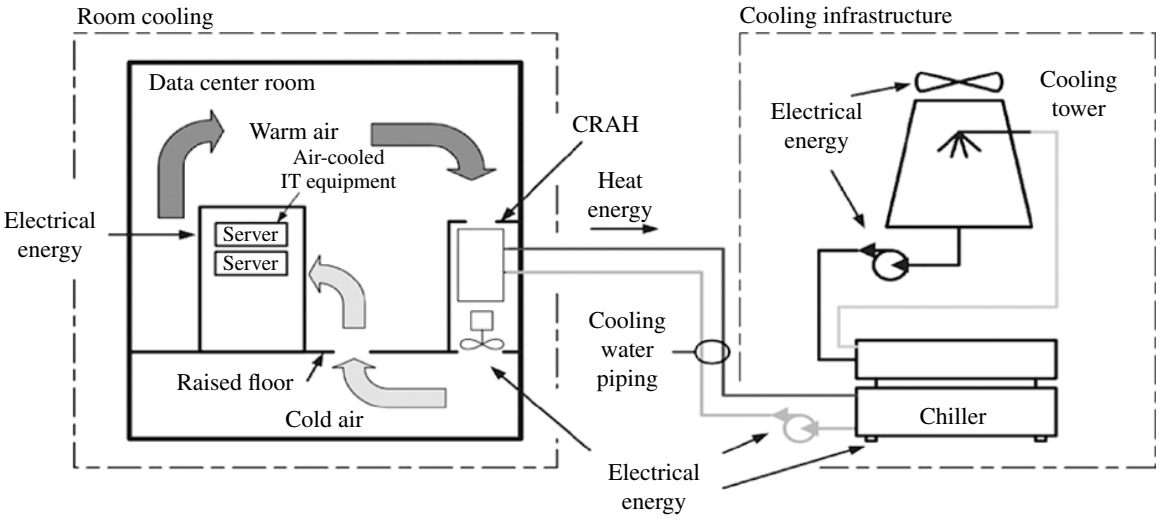


FIGURE 14.1 Data center cooling overview.

14.1.2.2 Transferring Heat

Most of the heat generated inside a data center originates from the IT equipment. As shown in Figure 14.2, electronic components are kept from overheating by a constant stream of air provided from internal fans.

Commercial IT equipment is typically mounted in what are termed "standard racks." A standard IT equipment rack has the approximate overall dimensions of 24 in wide by 80 in tall and 40 in deep. These racks containing IT equipment are placed in rows with inlets on one side and exits on the other. This arrangement creates what is termed "hot aisles" and "cold aisles."

14.1.2.3 Room-Level Cooling

Before we talk about rack- and server-level cooling, let's introduce conventional room-level cooling. The task of moving the air from the exit of the servers, cooling it, and providing it back to the inlet of the servers is commonly provided inside existing data center rooms by CRAHs arranged as shown in Figure 14.1. Note that the room may be cooled by CRACs that are water cooled or use remote air-cooled condensers; they are not shown in Figure 14.1.

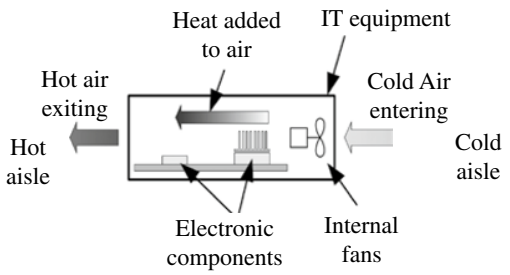


FIGURE 14.2 IT equipment cooling basics.

This air-cooling method worked in the past but can pose a number of issues when high-density IT equipment is added to or replaces existing equipment.

To understand how rack-level cooling equipment fits into room cooling, a brief listing of the pros and cons of conventional raised floor room cooling by CRAHs or CRACs is provided:

Pros: Cooling can be easily adjusted, within limits, by moving or changing the arrangement of perforated floor tiles.

Cons: Providing a significant increase in cooling at a desired location may not be practical due to airflow restrictions below the raised floor.

Raised floor cooling systems do not supply a uniform temperature of air presented at the IT equipment inlets across the vertical rack array due to room-level air circulation. Therefore the temperature of the air under the floor must be colder than it might otherwise be causing the external cooling infrastructure to work harder and use more energy.

If the existing room cooling systems cannot be adjusted or modified, the additional load must be met via another method, such as with a rack-level cooling solution. In the next section three common rack-level cooling solutions will be discussed.

14.2 RACK-LEVEL COOLING

In the last several years, a number of technologies have been introduced addressing the challenges of cooling high-density IT equipment. Before we look at a few common rack-level cooler types, three key functional requirements are discussed:

- Consistent temperature of cooling air at the IT equipment inlet:

  The solution should provide for a consistent temperature environment, including air temperature in the specified range and a lack of rapid changes in temperature. See the ASHRAE Thermal Guidelines (ASHRAE 2011) for these limits.
- Near neutral or slightly higher delta air pressure across the IT equipment:

  IT equipment needs adequate airflow via neutral or a positive delta air pressure to reduce the chance of issues caused by internal and external recirculation, including components operating above maximum temperature limits.
- Minimal load addition to the existing room air conditioning:

  Ideally a rack-level cooling solution should capture all the heat from the IT equipment racks it is targeted to support. This will reduce the heat load on the existing room cooling equipment.

There are a few distinct types of rack-level cooling device designs that have been installed in many data centers and proven over a number of years. The description of these designs along with pros and cons are discussed below. They are:

- Overhead
- In-Row™
- Enclosed
- Rear door
- Micro module

It should be noted that given the wide variety of situations where these devices might be considered or installed and with newer rack-level cooling devices frequently entering the market, there may be exceptions to the advantages or disadvantages listed.

14.2.1 Overhead Type

14.2.1.1 What Is Overhead Cooling and Its Principle

The term overhead, a type of Liebert XDO cooling modules, is located over the head of racks, taking in hot air through two opposite return vents and exhausting cool air from supply vent, going down into the cold aisle. The servers suck the cool air and exhaust hot air to overhead's return vent. Generally, an overhead cooling module uses for one cold aisle and consists of two cooling coils, two flow control valves, one filter dryer, and one fan. If there is only one row, the overhead cooling module should reduce by half. This cooling approach is closed to the cold aisle and supplies cool air directly into the server and has little resistance. The configuration is shown in Figure 14.3.

14.2.1.2 Advantages

It doesn't require floor space required and reduces cost. It is easy for installation. So the simple design contributes to the low maintenance and high reliability.

It is flexible to configure different cooling capability with standard modules. It can cool more than 5,400 W/m^2. It is excellent for spot and zone cooling, so it is highly energy efficient.

14.2.1.3 Disadvantages

When using multi-cooling modules in a row, there will be spacing between them and large distances between the top of the racks and the cooling module. Then it will result in hot air recirculation between the hot aisles and cold aisles. Therefore, blocking panels may also be required. If there aren't blocking panels between the overhead cooling

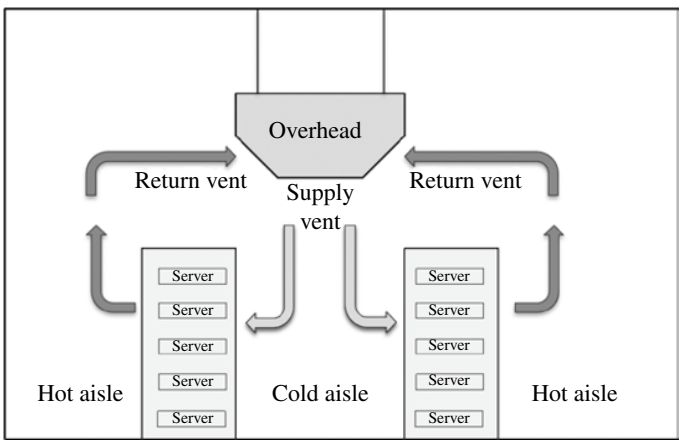


FIGURE 14.3 Overhead rack-cooler installation.

modules, you should use Computational Fluid Dynamics (CFD) to predict the thermal environment and make sure no hot spots.

14.2.2 In-Row™ Type

14.2.2.1 What Is In-Row Cooling and Its Principle

The term In-Row™, a trademark of Schneider Electric, is commonly used to refer to a type of rack cooling solution. This rack-level cooling design approach is similar to the enclosed concept, but the cooling is typically provided to a larger number of racks; one such configuration is shown in Figure 14.4. These devices are typically larger in size, compared with those offering the enclosed approach, providing considerably more cooling and airflow rate capacities. There are a number of manufacturers of this type of rack-level cooler, including APC by Schneider Electric and Emerson Network Power (Liebert brand).

14.2.2.2 Advantages

A wide variety of rack manufacturer models can be accommodated because the In-Row™ cooler does require an exacting mechanical connection to a particular model of rack. This approach works best with an air management containment system that reduces mixing between the hot aisle and cold aisles. Either a hot aisle or cold aisle containment method can be used. Figure 14.4 shows a plan view of a hot–cold aisle containment installation. Because In-Row™ coolers are often a full rack width (24 in), the cooling capacity can be substantial, thereby reducing the number of In-Row™ coolers needed. Half-rack-width models with less cooling capacity are also available.

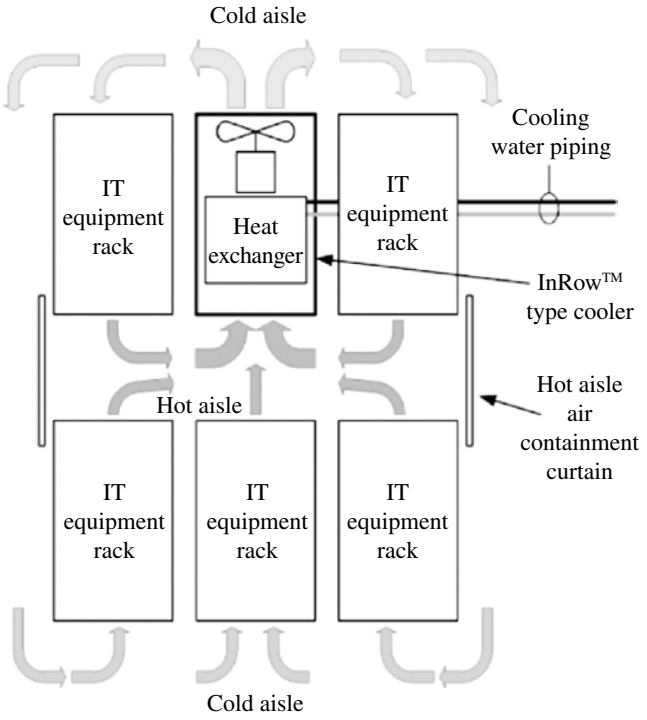


FIGURE 14.4 Plan view: In-Row™ rack-cooler installation.

14.2.2.3 Disadvantages

The advantage of the ability to cool a large number of racks of different manufacturers containing a wide variety of IT equipment also leads to a potential disadvantage. There is an increased likelihood that the temperature and air supply to the IT equipment is not as tightly controlled compared with the enclosed approach.

14.2.3 Enclosed Type

14.2.3.1 What Is Enclosed Type Cooling and Its Principle

The enclosed design approach is somewhat unique compared with the other two in that the required cooling is provided while having little or no heat exchange with the surrounding area. Additional cooling requirements on the CRAH or CRAC units can be avoided when adding IT equipment using this rack cooler type. The enclosed type consists of a rack of IT equipment and a cooling unit directly attached and well sealed. The cooling unit has an air-to-water heat exchanger and fans. All the heat transfer takes place inside the enclosure as shown in Figure 14.5. The heat captured by the enclosed rack-level device is then transferred directly to the cooling infrastructure outside the data center room. Typically one or two racks of IT equipment are supported, but larger enclosed coolers are available supporting six or more racks. There are a number of manufacturers of this type of rack-level cooler including Hewlett–Packard, Rittal, and APC by Schneider Electric.

Enclosed rack-level coolers require a supply of cooling water typically routed through the underfloor space. Overhead water supply is also an option. For some data centers installing a cooling distribution unit (CDU) may be recommended depending on the water quality, leak mitigation strategy, temperature control, and condensation management considerations. A CDU provides a means of separating water cooling loops using a liquid-to-liquid heat exchanger and a pump. CDUs can be sized to provide for any number of enclosed rack-level cooled IT racks.

14.2.3.2 Advantages

The main advantage of the enclosed solution is the ability to place high-density IT equipment in almost any location inside an existing data center that has marginal room cooling capacity.

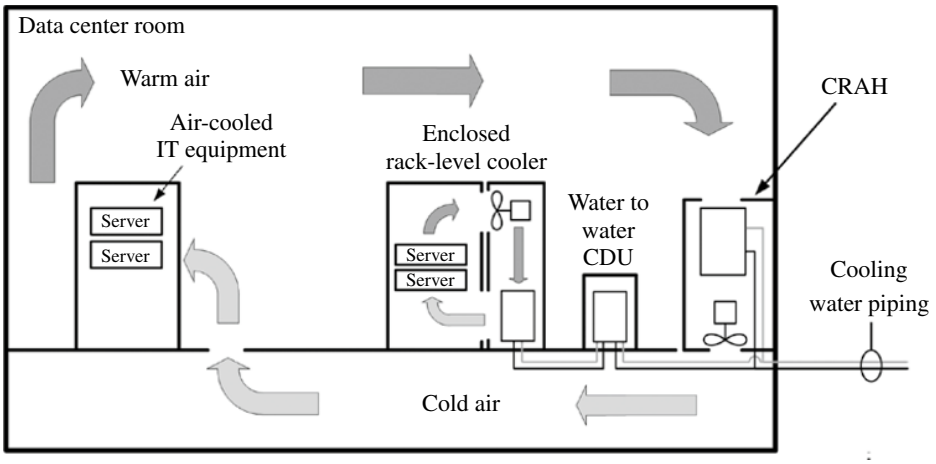


FIGURE 14.5 Elevation view: enclosed rack-level cooler installation.

A proper enclosed design also provides a closely coupled, well-controlled uniform temperature and pressure supply of cooling air to the IT equipment in the rack. Because of this feature, there is an improved chance that adequate cooling can be provided with warmer water produced using a cooling tower. In these cases the use of the chiller may be reduced resulting in significant energy savings.

14.2.3.3 Disadvantages

Enclosed rack coolers typically use row space that would normally be used for racks containing IT equipment, thereby reducing the overall space available for IT. If not carefully designed, low-pressure areas may be generated near the IT inlets.

Because there is typically no redundant cooling water supply, a cooling water failure will cause the IT equipment to overheat within a minute or less. To address this risk, some models are equipped with an automated enclosure opening system, activated during a cooling fluid system failure. However, CFD is a good tool to predict the hot spot or cooling failure scenarios. It may increase the reliability of the enclosed rack. See Figure 14.6, which is a typical enclosed rack called one-cooler-one-rack or one-cooler-two-rack. It can also be used in the In-Row type and Rear door type with more racks etc.

14.2.4 Rear Door

14.2.4.1 What Is Rear Door Cooling and Its Principle

Rear door IT equipment cooling was popularized in the mid-2000s when Vette, using technology licensed from

FIGURE 14.6 Front view: enclosed type rack level (hot airflow to the back of rack).

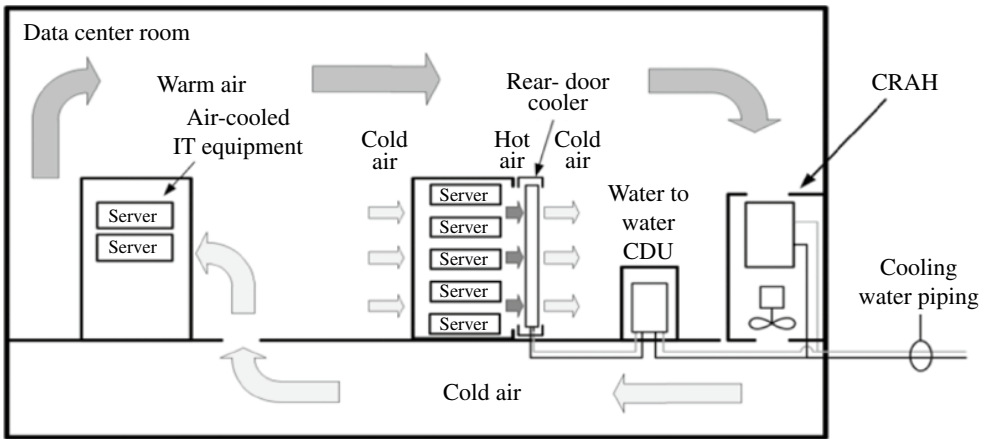


FIGURE 14.7 Elevation view: rear door cooling installation.

IBM, brought the passive rear door to the market in quantity. Since that time, passive rear door cooling has been used extensively on the IBM iDataPlex platform. Vette (now Coolcentric) passive rear doors have been operating for years at many locations.

Rear door cooling works by placing a large air-to-water heat exchanger directly at the back of each rack of IT equipment replacing the original rack rear door.

The hot air exiting the rear of the IT equipment is immediately forced to enter this heat exchanger without being mixed with other air and is cooled to the desired exit temperature as it re-enters the room as shown in Figure 14.7.

There are two types of rear door coolers, passive and active. Passive coolers contain no fans to assist with pushing the hot air through the air-to-water heat exchanger. Instead they rely on the fans, shown in Figure 14.2, contained inside the IT equipment to supply the airflow. If the added pressure of a passive rear door is a concern, "active" rear door coolers are available containing fans that supply the needed pressure and flow through an air-to-water heat exchanger.

14.2.4.2 Advantages

Rear door coolers offer a simple and effective method to reduce or eliminate IT equipment heat from reaching the existing data center room air-conditioning units. In some situations, depending on the cooling water supply, rear door coolers can remove more heat than that supplied by the IT equipment in the attached rack. Passive rear doors are typically very simple devices with relatively few failure modes. In the case of passive rear doors, they are typically installed without controls. For both passive and active rear doors, the risk of IT equipment damage by condensation droplets formed on the heat exchanger and then released into the airstream is low. Potential damage by water droplets entering the IT equipment is reduced or eliminated because these droplets would only be found in the airflow downstream of the IT equipment. Rear door coolers use less floor area than most other solutions.

14.2.4.3 Disadvantages

Airflow restriction near the exit of the IT equipment is the primary concern with rear door coolers both active (with fans) and passive (no fans). The passive models restrict the IT equipment airflow but possibly not more than the original rear door. While this concern is based on sound fluid dynamic principles, a literature review found nothing other than manufacturer reported data (reference Coolcentric FAQs) of very small or negligible effects that are consistent with users' anecdotal experience. For customers that have concerns regarding airflow restriction, active models containing fans are available.

14.2.5 Micro Module

14.2.5.1 What Is Micro Module Cooling and Its Principle

From the energy efficiency point of view, rack cooling has much higher energy utility percentage than room-level system. The cooler is much closer to the IT equipment (heat source). Here we introduce a free cooling type. The system draws outside air into the modular data center (MDC), a number of racks are cooled by free cooling air. Depends on the location of the MDC, the system includes primary filter that filters the bigger size of dust, medium efficiency filter that filters smaller size of dust, and high efficiency filter that may filter chemical pollution. The system will include fan walls with a matrix of fans depending on how many racks are being cooled. Figure 14.8 is a type of free cooling racks designed by Suzhou A-Rack Information Technology Company in China. Figure 14.9 is its CFD simulation

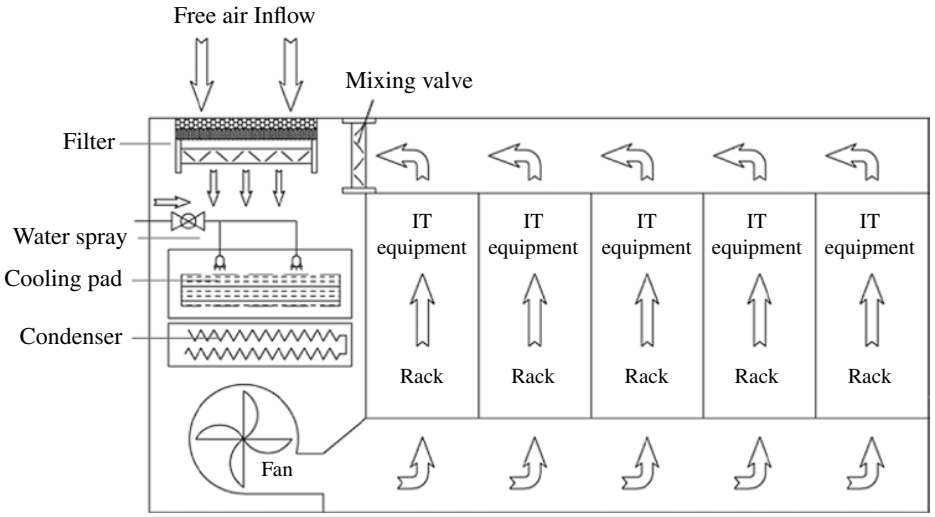


FIGURE 14.8 Overhead view: free cooling rack installation. Source: Courtesy of A-RACK Tech Ltd.

airflow, from www.rainspur.com. The cooling air from fan is about 26°C, enters into the IT equipment and the hot air about 41°C, and is exhausted from the top of the container.

14.2.5.2 Advantage

The advantage of free cooling rack level is its significant energy saving. The cool air is highly utilized by the system, which can produce very low PUE.

14.2.5.3 Disadvantage

The disadvantage of the free cooling type is that it may depend on the location environment. It requires good quality of free air and suitable humidity. If the environment is polluted, the change of the filter cost will be much higher. If the air humidity is high, the free cooling efficiency will be limited.

14.2.6 Other Cooling Methods

In addition to the conventional air-based rack-level cooling solutions discussed above, there are other rack-level cooling solutions for high-density IT equipment.

A cooling method commonly termed direct cooling was introduced for commercial IT equipment. The concept of direct cooling is not new. It has been widely available for decades on large computer systems such as supercomputers used for scientific research. Direct cooling brings liquid, typically water, to the electronic component replacing relatively inefficient cooling using air.

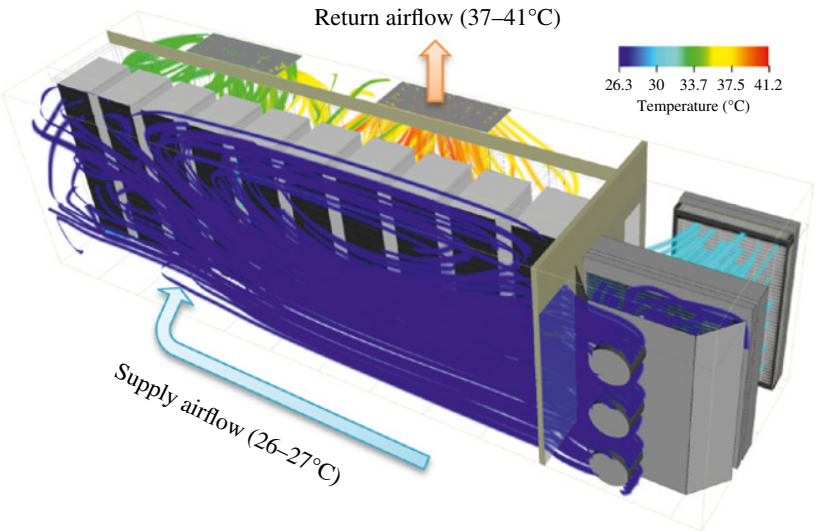


FIGURE 14.9 CFD simulation of free cooling rack. Source: Courtesy of Rainspur Technology Co., Ltd.

14.3 SERVER-LEVEL COOLING

Server-level cooling is generally studied by IT equipment suppliers. Air cooling is still a traditional and mature technology. However, for high-density servers, when installed in a rack, the total will be over 100 kW, in which air cooling cannot meet the server's environment requirement. So liquid cooling is becoming the cutting-edge technology for high-density servers. Two common ways of liquid cooling is discussed in this section, cold plate and immersion cooling.

14.3.1 Cold Plate and Its Principle

14.3.1.1 What Is a Cold Plate and Its Principle

Cold plate is a cooling method by conduction. Liquid flows inside the plate and dissipates the heat of a heat source. The liquid can be water or oil. These solutions cool high-heat-producing temperature-sensitive components inside the IT equipment using small water-cooled cold plates or structures mounted near or contacting each direct cooled component. Some solutions include miniature pumps integrated with the cool plates providing pump redundancy. Figure 14.10 illustrates cold plate cooling in a schematic view, and Figure 14.11 shows a cold plate server, design by Asetek, with water pipes and manifolds in server rack.

14.3.1.2 Advantages

High efficiency; the heat from electronic components is transferred by conduction to a cold plate that covers the server. Clustered systems offer a unique rack-level cooling solution; transferring heat directly to the facility cooling loop gives direct cooling, which is an overall efficiency advantage. The heat captured by direct cooling allows the less efficient room air-conditioning systems to be turned down or off.

14.3.1.3 Disadvantages

Most of these systems are advertised as having the ability to be cooled with hot water, and they do remove heat quite efficiently. The block in contact with the CPU or other hot body is usually copper with a conductivity of around 400 W/m*K so the temperature drop across it is negligible. If the water is pumped slowly enough, reducing pumping power, flow is laminar. Because water is not a very good conductor of heat, a temperature drop of around 5°C can be expected across the water copper interface. This is usually negligible but if necessary can be reduced by forcing turbulent flow by increasing flow rate. This could be an expensive waste of energy.

14.3.2 Immersion Cooling

14.3.2.1 What Is Immersion and Its Principle

Immersion liquid cooling technology mainly uses specific coolant as the heat dissipation medium to immerse IT equipment directly in the coolant and remove heat during the operation of IT equipment through the coolant circulation. At the same time, the coolant circulates through the process of heat exchange with external cold sources releasing heat into the environment. The commonly used coolant mainly includes water, mineral oil, and fluorinated liquid.

Water, mainly pure deionized water, is widely used in refrigeration systems as an easily available resource. However, since water is not an insulator, IT can only be used in indirect liquid cooling technology. Once leakage occurs, it will cause fatal damage to IT equipment.

Mineral oil, a single-phase oil, is a relatively low price insulation coolant. It is tasteless, nontoxic, not volatile, and an environmentally friendly oil. However due to its high viscosity property, it is difficult to maintain.

Fluorinated liquid, the original function of which is circuit board cleaning liquid, is applied in data center liquid cooling technology due to its insulating and noncombustible inert characteristics, which is not only the immersion coolant widely used at present but also the most expensive of the three types of coolant.

For immersion liquid cooling, the server is placed vertically in a customized cabinet and the server is completely immersed in the coolant. The coolant is driven by the circulating pump and enters the specific exchanger to exchange heat with the cooling water and then returns to the cabinet. The cooling water is also driven by the circulating pump into a specific exchanger to exchange heat with the cooling fluid and finally discharge heat to the environment through the cooling tower. Immersion liquid cooling, due to the direct contact between heat source and coolant, has higher heat dissipation

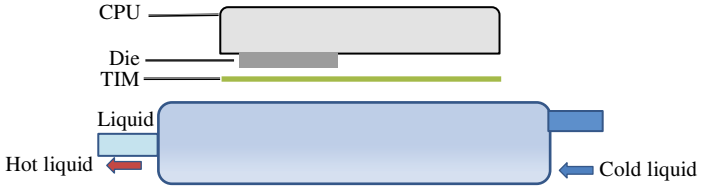


FIGURE 14.10 Cold plate cooling using thermal interface material (TIM).

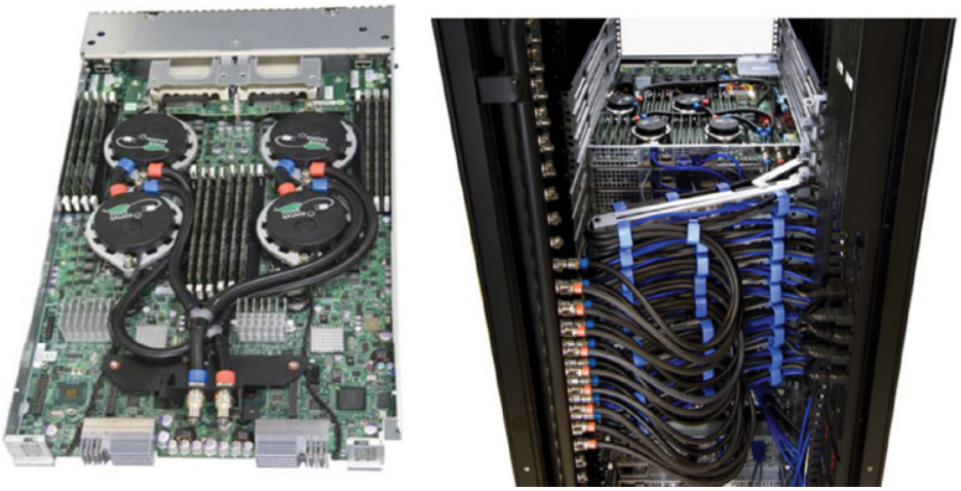


FIGURE 14.11 Cold plate server with water pipes and manifolds in rack. Source: Courtesy of Asetek.

efficiency. Compared with cold plate liquid cooling, it has lower noise (no fan at all), adapts to higher thermal density, and energy saving.

The operation of immersion liquid cooling equipment is shown in Figure 14.12.

In the application of immersion liquid cooling technology in data center, high-energy-consumption equipment such as CRAC, chiller, humidity control equipment, and air filtration equipment is not needed, and the architecture of the room is simpler. PUE value can be easily reduced to less than 1.2, the minimum test result can reach about 1.05, and CLF value (power consumption of refrigeration equipment/power consumption of IT equipment) can be as low as 0.05–0.1.

The main reasons are as follows: compared with air, the cooling liquid phase has a thermal conductivity 6 times that of air, and the heat capacity per unit volume is 1,000 times that of air. That is to say, for the same volume of heat transfer medium, the coolant transfer heat at six times the speed of air, heat storage capacity is 1,000 times the amount of air. In addition, compared with the traditional cooling mode, the cooling liquid has fewer heat transfer times, smaller capacity attenuation, and high cooling efficiency. This means that under the same heat load, the liquid medium can achieve heat dissipation with less flow rate and smaller temperature difference. The smaller medium flow can reduce the energy consumption needed to drive the cooling medium in the process of heat dissipation. The thermodynamic properties of air, water, and coolant are compared in Table 14.1.

TABLE 14.1 Thermodynamic properties comparison of air, water, and liquid coolant

Medium	Conductivity W/(m*K)	Specific thermal capacity kJ/(kg*K)	Volume thermal capacity kJ/(m^3*K)
Air	0.024	1	1.17
Water	0.58	4.18	4,180
Coolant	0.15	1.7	1,632

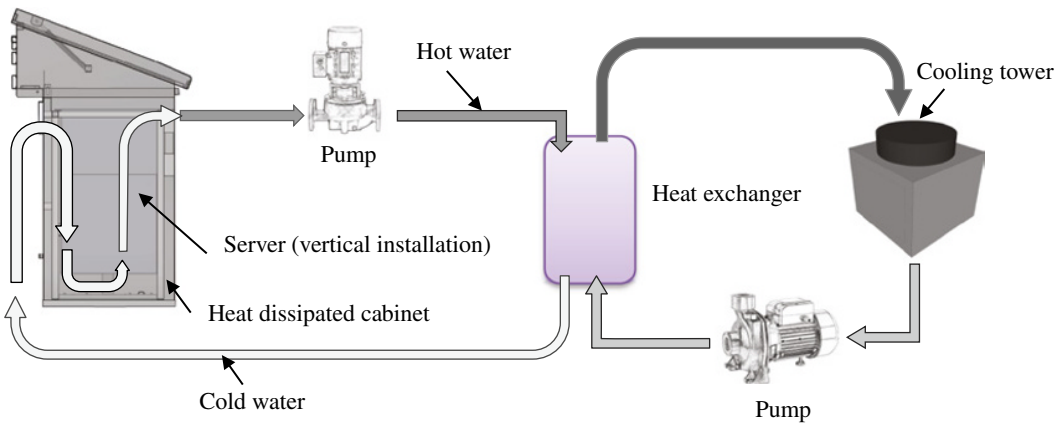


FIGURE 14.12 Immersion liquid cooling equipment operation chart. Source: Courtesy of Rainspur Technology Co., Ltd.

TABLE 14.2 Heat dissipation performance comparison between air and liquid coolant

Medium	Air	Liquid coolant
CPU power (W)	120	120
Inlet temperature (°C)	22	35
Outlet temperature rise (°C)	17	5
Volume rate (m^3/h)	21.76	0.053
CPU heat sink temperature (°C)	46	47
CPU temperature (°C)	77	75

Table 14.2 shows the comparison of CPU heat dissipation performance data in air-cooled and liquid-cooled environments. Under the same amount of heat load, liquid media can have less flow rate and smaller temperature difference to achieve heat dissipation. This reflects the high efficiency and energy saving of liquid cooling, which is more obvious in the heat dissipation process of high heat flux equipment.

14.3.2.2 Advantages

Energy saving: Compared with traditional air-cooled data center, immersion liquid cooling can reduce energy consumption by 90–95%.The customized server removes the cooling fan and is immersed in the coolant at a more uniform temperature, reducing energy consumption by 10–20%.

Cost saving: The immersion liquid-cooled data center has a small infrastructure scale, and the construction cost is not higher than the traditional computer room. The ultra-low PUE value can greatly reduce the operating cost of the data center, saving the total cost of ownership by 40–50%.

Low noise: The server can remove fans, minimize noise pollution sources, and make the data center to achieve absolute silence.

High reliability: The coolant is nonconductive; the flash point is high and nonflammable, which makes data center no fire risk of water leakage, IT equipment no risk of gas corrosion, and can eliminate mechanical vibration damage to IT equipment.

High thermal dissipation: The immersion liquid cooling equipment can solve the heat dissipation problem of ultrahigh density data centers. According to the 42U capacity configuration of a single cabinet, the traditional 19-in standard server is placed, and the power density of a single cabinet can range from 20 to 200 kW.

14.3.2.3 Disadvantages

High cost coolant: Immersion liquid cooling equipment needs coolant with appropriate cost, good physical and chemical properties, and convenient use, which means the cost of coolant is still high.

Complex data center operations: As an innovative technology, immersion liquid cooling equipment has different maintenance scenarios and operation modes from traditional air-cooled data center equipment. There are a lot of challenges such as, how to install and move IT equipment, how to quickly and effectively handle residual coolant on the surface of the equipment, how to avoid the loss of coolant in the operating process, how to guarantee the safety and health of maintenance personnel and how to optimize the design.

Balance coolant distribution: It is an important challenge in the heat dissipation design of submerged liquid cooling equipment to efficiently use the coolant to avoid local hot spots and ensure the accurate cooling of each IT equipment.

IT compatibility: Some parts of IT equipment have poor compatibility with coolant, such as fiber optic module, which cannot work properly in liquid due to different refractive index of liquid and air and needs to be customized and sealed. Ordinary solid-state drives are not compatible with the coolant and cannot be immersed directly in the coolant for cooling.

In addition, there are still no large-scale application cases of liquid cooling technology, especially immersion liquid cooling technology in the data center.

14.4 CONCLUSIONS AND FUTURE TRENDS

Rack-level cooling technology can be used with success in many situations where the existing infrastructure or conventional cooling approaches present difficulties. The advantages come from one or more of these three attributes:

1. Rack-level cooling solutions offer energy efficiency advantages due to their close proximity to the IT equipment being cooled. Therefore the heat is transferred at higher temperature differences and put into a water flow sooner.
   This proximity provides two potential advantages:
   a. The cooling water temperature supplied by the external cooling infrastructure can be higher, which opens opportunities for lower energy use.
   b. A larger percentage of heat is moved inside the data center using water and pumps compared with the

less efficient method of moving large volumes of heated air using fans.

Note: When rack cooling is installed, the potential energy savings may be limited if the existing cooling systems are not optimized either manually or by automatic controls.

2. Rack-level cooling can solve hotspot problems when installed with high-density IT equipment. This is especially true when the existing room cooling systems cannot be modified or adjusted to provide the needed cooling in a particular location.
3. Rack-level cooling systems are often provided with controls allowing efficiency improvements as the IT equipment workload varies. Conventional data center room cooling systems historically have a limited ability to adjust efficiently to changes in load. This is particularly evident when CRAH or CRAC fan speeds are not reduced when the cooling load changes.

As mentioned, new IT equipment is providing an increase in heat load per square foot. To address this situation, rack-level cooling is constantly evolving with new models frequently coming to the market.

Recent trends in IT equipment cooling indicate new products will involve heat transfer close to or contacting high heat generating components that are temperature sensitive.

Many current and yet-to-be-introduced solutions will be successful in the market given the broad range of applications starting with the requirements at a supercomputer center and ending with a single rack containing IT equipment.

Whatever liquid cooling technology is chosen, it will always be more efficient than air for two reasons. The first and most important is the amount of energy required to move air will always be several times greater than that to move a liquid for the same amount of cooling.

ACKNOWLEDGEMENT

Our sincere thanks go to Henry Coles, Steven Greenberg, and Phil Hughes who prepared this chapter in the first edition of the *Data Center Handbook*. We have reorganized the content with some updates.

FURTHER READING

ASHRAE Technical Committee 9.9. Thermal guidelines for data processing environments–expanded data center classes and usage guidance. Whitepaper; 2011.

ASHRAE Technical Committee 9.9. Mission critical facilities, technology spaces, and electronic equipment.

Bell GC. Data center airflow management retrofit technology case study bulletin. Lawrence Berkeley National Laboratory; September 2010. Available at https://datacenters.lbl.gov/sites/default/files/airflow-doe-femp.pdf. Accessed on June 28, 2020.

Coles H, Greenberg S. Demonstration of intelligent control and fan improvements in computer room air handlers. Lawrence Berkeley National Laboratory, LBNL-6007E; November 2012.

CoolCentric. Frequently asked questions about rear door heat exchangers. Available at http://www.coolcentric.com/?s=frequent+asked+questions&submit=Search. Accessed on June 29, 2020.

Greenberg S. Variable-speed fan retrofits for computer-room air conditioners. The U.S. Department of Energy Federal Energy Management Program, Lawrence Berkeley National Laboratory; September 2013. Available at https://www.energy.gov/sites/prod/files/2013/10/f3/dc_fancasestudy.pdf. Accessed on June 28, 2020.

Hewitt GF, Shires GL, Bott TR. *Process Heat Transfer*. CRC Press; 1994.

http://en.wikipedia.org/wiki/Stefan%E2%80%93Boltzmann_law.

http://en.wikipedia.org/wiki/Aquasar.

https://www.asetek.com/data-center/technology-for-data-centers. Accessed on September 17, 2020.

Koomey JG. Growth in data center electricity use 2005 to 2010. A report by analytics press, completed at the request of The New York Times; August 1, 2011. Available at https://www.koomey.com/post/8323374335. Accessed on June 28, 2020.

Made in IBM Labs: IBM Hot Water-Cooled Supercomputer Goes Live at ETH Zurich.

Moss D. Data center operating temperature: what does dell recommend?. Dell Data Center Infrastructure; 2009.

Rasmussen N. Guidelines for specification of data center power density. APC White Paper #120; 2005.

www.brighthubengineering.com/hvac/92660-natural-convection-heat-transfer-coefficient-estimation-calculations/#imgn_2.

www.clusteredsystems.com.

15

CORROSION AND CONTAMINATION CONTROL FOR MISSION CRITICAL FACILITIES

CHRISTOPHER O. MULLER
Muller Consulting, Lawrenceville, Georgia, United States of America

15.1 INTRODUCTION

Data Center \ ’dāt-ə (’dat-, ’dät-) ’sent-ər \ (circa 1990) n (i) a facility used to house computer systems and associated components, such as telecommunications and storage systems. It generally includes redundant or backup power supplies, redundant data communications connections, environmental controls (e.g., air conditioning, fire suppression), and security devices; (ii) a facility used for housing a large amount of computer and communications equipment maintained by an organization for the purpose of handling the data necessary for its operations; (iii) a secure location for web hosting servers designed to assure that the servers and the data housed on them are protected from environmental hazards and security breaches; (iv) a collection of mainframe data storage or processing equipment at a single site; (v) areas within a building housing data storage and processing equipment.

Data centers operating in areas with elevated levels of ambient pollution can experience hardware failures due to changes in electronic equipment mandated by several “lead-free” regulations that affect the manufacturing of electronics, including IT and datacom equipment. The European Union directive “on the ***R***estriction ***o***f the use of certain ***H***azardous ***S***ubstances in electrical and electronic equipment” (RoHS) was only the first of many lead-free regulations that have been passed. These regulations have resulted in an increased sensitivity of printed circuit boards (PCBs), surface-mounted components, hard disk drives, computer workstations, servers, and other devices to the effects of corrosive airborne contaminants. As a result, there is an increasing requirement for air quality monitoring in data centers.

Continuing trends toward increasingly compact electronic datacom equipment makes gaseous contamination a significant data center operations and reliability concern. Higher power densities within air-cooled equipment require extremely efficient heat sinks and large volumes of air movement, increasing the airborne contaminant exposure. The uses of lead-free solders and finishes used to assemble electronic datacom equipment also bring additional corrosion vulnerabilities.

When monitoring indicates that data center air quality does not fall within specified corrosion limits, and other environmental factors have been ruled out (i.e., temperature, humidity.), gas-phase air filtration should be used. This would include air being introduced into the data center from the outside for ventilation and/or pressurization as well as all the air being recirculated within the data center. The optimized control of particulate contamination should also be incorporated into the overall air handling system design.

Data centers operating in areas with lower pollution levels may also have a requirement to apply enhanced air cleaning for both gaseous and particulate contaminants especially when large amounts of outside air are being used for “free cooling” and results in increased contaminant levels in the data center. As a minimum, the air in the data center should be recirculated through combination gas-phase/particulate air filters to remove these contaminants as well as contaminants generated within the data center in order to maintain levels within specified limits.

General design requirements for the optimum control of gaseous and particulate contamination in data centers include sealing and pressurizing the space to prevent infiltration of contaminants, tightening controls on temperature and humidity, improving the air distribution throughout the data

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

center, and application of gas-phase and particulate filtration to fresh (outside) air systems, recirculating air systems, and computer room air conditioners.

The best possible control of airborne pollutants would allow for separate sections in the mechanical system for particulate and gaseous contaminant control. However, physical limitations placed on mechanical systems, such as restrictions in size and pressure drop, and constant budgetary constraints require new types of chemical filtration products.

This document will discuss application of gas-phase air and particulate filtration for the data center environment, with primary emphasis on the former. General aspects of air filtration technology will be presented with descriptions of chemical filter media, filters, and air cleaning systems and where these may be employed within the data center environment to provide for enhanced air cleaning.

15.2 DATA CENTER ENVIRONMENTAL ASSESSMENT

A simple quantitative method to determine the airborne corrosivity in a data center environment is by "reactive monitoring" as first described in ISA Standard 71.04-1985 Environmental Conditions for Process Measurement and Control Systems: Airborne Contaminants. Copper coupons are exposed to the environment for a period of time and quantitatively analyzed using electrolytic (cathodic, coulometric) reduction to determine corrosion film thickness and chemistry. Silver coupons should be included with copper coupons to gain a complete accounting of the types and nature of the corrosive chemical species in the environment. For example, sulfur dioxide alone will corrode only silver to form Ag_2S (silver sulfide), whereas sulfur dioxide and hydrogen sulfide in combination will corrode both copper and silver forming their respective sulfides.

15.2.1 ISA[1] Standard 71.04-2013

ANSI/ISA-71.04-2013 classifies several levels of environmental severity for electrical and electronic systems: G1, G2, G3, and GX, providing a measure of the corrosion potential of an environment. G1 is benign and GX is open-ended and the most severe (Table 15.1).

In a study performed by Rockwell Automation looking at lead-free finishes, four alternate PCB finishes were subjected to an accelerated mixed flowing gas corrosion test. Important findings can be summarized as follows:

1. The electroless nickel immersion gold (ENIG) and immersion silver (ImmAg) surface finishes failed early in the testing. These coatings are the most susceptible to corrosion failures and are expected to be much more susceptible than traditional hot air solder leveling (HASL) coatings. The use of these two coatings may make the PCB the weak link regarding the sensitivities of the electronic devices to corrosion.
2. None of the coatings can be considered immune from failure in an ISA Class G3 environment.
3. The gold and silver coatings could not be expected to survive a mid to high Class G2 environment based on these test results.

[1] International Society of Automation (www.isa.org).

TABLE 15.1 ISA classification of reactive environments

Severity level	G1 Mild	G2 Moderate	G3 Harsh	G*X* Severe
Copper reactivity level (Å)[a]	<300	<1,000	<2,000	≥2,000
Silver reactivity level (Å)[a]	<200	<1,000	<2,000	≥2,000

[a] Measured in angstroms after 1 month's exposure. Source: ISA.

According to a leading world authority on RoHS, ERA Technology, "Recent research has shown that PCBs made using lead-free materials can be more susceptible to corrosion than their tin/lead counterparts". Experts are working diligently to address these concerns, but they cannot be addressed overnight.

The Reliability and Failure Analysis group at ERA Technology has diagnosed failures in electronic devices due to interaction with low levels of gaseous sulfides—failures that caused both a financial impact to the manufacturers and safety issues with their customers. Recent work showed that corrosion could occur even with measured hydrogen sulfide levels as low as 0.2 μg/m^3 (0.14 ppb). Another reference describes the formation of a 200-Å thick layer of silver sulfide in 100 hours at a concentration of just 100 μg/m^3 [72 ppb].

15.2.2 Corrosive Gases

There are three types of gases that can be considered as prime candidates in the corrosion of data center electronics: acidic gases such as hydrogen sulfide, sulfur and nitrogen oxides, chlorine, and hydrogen fluoride; caustic gases, such as ammonia; and oxidizing gases, such as ozone. Of these, the acidic gases are of particular concern. For instance, it takes only 10 ppb (28.98 μg/m^3) of chlorine to inflict the same amount of damage as 25,000 ppb (17.40 mg/m^3) of ammonia.

Each site may have different combinations and concentration levels of corrosive gaseous contaminants. Performance degradation can occur rapidly or over many years, depending on the specific conditions at a site. Descriptions of

pollutants common to urban and suburban locations in which most data centers are located and a discussion of their contributions to IT equipment performance degradation follow.

15.2.2.1 Sulfur Oxides

Oxidized forms of sulfur (SO_2, SO_3) are generated as combustion products of fossil fuels and from motor vehicle emissions. Low parts per billion levels of sulfur oxides can cause reactive metals to be less reactive and thus retard corrosion. At higher levels, however, they will attack certain types of metals. The reaction with metals normally occurs when these gases dissolve in water to form sulfurous and sulfuric acid acids (H_2SO_3 and H_2SO_4).

15.2.2.2 Nitrogen Oxides (NO_X)

Some common sources of reactive gas compounds (NO, NO_2, N_2O_4) are formed as combustion products of fossil fuels and have a critical role in the formation of ozone in the atmosphere. They are also believed to have a catalytic effect on the corrosion of base metals caused by chlorides and sulfides. In the presence of moisture, some of these gases form nitric acid (HNO_3) that, in turn, attacks most common metals.

15.2.2.3 Active Sulfur Compounds

Active sulfur compound refers to hydrogen sulfide (H_2S), elemental sulfur (S), and organic sulfur compounds such as mercaptans (R-SH). When present at low ppb levels, they rapidly attack copper, silver, aluminum, and iron alloys. The presence of moisture and small amounts of inorganic chlorine compounds and/or nitrogen oxides greatly accelerate sulfide corrosion. Note, however, that attack still occurs in low relative humidity environments. Active sulfurs rank with inorganic chlorides as the predominant cause of electronic equipment corrosion.

15.2.2.4 Inorganic Chlorine Compounds

This group includes chlorine (Cl_2), chlorine dioxide (ClO_2), hydrogen chloride (HCl), etc., and reactivity will depend upon the specific gas composition. In the presence of moisture, these gases generate chloride ions that, in turn, attack most copper, tin, silver, and iron alloys. These reactions are significant even when the gases are present at low ppb levels. At higher concentrations, many materials are oxidized by exposure to chlorinated gases. Particular care must be given to equipment that is exposed to atmospheres which contain chlorinated contaminants. Sources of chloride ions, such as seawater, cooling tower vapors, and cleaning compounds, etc., should be considered when classifying data center environments.

15.2.2.5 Photochemical Species

The atmosphere contains a wide variety of unstable, reactive species that are formed by the reaction of sunlight with moisture and other atmospheric constituents. Some have lifetimes measured in fractions of a second as they participate in rapid chain reactions. In addition to ozone (O_3), a list of examples would include the hydroxyl radical as well as radicals of hydrocarbons, oxygenated hydrocarbons, nitrogen oxides, sulfur oxides, and water. Ozone can function as a catalyst in sulfide and chloride corrosion of metals.

15.2.2.6 Strong Oxidants

This includes ozone plus certain chlorinated gases (chlorine, chlorine dioxide). Ozone is an unstable form of oxygen that is formed from diatomic oxygen by electrical discharge or by solar radiation in the atmosphere. These gases are powerful oxidizing agents. Photochemical oxidation—the combined effect of oxidants and ultraviolet light (sunlight)—is particularly potent.

Sulfur dioxide (SO_2), hydrogen sulfide (H_2S), and sand active chlorine compounds (Cl_2, HCl, ClO_2) have all been shown to cause significant corrosion in electrical and electronic equipment at concentrations of just a few parts per billion in air. Even at levels that are not noticed by or harmful to humans, these gases can be deadly to electronic equipment.

15.3 GUIDELINES AND LIMITS FOR GASEOUS CONTAMINANTS

Established gaseous composition environmental limits, listed in Table 15.2, have been published in standards such as ISA 71.04, IEC 60721-3-3, Telcordia GR-63-CORE, and IT equipment manufacturers' own internal standards. These limits serve as requirements and guides for specifying data center environmental cleanliness, but they are not useful for surveying the corrosivity or predicting the failure rates of hardware in the data center environment for two reasons. First, gaseous composition determination is not an easy task. Second, predicting the rate of corrosion from gaseous contamination composition is not a straightforward exercise.

An additional complication to determine corrosivity is the synergy between gases. For example, it has been demonstrated that hydrogen sulfide alone is relatively not corrosive to silver when compared to the combination of hydrogen sulfide and nitrous oxide, which is very corrosive to silver. Correspondingly, neither sulfur dioxide nor nitrous oxide alone is corrosive to copper, but together they attack copper at a very fast rate.

TABLE 15.2 Published gaseous contaminant limits for IT equipment

Gas	IEC 60721-3-3	GR-63-CORE	ISA S71.04	Manufacturer's internal standard
Hydrogen sulfide (H_2S)	10 μg/m^3 (3.61273×10^{-13} lb/in^3) 7 ppb	55 μg/m^3 (1.987×10^{-12} lb/in^3) 40 ppb	4 μg/m^3 (1.44509×10^{-13} lb/in^3) 3 ppb	3.2 μg/m^3 (1.15607×10^{-13} lb/in^3) 2.3 ppb
Sulfur dioxide (SO_2)	100 μg/m^3 (3.61273×10^{-12} lb/in^3) 38 ppb	131 μg/m^3 (4.73268×10^{-12} lb/in^3) 50 ppb	26 μg/m^3 (9.3931×10^{-13} lb/in^3) 10 ppb	100 μg/m^3 (3.61273×10^{-12} lb/in^3) 38 ppb
Hydrogen chloride (HCl)	100 μg/m^3 (3.61273×10^{-12} lb/in^3) 67 ppb	7 μg/m^3 (2.52891×10^{-13} lb/in^3) 5 ppb[a]	—	1.5 μg/m^3 (5.41909×10^{-14} lb/in^3) 1 ppb
Chlorine (Cl_2)	100 μg/m^3 (3.61273×10^{-12} lb/in^3) 34 ppb	14 μg/m^3 (5.05782×10^{-13} lb/in^3) 5 ppb*	3 μg/m^3 (1.08382×10^{-13} lb/in^3) 1 ppb	—
Nitrogen oxides (NO_X)	—	700 ppb	50 ppb	140 μg/m^3 (5.05782×10^{-12} lb/in^3)
Ozone (O_3)	10 μg/m^3 (3.61273×10^{-13} lb/in^3) 5 ppb	245 μg/m^3 (8.85119×10^{-12} lb/in^3) 125 ppb	4 μg/m^3 (1.44509×10^{-13} lb/in^3) 2 ppb	98 μg/m^3 (3.54047×10^{-12} lb/in^3) 50 ppb
Ammonia (NH_3)	300 μg/m^3 (1.08382×10^{-11} lb/in^3) 430 ppb	348 μg/m^3 (1.25723×10^{-11} lb/in^3) 500 ppb	348 μg/m^3 (1.25723×10^{-11} lb/in^3) 500 ppb	115 μg/m^3 (4.15464×10^{-12} lb/in^3) 165 ppb
Volatile organics (C_XH_X)	—	5,000 μg/m^3 (1.80636×10^{-10} lb/in^3) 1,200 ppb	—	—

[a] Total HCl and Cl_2. Source: IEC, ISA, Telecordia.

Although Table 15.2 can be used to provide some indication of the possible harmful effects of several common contaminants, the data center environment needs a single set of limits, which will require considerable study and research. As the industry works toward a single set of limits, caveats or exceptions to generally accepted limits will exist. These exceptions will improve as the interactions of concentration, composition, and the thermal environment combine and become better understood along with their effects on the datacom equipment.

15.4 AIR CLEANING TECHNOLOGIES

Increasingly, enhanced air cleaning is being used in data centers to provide and maintain acceptable air quality with many options available for the control of particulate pollutants, and nearly as many options for the control of gaseous pollutants. Employing the proper level and type(s) of air filtration can effectively reduce airborne contaminants to well below specified levels and minimize equipment failure rates, but effective control of environmental pollutants requires the use of an air cleaning strategy optimized for both particulate and chemical removal.

15.4.1 Particulate Filtration

The control of particulates can be a considered a "mature" air cleaning application based on the number of technologies in everyday use and the relative ease of applying these technologies for a specific application. ASHRAE Technical Committee 9.9 has published recommended particulate filtration requirements for data centers.

15.4.2 Gas-Phase Air Filtration

Just as there are many options available for the control of particulate pollutants, there are nearly as many options for the control of gaseous pollutants. The problem is that for most data center designers this type of air cleaning is not as well understood and is not as easily applied. Also, most ventilation systems and computer room air conditioners/

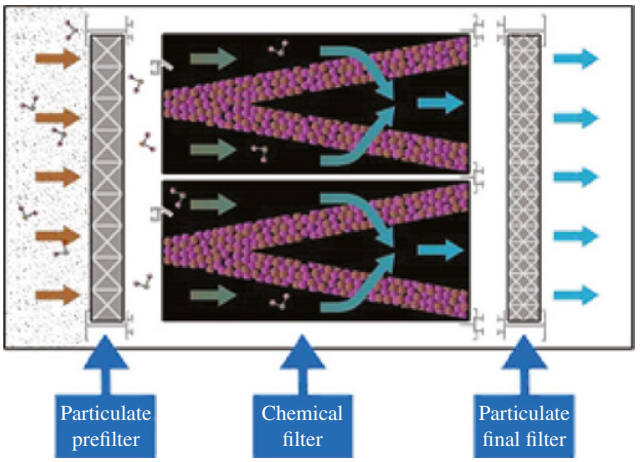


FIGURE 15.1 Schematic of an enhanced air cleaning system. Source: Courtesy of Purafil, Inc.

computer room air handlers/(CRACs/CRAHs) are not designed to readily accommodate this type of air cleaning technology.[2]

Gas-phase air filters employing one or more granular adsorbent media, used in combination with particulate filters, have proven to be very effective for the control of pollutants (Fig. 15.1). This "one-two punch" allows for the maximization of both particulate control and gaseous pollutant control within the same system. Physical limitations placed on these systems, such as restrictions in size and pressure drop, and constant budgetary constraints have spurred the development of new types of, and delivery systems for, gas-phase air filtration products. Foremost among these are filters using a monolithic extruded carbon composite media (ECC) and an adsorbent-loaded nonwoven fiber media (ALNF).

15.5 CONTAMINATION CONTROL FOR DATA CENTERS

There is no one standard for data center design, thus the application of air cleaning in a data center may involve several different technologies depending on whether the air handling system uses outdoor air to provide for ventilation, pressurization, and/or free cooling, or whether computer room air conditioning (CRAC) units are used as 100% recirculating air systems.

[2] Though they serve the same purpose, i.e., to provide precise temperature and humidity control, there is a fundamental difference between a CRAC and CRAH. A CRAC includes an internal compressor, using the direct expansion of refrigerant to remove heat from the data center. A CRAH includes only fans and a cooling coil, often using chilled water to remove heat from the data center. Although this document generically refers to CRAC units, the same design considerations can be applied for CRAHs.

The optimum control of airborne pollutants would allow for separate sections in the mechanical system for particulate and gaseous contaminant control. If this is not practical from a design or cost standpoint, air cleaning may be integrated directly into the fresh air systems or CRAC units or applied as stand-alone systems. Again, because most of these air handling systems already have particulate filtration as part of their standard design, the manufacturers would have to be consulted to determine what limitations there might be for the addition of gas-phase air filters. Most of these concerns would center on the additional static pressure from these filters.

The following sections will describe some basic steps for the optimization and application of enhanced air cleaning for the data center environment.

15.5.1 Basic Design Requirements

Before one considers adding enhanced air cleaning for either particulate or gas-phase contamination in a data center, there are specific mechanical design requirements which must be understood and considered.

15.5.1.1 Room Air Pressurization

In order to prevent contaminated air from infiltrating the data center, all critical areas must be maintained at a slight positive pressure. This can be achieved by pressurizing the room to ~0.02–0.04 iwg (inch of water gage) (5–10 Pa) by introducing ventilation (outdoor) air at a rate of 3–6 air changes per hour (5–10% of the gross room volume per minute).

15.5.1.2 Room Air Recirculation

Air cleaning systems can be designed that can function as pressurization-only systems or as pressurization and recirculation systems. Depending upon how well the data center environment is sealed, the amount of pedestrian traffic into and out of the space, and the level of other internally generated contaminants, pressurization only may be enough to provide an acceptable level of contamination control.

The general recommendation is the recirculation of tempered air through an air cleaning unit if:

1. The room is not properly sealed.
2. The space has high pedestrian traffic.
3. Sources of internally generated contaminants have been identified and source control is not practical.
4. The CRAC units or negative pressure ductwork are located outside the data center environment.
5. One or more of the walls of the data center are outside walls.

The rate of room air recirculation will be determined by the type of equipment used and the construction parameters of the data center. Typical recommendations call for 6–12 air changes per hour (approximately 10–20% of the gross room volume per minute).

15.5.1.3 Temperature and Humidity Control

The corrosive potential of any environment increases dramatically with increasing relative humidity. Rapid changes in relative humidity can result in localized areas of condensation and, ultimately, in corrosive failure.

ASHRAE Technical Committee 9.9 published Thermal Guidelines for Data Processing Environments which extended the temperature–humidity envelope to provide greater flexibility in data center facility operations, particularly with the goal of reducing energy consumption.

For high reliability, TC 9.9 recommends that data centers be operated in the ranges shown in Table 15.3. These guidelines have been agreed to by all major IT manufacturers and are for legacy IT equipment. A downside of expanding the temperature–humidity envelope is the reliability risk from higher levels of gaseous and particulate contamination entering the data center. Lack of humidity control is creating a controversy. Unfortunately, decisions are being based more on financial concerns than engineering considerations.

15.5.1.4 Proper Sealing of Protected Space

Without a tightly sealed room, it will be very difficult to control the four points mentioned above. It is essential that the critical space(s) be protected by proper sealing. Actions taken to accomplish this include the use of airlock entries/exits, sealing around doors and windows, door jambs should fit tightly or door sweeps should be used, closing and sealing all holes, cracks, wall and ceiling joints and cable, pipe, and utility penetrations with a fireproof vapor-retarding material. Care should be taken to assure that any space above a drop ceiling or below a raised floor is sealed properly.

15.5.2 Advanced Design Requirements

15.5.2.1 Particulate Control

Filtration is an effective means of addressing airborne particulate in the data center environment. It is important that all air handlers serving the data center have the appropriate particulate filters to ensure appropriate conditions are maintained within the room, in this case to meet the cleanliness level of ISO Class 8. The necessary efficiency is dependent on the design and application of the air handlers.

In-room process cooling with recirculation is the recommended method of controlling the data center environment.

TABLE 15.3 Temperature and humidity recommendations for data centers

	Equipment environmental specifications							
	Product operations[b,c]					Product power off[c,d]		
Classes[a]	Dry-bulb temperature (°C)[e,g]	Humidity range, noncondensing[h,i]	Maximum dew point (°C)	Maximum elevation[m]	Maximum rate of charge (°C/h)[f]	Dry-bulb temperature (°C)	Relative humidity (%)	Maximum dew point (°C)
Recommended (Applies to all A classes; individual data centers can choose to expand this range based upon the analysis described in this document)								
A1–A4	18–27	5.5°C DP to 60% RH and 15°C DP						
Allowable								
A1	15–32	20–80% RH	17	3,050	5/20	5–45	8–80	27
A2	10–35	20–80% RH	21	3,050	5/20	5–45	8–80	27
A3	5–40	−12°C DP and 8–85% RH	24	3,050	5/20	5–45	8–85	27
A4	5–45	−12°C DP and 8–90% RH	24	3,050	5/20	5–45	8–90	27
B	5–35	8–80% RH	28	3,050	NA	5–45	8–80	29
C	5–40	8–80% RH	28	3,050	NA	5–45	8–80	29

Note: Please visit original document for superscript symbols for footnotes. Source: ASHRAE 2016 Thermal Guidelines

Air from the hardware areas is passed through the CRAC units where it is filtered and cooled, and then introduced into the subfloor plenum. The plenum is pressurized, and the conditioned air is forced into the room through perforated tiles and then travels back to the CRAC unit for reconditioning. The airflow patterns and design associated with a typical computer room air handler have a much higher rate of air change than do typical comfort cooling air conditioners. This means that the air is much cleaner than in an office environment. Proper filtration can thus accomplish a great deal of particulate arrestance.

Any air being introduced into the data center for ventilation or positive pressurization should first pass through high-efficiency filtration. Ideally, air from sources outside the building should be filtered using high-efficiency particulate air (HEPA) filtration rated at 99.97% efficiency or greater at a particle size of 0.3 μm.

It is also important that the filters used are properly sized for the air handlers. For instance, gaps around the filter panels can allow air to bypass the filter as it passes through the CRAC unit. Any gaps or openings should be taped, gasketed, or filled using appropriate materials, such as stainless-steel panels or custom filter assemblies. The filtration requirements for CRAC units and the air coming into the data center are described in an article written by ASHRAE Technical Committee 9.9 on contamination limits for data centers and an ASHRAE white paper on gaseous and particulate contamination guidelines for data centers.

15.5.2.2 Gaseous Contamination Control

Assuming a data center's HVAC system is already equipped with adequate particulate filtration, gaseous air cleaning can be used in conjunction with the existing air handling systems. Gas-phase air filters or filtration systems employing one or more adsorbent and/or chemisorbent media can effectively reduce gaseous contaminants to well below specified levels. Properly applied gaseous air cleaning also has the potential for energy savings.

Computer Room Air Conditioning (CRAC) Units
Almost all CRAC units already have particulate filtration built in that can be retrofitted to use ALNF combination particulate and chemical filters (Fig. 15.2). With this type of filter, one can maintain the same level of particulate filtration while adding the chemical filtration required for the control of low levels of gaseous contamination. The pressure drops of these filters are slightly higher than the particulate filters they replace, but still well below the maximum terminal pressure drop. Most CRAC units use standard 1–4 in (25–100 mm) filters, however, the majority employ "non-standard" or proprietary sizes.

FIGURE 15.2 ALNF filters. Source: Courtesy of Purafil, Inc.

ECC filters (Fig. 15.3) are also being used in CRAC units to provide control of low-to-moderate levels of gaseous contamination. Standard 2 and 4 in (50 and 100 mm) filters are available, but do not provide particulate filtration. If this is required, a 2 in (50 mm) particulate filter can be supplied in front of a 2 in (50 mm) ECC filter as a packaged unit.

Makeup (Outdoor, Fresh) Air Handlers
If ventilation and/or pressurization air is being supplied by an existing makeup air handler, chemical filtration may be

FIGURE 15.3 ECC filters. Source: Courtesy of Purafil, Inc.

FIGURE 15.4 Bulk media modules. Source: Christpher O. Muller.

added as a retrofit using ECC filters or ALNF filters, depending on the capabilities of the air handler and the requirements for air cleaning. If this is not a practical solution, then one should consider integrating a separate air cleaning section onto the existing makeup air handler(s) incorporating ECC filters, ALNF filters, or refillable or disposable bulk media modules (Fig. 15.4). If additional air is required for pressurization, a stand-alone air cleaning system may be required incorporating ECC filters, modules, or bulk-fill media in deep beds.

Side Access System

A side access system (SAS, Fig. 15.5) is designed to remove both particulate and gaseous pollutants from outdoor (makeup) air for corrosion control. The SAS should be designed such that a positive seal is created to prevent air bypass and enhance filtration efficiency. When outdoor air is being delivered either directly to the data center or indirectly through a mechanical room, the SAS can be used as powered or non-powered units designed to control low-to-moderate levels of gaseous contaminants. This type of system can offer a wide range of particulate prefilters, chemical filters (media modules, ECC, ALNF), and particulate final filters to accommodate specific airflow requirements within the primary outside air handling system. A secondary unit can be used for recirculation in mechanical or equipment rooms.

FIGURE 15.5 Side access system installed at a bank data center. Source: Courtesy of Purafil, Inc.

Positive Pressurization Unit

A positive pressurization unit (PPU, Fig. 15.6) is designed to filter low-to-moderate concentrations of outside air contaminants. It is used to supply cleaned pressurization air to the critical space(s) and contains a particulate prefilter, two stages of 4 in (100 mm) ECC filters or media modules, a 90% particulate final filter, a blower section, and an adjustable damper for control of pressurization air into the air handling system or directly into the data center.

Recirculating Air Handler

A recirculating air unit (RAU, Fig. 15.7) is an in-room, self-contained unit used to provide increased amounts of recirculation air to areas with low-to-moderate gaseous contaminant levels. In data center applications, recirculating air handlers (RAHs) would contain a prefilter, two stages of 4 in (100 mm) ECC filters or media modules, a blower section, and a 90% final filter. These units are used

FIGURE 15.6 Positive pressurization unit. Source: Courtesy of Purafil, Inc.

FIGURE 15.7 Recirculating air system. Source: Courtesy of Purafil, Inc.

to further filter and polish room air in order to maintain very low contaminant levels.

Deep Bed Scrubber

A deep bed scrubber (DBS, Fig. 15.8) is designed for areas where higher levels of contaminant gases are present and other systems cannot economically provide the filtration required to meet air quality standards. The DBS provides protection against environmentally induced corrosion and is designed to provide clean pressurization air to the data center environment. Specific contaminant removal efficiencies can be met using up to three different chemical filtration media. The DBS is designed to be compatible with the PSA (Purafil Side Access) or CA (Corrosive Air) system when requirement require pressurization with recirculation.

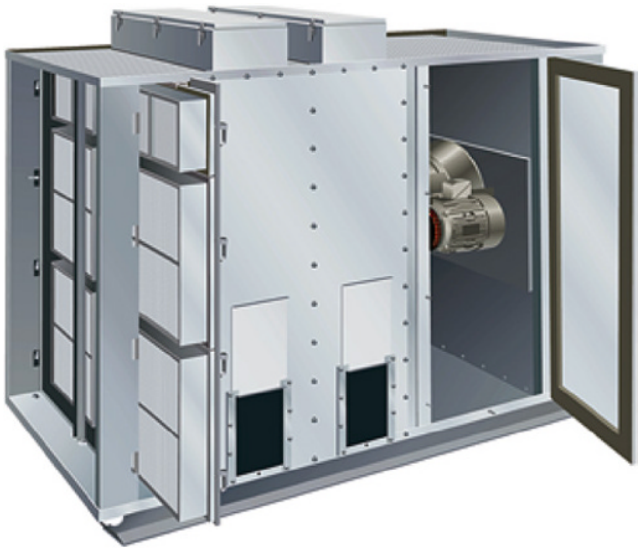


FIGURE 15.8 Deep bed scrubber. Source: Courtesy of Purafil, Inc.

Under-Floor Air Filtration

The latest innovation for the application of gas-phase air filtration is the use of ECC filters under the perforated panels on the cold aisles in raised floor systems. The filter is placed in a customized "tray" under the perforated panel and fits the dimensions of the existing access floor grid. Gasketing around the filter assembly assures that 100% of the air being delivered into the data center goes through the ECC filter for total gaseous contaminant control. Sealing the subfloor plenum will also help to maximize the amount of air going through the under-floor ECC filters and ultimately the amount of clean air being delivered to data center.

There are many types of commercially available floors that offer a wide range of structural strength and loading capabilities, depending on component construction and the materials used. The general types of raised floors include stringerless, stringered, and structural platforms. For installation of the ECC filter into a raised floor system, the stringered floor system is most applicable. The ECC may also be used with structural platforms, but there are more restrictions in their application.

A stringered raised floor (Fig. 15.9) generally consists of a vertical array of steel pedestal assemblies (each assembly is comprised of a steel base plate, tubular upright, and a head) uniformly spaced on two-foot centers and mechanically fastened to the concrete floor.

Gas-phase air filtration can be applied in several locations within and outside the data center environment (Fig. 15.10). Filters can be added to existing air handling equipment given proper design considerations or supplied as stand-alone pressurization and/or recirculation equipment.

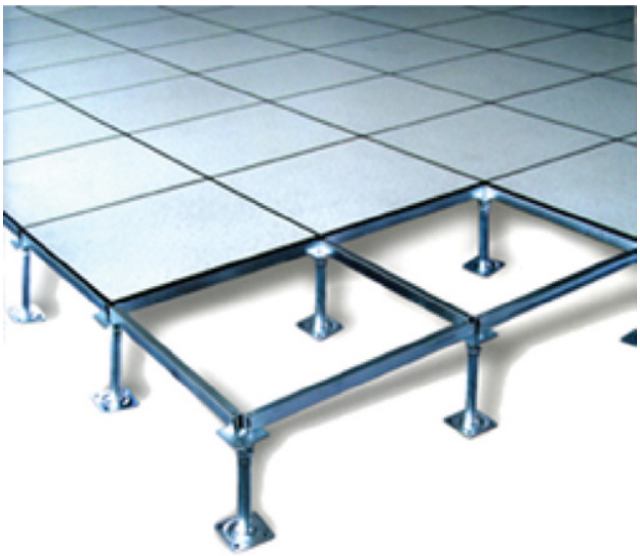


FIGURE 15.9 Raised access floor system. Source: Courtesy of Purafil, Inc.

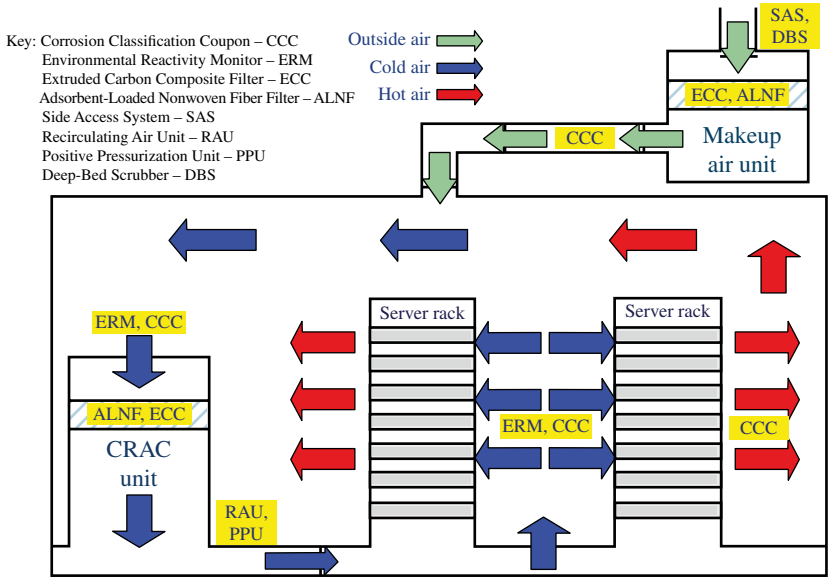


FIGURE 15.10 Data center schematic with possible locations of enhanced air cleaning and corrosion monitoring systems. Source: Courtesy of Purafil, Inc.

15.5.2.3 Air-Side Economizers

An economizer is a mechanical device used to reduce energy consumption. Economizers recycle energy produced within a system or leverage environmental temperature differences to achieve efficiency improvements. The primary concern in this approach to data center cooling is that outside air contaminants—both particulate and gas-phase—will have a negative impact on electronics.

Research performed by Lawrence Berkley National Laboratory stated "The pollutant of primary concern, when introducing particulate matter to the data center environment, is fine particulate matter that could cause conductor bridging," The study also concluded that ". . .filtration systems in most data centers do just fine in keeping contaminants out." However, this was referencing particulate filtration and not the use of gas-phase air filtration to address the potential damage to electronic equipment from the introduction of unwanted gaseous contaminants.

Air-side economizers typically include filters with a minimum ASHRAE-rated particulate removal efficiency of 40% (MERV 9) to reduce the amount of particulate matter or contaminants that are brought into the data center space. However, in areas with high ambient particulate levels, ASHRAE 60–90% (MERV 11–13) filter may be required.

Some references have been found describing the use of gas-phase air filter in economizers, especially since the advent of RoHS and other "lead-free" regulation. However, with the increasing pressure to reduce energy consumption in data centers, and the increasing interest in the use of air-side economizers for "free cooling," data centers located in regions with poor ambient air quality will struggle to maintain an environment conducive for the protection of sensitive electronic equipment. ECC filters and/or ALNF filters can be easily applied in these systems to address this serious contamination issue.

15.6 TESTING FOR FILTRATION EFFECTIVENESS AND FILTER LIFE

Once enhanced air cleaning has been specified and installed, one must be able to determine the effectiveness of the particulate and gas-phase air filters. One must also be able to replace the filters on a timely basis so as not to compromise the data center environment.

15.6.1 Particulate Contamination Monitoring

Filtration effectiveness can be measured using real-time particle counters in the data center. Excess particle counts or concentrations can indicate filter failure, filter bypass, and/or internal sources of particle generation, e.g., CRAC belt dust, tin whiskers. Particle monitoring in a data center is generally needed daily; it is usually done only when there is a notable problem that could be caused by particulate contamination.

Particulate filters have specified initial and final pressure drops at rated air flows and differential pressure

gauges can be used to observe filter life and set alarm limits. Timely replacement of prefilters, primary and final filters not only protects the electronic equipment but also maintains optimum performance of the air handling equipment.

15.6.2 Gaseous Contamination Monitoring

The primary role of gas-phase air filters and filtration systems in the data center is to prevent corrosion from forming on sensitive electronic equipment. Therefore, the best measure of their effectiveness would be to use either passive Corrosion Classification Coupons (CCCs, Fig. 15.11) or real-time environmental reliability monitors (ERMs) such as the OnGuard® Smart (Fig. 15.12) to monitor the copper and silver corrosion rates. Copper and silver coupons and their use are described in in a white paper published by ASHRAE Technical Committee 9.9 titled "2011 Gaseous and Particulate Contamination Guideline for Data Centers".

CCCs can be placed upstream and downstream of the gas-phase air filters to gauge the efficiency of the systems for reducing total corrosion and against individual corrosion species and to determine when filter replacement is required. They can also be placed throughout the data center to provide ongoing verification of environmental specifications. ERMs can be placed in the controlled environment and on or in server cabinets to provide real-time data on corrosion rates and the effectiveness of various gaseous contaminant control strategies—whether they involve the use of gas-phase air filtration or not (Fig. 15.12).

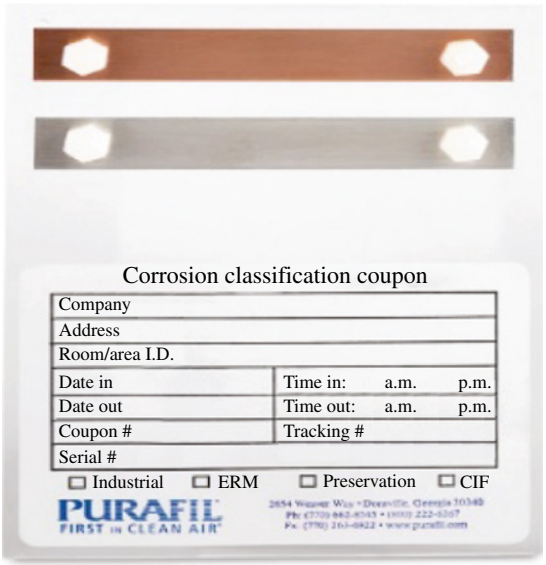


FIGURE 15.11 CCC. Source: Courtesy of Purafil, Inc.

FIGURE 15.12 ERM. Source: Courtesy of Purafil, Inc.

While proper operation and maintenance of the particulate and gas-phase filtration system may require monitoring at various locations within and outside the data center, ASHRAE specifically recommends monitoring at the primary locations of concern, which are in front of the computer racks at one-quarter and three-quarter height of the floor.

15.7 DESIGN/APPLICATION OF DATA CENTER AIR CLEANING

15.7.1 Basic Data Center Designs

There are no universal specifications for data center design. However, for the purposes of this design guide, data centers can be categorized into three basic types: closed systems, ventilation air without no pressurization, and ventilation air with pressurization. A brief discussion of each type will follow with specific details pertaining to the application and use of enhanced air cleaning for chemical contamination control.

Datacom equipment rooms can be conditioned with a wide variety of systems, including packaged CRAC units and central station air-handling systems. Air-handling and refrigeration equipment may be located either inside or outside datacom equipment rooms. A common ventilation scheme uses CRAC units set on a raised-access floor with vents in the ceiling or top parts of the walls for exhaust hot-air removal.

15.7.1.1 Closed Systems

The HVAC systems for many small data centers and server rooms are designed to be operated as 100% recirculation systems meaning that there is no outside ventilation (or makeup) air being delivered to the space. All the air is continuously recirculated (typically) through CRAC units or

FIGURE 15.13 ALNF filter installed in a CRAC unit. Source: Courtesy of Purafil, Inc.

FIGURE 15.14 RAU installed in a data center room. Source: Courtesy of Purafil, Inc.

other types of precision air conditioning designed to protect the datacom equipment, not people. ALNF filters that are available in various sizes or 2 in or 4 in ECC filters may be added to these systems to provide continuous cleaning of the air (Fig. 15.13). Stand-alone recirculation systems can be used to provide local filtration (Fig. 15.14) if contamination cannot be adequately controlled otherwise.

15.7.1.2 Outside Air: No Pressurization

Outside air is introduced to the data center space for one of the following four reasons: to meet and maintain indoor air quality requirements, to pressurize the space to keep contaminants out, as makeup air for smoke purge, or to conserve energy when outside air conditions are conducive to free cooling.

Some larger datacom facilities use central station air handling units. Specifically, many telecommunications central offices, regardless of size, utilize central systems. With these systems there are opportunities for the use of ECC filters, ALNF filters, or bulk media modules (Fig. 15.15) for primary control of chemical contaminants. Often the size and location of these air handling units will dictate which type of chemical filter can be applied.

Where the data center is not maintained under a positive pressure, installation of ALNF filters or ECC filters in the CRAC units may be required to further reduce contamination (corrosion) levels to within manufacturers' guidelines or specifications. Consideration should be given to providing additional outdoor (ventilation) air to prevent infiltration of contaminants either through the central HVAC system or using positive pressurization units (PPUs). To provide clean pressurization air in locations with high levels of outdoor contaminants, deep-bed bulk media air filtration systems (DBSs) may be required (Fig. 15.16).

FIGURE 15.15 SAS installed on outdoor air intake. Source: Courtesy of Purafil, Inc.

FIGURE 15.16 DBS installed on roof of building to provide clean pressurization air to data center. Source: Courtesy of Purafil, Inc.

15.7.1.3 Outside Air: With Pressurization

Datacom facilities are typically pressurized to prevent infiltration of air and pollutants through the building envelope. An airlock entry is recommended for a datacom equipment room door that opens directly to the outside. Excess pressurization with outdoor air should be avoided, as it makes swinging doors harder to use and wastes energy through increased fan energy and coil loads. Variable-speed outdoor air systems, controlled by differential pressure controllers, can ramp up to minimize infiltration and should be part of the HVAC control system.

In systems using CRAC units, it may be advantageous to introduce outside air through a dedicated HVAC system serving all areas. This dedicated system will often provide pressurization control and control the humidity in the datacom equipment room based on dew point, allowing the primary system serving the space to provide sensible-only cooling.

Chemical contaminants can be removed from the outdoor air using ECC filters, ALNF filters, or bulk media modules may be used in the makeup air handlers. Additional air cleaning (as necessary) can be applied in the CRAC units and/or with individual recirculation units. Most often chemical filtration in the makeup air handlers and the CRAC units is required for optimum contamination control. If additional amounts of outside air are required to maintain adequate pressurization, the use of PPUs or DBSs may be considered.

15.7.2 Contamination Control Process Flow

Although each data center will be different in terms of construction, layout, size, HVAC system components, etc., the fundamental steps one takes to establish the risk potential of the data center environment toward the datacom equipment in use are relatively straightforward.

Assess data center environment with CCCs and/or ERMs. Establish the severity levels for copper and silver according to ANSI/ISA-71.04-2013. This is a primary requirement for all applications.

Confirm that basic design requirements have been addressed. Assure that temperature and humidity are within design specifications. Air leaks into room are sealed: adequate sealing for doors, windows, walls, ceilings, and floors. Internal sources of contamination have been eliminated.

Install ALNF filters or ECC filters in CRAC units. These will be used to replace the existing particulate filters. The number of filters and their exact sizes (not nominal dimension) will be required as most are nonstandard sizes proprietary to the CRAC manufacturer. A combination 30% (MERV 6–8) (4 in, 100 mm) ALNF filter will be typically used. If ECC filters are used, use a 2 in deep filter in tandem with a 2 in (50 mm) particulate filter on the downstream side to maintain the required level of particulate filtration.

If outdoor air is being delivered to the data center, ***add gas-phase air filters to the makeup (fresh) air units*** if possible or install a (powered) air filtration system. ECC filters, ALNF filters, or bulk media modules may be used depending on the system design, available pressure drop, and size/weight restrictions.

If additional outdoor air is required for pressurization, PPUs or powered air filtration systems can be used depending on the layout of data center and any mechanical/equipment rooms. If contaminant levels are extremely high, ISA Class GX, or if active sulfur and/or inorganic chlorine are present, bulk-fill PPUs, SASs, or deep-bed scrubbers may be considered.

If data center has no outdoor air or if cleaning the outdoor is not possible, ***install stand-alone air cleaning units to provide clean recirculation air***. It is preferable to apply several smaller units as opposed to fewer larger units to provide for more complete coverage of the data center and assure clean air is being distributed throughout. Bulk-fill recirculation units may be used for high contaminant levels.

It is very important that after each step of the contamination control process, ***the data center environment must be assessed using CCCs or ERMs*** to assure that contaminants are being reduced and controlled and that corrosion rates are below the specifications established for that particular data center. In most cases this will be the ISA Class G1 level of <300 Å/30 days for copper and <200 Å/30 days silver corrosion as described in the ASHRAE guidelines and in manufacturers' warrant requirements. Ongoing monitoring is essential to guarantee the risk to electronic equipment has been eliminated.

A flow chart for the process of monitoring and controlling corrosive chemical contaminants in a data center is provided in Section 15.10.3

NOTE: Even with the steps taken above, it is still possible to experience corrosion-related failures especially if the environmental conditions in the data center were particularly

severe (ISA Class G3, GX) prior to any corrective actions. Providing a benign environment for electronic equipment (ISA Class G1) will not reverse any corrosive damage that may have already occurred but will serve to extend the life of those devices that would have failed without these actions being taken.

15.8 SUMMARY AND CONCLUSION

Electronic equipment used in data centers are protected against the potential threats posed by fire, power, shock, humidity, temperature, and (to a degree) particulate contamination. Unfortunately, the potential damage to this equipment caused by the corrosive effects of gaseous contaminants has still not been fully recognized or addressed.

Recognizing the severity of the problem, the world's leading manufacturers of computer systems jointly published a white paper titled "2011 Gaseous and Particulate Contamination Guideline for Data Centers". That, along with the ASHRAE Handbook on "Particulate and Gaseous Contamination in Datacom Environments" summarizes the acceptable levels of data center contamination as shown in Table 15.4.

Enhanced air filtration systems for data centers should be designed to the ASHRAE guidelines as described or to the IT/datacom equipment manufacturers' site planning guides or environmental specifications. The length of time that these levels of cleanliness can be provided is a function of the total contaminant load, air cleaning system design, and filters/equipment employed. Factors that may cause the data center environment to exceed these classifications include critical space(s) not being properly sealed, high pedestrian traffic, high levels of internally generated contaminants, the air conditioning system and/or negative pressure ductwork being located outside the protected space, and air filtration system being undersized. In properly sealing a room, often the spaces beneath raised floors and/or the space above dropped ceilings are neglected. These areas are critical, especially when they are used as supply and return plenums. Long forgotten cable penetrations, floor drains, cracks in the walls and ceiling, etc., can cause infiltration of contaminated air. The solution is to seal all penetrations, supply an adequate amount and distribution of supply and recirculation air, provide clean outdoor air to achieve positive pressure, and balance the system.

With the increasing pressure to reduce energy consumption in data centers, and the increasing use of air-side economizers, data centers located in regions with poor ambient air quality will struggle to maintain efficient operations without the application of enhanced air cleaning. This means increasing particulate filter efficiencies to at least ASHRAE 85% and the addition of gas-phase air filtration designed to control specific contaminants of concern.

The issue and potential problem of corrosion in data centers has been presented (ASHRAE 2009a;). The problem needs to be addressed by monitoring of the environment and removal of contaminants where needed. Ultimately, the successful implementation of a contamination control program requires:

1. Knowledge and understanding that corrosion of electronic equipment is a serious problem.
2. Commitment to a monitoring program to detect the potential for IT equipment failure before this equipment is damaged and costly shutdowns occur.
3. Commitment to an integrated contamination control system.
4. Commitment to take corrective action whenever necessary.

TABLE 15.4 Particulate and gaseous contamination guidelines for data centers

Data centers must be kept clean to ISO 14644-1 Class 8. This level of cleanliness can generally be achieved by an appropriate filtration scheme as outlined here:

1. The room air may be continuously filtered with MERV 8 filters as recommended by ANSI/ASHRAE Standard 127-2012, Method of Testing for Rating Computer and Data Processing Room Unitary Air Conditioners (ASHRAE 2012).
2. Air entering a data center may be filtered with MERV 11 or MERV 13 filters as recommended by ASHRAE handbook titled "Particulate and Gaseous Contamination in Datacom Environments".

Sources of dust inside data centers should be reduced. Every effort should be made to filter out dust that has deliquescent relative humidity greater than the maximum allowable relative humidity in the data center.

The gaseous contamination should be within the ANSI/ISA-71.04-2013 severity level G1-Mild that meets:

1. A copper reactivity rate of less than 300 Å/month, and
2. A silver reactivity rate of less than 200 Å/month.

For data centers with higher gaseous contamination levels, gas-phase filtration of the inlet air and the air in the data center is highly recommended.

Source: Christpher O. Muller.

15.9 APPENDIX 1: ADDITIONAL DATA CENTER SERVICES

Many companies can, upon the request of the data center owner/operator, provide any or all the following services. These services should be considered if air quality/equipment reliability problems are discovered and/or are ongoing in the data center environment.

15.9.1 Assessment

- Perform data center environment survey with CCCs, ERMs to determine compliance with ISA Standard 71.04-2013. A detailed analysis and interpretation of all data obtained should be provided along with recommendations for reducing corrosion rates (if warranted).
- Perform system TAB (Test, Adjust, Balance) by checking and adjusting all data center environmental systems to produce the design objectives. Testing to determine the quantitative performance of air handling equipment, adjustment of equipment to proportion airflows to specified design quantities, and balancing to regulate the specified airflow rate at the terminal equipment.
- Measure airflow through raised access floor along the length of the cold and hot aisles. Repair access flooring as required.
- Install room pressure sensor and monitor to verify minimum pressure differential of +0.05–0.10 iwg (12.5–25.0 Pa) is being maintained inside the data center.

15.9.2 Control

- Seal around all doors leading into/out of the data center. This includes adding sweeps at the bottom of the doors, gasketing around the doorframe, etc. Put automatic door closers on any doors.
- Seal around all wall penetrations including areas are under raised floors or above drop ceilings. Seal around windows.
- Install as needed and maintain chemical filtration for CRAC/CRAH units.
- Install as needed and maintain chemical filtration for makeup (outside) air handlers, air-side economizers, RAHs, pressurization air handlers, and/or raised floor systems.

15.9.3 Testing

- Continuous monitoring of the data center environment with ERMs and/or CCCs and certification of ISA Severity Level.
- Perform airborne particle counts and certify to ISO class.
- Develop a temperature and humidity profile for the data center.

All work performed would be typically performed as part of a fee-based service contract issued by the provider. All warrantees and guarantees for the work performed would be based on negotiations between the service provider and the data center owner/operator.

15.10 APPENDIX 2: DATA CENTER HISTORY

Data centers have their roots in the huge computer rooms of the early ages of the computing industry. Early computer systems were complex to operate and maintain and required a special environment in which to operate. Many cables were necessary to connect all the components and methods to accommodate and organize these were devised, such as standard racks to mount equipment, elevated floors, and cable trays (installed overhead or under the elevated floor). Also, old computers required a great deal of power, and had to be cooled to avoid overheating. Security was important, computers were expensive and were often used for military purposes. Basic design guidelines for controlling access to the computer room were therefore devised.

During the boom of the microcomputer industry, and especially during the 1980s, computers started to be deployed everywhere, in many cases with little or no care about operating requirements. However, as information technology (IT) operations started to grow in complexity, companies grew aware of the need to control IT resources. With the advent of client-server computing, during the 1990s, microcomputers (now called "servers") started to find their places in the old computer rooms. The availability of inexpensive networking equipment, coupled with new standards for network cabling, made it possible to use a hierarchical design that put the servers in a specific room inside the company. The use of the term "data center," as applied to specially designed computer rooms, started to gain popular recognition about this time.

The boom of data centers came during the dot-com bubble. Companies needed fast Internet connectivity and nonstop operation to deploy systems and establish a presence on the Internet. Installing such equipment was not viable for many smaller companies. Many companies started building very large facilities, called Internet data centers (IDCs), which provide businesses with a range of solutions for systems deployment and operation. New technologies and practices were designed to handle the scale and the operational requirements of such large-scale operations. These practices eventually migrated toward the private data centers and were adopted largely because of their practical results.

Today, data center design, construction, and operation are well-known disciplines. Standard documents from accredited professional groups, such as the Telecommunications Industry Association (TIA), specify the requirements for data center design. Well-known operational metrics for data center availability can be used to evaluate the business impact of a disruption. There is still a lot of development being done in operation practice, and in environmentally friendly data center design. Data centers are typically expensive to build and maintain.

15.10.1 Requirements for Modern Data Centers

IT operations are a crucial aspect of most organizational operations. One of the main concerns is business continuity; companies rely on their information systems to run their operations. If a system becomes unavailable, company operations may be impaired or stopped completely. It is necessary to provide a reliable infrastructure for IT operations, in order to minimize any chance of disruption. Information security is also a concern, and for this reason a data center has to offer a secure environment which minimizes the chances of a security breach. A data center must therefore keep high standards for assuring the integrity and functionality of its hosted computer environment. This is accomplished through redundancy of both fiber optic cables and power, which includes emergency backup power generation.

Telcordia GR-3160, *NEBS Requirements for Telecommunications Data Center Equipment and Spaces*,[3] provides guidelines for data center spaces within telecommunications networks, and environmental requirements for the equipment intended for installation in those spaces. These criteria were developed jointly by Telcordia and industry representatives. They may be applied to data center spaces housing data processing or IT. The equipment may be used to:

- Operate and manage a carrier's telecommunication network
- Provide data center-based applications directly to the carrier's customers
- Provide hosted applications for a third party to provide services to their customers
- Provide a combination of these and similar data center applications.

Effective data center operation requires a balanced investment in both the facility and the housed equipment. The first step is to establish a baseline facility environment suitable for equipment installation. Standardization and modularity can yield savings and efficiencies in the design and construction of telecommunications data centers.

Standardization means integrated building and equipment engineering. Modularity has the benefits of scalability and easier growth, even when planning forecasts are less than optimal. For these reasons, telecommunications data centers should be planned in repetitive building blocks of equipment, and associated power and support (conditioning) equipment when practical. The use of dedicated centralized systems requires more accurate forecasts of future needs to prevent expensive over construction, or perhaps worse—under construction that fails to meet future needs.

[3] http://telecom-info.telcordia.com/site-cgi/ido/docs.cgi?ID=SEARCH&DOCUMENT=GR-3160&.

15.10.2 Data Center Classification

The TIA-942: Data Center Standards Overview[4] describes the requirements for the data center infrastructure. The simplest is a Rated 1 data center, which is basically a server room,[5] following basic guidelines for the installation of computer systems. The most stringent level is a Rated 4 data center, which is designed to host mission critical computer systems, with fully redundant subsystems and compartmentalized security zones controlled by biometric access controls methods. Another consideration is the placement of the data center in a subterranean context, for data security as well as environmental considerations such as cooling requirements. The rated levels describe the availability of data from the hardware at a location—the higher the rating the greater the accessibility. The levels are:

Rated level	Requirements
1	• Single nonredundant distribution path serving the IT equipment • Nonredundant capacity components • Basic site infrastructure guaranteeing 99.671% availability
2	• Fulfills all Rated 1 requirements • Redundant site infrastructure capacity components guaranteeing 99.741% availability
3	• Fulfills all Rated 1 and Tier 2 requirements • Multiple independent distribution paths serving the IT equipment • All IT equipment must be dual-powered and fully compatible with the topology of a site's architecture • Concurrently maintainable site infrastructure guaranteeing 99.982% availability
4	• Fulfills all Rated 1, 2, and 3 requirements • All cooling equipment is independently dual-powered, including chillers and Heating, Ventilating and Air Conditioning (HVAC) systems • Fault-tolerant site infrastructure with electrical power storage and distribution facilities guaranteeing 99.995% availability

15.10.3 Physical Layout

A data center can occupy one room of a building, one or more floors, or an entire building. Most of the equipment is often in the form of servers mounted in 19-in rack cabinets, which are usually placed in single rows forming corridors between them. This allows people access to the front and

[4] http://www.adc.com/Attachment/1270711929361/102264AE.pdf.

[5] A server room is a room that houses mainly computer servers. In IT circles, the term is generally used for smaller arrangements of servers; larger groups of servers are housed in data centers.

rear of each cabinet. Servers differ greatly in size from 1U servers[6] to large freestanding storage silos which occupy many tiles on the floor. Some equipment such as mainframe computers and storage devices are often as big as the racks themselves and are placed alongside them. Very large data centers may use shipping containers packed with 1,000 or more servers each; when repairs or upgrades are needed, whole containers are replaced (rather than repairing individual servers).

Local building codes may govern the minimum ceiling heights.

The physical environment of a data center is rigorously controlled:

- For earlier thermal guidelines, the purpose of the recommended envelope was to give guidance to data center operators on maintaining high reliability and operating their data centers in the most energy-efficient manner. This envelope was created for general use across all types of businesses and conditions. However, different environmental envelopes may be more appropriate for different business values and climate conditions. Therefore, to allow for the potential to operate in a different envelope that might provide even greater energy savings, ASHRAE's third whitepaper on thermal guidelines[7] provided general guidance on server metrics that will assist data center operators in creating a different operating envelope that matches their business values. Each of these metrics is fully described, with more details to be provided in the fourth edition of the "Thermal Guidelines for Data Processing Environments"[8] Datacom Book. Any choice outside of the recommended region will be a balance between the additional energy savings of the cooling system versus the deleterious effects that may be created in reliability, acoustics, or performance.

  A flow chart (Fig. 15.17) is provided to help guide the user through the appropriate evaluation steps. Many of these server metrics center around simple graphs that describe the trends. However, the use of these metrics is intended for those that plan to go beyond the recommended envelope for additional energy savings. To do this properly requires significant additional analysis in each of the metric areas to understand the total cost-of-ownership impact of operating beyond the recommended envelope.

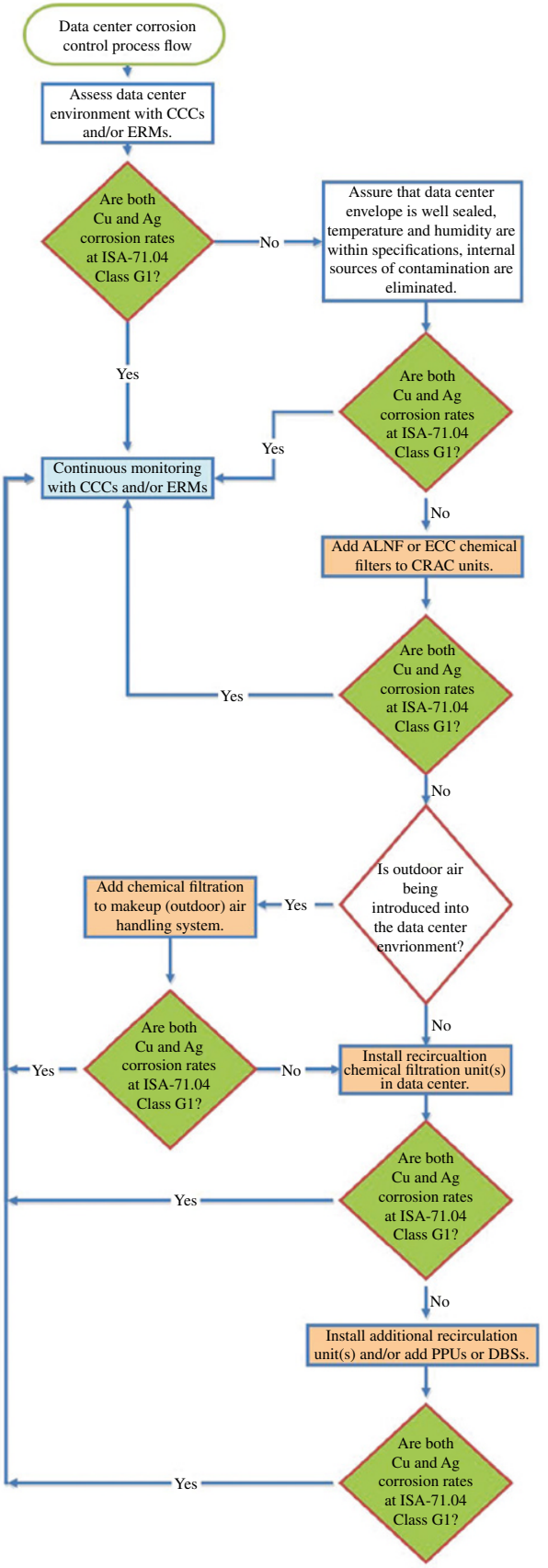


FIGURE 15.17 Data center corrosion control process flow.

[6] The size of a piece of rack-mounted equipment is frequently described as a number in "U." For example, one rack unit is often referred to as "1U", two rack units as "2U" and so on.

[7] http://tc99.ashraetcs.org/documents/ASHRAE%20Whitepaper%20-%202011%20Thermal%20Guidelines%20for%20Data%20Processing%20Environments.pdf.

[8] https://www.techstreet.com/ashrae/standards/thermal-guidelines-for-data-processing-environments-4th-ed?product_id=1909403.

The intent of outlining the process herein is to demonstrate a methodology and provide general guidance. This paper contains generic server equipment metrics and does not necessarily represent the characteristics of any particular piece of IT equipment. For specific equipment information, contact the IT manufacturer.

The other major change in the environmental specification is in the data center classes. Previously there were two classes applying to IT equipment used in data center applications: Classes 1 and 2. The new environmental guidelines have more data center classes to accommodate different applications and priorities of IT equipment operation. This is critical because a single data center class forces a single optimization whereas each data center needs to be optimized based on the operator's own criteria (e.g., fulltime economizer use versus maximum reliability).

- Today's data centers try to use economizer cooling, where outside air is used to keep the data center cool. Washington State now has a few data centers that cool all the servers using outside air 11 months out of the year. They do not use chillers/air conditioners, which creates potential energy savings in the millions.
- Backup power consists of one or more uninterruptible power supplies and/or diesel generators.
- To prevent single points of failure, all elements of the electrical systems, including backup system, are typically fully duplicated, and critical servers are connected to both the "A-side" and "B-side" power feeds. This arrangement is often made to achieve $N+1$ Redundancy in the systems. Static switches are sometimes used to ensure instantaneous switchover from one supply to the other in the event of a power failure.
- Data centers typically have raised flooring made up of 2×2 ft (60×60 cm) removable square tiles. The trend is toward 31–39 in^2 (80–100 cm^2) void to cater for better and uniform air distribution. These provide a plenum for air to circulate below the floor, as part of the air conditioning system, as well as providing space for power cabling.

  Telcordia GR-2930, *NEBS: Raised Floor Generic Requirements for Network and Data Centers*,[9] presents generic engineering requirements for raised floors that fall within the strict NEBS guidelines. Data cabling is typically routed through overhead cable trays in modern data centers. But some are still recommending under raised floor cabling for security reasons and to consider the addition of cooling systems above the racks in case this enhancement is necessary. Smaller/less expensive data centers without raised flooring may use antistatic tiles for a flooring surface. Computer cabinets are often organized into a hot aisle arrangement to maximize airflow efficiency.

[9] https://telecom-info.telcordia.com/site-cgi/ido/docs.cgi?ID=SEARCH&DOCUMENT=GR-2930&.

15.10.4 Applications

The main purpose of a data center is running the applications that handle the core business and operational data of the organization. Such systems may be proprietary and developed internally by the organization or bought from enterprise software vendors. Such common applications are ERP and CRM systems.

A data center may be concerned with just operations architecture or it may provide other services as well.

Often these applications will be composed of multiple hosts, each running a single component. Common components of such applications are databases, file servers, application servers, middleware, and various others.

Data centers are also used for offsite backups. Companies may subscribe to backup services provided by a data center. This is often used in conjunction with backup tapes. Backups can be taken of servers locally on to tapes; however, tapes stored on site pose a security threat and are also susceptible to fire and flooding. Larger companies may also send their backups off site for added security. This can be done by backing up to a data center. Encrypted backups can be sent over the Internet to another data center where they can be stored securely.

For disaster recovery, several large hardware vendors have developed mobile solutions that can be installed and made operational in very short time. Vendors such as Cisco Systems, Sun Microsystems, IBM, and HP have developed systems that could be used for this purpose.

15.11 APPENDIX 3: REACTIVITY MONITORING DATA EXAMPLES: SAMPLE CORROSION MONITORING REPORT

World Data Center, Inc.
Corrosion Monitoring Report
Report for data collected Date 1 – Date 2

15.11.1 Executive Summary

Seven copper/silver CCCs were placed at World Data Center, Inc. (Anytown, USA) to provide an assessment of the air

outside and inside the data center. All CCCs were analyzed via electrolytic reduction to identify and quantify corrosive contaminants to which the coupons had been exposed. The electrolytic reduction analysis shows the presence of high contaminant concentrations.

Analysis results indicate that the air outside the data center would be classified as *GX – SEVERE* according to *ASHRAE TC 9.9 ANSI / ISA Standard S71.04* which significantly exceeds the recommended severity level of *G1 – MILD*. The air inside the data center would be classified, as *G3 – HARSH* which, on average, also exceeds the recommended severity level for these environments. The presence of oxidized forms of sulfur and active sulfur compounds was detected, and *it is estimated that the concentrations in air would be 10–100 parts per billion (ppb) and >10 ppb, respectively, for these two sulfur species*. It is also suspected that there are significant levels of nitrogen oxides and ozone present in the ambient (outdoor) environment.

Oxidized forms of sulfur include sulfur dioxide (SO_2) and sulfur trioxide (SO_3). Active sulfur compounds include elemental sulfur, hydrogen sulfide (H_2S), organic sulfur compounds (e.g., mercaptans), as well as sulfuric and sulfurous acids. These contaminants will cause corrosion-related problems should adequate control measures not be put in place.

The active sulfur contamination (as Cu_2S) detected on all CCCs placed both outside and inside the data center indicates a SEVERE risk potential if steps are not taken to reduce and maintain lower levels of these contaminants in the data center.

It is recommended that reactivity monitoring be continued—either with CCCs or a real-time environmental reliability monitor (ERM)—to provide a continuous environmental assessment of the data center air quality and to assure that chemical contamination is being maintained at acceptable levels.

15.11.2 Background

Corrosion of metals is a chemical reaction caused primarily by attack of gaseous contaminants and is accelerated by heat and moisture. Rapid shifts in either temperature or humidity cause small portions of circuits to fall below the dew point temperature, thereby facilitating condensation of contaminants. Relative humidity above 50% accelerates corrosion by forming conductive solutions on a small scale on electronic components. Microscopic pools of condensation then absorb contaminant gases to become electrolytes where crystal growth and electroplating occur. Above 80% RH, electronic corrosive damage will occur regardless of the levels of contamination.

In the context of electronic equipment, corrosion is defined as the deterioration of a base metal resulting from a reaction with its environment. More specifically, corrosive gases and water vapor coming into contact with a base metal result in the buildup of various chemical reaction products. As the chemical reactions continue, these corrosion products can form insulating layers on circuits which can lead to thermal failure or short-circuits. Pitting and metal loss can also occur.

15.11.2.1 Corrosive Gases

Three types of gases are the prime suspects in the corrosion of electronics: acidic gases, such as hydrogen sulfide, sulfur and nitrogen oxides, chlorine, and hydrogen fluoride; caustic gases, such as ammonia; and oxidizing gases, such as ozone. Of the gases that can cause corrosion, the acidic gases are typically the most harmful.

Each site may have different combinations and concentration levels of corrosive gaseous contaminants and performance degradation can occur rapidly or over many years, depending on the concentration levels and combinations present at a site. The following paragraphs describe how various pollutants contribute to equipment performance degradation.

Active sulfur compounds (H_2S). This group includes hydrogen sulfide (H_2S), elemental sulfur (S), and organic sulfur compounds such as the mercaptans (RSH). When present at low parts per billion levels, they rapidly attack copper, silver, aluminum, and iron alloys. The presence of moisture and small amounts of inorganic chlorine compounds and/or nitrogen oxides greatly accelerate sulfide corrosion. Note, however, that attack still occurs in low relative humidity environments. Active sulfurs rank as one of the predominant causes of atmospheric corrosion in datacom equipment.

Sulfur oxides. Oxidized forms of sulfur (SO_2, SO_3) are generated as combustion products of sulfur-bearing fossil fuels. Low parts per billion levels of sulfur oxides can passivate reactive metals and thus retard corrosion. At higher levels, however, they will attack certain types of metals. The reaction with metals normally occurs when these gases dissolve in water to form sulfurous and sulfuric acid.

Nitrogen oxides (NO_X). NO_X compounds (NO, NO_2, N_2O_4) are formed as combustion products of fossil fuels and have a critical role in the formation of ozone in the atmosphere. They are also believed to have a catalytic effect on corrosion of base metals by chlorides and sulfides. In the presence of moisture, some of these gases form nitric acid that, in turn, attacks most common metals.

Inorganic chlorine compounds. This group includes chlorine (Cl_2), chlorine dioxide (ClO_2), hydrogen chloride

(HCl), etc., and reactivity will depend upon the specific gas composition. In the presence of moisture, these gases generate chloride ions that react readily with copper, tin, silver, and iron alloys. These reactions are significant even when the gases are present at low parts per billion levels. At higher concentrations, many materials are oxidized by exposure to chlorinated gases. Particular care must be given to equipment that is exposed to atmospheres which contain chlorinated contaminants. Sources of chloride ions, such as bleaching operations, sea water, cooling tower vapors, and cleaning compounds, etc., should be considered when classifying data center environments.

Hydrogen fluoride (HF). This compound is a member of the halogen family and reacts like inorganic chloride compounds.

Ammonia and derivatives. Reduced forms of nitrogen (ammonia (NH_3), amines, ammonium ions (NH_4^+)) occur mainly in fertilizer plants, agricultural applications, and chemical plants. Copper and copper alloys are particularly susceptible to corrosion in ammonia environments.

Photochemical species. The atmosphere contains a wide variety of unstable, reactive species that are formed by the reaction of sunlight with moisture and other atmospheric constituents. Some have lifetimes measured in fractions of a second as they participate in rapid chain reactions. In addition to ozone (O_3), a list of examples would include the hydroxyl radical as well as radicals of hydrocarbons, oxygenated hydrocarbons, nitrogen oxides, sulfur oxides, and water. Because of the transient nature of most of these species, their primary effect is on outdoor installations and enclosures.

Strong oxidants. This includes ozone plus certain chlorinated gases (chlorine, chlorine dioxide). These gases are powerful oxidizing agents. Ozone can function as a catalyst in sulfide and chloride corrosion of metals. Photochemical oxidation—the combined effect of oxidants and ultraviolet light (sunlight)—is particularly potent.

Hydrogen sulfide (H_2S), sulfur dioxide (SO_2), and active chlorine compounds (Cl_2, HCl, ClO_2) have all been shown to cause significant corrosion in electrical and electronic equipment at concentrations of just a few parts per billion in air. Even at levels that are not noticed by or harmful to humans, these gases can be deadly to electronic equipment. Most of the odor threshold levels are much higher than the levels at which corrosive damage will occur.

15.11.2.2 Environmental Classifications

Table 15.5 lists a standard classification scheme that directly correlates corrosion rates to environmental classifications. These classifications are being refined based on the results of testing and the specific needs of this market. Typical uses of reactivity monitoring in data centers have been for the characterization of outdoor air used for ventilation and pressurization, the identification of "hot spots" within a facility, and the effectiveness of various preventive measures. Reactivity monitoring is being used for the purpose of developing the cause-and-effect relationship between gaseous pollutants and the damage it may cause within data centers to sensitive electronic equipment.

Generally speaking, the silver and copper corrosion rates should be class G1 or better unless otherwise agreed upon. The individual corrosion films quantified using reactivity monitoring may be used to further characterize the environment and to determine the proper control strategies. Based upon these recommended control levels and test results from laboratory and field-exposed silver coupons, acceptance criteria relevant to these applications have been determined.

These criteria consider total corrosion as well as the relative contribution of each individual corrosion film.

TABLE 15.5 Classification of reactive environments

Severity level	G1 Mild	G2 Moderate	G3 Harsh	GX Severe
Copper reactivity level (in angstroms)[a]	<300	<1,000	<2,000	≥2,000
Silver reactivity level (in angstroms)[a]	<200	<1,000	<2,000	≥2,000

[a] Normalized to a 30-day exposure. See ANSI/ISA-71.04-2013 Annex C, item numbers C.2, C.3.

TABLE 15.6 General reactivity monitoring acceptance criteria

Copper reactivity acceptance criteria		Silver reactivity acceptance criteria	
Copper corrosion reaction products	Corrosion film thickness	Silver corrosion reaction products	Corrosion film thickness
Copper sulfide, Cu_2S	***0 Å/30 days***	***Silver chloride, AgCl***	***0 Å/30 days***
Copper oxide, Cu_2O	<300 Å/30 days	Silver sulfide, Ag_2S	<200 Å/30 days
Copper unknowns	0 Å/30 days	Silver unknowns	0 Å/30 days
Total copper corrosion	<300 Å/30 days	Total silver corrosion	<200 Å/30 days

The control specifications for the individual corrosion films are listed in Table 15.6. These specifications are more general in their application than those listed above and are most often used for the characterization of an environment prior to the implementation of pollutant control measures.

If the total corrosion *AND* each individual corrosion film meets the recommended criteria, the local environment in which that particular coupon has been exposed meets the requirements of a Class G1 classification. *ANY* of the criteria which are not met indicates that the local environment may not be sufficiently well-controlled to minimize the corrosion of sensitive electronic equipment due to the presence of gaseous pollutants. Steps should be taken to determine what problems exist and what corrective actions may be appropriate.

15.11.3 Results and Discussion

As shown in Table 15.7 the total copper and silver corrosion rates for all the CCCs placed in the data center at World Data Center, Inc. (Anytown, USA) exceed the generally recommended severity level of Class G1.

When interpreting the analysis results for the individual corrosion films, the detection of a silver sulfide (Ag_2S) film without a corresponding copper sulfide (Cu_2S) film indicates the presence of oxidized forms of sulfur such as sulfur dioxide (SO_2) and sulfur trioxide (SO_3). All silver coupons showed evidence of sulfur oxide contamination.

Oxidized forms of sulfur are generated as combustion products of sulfur-bearing fossil fuels. Low parts per billion levels of sulfur oxides can passivate reactive metals and thus retard corrosion. At higher levels they attack certain types of metals, elastomers, and plastics. The reaction with metals normally occurs when these gases dissolve in water to form sulfurous and sulfuric acid.

When both films are detected, as highlighted above, it more often than not indicates the presence of active sulfur compounds such as elemental sulfur, hydrogen sulfide (H_2S), organic sulfur compounds (e.g., mercaptans), and sulfuric and sulfurous acids as well. When both films are present and the amount of Cu_2S is greater than 50% of the total corrosion, this is further evidence of the presence of active sulfur compounds in the subject environment.

Active sulfur compounds include hydrogen sulfide, elemental sulfur, and organic sulfur compounds such as the mercaptans. When present at low parts per billion levels, they rapidly attack copper, silver, aluminum, and iron alloys. The presence of moisture and small amounts of inorganic chlorine compounds greatly accelerates sulfide corrosion. Note that corrosive attack still occurs in low relative humidity environments.

Chloride corrosion (AgCl) which was not detected on any of the silver coupons would have indicated the presence of (an) inorganic chlorine compound(s), e.g., chlorine (Cl_2), chlorine dioxide (ClO_2), and/or hydrogen chloride (HCl). High levels of chloride (halogen) contamination can also serve to effectively mask evidence of sulfur contamination on the corresponding copper coupons and can cause a large "unknown" copper corrosion film to appear.

Chlorine contamination, whether as chlorine or hydrogen chloride, is a most dangerous contaminant for metals. At elevated levels, many materials are oxidized by exposure to chlorinated gases.

Based on these results, it is estimated that the active sulfur concentration is >10 ppb and sulfur oxides are in the range of 10–100 ppb. Further, the ratio of Cu_2S to Ag_2S indicates a high likelihood that significant levels of nitrogen oxides are also present in the local environment. Fluctuating humidity may also play a role in the high copper corrosion rates.

TABLE 15.7 CCC monitoring results for World Data Center, Inc. (Anytown, USA)

		Copper corrosion				Silver corrosion					
CCC panel #	Location	Cu_2S	Cu_2O	Cu-unk	Total	AgCl	Ag_2S	Ag-unk	Total	Class	ISA class
60127	Outdoor air	***4,255***	0	0	4,255	0	***2,212***	0	2,212	GX	GX
60142	Office area	***1,937***	294	0	2,231	0	***337***	0	337	G2	GX
60126	Entrance to data center	***934***	112	130	1,176	0	***1,590***	0	1,590	G3	G3
60123	Mux (Maux) room	***2,127***	131	0	2,258	0	***655***	0	655	G2	GX
60125	Low density room	***2,051***	196	0	2,247	0	***1,262***	0	1,262	G3	GX
60128	Medium density room	***1,142***	246	0	1,388	0	***542***	0	542	G2	G3
60124	High density room	***675***	134	0	809	0	***692***	0	692	G2	G2

15.11.4 Conclusions

The amount of corrosion forming over any given period is a primary indicator of how well-controlled an environment may be. Where gas-phase air filtration is used for the control of gaseous pollutants, corrosion levels well within the general and specific acceptance criteria can be easily attained. It is felt that if an environment exhibits reactivity rate of G1 (<300 Å/30 days for each copper and silver corrosion), there is little else that can be done, economically, to improve the environment.

As a minimum, the broad guidelines for applying chemical contamination control in data centers would be in all makeup air units and all recirculation systems serving data storage and data processing areas. Makeup (fresh) air systems must typically be designed to control SOx, NOx, ozone, volatile organic compounds (VOCs), and some site-specific contaminants such as chlorine. Chemical filtration in recirculation systems should be able to remove a wide array of VOCs and acid gases.

The fact that the total amount of corrosion measured on this CCC is significantly higher than what is recommended for this type of application, and that the presence of sulfur oxides and inorganic chlorine compounds has been confirmed indicates that the outdoor (makeup) air should be treated in order to prevent damage from corrosive chemical contaminants and adversely affecting exposed electronic equipment in the data center.

The levels of active sulfur and sulfur oxide contamination (as Cu_2S and Ag_2S, respectively) measured on ALL CCCs exposed at this location indicates a SEVERE risk potential for these types of contamination being introduced and distributed through the data center through the air handling system. Special consideration should be given for continuous monitoring of the air in these spaces.

World Data Center, Inc. should use chemical filtration in the fresh (makeup) air units, and the air conditioning systems serving the data center in order to reduce and maintain chemical contamination at acceptable levels. *The detection of sulfur contamination is particularly problematic to electrical and electronic equipment—even at very low levels* and steps should be taken to reduce contaminant levels to those which would not have effects on electronic equipment.

It is further recommended that reactivity monitoring be continued—either with CCCs or a real-time environmental reliability monitor (ERM)—to provide a continuous assessment air quality. While information on individual contaminant species can be obtained using the CCCs, real-time reactivity monitoring can provide a more accurate assessment of the total corrosion being formed due to the presence of chemical contaminants. Reactivity monitoring can also be used to measure the performance of the chemical filter systems (if installed) and serve as a guide for media replacement.

Direct gas monitoring may be indicated in order to determine the source(s) of the sulfur corrosion reported. This could help determine if these results were an anomaly, due to episodic events, or if there are some other gaseous contaminants present which may have synergistic effects and need to be accounted for.

15.12 APPENDIX 4: DATA CENTER CASE STUDY

15.12.1 Sample ERM Monitoring Data

A data center had experienced hardware failures due to corrosion and the owners wanted to improve to prevent future failures. Previous reactivity (corrosion) monitoring using CCCs had indicated elevated levels of sulfur contamination (both as Cu_2S and Ag_2S) in the outdoor ventilation air and in the data center itself (Fig. 15.18). The copper severity level was generally G1; however, the corresponding silver severity level was a high G2 with an average silver corrosion rate almost 4½ times higher than that of the copper. The owners wanted to bring the overall corrosion rate to a G1 level.

After a site visit and room survey, recommendations were made for improvements to the data center envelope (room sealing, pressurization, etc.) and for the addition of enhanced air cleaning systems to control the gaseous contamination responsible for the elevated corrosion rates. A side access chemical filtration system (SAS) was installed to clean the outdoor ventilation air being delivered to the data center and four recirculating air filtration systems (RAUs) were installed inside the data center to provide additional air cleaning along with increased recirculation.

The SAS and RAU systems were installed in December and turned on in January. There was a dramatic and immediate improvement in the 30-day corrosion rates for silver which indicated the sulfur contamination was being removed. Improvements to the data center envelope in April resulting in even lower corrosion rates.

Given the improvements made to the air quality in the data center and with the air cleaning systems in operation, a management decision was made in early June to turn off the SAS to determine if this additional air cleaning was required. Immediate and significant increases were observed in both the copper and silver corrosion rates which led to the SAS being turned back on in July. Corrosion rates began to decline but it was determined by life analysis that the media in both the SAS and RAUs were nearing exhaustion and all media was replaced in August. At the same time, ALNF filters were added to one of the CRAC units. Corrosion rates again dropped considerably, and the data center environment achieved and maintained a mid-to-low G1 for both copper and silver.

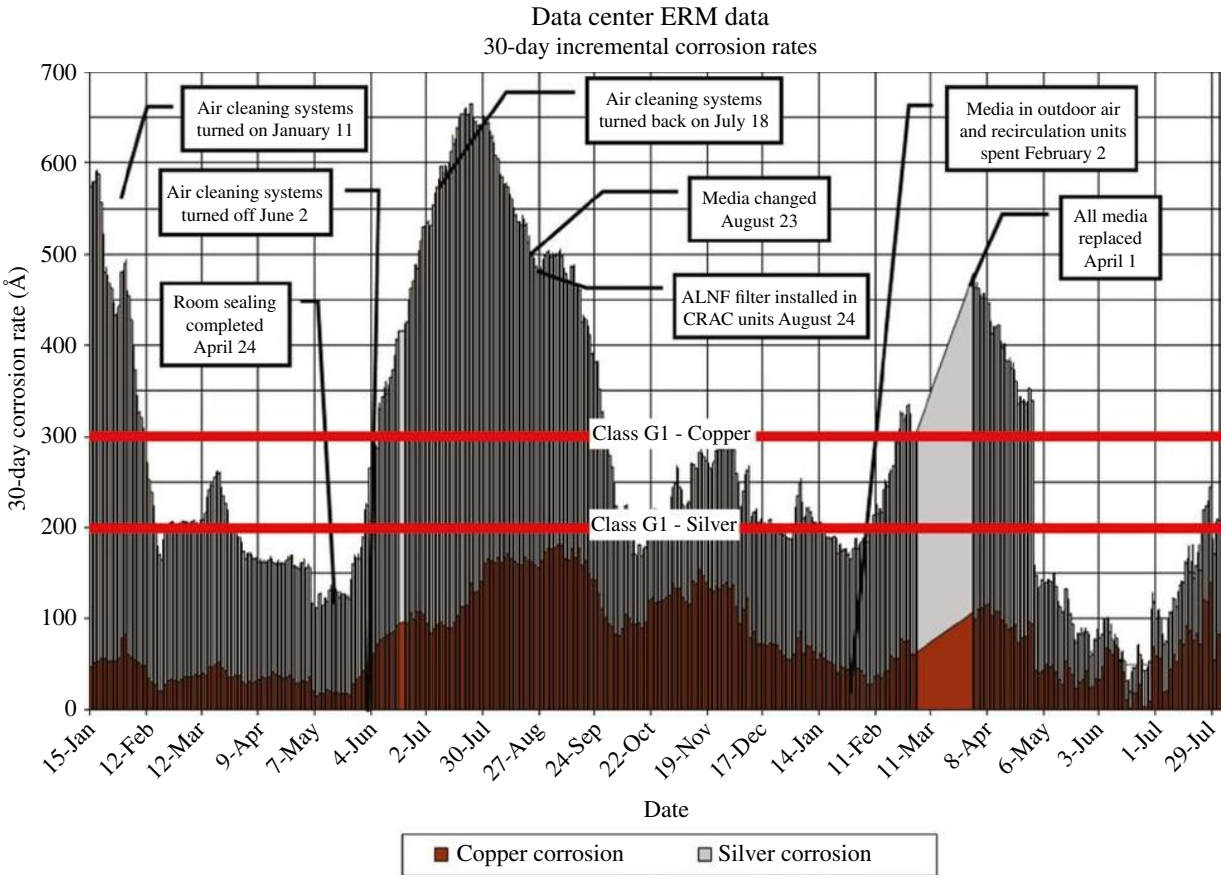


FIGURE 15.18 Data center reactivity monitoring results. Source: Courtesy of Purafil, Inc.

FURTHER READING

DIRECTIVE 2011/65/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 8 June 2011.

ANSI/ISA–71.04–2013 – *Environmental Conditions for Process Measurement and Control Systems: Airborne Contaminants*. Research Triangle Park: International Society for Automation; 2013.

IEC. IEC 60721-3-3 Ed. 3.0 b:2019, *Classification of Environmental Conditions – Part 3-3: Classification of Groups of Environmental Parameters and Their Severities –Stationary Use at Weatherprotected Locations*. Geneva: International Electrotechnical Commission; 2019.

Telcordia. *GR-2930, NEBS: NEBS: Raised Floor Generic Requirements for Network and Data Centers*. Piscataway, NJ: Telcordia Technologies, Inc; 2012.

Muller CO, England WG, Affolder CA. Multiple contaminant gas effects on electronic equipment corrosion: further studies. Adv Instrum Control 1991;46:27–31.

Volpe L, Peterson PJ. Atmospheric sulfidation of silver in a tubular corrosion reactor. Corros Sci 1989;29(10):1179–1196; Volpe L., IBM Corp., Environmental factors in indoor corrosion of metals, IBM Internal Technical Report, 1989.

Johansson L. Laboratory study of the influence of NO2 and combination of NO2 and SO2 on the atmospheric corrosion of different metals. Electrochem Soc 1985;Extended Abstracts, 85(2):221–222.

ASHRAE. *Particulate and Gaseous Contamination in Datacom Environments*. 2nd ed. Atlanta: American Society of Heating, Refrigeration and Air-Conditioning Engineers, Inc; 2011.

Muller CO, Jin L. Clearing the air: advances in affordable filtration for IAQ. Proceedings of ISHVAC 2009, the 6th International Symposium on Heating, Ventilating and Air Conditioning; November 6–9, 2009; Nanjing, China: Southeast University.

Middlebrooks MC, Muller C. Application and evaluation of a new dry-scrubbing chemical filtration media. Proceedings of the Air & Waste Management Association 94th Annual Meeting and Exhibition; June 24–28, 2001; Orlando, FL.

ASHRAE TC9.9 (2016). Data Center Power Equipment Thermal Guidelines and Best Practices. https://tc0909.ashraetcs.org/documents/ASHRAE_TC0909_Power_White_Paper_22_June_2016_REVISED.pdf. Accessed on September 22, 2020.

ISO. ISO 14644-1:2015, *Cleanrooms and Associated Controlled Environments—Part 1: Classification of Air Cleanliness*.

Geneva: International Organization for Standardization; 2015.

U.S. Department of Defense (USDOD). 1995. Method 102.9.1: DOP-smoke penetration and air resistance of filters. Military standard, filter units, protective clothing, gas-mask components and related products: performance test methods. MIL-STD-282.

Singh PJ. Gaseous and particulate contamination limits for data centers. The Air Conditioning and Refrigeration Journal, 2010. Delhi: Indian Society of Heating, Refrigerating and Air Conditioning Engineers (April-June 2010 edition).

Shehabi A, Tschudi W, Gadgil A. 2007. Data center economizer contamination and humidity study. LBNL/Pacific Gas & Electric. Available at https://buildings.lbl.gov/sites/default/files/2424e.pdf. Accessed September 2020.

ASHRAE. *Standard 52.2: Method of Testing General Ventilation Air-Cleaning Devices for Removal Efficiency by Particle Size*. Atlanta: American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc; 2017. https://standards.globalspec.com/std/10148034/ASHRAE%20STD%2052.2. Accessed 9/22/2020.

Muller C. There's no such thing as a free ride: the real costs of free cooling. Proceedings of the Green Data Center Conference and Exhibition; February 24–26, 2015; San Diego, CA.

Shah JM, Oluwaseun A, Agarwal P, Akhigbe I, Agonafer D, Singh P, Kannan N, Kaler M. Qualitative study of cumulative corrosion damage of IT equipment in a data center utilizing air-side economizer. Proceedings of ASME 2016 International Mechanical Engineering Congress and Exposition; November 11–17, 2016; Phoenix, AZ.

England WG, McShane WJ, Muller CO. Developments in measurement and control of corrosive gases to avoid electrical equipment failure. Proceedings of PITA Annual Technical Conference; September 14–16, 1999; Manchester, England.

Muller C. *What's Creeping Around in Your Data Center? ASHRAE Transactions*. Atlanta: American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc.; 2010.

Gaseous and particulate contamination guidelines for data centers. Whitepaper prepared by ASHRAE Technical Committee (TC) 9.9 Mission Critical Facilities, Technology Spaces, and Electronic Equipment. https://www.ashrae.org/File%20Library/Technical%20Resources/Publication%20Errata%20and%20Updates/2011-Gaseous-and-Particulate-Guidelines.pdf. Accessed 9/22/2020.

ISA Standard ANSI/ISA-71.04-2013. *Environmental Conditions for Process Measurement and Control Systems: Airborne Contaminants*. Research Triangle Park, NC: International Society for Automation; 2013.

16

RACK PDU FOR GREEN DATA CENTERS

Ching-I Hsu[1] and Ligong Zhou[2]

[1] *Raritan, Inc., Somerset, New Jersey, United States of America*
[2] *Raritan, Inc., Beijing, China*

16.1 INTRODUCTION

The rack power distribution unit (rack PDU) is emerging from obscurity. As the last link of the elaborate data center power chain, the traditional role of the rack PDU has been to deliver stable, reliable, and adequate power to all the devices in the rack or cabinet—servers, storage, network equipment, etc., which are plugged into it. And while it provides the electrical heartbeat to all the systems that run the critical applications that support the operation of the business (or that, in some cases, are the business), it was often considered a simple commodity—just a power strip. Typically, IT merely told facilities how much power was needed, based on device nameplate specs and often with redundancy, so there was plenty of headroom and minimal risk of downtime. Little thought was given to efficiency or what other value a rack PDU could provide. That was yesterday.

Over the past few years, system availability has become a "given," and now, data center management attention is being focused on operational costs, efficiency improvements, and resource optimization. With the annual expenditure for powering the average data center surpassing the cost to purchase the Information Technology equipment (ITE) itself, the use (and waste) of energy is now targeted as a priority. And beyond the actual cost to power the data center, there are the related issues that impact both current operations and future expansion—for example, physical space and utility power availability, CO_2, footprint, and potential government regulation. Since almost all of the power delivered from the utility to the data center is consumed either directly by the devices plugged into rack PDUs or indirectly by the infrastructure to bring power to the rack and cool the devices, the once obscure rack PDUs have become visible on the data center management radar.

Not surprisingly, many of the major strategies to address the above issues and improve overall data center efficiency depend on new capabilities not available in the commodity outlet strips of a few years ago. To consider a few of these capabilities:

- In order to maximize the use of data center space and other resources, there has been a trend to deploy densely packed 1U servers or power-hungry blade servers. Today's rack PDUs typically handle loads of 5–10 kW with 20 outlets and there are PDUs now designed to support 20+ kW and 36 or more outlets.
- To increase IT staff productivity and conserve energy by employing lights-out and/or remote data center operation, some rack PDUs provide real-time monitoring, reporting and alerts, as well as secure, reliable outlet switching.
- To identify ghost (no function), underutilized, or grossly inefficient servers for elimination, replacement, consolidation, or virtualization, rack PDUs provide individual outlet monitoring.
- To create individual awareness, accountability, and/or charge-back for power usage and CO footprint, some rack PDUs are equipped with highly accurate, real-time power measurement capability at the PDU and outlet levels.
- To optimize IT workload and make informed decisions for infrastructure capacity planning, IT and facilities managers need rack PDU management software that continually collects data on power consumption, analyzes trends, and correlates with IT workload data.

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

These are but a few of the reasons that the selection of rack PDUs has become important.

A wide variety of rack PDU configurations is available based on parameters such as the number of phases, voltage, total ampere, branch circuits, number of outlets, socket type, plug type, rack units consumed, and physical dimensions. Beyond the functions of the basic rack PDU, additional capabilities are available in rack PDU categories, or types, we call metered, switched, and intelligent (Fig. 16.1). Furthermore, if you cannot find an off-the-shelf rack PDU that matches your specific requirement, some vendors will assemble or even design a custom rack PDU, also called BTO/ETO: built-to-order/engineered-to-order.

16.2 FUNDAMENTALS AND PRINCIPLES

ITE is normally mounted in racks or cabinets with provisions for all necessary cables, ventilation, cooling, and convenient access. Previous chapters of this handbook have discussed the large data center PDUs that are used earlier in the power chain and take the form of panel boards mounted on walls or freestanding pedestals. In this chapter, we're discussing only the rack PDU, at the end of the chain, which supplies power to the ITE in the rack. *Unless otherwise stated, any reference to "PDU" for the remainder of this chapter means "rack PDU."*

	Non-intelligent PDUs		Intelligent PDUs			
Primary features	**Basic**	**Monitored**	**Metered input**	**Metered output**	**Switched**	**Switched and outlet metered**
Power distribution	●	●	●	●	●	●
Input metering	○	●	●	●	●	●
Outlet metering	○	○	○	●	○	●
Network connectivity	○	○	●	●	●	●
Switching	○	○	○	○	●	●
Secondary features						
Environment sensor support	○	○	◐	◐	◐	◐
Strong passwords	○	○	◐	◐	◐	◐
Encryption	○	○	◐	◐	◐	◐
User permissions	○	○	◐	◐	◐	◐
Additional ports, e.g. USB	○	○	◐	◐	◐	◐

FIGURE 16.1 Types of rack PDUs. Source: Courtesy of Raritan, Inc.

Rack PDUs come in many configurations with respect to the number and type of receptacles, voltage, load capacity, and physical mounting (horizontal or vertical). A unit may perform no function other than providing power to the devices plugged into it; or it may also provide additional functions—for example, turning power off and on remotely, monitoring power consumption, and sensing the temperature in the ITE rack.

16.2.1 Overview and Class of Devices

A rack PDU is mounted in an ITE rack and provides electrical power to various IT devices such as servers, networking, and storage equipment. Today, rack PDUs are available in a number of configurations. We describe in the following the basic characteristics of four types of rack PDUs using frost and Sullivan's classifications as a general guide (Fig. 16.1). In Section 16.4.3, we will discuss the strengths and weaknesses of each PDU type as well as their typical applications.

16.2.1.1 Types of Rack PDUs

Rack PDUs can be divided into two categories: Non-Intelligent PDUs and Intelligent PDUs.

Non-Intelligent PDUs

- *Basic PDUs: Basic PDUs* are power strips constructed out of high-quality components for use in critical environments such as data centers. They distribute desired voltage and current to multiple outlets.
- *Monitored PDUs: Monitored PDUs* allow users to view a local display that typically provides electric current information only. However, this information cannot be accessed remotely as the units have no network connectivity capabilities.

Intelligent PDUs

- *Metered Input PDUs: Intelligent PDUs* meter power at the PDU level and can display data both locally and over a network. Metering helps users to determine power usage and available capacity at the rack and facilitates provisioning. By metering at the input level, users can avoid overloading circuits and easily calculate efficiency metrics such as Power Usage Effectiveness (PUE).
- *Metered Outlet PDUs:* Meter power at the outlet level and can display the data both locally and over a network. Like metered input PDUs, outlet-metered models help users to determine power usage and available capacity at the rack, and also facilitate provisioning. Most importantly, outlet-level metering allows users to understand power consumption at the device or server level which makes it possible to allocate costs to specific business units or customers.
- *Switched PDUs:* These PDUs offer features of Metered Input PDUs and also provide controlled on/off switching of individual outlets. They enable authorized users to securely power cycle devices remotely in a specific order, offer power sequencing delay to minimize inrush currents, prevent unauthorized device provisioning, and can power off devices that are not in use to conserve energy.
- *Switched PDUs with Outlet Metering:* These PDUs combine the capabilities of Switched PDUs and Metered Outlet PDUs.

16.2.2 Electrical Power Distribution to the Rack

16.2.2.1 Branch Circuits

Power is distributed to the rack over one or more electrical branch circuits. Branch circuits are power feeds that originate from a panel, switch, or distribution board and terminate into an electrical receptacle mounted in a junction box near the ITE rack. Branch circuit wiring can be overhead, underneath a raised floor, or both. The rack PDU itself could have multiple branch circuits. See Section 16.2.5 for details regarding branch circuit protection requirements.

16.2.2.2 Branch Circuit Load Capacity

The power that can be delivered by a branch circuit depends on the electrical characteristics of the circuit. A key factor in delivering power to a rack is whether the power is single phase or three phases. The amount of electricity delivered to a rack is often referred to as the load capacity and is the product of the rated voltage and the rated current and is presented as volt-amps (VA) or kVA (VA × 1,000). Given the rated voltage and current, the load capacity that can be delivered by a branch circuit is determined using these formulas:

- Single-phase: load capacity = rated voltage × rated current
- Three-phase: load capacity = $\sqrt{3}$ × rated voltage × rated current

16.2.2.3 Branch Circuits: Rated Voltage

The rated voltage of a branch circuit specifies both its magnitude (volts) and number of phase conductors (Table 16.1). Single-phase wiring is straightforward and consists of two wires (plus safety ground) where the ac voltage is a single sinusoidal wave as measured across the two wires.

Three-phase wiring is more complicated and consists of either three (three-phase conductors) or four (three-phase and one neutral) wires, plus safety ground (Fig. 16.2). Three-phase branch circuits deliver more power but require a rack PDU specially designed for three-phase branch circuits.

TABLE 16.1 Branch circuit rated voltage and wire requirements

Rated voltage	Location	# of wires	Outlet voltage(s)
120 V	North America	2 (phase + neutral)	120 V
208 V	North America	2 (phase + phase)	208 V
230 V	International	2 (phase + neutral)	230 V
208 V 3Ø	North America	3 (three-phase lines)	208 V
208 V 3Ø	North America	4 (three-phase + neutral)	Mixed 120 and 208 V
400 V 3Ø	International	4 (three-phase + neutral)	230 V

Source: Courtesy of Raritan, Inc.

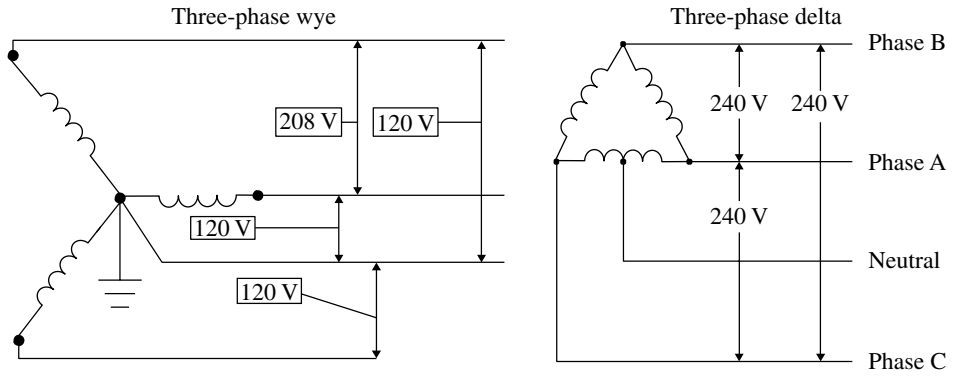


FIGURE 16.2 Three-phase wiring diagram. Source: Courtesy of Raritan, Inc.

Internally, a three-phase rack PDU divides the three or four branch circuit wires into pairs of single-phase circuits—and these single-phase circuits are wired to the rack PDU's single-phase outlet receptacles.

The three-phase conductors have the same voltage magnitude, but the sinusoidal ac waveforms are out of phase with each other by 120°. Regardless of the number of wires, the rated voltage of three-phase wiring is always the measured voltage difference between any two-phase conductor wires—not the difference between a phase wire and neutral. Just as with single-phase power described earlier, connecting across one 120 V hot line and the neutral provides 120 V ac. But connecting across any two 120 V hot lines, say, L1 and L2, provides 208 V ac, not 240 V ac. Why? Because the phase of L1 is offset 120° from L2, the voltage is not 240 V (120 V × 2), as it is for single phase, but is 120 V × square root 3 or 120 V × 1.732 = 208 V. A three-phase PDU can deliver three circuits of 208 V each. Some rack PDUs take advantage of a neutral wire to provide three circuits of both 120 and 208 V. But as mentioned in the preceding paragraph, regardless of the number of wires, or whether or not both higher and lower voltages are supplied as outputs, a three-phase rack PDU is rated at the voltage between two phases, for example, L1 and L2, which in the example here is 208 V.

A rack PDU can also provide 400 V ac. Just as with the 208 V three-phase rack PDU, if one of those lines is connected to a neutral instead of another line, this provides a single-phase output circuit that for a 400 V rated PDU is 230 V ac (400 V/1.732 = 230 V). This is a common deployment in Europe and is becoming more common for high-power racks in North America.

Three-phase rack PDU specifications often use the terms Wye and Delta or the Greek letters γ and Δ. These terms or letters were chosen because the electrical configuration diagram of a Delta transformer looks like a Δ and the electrical configuration diagram of a Wye transformer looks like a γ. A rack PDU that does not convert a higher input voltage, for example, 208 or 400 V, to a lower output voltage, for example, 120 or 230 V, but instead retains the higher voltage throughout uses a Delta transformer. A Delta transformer has three connection points, one at each corner of the triangle. Each of these points is a connection for one of the three lines. Connecting any point to any other point provides a line-to-line connection, for example, L1 to L2, and provides 208 or 400 V as described in the earlier examples.

A rack PDU that does convert a higher input voltage to a lower output voltage uses a Wye transformer. A Wye transformer has three connection points for the lines, one at the end of each "arm" of the γ and one at the "foot" of the γ.

The center intersection point of the γ is a fourth connection point and is where the neutral wire is attached. Connecting any two of the three line connections together, for example, L1 and L2, provides 208 or 400 V. Connecting any one of the three line connections to the neutral, for example, L1 and neutral, provides 120 or 230 V as described in the examples.

16.2.2.4 Branch Circuits: Rated Current

The current flowing in a circuit is determined by the size (thickness) of its wire and terminating receptacle. Branch circuits are required to be overcurrent protected using a circuit breaker or fuse. The rating of the circuit breaker is sized to the current-carrying capacity of the branch circuit's wiring and receptacle. For example, 10 AWG (American Wire Gauge) wire and a NEMA L21-30R receptacle are both specified at 30 A—so a circuit using these components must be protected by a 30 A circuit breaker.

In North America, the National Electric Code (NEC) for data centers (NEC Article 645) requires branch circuit wiring to be rated 125% greater than the total connected load. To ensure this requirement is met without running heavier gauge wires, all electrical devices (rack PDUs, computers, etc.) used in North American data centers must be certified to Underwriters Laboratories (UL) 60950-1. UL 60950-1 limits a device to draw no more than 80% of the rating of its input plug. For example, a rack PDU containing a 30 A NEMA L21-30P plug must not draw more than 24 A. This 80% limitation is commonly known as "derated" current.

Table 16.2 summarizes power available for various branch circuits.

16.2.3 Plugs, Outlets, and Cords

Rack PDUs are available with several types of plugs and receptacles (or outlets), designed so that only the appropriate rack PDU plug will fit into the appropriate circuit outlet and only the appropriate device plug will fit into the appropriate rack PDU receptacle. This is done to protect equipment, for example, so a device designed for 120 V only isn't plugged into a 208 V circuit, and for safety reasons, for example, a server that draws 30 A doesn't overload a circuit designed to handle only a maximum of 15 A.

The major classifications of plugs and receptacles in data centers are the International Electrotechnical Commission (ICE) and National Electrical Manufacturers Association (NEMA).

IEC plugs and receptacles (Fig. 16.3) are most common in Europe, and NEMA plugs and receptacles (Fig. 16.4) are most common in North America. However, many data centers in North America use IEC plugs and receptacles, and there are many families of plugs and receptacles in use in data centers around the world.

A significant concern in data center power distribution is unintentional disruption of power by accidentally disconnecting cords. Solutions exist that lock the plug into the receptacle and prevent the cord separating from the receptacle. There are three methods of securing the plug in the receptacle:

- Plug with tabs snaps into the receptacle locking them together
- Plug inserted into a receptacle with a locking mechanism that grips the plug ground blade
- Wire retention clips mounted to the PDU chassis hold the plug in the receptacle

TABLE 16.2 Branch circuit available power

Location	Rated voltage	Rated current (A)	Derated current (A)	Available power/branch circuit (kW)
North America	120 V	20	16	1.9
	208 V			3.3
	208 V 3Ø			6.7
International	230 V	16	16	3.7
	400 V 3Ø			11.0
North America	120 V	30	24	2.9
	208 V			5.0
	208 V 3Ø			8.6
International	230 V	32	32	7.4
	400 V 3Ø			22.1

Source: Courtesy of Raritan, Inc.

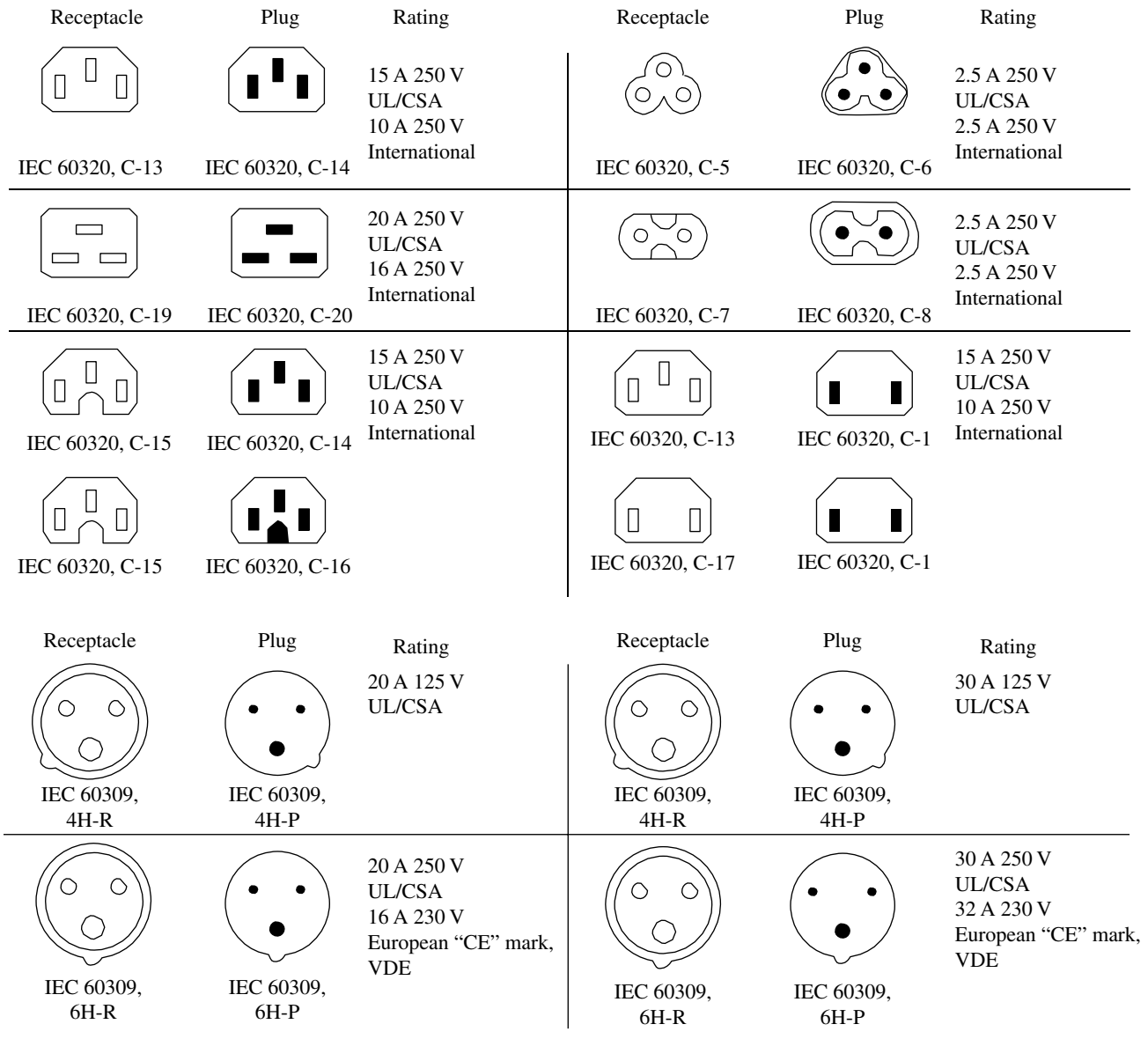


FIGURE 16.3 IEC plugs and receptacles. Source: Courtesy of Raritan, Inc.

The higher the current-carrying capability of a plug, receptacle, or cord, the greater the amount of wire conducting material, typically copper, required to prevent overheating the wire, which could lead to a fire. Note that the smaller the wire gauge number, the greater the diameter of the conductor.

The conductors are surrounded by insulating material and jacket, which may have special properties. For example, the jacketing may be designed to resist damage from exposure to oil. Typical insulating and jacket materials are PVC, rubber, and neoprene.

The number of wires in a cable can vary. Below are some typical data center configurations:

- Two: a hot wire and a neutral wire without a ground wire
- Three: a hot wire, a neutral wire, and a ground wire
- Four: three hot wires (L1, L2, and L3) and a ground wire
- Five: three hot wires (L1, L2, and L3), a neutral wire, and a ground wire

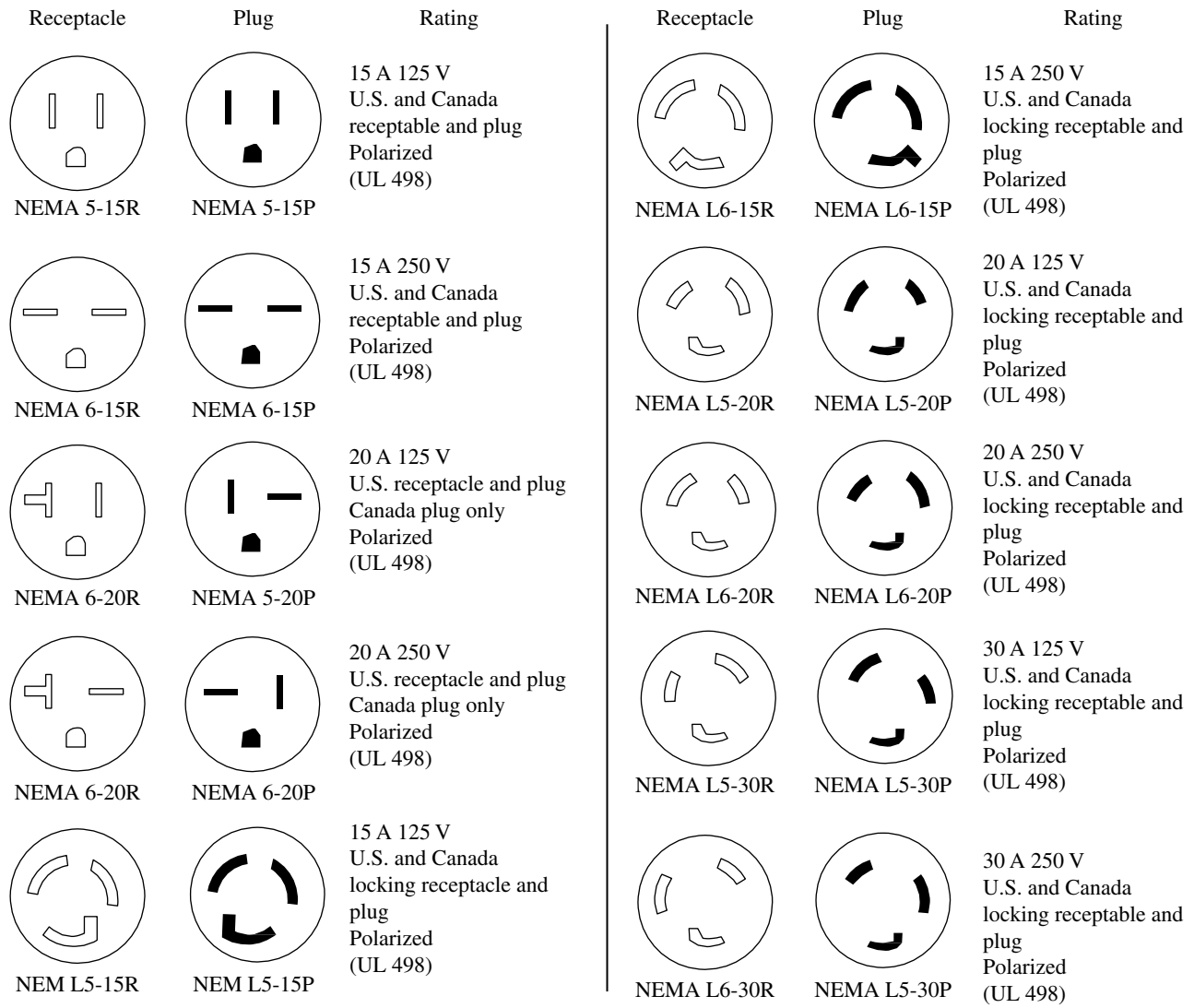


FIGURE 16.4 NEMA plugs and receptacles. Source: Courtesy of Raritan, Inc.

16.2.4 Ratings and Safety

Rack PDUs, like all other electrical equipment, are subject to many general and specific safety standards. Furthermore, there are general industry terms and conventions that should be understood in order to ensure a reliable and safe data center. These are discussed in detail in the following.

16.2.4.1 Nameplate Data

Nameplate data is the electrical power consumption information specified by the equipment manufacturer. It is typically a conservative estimate of the maximum amount of power the device could draw. This information is found on a label near the electrical power input to the device. More discussion of the use of nameplate data will follow.

16.2.4.2 Power Rating versus Load Capacity

There can be confusion about power capacities and load capacity. This stems from misunderstanding approval agency regulations and from some manufacturers who may use misleading terminology. In North America, typical circuits have a maximum current-carrying capability and use circuit breakers or fuses rated at 15, 20, 30 A, etc. In other words, a 20 A fuse will blow or a 20 A circuit breaker will trip if a 20 A circuit experiences more than 20 A for some period of time.

The period depends on the magnitude of the current and the type of fuse or circuit breaker protecting the circuit.

In North America, circuits are to be loaded to 80% of their maximum capacity. So, for example, a 15 A circuit should not carry more than 12 A, a 20 A circuit not more than 16 A, a 30 A circuit not more than 24 A, etc. The 80% value, for example, 16 A for a 20 A circuit, is often referred to as the derated value or the load capacity. In North America, a rack PDU vendor's specifications sheet may have a few current-carrying specifications. The specifications provided and the terminology used may vary by vendor, but the following are typical examples:

- Maximum line current per phase: 30 A
- Rated current: 24 A (30 A derated to 80%)
- Maximum current draw: 6 × 16 A

(Six circuits, each capable of carrying up to 16 A)

In Europe and elsewhere, circuits are simply described at their rated capacity, for example, 16 and 32 A.

Apparent power is specified in VA, which is volts × amps. Load capacity is specified in VA, where amps are the rated current, that is, the derated value. For example, for a single-phase rack PDU with 208 V and rated (not maximum) current of 24 A, the load capacity is 5.0 kVA (208 V × 24 A).

16.2.4.3 Approval Agencies

In order to meet applicable local and NEC, rack PDUs must be safe and not emit electromagnetic radiation. Standards exist and recognized approval agencies are contracted by manufacturers to test products. A product that passes agency testing receives an approval listing number, and the manufacturer can then affix the agency approval listing logo on each product. The listing logo is your assurance that the product meets applicable safety and electric codes (Table 16.3). The manufacturer is required, upon request, to provide you the listing number and a copy of the testing report. You can also submit the listing number to the approval agency to verify compliance.

16.2.4.4 Proper Grounding

The NEC (NEC Article 645.15) requires all exposed non-current-carrying metal parts of an IT system to be grounded. This means all equipment within a rack and the metal rack itself must be grounded. The inlet plug of a PDU contains a ground pin. When this plug is connected to a properly wired receptacle, the PDU becomes the grounding point for the equipment plugged into it. The PDU can also be used to ground the metal rack—and most PDUs contain a special threaded hole for this purpose. Typically, a grounding wire is connected to the rack and the PDU using screws. Care should be taken to ensure paint on the rack is scraped off where the grounding wire is attached to ensure proper electrical conduction. There are special grounding screws with teeth under the head to ensure a good ground.

16.2.5 Overload Protection

The standard UL 60950-1 applies to the safety of ITE and requires the use of branch circuit overcurrent protection for ITE PDUs greater than 20 A. Typically, ITE PDUs greater than 20 A and certified after April 2003 must have built-in UL 489 circuit breakers or fuses (e.g., UL 248-5 fuses) suitable for branch circuit protection.

UL 60950-1 permits products at a maximum current of 15 and 20 A without circuit breakers or fuses, since the 15 or 20 A circuit breakers in the building are considered sufficient to protect the PDU; however, supplementary protection in the PDU provides additional protection. UL also "grandfathers" PDUs at more than 20 A that were certified prior to April 2003. Although such PDUs are still being

TABLE 16.3 Safety and electromagnetic approval agencies

Approval	Description	Standard/revision/year	Comment
UL	Safety	UL 60950-1	Required in the United States
cUL/CSA	Safety	CAN/CSA-C22.2 No.	Required in Canada
CB	Safety	IEC 60950-1	Common replacement for UL, CSA, and CE in countries that accept CB
CE	Electromagnetic interface (EMC)	EN 5502:2006	Europe
CE	Safety	EN 60950-1	Europe
FCC-A or FCC-B	EMC	FCC 47 CFR Part 15	United States
ICES-003	EMC	ICES-0003 issue-004	Canada

Source: Courtesy of Raritan, Inc.

sold, their use should be avoided if they are to be incorporated in larger ITE systems designed to the latest UL 60950-1 standard.

Newly certified ITE PDUs at more than 20 A are required to use overcurrent protection that meets branch circuit protection requirements in accordance with the National Electrical Code, ANSI/NFPA 70. In effect, this means these products are required to have circuit breakers listed under UL 489, "Standard for Molded-Case Circuit Breakers, Molded-Case Switches and Circuit Breaker Enclosures," or fuses, such as those listed to UL 248-5, "Low-Voltage fuses—Part 5: Class G fuses."

In addition to standard UL 489, UL also publishes the standard UL 1077, "Standard for Supplementary Protectors for Use in Electrical Equipment." Devices certified to this standard are called "Supplementary Protectors" and are called "Recognized" components, not "Listed" devices, as are UL 489 breakers. UL Listed Circuit Breakers meet more stringent requirements for branch circuit protection than Supplementary Protectors with UL Recognition.

Circuit breakers are used in a variety of ways. They are mounted in panel boards (also referred to as building PDUs) and rack PDUs to protect branch circuit wiring. They are also built into equipment to protect components and systems. Interrupting a short circuit—current flow limited only by the resistance of wiring—is a severe test of a circuit breaker. If the interrupting capacity of the breaker is not adequate, the device can literally explode.

UL 489 requires the breaker to be functional after being subjected to a short-circuit test. UL 1077 and the IEC standard EN 60934 allow for breakers to clear a short-circuit condition but become safely destroyed in the process. UL 489 breakers can interrupt short circuits of 5,000 A or more. Typically, UL 1077 breakers can interrupt fault currents of 1,000 A.

Overloads can be short term or long term. The protective device must not trip with a momentary overcurrent event that is normal for the piece of equipment being protected. Servers, for example, may create inrush currents as their internal power supply and filter circuits start. These inrush currents typically last only a fraction of a second and seldom cause a problem. If an overload lasts longer than a few minutes, the breaker should open to prevent overheating and damage. What gives a breaker the ability to discriminate between normal and damaging over currents is its delay curve.

16.3 ELEMENTS OF THE SYSTEM

Rack PDUs are the final endpoint of power supplied to ITE from incoming building feeds through a chain of equipment including UPS, transformers, and larger PDUs and circuit panels. IT and facilities management are increasingly viewing their rack PDUs not merely as a collection of power outlets for ITE but as a network of critical devices that significantly impact the overall efficiency and effectiveness of the data center. As such, they need to be properly managed like the ITE they power. This is driving the trend for use of more intelligent PDUs in data centers with environmental sensors and even integration with higher-level data center management systems. This section describes not only the components of the physical rack PDU and basic environmental sensors but also the rack PDU management system that leverages the intelligence in PDUs for operational improvements and energy-use reduction. Further, this system can interface with and become part of a larger ecosystem of enterprise IT and facilities management systems.

16.3.1 Rack PDU Anatomy

Over the past few years, average power consumption per server has rapidly increased with the adoption of high-power computing equipment like blade servers or data center containers. In addition, ongoing deployment of densely packed storage, virtualization, and cloud computing results in data centers with greater watts per square foot requirements from more densely packed racks such as a rack filled with 1U servers. To support new, power-hungry ITE, data center managers have to deliver more power to the ITE rack. Over the last decade, the typical power required at a rack has increased from 2 to 12 kW and continues upward.

16.3.1.1 Single-Phase or Three-Phase Input Power for Rack PDU

To accommodate the increased power demands at ITE racks, data center managers deploy rack PDUs capable of supplying multiple circuits, higher voltages, and higher currents. One way to increase the power at the rack is to increase the number of circuits and the voltage coming to the rack.

The amount of power available for use is referred to as apparent power and is calculated as volts × amps and is described as VA. A 120 V, 20 A circuit has an apparent power of 2,400 VA or 2.4 kVA. A 208 V, 20 A circuit has an apparent power of 4,160 VA or 4.2 kVA. Thus, one 208 V circuit provides almost twice as much power as one 120 V circuit assuming the current (amperage) remains the same. With three 208 V circuits, a substantial amount of power can be deployed in one three-phase PDU.

The cable to provide power to a three-phase PDU is thick and heavy but not as thick and heavy as the multiple, individual cables required to provide the same amount of power using either single-phase 120 V or single-phase 208 V. Running a single three-phase power cable to each three-phase rack PDU reduces both the number of cables, making installations easier, and the physical bulk of the cables, so

less space is filled with cables blocking necessary cooling airflow under raised floors and within racks.

In cases where power needs to be provided at 120 V for devices such as routers, hubs, and switches, as well as at 208 V for demanding servers, three-phase PDUs can provide outlets with both 120 V (one of the three lines and a neutral) and 208 V (two of the three lines). Three-phase power at the rack is a convenient way to efficiently deploy both greater power capacity and flexibility.

16.3.1.2 Form Factor

Rack PDUs are available in heights of one rack unit (1U, 1.75 in) or two rack units (2U, 3.5 in) for horizontal mounting in a 19 in equipment rack.

Zero U rack PDUs mount vertically, typically to the vertical rails at the back of the rack. This can offer advantages. Zero U PDUs don't consume any rack unit spaces, and since the receptacles on the Zero U PDU line up better with the power cords for each IT device in the rack, they allow for the use of shorter power cords. This results in neater cable arrangements contributing to better airflow within the rack, which can improve cooling efficiency. Depending on the rack cabinets, Zero U rack PDUs can be mounted with screws or hung into the cabinet via buttons that are spaced 12.25 in apart.

High-power rack PDUs are commonly equipped with circuit breakers for branch circuit protection. These circuit breakers may cause the rack PDUs to extend deeper into the racks. Consider how these PDUs are mounted in the rack, whether outlets facing center or back, to allow for cable management, airflow, and easy accessibility and serviceability of the ITE.

16.3.1.3 Outlet Density and Types

Rack PDUs vary in the number of outlets supported based on the physical size (length, width, and depth) and thus the total space available for mounting outlets and internal components and the power-handling capacity of the PDU. For example, a 1U rack-mounted PDU may have enough space for eight 120 V/15 A NEMA 5-15R outlets, whereas, a 2U rack-mounted PDU may have enough space for 20 NEMA 5-15R outlets. On the other hand, a Zero U PDU may have 24 IEC C-13 230 V/10 A outlets or just four 250 V/30 A NEMA L15-30R outlets to support blade servers.

In the case of a large number of devices, each demanding a moderate amount of power, a large number of moderate power outlets is required. A typical dense "pizza box" deployment would include two rack PDUs for redundant power where each PDU is loaded to 40% so that if one power feed fails, the other feed will not exceed the NEC requirement of 80% (for North America). Typical outlets for "pizza box" servers are IEC C-13 (up to 250 V, 16 A) and NEMA 5-20R (up to 125 V, 20 A, 16 A rated).

In the case of high power consumption at a rack for a few devices, each of which consumes a lot of power, such as blade servers, storage, or network devices, the total amount of power required might be comparable to the high outlet density example given earlier, but the number and type of outlets may be different. Density for devices such as blade servers depends on their number of power supplies (often between two and six for redundancy), how the power supplies are configured (power supplies are most efficient when they operate close to their maximum level), and how many devices will be deployed.

In the case of a few devices demanding a lot of power, a large number of outlets may not be needed but outlets capable of delivering substantial power will be required. Typical outlets for high-demand devices such as blade servers at 208 or 230 V are IEC C-13 (16 A) or C-19 (32 A) or, less commonly, NEMA L6-20R (20 A, 16 A rated) or L6-30R (30 A, 24 A rated) locking outlets.

16.3.1.4 Connectors: Ethernet, Serial, Sensor, USB, and Others

Today, only the Non-Intelligent PDUs (very basic rack PDUs) have no external connectors—an input plug and outlets, much like a common power strip. Most rack PDUs now include a variety of connectors based on application requirements. Below, we describe four rack PDU connector configurations and general applications:

1. No connectors for external management or remote alarms and may not even have a local display. Not suitable for most data center applications today.
2. Local buttons allow navigation to see basic unit and outlet data.
3. A serial RS232 connector for local metering; local meter may be an LCD or LED. Can be plugged into a terminal or console server for Telnet or SSH remote access. Access via a menu or CLI using terminal emulation. Local buttons allow navigation to see basic unit data. No SNMP support available for alarms, unless via a specially developed serial console server. Typically non-switched.
4. Ethernet (RJ-45) and RS232 serial (DB-9M) connectors allow remote metering for the PDUs, circuit breakers, and outlets. Two Ethernet (RJ-45) connectors are useful for redundant network connections, PDU-to-PDU cascading. Multiple USB connectors support PDU-to-PDU cascading, wireless networking, webcams, door lock security, and Zero-Touch deployment. SNMP port is useful for alarms, Telnet or SSH access possible for command line access. Support for sensors (temperature, humidity, airflow,

air pressure, and others) may be available on the PDU or with an add-on external device. Remote metered models typically have an LCD or LED display as well with buttons for navigation to see basic unit and outlet data.

16.3.1.5 Branch Circuit Protection

Since April 2003, UL requires branch circuit protection, circuit breakers, or fuses for PDUs where the inlet current is greater than the outlet current, for example, 30 A (24 A rated) plug and 20 A (16 A rated) outlets. 15 and 20 A (12 and 16 A rated) rack PDUs can be supplied without branch circuit breakers because circuit breakers in upstream panel boards are deemed to provide the necessary protection. Rack PDUs with breakers or fuses are like mini-subpanels. For example, a 208 V, 30 A (24 A rated) three-phase PDU has three circuits, and each circuit/set of outlets has a 20 A circuit breaker.

There are four types of circuit breakers: thermal, magnetic, thermal–magnetic, and hydraulic–magnetic.

Thermal circuit breakers incorporate a heat-responsive bimetal strip. This technology has a slower characteristic curve that discriminates between safe, temporary surges and prolonged overloads.

Magnetic circuit breakers operate via a solenoid and trip nearly instantly as soon as the threshold current has been reached. This type of delay curve is not ideal for servers that typically have inrush currents anywhere from 30 to 200% above their normal current draw.

Thermal–magnetic circuit breakers combine the benefits of thermal and magnetic circuit breakers. These devices have a delay to avoid nuisance tripping caused by normal inrush current and a solenoid actuator for fast response at higher currents. Both thermal and thermal–magnetic circuit breakers are sensitive to ambient temperature. A magnetic circuit breaker can be combined with a hydraulic delay to make it tolerant of current surges.

Hydraulic–magnetic breakers have a two-step response curve. They provide a delay on normal over currents but trip quickly on short circuits and are not affected by ambient temperature.

Circuit breakers used in rack PDUs are typically thermal–magnetic or hydraulic–magnetic with delay curves that allow for reasonable inrush currents (servers typically have inrush currents 30–200% above their normal operating load) while protecting devices from excessive fault currents.

Fuses are also acceptable for PDU circuit protection. However, replacing a fuse can be time-consuming and may require an electrician leading to longer mean time to repair (MTTR). Spare fuses must be stocked in inventory and the correct fuse must be used to ensure reliability and protection.

The following are some points to consider when selecting a rack PDU:

- Compliance with the latest fuse and circuit breaker standards
- The acceptable MTTR for fuse replacement versus circuit breaker resetting
- Impact on uptime service-level agreements (SLAs) if a fuse blows versus if a circuit breaker trips

16.3.1.6 Circuit Breakers: Single Pole versus Double and Triple Pole

The reliability and flexibility of the branch circuit breaker configuration are important. Typically, circuit breakers are available as single-, double-, or triple-pole devices. Single-pole breakers are appropriate for circuits comprised of a hot wire and neutral, for example, 120 V at 20 A or 230 V at 16 A. Single-pole breakers provide a disconnect for the single hot wire used in circuits with a hot wire and neutral. Double-pole breakers provide a disconnect for circuits comprised of two hot wires, for example, 208 V at 20 A. Some PDU designs use double-pole (or triple-pole) breakers to provide protection for two different circuits, for example, two different hot wires. Since a single double-pole breaker is less expensive than two single-pole breakers, this type of design will lower the cost. Double-pole breakers will trip if either of the two circuits they protect is overloaded. It is less expensive than two (or three) single-pole breakers, but unless the poles can be operated independently, in a maintenance shutdown or trip, all two or three circuits are de-energized.

For example, assume a rack PDU with six branch circuits is protected by circuit breakers. Some rack PDUs in this configuration may protect the six circuits with three double-pole circuit breakers—one double-pole circuit breaker for the circuits with Line 1, one for the circuits with Line 2, and one for the circuits with Line 3. It is less expensive to use double-pole circuit breakers, but there are some drawbacks. Double-pole breakers will trip if either of the two circuits they protect is overloaded. This means double-pole breakers are less reliable. Double-pole breakers are also limiting because if you choose to shut off a circuit, for maintenance, for example, you have no choice but to shut off both circuits. Alternatively, some rack PDUs protect the six circuits with six single-pole circuit breakers—one breaker per circuit. This is more expensive but single-pole breakers are more reliable and less limiting. Look for rack PDUs that allow only one circuit to be de-energized for improved reliability and flexibility.

16.3.1.7 Circuit Breaker Metering

Circuit breaker metering is a useful feature on any rack PDU, but it is particularly important when dealing with high power because the consequences of tripping a breaker can be disastrous if it means losing several blade servers. With circuit breaker metering, the end user sets a threshold. When that threshold is crossed, an alert is delivered so the end user knows power demand needs to be reduced or there is the risk of tripping a circuit breaker. Monitoring branch circuit breakers is important since high-power draw means a greater chance of tripping a breaker.

Line metering, intended for three-phase rack PDUs, is very useful for balancing the power drawn over each line. Overdrawing power from one line relative to another line wastes available power, and unbalanced lines can place excessive demands on the neutral in Wye-configured PDUs.

16.3.1.8 Cord Length, Feed, and Retention

Rack PDU power cords vary in length depending on the whips (power cables from a building PDU) and the location of the racks. The rack PDU power cord must be long enough to reach its power source, which is typically a whip located under the raised floor or an outlet above the rack. A common power cord length is 10 ft (3 m), but other lengths can be specified, to a UL maximum of 15 ft (4.5 m).

Rack PDU power cords may exit the PDU itself from the rear, the front, the top, or the bottom. With the power cable exiting the bottom, a Zero U PDU, the data center manager will need to ensure sufficient space for the bend radius of the cable. In general, a bend radius of 5.25 in (3U) will be sufficient, but this should be confirmed as bend radii will depend on the gauge (AWG) of the cord. A smaller bend radius may be acceptable for thin cables and a larger bend radius may be required for heavy-duty cables. The orientation of the PDU power cord may seem trivial, but it can be a potential problem depending on the physical rack and the location of the power source for the rack. Consider the orientation of the power cord and how it will be routed to connect to the whip. For example, does the power come up from the raised floor or down from cable trays above the racks and is there room inside the rack to route the cable so as not to block airflow?

16.3.1.9 Cord Retention

Good PDU cord retention practices, just like rack cable management, can make a big difference in operational efficiency and reliability. Taking steps to support, organize, and secure many power cords using some method of cord retention will dramatically improve your ability to access and manage the equipment connected to PDUs inside the rack. This will also minimize the chance of inadvertently unplugging power cords from rack PDUs. You should neatly arrange and secure the power cords between equipment and the rack PDU to allow for maximum airflow.

There are a few different methods to secure IT equipment power cords. Two of the most common are to design the rack PDU with locking IEC C-13 and C-19 outlets that work on standard IEC power cords. Another option, which is less expensive, is to use specially designed outlets and power cords with tabs that securely hold the cord in place until the tabs are released (Fig. 16.5). These cords can be purchased in different colors to help identify which power supply is plugged into the PDU.

16.3.1.10 Local Display and User Interface

Virtually, all rack PDUs designed for data center use have built-in displays, typically LCDs, to show current draw for the entire PDU unit. Local displays have limited functionality compared to the information and control available from a remote interface, but they can be convenient and useful when working at the rack itself. The local display might allow an IT admin to toggle between current draw and voltage or, for those rack PDUs that monitor individual outlets, to sequence through the outlets to determine the current being drawn by each device. Some switched and intelligent models have LCD indicators next to each outlet to display status, whether it is on/off, booting, firmware upgrade, or fault.

In addition to a local display on the rack PDU itself, some PDUs offer a serial interface for local terminal connectivity via a laptop for configuration, diagnostics, or connectivity to a serial console server that concentrates multiple connections.

16.3.1.11 Remote User Interface

For a remotely accessible rack PDU (all but the basic PDU or PDU with metering and only local display), there are typically two choices for a remote user interface to the rack PDU over an IP network. The most common is a Web-based graphical user interface (GUI) (Fig. 16.6) to an Ethernet-enabled PDU. Some PDUs support SSL-encrypted access (using https), while others support only unencrypted access

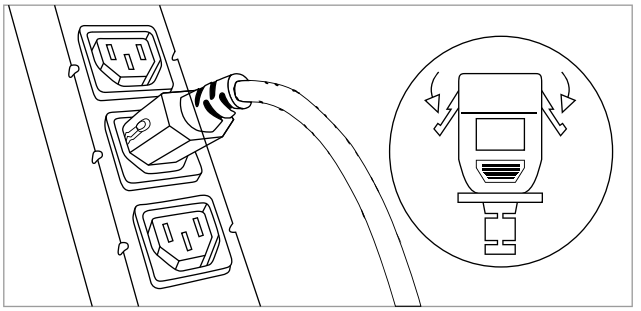


FIGURE 16.5 SecureLock. Source: Courtesy of Raritan, Inc.

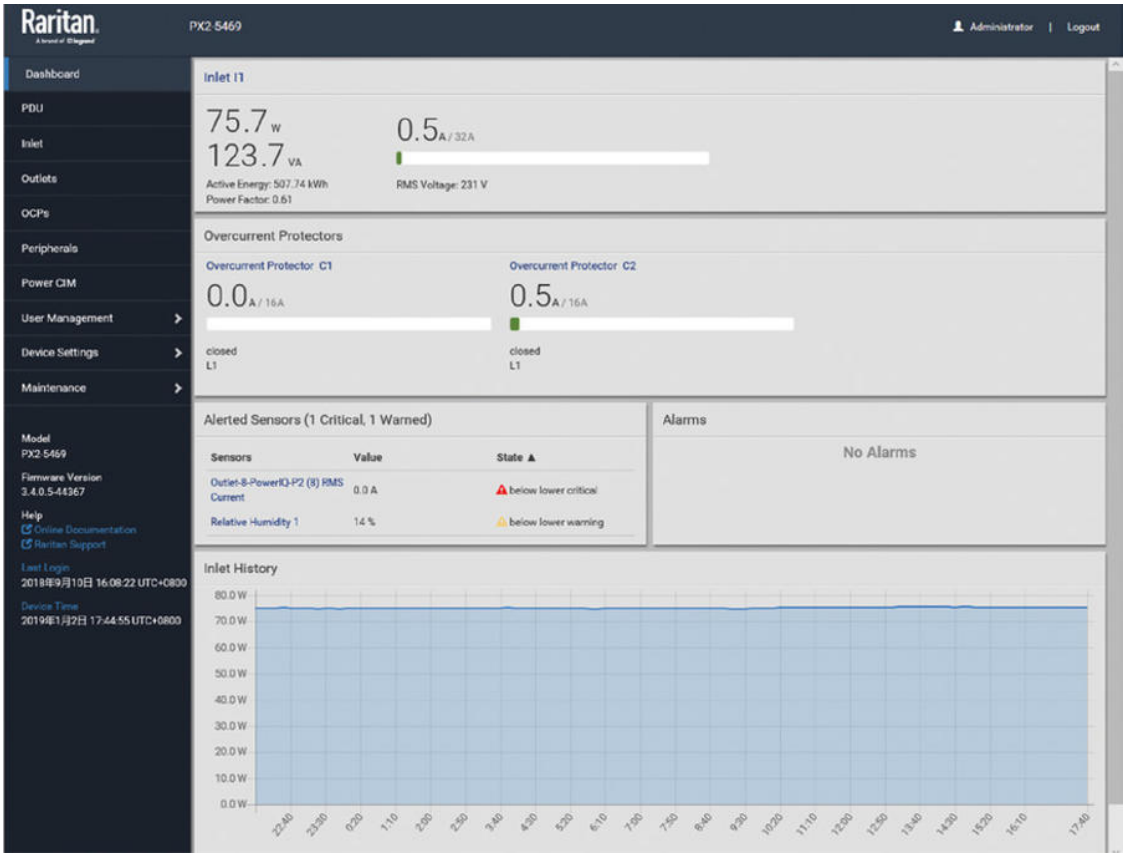


FIGURE 16.6 Rack PDU management system graphical user interface. Source: Courtesy of Raritan, Inc.

(using http). Check your organization's security requirements when selecting a PDU. The PDU can also be accessed via Ethernet over IP using SSH (encrypted) or Telnet (unencrypted) with a CLI. Security considerations should be kept in mind before enabling/disabling Telnet access. Some PDU manufacturers provide a serial console server that connects to the PDU locally via serial (RS232) and allows access to the unit remotely using SNMP or CLI.

One factor to consider is integration with central directory services for user authentication and access control. This becomes especially important when the rack PDU offers the ability to remotely turn on/off/recycle individual outlets or groups of outlets. Finally, remote access to the PDU does not eliminate the need for some local access to the PDU with an LED/LCD and associated buttons.

16.3.2 Environmental Management

With the IT industry's increasing focus on improving data center efficiency, more rack PDU manufacturers are offering environmental sensors. These include sensors to measure rack air temperature at the server inlets, humidity, airflow, vibration, smoke, water, and air pressure. Some PDUs will have preinstalled sensors; others provide for optional, plug-in external sensors. Another common approach is to deploy a completely independent rack management system, choosing from a wide range of environmental sensors; however, this has the disadvantage of consuming additional rack space for the rack management system as well as the cost of a separate infrastructure—for example, IP addresses, Ethernet ports, and cabling. Connectivity for sensors is typically either via RS485 or 1-Wire®.

16.3.2.1 Temperature Sensors

Temperature sensors monitor the air inlet temperatures at IT devices such as servers. (See the ASHRAE sensor placement diagram in Fig. 16.7.) Since ITE generates considerable heat, manufacturers specify a range of acceptable temperatures for proper operation. A sensor-capable PDU should allow thresholds to be set for sending automatic alerts when the inlet temperature approaches the vendor-specified maximum to prevent servers from shutting down or failing due to overheating. In addition, it is also a good practice to set a minimum to provide alerts when the inlet temperature is colder than necessary. From a data center plant perspective, the cost of cooling and moving air is the largest infrastructure expense, so maintaining IT inlet air temperatures colder

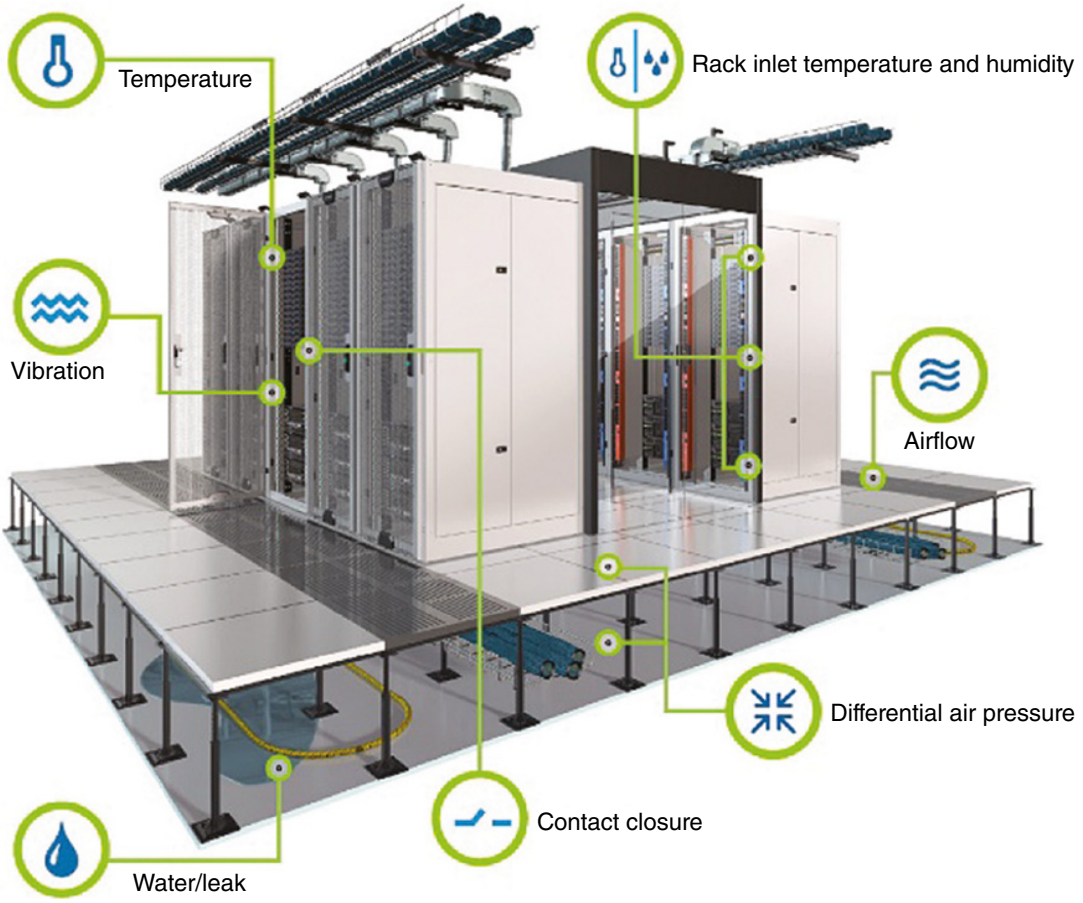


FIGURE 16.7 Temperature sensors monitor air inlet temperature at IT devices. Source: © Raritan, Inc.

than necessary merely wastes energy and money. Temperature sensors at the rack also provide early warning about temperature extremes, hot spots, or cold spots and can help identify when an HVAC system is becoming unbalanced. To ensure that ITE is getting enough cool air, the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) has recommended that temperature probes be placed at specific locations at the inlets of equipment racks.

16.3.2.2 Humidity Sensors

Understanding the basics of what humidity is and how it affects your server room can impact how long your computer equipment lasts and how much your electricity bill costs. Humidity is a measurement of moisture in the air. High humidity can cause condensation buildup on computer components, increasing risks of shorts. Likewise, if the humidity is too low, data centers can experience electrostatic discharge (ESD). Humidity can be monitored per area or zone to ensure that it is in the safe range. ASHRAE has recommended ranges for the data center that should be consulted. Appropriate thresholds and alarms should be set to indicate a potential problem.

16.3.2.3 Airflow Sensors

Airflow sensors will detect a reduction of air movement that might create the potential for overheating, which can destroy ITE. There are two primary areas for monitoring airflow in the data center—above the floor (monitored at a number of points) and below the floor (monitored at select points). Differential airflow sensors are used to ensure that the pressure differential between the subfloor and the floor is sufficient to control air flowing from the subfloor to the floor above. Blockages in underfloor supply plenums can cause high pressure drops and uneven flow, resulting in cold spots in areas where cooling air is short-circuiting to the return path. Airflow sensors should have thresholds set, and alarms enabled, like other envi-

ronmental sensors, to ensure that data center managers are alerted when conditions are less than optimal for efficient cooling.

16.3.2.4 Air Pressure Sensors

It is important to have the appropriate air pressure in underfloor supply plenums, but sometimes, this is treated as an afterthought. Air pressure that is too high will result in both higher fan costs and greater leakage, which can short-circuit cooling air, while pressure that is too low can result in hot spots at areas distant from the cool air supply point. This can lead to poor efficiency "fixes" to correct the problem such as lowering the supply air temperature or overcooling the full space just to address a few hot spots. Differential air pressure sensors can be used to ensure that the pressure differential between the subfloor and the floor is sufficient. Maintaining appropriate room pressure prevents airborne particulates from entering the data center.

16.3.2.5 Contact Closure Sensors

Contact closure sensors can be used for a variety of applications. For example, a contact closure could send an alert when a cabinet door is opened and trigger a webcam to take a picture. Contact closure sensors can be connected to any device that can open or close a contact.

16.3.2.6 Other Sensors

There are a variety of other sensors that can be used in the data center. Examples include in-cabinet smoke, water, and vibration sensors. Like the other sensors mentioned earlier, these are used to send alarms when measured conditions are outside the range for proper data center operation.

16.3.3 System Connectivity

16.3.3.1 Physical Topology

Like many functions of data center management, the best practices for remote management of rack PDUs are evolving. The current best practice is to connect all remotely accessible rack PDUs to the "management network" (separate from the "production network") directly in order to collect periodic meter readings, get immediate notifications of any faults or potential problems, and enable remote power cycling of ITE (depending on the intelligence of the rack PDU). When planning for a new facility, provide for a minimum of two Ethernet drops for each cabinet to support rack PDUs, since each rack will typically require two PDUs.

However, some data centers try to reduce the number of Ethernet drops and IP addresses to minimize cabling and costs. In these cases, the data center deploys PDUs that can be daisy chained or cascaded so that one Ethernet connection can communicate with all the PDUs in the chain.

16.3.3.2 Communication Protocols

The communication protocols used are typically TCP/IP when PDUs are Ethernet connected, and proprietary protocols for PDUs serially connected to a console server, which, in turn, connects to the TCP/IP network via Ethernet. Most often, SNMP protocol is used for management, while LDAP and Active Directory are used for authentication, authorization, and access control. SSH and Telnet may be used for command line management and HTTP/HTTPS for Web-based access. There are rack PDUs now equipped with Multiple USB (device) ports that can be used to support USB devices such as webcams and Wi-Fi modems. Some rack PDUs support MODBUS, a common, older building management communication protocol, and some rack PDUs support the GSM modem protocol so that cell phones can receive one-way text alerts.

16.3.3.3 Managing the Rack PDU

The management system for data center power is often run on a "management network" separate from the production network. This reduces the likelihood of a Denial of Service (DOS) or other attack that would affect this critical function. In mission-critical facilities, there are often two connections to each rack PDU equipped with remote communications: one for syslog, SNMP traps, access via Web browser, and kilowatt-hour logging and another for critical functions like remote power cycling, status of circuit breakers, and load monitoring. In some cases, administrative functions, like rack PDU configuration, are performed via command line scripting through a secondary interface such as a serial port, while Ethernet remains the primary interface for all other functions.

Some important management functions are listed below:

- Audit logging to track all activity—like switching of outlets and configuration changes. Two or more syslog servers are often used for this function.
- Fault management—via SNMP with tools like HP OpenView, IBM Tivoli, and others. SNMP V2 is still the most commonly used, but SNMP V3, with its built-in security, is recommended for applications requiring outlet control.
- Configuration—via Web browser, SNMP, command line, or a central software tool.
- Firmware upgrades—not an issue for older PDUs with minimal functionality, but something that may be required for Ethernet-enabled PDUs. A central tool is essential to manage potentially large numbers of PDUs to simplify management and reduce cost of ownership.
- Alerts—via SMTP messages.

A combination of some or all of the aforementioned capabilities is required to effectively manage a data center. Check your application requirements and choose the PDU type appropriate for your application. If, however, you have multiple rack PDUs (40+), you will want to consider a comprehensive rack management system, discussed in Section 16.3.4.

16.3.4 Rack PDU Management System

A rack PDU management system is a software application (sometimes delivered as a software appliance) that consolidates communication with all your rack PDUs and in-line meters equipped for remote communication (Fig. 16.8). Its main functions are data collection, reporting, power control, element management, and fault management. The system collects and converts detailed power data into useful information and provides a central point for secure access and control across multiple rack PDUs with operation validation and an audit log. It simplifies the management of rack PDUs and alerts you to potential incidents. We include it in this section because for larger data centers (with more than 40 racks), it is a "must-have" to realize many of the benefits offered by metered, switched, or intelligent PDUs—improved energy efficiency, increased uptime, and lower operational costs.

16.3.4.1 Data Collection

Data collection is the fundamental component that enables reporting and most other management functions. The management system can collect only the data elements provided by the managed PDUs. As discussed earlier, basic PDUs provide no data, metered PDUs provide total unit data, and intelligent PDUs provide individual outlet data and more, so it is important to understand what data you will want to analyze when selecting rack PDUs. Typical data elements you will probably want to collect, as available from the PDUs, include total unit active and apparent power, line current and capacity, outlet-level current and active power, environmental sensor data, and real-time kilowatt-hour metering data.

Next, you will want to determine the granularity of the data. Your management system should offer a user-configurable data polling interval. For most applications, a normal polling interval is 5 min, which means the system will collect data points every 5 min, but if greater granularity is required, the rack PDU may need to store data readings so that the network is not overloaded with polling traffic. Finally, you will want to use a roll-up algorithm to collect data for long periods without causing the database to balloon and affect performance. For most energy management applications, data readings are rolled up to a maximum, average, and minimum—hourly, daily, and monthly.

Advanced polling options enable a customer to minimize network traffic while still enabling granular data collect. Advanced polling requires a rack PDU that has the memory capacity to record and store readings called samples. For example, a rack PDU might be able to store 120 samples. The rack PDU management system should offer the ability to configure optional sample rates for each rack PDU and

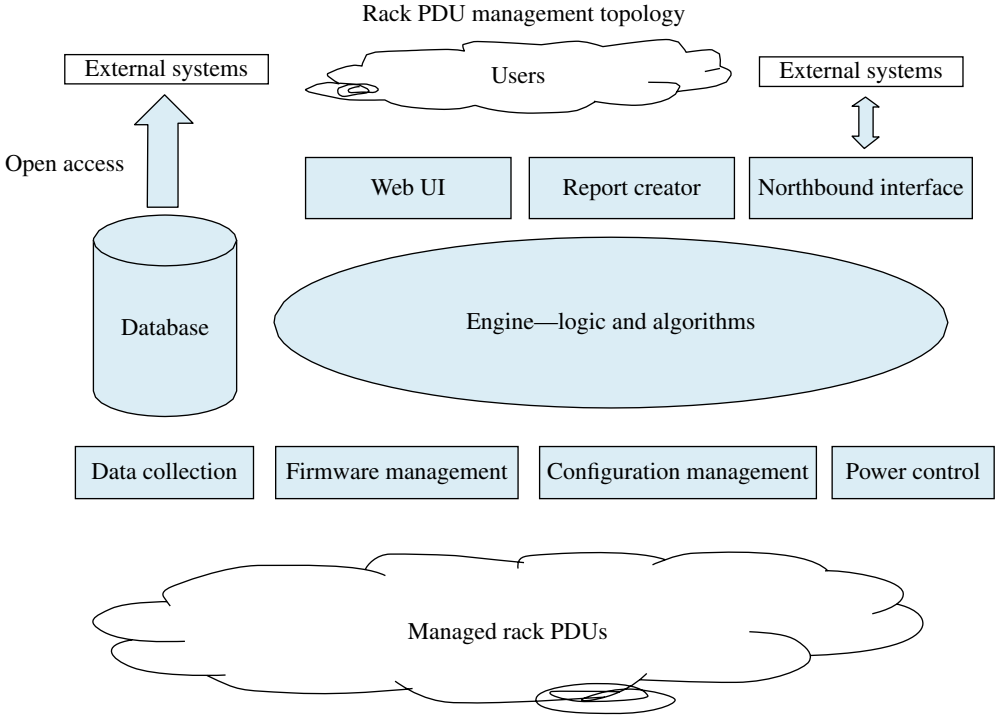


FIGURE 16.8 Rack PDU management topology. Source: Courtesy of Raritan, Inc.

also set optional polling intervals for the management system itself to collect the stored samples at each rack PDU not previously collected. For example, a rack PDU can be configured to record and store 1 min samples. The rack PDU management system can be configured to poll the rack PDU once an hour. In each poll, it will pull the 60 min samples since the last poll with the intelligence to know the last reading is recorded on the previous poll.

16.3.4.2 Reporting and Analytics for Power Monitoring and Measurement

Reporting and graphing should include active power, current, temperature, humidity, and information derived from the basic collected data such as energy usage, cost, and carbon emitted due to the energy consumed for standard and selected time periods.

Reports on maximums and minimums for current, temperature, humidity, and active power simplify key tasks and ensure that you are not in danger of exceeding circuit breaker ratings, overcooling, or undercooling. This environmental information can give data center operators the confidence to raise temperature set points without introducing risk to the ITE. Analyzing trend line graphs and reports and "what-if" modeling can help you do capacity planning based on real-world data.

Outlet-level data and reporting granularity can help you become more energy efficient. It enables you to determine the potential savings of upgrading to more energy-efficient servers or the benefits of server virtualization. Consolidating several low-utilization physical servers as virtual servers on one high-utilization physical server can reduce overall expenses, but you will need to understand the resulting power demand of the host servers. You can also establish objectives, report on usage, and implement changes for both physical components of the data center (floor, room, row, rack, and IT device) and also logical groupings (customer, department, application, organization, and device type). This level of detail creates visibility and accountability for energy usage, and some IT organizations issue energy bill-back reports to users/owners of the ITE.

16.3.4.3 GUI

The GUI (Fig. 16.9) is your window into all of the rack PDU management system functions. This should be clean,

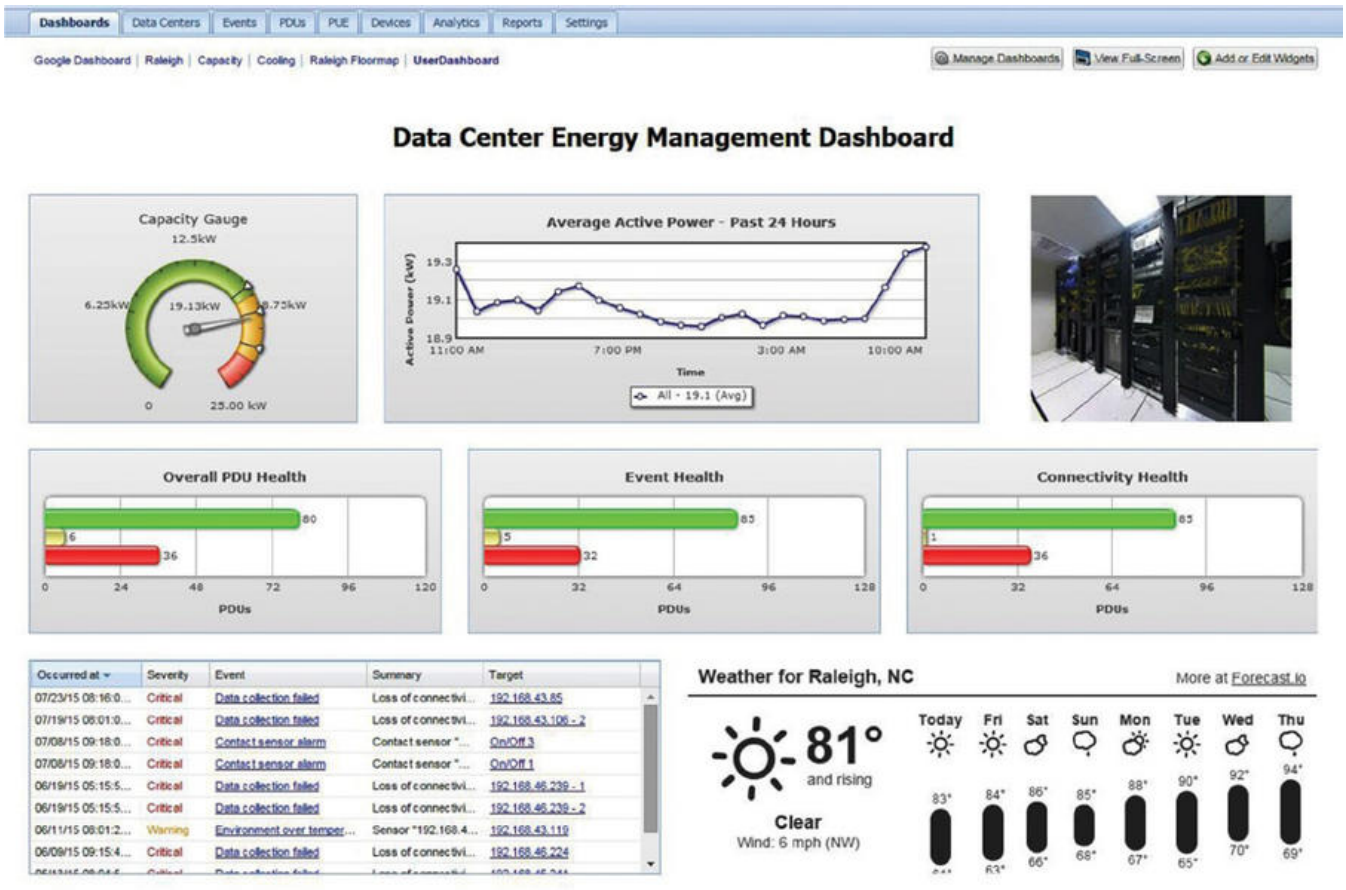


FIGURE 16.9 Data center infrastructure management (DCIM) monitoring software. Source: Courtesy of Raritan, Inc.

intuitive, and Web-based, functioning with all major Web browsers. A Web-based system provides you more remote access options and is easier to support and upgrade. The GUI will most likely include a user-configurable dashboard. The dashboard can be displayed in the data center network operations center for an easy at-a-glance view of the status of the data center power and environmental conditions. This will give your customers, either internal or external, a good indication of your data center management capabilities.

16.3.4.4 Element Management

The main components of element management include centralized rack PDU access and control, firmware management, and bulk configuration.

You can view all your managed rack PDUs from one Web browser window and get a summary view of the name, location, status, manufacturer model, and firmware level. You will want to be able to drill down to manage at the PDU unit level, line level, and, in the case of intelligent PDUs, outlet level. Finally, one-click, sign-on access to each managed PDU can give you control through the PDU's own GUI.

Since Intelligent PDUs run firmware with many configuration options, the rack PDU management system should allow you to centrally store rack PDU firmware/configuration versions and facilitate distribution to multiple PDUs. Configuration template storage and distribution will simplify initial PDU installation as well as future unit additions and replacements.

16.3.4.5 Fault Management

Rack PDU management systems often provide a map view and a floor layout view and use a color scheme to provide an at-a-glance view of the health of all managed PDUs. Health problems are discovered in a several ways. The system can receive an SNMP trap or a syslog event so that you become aware of the problem as it happens. Also, a management system can poll the rack PDU at set intervals to collect the health status of the communication path to the PDU or critical failures and forward events to a higher-level enterprise management system.

16.3.4.6 Local and Remote Control/Switching

Switched rack PDUs allow for outlet control including on/off power cycling. However, most IT devices have more than one power supply for power redundancy purposes, and these supplies are connected to outlets on separate rack PDUs. Through the management system, you can power cycle at an IT device level, which will programmatically switch outlets from multiple PDUs. The system should allow for grouping IT devices into racks such that you can control a full rack. Finally, any switched PDU must allow for flexible sequencing and delay so an inrush current spike does not trip a circuit breaker and so that application intra-system dependencies are taken into account during start-up and shutdown.

16.3.4.7 Security of Data and User Access

Remote monitoring, metering, and management require secure remote access via Ethernet and/or serial connections. To ensure security, an intelligent rack PDU should have strong encryption and passwords and advanced authorization options including permissions, LDAP/S, and Active Directory. A Web session timeout will protect against leaving an authenticated session live while not in use.

16.3.4.8 Administration and Maintenance

Most of the administration is initial setup. All systems will allow for GUI entry of these data but that can be time-consuming. Systems should also allow for the import of configuration information, for example, via CSV files. During the setup, you will add your rack PDUs and hierarchical and logical relationships. Hierarchical relationships include data center, floors, rooms, rows, racks, rack PDUs, and IT devices. Logical associations include owners/customers of the IT device and IT device type. The administrator will also set the data pruning intervals to ensure unnecessary data is pruned from the system.

16.3.4.9 Open Point of Integration

Most data centers have some other management systems already in use, so it is important that the rack PDU management system can be integrated into these systems to minimize the amount of duplicate data entry and collection. Asset management and enterprise reporting systems are two typical systems that should logically interface with the rack PDU management system. The asset management system will automatically add rack PDUs, IT devices, and their associated connections to the rack PDU management system's inventory of managed devices. Integration with an enterprise reporting system enables the creation of custom reports with the additional ability to correlate data that exist in other systems. Finally, in recent years, a more comprehensive class of products called the Data Center Infrastructure Management (DCIM) have been introduced that normally include the aforementioned functions, overall capacity planning tools, and more.

16.4 CONSIDERATIONS FOR PLANNING AND SELECTING RACK PDUs

The following paragraphs in this section will address some of the basic considerations and options you will have when designing and deploying your rack PDU system.

16.4.1 Power Available and Distributed to Racks

There are several approaches to deploying power to ITE racks that affect rack PDU selection and configuration. Some approaches provide degrees of redundancy and hence higher reliability/availability than others but may not be appropriate for certain types of equipment. Redundancy and higher availability require resources, so managers of data centers that have limited power resources need to decide what ITE justifies redundant power, for example, production servers, and what equipment does not, for example, nonproduction equipment being tested or evaluated.

16.4.1.1 Single Feed to Single Rack PDU

The simplest power deployment to an ITE rack is a single appropriately sized power feed to a single rack PDU (Fig. 16.10). ITE with one or more power supplies would plug into this single rack PDU. If that single feed or single rack PDU should fail, for whatever reason, power to the equipment in the rack will be lost. The failure could occur at the rack PDU itself or farther upstream, perhaps a main feed fails or a building PDU circuit breaker trips.

As noted earlier, the NEC requires that circuits be loaded to no more than 80% of their maximum capacity. For example, if a 30A feed and rack PDU were deployed in this configuration, the load allowed (the rated current) would be 24A (30A × 80%). The NEC would expect the feed and PDU to handle a maximum of 30A, but the circuit should be loaded to only 24A.

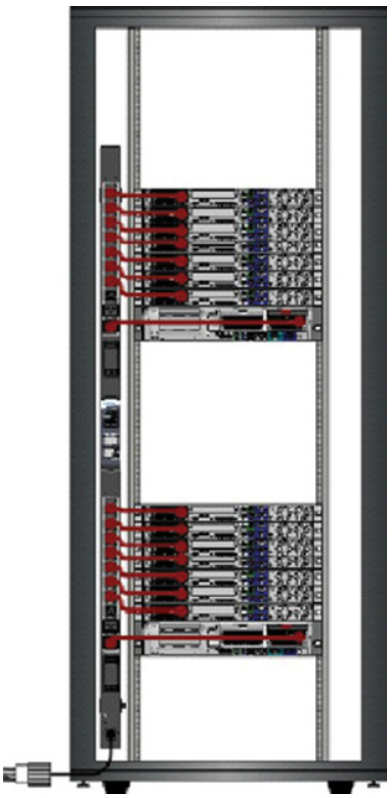


FIGURE 16.10 Single feed to single rack PDU. Source: Courtesy of Raritan, Inc.

16.4.1.2 Dual Feed to Single Rack PDU with Transfer Switch

The next step up in availability is still a single feed to a single rack PDU with the addition of a Transfer Switch, which typically has two feeds from the same or different building feeds (Fig. 16.11). If a feed to the transfer switch fails, it automatically switches to the other power feed and the rack PDU continues to power the ITE. However, if the single rack PDU fails, the power to the ITE is lost.

There are two types of transfer switches: static transfer switch (STS) and automatic transfer switch (ATS). An STS is based on static electronic component technology (silicon-controlled rectifier), which results in faster and better controlled transfer between sources. An ATS is less expensive and is based on electromechanical relay technology, which results in slower transfer times.

Again, with this arrangement, the rack PDU is still loaded to 80% of the maximum, but the electrical power capacity required has doubled—one feed is operational and the second feed is a backup. It has also doubled the amount of upstream equipment necessary to supply the additional feed. Two power feeds to an ATS and then to a

FIGURE 16.11 Dual feed to single rack PDU with transfer switch. Source: Courtesy of Raritan, Inc.

single rack PDU are generally used only where reliability is a concern but the ITE itself, for example, a server, has only one power supply.

16.4.1.3 Dual Feed to Dual Rack PDUs

Today, many servers, network devices, storage systems, even Keyboard, Video, Mouse (KVM) switches, and serial console servers are available with dual power supplies. Some larger servers may have as many as four or even six power supplies. The most reliable deployment here is to use two power feeds to two rack PDUs (Fig. 16.12). With this configuration, if one rack PDU or power feed fails, there is a second one available to maintain power to the ITE in the rack. A common practice when using dual feeds is to use rack PDUs with colored chassis such as red and blue. The colored chassis enables a visual control for installation of or changes to the PDU and connections. The rack will have a red chassis PDU fed by input circuit "A" and a blue chassis PDU fed by input circuit "B." The colored chassis helps to eliminate confusion about which PDU is fed by circuit "A" or "B."

But it is important to remember the requirement that each circuit be loaded to no more than 40%. If the two circuits feeding the rack are both loaded to 80%, the NEC requirement will be met, but what would happen if one of the circuits failed? The power demand to the second circuit would jump from 80 to 160%, and the circuit breaker for that feed would trip so the second circuit to the rack would also lose power. To prevent this, both feeds should be loaded to no more than 40% so that if one fails, the remaining circuit won't be loaded to more than 80%. Compared to the previous case with ATS in Section 16.4.1.2, where one feed is backup, in this configuration, both feeds are powering ITE.

Note that if you intend to perform remote switching for ITE with dual power supplies, you will want to use a rack PDU that supports outlet grouping; that is, two or more outlets are controlled as though they are a single outlet.

16.4.1.4 Multiple Power Supplies

ITE with two or more power supplies can vary in the way power is delivered to the equipment (Fig. 16.13). Some devices have a primary and backup power supply; some alternate between the power supplies; and some devices share power demand across all the power supplies. For example, a blade server with four power supplies in a 3 + 1 redundancy configuration would draw one-third of its power from each of its three primary power supplies, leaving one for redundancy in the event any one of the three fails. Finally, some more sophisticated devices have multiple power supplies that are designed for both redundancy and efficiency. For example, some devices might drive utilization rates higher on specific power supplies to drive higher efficiency. You will need to check with each equipment manufacturer to understand how the power supplies work so that optimal balanced load configurations can be achieved on the rack PDU, especially those with branch circuits and three-phase models.

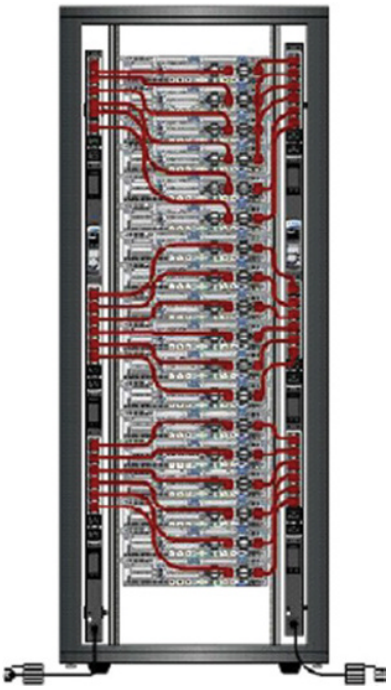


FIGURE 16.12 Dual feed to dual rack PDUs. Source: Courtesy of Raritan, Inc.

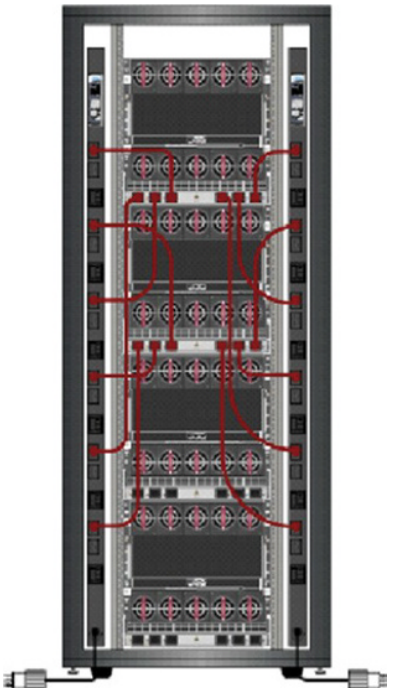


FIGURE 16.13 Multiple power supply configuration. Source: Courtesy of Raritan, Inc.

16.4.1.5 Load Balancing

Load balancing attempts to evenly distribute the rack equipment's current draw among the PDU's branch circuits so that as you come closer to perfect balance, more total current can be supplied with the greatest headroom in each branch circuit. For example, consider a PDU with two 20A circuit breaker-protected branch circuits—where each branch contains a number of outlets. The total current capacity of the PDU is 40A with the limitation that no branch circuit of outlets can exceed 20A. If the total load of all devices plugged into the PDU is 30A, perfect balance is when the load is exactly divided between the two branches (15A each branch). The headroom in each branch is then 5A (20A circuit breaker less 15A load). Any other distribution of the load (16A : 14A, 17A : 13A) results in less headroom.

Load balancing has similar benefits for three-phase PDUs. As the load comes closer to perfect balance, the current draw is more evenly distributed among the three-phase lines (more headroom), and total current flowing in the three lines is minimized. For example, consider a 24A three-phase Delta-wired PDU with three branch circuits. When an 18A load is balanced across the three branch circuits (6A load in each branch), the current flowing in each input phase line is 10.4A, and the total current in all three lines is 31.2A. If the entire load was carried by one branch circuit (totally unbalanced), the current in the three-phase lines are 18, 18, and 0A, respectively, and the total current is 36A. When the load is balanced across all three lines, the PDU has 7.6A (18.0–10.4A) more headroom.

Load balancing can be tricky because many IT devices draw power in varying amounts depending on the computational load. For devices with single power supplies, an estimate of the power consumption should be made for each device and then the devices plugged into the several circuits so that the circuits are loaded evenly. This is true both within a rack and across multiple racks. For devices with dual power supplies, they should be plugged into different circuits. A typical deployment would be the dual feed to dual rack PDUs mentioned earlier.

For IT devices with more than two power supplies, such as blade servers, load balancing can become even more complicated, especially if the rack PDUs are three-phase models.

As an example, assume four blade chassis are to be installed in a rack, each chassis has six power supplies, and two three-phase rack PDUs will be installed in the rack for redundant power. The first blade server will have power supplies (PS) #1, #2, and #3 plugged into circuits (C) #1, #2, and #3, respectively, on PDU A and power supplies (PS) #4, #5, and #6 plugged into circuits (C) #1, #2, and #3, respectively, on PDU B. Since we want to try to balance the load across all circuits and lines and we can't be sure that each of the four blade servers will be performing tasks that equally load the circuits, we will stagger the second blade server power supplies. So the second server will have PS #1 plugged into C #2, PS #2 plugged into C #3, and PS #3 plugged into C #1 on PDU A and PS #4 plugged into C #2, PS #5 plugged into C #3, and PS #6 plugged into C #1 on PDU B. Circuit-level metering, phase-level metering, and outlet-level metering will be very helpful for (re)balancing loads in the rack.

16.4.1.6 Inrush Current

Servers draw more current when they are first turned on, known as inrush current. As discussed in the section on overload protection, rack PDUs with circuit breakers are designed not to trip during very short periods of high currents. However, it is better for upstream circuits if sudden surges during equipment powering on are minimized. For this reason, some rack PDUs provide outlet sequencing and allow users to configure both the sequence and the delay time in which the outlets are turned on. Some rack PDUs may allow programming of outlet groups and allow sequencing of groups of outlets.

16.4.2 Power Requirements of Equipment at Rack

Section 16.4.1 deals with ways to deploy electrical power to a rack. This section deals with determining how much power to deploy to a rack. Typically, the starting point is an IT device's nameplate power requirement data (see Section 16.2.4.1) that specifies a voltage and current (amps), which is typically higher than what is usually seen during actual deployment. Often, a percentage of the nameplate value, for example, 70%, is used when computing the maximum PDU load capacity required: PDU load capacity = sum (or Greek letter sigma) (device nameplate in VA × 70%). For example, 208 V × 2.4 A × 70% × 14 servers = 4.9 kVA.

For the aforementioned example, if you run 208V, you need a 30A (5 kVA) rack PDU since you will load it to 80% to meet North American requirements (4.9 kVA/208 V = 23.5 A; 23.5A is approximately 80% of 30A). If you want redundancy, add a second 5 kVA rack PDU and load both PDUs up to 40%. You will need to specify the appropriate number of outlets. It is a good idea to have a few spare outlets for other devices even if the rack PDU will be at its maximum capacity. More efficient or different equipment might be installed in the rack in the future or servers may not run near full capacity, leaving additional power capacity to power more equipment. The current best practice is to standardize on IEC C-13 and/or C-19 PDU outlets and 208V. Most servers and data center devices can run at 208V (even up to 240V). Remember that the derating factor of 70% was just an estimate. Research has been done with sophisticated rack PDUs that accurately measure power at the outlet. The findings

were surprising. Even at peak power consumption, 15% of the servers drew 20% or less of their nameplate rating. Equally surprising was that nearly 9% drew 81% or more of their nameplate rating. The point here is that the actual power consumed as a percentage of nameplate rating can vary widely. Ideally, data center managers should measure the actual power consumption rather than use a rule-of-thumb average such as 70%. If the actual overall average is closer to 40%, as it was in the study, deploying power at 70% of nameplate is wasteful and strands unused power.

If a cabinet populated with 30 1U servers has dual power feeds and the servers require an average of 150 W each, then the total power requirement for a cabinet is 150 W × 30 servers = 4.5 kVA. Assuming 250 VA for additional equipment, like an Ethernet switch and a KVM switch, this brings the total to 4.75 kVA. So a 208 V, 30 A PDU, which is rated at 5 kVA, would be sufficient. Such a PDU can carry the full load of 4.75 kVA in a failover situation when the power feed to one side of the cabinet fails or is taken down for maintenance. Typically, each PDU would be carrying only 40% of the 4.75 kVA.

It is also important to note that three-phase Wye 208 V rack PDUs are able to support both 120 and 208 V in the same PDU. This can be handy for situations where a variety of equipment types with different voltage requirements need to be racked together.

16.4.2.1 208 V Single-Phase versus 208 V Three-Phase Rack PDU

In a rack of 42 1U servers, if each server consumes an average of 200 W, then the total power consumption is 42 × 200 W = 8.4 kW. To allow for the NEC requirement of 80%, the rack needs 10.5 kVA (8.4 kW/0.8). To allow for redundant power feeds, two rack PDUs able to provide 10.5 kVA are required. The 208 V single-phase at 60 A (48 A rated) can deliver 10.0 kVA. This could suffice, particularly if the 200 W per server estimate is on the high side. Another alternative is the 208 V three-phase at 40 A (32 A rated), which can deliver 11.5 kVA. The 208 V three-phase alternative provides headroom to add higher-power-demand servers in the future and can handle the existing servers even if their average power consumption increases from 200 to 220 W.

The use of three-phase power enables one whip or rack PDU to deliver three circuits instead of just one. The whip or input power cord on the rack PDU will be somewhat larger for three-phase power than single-phase power because instead of three wires (hot, neutral, and ground), a three-phase cable will have four or five wires.

The two three-phase alternatives are Delta and Wye. A three-phase Delta system will have four wires: Line 1 (hot), Line 2 (hot), Line 3 (hot), and a safety ground. Individual circuits are formed by combining lines. Three circuits are available—L1 + L2, L2 + L3, and L1 + L3. The power on each of the lines is a sine wave (this is also the case for single-phase power), but each of the three sine waves is 120° out of phase with the other two.

For three-phase power, the sine waves are 120° out of phase, so calculating VA is slightly more complex because we need to include the square root of 3, which is 1.732. The apparent power formula for three-phase is V × derated A × 1.732 = VA. As an example, 208 V, 40 A (32 A derated) three-phase is 208 V × 32 A × 1.732 = 11.5 kVA. In other words, the three-phase Delta deployment provides more than 170%, or 70% more, than the comparable single-phase, single-circuit deployment.

A three-phase Wye system will have five wires: Line 1 (hot), Line 2 (hot), Line 3 (hot), a neutral, and a ground. Individual circuits are formed by combining lines and a line with the neutral. As an example, a three-phase 208 V Wye rack PDU supports three 208 V circuits (L1 + L2, L2 + L3, L1 + L3) and three 120 V circuits (L1 + *N*, L2 + *N*, L3 + *N*). Three-phase Delta and three-phase Wye have the same apparent power, but the three-phase Wye can provide two different voltages.

In North America, there may be a requirement for 120 V convenience outlets such as NEMA 5-15R (120 V, 15 A, 12 A rated) or 5-20R (120 V, 20 A, 16 A rated). These can be supported by 208 V three-phase Wye PDUs where wiring between lines (L1, L2, L3) and lines and the neutral can provide power to both 208 and 120 V outlets. Whether the three-phase wiring is Delta or Wye, the voltage is always referenced to the line-to-line voltage, not the line-to-neutral voltage. This is even true in the following 400 V example where all the outlets are wired line to neutral.

Since the Wye system adds a neutral wire, many data centers are wired for Wye and use whips terminated with Wye receptacles such as NEMA L21-30R. This means the data center can use Wye PDUs that support 120/208 V or use Delta PDUs that support only 208 V without needing to change the data center wiring. A Delta PDU would use a NEMA L21-30P (the mating Wye plug) but would not use a neutral wire inside the PDU. This is a perfectly acceptable practice. For example, a data center could deploy Delta PDUs to racks where there is only a need for 208 V and Wye PDUs to racks where there is a need for both 120 and 208 V. Three-phase cables may be slightly larger than single-phase cables, but it is important to remember that one slightly thicker three-phase cable will be significantly smaller and weigh less than three single-phase cables for the same voltage and amperage.

16.4.2.2 Rack PDU 400 V Three-Phase

As shown in the 208/120 V example, three-phase Wye wiring is a convenient way to step down voltage. This is particularly true for 400 V power. A generally accepted method of delivering substantial power to densely packed racks is via 400 V

three-phase Wye rack PDUs. A 400 V power distribution from panels to racks is now an accepted practice. A data center designer could specify 400 V Wye whips to 400 V Wye rack PDUs. Since much data center equipment can safely operate on voltages ranging from 100 to 240 V, the 400 V Wye PDU can provide three circuits—L1 + *N*, L2 + *N*, L3 + *N*—each supplying 230 V (400 V/1.732). The 400 V Wye rack PDUs do not lend themselves to supporting 120 V outlets as do 208 V Wye rack PDUs.

16.4.3 Rack PDU Selection

16.4.3.1 Rack PDU Selection and Special Application Requirements

There are many factors involved in selecting a rack PDU. Data center location, application, and ITE requirements, available power, cabinet, energy management and efficiency objectives, etc., will combine to dictate what type of PDU should be used. Some of the considerations in the following will guide you to select the feature set and hence the type of PDU you will need to satisfy your requirements.

What is the type of equipment and how many devices are going into the cabinets, for example, 42 × 1U servers with a single feed per device versus three 10U high blade servers with six power supply feeds per server? The answer will help define the physical configuration, for example, number and type of outlets, and capacity of your PDU(s), for example, how much power (kW) the PDUs need to support.

Clearly, decision criteria for 24/7 manned sites will be different than remote management of lights-out facilities. If you need remote or lights-out management of a facility, then you will probably need a switched PDU, which will require more security and user access management. Remote applications may also call for SNMP management.

Integration with directory services like LDAP or Microsoft's Active Directory is increasingly a requirement for controlling access to resources, rather than requiring a separate access control system. This capability is applicable to all applications, requiring central authentication, local or remote. And for many data center applications, for example, federal government and financial institutions, encryption and strong password support are necessary for remote access.

The rack PDU must supply uninterrupted power to each device plugged into it. You will want to prevent or mitigate any events that can potentially cause the circuit breaker on the rack PDU or upstream to trip Outlet sequencing is a valuable feature to prevent inrush current from tripping a circuit breaker by establishing a sequence and appropriate delay for powering multiple devices. Outlet sequencing not only prevents the undesired tripping of a circuit breaker but also lets the user specify the order in which services (device(s)) come on line or are shut down during power cycling. For example, you will want to power the database service before the Web servers. This capability is most useful when used in conjunction with the outlet grouping capability (see in the preceding text).

For some applications and equipment, you many need a customizable alarm threshold for each outlet, with the capability to switch off an outlet should it exceed a certain power draw. This would prevent a temperature or other sensor (see Section 16.3.2) from causing a shutdown of servers. An advanced application is HVAC control using the temperature reported by a PDU's temperature sensor.

In many mission-critical environments, managed devices often have multiple feeds, which will be fed from different feeds or circuits for failover and redundancy. The device needs to be managed as a single device regardless of the number of power supplies/plugs, and all outlets must be handled simultaneously. This capability is applicable to all applications, local or remote.

Event-driven power cycling of an outlet/device is required for some applications, particularly for remote or unmanned sites. For example, if a device in a remote location fails to respond and the WAN is not operational, there are basically two options: first, an expensive, time-wasting truck roll to restart and, second, a rack PDU with the intelligence to trigger a restart of a malfunctioning device, for example, if the device has not responded for 20 min recycle power to the device.

If there is a need to maximize power efficiency, then rack PDUs can provide valuable data to support those efforts. Look for current, voltage, and power factor measurements at the PDU, line, breaker, and outlet level. Look for accurate kilowatt-hour metering at the outlet level, especially if you intend to report or charge back individuals or groups for usage. Metering accuracy can vary significantly, and for some rack PDUs, calculations may be based on assumptions and not actual real-time measurements.

16.4.3.2 Rack PDU Functionality

Rack PDUs can vary significantly, not only in operational functions they offer, but also in their monitoring and data collection. The following is an overview of the strengths and weaknesses of the four types/classes of rack PDUs previously defined in Section 16.2.1.1. Clearly, our class definition is not rigid, since features offered by vendors will vary and you will want to select PDUs based on the total fit to your requirement, but this can be a useful guide in your selection. The strength and weakness of different rack PDUs are explained:

Non-Intelligent PDUs

- *Strength*: A basic, proven, low-cost technology provides reliability
- *Weakness*: Lack instrumentation and are not manageable on any level

Intelligent PDUs

1. *Metered Input PDUs*
   - *Strength:* Provide real-time monitoring of PDU current draw. User-defined alarms alert IT staff of potential circuit overloads before they occur.
   - *Weakness:* Limited data, for example, no outlet level or environmental data and no outlet switching.
2. *Switched PDUs*
   - *Strength:* Offer some or all the features of metered PDUs plus remote power on/off capabilities, outlet-level switching, and sequential power-up.
   - *Weakness:* Must be managed carefully and risk of inadvertent power cycling. May not be appropriate for some environments, such as blade servers. Usually limited data, for example, no outlet-level monitoring for critical environmental data.
3. *Switched PDUs with Outlet Metering*
   - *Strength:* State-of-the-art devices that are remotely accessible via web browser or log on to *CLI (Command Line Interface)*. Models include all features of switched devices (though they may be switched or unswitched) plus outlet-level monitoring, standard-based management, integration with existing directory servers, enhanced security, and rich customization. Provide comprehensive data including current, voltage, apparent power, active power, real-time environmental data, and often real-time kilowatt-hour (kWh) metering.
   - *Weakness:* Higher initial cost due to their greatly enhanced feature set.

16.4.3.3 Benefits of an Intelligent Rack PDU

The IT industry has dramatically chosen to move to more sophisticated, manageable systems. This fact is no more in evidence than the dramatic trend to the use of intelligent PDUs.

A truly intelligent PDU will provide real-time outlet-level and PDU-level power monitoring, remote outlet switching, and rack temperature and humidity monitoring. For top-tier data centers, deployment of intelligent PDUs can make a significant difference in the ability of IT administrators to improve uptime and staff productivity, efficiently utilize power resources, make informed capacity planning decisions, and save money. And, in so doing, they will operate greener data centers. Clearly, if your data center has dozens of racks, then the greatest benefits will be realized by using a rack PDU management system to consolidate data acquisition, reporting, as well as PDU administration and control. Here are a few practical reasons to be selecting intelligent PDUs for your racks:

Improve Uptime and Staff Productivity

- Monitoring power at a PDU and individual outlet level, with user-defined thresholds and alerts via e-mail or SNMP, provides awareness of potential issues before they occur.
- Remote reboot of servers and ITE from anywhere in the world via a Web browser reduces downtime and personnel costs.

Use Power Resources Safely

- User-configurable outlet-level delays for power sequencing prevent circuits from tripping from ITE inrush currents.
- Control of outlet provisioning prevents accidently plugging ITE into circuits that are already heavily loaded and are at risk of tripping circuit breakers.

Make Informed Power Capacity Planning Decisions

- Outlet-level monitoring may identify some simple rearrangements of equipment to free up power resources by balancing power demands across racks.
- Monitoring power at the outlet level can identify equipment that may need to be changed to stay within the margin of safety of defined thresholds.
- Monitoring rack temperature and other environmental conditions can prevent problems, especially when a data center is rearranged and airflow patterns change.

Save Power and Money

- Monitoring power at the outlet level combined with trend analysis can identify ghost or underutilized servers that are candidates for virtualization or decommissioning.
- Remote power cycling enables IT managers to quickly reboot hung or crashed ITE without incurring the cost of site visits.
- Temperature and humidity sensors help data center managers optimize air-conditioning and humidity settings and avoid the common practice of overcooling and related waste of energy.

16.4.4 Power Efficiency

16.4.4.1 PUE Levels

The Green Grid defines three levels of PUE: Basic or Level 1, Intermediate or Level 2, and Advanced or Level 3. Many industry analysts recommend measuring IT power consumption at the Intermediate, Level 2, that is, at the PDU level. While it is true that PDU-level power consumption will provide the denominator needed to calculate PUE, this information alone is unlikely to be sufficient to drive the best

efficiency improvement decisions. Regardless of the PUE level you choose to employ, the best practice is to gather data over a time period of "typical" power usage to ensure that the peaks and valleys have been captured in calculating your PUE to establish a baseline and to track your improvements. There are many tools for collecting the data you need, described elsewhere in this book.

16.4.4.2 Why Advanced Level 3 PUE?

An improved (lower) PUE can be misleading since that can result from inefficiencies in the power consumed by ITE, which merely increases the denominator. A lower PUE is generally better than a higher one, but it is possible to implement measures that reduce data center energy consumption yet actually increase your PUE. For example, if you were to replace older, less efficient servers with more efficient ones, or eliminate ghost servers, or turn off servers that were idle during the night, or employ server virtualization, the net result is power reduction, but your PUE would actually increase. The detailed IT load data from Level 3 provides the granularity of information to reduce energy consumption, not just improve the PUE metric. Clearly, the PUE (and its inverse DCIE) becomes a more useful beacon once you have built efficiency into the ITE performance, and to do that, you will want the granular power usage data for the Advanced, Level 3 PUE metric. Then you can attack the numerator and squeeze inefficiencies out of the infrastructure.

16.4.4.3 The Advantages of High Power

A single-phase 120 V at 100 A (80 A rated) circuit provides 9.6 kVA. A single-phase 208 V at 60 A (48 A rated) circuit provides 10.0 kVA. A three-phase 208 V at 40 A (32 A rated) circuit provides 11.5 kVA. A single-phase 230 V at 60 A (48 A rated) circuit provides 11.0 kVA. A three-phase 400 V at 20 A (16 A rated) circuit provides 11.1 kVA.

Running higher voltages at lower currents means smaller cables that use less copper, weigh less, take up less space, and cost less. Running three-phase power instead of single-phase power means fewer cables, which simplifies deployment as well.

Plugs and receptacles are also less expensive at higher voltages and lower current ratings. For example, a 30 A 400 V three-phase Wye (16.6 kVA) plug (Hubbell NEMA L22-30P) costs $32 and the receptacle costs $41. A 60 A 208 V three-phase Delta (17.3 kVA) plug (Mennekes IEC309 460P9W) costs $166 and the receptacle costs $216. The plug/receptacle combination is $73 versus $382.

There are other benefits to higher voltages. A 400 V power circuit will eliminate voltage transformations and can reduce energy costs by approximately 2–3% relative to 208 V distribution and approximately 4–5% relative to 120 V distribution. Consolidating data centers will generally reduce total power consumption but may create opportunities for use of high-density racks and high-power rack PDUs. For example, a 42U rack filled with 1U servers consuming 250 W each draws 10.5 kW, which would require two three-phase 208 V, 50 A circuits providing 14.4 kVA each. Taking advantage of blade servers might lead to deploying five blade chassis in one rack, which would require two three-phase 208 V, 80 or 100 A or two three-phase Wye 400 V, 50 or 60 A rack PDUs. These examples allow sufficient headroom should one of the feeds fail. They also support the North American requirement for 80% derating.

High-density racks can be deployed in small, medium, or large data centers. Even small data centers benefit from high-power racks for multiple blade servers or densely packed 1U servers.

16.5 FUTURE TRENDS FOR RACK PDUs

Two primary forces are influencing rack PDU development and innovation trends. First is the demand for increasing power and density of ITE at the rack or compute density per U of rack space. Second is the industry-wide goal, even mission, to create energy-efficient (often called "green") data centers, including carbon footprint reduction. Both trends challenge the PDU vendors to improve both hardware and software design; and the second requires all IT and facilities organizations to better understand how the data center power is consumed and take active measures to reduce it.

The aforementioned trends are underlined in the IHS 2018 survey and report on the World Power Distribution Unit Market. The report shows a healthy 5.4% compounded annual growth rate (CAGR) for overall rack PDU revenue from 2017 to 2022. However, they project 9.0% CAGR for three-phase PDUs versus 4.3% CAGR for single-phase PDUs and predict 6.1% CAGR for PDUs with intelligence and 4.4% CAGR for basic PDUs. The "high impact" factors that are driving PDU demand are increasing power consumption and higher densities along with increasing need for PDU intelligence.

16.5.1 Higher-Density, Higher-Power Rack PDUs with Sensors

The growing popularity of 1U servers, blade servers, network-attached storage, storage area networks (SANs), and multi-gigabit, chassis-based network communications gear places enormous demands on rack PDUs. For example, four blade server chassis in a single rack could draw in excess of 20 kW of power creating power and cooling challenges for data center managers. From a power perspective, racks will

require three-phase power with 60, 80, even 100 A of service. There are some data centers bringing 400 V three-phase service to the rack, to accommodate power demand while increasing efficiency from reduced voltage step-downs. Similarly, end users are packing dozens of 1U servers into a single rack and pressing rack PDU vendors to support 40+ outlets and 20+ kW.

Server virtualization is a major trend in data centers, leading to improved efficiency and cost reduction. However, running multiple virtual machines on one server will drive up its total power consumption; and a rack containing several such servers could experience a lot more power consumption driving the need for additional power load visibility to optimally manage power capacity.

More power consumption means more cooling to remove the additional heat. PDU vendors will be expected to supply the basic environmental sensors for heat, humidity, and airflow to help understand the overall environmental conditions and to identify zones that must be fine-tuned or supplemented with dedicated or specialized cooling.

16.5.1.1 Customizing ITE for Power Efficiency

One trend to watch is the design and deployment of custom servers, power supplies, rack PDUs, etc., to maximize power usage efficiency. For example, Facebook along with Open Compute has begun to deploy 480 V three-phase Wye power where each line is wired to the neutral so the outlets deliver 277 V. This Wye configuration with lines wired to the neutral is the same wiring configuration as the 400/230 V wiring described earlier. This approach is very efficient, but it is highly customized since most ITE today are not built with power supplies that support 277 V. Furthermore, common data center receptacles are IEC C-13 and C-19, which do not support 277 V.

The savings and efficiencies (1–2% over 400/230 V three-phase systems) are sufficient that Facebook/Open Compute can justify building custom triplet racks, custom servers with custom power supplies, custom battery/UPS, and 480/277 V rack PDUs with custom Tyco 3-pin Mate-N- Lock outlets.

16.5.2 Increased Intelligence at the Rack to Support Efficiency Initiatives: "Smart Rack"

Many data centers have grown larger and more complex in recent years as the consolidation trend continues. With increasing size and complexity, there is a greater need to drive intelligence to the ITE at the rack to create what industry people are beginning to think of as the "Smart Rack."

Every data center, regardless of size, is designed to support the servers at the rack where the actual computing is taking place. It is also where the vast majority of the power is being consumed. Proper monitoring and metering of the ITE along with environmental sensors at the rack will collect the data necessary to produce the most significant overall efficiency, savings, and operational improvement. Collection and analysis of actual energy data will enable you to maximize the use of current resource capacity and take advantage of the capacity of planning tools to "right size" the data center for future requirements. This will allow you to eliminate or defer capital expenses of data center expansions while improving day-to-day energy efficiency and overall IT productivity.

Capacity planning based on nameplate data is no longer sufficient. Efficiency improvement is an information-driven activity. In order to formulate and drive the most effective decisions, you will need to collect IT device CPU utilization and their corresponding actual power usage. More energy efficiency will be gained if such planning is based upon the trends observed from the actual data over time. Furthermore, the actual data collected at the rack level can be integrated with the overall data center infrastructure management (DCIM) systems and data center energy management systems for complete data center and power chain visualization, modeling, and planning, which can lead to further improvements in the data center ecosystem, for example, computing carbon emissions generated by IT devices to report on and take steps to lower your carbon footprint.

Efficiency can also be gained from software that offers policy-based power control to automatically turn servers on and off based on granular power consumption data and a set of preestablished static or even dynamic rules. These power-saving applications can be found in development labs, Web server farms, and cloud computing environments. They enter into mainstream data centers where the deployment of intelligent PDUs will enable their functions.

Creating energy-efficient behavior throughout your organization is a key factor in reducing waste and costs; and the essential ingredient to affect behavior is individual awareness/accountability for energy usage. Of course, to be effective, any such energy reporting or charge-back system must be based on credible, comprehensive, and coherent usage data, so PDU vendors will be expected to deliver the highest accuracy for energy usage at every level of the organization.

16.5.3 Integration with Higher-Level Data Center Management Systems

In recent years, a variety of software products have been introduced to help both IT and facilities people manage the data center. While the category name may differ—Physical Infrastructure Resource Management (PRIM), DCIM, Data Center Service Management (DCSM)—these applications provide most of the following major functions: database of all physical data center assets with detailed data for IT, power and HVAC equipment, physical

data center layout, and cable connections; change management; 2D or 3D visualization of the data center building with drill down to lowest-level data element; and capacity planning based on availability of floor and rack space, power, cooling, etc.

The data required to manage data center infrastructure and energy effectively are collected from power devices along the entire power chain up to the IT devices, from the IT devices themselves, from the environmental sensors, and from data center layout maps, cable plans, and cooling system design documents. The more data collected, and the more accurate the data will be, the better the data center personnel are enabled to manage the data center to support critical IT operations reliably, efficiently, and cost-effectively.

The following is a simplified view of data measurement, collection, compilation, analysis and correlation, and decision support:

- Intelligent rack PDUs measure essential power data at a predefined frequency and store such data in memory.
- Data collection from rack PDU management (or power management) system polls the intelligent PDUs through industry standard management protocols such as SNMP.
- The data collection service can be part of the intelligent PDU vendor's rack PDU management system (Raritan's Power IQ is an example), or it can be part of a DCIM (Vertiv Trellis is an example) or energy management system (Raritan's Power IQ is an example). For scalability reasons, data collection is typically delegated to specific PDU vendor's PDU management system, which is deployed along with intelligent PDUs to administer, maintain, and troubleshoot the PDUs, as well as to collect power statistics from these intelligent PDUs.
- The rack PDU management system can use the collected data to perform first level of analysis. This will help to visualize the power trends and pinpoint some potential issues. Then the collected data as well as the compiled information can be used by DCIM or energy management system for further analysis.
- The energy management or DCIM system has visibility beyond the PDU management system. They can, for example, poll information from upstream smart power devices; and they typically also feature the static information such as data center physical layout, cable plan, and HVAC deployment information, making them more suitable for analysis that must take into consideration many more factors beyond the intelligent PDUs.

With the advanced analysis conducted by a DCIM or energy management system, data center management staff can make their day-to-day operational decisions as well as longer-term strategic planning, to provide reliable and high-quality power for business applications while reducing waste in data center energy consumption.

FURTHER READING

Alger D. *Build the Best Data Center Facility for Your Business: A Comprehensive Guide to Designing and Operating Reliable Server Environments*. Indianapolis: Cisco Press; 2005.

ASHRAE. *Thermal Guidelines for Data Processing Environments*. 2nd ed. Atlanta: ASHRAE; 2004.

ASHRAE. *ASHRAE Workshops on Improving Data Center Energy Efficiency and Best Practices*. NYSERDA Sponsored Workshop. New York: ASHRAE; November 6, 2008a.

ASHRAE. *2008 ASHRAE Environmental Guidelines for Datacom Equipment*. Atlanta: ASHRAE; 2008b.

ASHRAE. *High Density Data Centers: Case Studies and Best Practices*. Atlanta: ASHRAE; 2008c.

ASHRAE. *Best Practices for Datacom Facility Energy Efficiency*. 2nd ed. Atlanta: ASHRAE; 2009.

Cuthbertson D. *Practical Data Center Management Training Workshop, Part 1—Managing the Facility*. Workshop; Somerset: Square Mile; July 24, 2008.

Data Center Users Group, Emerson Network Power. *Data Center Users' Group Special Report: Inside the Data Center 2008 and Beyond*. Columbus: Data Center Users Group, Emerson Network Power. White Paper WP165-118, SL-24634; 2008.

Digital Realty Trust. *kW of IT Load. The New Chargeback Mechanism*. San Francisco: Digital Realty Trust.

Frost and Sullivan. *Worldwide Power Distribution Unit Market, N2fE-27*. Rockville: Frost and Sullivan; December 2008.

Haas J, Monroe M, Pflueger J, Pouchet J, Snelling P, Rawson A, Rawson F. *Proxy Proposals for Measuring Data Center Productivity*. Beaverton: The Green Grid. White Paper #14; 2009.

Information Technology Equipment—Safety—Part 1: General Requirements. Northbrook: Underwriters Laboratories, Inc.; 2007. UL 60950-1.

NFPA 70: National Electric Code. Quincy: National fire Protection Association; 2008.

Raritan Inc. *Data Center Power Overload Protection: Circuit Breakers and Branch Circuit Protection for Data Centers*. Somerset: Raritan Inc. White Paper; 2009a. Available at http://www.raritan.com/resources/white-papers/power-management/. Accessed on May 22, 2014.

Raritan Inc. *Data Center Power Distribution and Capacity Planning: Understanding What You Know—and Don't Know—About Power Usage in Your Data Center*. Somerset: Raritan Inc. White Paper; 2009b. Available at http://www.raritan.com/resources/white-papers/power-management/. Accessed on May 22, 2014.

Raritan Inc. *Power Distribution Units (PDUs): Power Monitoring and Environmental Monitoring to Improve Uptime and*

Capacity Planning. Somerset: Raritan Inc. White Paper; 2009c. Available at http://www.raritan.com/resources/white-papers/power-management/. Accessed on May 22, 2014.

Raritan Inc. *Deploying High Power to IT Equipment Racks*. Somerset: Raritan Inc. White Paper V1156; 2012.

U.S. Environmental Protection Agency. *Report to Congress on Server and Data Center Energy Efficiency Public Law 109–431*. Washington: U.S. Environmental Protection Agency, ENERGY STAR Program; 2007.

Verdun G, editor. *The Green Grid Metrics: Data Center Infrastructure Efficiency (DCiE) Detailed Analysis*. Beaverton: The Green Grid. White Paper #14; 2008.

Wikipedia. "1-Wire," of Dallas Semiconductor Corp. Available at http://en.wikipedia.org/wiki/1-Wire. Accessed on May 16, 2020.

17

FIBER CABLING FUNDAMENTALS, INSTALLATION, AND MAINTENANCE

ROBERT REID
Panduit Corporation, Tinley Park, Illinois, United States of America

17.1 HISTORICAL PERSPECTIVE AND THE "STRUCTURED CABLING MODEL" FOR FIBER CABLING

17.1.1 What Is Point-to-Point (PtP) Cabling

Point-to-point (PtP) cabling refers to a data center (DC) cabling system comprised of "jumper" fiber cables that are used to connect one switch, server, or storage unit directly to another switch, server, or storage unit. A PtP cabling system is adequate for a small number of connections. However, as the number of connections in a data center increases, PtP cabling lacks the flexibility necessary when making additions, moves, or changes to data center infrastructure.

When the data centers were first built, end user terminals were connected via PtP connections. This was a viable option for small computer rooms with no foreseeable need for growth or reconfiguration. As computing needs increased and new equipment was added, these PtP connections resulted in cabling chaos with associated complexity and higher cost. Therefore, there is a downside to PtP cabling. However, the PtP cabling is surfacing again with the use of top of rack (ToR) and end of row (EoR) equipment mounting options. ToR and EoR equipment placement rely heavily on PtP cables, which can be problematic and costly if viewed as a replacement for standards-based structured cabling systems (Fig. 17.1).

17.1.2 What Is Structured Cabling

As it has been mentioned before, PtP cabling had aroused many problems. In response, data center standards like TIA-942 and ISO 24764 recommended a hierarchical structured cabling infrastructure for connecting equipment. Structured cabling is a comprehensive network of cables, equipment, and management tools that enables the continuous flow of data, voice, video, security, and wireless communications (Fig. 17.2). Instead of PtP connections, structured cabling uses distribution areas that provide flexible, standards-based connections between equipment, such as connections from switches to servers, servers to storage devices, and switches to switches. Structured cabling is designed to meet Electronic Industries Association/Telecommunications Industry Association (EIA/TIA) and American National Standards Institute (ANSI) standards related to design, installation, maintenance, documentation, and system expansion. This helps to reduce costs and risk in increasingly complex information technology (IT) environments.

17.1.3 Comparison Between Point-to-Point and Structured Cabling

Traditionally, PtP cabling has been used in the manufacturing sector to establish a direct connection between devices and automation and control systems. However, PtP cabling lacks the flexibility, reliability, manageability, and performance required for the exploding number of connections within today's networks.

Structured cabling provides the flexibility that PtP does not, as well as the capability to support future technologies, faster connections, and more intelligent networks. Although structured cabling has long been the preferred approach in IT, we cannot deny PtP cabling completely. Here, the pros and cons of selecting a structured cabling implementation versus PtP implementation are listed in Table 17.1.

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

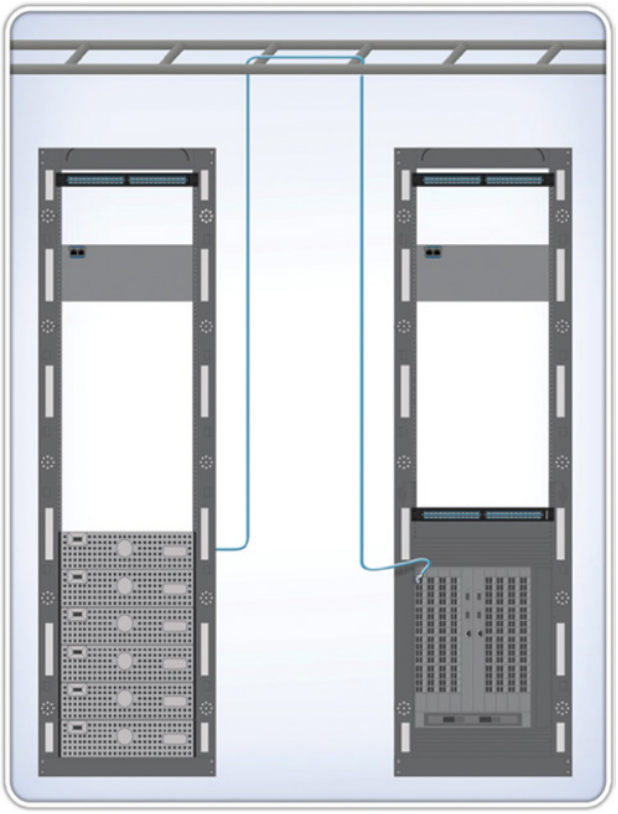


FIGURE 17.1 "Unstructured" cabling with PtP wiring.

17.2 DEVELOPMENT OF FIBER TRANSPORT SERVICES (FTS) BY IBM

Fiber Transport Services (FTS) was an IBM Global Services (IGS) connectivity system that provided a structured, flexible, and modular fiber cabling system. It incorporated planning and design, fiber trunking components, and installation activities, all performed by IBM personnel. It consisted of a variety of connectivity solutions including small form factor (SFF) connectivity and array solutions primarily MPO (multi-fiber push-on) connector-based and modular panel-mount connectivity solutions.

FTS was first to deploy multi-fiber trunks and modular fiber breakouts for the central patching locations (CPL) and zone patching locations (ZPL). Patch panels in FTS are modular panels of duplex fiber couplers connected to rear MPO connectors for attachment to MPO multi-fiber trunk cables.

The FTS system organizes fiber cabling for the large-system data center and network. It is based on a quick connect/disconnect trunking strategy using the MPO connector as shown in the Fig. 17.3.

It is possible to attach 12 fibers (six duplex channels) with a single connection operation. The MTP connector enables FTS to be disconnected and relocated very quickly. This system became the core structured cabling model for modern "plug-and-play" data center fiber connectivity systems and the practical basis for the cable models espoused in ANSI TIA 942 "Telecommunications Infrastructure Standard for Data Centers."

17.2.1 Discussion

PtP cabling, common in many DC environments (particularly SANs), are the least expensive and most available solutions (many vendors/cord houses can rapidly supply duplex jumpers from stock). This solution also exhibits the lowest channel loss because there are no physical interconnect points in the channel.

TABLE 17.1 Scorecard of DC cabling methodologies.

Primary considerations	Structured cabling	Point-to-point cabling
Meet design specifications	• High cable density—Many cables from panel to panel • Testability at the panel can provide assurance for commissioning new ports and may yield potentially longer warranty terms	• Low cable density—Few cables from panel to equipment • Ring or linear topology for reach beyond 100 m where distance between connections is <100 m • Point Coordination Function for long reach or noise mitigation
Network longevity (future proof)	• Designed in spare ports (no need to re-pull new cables for "adds") • Fiber backbones with higher grade fiber such as OM3 or OM4	• Impractical to have spare cable runs laying loose and/or unprotected • Higher performance with fewer connectors
Maintainability (moves, adds, changes)	• Environments with multiple changes occurring • Cable slack is required	• Environments with minimal changes occurring • Slack cabling is undesired, and precise cable lengths are required
Installation	• Multiple points of connectivity • Backbone and horizontal cabling is largely untouched	• Quick installation • Use where tight bends or moderate flexing is required • Use in areas where it is impractical or impossible to mount a patch panel or other cable connector interface

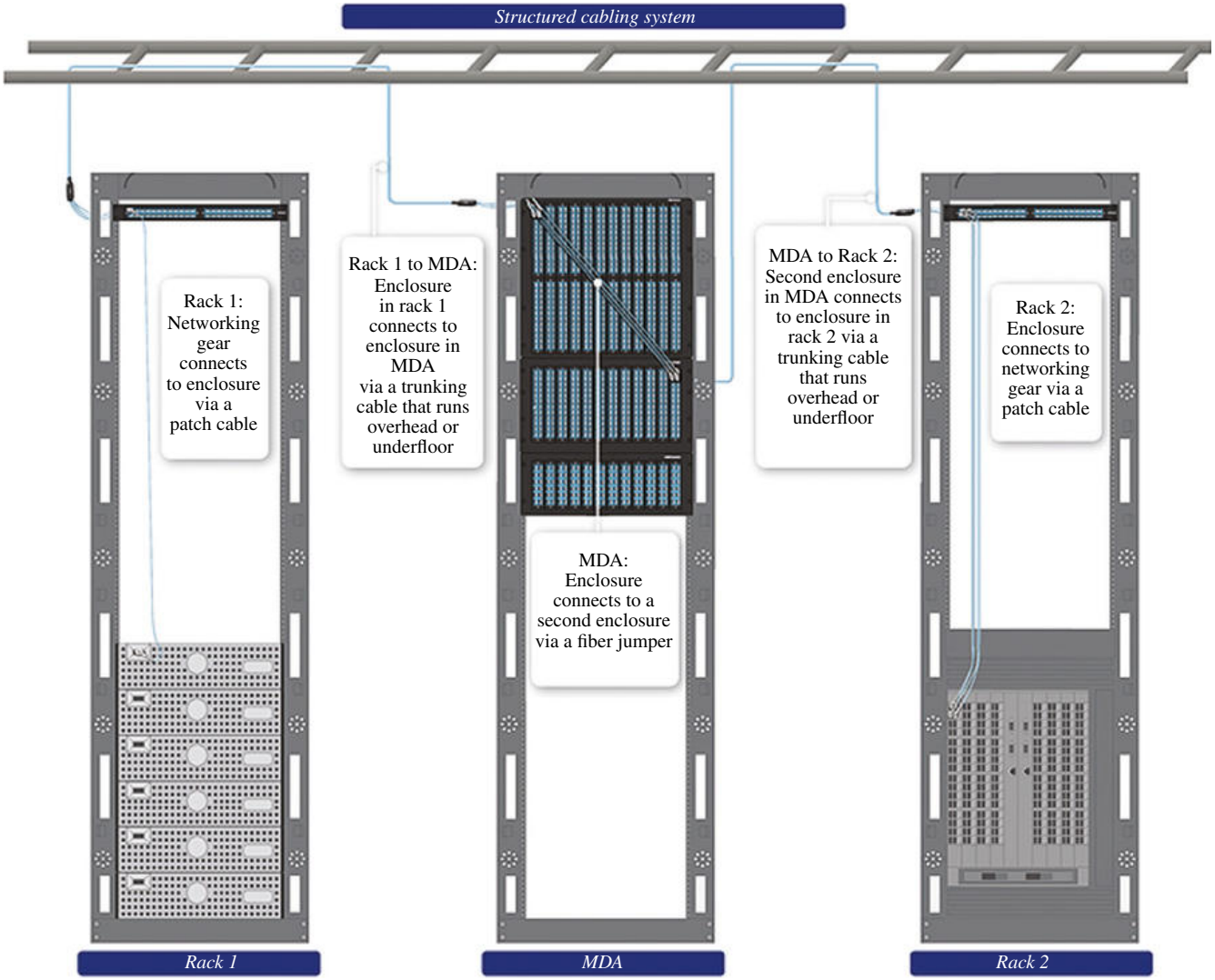


FIGURE 17.2 Structured cabling model.

The main disadvantage of this approach is that it is an "unstructured" cabling model and that incremental and organic growth in equipment is difficult.

Moves and changes on the equipment create congestion/confusion, and potential identification issues as many individual cables must be disconnected and rerouted through the vertical pathways and horizontal managers on the frame.

PtP cabling based on duplex zipcord is the most difficult to install as individual cables must be handled, identified, and placed or pulled through pathways.

Congestion issues are amplified on the switch side as this model is usually propagated there. Many individual cables coming from host equipment arrive at the director chassis en masse and must be managed in close proximity to the switch. In such high-density fabrics, it can be very difficult to locate and alter cables. Discreet zipcord cables can obscure equipment LEDs and make it difficult to remove neighboring blades. Cabling done in this way can also block airflow or access to the power supply chassis in the director/switch.

Duplex zipcord cable is not robust, and long runs of this type of media increase risk of physical compromise during installation and in situ.

PtP cabling systems usually consist of many cable part numbers (depending of design lengths), and cable length manufacturing tolerance can create differences in length going to and from the same locations (even in cables with the same part number). In this case, slack cable "storage" is usually done (by necessity) in and around the vertical and horizontal managers.

More advanced systems for fiber horizontal cabling between switches can be built with array-based (MTP) factory pre-terminated "trunk" assemblies that are mated into MTP/MPO adapter panels or breakout cassettes on each side of the permanent link. These solutions present the absolute

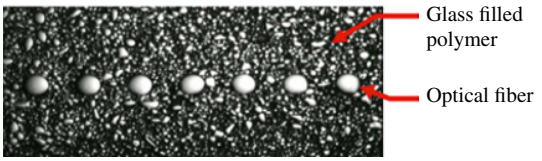
 MPO connector	The MPO connector family is defined by two existing standards. Internationally the MPO is defined by IEC-61754-7. In North America the MPO is defined by TIA-604-5 Also called FOCIS 5 (fiber optic connector intermateability standard) The MTP® brand multi-fiber connector is the trademarked name for US Conec's MPO connector. The MTP® connector is fully compliant with both FOCIS 5 and IEC-61754-7. The MTP® connector is fully intermateable with any FOCIS 5 or IEC-61754-7 compliant MPO connector
   *MPO connector ferrule detail*	• Rectangular shape with guide pin holes (female ferrule)/pins (male ferrule) • Monolithic, high precision, molded component (2–72 fibers) • Highly (60–80%) glass filled engineering polymer (typically polyphenylene sulfide) • Large amount of glass filler improves thermal stability relative to glass. Compare with typical single fiber ferrules which are ceramic or zirconia

FIGURE 17.3 MPO connector families.

minimum cabling cross section in the vertical managers and in the pathways serving to reduce vertical and horizontal pathway space.

The main benefit for the customer is a modular array connector solution that allows easy connection to storage systems (plug and go). MTP connector fiber capacity will be sized to match the quantities of transceiver groups with the hosts (Base 2–4, 8, 16 fibers for both SAN and HPC environments). Such modularity (which is native to the design philosophy of newer host systems) allows the customer to scale host capacity by adding MTP trunking assemblies in the correct Base2 fiber increments.

Cabling is among the most important considerations for organizations managing a data center, and hence investing in the right technologies to enable flexibility and optimal performance is key. Although there are several instances where PtP ToR or EoR connections make sense, an overall study that includes total equipment cost, port utilization, maintenance, and power cost over time should be undertaken. This should involve both facilities and networking personnel in order to make the best overall decision.

The most apparent benefit of a structured cabling system is the large reduction in the number of individual fiber cables under the raised floor or in overhead pathways. Fewer individual cables make tracking the cable destinations easier and facilitate the tracing of a fiber links during problem determination.

A structured cabling system provides for cost-effective moves, adds, and changes (MACs) while minimizing disruption in cabling pathways. With individual hardware devices such SAN directors and core and spine switches aggregated through local patch panels (port replication), hardware connectivity is handled in the vicinity of the hardware/patch panels, rather than at a distant location. Structured cabling provides for faster, less disruptive installation and removal of equipment, ease of reconfiguration of equipment, and more efficient use of pathways (potential to improve air movement).

Overall, PtP cabling can present data center many problems. Structured cabling is a better choice over PtP cabling.

17.3 ARCHITECTURE STANDARDS

17.3.1 DC Cabling Standards (TIA, IEC, BICSI)

17.3.1.1 ANSI/TIA-942-A

"Infrastructure Standard for Data Centers"—This standard references the TIA-568 series of the standards but also contains information appropriate for data centers. It provides an

outline of the specific functional areas of the data center and provides recommendations for pathways and spaces, backbone and horizontal cabling elements, cable management, resilience, and considerations for environmental controls.

17.3.1.2 CENELEC EN 50173-5 Information Technology

"Generic Cabling Systems Part 5: Data Centers"—This European Union (EU) standard is harmonized with TIA-942 and specifies requirements for data center cabling support existing and emerging applications.

17.3.1.3 ISO/IEC 24764 Information Technology

"Generic Cabling Systems for Data Centers"—This international standard specifies cabling for use within the data center (based on both TIA-942 and EN 50173-5). It references the cabling requirements of ISO/IEC 11801 and provides guidance and requirements for data centers.

17.3.1.4 ANSI/BICSI 002-2019

"Data Center Design and Implementation Best Practices"—This document is not so much a standard as it is a data center design and operation guide. This document describes planning, construction, commissioning, management/maintenance, cabling infrastructure, pathways, and spaces. It also gives guidance regarding modular data centers and includes a class structure of availability for determining reliability.

17.3.2 Generic DC Cabling System Elements

The TIA-942 "Telecommunications Infrastructure Standard for Data Centers" addresses cabling infrastructure in data centers and defines the functional elements of the cable plant that determines the placement of cabling, components, and equipment.

Following TIA-942 for functional layout is a first step in establishing a well-designed cabling data center infrastructure physical layer. The main cabling elements of TIA-942 are:

- Entrance room (ER)
  ERs provide a location for carrier equipment, demarcation points for outside plant cabling and can serve as "Meet-Me Rooms" and cross connect of fiber cabling delivered to Data Center data halls.
- Main distribution area (MDA)
  Houses the main cross connect (MC) (CPL), core routers/switches, and SAN directors.
- Horizontal distribution area (HDA)
  Houses cross connects and active equipment (LAN, SAN, KVM switches) for connecting to the equipment distribution area (EDA).
- Zone distribution area (ZDA)
  Optional ZDA acts as a consolidation point for horizontal cabling between the HDA and EDA(s).
- Equipment distribution area (EDA)
  The EDA is where cabinets and racks house end equipment (servers) and where horizontal cabling from the HDA is terminated at patch panels.

17.3.2.1 Centralized Patching Location (CPL)

As data center switching gear increases density and port counts increase to hundreds and thousands, the management of cables connected to these devices becomes a challenge. Historically, direct connection of cables to individual switch ports on lower density equipment was considered manageable. Applying this practice to newer high-density switch equipment makes the task difficult, making it nearly impossible to add or remove cables that have been connected directly to the equipment ports and whose cables may be buried under others in serpentine pathways across the data center.

Without proper physical infrastructure planning, execution, and best practices, customers may experience network performance issues because the physical infrastructure impacts the performance of critical business applications.

Structured cabling uses optical fiber connector enclosures connected through permanent links of optical fiber cable, typically configured in a hierarchical topology connecting the various cabling areas within the data center (network and SAN). Utilization of pre-terminated MPO/MTP trunking cables from these areas to a central patching area provides cabling infrastructure in which ports from any device can be connected to any other port.

Customers should perform MACs in high-density switch implementations at the CPL or cross connect. All physical fiber ports on all modular transceivers within the switch line cards are represented ports on a front side of a patch panel located at the CPL.

When additional chassis or line cards are added to the network, they are connected via appropriate fiber assemblies to a new/existing patch panel located at the CPL facility. The switch-to-switch and switch device connections occur using short jumpers at the CPL to structured cabling reaching out into the data center.

In a CPL design, SFF cabling solutions (harness breakout assemblies) coupled with array connector (MTP) patch panels minimize physical cable congestion on the front of the switch blade. Hydra assemblies on the switch side connect to MTP trunking assemblies through patch panels in the switch rack. These (typically short) trunking assemblies connect the switching fabric through a modular fiber cassette equipped MC (or CPL) to the HBAs/NICs present in host equipment present in the server EDA.

LAN/SAN cabling consists of the connectivity used to connect the switch ports to hosts (storage/servers) (Fig. 17.4). The connectivity equipment shall include the following options:

1. Modular cassette and MTP trunk management solutions
2. High-performance/density connectivity to support higher speed channels
3. High-density port-mapping functionality to common OEM blades

High density is a CPL requirement particularly in market segments where floor space is either predefined or otherwise limited (read brownfield and technology change) or where said is expensive on a volumetric basis. Anticipation of future connectivity (more fiber, breakouts, etc.) and media systems that require forethought with respect to cabling pathways (trunk and patch cord cross section) and static long life cycle infrastructure weigh heavily here.

There is a preference to replicate and "map" HD ports off of a high-density core switch or SAN director to a "safer" and more manageable area in CPL. The majority of higher speed applications, for instance, are those where ports on an HD switch get replicated to lower speed ports through a breakout assembly that "map" the switch ports logically in the CPL.

Risk reduction indicated this as a primary driver in selecting products for patch field MAC work at a CPL. The cost of mistakes (depending on vertical market) can be high. Network resiliency planned for through switch A/B segregation and separate cable runs etc. should be reflected at the patch field and should mitigate such risk (disturbing adjacent circuits associated with redundancy). For many users, movement of patching areas and planning for cable movement are not preferred.

17.3.2.2 Managing Horizontal Cabling and the MDA

Many DC deployments today use a modified topology described in TIA-942 wherein the HDAs collapse into the MDA using port replication cross connect. In such collapsed architectures, fiber distribution elements (cabling and connectivity) are installed between the MDA and multiple EDAs.

This horizontal cabling provides the connection between the HDA and EDA and SAN, including the optional ZDA.

There are several recognized horizontal fiber cabling media solutions:

- 850 nm laser-optimized 50/125 μm multimode fiber cable OM3 or OM4 (ANSI/TIA-568.3-D), with OM4 recommended; OM5 fiber solutions to support SWDM applications
- Single-mode optical fiber cable (ANSI/TIA-568.3-D)

Fiber cabling systems deployed today should be selected to support future data rate applications, such as 100G/400G Ethernet and Fiber Channel ≥32G. For multimode solutions, OM3 is a starting point. In addition to being the only multimode fibers included in the 40G and 100G Ethernet standard, OM3 and OM4 fibers provide high performance as well as the extended reach often required for structured cabling installations in the data center.

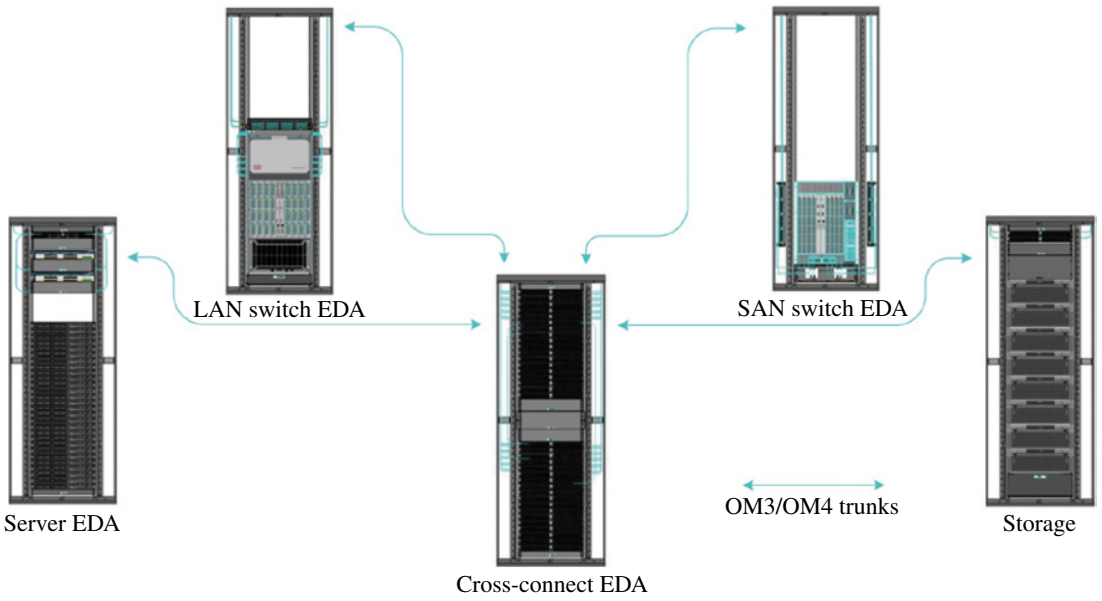


FIGURE 17.4 LAN/SAN cabling connects switch ports to hosts (storage/servers).

The MDA in a data center includes the MC, which is the central point for distribution of the structured cabling system. An SFF and high-density cabling system can be designed into this distribution area with modular cassettes (or port-mapping cassettes) and pre-terminated array connector harness assemblies that span MC to host/switch electronics with a consolidated and SFF cabling footprint.

A solution approach to data center design focuses on the MDA as a high-density fiber distribution solution. It provides a flexible range of pre-terminated high-density and feature-rich modules (labeling/ID options as example) for more rapid and reliable fiber cable and connectivity installation into SAN/LAN DC applications that use CPL in the proximity of the EDA described above.

The fiber distribution system serves the following purposes and exhibits these characteristics within the data center:

- Modular patch capability for connecting "any to any" (switch to switch/switch to host)
- Port mapping of OEM blades to manageable patch field
- Clear, easy-to-use, and future-ready port labeling/ID system
- Horizontal and vertical storage capability for patch cords used for MACs
- Modular, logical, and easy-to-install facility for high-density fiber trunking
- Growth path for higher data rate channels (straightforward and cost-effective migration from serial duplex to parallel optics)
- Described solutions for high-density trunk management (verticals and pathway ingress/egress)—of particular interest is articulation of a trunk slack storage solution
- Modular components (cassettes, trunk furcations, management systems, etc.) that are eventually deployed across several DC cabling solutions (zone, floor boxes, preconfigured systems, etc.)
- Easy access for technicians to "service" side of the fiber distribution system in the event of trunk physical failures or technology migration

In addition to the performance requirements discussed, the choice in physical connectivity is most import to assure easy migration to higher data rates in the face of new media device interfaces on transceiver modules.

17.3.2.3 Port Replication as a CPL Cabling Element

In certain applications, customers desire to use the SFF of higher speed QSFP parallel optic transceivers to break out physical fiber lanes from the transceiver and represent them as independent lower speed ports at a patch panel in the CPL or local to the high-density switch.

For the high-density switch port-mapping application (Fig. 17.5), components within the fiber cabling system facilitate logically structured cabling elements to present switch ports in the CPL.

For the QSFP native interfaces on the line card (MPO for 100GE/40GE parallel optics), SFF MPO cabling and MPO patch panels present mapped MPO port interfaces at the cross connect. For bidirectional (BiDi) and parallel breakout use cases, MPO cabling and purpose-built modular cassettes present mapped LC (Lucent Connector) port interfaces at the cross connect.

Delivering the ports to a CPL in a logical and organized fashion can be done by using short MPO to MPO interconnect cordage and the deployment of 4×1 breakout cassettes in the HD Flex enclosure/patch panel using the arrangement as shown in Figure 17.6.

With this system, the 40GE ports are off the switch with an MPO fiber assembly connected to the 4×1 breakout cassette in the CPL. The LC ports are in three four-port groups

- Generally, the application resides in high-density switch applications with parallel optics transceiver modules (separate blades with QSFP modular optics).
- It is desired to have a logical mapping of 40/100GE switch port optics as they are broken down through passive fiber breakout harnesses and mapped to 10/25GE ports (for example, 36, 40G switch ports can become 144, 10GE ports).
- The breakout command works at the module level and splits the 40G/100G interface of a module into four 10G/25G interfaces. The module is reloaded, and the configuration for the interface is removed when the command is executed.
- Connections between switches are 40/100GE, while connections between switches and servers are generally 10/25GE (both can be converted to base lane rate using the same 4:1 breakout).

FIGURE 17.5 High-density switch port mapping application.

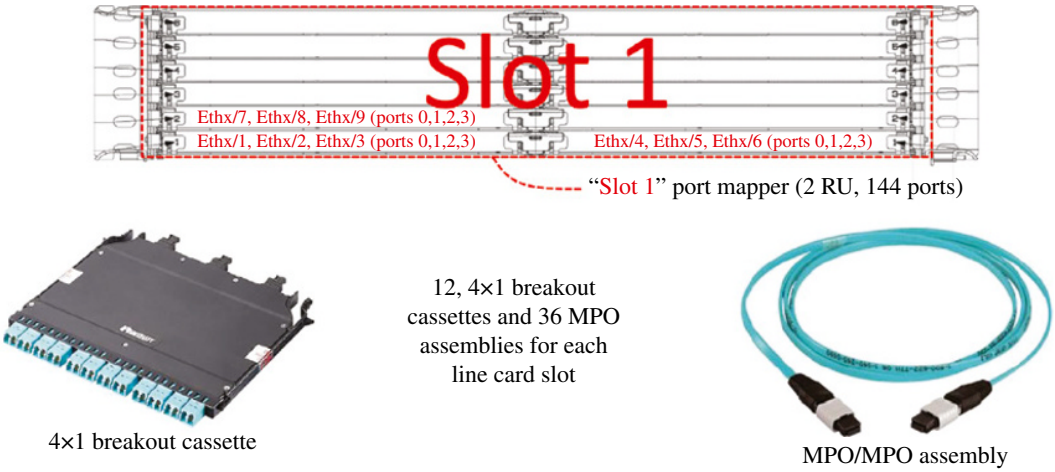


FIGURE 17.6 MPO to MPO interconnect cordage and the deployment of 4×1 breakout cassettes in the HD Flex enclosure/patch panel.

on the front of the cassette to be mapped as logical 10GE ports from the 40GE QSFP module in the line card.

The deployment of eight of the 2RU enclosures or patch panel fiber distribution systems exactly maps a switch populated with eight 36-port 40 Gb/s blades in 10GE breakout mode. Alternatively, the deployment of four of the 4RU enclosure or patch panel fiber distribution systems exactly maps the switch with slightly less modularity.

The addition of these ports brought off the switch and mimicked in the CPL allows for efficient cross connection with LC duplex jumpers to a 10GE serial duplex structured cabling system downstream.

17.3.2.4 Zone Cabling (ZDA) as a Flexible/Extensible Fiber Distribution Element

These ZDA elements are typically deployed as a floor enclosure and represent a "cable consolidation point" (CP) in that single cables (duplex LC jumpers, for instance) enter and are gathered through patch panels into larger distribution cables headed for the MDA. Floor deployments usually serve uplinks from TOR or EOR switches in server racks and drop through a single tile opening and are affixed to raised floor pedestals (Fig. 17.7).

These enclosures must be properly grounded and bonded to the raised floor system when installed and meet UL2043 requirements for use in a plenum space. The enclosure and cable openings have the capacity to support a maximum of 288 fiber ports per enclosure. Such enclosures should be capable of being secured and protected from unauthorized access with a lockable cover.

17.4 DEFINITION OF CHANNEL VS. LINK

17.4.1 Channel Definition

ISO/IEC and TIA standards define the channel as the completed fiber-structured cabling over which the active equipment must communicate. This end-to-end link includes

Optional zone distribution area (LDP/CP)

- The ZDA is an optional interconnection point within the horizontal cabling located between the HDA and the EDA to allow frequent reconfiguration and added flexibility.
- Cross connection shall not be used in the ZDA.
- No more than one ZDA shall be used within the same horizontal cable run.
- There shall be no active equipment in the ZDA.

FIGURE 17.7 Zone Distribution Area interconnects between HDA and EDA.

equipment patch cords to connect the active network devices in EDAs (typically switch to switch or switch to host) and the patch cords in the cross connect patch (optional and located in the HDA and/or MDA) (Fig. 17.8).

17.4.2 Permanent Link Definition

ISO/IEC and TIA standards define the permanent link as the permanent fiber cabling infrastructure over which the active equipment must communicate. This does not include equipment patch cords to connect the active network devices in EDAs or the patch cords in the cross connect patch areas (Fig. 17.9).

17.4.3 Application Requirements vs. Field Test Requirements (Channel vs. Link)

The performance and reliability of cabling infrastructure within the data center and in premises applications are of paramount importance. As the performance requirements for

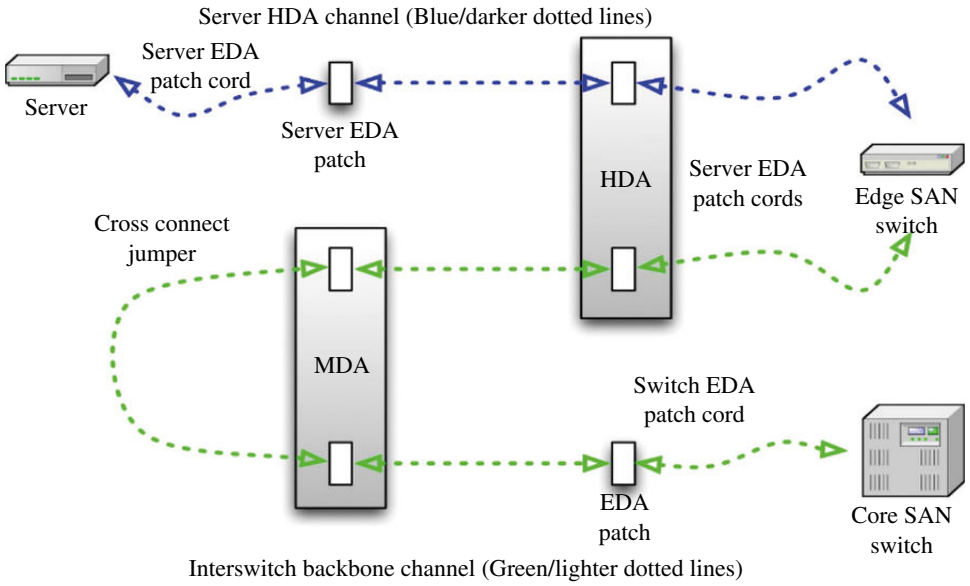


FIGURE 17.8 End-to-end link includes equipment patch cords to connect the active network devices in equipment distribution areas, and the patch cords in the cross connect patch. Source: Panduit Corporation.

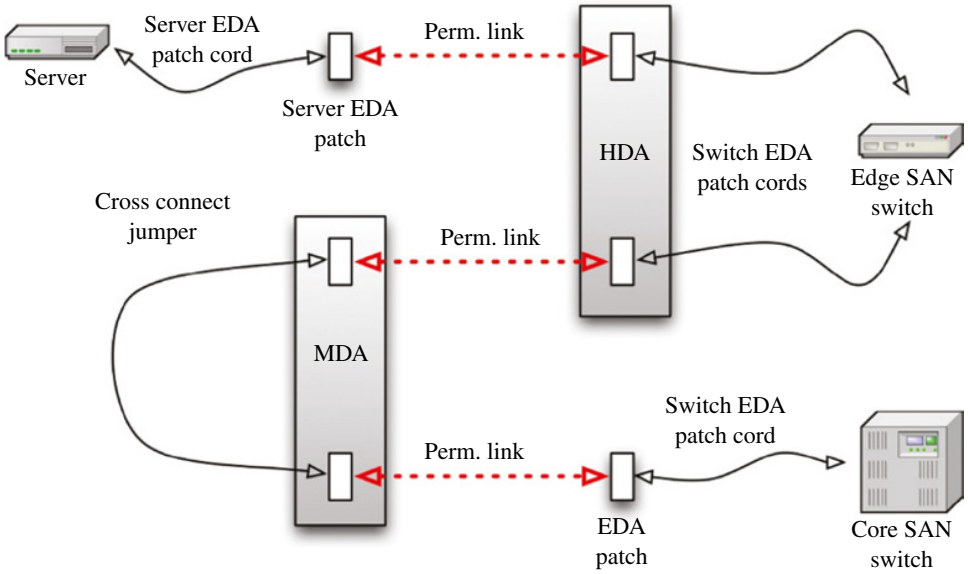


FIGURE 17.9 Permanent fiber cabling infrastructure links and communicates active equipment. Source: Panduit Corporation.

networks have advanced, the specifications on the constituent components (connectors deployed in permanent links) have become more stringent. For new high-speed optical networks, it is critical to have an accurate knowledge of the performance of the permanent links deployed in the network. It is also very important to assure that links deployed by customers present a warrantable solution when measured against cabling standards.

17.5 NETWORK/CABLING ELEMENTS

17.5.1 Fiber Selection Criteria and Transceiver Options

In constructing a 10 GbE channel link, the first requirement that must be identified is the maximum fiber length or "reach." Once the maximum reach is specified, the fiber type and physical media device (PMD) can be selected. In Table 17.2, we list the maximum reach for each standard fiber type and 10 GbE PMD option as specified in IEEE 802.3.

The two bandwidth values for each fiber type in Table 17.2 are for short and long wavelengths (S and L). The bottom two PMD options LR/LW and ER/EW are for single-mode fiber and are typically used for long-haul transmission. Single-mode options can be used for short reach, but the transceivers are generally more expensive. The 10GBASE-LX4 PMD is a four-lane parallel data transmission over a single fiber using 4 discrete long-wavelength lasers. The four wavelengths (1,275.7, 1,300.2, 1,324.7, and 1,349.2 nm ±6.7 nm) are multiplexed into the single fiber using a wavelength division multiplexer (WDM). For this PMD, each lane operates, and 2.5 Gb/s providing the means of transmitting a 10 Gb/s aggregates data rate over low bandwidth legacy fibers. The LX4 option can also be used with single-mode fiber. For 10 Gb/s serial transmission, there are two PMD options, 10GBASE-SR/SW and 10GBASE-LRM (to be published).

In Figure 17.10, we summarized the maximum reach for each fiber bandwidth using the IEEE 802.3 10 Gb/s PMD options.

Close examination of Figure 17.10 reveals that for low bandwidth fibers 500/500 MHz·km and less, the maximum reach for long wavelength is only 220-m as compared with 300-m for OM3 fiber that has the same bandwidth at 1,310 nm. This discrepancy arises from the fact that some legacy fibers were poorly manufactured (imperfection in refractive index profile) and a 300-m reach cannot be guaranteed (in some extreme cases a 220-m reach might not be attained).

We can use Table 17.2 and Figure 17.10 to select the fiber type and PMD option best suited for a given application. For example, for a maximum reach of 125-m, the most cost-effective solution would be 10GBASE-SR over enhanced OM2 (950 MHz·km) fiber. If a 300-m reach over a legacy fiber with unknown bandwidth is required, we have only one option, the 10GBASE-LX4 PMD. For

TABLE 17.2 Maximum reach for each standard fiber type and 10 GbE PMD option as specified in IEEE 802.3

		Minimum modal bandwidth (MHz·km) for five-multimode fiber types					
IEEE 802.3 PMD options	Wavelength (nm)	160/500 FDDI	200/500 OM1	400/400 FDDI	500/500 OM2	2000/500 OM3	Single-mode 9/125
Serial							
10GBASE-SR/SW	850	26 m	33 m	66 m	82 m	300 m	NA
CWDM							
10GBASE-LX4	1,310	300 m	300 m	240 m	300 m	300 m	10 km
Serial							
10GBASE-LRM	1,310	220 m	220 m	100 m	220 m	220 m	NA
Serial							
10GBASE-LR/LW	1,310	NA	NA	NA	NA	NA	10 km
Serial							
10GBASE-ER/EW	1,550	NA	NA	NA	NA	NA	40 km

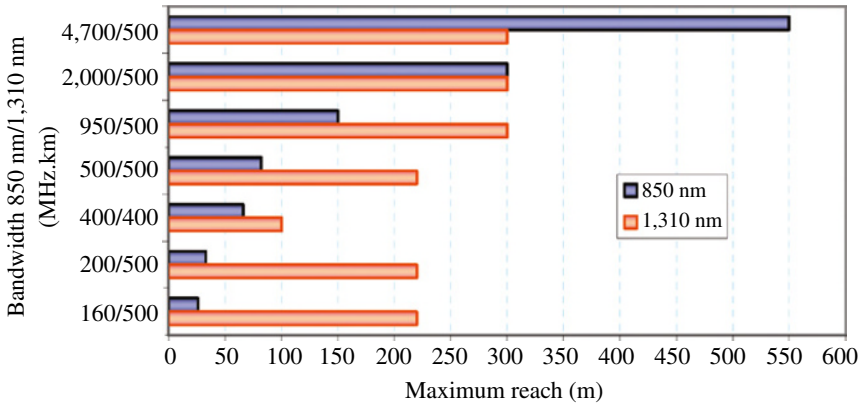


FIGURE 17.10 Maximum reach for each fiber bandwidth using the IEEE 802.3 10 Gb/s PMD options.

high-bandwidth OM3 or extended reach OM3 fibers, we have two options: 10GBASE-SR or the new 10GBASE-LRM long-wavelength PMD. Although we can use 10GBASE-LX4, it is not a cost-effective solution.

Until now we have not discussed link attenuation. Attenuation is an important parameter, but it does not typically limit minimum reach. The link budget as specified in IEEE 802.3 is 1.0 dB for fiber attenuation and 1.5 dB for connector loss. Each fiber type in Table 17.2 can meet the attenuation requirement, but not all fibers will support a maximum reach of 300-m. The link distance is limited by modal dispersion and other power penalties contributing to signal noise. If the total connector loss is less than 1.5 dB and the link attenuation is within spec, it is still possible for a "mode-selective loss" to significantly degrade link performance. Lateral offsets at connector interfaces can selectively attenuate high-order modes causing fluctuations (noise) in the received optical power increasing the optical power penalty. Low loss connectors that limit mode-selective loss are important in 10 Gb/s optical channel links.

17.5.2 Fiber Type Selection (10G Multimode Fiber Example)

With the availability of several grades of high-bandwidth laser-optimized fibers, the wide deployment of various types of legacy fiber and the specification of five IEEE 802.3 10 Gb/s Ethernet PMD options, selecting the most suitable PMD and multimode fiber type for your application, can be challenging. In this technical reference we discuss the various fiber types, 10 GbE PMD options, the effect of mixing fiber types, and key performance parameters to help you make well-informed choices in selecting your network connectivity.

In general, graded index multimode fibers differ by core diameter and a parameter known as "bandwidth." Bandwidth is a measure of the fiber's ability to propagate optical pulses at a given bit rate over a specified distance (given in units of MHz·km). Bandwidth (i.e., effective modal bandwidth [EBM]) is calculated using the industry standard test method (TIA-455-220), which measures differential mode delay (DMD). DMD characterizes the modal structure of a graded index multimode fiber by measuring the delay in picoseconds per meter (ps/m) between the fastest and slowest modes that propagate through a given length of fiber. Propagation delay between discrete modes that make up the optical pulse will result in pulse broadening or "modal dispersion," which in turn degrades system performance. The amount of modal dispersion depends on the detailed structure of the fiber core's refraction index profile. To reduce modal dispersion, the refractive index of multimode fiber must closely follow a parabolic distribution as shown in Figure 17.11 (where α~2).

If a fiber's refractive index profile is near perfect, there will be little modal dispersion, and the fiber will have "high bandwidth." In practice however, a fiber's radial index profile typically has imperfections and discontinuities (or combinations of), which cause modal dispersion and hence

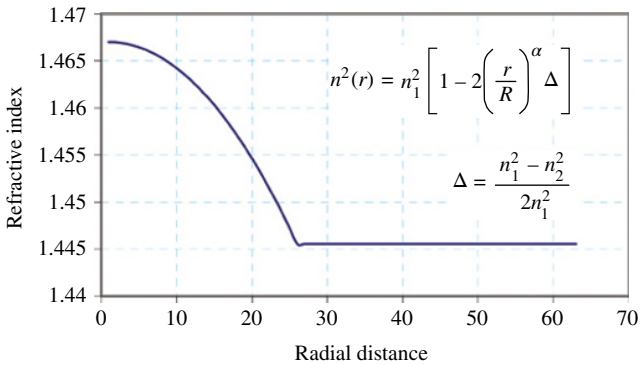


FIGURE 17.11 The refractive index of multimode fiber must closely follow a parabolic distribution.

reduced bandwidth. Fiber manufacturers use DMD measurements to characterize their multimode fiber as part of their manufacturing process and sort their fibers into several categories. An "enhanced" OM3 fiber is one that has a bandwidth that far exceeds the minimum OM3 requirement of 2,000 MHz·km (bandwidth for a restricted launch condition). Likewise, an enhanced OM2 fiber is one that exceeds the OM2 bandwidth requirement of 500 MHz·km (for an overfill launch condition).

Legacy graded index multimode fibers were originally optimized for long-wavelength (1,310 nm) transmission with a core diameter of 62.5 µm. Higher bandwidth was achieved by reducing the fiber core diameter from 62.5 to 50 µm and improving the manufacturing processes to produce a more controlled refractive index profile. With the development of low-cost short-wavelength (850 nm) vertical cavity surface emitting lasers (VCSELs), today's fibers are optimized for the 850 nm wavelength window while maintaining a minimum bandwidth of 500 MHz·km for the 1,310 nm window.

17.5.3 Cable-Type Selection

In choosing fiber-optic cables for an application, you should consider installation requirements and environmental/long-term requirements. Installation requirements define how the cable will be pulled in conduits or placed in cable trays. Considerations for water ingress, temperature variation, tension (for aerial installations), etc. should be resolved to assure long-term service in the designated environment.

You should contact cable manufacturers and share your requirements. Manufacturers will want to know the installation environment for the cable, fiber counts (consider "spares"), flame ratings, fiber types (multimode/single mode), maximum cable run, etc. and may recommend a "hybrid" cable (multimode fibers for today and single-mode fibers that remain dark for tomorrow's applications). The cable companies will evaluate your requirements and make suggestions before you can seek competitive bids.

The cable plant design will call for a defined fiber count, and it may be advantageous to consider adding "spare" fibers to the build (terminated in panels, but remaining "dark") since a few extra add marginal cost compared with the labor and incremental cost of installing more cable (to mitigate against broken fibers during installation or connectors at distribution panels that get damaged through use) (Table 17.3).

17.5.4 Connectivity Components and Systems

17.5.4.1 Fiber Connectors

Although dozens of fiber-optic connectors have been deployed over the past 30–40 years, none have taken hold in the data center as much as the "LC" and "MPO" connectors. These are documented in the TIA Fiber Optic Connector Intermateability Standard (FOCIS) as FOCIS 10 and FOCIS 5, respectively.

This was driven by the SFF requirements of the transceiver manufacturers (LC connectors) and the desire for physical fiber cable aggregation into "trunking assemblies" for connection to breakout fiber cassettes (MPO connector facilitates this) (Table 17.4).

17.5.4.2 Connectivity: Field Terminated vs. Pre-terminated

Field termination describes the termination of fiber connectors at the jobsite. Newer "cam-style" connectors designed for field termination are epoxyless and require no field polishing. These connectors are higher insertion loss (IL) and poorer return loss as there is a mechanical splice joint and incremental losses associated with such are imbedded in the total loss of the installed connector. Additional field termination options include fusion splicing of pre-terminated pigtails. Trained technicians at the jobsite perform field terminations.

In general, field termination methods involve higher material and labor costs and tend to extend installation times.

Pre-terminated assemblies and modules/cassettes are connectorized and built to exact length and performance requirements under controlled conditions in the factory and can include custom staggering of connectors, custom labels, and other options. These connectors are terminated in a facility specifically designed for that purpose. Manufacturers use automated cable preparation and polishing machines, inspection scopes, and precision equipment to test IL, return loss, and end-face geometry.

Termination efficacy and optical/mechanical performance as well as reduced material and labor costs of a factory pre-terminated fiber trunk assemblies make this approach a cost-effective solution when length requirements for the channels are known with a modicum of precision (Fig. 17.12).

17.5.4.3 Enclosure and Fiber Distribution Elements

Rack mount fiber enclosures house, organize, manage, and protect fiber-optic cable, terminations, splices, connectors, and patch cords and provide logical representations

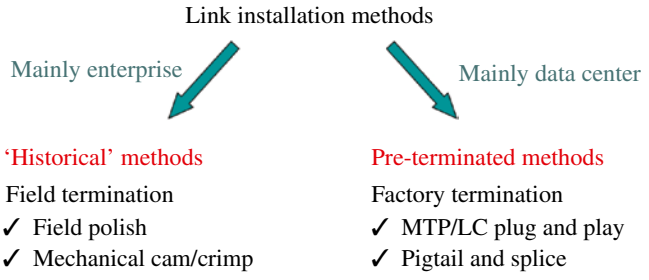


FIGURE 17.12 Field terminated vs. pre-terminated connectivity.

TABLE 17.3 Compendium of fiber-optic cables for data center applications

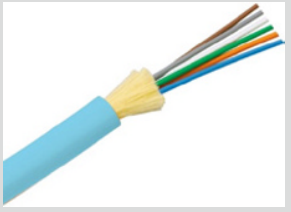
Simplex and zip cord Simplex cables are single fiber, tight-buffered (900 µm buffer over the primary fiber buffer coating) with aramid fiber (Kevlar) strength members jacketed for indoor use. Zipcord (mostly used for patch and equipment applications) is two of these simplex elements joined with a thin web.	
Distribution cables Distribution cable are small form factor light weight multi-fiber indoor cables containing tight-buffered fibers bundled under the same jacket with Kevlar strength members. These cables are typically used for short, dry conduit runs, riser, and plenum applications. The individual fibers are not reinforced, so for termination these cables need to be broken out and protected in an enclosure.	
Breakout cables Breakout cable facilitates direct connector termination without the aid of extra hardware (enclosures) to protect the fibers. They are made of several simplex cables bundled together inside a common jacket (cables within a cable). They are typically larger and more expensive than an equivalent distribution style cable. These are suitable for conduit runs, riser, and plenum applications. Breakout cable is more economical when fiber count isn't too large and/or distances are not too long, because the design requires much less labor and materials to terminate.	
Loose tube cables Loose tube cables are widely deployed as outside plant trunks protected from moisture with water-blocking gel or tapes. These cables are typically designed as fibers inside a plastic tube, wound around a central strength member, surrounded by strength members and jacketed. They provide a small, high fiber count cable that can be used in conduits, lashed overhead, or directly buried into the ground. Loose tube cables with single-mode fibers are generally terminated by splicing pigtails onto the fibers and protecting them in a splice closure.	
Ribbon cable Ribbon cable provides the highest fiber counts in a small form factor. Fiber ribbon elements, typically of 12 fibers, can be "stacked" to form cables with upward of 1,728. Mass fusion splicing can join 12 fibers at once, making installation fast and easy, and connectorized ribbon pigtails can be spliced onto the cable for quick termination. Some manufacturers have introduced fibers joined with periodic connections of matrix along the fiber length rather than solid ribbon matrix. This allows for more flexibility and the ability to roll up the "ribbons" to ease fiber management in trays and enclosures (standard ribbons have different bend radii depending on direction).	
Armored cable Armored cable can be used inside plant applications where a rugged cable and/or rodent resistance is required. Armored cables can be installed in open pathways underfloor with no worry about fiber cable being mechanically compromised. Armored cable is conductive, so it must be grounded properly.	

TABLE 17.4 Connectivity components and systems

	• The LC product family includes single and multimode versions, in simplex and duplex. • Developed by Lucent, the LC is based on the SC technology (1/2 size). • Includes a familiar RJ45 style housing. • Multiple suppliers have adopted this connector for their SFF transceivers.
	• The MPO (dev. by NTT) was used for years in Japan telco closets before U.S. acceptance. • Higher-density variants are emerging with 24 fibers terminated into a single MPO connector (variants up to 72 fibers have been documented in TIA FOCIS standards). • USConec manufactures the MTP®, which is fully intermateable with the MPO. The two terms are often used interchangeably, and both connectors are fully intermateable.

of connectorized fibers on their front patching fields (Table 17.5).

Because of this versatility, such enclosures serve as a transition from backbone cabling to distribution switching, as interconnect to active equipment, or as a cross connect or interconnect in a main or HDAs. Such enclosures should provide user (front) and service access (rear) such that initial installation of fiber connectivity via plug/play cassettes, field-terminable connectors or fusion splicing, and ongoing MAC work (jumper cross connection or equipment connection) is robust against unintentional errors.

17.6 PLANNING FOR FIBER-OPTIC NETWORKS

17.6.1 Cable Plant Life Cycle: Migration Plan

As switch/director port density increases and hence per rack port counts, documenting and executing a cable management plan is very important during equipment service or technology change and can help in fabric problem resolution. Such plans should include current/future fabric design requirements (transceiver types, reach, port counts, etc.). Equipment jumpers, cables, and harnesses can be managed on equipment frames/cabinets in a variety of ways, such as by routing and managing cables in horizontal pathways below the chassis, to either side of the chassis routed through vertical managers, through cable channels on the sides of the cabinet, and/or by deploying local patch panels near the chassis to minimize pathway and cable congestion. When planning cable management solutions and defining cable pathways, it is important to consider the location(s) of rack/cabinet power distribution units and director/switch power supplies, thereby eliminating the potential for cable interference when servicing power cords/supplies.

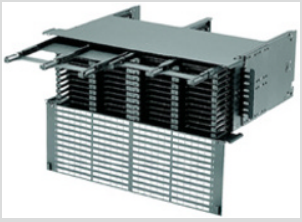
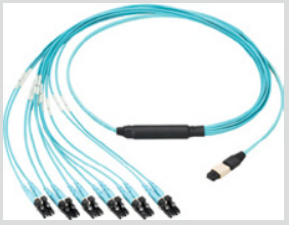
Cable management plans may involve greenfield fiber cable build in new data center or a "forklift" or incremental upgrade of fiber cabling in an existing data center:

- For existing data centers being upgraded, it is imperative to evaluate existing cabling and to document the present fiber infrastructure thoroughly (TIA 606 labeling and identification provides guidance here).
- Document current and projected network architecture using Microsoft Visio, Excel, or the like (there are also several vertical market software specifically for this task).
- Equipment interfaces should be documented in the same way with relation to current and projected needs, similarly, for various current/projected cable types and port counts present, and approximations of routed distances to distribution areas and equipment.
- Develop a plan to accommodate future growth. Build in flexibility into the CPL, such that cross-connect patching will allow "any-to-any" connections. This allows for ultimate flexibility in placement of current and future device needs within the data hall.

Higher fiber cable count requirements, enhanced performance demanded from fiber solutions, and longer distances required as data centers grow have placed strain on IT organizations, which require miles of cable infrastructure for interconnect servers, storage, and Fiber Channel fabrics. Unfortunately, many organizations have chosen the reactionary and incremental cabling approach of traditional PtP interconnect, deploying fiber cables singly to satisfy immediate needs. The resulting cable chaos impedes intelligent rational growth and contributes to an ineffective growth strategy that only deteriorates over time.

Verifying connectivity, troubleshooting, and managing change become more complex and time consuming with a PtP interconnect model and can affect negatively network uptime of critical business applications (during technology upgrades and scheduled MACs, for example). Sometimes this approach at its extreme can contribute to the energy load and efficiency of data centers, particularly in raised floor

TABLE 17.5 Rack mount fiber enclosures contain fiber-optic cable, terminations, splices, connectors, and patch cords

MTP trunk assembly		MTP terminated optical fiber trunk assemblies are typically 12–144 fibers and create the permanent fiber links between patch panels in a structured cabling environment. They are pre-terminated from the manufacturer with MTP connectors at a specified length and have a pulling grip for easy installation.
Fiber distribution enclosure		Connector housings are physically mounted in a 19 in rack or cabinet. They are typically offered in various sizes such as 1U, 2U, or 4U, which refers to the amount of rack space required for mounting.
Plug/play cassette		MTP to LC cassettes are installed into fiber distribution enclosures. They break out MTPs in the trunk assemblies into LC adapters. Trunk cables plug into the rear of the cassette, and LC jumpers plug into the adapters on the front of the cassette.
QSFP breakout cassette		MTP to LC modules are installed into the connector housings. They breakout the MTP connection from the trunk cables into LC connectivity. Thus, the trunk cables plug into the rear MTP of the module, and LC jumpers plug into the front of the module.
MTP fiber adapter panel		MTP fiber adapter panels (FAPs) are installed into fiber distribution enclosures. They offer a connection point between the MTP trunks and MTP jumpers or breakout harnesses. Thus, the trunk cables will plug into the rear of the panel, and the MTP jumpers or harnesses will plug into the front of the panel.
MTP-LC harness (breakout cable)		Harness assemblies are used for breaking out the MTP connector into multiple LC connections. In the case of breaking out a QSFP into four lanes at the QSFP base line rate, these assemblies are wired identically to QSFP breakout cassettes.
MTP or LC jumpers		MTP jumpers serve to interconnect QSFP active ports, connect cassettes for permanent links, and provide jumper connection between active QSFP ports and structured cabling systems.

environments and around the racks/cabinets where cable clutter primarily occurs and can impede airflow.

17.6.2 Fiber Cable Management: Best Practices

Structured cabling is expected to sustain growth and change over the 10–15-year life cycle of the data center. Effective cable management is considered key to the reliability of the data center network infrastructure. However, the relationship between cabling and facilities systems is often overlooked. This relationship centers on the successful deployment of structured cabling along pathways that complement facilities systems. Effective cable pathways protect cables to maximize network uptime and showcase your data center investment.

Once cable pathways are in place, attention can be directed to placing the cables in the pathways. Fiber cable bundling strategies have been developed to effectively and neatly route cables through pathways, retain flexibility for operators to make frequent MACs, and not obstruct airflow. These strategies include use of cable ties and pathway accessories that protect and manage high-performance copper and fiber cabling, in accordance with TIA/EIA-568 and GR-1275-CORE Section 13.14, in order to maintain network integrity. Cable ties and accessories must be operator safe without protruding sharp ends or edges that can potentially cut or abrade cables. Plenum-rated cable tie designs are required for cable bundling within air handling spaces such as ceiling voids and underfloor areas, in accordance with National Electrical Code (NEC) Section 300-22 (C) and (D).

A variety of cable bundling solutions are effective in high-density data center cabling environments (Fig. 17.13). For example, hook-and-loop cable ties can be used to bundle cables across overhead areas and inside cabinets and racks and are approved for use within air handling spaces in accordance with the NEC. They are adjustable, releasable, reusable, and soft, enabling installers to deploy bundles quickly in an aesthetically pleasing fashion and to address data center scalability requirements.

Cable rack spacers are used with ladder racks as a stackable cable management accessory that helps ensure proper cable bend radius and minimize stress on cable bundles. Also, waterfall accessories provide bend radius control as cables transition from ladder rack or conduit to cabinets and racks below.

17.6.3 Fiber Cable Routing/Pathways: Protection, Management, and Fill

The primary value of pathways in a data center is to provide functional, protective containment for the structured cabling infrastructure in an often dense cabling environment. Pathways that are versatile and accessible accommodate data center growth and change and protect cables from

Wire basket with hook and loop cable ties	Ladder rack with rack spacers and waterfall accessories

FIGURE 17.13 Cable bundling in wire basket and ladder rack pathway.

physical damage. Well-designed cable pathways also strengthen the visual impact of your data center.

A cable routing system is a collection of channels, fittings, and mounting brackets that can be assembled to create a structure that protects fiber-optic cabling from physical damage that can disrupt or cut off signal transmission. It also provides a versatile, scalable pathway that reduces the costs associated with maintaining existing network operations and implementing new services. Cable routing systems are not just a means of containing cable deployed in data centers or central offices. They are an integral component of the overall cable management system needed to ensure optimum network performance.

The key capacity planning issue is an accurate estimation of cable count and volume in order to specify pathway size. For initial deployments, maximum fill should be 25–40% to leave room for future growth. A calculated fill ratio of 50–60% will physically fill the entire pathway due to spaces between cables and random placement.

Most vendors of pathways offer online fill calculator tools can help you determine the size of pathway needed for a specified cable quantity and diameter.

17.6.4 Designing Fiber Cable Pathways

ANSI/TIA-942—"Telecommunications Infrastructure Standard for Data Centers"—and ANSI/TIA-569—"Commercial Building Standard for Telecommunications Pathways and Spaces" are reference methods of separating and routing data cable in overhead/underfloor cable routing applications.

The variety and density of data center cables mean that there are no "one-size-fits-all" solutions when planning cable pathways. Designers usually specify a combination of pathway options. Many types and sizes are available for designers to choose from, including wire basket, ladder rack, J-hooks, conduit, solid metal tray, and fiber-optic cable routing systems. Factors such as room height, equipment cable entry holes, rack and cabinet density, and cable types, counts, and diameters also influence pathway decisions.

One effective pathway strategy is to use overhead fiber-optic cable routing systems to route horizontal fiber cables and use underfloor wire baskets for horizontal copper and backbone fiber cables. This strategy offers several benefits:

- The combination of overhead and underfloor ensures physical separation between the copper and fiber cables, as recommended in TIA-942.
- Overhead pathways protect fiber-optic jumpers, ribbon interconnect cords, and multi-fiber cables in a solid, enclosed channel that provides bend radius control, and the location of the pathway is not disruptive to raised floor cooling (Fig. 17.14).

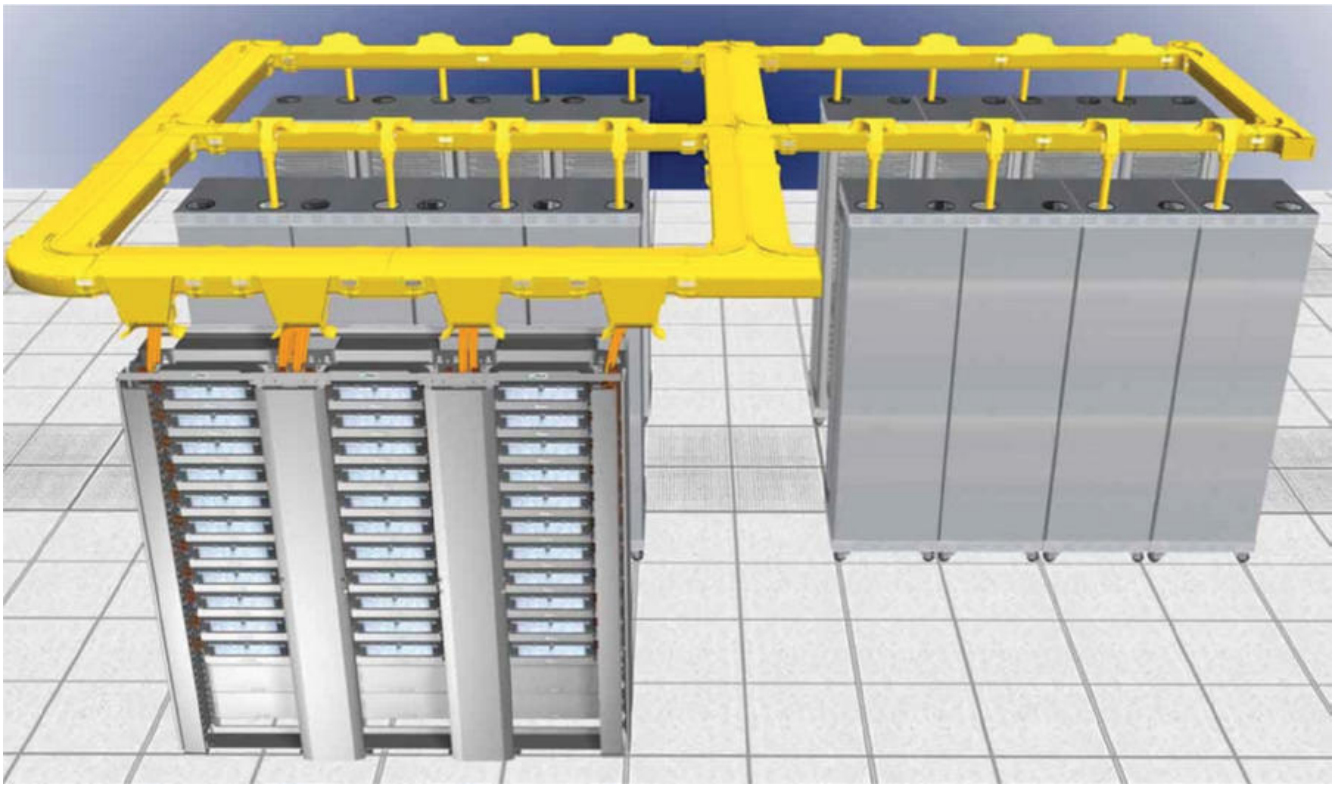


FIGURE 17.14 Overhead fiber cabling pathway system.

- Underfloor pathways hide the bulkiest cabling from view; also, copper cables can be loosely bundled to save installation cost, and each underfloor pathway can serve two rows of equipment.

Underfloor cabling pathways should complement the hot aisle/cold aisle layout to help maintain cool airflow patterns. TIA/EIA-942 and TIA/EIA-569 state that cable trays should be specified for a maximum fill ratio of 50% to a maximum of 6 in (150 mm) inside depth. TIA-942 further recommends that cable trays for data cables should be suspended from the floor under hot aisles, while power distribution cables should be positioned in the cold aisles under the raised floor and on the slab.

These underfloor pathway strategies are recommended for several reasons:

- Pathways for power and twisted pair data cables can be spaced as far as possible from each other (i.e., 6–18 in) to minimize longitudinal coupling (i.e., interference) between cables.
- Copper and fiber cable pathways are suspended under hot aisles, the direction toward which most server ports face.
- Cable pathways do not block airflow to the cold aisles through the perforated tiles.

Improperly routed or unprotected fiber-optic cable is susceptible to various types of damage. Crushing, pinching, or micro-bending can result in impeded signal transmission and cable breakage. Bend radius violations, or macrobending, in fiber-optic and copper cables can increase attenuation affecting overall system performance and cause fatigue leading to long-term signal failure.

Fiber cables are at risk of being damaged and can result in service interruption and downtime. Identifying, testing, removing, and replacing a damaged cable are costly in terms of labor and network service interruptions. A properly designed and installed cable routing system carries cabling along a logical route to minimize bends and optimize cable lengths while providing easy access to make MACs.

17.6.5 Fiber Cable Plant Labeling and Identification

A properly identified and documented infrastructure allows managers to quickly reference all telecommunication and facility elements, reduce maintenance windows, and optimize the time spent on MACs. TIA-942 recommends that data center identification start with the floor tile grid system. Each 2 × 2 ft floor tile is assigned an alphanumeric grid identifier, so that a lettered system for rows (AA, AB, AC, etc.) and a numbered system for columns (01, 02, 03, etc.) can be used to reference any given component in the data center by specific location (i.e., rack located at grid location AB03). Grid identifiers can range from computer-printable adhesive labels to engraved marking plates.

A thorough identification strategy will include the following: labels for cabling infrastructure (cables, panels, racks, cabinets, and pathways); labels for active equipment (switches, servers, storage); labels for cooling pipe, electrical, and grounding systems; and floor grid markers, voltage markers, fire-stops, and other safety signage. TIA/EIA-606-A is the standard for labeling and administration of structured cabling, and TIA-942 Annex B provides supplemental recommendations for data centers. "TIA/EIA-606-A Labeling Compliance" discusses how to implement standards-based labeling solutions.

Patch panel connectivity defines the connections between the near end ports and the far end ports. This labeling/ID system can define the connection of a range of ports on a panel or just define the connection for two individual ports.

Patch cord/equipment cord labels are identified with information that defines the connection between the near end patch panel front connections and the far end patch panel front connections or equipment connections. A near end connection identifier would consist of the cabinet/rack location, panel location, and port location. The far end connection identifier would consist of the cabinet/rack location, panel location, and port location.

Trunk cable labels are identified with information that defines the connection between the near end panel connection and the far end panel connection. A near end connection identifier would consist of the cabinet/rack location, panel location, and port location. The far end connection identifier would consist of the cabinet/rack location, panel location, and port location.

At a minimum, the labeling system shall clearly identify all components of the structured cabling system: enclosures, cables, cassettes, and ports.

The labeling system shall designate the cables' origin and destination and a unique identifier for the cable within the system. Enclosures and cassettes shall be capable of being labeled to identify the location within the structured cable system infrastructure.

17.6.5.1 Enclosure Labeling

There must be a land area defined on the enclosure system to locate the individual unit in the context of the racks and rows within the data center. The designation for the enclosure position within a cabinet/rack can be either an alphabetic designation or a two-digit number (in addition to DC

location—rack/row information) that represents the rack unit (RU) number. Real estate for this rack Location designation must be chosen such that it is not obscured by patch cords populating the front of the enclosure.

Enclosure location information typically would be in the form of a six-digit hyphenated code such as "ABCD-EF," where "AB" is row number, "CD" is rack number, and "EF" is RU location.

17.6.5.2 Port (Cassette) Labeling

Identifiers must be established for each port on the cassette to define the connectivity of cabling within the data center infrastructure. A flexible labeling system will define the location of the port within the context of the DC space. It is desired to have a landing space on the cassette for the placement of such labels. The numbering sequence should proceed from left to right and top to bottom for all ports on the enclosure. The number of digits used for all numbers on a populated enclosure system will be consistent with the total number of ports on the system. For example, a 48-port patch panel should be labeled 01 through 48, and a 144-port patch panel should be labeled 001 through 144.

Port location information typically would be in the form of an eight-digit hyphenated code such as "ABCD-EF-GH," where "AB" is row number, "CD" is rack number, "EF" is RU location, and "GH" is the port location.

17.6.5.3 Trunk Labeling

Cable labels are identified with information that defines the connection between the near end panel connection and the far end panel connection. A near end connection identifier would consist of the cabinet/rack location, panel location, and port location. The far end connection identifier would consist of the cabinet/rack location, panel location, and port location.

Port location information typically would be in the form of a 16-digit hyphenated code such as "ABCD-EF-GH/IJKL-MN-OP," where "AB" is row number, "CD" is rack number, "EF" is RU location, and "GH" is the port location, and on the other side of the trunk "IJ" is row number, "KL" is rack number, "MN" is RU location, and "OP" is the port location

17.7 LINK POWER BUDGETS AND APPLICATION STANDARDS

17.7.1 IEEE and ANSI Fiber Channel Link Power Budgets

The overall power budget for an optical channel link is determined during the development phase of the associated application standard and is based on the magnitude of several optical impairments (or power penalties), as well as the maximum channel reach. For a multimode-based Fiber Channel, these penalties include inter-symbol interference (ISI), mode partition noise (MPN), modal noise (MN), relative intensity noise, (RIN) reflection noise (RN), polarization noise (PN), and IL.

Typically, most of these optical impairments are small (<0.3 dB) and will not be considered here. However, ISI and IL do contribute large optical penalties and therefore are the two primary impairments that limit channel performance (or channel reach) and are strongly influenced by the quality and practices used in the construction of the physical link.

When an optical pulse propagates through a Fiber Channel, its shape will broaden in time due to bandwidth limitation in the transmitter, fiber, and receiver. The optical pulse representing each data bit or "symbol" will spread in time and overlap the adjacent symbols to the degree that the receiver cannot reliably distinguish between changes in the individual symbols or signal elements. The power penalty due to this effect is called ISI. ISI therefore affects the temporal characteristics of the signal pulses, which results in signal dispersion and timing jitter at the receiver. ISI typically contributes the largest optical power penalty in high-speed MMF transmission systems.

To meet the ISI channel requirement, each standard such as 10 Gb/s Ethernet (IEEE 802.3ae) or 8 Gb/s Fiber Channel (FC-4) specifies the minimum fiber bandwidth (or maximum dispersion) necessary to comply with the system ISI requirements and ensure error-free system performance. The fiber bandwidth is specified in terms of EMB, and high-speed systems (>10 Gb/s) must achieve a minimum EMB of 2,000 MHz·km for laser-optimized OM3 MMF and 4,700 MHz·km for OM4 MMF.

IL is the second critical parameter that determines the performance of a channel link. There are two sources of IL: loss at the connector-to-connector interfaces and loss or attenuation within the fiber itself due to the absorption and scattering of light as it propagates. For high-performance and reliable 10 Gb/s network operation, both loss sources should be minimized by selecting high quality, low IL connectors, patch cords, and cassettes plus high-performance MMF. In Figure 17.15, we compare the optical power penalties for a 10 Gb/s Ethernet channel link as specified in IEEE 802.3ae for 10GBASE-SR. The total power budget for this channel link is 7.3 dB.

17.7.2 Worked Example Power Budget

In principle, one can trade off cable attenuation for connector IL, or ISI power penalties for IL; however this must be

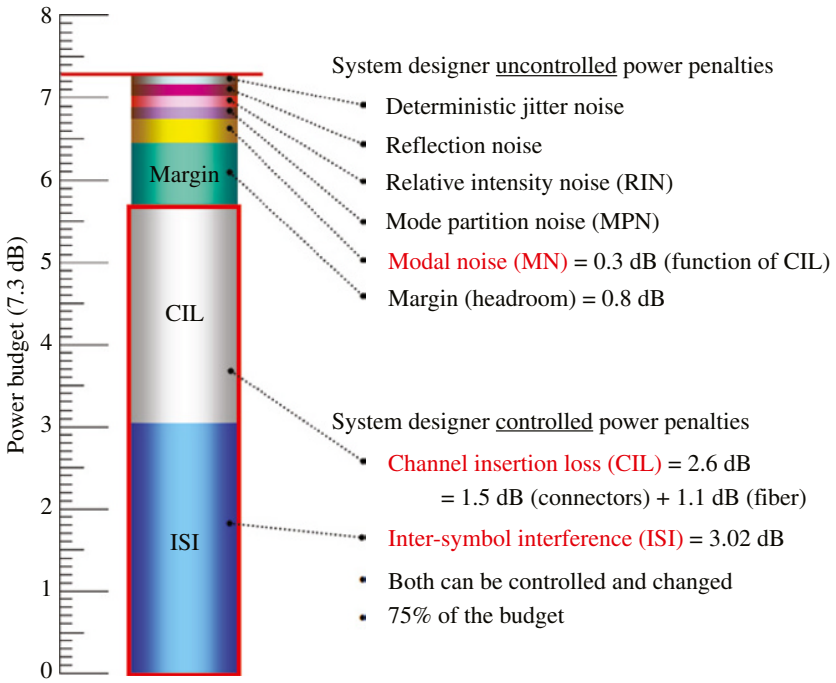


FIGURE 17.15 Optical channel budget for 10 Gb/s Ethernet (10GBASE-SR).

TABLE 17.6 16 Gb/s Fiber Channel reach/power budget vs. total connector insertion loss

	Distance (m)/loss budget (dB)				
	Connection loss				
Fiber type	3.0 dB	2.4 dB	2.0 dB	1.5 dB	1.0 dB
M5F (OM4)	NA	50/2.58	100/2.36	125/1.95	150/1.54
M5E (OM3)	NA	40/2.54	75/2.27	100/1.86	120/1.43
M5 (OM2)	NA	NA	25/2.09	35/1.63	40/1.14

done with caution. "Engineered links" are those channels designed making trade-offs of parameters.

As an example, consider an OM4 (M5F in Fiber Channel), 16 Gb/s Fiber Channel link with an installed reach of 50 m (with 2.4 dB of total connector IL); this is a third of the maximum specified reach of a 150 m engineered link with 1.0 dB of connector IL (see Table 17.6).

ISI for this channel is significantly less than when it is at 150 m. As a result, a larger connector IL of 2.4 dB can be tolerated. Alternatively, the ISI penalty can be reduced by using the increased fiber bandwidth of OM4.

It is important to understand and quantify the permanent link certification limits within the LSPM test procedure. The setup values for connector loss for "engineered links" in particular must be selected to be in compliance with the application standard if these limits are tighter than the relevant cabling standard. For example, the TIA/ISO typically states 0.75 dB maximum per connector pair, but an "engineered link" may call for 0.5 dB maximum; therefore the application standard takes precedence.

17.7.3 "Engineered Link": Working Outside of the Application Standards

Customers are designing "engineered channels" for solid reasons:

- The reach of the standards-based solutions for Ethernet and Fiber Channel does not fulfill requirements.

- Customers like the flexibility and scalability of fiber-structured cabling and by default will specify a CPL that functions as a cross-connect facility for "any-to-any" MACs. Certain customers will also propagate this model into cross-connect zones/pods, which results in an MC and a zone cross connect that are concatenated. This pushes the boundaries of what can be done with the espoused standards and brings us to the realm of engineering channels to suit, based on the deployment of very high-performance fiber and/or "ultra" low loss connector systems.
- The customer is designing a "migratable" cable plant that will be used for higher speed optics at some point and is anticipated (based on industry trends) more loss constrained channels.

The core problem and crux of this article sit with bullet point #2. Customers expect that "ultra" low loss connectivity (built into an engineered channel) should be able to be validated in the field to the same performance as is measured to in the factory.

This customer has chosen to design a full cross connect into his 40G SR4 channel to support port-mapping 40G core switches within a CPL proximal to the core (Fig. 17.16). 12-fiber ribbon cable terminated with MPO connectors throughout (trunks and patch cords) reach out from the switch core to Top of Rack switches within the server pods.

This customer's longest channel for the end-to-end optics is 170 m, which puts him outside of the capability of the 150 m OM4 channel designed with 1.0 dB maximum connector loss in the IEEE 802.3ba standard (see plot of connector loss and fiber type vs. reach for the ba standard in Fig. 17.17), so he chooses to deploy "ultralow loss" MPO connectors to assure channel integrity. These MPO connectors demonstrate a maximum IL of 0.25 dB (factory test results).

The design calls for long trunk assemblies reaching out to the servers (150 m max) and shorter trunks to connect between the core and cross-connect area (10 m max). The customer then wants to qualify the two trunks (when mated into MPO fiber adapter panels) as permanent infrastructure (links; see Fig. 17.16) to the manufacturers' "ultra" specifications (and not the TIA limits). So, the long trunk (Fig. 17.18 on the left) and the short trunk (Fig. 17.18 on the right) would yield the following "engineered limits":

$$\text{Server-side trunk test limit} = 2 \times 0.25\,\text{dB} + 0.15\,\text{km} \times 3.5\,\text{dB/km} = 1.03\,\text{dB}$$

$$\text{Core-side trunk test limit} = 2 \times 0.25\,\text{dB} + 0.01\,\text{km} \times 3.5\,\text{dB/km} = 0.54\,\text{dB}$$

17.7.4 Cabling Standards Expectations

If these links were tested against the TIA/IEC guidelines, we would yield the following:

$$\text{Server-side trunk test limit} = 2 \times 0.75\,\text{dB} + 0.15\,\text{km} \times 3.5\,\text{dB/km} = 2.03\,\text{dB}$$

$$\text{Core-side trunk test limit} = 2 \times 0.75\,\text{dB} + 0.01\,\text{km} \times 3.5\,\text{dB/km} = 1.54\,\text{dB}$$

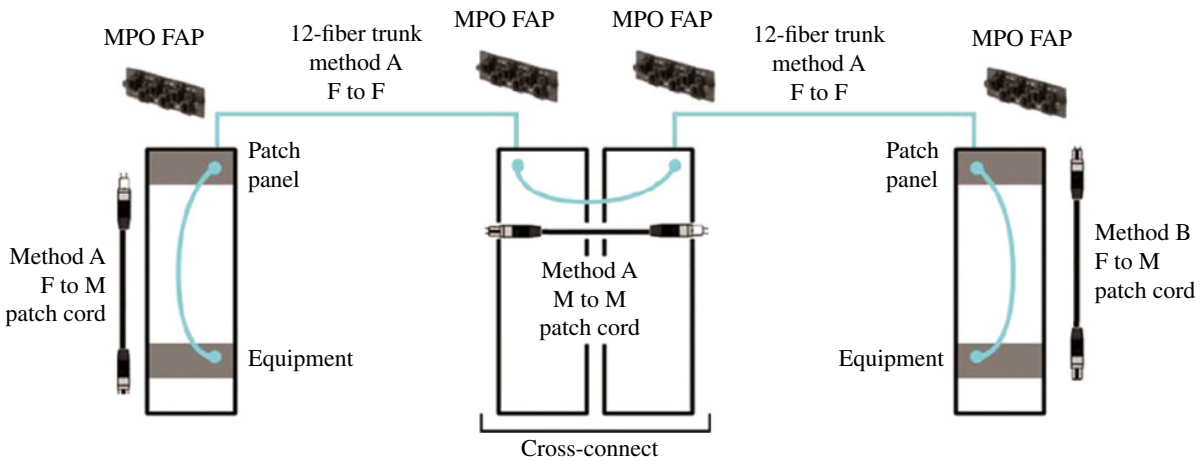


FIGURE 17.16 40GBASE-SR4 "engineered channel" example.

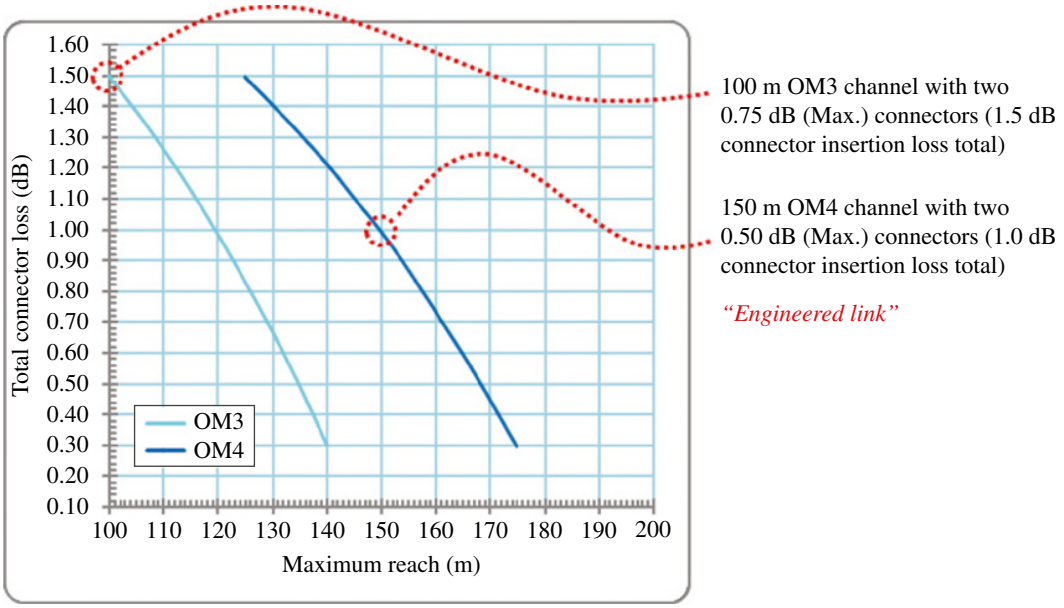


FIGURE 17.17 IEEE model for various connector insertion loss values.

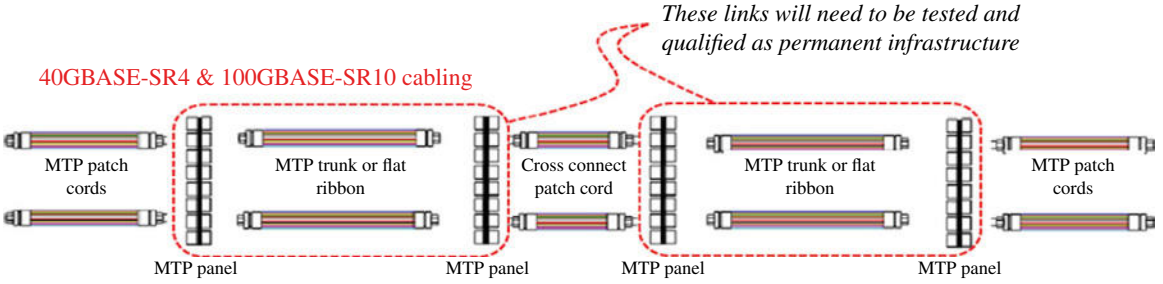


FIGURE 17.18 Long trunk (left) and short trunk (right) would yield different "engineered limits."

17.8 LINK COMMISSIONING

The most widely accepted method of measuring the loss of a permanent link is the one reference patch cord method. Like component testing per TIA-FOTP 171, this method involves a single well-controlled, nearly ideal patch cord used as the test interface. Single reference cord methods for permanent link qualification yield a high degree of internal measurement repeatability and reproducibility between multiple test sets, light source/power meters (LSPM/PMLS), and across many operators.

Recent changes in domestic fiber cabling standards (ANSI/TIA 568.3-D) have tightened the requirements on link test limits and have differentiated the limits for reference vs. standard grade test cords. This effort harmonizes the domestic standard with the important questions to ask prior to specifying operating procedures for testing:

- What is the most capable and accurate measurement methodology for higher speed multimode and single-mode links?
- What is the industry's best practice to assure that the frequency of measurement errors is minimized, saving time and money?

Test method and field practice have strong impacts on measurement accuracy, repeatability, and reproducibility to enable installers to decide on the most effective methods for their needs. Used cases for testing cable plant supporting higher speed applications are presented.

17.8.1 Tier I vs. Tier II Testing

Many individuals responsible for performing link testing have questioned whether they should perform optical time domain reflectometer (OTDR) testing for data center cabling as specified in ANSI/TIA-568-C. A subset of these individuals also questions if this type of testing can supplant traditional PMLS testing.

Industry standards require Tier I PMLS testing as the minimum regimen for a compliant installation. Tier II OTDR testing (i.e., extended testing) is not a substitute for PMLS

testing but is complementary and although highly recommended is ultimately performed at the discretion of the network owner and system designer. OTDR testing does not replace Tier I PMLS testing as the only type of testing required by domestic and international standards bodies for the commissioning testing for permanent links.

Together, PMLS and OTDR testing provide both the absolute loss measurements in comparison with the loss budget and individual measurement of events on a fiber link. When measuring a simple, short data center link using PMLS, only total loss for the link is obtained (not component-level loss information). By contrast, in addition to link loss, OTDR testing reveals component IL and reflectivity of connectors, splices, and other fiber attenuation discontinuities in the link.

The combined results of Tier I and Tier II testing are beneficial in that they can be used to validate individual component specifications. For links that marginally fail, the typical issue that people performing link testing run into is the decision of which connector to remediate (retest and/or cut off, re-terminate and retest). The information to make these types of decisions cannot be gleaned from PMLS tests alone, but can be obtained from OTDR tests.

Specifically, for field installed connector systems deployed in permanent links, the decision to re-terminate connectors after a link fails Tier I PMLS testing can be a decision that brings added cost beyond that of the scrap parts, extra consumable materials, and the labor to perform the re-termination and retest. Since PMLS testing yields only link loss and not component loss, it occasionally becomes a "guessing game" on marginally failing links as to which connector is causing the link test failure. A portion of the time, compliant connectors will be cut-off and re-terminated, thus not fixing the root cause issue for the failing link. This effect is exacerbated by installer "first pass yield" (the yield in percent of individual single fiber connectors terminated successfully), by presenting more opportunities for these types of errors.

17.8.2 Tier 1 Applicable Equipment

Fiber verification testing capability (including end-face inspection/cleaning) should be a part of cabling technician's capabilities. Throughout the cable installation process and before cable link certification and commissioning, individual losses of cabling segments should be executed to ensure the ongoing quality of installation workmanship. (Don't test everything at the end of the build!)

The most basic link qualification test is performed with a LSPM test set (referred to as a Tier I test) (Fig. 17.19). Fiber verification tools (such as a visual fault locator laser) are typically less expensive but highly useful tools; they're also most effective in quickly isolating broken or bent fibers in a troubleshooting mode. Certain aspects of end-to-end link loss measurements may indicate if the optical fiber cable is suspect or that other network elements are the cause of failures.

17.8.3 Link Attenuation Equation and Graphs

A link testing budget may be drawn from the standards as follows (Reference [1], the subclause 11.3.3.4, TIA 568):

11.3.3.4 Link attenuation equation and graphs

Link attenuation is calculated as:

Link attenuation = *Cable attenuation* + *Connector insertion loss* + *Splice insertion loss* (16)

where:

Cable attenuation (dB) = Attenuation coefficient (dB/km) × Length (km)

Attenuation coefficients are:
- 3.5 dB/km @ 850 nm for multimode — Fiber attenuation max = 3.5 dB/km
- 1.5 dB/km @ 1300 nm for multimode
- 0.5 dB/km @ 1310 nm for singlemode outside plant cable
- 0.5 dB/km @ 1550 nm for singlemode outside plant cable
- 1.0 dB/km @ 1310 nm for singlemode inside plant cable
- 1.0 dB/km @ 1550 nm for singlemode inside plant cable

Connector insertion loss (dB) = number of connector pairs × connector loss (dB)

Example:
= 2 × 0.75 dB — Connector loss max = 0.75 dB;
= 1.5 dB

Splice insertion loss (dB) = number of splices (S) × splice loss (dB) — Splice loss max = 0.30 dB

Example:
= S × 0.3 dB

ISO/IEC and TIA standards define permanent link testing to verify the performance of the fixed (permanent) segments of installed cabling as accurately as possible using LSPM testing as shown in Figure 17.20.

Completion of this testing provides assurance that permanent links that pass standards-based (or application-based) limits can reliably be configured into a passing channel by adding good quality patch cords.

It does not assure that all transmission standards will run over links that are tested against this cabling standard. Each application standard places requirements on the individual IL of all connector mated that constitute the channel (see requirements for IEEE 802.3 ae 10GBASE-SR subclause 52.14.2.1):

> 52.14.2.1 Connection insertion loss
>
> The insertion loss is specified for a connection, which consists of a mated pair of optical connectors.
>
> The maximum link distances for multimode fiber are calculated based on an allocation of 1.5 dB total connection and splice loss. For example, this allocation supports three connections with an insertion loss equal to 0.5 dB (or less) per connection, or two connections (as shown in Figure 52–14) with an insertion loss of 0.75 dB per connection. Connectins with different loss characteristics may be used provided the requirements of Table 52–24 are met.

Typical LSPM test set shown consisting on main unit (power meter and source with display) and remote (power meter and source only). Some units provide both copper and fiber certification testing that guarantee that cabling installations comply with all TIA/ISO standards:

- Such testers should comply with ISO Level IV and TIA Level IIIe accuracy requirements.
- Length measurement capability built in to estimate fiber attenuation loss on the basis of set attenuation coefficients.
- Should be capable of analyzing test results and creating test reports for interpretation by third-party software.
- Some units combine extended fiber certification (Tier 2) as an option to basic LSPM capability.

FIGURE 17.19 Light Source and Power Meter (LSPM) Tier I test set.

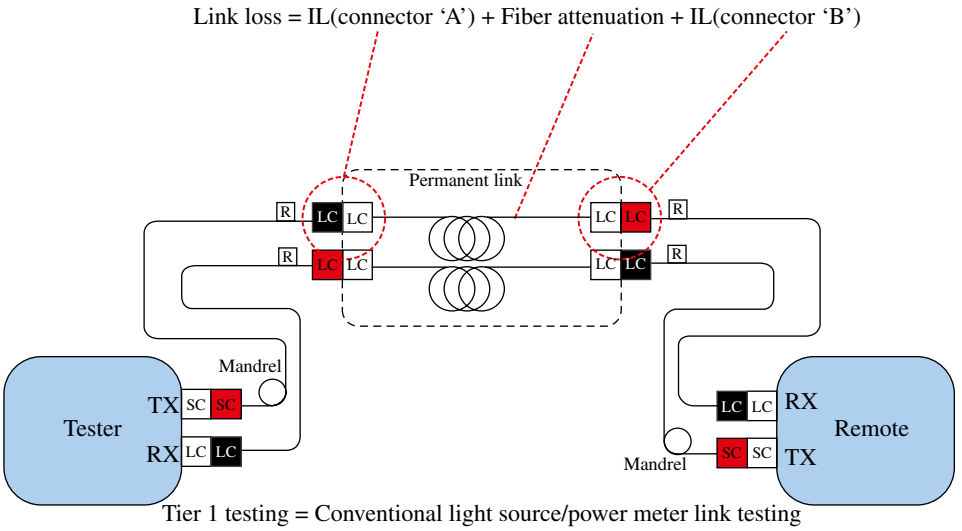


FIGURE 17.20 Installed cabling as accurately as possible using LSPM testing.

So, this places the additional requirement on the link to assure that all connectors have mated pair IL of no more than 0.75 dB/connection (Fig. 17.21). This requirement can cause a dichotomy between links that pass test limits and channel requirements as noted above. A scenario for this is presented below (200 m, 10G MM backbone as example with a cable standards-based loss budget = 2.2 dB), representing worst-case budget allowance using 568 backbone testing requirements of 0.7 dB (fiber attenuation) + 0.75 dB × 2(connectors).

Some import things to note from the above example that indicates a passing link to the cabling standards are the following (2.17 dB link loss vs. 2.2 dB for the test limit):

- Marginally passing link but individual connector is far out of 10G standards compliance (1.6 dB vs. 0.7 dB max for the IEEE application standard).
- This is a field termination issue for higher speed apps when performing only Tier I (LSPM); Tier II testing in addition to this would identify the individual discrepant connector loss.
- Factory pre-terminated systems (based on MPO connectivity) don't usually suffer from this issue because individual connector loss is validated in the factory (in addition to return loss, which cannot be effectively evaluated in the field without OTDR gear).

Ultimately, network functionality and signal integrity rely on the performance of the Channel (the completed end-to-end link). Installation and test personnel do not typically measure end-to-end loss of the complete Channel with all EDA cords and cross-connect cables in place.

EDA cords and cross-connect cables are generally installed after the "permanent" cabling installation has been completed and tested and then are subject to MACs throughout the cabling system's lifetime. It is therefore compulsory to certify that the PL cabling infrastructure meets performance level defined by standards and/or applications (generally whichever is stricter) to assure adequate system headroom when MACs are performed by IT personnel at a later date.

There are no standards that talk to application-based Channel test limits other than the extension of the permanent link test limits (with the addition of the connector losses in patch cords). The application link power budget (Ethernet, Fiber Channel, etc.) does not include the connectors that are attached to equipment on either end of the link as IL "milestones." These are built into the link power budget as minimum transmitter power into (−dBm) the fiber and receiver minimum sensitivity in Amps/Watt. So, strictly speaking, the number of connectors in the Channel is the total number of "mated pairs" of connectors (connector terminations into the receptacles of the transceivers are not "mated pairs").

Typically, channel certification using LSPM methods is at the behest of network owners and/or specifiers and brings no real additional value beyond the initial permanent link testing and is best deployed as a troubleshooting tool for channel functionality.

17.8.4 LSPM Testing Best Practice: Increasing the Efficacy of Field Test

Use TIA-526-14-B Annex "A" (one jumper method) as the default method of validating permanent link performance for data center links with multimode fiber. Test equipment (receive head) must be equipped with link under test connectors.

Use Encircled Flux launch conditioning cords (or mandrel wraps) per test equipment manufacturers' guidelines to produce standards-compliant launch conditions. This serves to reduce variability of test particularly between test sets.

Use precision or reference-grade launch jumpers in all cases. Make sure that mechanical and optical characteristics of these conform to local standards. Reference-grade patch

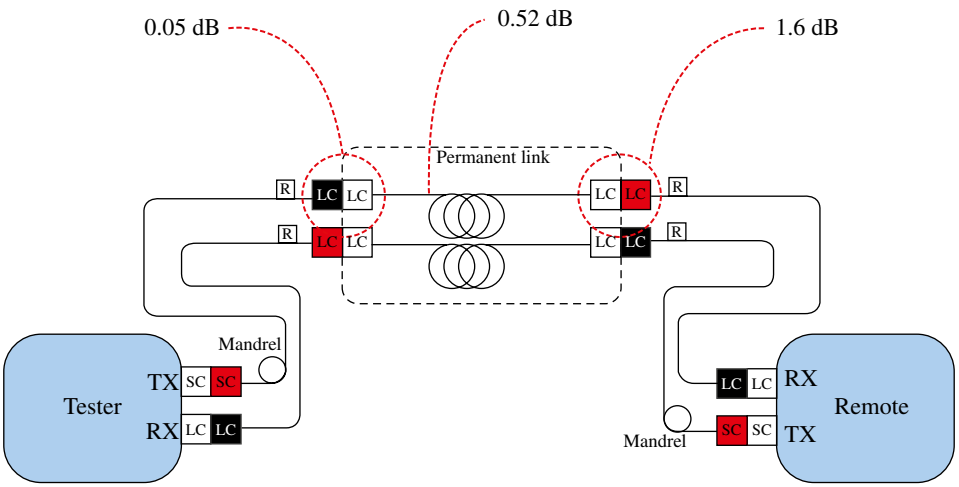


FIGURE 17.21 Connection insertion loss.

cords are required for accurate characterization of link loss in fiber-based permanent links. These cords are typically used as consumable items in the commissioning and qualification of links (after initial installation). Reference-grade patch cords minimize total installed cost by providing excellent measurement capability in the face of tight application power loss budgets required for higher speed channels.

A reference patch cord is a cord that contains connectors that minimize the mean and standard deviation of IL when mated against a large population of sample connectors. These reference connectors are then connectors that have nominal optical and geometrical characteristics (numerical aperture and core/ferrule concentricity, for example), such that when mated against other reference connectors produce "near-zero" loss. Use of such reference-grade patch cords is a necessity to assure accuracy (in referencing) and gage repeatability (replication of link tests under same reference) and reproducibility (replication of test results across multiple test sets and references):

- Use FOTP 171 (one jumper method) to qualify precision jumper connectors on a component basis (instead of a fixed number of mating cycles).
- Guidance with respect to the longevity and durability of such cords is also discussed in standards (Telcordia GR 326, as an example) with the aim of providing guidance with respect to maintenance of working reference cords. Here it is generally left to the individuals performing testing to assess the integrity of the reference cords.
- Use FOTP 171 (one jumper method) to qualify reference cords on a "schedule" and when reference cords are in question (instead of a fixed number of mating cycles)—deciding when a reference cord is best taken out of service can be best done by performing one jumper component IL on all reference cord ends that interface to links under test with a "master" cord that is purpose-built to qualify working reference cords. If possible, chart (or at least log) these measurements to determine the state of control of the reference.

Be sure to allocate the actual number of mated pairs of connectors present in link into the link power budget (measured against reference connectors), irrespective of link measurement technique chosen.

For loss challenged links (tight engineered links), assess the test limits against the GR&R of the test set, and if the GR&R significantly infringes the capability to test to said limits, negotiate with the customer to modify limits upward by one-half of the GR&R. This is a good point to engage the structured cabling supplier to provide guidance and a technical bridge between the end customer and the SI/Installer.

And most importantly, adhere to good cleaning and inspection practices as outlined in connector component and test equipment manufacturers' guidelines—"When in doubt, clean it." This goes for anything that touches the link under test, including the test equipment reference cords, visual inspection equipment, etc.

17.8.5 Interpreting LSPM Results/Remediation

If there are link test failures during single-ended LSPM testing (high loss in a link above the test limit), make sure to test the link in the opposite direction using a single-ended method that isolates the far end connector. Since the single-ended test only tests the connector on one end, you can easily isolate a bad connector.

High loss in the double-ended test (with launch and receive jumpers) should be isolated by retesting single-ended and reversing the direction of test to see if the end connector is bad. If the loss is the same, you need to either test each segment separately to isolate the bad segment or, if it is long enough, use an OTDR.

If you see no light through the cable (very high loss and indicating darkness when tested with a visual fault locator), it's probably one of the connectors, and in this you have few options. The best one is to isolate the problem fiber in the cable, cut the connector off of one end (field re-terminate if possible), and hope it was the bad one. If you have access to an OTDR, the failing connector can be pinpointed, and the potential error just described can be mitigated by properly identifying and remediating the failing connector.

In "dual-window" LSPM testing (dual wavelength testing of a multimode fiber link at 850 and 1,300 nm, for instance), it is important to look at the difference of the two loss results. Fibers can fail at one or both "windows." Significant differences in these loss figures can point to one of two things:

1. Loss at lower wavelength significantly higher than higher wavelength—potential connector issue
2. Loss at higher wavelength significantly higher than lower wavelength—potential cable "macrobend" issue—inspect the cable plant with a visual fault locator and if possible walk the cable run, inspecting for cable bends as you go.

17.9 TROUBLESHOOTING, REMEDIATION, AND OPERATIONAL CONSIDERATIONS FOR THE FIBER CABLE PLANT

17.9.1 MACs

Hierarchical, distributed switching architectures tend to reduce the amount of fiber cabling to the core switching/CPL locations but with a corresponding increase in the total number of switches in the network. An alternate "structured cabling system" (SCS) approach focused on standardizing

the ways in which cables are run along with standardized cabling processes; SCS makes use of fiber distribution patch panels at each aggregation point or intermediate distribution frame (IDF).

Manual cross-connect patches or equipment patches in EDAs allow for the completion of channels throughout the data center, but without strict adherence to the guidelines and best practice documentation by every data center employee or contractor, these distribution areas can quickly become just as difficult to maintain and manage as traditional methods. In each case, manual MACs introduce the human error factors, both in performing MACs and documenting such when the work is done.

17.9.2 Jumper Storage

Horizontal and vertical cable management installed both within and between racks/frames/cabinets enable jumper cable management and provision orderly growth of the cable plant. Built-in jumper storage panels mounted in the vertical space between racks (outside of the rack space/patching area) help to store excess jumper slack and serve to minimize the number of discrete jumper lengths required while maintaining controlled fiber bend radius.

In runs of fiber (both in the cable plant and in patch cords), a rule of thumb of minimum 1.5-in bend radius for patch cords is needed, or a bend radius of no less than 10 times the fiber cable outside diameter. Bends with increasingly smaller radii than the specified minimum bend radius will suffer from static fatigue and will fail over time.

Bend radius violations may cause macrobending of fibers inside of the cable, resulting in signal attenuation (this will be particularly noticeable at higher wavelengths). Severe bending of fiber cables in situ present a long-term reliability issue (fibers can break completely over time due to static fatigue). Optical "bend-insensitive" fiber variants can be bent to a less than 10 mm radius without significant excess loss, but that does not eliminate the need for proper cable management and bend radius control. Remember, any bend-enhanced fiber only addresses the optical performance at tight bends. It does not change the mechanical capabilities of the fiber. This is true for any commercially available telecommunications fiber (single mode or multimode).

Jumper storage systems are designed to simplify the management of extra length of jumpers particularly at cross connects and ultimately save money by reducing the need for holding an inventory of different jumper part numbers. These systems should be designed to ensure easy cable access minimizing the number of fiber crossover points.

17.9.3 Tier II Testing: OTDR Basics

It's not particularly onerous to measure optical physical plant for total link loss using Tier I methods given today's test gear options. Documented Tier I methods guide us in using stabilized light source to excite one end of the link and how to measure the amount of light emanating from the other end using a calibrated optical power meter. This radiometric method, if done properly, gives a good measure of the overall end-to-end link loss. If the loss you measure in this way is excessive, what can be done to pinpoint the source of the excess loss?

This loss could be a dirty connector, a marginal splice, excessive fiber bending, or a myriad of other optomechanical issues. The LSPM allow you to total link loss but provide no way to componentize the loss to where the problem originates.

PMLS methods are extremely useful for quick performance checks and are required for link commissioning, but unfortunately do not allow for identification of nature or location of an optical loss problem in the link. An OTDR (Fig. 17.22) is a piece of test equipment that can measure the fiber loss, splice loss, connector attenuation, and the optical return loss (ORL). Most modern OTDRs provide detailed reports that describe all optical component performance metrics and where in the link excess losses are located. Measurements are made from one end of the link (Tier 1 measurements with source/power meter require simultaneous access to both ends).

We can measure these small incremental and independent connector and splice losses and other optical fiber characteristics (detecting excessive bends, assessing return loss at connectors, etc.) using an OTDR. This instrument makes measurements by examining the optical fiber like an optical radar with intense, ultrashort pulses of laser radiation.

Using an OTDR with access to an open fiber port, you can:

- Measure the distance to and the loss across a fusion splice, connector, or significant bend in the fiber

FIGURE 17.22 Optical time domain reflectometer.

- Measure the ORL of discrete components, such as connectors and fiber-optic modules such as splitters
- Measure the integrated (total) return loss of a complete fiber-optic system
- Measure fiber attenuation (expressed in dB/km) and the uniformity of attenuation
- Measure the link loss or end-to-end loss of the fiber network
- Provide active monitoring on live fiber-optic systems

17.9.4 Interpreting a Basic OTDR Trace

Figure 17.23 shows a typical OTDR signal trace, commonly referred to as a waveform.

This downward sloping waveform results from cumulative (as the laser pulse transits the length of the fiber) Rayleigh backscatter intrinsic to the fiber, and the spikes shown in the trace result from discrete points of reflection or local change of reflection on the fiber ("events"). The spike near the beginning of the waveform, on the left, is a reflection from the front-panel connector on the OTDR. The second spike (very close to the first one) is caused by the connector on the jumper connected at the patch panel (see Fig. 17.23), which terminates the connects to the cable plant for management and access to the communication equipment.

Along the sloping portion of the waveform are two points where the Rayleigh scattering level drops slightly, but abruptly. These drops in backscatter result from a pair of fusion splices that are nonreflective events (unlike connectors which, if poorly processed, can be highly reflective). The large spike at the end of the waveform is caused by reflected light off the far end patch panel's connector. This is an "open" connector in this case (not mated to another connector) and hence is highly reflective due to the glass-to-air interface.

In troubleshooting or performing higher level (Tier II) testing to more fully document a links performance, you want the IL (in decibels) of any splices and/or connectors deployed along the link and the distance to these "events" (meters or feet). Notice from Figure 17.23 that the OTDR vertical scale is marked in dB and that the horizontal scale is marked in distance units (in this case km). In providing measurements in dB and meters, the OTDR makes internal conversions as its receiver measures optical power linearly (not in decibels) and as a function of time (not distance).

The most easily accessible results calculated from an OTDR (outside of the presentation of length of link segment lengths) are the loss of discrete elements along the length of the link (connector and fusion splice loss).

These are categorized as "reflective" (connectors) and "nonreflective" (fusion splices) events on the OTDR trace (Fig. 17.24). Modern OTDRs present these results as computed loss for each of these types of events in the following way.

17.9.5 Connector Inspection and Cleaning Processes

At present, there are no published industry standards or guidelines with which to compare connector visual inspection criteria and link performance test methods. Standards created by inspection equipment manufacturers and connector manufacturers typically state tolerances concerning the locations and severity of the defects as well as specifying the magnification used for the inspection. The primary consideration that must be justified in any inspection process is that it must protect the end user (customer) of the product from defects that could potentially lead to field failure. On the other hand, the process

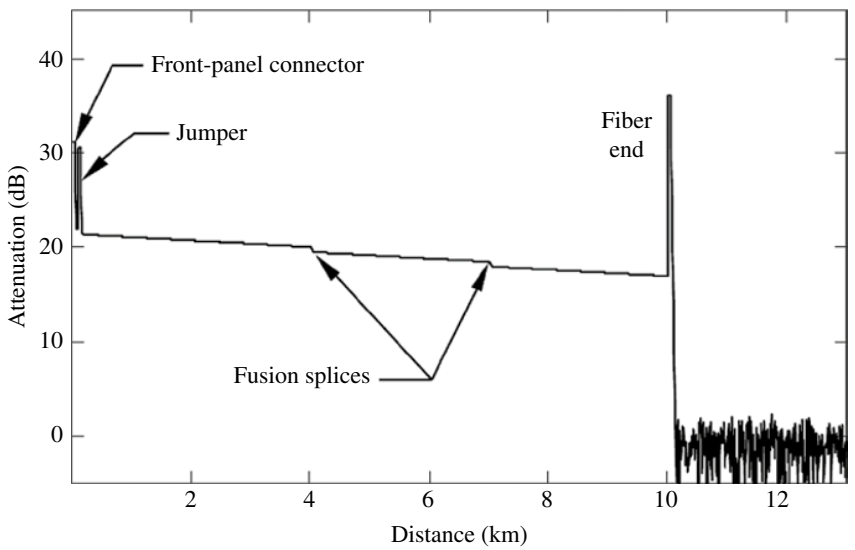


FIGURE 17.23 Waveform as shown on a typical OTDR signal trace.

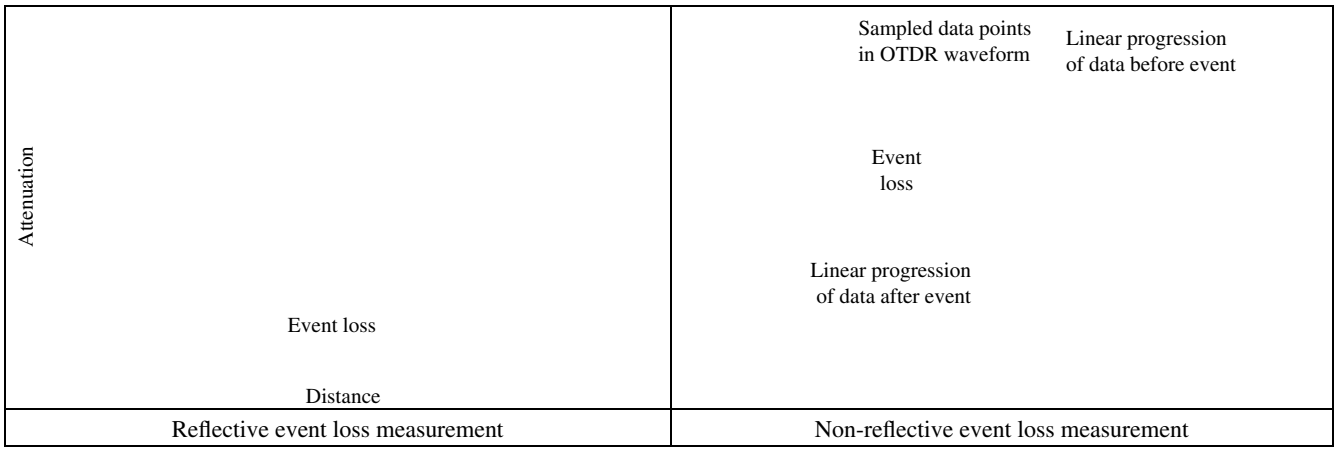


FIGURE 17.24 Reflective and nonreflective event loss measurement.

must also protect the manufacturer from discarding their product due to overly tight specifications that do not correlate with product performance and reliability.

There is standardization around the expectations of "clean" connector end faces that is outlined primarily in IEC 61300-3-35 in the form of inspection zones and acceptable particulate size and counts in each of these inspection zones (Fig. 17.25).

When cleaning fiber components, procedures must be followed precisely and carefully with the goal of eliminating any dust or contamination. Clean connectors function properly, while contaminated connectors may transfer contamination to other components or damage optical surfaces. Inspection and cleaning are critical steps that must be performed before making any fiber connection.

17.9.5.1 Challenges and Industry Norms

We have noted (as does most of the fiber-structured cabling industry) that lack of initial connector efficacy and maintenance of efficacy prior to deployment is a singularly large problem, yielding "plug-and-play" systems (pre-terminated in the factory) that are not truly so. In summary:

- Connector contamination and damage are the leading root causes of fiber-optic network failures.
- Network failures cause downtime and truck rolls.
- These cost money. . . lots of money.
- Inspecting and cleaning before connecting saves troubleshooting costs and downtime and improves performance. Period!

The industry has responded with practice and tools that are focused on troubleshooting and remediation. The basic accepted process across our industry is that a port must be inspected prior to use.

This typically involves the following paradigm: "inspect, clean, and connect" (Fig. 17.26).

This is an "INSPECT-BEFORE-YOU-CONNECT" paradigm to ensure fiber end faces are clean prior to mating connectors:

- Inspect first to determine need for cleaning
- Dry cleaning is quite effective, but is not perfect—so inspect after clean

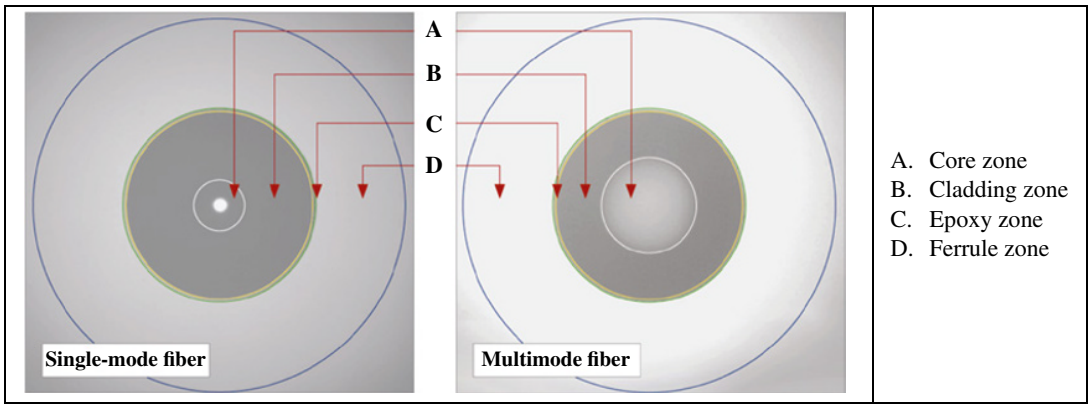


FIGURE 17.25 IEC 61300-3-35 in the form of inspection zones and acceptable particulate size and counts in each zones.

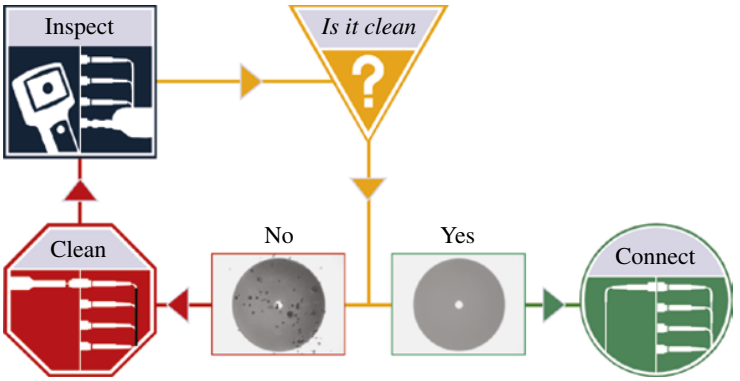


FIGURE 17.26 "Inspect Before You Connect" paradigm.

- Many customers now require proof of inspection (photos) to certify installations
- Verifies pre-connectorized products have been supplied as needed
- Saves time and money in the long run

Historically, for single fiber connectors such as the SC or LC, this process has not been particularly onerous because data centers did not contain tens and/or hundreds of thousands of fibers and, most importantly, these fibers were not deployed in multi-fiber array connectors (such as MPO). Presently, the ecosystem for testing, inspecting, and especially cleaning MPOs is less mature than the comparable ecosystem for single fiber connectors. We would call this ecosystem emergent.

Particularly troublesome is the lack of tools to effectively inspect and clean the pinned MPO. MPO inspection and cleaning tools focus on the contact area of the connector and fail to address contamination that may be present outside of the field of view. Antistatic, reel-based handheld cleaning tools for MPOs (like Optipop etc.) do not effectively clean the areas around the base of the alignment pins. Incompressible debris present in these areas are one of leading causes for open connectors (connectors that display high IL and unstable return loss).

17.9.6 Link Extension and Mixing Fiber Grades

Standards compliant 10 Gb/s networks should deploy consistent fiber types and grades throughout individual channels (ISO OM3 grade fiber). For some applications, it is either impractical or cost prohibitive to remove and replace existing low bandwidth fiber media solutions with a higher performing one (the example of large investment in lower bandwidth jumpers comes to mind and the wish to "recycle" these). In these cases, it becomes important for system designers to evaluate the practicality of hybrid solutions containing mixed fiber grades.

Since the fiber geometries are the same in the case of such a mixed fiber grade network (OM2 and OM3 50/125 μm MM fiber), laser bandwidth of the fiber (stated in EMB) is the differentiating factor that determines the ultimate functionality of a mixed fiber cable solution.

When mixing different grades of 50/125 μm multimode fiber is desired (OM2 and OM3 in particular), a general formula for the OM3 effective/equivalent length of the mixed fiber system is obtained:

$$L_{\text{eff}} = L_{\text{OM2}}\left(\frac{\text{BW}_{\text{OM3}}}{\text{BW}_{\text{OM2}}}\right) + L_{\text{OM3}}$$

where:

L_{OM2} = Length of 50/125 μm OM2 fiber deployed in meters
L_{OM3} = Length of 50/125 μm OM3 fiber deployed in meters
BW_{OM2} = Laser bandwidth of 50/125 μm OM2 fiber deployed in MHz·km
BW_{OM2} = Laser bandwidth of 50/125 μm OM3 fiber deployed in MHz·km

Using the formula above, we can form a matrix of supported distances based on different combinations of discrete lengths of OM2 and OM3 fiber (Table 17.7). For 10 Gb/s serial channels to be supported, the effective length must be less than 300 m. The gray areas in the table below denote those combinations that exceed the channel length specified for a 10 Gb/s serial link built solely of OM3 fiber.

Example:

Would a permanent link formed with 30 m of OM3 fiber cross connected to a 100 m OM4 fiber link with 3 m OM4 jumpers at both ends and at the cross connect be able to support 10 Gb/s?

$$L_{\text{OM3}} = 30\,\text{m}$$

$$L_{\text{OM4}} = 100\,\text{m} + 3(3)\,\text{m} = 109\,\text{m}$$

Therefore,

$$L_{\text{eff}} = 30(4{,}700/2{,}000) + 109 = 180\,\text{m}$$

TABLE 17.7 OM3 "effective link length"

L_{OM3} (m)	L_{OM2} (m)			
20	20	40	60	80
40	100	180	260	
60	120	200	280	
80	140	220	300	
100	160	240		
120	180	260		
140	200	280		
160	220	300		
180	240			
200	260			
220	280			
	300			

Such a channel would be 10 Gb/s standards compliant, since the effective channel length is less than 300 m.

17.10 CONCLUSION

As the requirements for higher speed data center networks have advanced, the specifications on the constituent components have become more stringent. It is critical therefore to have implicit knowledge of the performance of the fiber infrastructure deployed in the network.

Choosing the right components (fiber type, cables, connector, etc.) for higher speed data center applications is a critical decision in futureproofing fiber infrastructure that may have to perform for several iterations of optoelectronics.

Understanding the current requirements and having a view to potential future transceiver technologies must guide the selection of appropriate fiber systems. This will not only assure lowest life cycle costs but also will give the greatest flexibility in supporting higher data rates.

Practices outlined in this chapter have illustrated proper specification of fiber components, effective deployment, management, and test methodologies that minimize risk to customer networks.

REFERENCE

1. ANSI/TIA/EIA-568-B.1-2001, Commercial Building Tel ecommunications Cabling Standards, Telecommunciations Industry Association, May 2001.

FURTHER READING

BOOKS

Hecht J. *Understanding Fiber Optics*. 5th ed. Great mix of basic fiber optic theory, applications, testing and a good view of current and future applications for fiber optics.

Oliviero A, Woodward B. *Cabling: The Complete Guide to Copper and Fiber - Optic Networking*. 5th ed. Text focuses on the practical aspects of fiber cabling systems from an application-based viewpoint.

PAPERS

Reid R. Field testing multimode 10 Gb/s (and beyond) fiber permanent links - Best practices to minimize costs by ensuring measurement repeatability, reproducibility and accuracy. Panduit White paper; April 2011.

Roberts C, Ellis R. Fiber selection and standards guide for premises networks. Corning White Paper; November 2012.

Jonathan Jew, "Data Center Standards: How TIA-942 and BICSI-002 Work Togethern" J&M Consultants, Inc., 2017 BICSI Fall Conference, https://www.bicsi.org/docs/default-source/conference-presentations/2017-fall/data-center-standards.pdf?sfvrsn=367f9892_2 (accessed on 9/28/2020)

WEB

FOA (Fiber Optics Association). Lennie lightwave's guide to fiber optics – hosted by the FOA (Fiber Optics Association) – practical guide to specifying and assessing performance of fiber optic cabling systems.

TIA FOTC (Fiber Optics Technology Consortium). Clearinghouse of white papers, webinars and vendor-sourced materials on fiber structured cabling and data center trends.

18

DESIGN OF ENERGY-EFFICIENT IT EQUIPMENT

Chang-Hsin Geng
Super Micro Computer, Inc., San Jose, California, United States of America

18.1 INTRODUCTION

Computing has changed a lot since its earliest days. The cell phone we hold in our hands has more computing power than the computer used on the Apollo 11 that landed on moon in 1969. Not only has the computing power increased substantially, but also the size of the computer has decreased substantially. In the early years of computing, most people were less concerned with the impact of computer power usage. In those times, improving the performance of compute, storage, and network were the main goals regardless of energy consumption. Today, computing resources are in use in all facets of life including communication, the Internet of Things (IoT), artificial intelligence (AI), manufacturing, Governments, and many other industries. As of today, energy efficiency has become the new mantra but each industry has its own definition and use case. For example, a compute server may be running continuously at a high load, whereas an IoT server may be asked to do something intermittently. Every architecture in computer systems can be made more energy-efficient from design to operation to end of life. In this chapter, we will delve into the process to enable energy-efficient IT equipment and the concept of the design.

18.1.1 History of the Energy-Efficient IT Equipment

ENERGY STAR® was a program launched in 1992 by the U.S. Environment Protection Agency (EPA) to encourage efficient use of energy by IT equipment (ITE). EPA ensures each product earning the ENERGY STAR label has independently certified and delivered the efficiency performance and savings that consumers have come to expect. The ENERGY STAR program is applicable to various electronic products such as televisions, refrigerators, printers, and light bulbs ITE to promote energy efficiency and cost savings.

In a computer server, the CPU (central processing unit) usually consumes the most energy. Intel developed Speed Step technology to tune performance and power consumption in an automated fashion. The firmware and OS (operating system) must leverage the technology to optimize power consumption. During low usage, the CPU enters a low-power mode by decreasing its frequency which consequently decreases the power consumption. Conversely, depending upon the cooling available, the CPU can also enter the "Turbo" mode which increases the frequency beyond the nominal value. ENERGY STAR program facilitates the impetus for energy-efficient design.

18.1.2 The Benefit of Energy-Efficient IT Equipment

Energy-efficient ITE will significantly improve the operating cost of running a data center. If an ITE is designed with high performance and uses less energy, less heat is generated to do the same amount of work and then the operating costs will be lower. An immediate benefit of lower heat generation uses less cooling, thus less cooling cost, to maintain desired operating temperatureas well as longer life of ITE. Another benefit is that the equipment density can be increased based on a fixed amount of cooling which enables more functionality of the ITE. Data center power consumption is approximately 1% of worldwide power consumption and more than 50% [1] of the data center energy is consumed by ITE. Building an energy-efficient computer server can directly reduce the use of electricity in a data center and reduce harmful greenhouse gases.

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

18.2 ENERGY-EFFICIENT EQUIPMENT

Enabling energy-efficient ITE in a data center is challenging because of the complex applications and workloads. This process can either be driven top-down (design of a data center to ITE) or bottom-up (ITE fit to a data center) depending upon the situation. There isn't a single design that can optimize the energy efficiency of ITE because of the complexity of tasks that the Data Center operates. This chapter will focus on the design of ITE used in the data center such as enterprise servers, work stations, other computer systems, cooling solutions (air and liquid), and operations management. The core concept, "more with less," allows computer systems to run more efficiently with equal or better performance and function that will reduce the overall energy usage and consequently carbon emissions.

18.3 HIGH-EFFICIENT COMPUTE SERVER CLUSTER

The core ITE of a data center is the compute servers. A compute server is a catch-all term that includes the storage system, compute system, network system and anything else that directly processes digital data. The cooling system, of a data center for example, would not be part of the cost of doing compute work but would be included in running the data center.

Because each system has its own use case, each system has its own definition of efficiency. One of key metrics is performance per watt where the performance depends upon the use case. For example, an IoT system will have very different system requirements than an HPC (high performance computing) system since the IoT system at edge or fog requires a small footprint with a low wattage CPU. The HPC system design must fit into 19 in wide four-post rack and use high-end CPUs. Online gaming system and software development server are other examples of potential use cases where the online gaming system needs a dedicated GPU (Graphics Processing Unit) for each CPU and the software development server requires high clock speed single-thread CPU. Each use case has a specific system configuration. This gives the system manufacturers a method to quantify their green computing design portfolio. Benchmarks such as LINPACK and PCMark as well as the SPEC Benchmark Suite are well-known benchmark applications that are used to characterize system performance. The server designer can use the benchmark to pre-scan the performance and calculate the efficiency of a server. However, those benchmarks focus on CPU and GPU performance. If the use case is different, the performance score will be based on different criteria defined by the data center engineers and managers. For example, in the bitcoin mining process, the performance of the ASIC (application-specific integrated circuit) to calculate the encrypt hash will be the key to the design of the server.

An equation to calculate the efficiency:

Energy efficiency = performance/total power

Based on the equation, performance is a function of power consumption and efficiency. High performance CPU may not provide the best efficiency due to core count, processor frequency, and memory subsystem efficiency constraints.

Not all data centers can optimize for a single-use case because different applications of systems have varying requirements. For example, for a DL (deep learning) and AI system needs a high processor core count to reduce training time but an inference engine used in AI can do its job with a low core count. If a computer server is running at full load in a general-use case, a standard method will help the IT professional define the energy efficiency of computer servers. The ENERGY STAR program establishes three testing configurations and the results are averaged with different weight of each component to calculate the efficiency [2].

A. **Low-end Performance Configuration**: The combination of processor socket power (PSP), power supply units (PSUs), memory, storage devices, and I/O (input/output) devices is designed to the lowest performance computing platform within the Product Family. This configuration includes the lowest performance processor per socket, as represented by the lowest numerical value resulting from the multiplication of the core count by the frequency in GHz, offered for sale and capable of meeting ENERGY STAR requirements. It shall also include a memory capacity at least equal to the number of memory channels in the server multiplied by the smallest DIMM size offered in the family.

B. **High-end Performance Configuration**: The combination of PSP, PSUs, memory, storage device, and I/O devices is designed to the highest performance computing platform within the Product Family. This configuration shall include the highest processer performance per socket, as represented by the highest numerical value resulting from the multiplication of the core count by the frequency in GHz, offered for sale and capable of meeting ENERGY STAR requirements. It shall also include a memory capacity equal to the value found in equation below:

Minimum Memory Capacity of High-end Performance Configuration =

$$Mem_Capacity_High \geq 3 \times (\#of\ Sockets \times \#of\ Physical\ Cores \times \#Threads\ per\ Core)$$

C. **Typical Configuration**: A product configuration that lies between the Low-end Performance and High-end Performance configurations and is representative of a deployed product with high volume sales. It shall also include a memory capacity equal to the value found in equation below:

Minimum Memory Capacity of Typical Configuration =

$$Mem_Capacity_Typ \geq 2 \times (\#\,of\ Sockets \times \#\,of\ Physical\ Cores \times \#\,Threads\ per\ Core)$$

Active State Efficiency (ASE) Requirements: Calculated Active State efficiency score (Eff_{ACTIVE}) shall be greater than or equal to the minimum ASE thresholds listed in the Table 18.1 for all configurations submitted for certification within a product family, as well as any additional configurations within the product family shipped as ENERGY STAR certified products. The

Eff_{ACTIVE} is calculated as follow:

$$Eff_{ACTIVE} = EXP\left(0.65 \times \ln\left(Eff_{CPU}\right) + 0.30 \times \ln\left(Eff_{MEMORY}\right) + 0.05 \times \ln\left(Eff_{STORAGE}\right)\right)$$

TABLE 18.1 Active state efficiency thresholds for typical of computer servers [2]

Product type	Minimum Eff_{ACTIVE}
One installed processor	
Rack	11.0
Tower	9.4
Blade or multi-node	9.0
Resilient	4.8
Two installed processors	
Rack	13.0
Tower	12.0
Blade or multi-node	14.0
Resilient	5.2
Greater than two installed processors	
Rack	16.0
Blade or multi-node	9.6
Resilient	4.2

where Eff_{ACTIVE} comprises Eff_{CPU}, Eff_{MEMORY}, and $Eff_{STORAGE}$. The **Eff_{CPU}** is calculated as follows:

$$Eff_{CPU} = GEOMEAN\ (Eff_{COMPRESS}, Eff_{LU}, Eff_{SOR}, Eff_{CRYPTO}, Eff_{SORT}, Eff_{SHA256}, Eff_{HYBRIDSSJ})$$

where

$Eff_{COMPRESS}$ = the calculated Compression worklet score
Eff_{LU} = the calculated LU worklet score
Eff_{SOR} = the calculated SOR worklet score
Eff_{CRYPTO} = the calculated Crypto worklet score
Eff_{SORT} = the calculated Sort worklet score
Eff_{SHA256} = the calculated SHA256 worklet score
$Eff_{HYBRIDSSJ}$ = the calculated Hybrid SSJ worklet score

Above sever efficiency formula can be found in document "Server Efficiency Rating Tool (SERTTM) Design Document" [3].

The Eff_{MEMORY} is calculated as follows:

$$Eff_{MEMORY} = GEOMEAN\ \left(Eff_{FLOOD2}, Eff_{CAPACITY}\right)$$

where Eff_{FLOOD3} is the calculated Flood3 worklet score and $Eff_{CAPACITY3}$ is the calculated Capacity3 worklet score.

$Eff_{STORAGE}$ is calculated as follows:

$$Eff_{STORAGE} = GEOMEAN\ \left(Eff_{SEQUENTIAL}, Eff_{RANDOM}\right)$$

where $Eff_{SEQUENTIAL}$ is the calculated Sequential worklet score and Eff_{RANDOM} is the calculated Random worklet score.

To calculate Eff_i is calculated as follows:

$$Eff_i = 1{,}000 \frac{Perf\left(i\right)}{Pwr\left(i\right)}$$

where

i = Represents each workload referenced in result of **Eff_{CPU}**, **Eff_{MEMORY}**, and **$Eff_{STORAGE}$**.
Perf(i) = Geometric mean of the normalized interval performance measurements.
Pwr(i) = Geometric mean of the calculated interval power values.

18.3.1 Anatomy of a Server

Core components of a typical server comprise of motherboard (MB), CPU, memory, storage media, and PSU. Special-use

cases require additional components to streamline the compute process that include GPU, ASIC, and other add-on cards. In today's ITE, most core components are standardized and optimized by individual vendors except the MB and daughter and add-on cards.

18.3.1.1 Motherboard (MB)

The MB is one of the most critical components in a computer server. Anatomically, the MB is like a human body. The CPU, memory, storage media are parts of the brain network in the neural system, and Ethernet is a doorway used to communicate with others. The MB is a printed circuit board Assembly (PCBA). It is the foundation of a computer that connects all other computer components together. The MB connects power and enables the communication between the components such as CPU, memory RAM (random-access memory), storage media, PSU, NIC (network interface card), and all other hardware components.

In an enterprise rack mount server, the MB doesn't have a strict standard like most desktop computers. Hardware manufacturers use their creativity to design motherboards to fit into customized rack mount server chassis. Besides the major components, most of the other I/O interface and function chips such as PCIe (peripheral component interconnect express) slot, fan control pin, USB port, onboard sound chip, and onboard video chip will be modified or removed from the MB. Designing a MB is driven by the use case of the server and limited space in the chassis to maximize efficiency.

18.3.1.2 CPU

CPU is the core component that the main computing performance comes from. Each generation of CPU improves the performance per watt, yet, the CPU in the same family also has its own "sweet spot" for green computing. Top-end CPUs may achieve the top-end performance but may not possess ideal performance per watt. The factors contributing to this inefficiency could be the workload place upon the system. The SPEC's GEOMEAN (geometric mean) benchmark comparison in Figure 18.1 illustrates the point.

There are many other factors such as core optimized, core speed demand, cross CPU compute, and the memory bandwidth. In the end, the main application used in data centers will determine the type of CPU needed to achieve optimal energy efficiency.

18.3.1.3 Memory

Memory is the second component that impacts system performance in a server. While a CPU performs actual work, the memory is the brain that feeds the CPU. Memory subsystem technology has not kept pace with the CPU performance gains and is lagging in the advancement of system performance. Nevertheless, newer memory products increase efficiency by improving performance per watt similar to CPU. For example, standard DDR4 SDRAM (double data rate 4 synchronous dynamic random-access memory) power specification is 1.2/1.05 V and DDR3 is 1.5/1.35 V in the same bandwidth (at the transition between generations point) at initial release. As the technology develops, DDR4 memory has advance on both power consumption and bandwidth over DDR3 memory.

18.3.1.4 Storage Media

Storage media hold long-term data that has been processed by the computer system. While a memory is usually volatile and temporarily holding process data, storage media are mostly nonvolatile. This includes any components in any form factor that store digital data in the computer server. Today's data center mainly uses SSD (solid state drive) and HDD (hard disk drive). Additionally, I/O technologies such as NVMe (non-volatile memory express), SAS (serial attached SCSI), and SATA (serial advanced technology attachment) are used. In order to store complete data from the computing server to the storage server, a storage server requires a large quantity of storage media. Depending upon capacity and the implementation (HDD vs. SSD), some enterprise storage servers (Fig. 18.2) consume more energy than any other equipped computer component. Having large storage capacity introduces another issue: how to place the storage—front-loading or top-loading. A front-loading design system is used where rack density is critical. Failed drives can be replaced easily because it is readily accessible. Front-loading designs have the propensity to optimize the airflow as the drives are placed in front of processors and the memories. A top-loading design system requires an extra step to pull out the server for replacement but enables a higher storage density and energy efficiency (based on data per server). The tradeoff is a top-loading design requiring lower ambient temperature due to additional preheat of storage media and small air path to the processor and memories.

When SSDs were first introduced, their cost performance was disproportionately higher than HDDs. Factors that limit the SSD capacity and cost include short endurance of the flash memory cell scaling. Flash memory cell is a nonvolatile memory chip designed for storing permanent data. It is a floating gate memory cell that can be programmed electrically. It is designed using the idea of EPROM (erasable programmable read-only memory). The benefit of using memory cell is it has less access latency comparing to read and write data by magnetic heads on platters in a HDD. Over time, the SSD capacity and endurance times have significantly improved as the manufacturing technology advanced. Currently, the SSD is at near cost parity with platter-based HDD storage media and has a better performance with lower power consumption (Table 18.2).

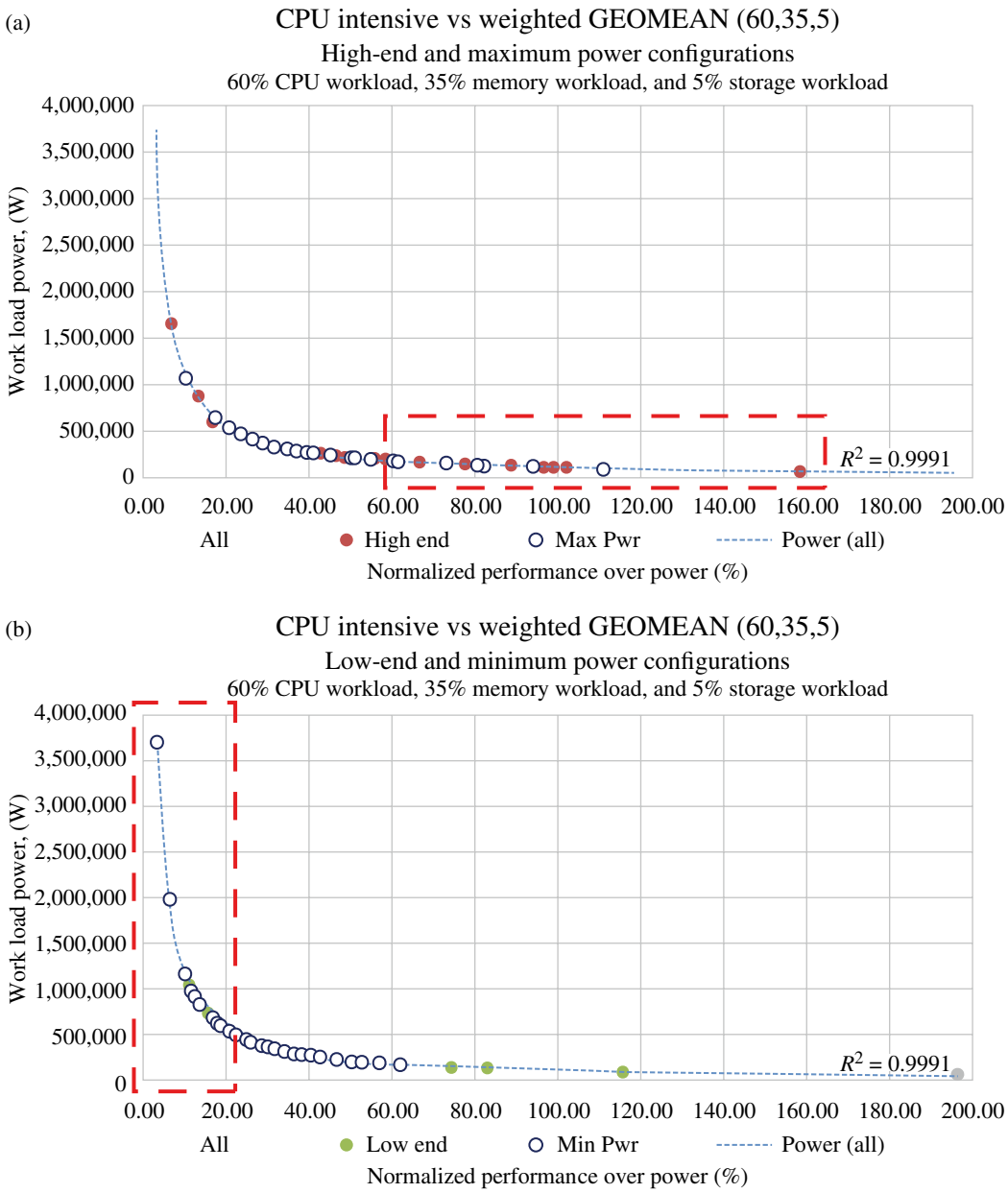


FIGURE 18.1 (a) High-end CPU configurations are more efficient with workload of 60% or higher (in dotted-line block). (b) Low-end CPU configurations are more efficient with workloads of 20% or lower (dotted-line block) [4].

18.3.1.5 NIC (Network Interface Card)

The NIC, also called Network Interface controller enables communication and data exchange between other computers. From the 1980s to early 1990s, computer motherboards did not include built-in networking capabilities. A NIC could be installed in a PCI (peripheral component interconnect) or ISA (industry standard architecture) slot on the MB to provide networking. Today, a 1 Gbps Ethernet NIC is a standard component built-in a MB. Compared to the other main components, the NIC consumes the least power in the overall system. Some high-speed network devices, such as 100G InfiniBand, are not part of the compute server by default. The simplest way to enable energy saving in a network is to pick the network chipset that requires the least power while providing desired functionality.

18.3.1.6 PSU

PSU is the component that converts, controls, and maintains the electricity to a server from data center's outlet and/or

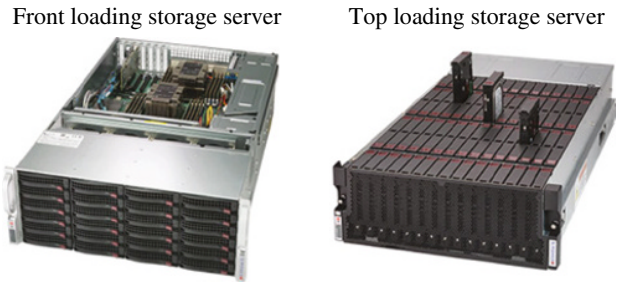


FIGURE 18.2 Front load (left graph) vs. top load (right graph) storage server that is able to install 24–90 3.5 in of SAS3/SATA3 drive consuming 276–1,035 W on drive [5]. Source: Courtesy of Super Micro Computer, Inc.

power bus line. While a PSU does not perform any data-related functions, it is one of the most critical computer system components. The voltage drop, the ripple, and current flow all directly affect the stability, functionality, and life span of a computer server. A good PSU must be designed to minimize heat waste when converting AC (alternating current) to DC (direct current). While the best PSU power load efficiency is fixed at a certain range, a good understanding and calculating the average and peak power consumption of a server before installing any PSU is a critical step to optimize the efficiency of PSU.

The 80 PLUS® is a voluntary certification program that helps identify PSU efficiency. There are six levels of 80 Plus certification: 80 PLUS®, 80 PLUS® Bronze, 80 PLUS® Silver, 80 PLUS® Gold, 80 PLUS® Platinum, and 80 PLUS® Titanium. The standards used by the certification program perform tests on PSU load efficiency at a PSU usage of 10, 20, 50, and 100% rated load. Higher PSU efficiency has less loss during power conversion and consequently less heat generation that reduces electricity costs directly and data center cooling costs indirectly (Table 18.3).

Although AC power is most widely used in the electrical grid, some data centers use DC power infrastructure. Using DC power improves the energy efficiency because power conversion loss to convert AC power to DC for each computer is eliminated. Facebook, one of the technology leaders,

TABLE 18.2 SSD vs HDD in SPEC in February 2020 [6, 7]

Name	Seagate Nytro 1000 960 GB 2.5 in	Seagate Exos 15E900 900 GB 2.5 in
Model	XA960ME10063	ST900MP0006
Type	SSD (SATA 3)	HDD (SATA 3)
Capacity	960 GB	900 GB
RPM	N/A	15,000
Max transfer data rate	564 MBps read	300 MBps read
Active watt	3.2 W	7.6 W
Read perf./W	176.25 MBps/W	39.47 MBps/W

TABLE 18.3 80 PLUS® standard in different rated load with different input voltage

80 Plus	115 V internal nonredundant			230 V internal redundant			230 V EU internal nonredundant		
Percentage of rated load	20%	50%	100%	20%	50%	100%	20%	50%	100%
80 Plus	80%	80%	80%				82%	85%	82%
80 Plus Bronze	82%	85%	82%	81%	85%	81%	85%	88%	85%
80 Plus Silver	85%	88%	85%	85%	89%	85%	87%	90%	87%
80 Plus Gold	87%	90%	87%	88%	92%	88%	90%	92%	89%
80 Plus Platinum	90%	92%	89%	90%	94%	91%	92%	94%	90%
80 Plus Titanium	92%	94%	90%	94%	96%	91%	94%	96%	94%

Source: https://www.plugloadsolutions.com/80PlusPowerSupplies.aspx#.

launched the Open Compute Project that used 480/277VAC distribution to IT equipment and implemented 48VDC UPS System as a new standard for the data centers to maximize efficiency. This is designed from facility to server thus removing the unnecessary AC to DC or DC to AC conversation.

18.3.1.7 GPU

GPU is a specialized accelerator (or "soul" called by NVIDIA) for applications such as entertainment games, data analytics, and AI. While CPU core count reaching one hundred is considered a high-end product, GPU's provide a much higher core count. For example, while AMD EPYC 7702P has 64 cores but NVIDIA Quadro GV100 has 5,120 cores. Consequently, the power consumption in a GPU is much higher than a CPU that one can be benchmarked by using speed tester such as GPU UserBenchmark (https://gpu.userbenchmark.com/Software).

18.3.1.8 ASIC

ASIC is an integrated circuit chip used to designed and customized for a particular purpose. This can be data encryption/decryption or uniquely patterned data processing such as high-efficiency bitcoin miner. Today's data centers are moving to highly efficient SDDC (software-defined data center) for two main reasons. First, a data center is not locked into one specific vendor. Second, ASICs can be used to accelerate certain tasks such as system-level bottlenecks. ASIC applications could be server load balance or a security gateway. Routing is another application that can benefit from the performance of an ASIC. Compute servers combined with ASICs can optimize the performance per watt providing a sweet spot of server efficiency.

18.3.2 Concepts of Design

A use case determines the design of compute servers in a particular data center. Different types of systems have different designs and application focus thus different energy efficiency even though the basic components discussed previously may be the same. Instead of single computing performance from a CPU or benchmark of any single component in a server, the GEOMEAN metric puts different percentage weight to each component to simulate a complex environment that combines CPU demand application, memory bandwidth demand application, memory capacity demand application, and storage demand application. The total benchmark score divided by total power consumption is the general energy efficiency of the server. A good example could be found in the "Recommendation for Server Activity Efficiency Metric" by ENERGY STAR® [4]. Most IT professional can use the baseline Metric from Energy Star to determinate and design the best server configuration for their data center.

18.3.2.1 CPU and Memory Computer Server

A compute-oriented server has sockets for the 1/2/4 CPU, memory, and local storage. Available CPU TDP (thermal design power) has an upward trend from about 55 W for the Pentium to 205 W for the Xeon Scalable 2nd generation processor. Designers have had a lot of experience in designing such systems with the inlet air flowing over passive heat sinks of the CPU. In general, the air flows first pass over storage media at front of the server. The inlet air also flows over the DIMMs which are in-line with the CPU. Another trend has been the use of heat-spreaders over the DIMMs. The air then flows over VRM (voltage regulator module) and other parts on MB before being expelled from the back of the system. This has added benefit as it allows a data center to alternate cold and hot aisles thus optimizing hot and cold air management.

Since server cooling is a critical issue, designers are looking to leverage alternate cooling solutions, such as active cooling which incorporates a fan in the CPU heat sink. However, this causes turbulent airflow over the MB components and limits cooling efficiency. Liquid cooling is an alternative approach designers considering which is ideal for high TDP processors but it increases design complexity at the rack level. This is partly because the liquid cooling invariably must be isolated from electrical components.

18.3.2.2 GPU Computer Server

The GPU-based system provides extreme challenges for designers. Each GPU has a TDP up-to 300 W nowadays not counting watt needed by the CPU. Total power consumption of a single system can be approximately 3 kW in a 4U rack-mounted server. This means the GPU will generate a lot of preheat for components placed after it, and the placement of the GPU will be creative. Additionally, the GPUs reside in the PCIe slots, which have to be in-line to the airflow because most enterprise GPU cards usually has passive cooling (heat sink on GPU). Spacing between the PCIe slots becomes a competing requirement as insufficient space may cause system performance issues and different GPU cards have different heights in slots.

18.3.2.3 Storage Computer Server

Storage servers have a bigger challenge of blocked airflow because of the U.2 interface standard between SSDs and computers. The CPUs are not the limiters of airflow. The front to back airflow design allows cold inlet air flowing over storage blocks then over CPUs and memory then exit out of a server. The result is the storage blocks in front of server need to be well design, overwhelming storage blocks increase the number of storage media but will block the inlet air for the entire server. Newly proposed form factors, such as developed by the EDSFF (enterprise & datacenter SSD Form Factor) Working Group, help increase density while limiting airflow-related issues.

Different than the early defined 2.5 in HDD/SDD form factor, it's thin, flat, and long form factor looks like ruler, which Intel simply gives it the name ruler before it formally announces.

18.3.2.4 Network Computer Server

Network-oriented computer server primarily handles network traffic for applications such as security scanning, load balancing, or routing. Typically, a large number of high-throughput network interfaces will be placed in the front of a network server because the network chipset is sensitive to heat, and thus should be placed before any heat-generating sources. Network-oriented servers are fewer compare to other type of server in most data centers. With the move to SDDCs that are designed with more powerful CPU and converged infrastructure, this specific problem is being mitigated as the similar cooling issues affect storage, networking, and compute components in data centers.

18.3.2.5 ASIC Computer Server

An ASIC-oriented server can have conflicting requirements. Sometimes the ASIC can consume a large amount of power and can be paired with a lower power CPU to enable boot up. The overall system power is lowered even though the total power of one component is high. In this scenario, the inlet cool air would flow over the ASIC first then subsequently the CPU. The memory placement is critical as either the number of cores is larger or the frequency of each core is higher (both leading to higher power consumption). Locating the memory farther from the computing unit degrades performance. Locating the memory closer to the computing unit improves performance; however, more consideration of optimization is required in the cooling solution.

18.3.2.6 Expandable Compute Server

An expandable compute server is designed to handle as many different functions as possible but not to a specific use case. These servers have various expansion slots that can be customized as desired. A GPU PCIe add-on card is a good example of such a system. It can be used to enable AI applications. However, the power required by GPUs can exceed the demand for CPU requirements. The PCIe slot that provides enough power must also provide airflow for system cooling. Thus, these systems are optimized differently and may not be ideal energy efficient for *every* application. Some more examples include: accelerator card can be added to expansion slots to *accelerate* certain applications such as FPGA accelerate network encrypt/decrypt speed. FC (fiber channel) PCIe card enables the capability to connect to an external storage server. High-speed fiber optic network card enables large bandwidth of data exchange between servers. The board and server layout must consider required factors and enable cooling on all scenarios. As a result, this approach can't be optimized to accommodate extra or specific applications.

18.3.3 High Efficiency Operation

A compute server may not run at optimal efficiency even devising an optimal design. This depends on how the operation of the server utilizes the compute resource in CPU, memory, storage, and other components.

18.3.3.1 Efficient Compute Resource Usage

Building an energy-efficient computer does not mean it is operating at an optimized setting or workload. This requires the administrator (or user) to research and control the workload on an individual server to run in an optimal configuration to enable the highest performance per watt. For example, an 80 PLUS® Titanium certified device has the best power conversion efficiency at 50% load at 94% efficiency compared to 100% load at 90% efficiency under 115 V internal nonredundant scenario. Managing resource usage at the system level (hardware and software) is the most important phase in building energy-efficient equipment. This can be best summarized as: Minimize operating and acquisition costs (air condition, lighting, electrical system) while maximizing computing of CPU, memory, and storage.

Many data centers are equipped with abundant energy-efficient equipment, but if the workload is run on a wrong system, inefficiency or lower performance per watt results. For example, running GPU optimized rendering application on CPU optimized server. Another way to optimize data center power consumption is to automate a server cluster via the BMC (baseboard management controller) to power on servers as needed and shut down when the job completes. However, this approach of switch computer/server on and off, will cause power ripples and wear on the systems. Virtualizing physical hardware resources and running multiple VMs (virtual machines) on the same physical hardware is a prevailing way to efficiently utilize resources.

18.3.3.2 SDDC

SDDC is an inevitable way to design the data center infrastructure. Starting from converged infrastructure to hyper converged infrastructure, the main concept is to pool the CPU, memory, and storage resources and utilize them together rather than as individual components. This flexibility enables the same server to be used for storage or as a compute server depending on the software configuration. The hypervisor technology providing virtualization has been evolving and is easy to use while improving efficiency. Container technology (CT) makes the deployment easier overall. When the entire application is containerized, deployment time reduces and system-level efficiency increases.

VDI (virtual desktop infrastructure) consolidates many computer workstations into one server host in a data center. Most of workstation workloads do not consume 100% of the CPU power. Hypervisor and VDI allows CPU cycles to be shared to keep the computing load at 80% and temporary peak load at 100% on an average. A CPU average load of 90–100% is a warning sign that the CPU is approaching an overloaded condition. Exceeding the load may lead to the system being throttled as the cooling may be unable to maintain CPU junction temperature. Lastly, the best hardware or software control is based on the application running in a data center.

18.3.3.3 Recycle Energy

Recycling thermal energy generated from the server is another pillar of green computing that data center facility design must be considered. For example, the heat generated from server systems can be used to heat facility water or maintain office temperature.

18.4 PROCESS TO DESIGN ENERGY-EFFICIENT SERVERS

Designing energy-efficient compute server begin by understanding the IT equipment and data center requirements. As one of the core equipment that consumes up to 90% of data center energy, the server design must be carefully reviewed and validated with defined specifications and standards. A recommended new product management process is shown in Figure 18.3.

18.4.1 Initiation

Key members to initiate the design of a computer server include stakeholders from customer, product manager, software/hardware design engineer, thermal engineer, mechanical engineer, and compliance team. This team is responsible for defining product vision, project scope, design, and release.

Core engineers are crucial to ensure delivered voltage is properly conditioned to avoid voltage ripples or any electric noise which may cause system instability and consequently data loss. System thermal dynamics is another important aspect to be considered. Unclear product vision from marketing or customer will lead to product failure.

18.4.2 Planning

Once IT server requirements are defined by the customer, system specification and project schedule will be discussed with computer server manufacturer from the very beginning. All the customer's data center background information such as application workloads, elecrrical network infrastructure, thermal solution, operations preference, and main goals are collected to optimize IT server design.

New product initiation or kick-off defines main purpose of the desired server, the product scope and life-cycle product concept, release schedule, requirements of new technology or design, and development cost. Subject matter experts (SMEs) at design house brainstorm the best possible solution to the project. Issues such as the major components, form factor, and thermal solution are aligned with the data center where the servers will be installed, and the baseline of performance per watt are also determined.

As a team, SMEs optimize the design based on past experience and innovative solutions. Applicable technologies for CPU, memory, I/O standard, materials, networking, connect interfaces, and VRM (voltage regulator module) and thermals are validated. The team reviews all proposed solutions, converges on a specification, and moves to prototype phase.

18.4.2.1 Embracing New Technology in Computing

Technologies in general are evolving at an unbelievable rate in the high-tech industry. AI has been around for almost three decades but was no closer to practical application until the middle of the last decade. Suddenly, ML (machine learning) and DL (deep learning) technologies are in the mainstream. Less than a decade ago, Alexa, the VA (virtual assistant), applied

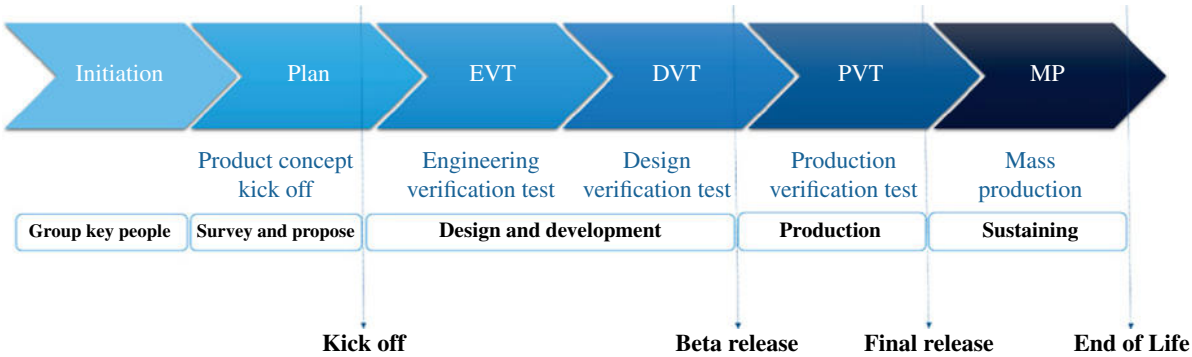


FIGURE 18.3 A process to produce a qualified energy-efficient server from initiation to production release. Source: © 2020 Chang-Hsin Geng.

AI technology. Since then, it has captured a significant market share in homes. Other competitors such as Google Home (Google) and Portal (Facebook) have introduced their own version of virtual assistants. It is important to note that about a decade ago, using VAs were nearly impossible because the computing resources used to drive ML or DL were unfathomable. The performance gains and system efficiency led to reduction in energy while increasing compute power. Successive generations of CPU's provide a 15% performance gain and efficiently managing power usage. The overall performance per watt is thus higher which implies a saving in electricity costs. The new design enables component reuse and reduces e-waste when the computer is refreshed. Another good example is the power delivery network. Normally, power is delivered in the AC mode but the system uses DC power which leads to power conversion losses. In the Open Computing Project® initiated by Facebook, specifies data center uses 480 V DC PSU thus reducing AC to DC conversion energy loss.

18.4.2.2 The Longevity of Product Life

Some parts of a compute server in a data center can be reused across generations. The PSU is one of the most common parts that can be reused over many years. Simplifying the PSU connector across system generations promotes reuse. Designed for higher efficiency PSU reduces energy loss during power conversion from AC to DC. Beside PSU, other components' design can be modularized for reuse. However, this comes at a cost, mainly the connection speeds across the blocks. A blade system is a good example of modular design for PSU's, chassis, switch, and central chassis management. This enables longer life cycle which means most of the modules can be used across generations. The only trade-off is the higher initial investment cost of the IT system. However, that may not always be the case. In some case, a fully populated blade system will be equivalent to a similarly configured rack-based system. The product longevity benefit will be realized during the refresh cycle.

18.4.2.3 Server System Design

Taking away the major standard components such as CPU, memory, storage media, and NIC that are developed by major vendors, server system structure layout design has major impacts on a server's functionality and efficiency. To create a successful server, the design must meet key criteria such as product spec, form factor, MB layout, and thermal design. An initial concept design of the computer server block diagram and spec is defined in this stage. Oftentimes, this stage is where most of the projects fail, redesign, or redefine the product spec.

All hardware, software, mechanical, and thermal engineers review the block diagram and validate the design based on their expertise. If the prototype cannot meet the defined spec and standards, engineers will provide feedback and restart collaborations.

18.4.2.4 Server Mechanical Design

The server chassis form factor is the first design component that requires team consensus. It directly impacts MB layout and thermal design. The chosen form factor can lead MB and thermal design of the server in a different direction than originally intended. To create a space-efficient solution, AT&T first established a standard for racks in 1922 that is a 19-in rack format with rack unit (RU) of 1.75 in height and 19 in wide. After several evolutions, today's data center follows the IEC 60297. This changes the data center computer form factor from tower server to rack mount server.

In general, today's servers leverage a 1, 2, or 4 RU's chassis. Most of the boards in use leverage single or dual air-cooled CPUs where the inlet air directly cools the CPU. More exotic cooling solutions for high performance CPUs use liquid cooling at the rack level.

18.4.2.5 Motherboard and Extension Card Design

MB layout design impacts the server thermal solution. In some cases, it can lead to a redesign of the chassis mechanical design. A good layout design will have signal integrity, thermal stability, and minimize power supply ripple. Components such as VRM used on the MB impacts electric efficiency of a MB. High-speed signals call for shorter line length to minimize power consumption and noise due to crosstalk. Crammed system components penalize airflow. These competing issues are a constant challenge to MB and system designers.

Modern data centers are designed to enable higher cooling and higher power capacity to deploy higher density of TE. Blade servers, have a higher density and energy efficiency. It consolidates the power module, switch module, system fan module, and a central management module for multiple servers. The limited space in the chassis presents challenges for MB layout as well as to thermal engineering. Twin server released in 2007 from Super Micro Computer, Inc. is an alternated design that increases the server density in a data center on standard 19 in rack. This design provided more space saving and server density.

Every server MB design is based on the specification defined during the planning stage. A system-level block diagram determines the system components' placement on the MB. Other issues such as electrical and thermal design along with costs are traded off at this stage. Any mistakes not accounted for will need a redesign and thus delay bringing the product to market. A suboptimal board design would impact system performance. Higher number of PCBA layers can be used to limit the signal routing but will increase board costs and negatively impact thermal design.

18.4.2.6 Thermal Solution Design

Cooling is a critical consideration that plays an important role to keep the server running at a stable temperature. Cooling is the number one bottleneck that needs to be addressed for HPC. The most common solution is to streamline airflow over a passive heat sink attached to the CPU by thermal paste. In some scenarios, the fan is built in the heat sink to target the cooling solution close to the heat source. Liquid or hybrid cooling is another solution enabling even higher performance as it improves over conventional air-cooled solution.

Air cooling with passive heat sink is one of the oldest, yet stable thermal solutions for CPU's. Different components of a server system have different needs of cooling. Power consumed which translates to heat is directly proportional to frequency. The CPU normally has the highest frequency thus the highest TDP and needs the most cooling. Various components on the MB, CPU, memory, storage media, and NIC all have different tolerance to temperature. Different components need different cooling, and effectively guiding airflow is needed. For example, memory requires less cooling compared to CPU, so laminar airflow over CPU heat sink is a way to optimize the thermal solution. Without proper cooling for each major component, it will lead to poor system performance or failure. The active air cooling depends on data center's ambient air temperature and the system fan module and speed in a server. The higher the fan speed or more fan modules, the higher the wattage use (also noisier to the system) thus less efficient the server. If a system demands lower inlet air temperature, higher refrigeration will be required. Both outcomes are energy inefficient and design should be avoided.

As shown in Figure 18.4, Blade server 1 has air cooling with a memory air shroud to guide and focus the air over the CPU heat sink and minor path over memory to transfer more heat out. This configuration reduces the system fans or fan speed required to cool, and thus reduces unnecessary wattage use and increases the overall system energy efficiency. With the air shroud removed in Blade server 2, the inlet air is not efficiently transferred to the necessary components in the server and causes rear CPU to overheat easily. The CPU is the main component that requires additional cooling compared to the memory. The gap between memory DIMMs is larger than CPU heat sink fins, therefore more air will pass through the memory DIMMs. Due to the airflow being distributed over all the system components in Blade server 2, more heat will be retained in CPU heat sink subsequently the CPU runs hotter. The overall CPU cooling system is suboptimal leading to the exit air being warm compared to hot in Blade 1. In the past, data centers had very low ambient temperature so the cooling solution didn't need to be optimized in a computer server but it demands higher energy consumption in data centers. With optimized cooling in the thermal solution design, data centers will use less energy to maintain the operations.

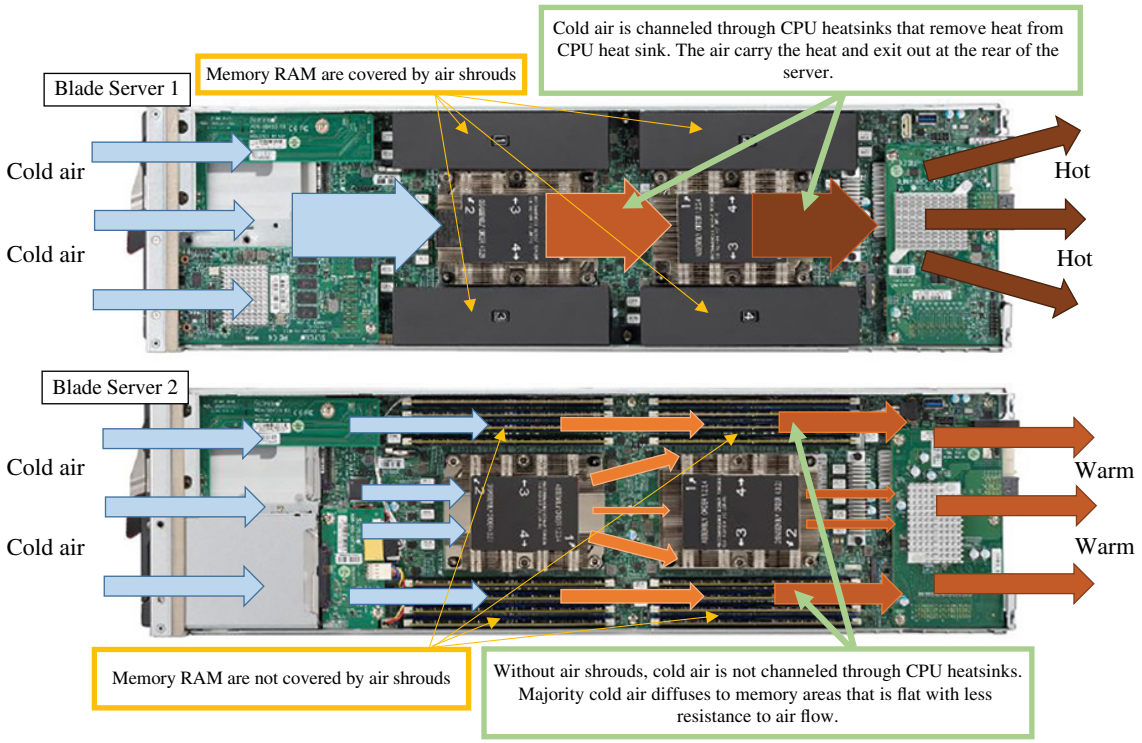


FIGURE 18.4 The Blade 1 motherboard design is a preferred design due to airflow is focused on CPU heat sinks that remove more heat from the memory. The Blade 2 has more heat trapped inside the blade due to air diffusion [8]. Source: Courtesy of Super Micro Computer, Inc.

The more uninterrupted the airflow is, the higher cooling efficiency will be. Yet, many times there are components and daughterboards in a computer server blocking airflow leading to air turbulence. That complexity creates dead zone or high wind force resistance or friction resulting in insufficient air cooling. A simulation to validate the concept of a design in developing and engineering stages becomes critical. In this situation, CFD (Computational fluid dynamics) software is an extremely important software tool that helps design the thermal solution for the computer server. AutoDesk CFD, Future Facilities, and Simcenter STAR-CCM+ are some of the well-known simulation software that helps define thermal management for electronics cooling.

Blades typed computer system have an advantage over conventional rack-based systems due to their modular design. In the design of the blade-form factor server, the air cooling will be pushed to the next stage. Instead of many individual system fans, blade-modularized system fans enable the ability to share the fans over many blade nodes (Fig. 18.5) and reduce energy usage.

The simplicity and minimal I/O motherboard design are the other key factors to reduce electric usage. Blade systems have this trait and are proven in SERT (server efficiency rating tool) as shown in Figure 18.6.

However, when the MB component density increases or a CPU with a higher TDP is used, air cooling becomes difficult and liquid cooling must be considered. The main concept of liquid cooling is to remove heat with liquid. Major CPU and other components are placed under a water block and connected to each other to build a liquid loop. Vendors provide different solutions; some liquid cooling solutions build the liquid pump into the water block on the CPU while others build a passive water block with a centralized water pump host in the facility. There are two typical water loops: a loop is within the rack cooling radiator or a loop transports water to an off-rack facility where the water is chilled before returning to the rack to cool the systems.

18.4.2.7 Proof of Concept

Before proceeding to test and validation, the next step is to build a prototype or "engineering sample," which is essentially a low quantity (1–10 units) run combining engineering and design. The goal is to create an MVP (minimum viable product) the engineers can confidently claim. This is only a few iterations away from mass production. The cost of a prototype will be much higher than final production run because of the small lot size and lack of production optimization. Building a prototype will provide the design team with a database of parts and model files that can be sent to the CM (contract manufacturer) for an RFQ (request for quote).

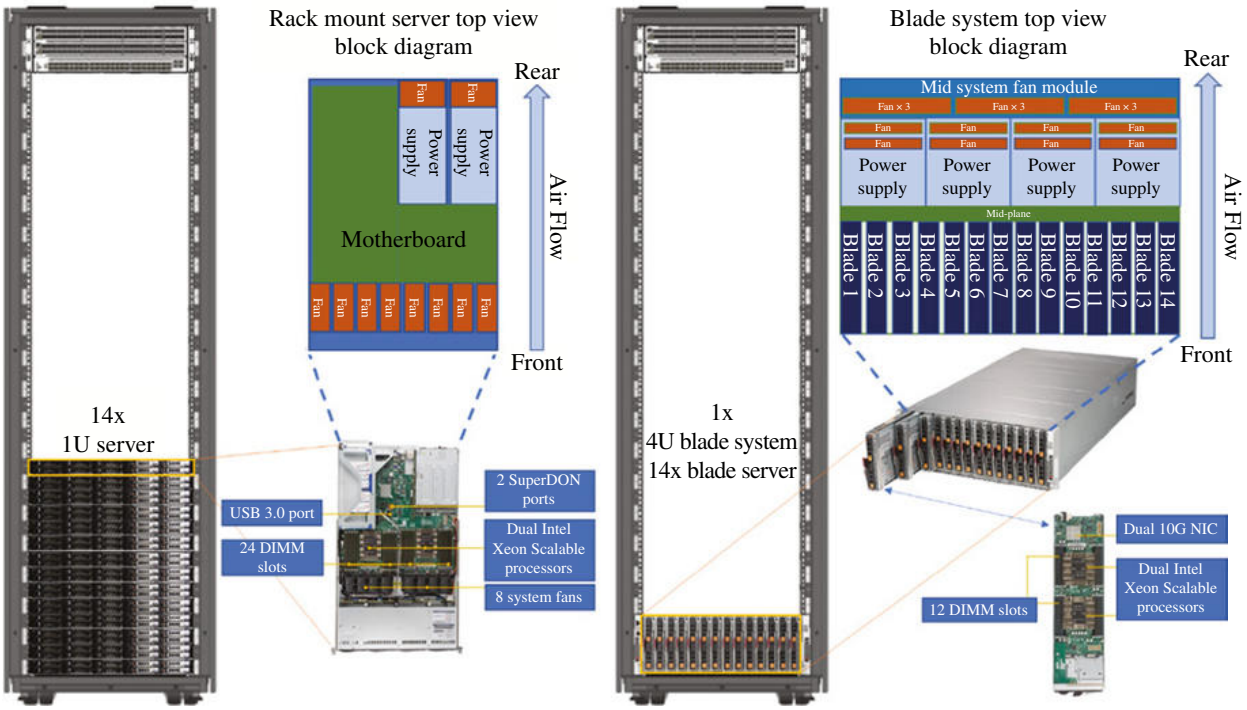


FIGURE 18.5 Both 14x 1U servers (left graph), each server has eight system fans and two PSU fans and 1x 4U Blade system with 14x blade server (right graph) equipped with three system fans plus eight PSU fans that have same compute performance in same elapsed time. But the blade system (right graph) uses less electric power and smaller space. Source: © 2020 Chang-Hsin Geng.

	CPU							Memory		Storage			Power	
Server	Compress	CryptoAES	LU	SOR	XML validate	SORT	SHA256	Flood	Capacity	Sequential	Random	SPECPower SSJ hybrid	Maximum power	Idle Power
4S - Resilient median:	**20**	**15**	**19**	**26**	**20**	**30**	**19**	**336**	**1114**	**23**	**7**	**27**	**1156**	**752**
4S - Blade median:	**34**	**27**	**27**	**38**	**26**	**34**	**25**	**719**	**1628**	**94**	**68**	**36**	**3188**	**766**
4S - Rack median:	**28**	**113**	**24**	**22**	**22**	**31**	**29**	**183**	**665**	**46**	**22**	**38**	**718**	**221**
%CH - Rack and Blade	–18%	–319%	–11%	–41%	–14%	-9%	18%	–75%	–59%	–51%	–68%	3%	–77%	–71%
2S - Resilient median:	**16**	**12**	**15**	**21**	**16**	**23**	**15**	**288**	**760**	**50**	**17**	**21**	**591**	**300**
2S - Blade median:	**39**	**32**	**31**	**43**	**30**	**38**	**29**	**107**	**358**	**129**	**79**	**44**	**536**	**199**
2S - Rack median:	**32**	**30**	**30**	**41**	**27**	**36**	**32**	**89**	**180**	**83**	**63**	**34**	**307**	**134**
%CH - Rack and Blade	–18%	–4%	–4%	–6%	–11%	–5%	10%	–17%	–50%	–35%	–20%	–23%	–43%	–33%
1S - Rack median:	**33**	**36**	**30**	**35**	**29**	**31**	**31**	**25**	**54**	**113**	**60**	**38**	**117**	**46**

FIGURE 18.6 Median SERT (server efficiency rating tool) values by classification. SERT results in different configurations and classes show high-density design such as blade system having greater energy efficiency than rack systems. Nevertheless, the resource sharing, optimized cooling, and space optimization apply on a general rack mount server and improve the individual server's energy efficiency [9].

18.4.3 EVT (Engineering Validation Test)

This stage is mainly a concept and basic functionality validation stage. In this stage, about 20–50 set of units will be built for validation, yet, the number varies and depends on individual company's strategy, resource availability, and budget. Usually, EVT validates the first version of MB, cable, extension cards, thermal solutions, and mechanical mock-up. Multiple teams such as hardware engineer, software engineer, thermal engineer, and validation lab will work together to validate with the function of the server. The validation tool may vary and depend on both of vendors and companies such as mechanical interference, signal integrity, power, and thermal limitation. If an engineer encounters an issue that cannot meet the design specification and cannot be resolved by the information, the design must be returned and modified.

18.4.4 DVT (Design Validation Test)

In this stage, about 50–200 sets of units will be built for validation. The unit number varies and depends on individual company's strategy, resource availability, and budget. Validations at complete design include vibration test, performance test, thermal test, safety test, reliability test, compliance test, and software certifications. Based on test results, most functional issues will be found and alpha samples will be available for customers to review. Major mechanical tooling will be reviewed and prepared by CM at this stage. Minor design change may be released based on the feedback.

18.4.5 PVT (Production Validation Test)

This stage is to validate the designed products along with production process from sales orders, manufacture orders, production tools, and shop floor systems to logistics. Sample units are intended to be as close to production unit and can be manufactured at volumes and target cost. If all PVT units pass the validation test, the next phase will be the pilot run. Ideally, these units will be suitable to sell and will become part of a volume ramp. Furthermore, the QA (quality assurance) and QC (quality control) procedures are developed in this stage that allow the PE (production engineer) and CM to check for any failures throughout the manufacturing process. This is the last opportunity for adjustments to be made. Once performance and quality have been verified and signed off, the product will move to mass production.

18.4.6 MP (Mass Production)

System design to be used for mass production are posted in PVT. Lessons learned during the design process are captured into the SOP (Standard Operating Procedure). Minor changes may be made to improve production yield. Information about defects and defective parts in the field will be incorporated into the manufacturing process to further improve system quality. Furthermore, as test data becomes available, any system-level errors that require fine-tuning the design can be incorporated into the design after appropriate validation.

18.5 CONCLUSION

In today's data center, servers consume most of the electrical power and thus the energy efficiency of its operation is paramount. Designing an efficient server requires a holistic approach, combining with a heuristic approach when it is impossible or impractical to be optimal, that includes system requirements, mechanical and thermal

design, prototyping, and validation. Considerations of facility thermal and power solutions in a data center drive additional server system requirements. Conceptual design and simulation are key aspects to design energy-efficient servers that conserve resources, minimize operating costs, and save our earth.

ACKNOWLEDGEMENT

The author wishes to thank Dennis Oshiba at A10 Networks, Inc., Vivek Joshi and Chi Kuo Alexander Yen at Super Micro Computer, Inc. for their time and help in preparation of this chapter.

REFERENCES

[1] IT Equipment Electricity Use of Datacenter. Available at https://energyinnovation.org/2020/03/17/how-much-energy-do-data-centers-really-use/. Accessed on May 16, 2020.

[2] ENERGY STAR® Program Requirements for Computer Servers. Available at https://www.energystar.gov/sites/default/files/ENERGY%20STAR%20Version%203.0%20Computer%20Servers%20Program%20Requirements.pdf. (accessed 2/22/2020).

[3] Server Efficiency Rating Tool (SERTTM) Design Document SERT. Available at https://www.spec.org/sert2/SERT-designdocument.pdf. Accessed on February 22, 2020.

[4] ENERGY STAR®. Recommendation for Server Activity Efficiency Metric Version 3 Draft 1. Available at https://www.energystar.gov/sites/default/files/ITI%2C%20SPECpower%20Committee%20and%20TGG%20WG%20Comments%20-%20Part%203.pdf. Accessed on July 4, 2020.

[5] SuperMicro Front Load Storage Server and Top Load Storage Server. Available at https://www.supermicro.com/en/products/general-purpose-storage https://www.supermicro.com/en/products/system/4U/6048/SSG-6048R-E1CR90L.cfm. Accessed on May 16, 2020.

[6] SSD SPEC. Available at https://www.seagate.com/www-content/datasheets/pdfs/nytro-1351-1551-sata-ssdDS1992-4-1907US-en_US.pdf. Accessed on February 22, 2020.

[7] SDD SPEC. Available at https://www.seagate.com/www-content/datasheets/pdfs/exos-15-e-900DS1958-2-1710US-en_US.pdf. Accessed on February 22, 2020.

[8] SuperMicro Blade server SBI-4429P-T2N. Available at https://www.supermicro.com/en/products/superblade/module/SBI-4429P-T2N.cfm. Accessed on May 16, 2020.

[9] ENERGY STAR Data Center Server Meeting Initial Insights from SERT Server Results. Available at https://www.energystar.gov/sites/default/files/specs/EPA%20Analysis%20of%20SERT%20Data%20.pdf. Accessed on February 22, 2020.

FURTHER READING

Energy-Efficient Enterprise Servers. Available at https://www.energystar.gov/products/data_center_equipment/enterprise_servers. Accessed on May 16, 2020.

Green IT Factsheet. Available at http://css.umich.edu/factsheets/green-it-factsheet. Accessed on May 16, 2020.

PERFORMANCE STANDARD

https://www.cpubenchmark.net/.

https://www.spec.org/cgi-bin/osgresults.

https://www.techpowerup.com/review/nvidia-geforce-rtx-2080-ti-founders-edition/34.html.

PSU 80 PLUS

https://www.plugloadsolutions.com/80PlusPowerSupplies.aspx.

GPU

https://www.nvidia.com/en-us/design-visualization/technologies/rtx/.

WATER COOLING SEVERS

https://www.ibm.com/support/knowledgecenter/POWER8/p8had/p8had_wc_overview.htm.

https://www.gigabyte.com/solutions/cooling/asetek-liquid-cooling.

https://www.asetek.com/.

https://www.coolitsystems.com/rack-dlc/.

19

ENERGY-SAVING TECHNOLOGIES OF SERVERS IN DATA CENTERS

Weiwei Lin[1], Wentai Wu[2] and Keqin Li[3]

[1] *School of Computer Science and Engineering, South China University of Technology, Guangzhou, China*
[2] *Department of Computer Science, University of Warwick, Coventry, United Kingdom*
[3] *Department of Computer Science, State University of New York, New Paltz, New York, United States of America*

19.1 INTRODUCTION

We are enjoying various kinds of online services that offer so much fun and convenience to our life—in a large part, these should be credited to those enterprise-owned data centers though we may never be conscious of where they are. Speaking more precisely, it is those metal "labors" racked row upon row working 24 hours a day that uphold every demand of us. When you upload the latest holiday photos to Facebook, there's a chance they'll end up stored in one of the tens of thousands of servers in Prineville, Oregon, a small town where the company has built three giant data centers and two more are in plan. Undoubtedly, data centers are making life better, but benefits always come at a cost—data centers are gobbling up our energy. Already, data centers use an estimated 200 terawatt hours (TWh) each year, while Anders Andrae, a specialist in sustainable Information and Communication Technology (ICT), forecasts that the energy demand of ICT will accelerate in the 2020s, approaching 9,000 TWh by 2030 among which data centers will take a large slice. With this trend, topics related to energy conservation have gradually become the focus when people are talking about data centers.

For many reasons, people have designed multiple architectures of data centers such as containerized data centers, industrialized data centers, and traditional data centers. However, the core component in every data center never changed—the primary reason why we built data centers is always to retain servers. Despite the complexity of supporting infrastructures and facilities like Power Distribution Units (PDUs), cooling systems and lightings, data centers are just like shells that provide space for racking physical servers in rows and all necessary conditions to maintain them running 365 × 24 all year around. Playing the most important role in data centers, servers also account for the majority of data center energy consumption—at around 40% of the total and even higher in well-planned data centers with excellent natural/artificial cooling conditions. A single rack of servers can consume more than 20 kW, which is equal to the average power of 35 common households in Austria. In addition to that, the reported power density of server racks is still growing as engineers continue to compact the space but add more servers into modern data centers. All these facts indicate that improving server energy efficiency is a top-priority task when it comes to the effort on energy conservation (and cutting electricity bills, of course) for a data center that has already been operating in the first place.

Energy efficiency can hardly be achieved unless we are able to be precisely aware of how much energy has been used and how quickly our servers are consuming electricity. This means that appropriate setup of power/energy consumption monitoring modules serves as the prerequisites of data center sustainability. Nevertheless, more focus should be cast on how to implement a flexible, fine-grained monitoring system that is easy to scale up with the data center. On this point, we argue that software-based implementation is the right path considering all its advantages (e.g., low cost, scalability and compatibility) match what we need for a decent data center power monitoring system. So, the primary purpose of this chapter is to provide useful guidance and as much information as it is concerned about modeling and then reducing power usage at the server

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

level—we believe it is a fine granularity considering the massive system of a data center. In the following content, we begin with introducing the common methodologies for server power modeling in the form of taxonomy, including some mathematical stuff by formulating a number of representative power consumption models that are prevailingly used in engineering. Besides, we will look into the problem using some more comprehensive metrics rather than joules and watts by introducing the notion of energy efficiency and discuss the ways to define it. The last part of this chapter is a bit algorithmic as it is all about the topic of energy conservation technology and strategies involving optimization policies from practice and cutting-edge solutions from research, which we believe are of great practical use in establishing a green data center.

19.2 ENERGY CONSUMPTION MODELING OF SERVERS IN DATA CENTERS

19.2.1 Energy Efficiency of Servers in Data Centers

There have been broad concerns about the rapid growth of data center energy consumption as well as their low energy efficiency. The electricity consumption of the ICT sector (within which data centers are the most important operation in industrialized countries) accounts for 5–10% of the total. Improving energy efficiency of data center is an important subject on the way to achieving Green IT, which typically means less greenhouse gas emissions, less harmful material, and encouraging the use of renewable energy.

Using the most intuitive definition, energy efficiency can be measured as the ratio of the amount of work done to the amount of energy consumption over a period of time. The metric is crucial in case you want to look into how efficient and productive your servers are and before you decide to optimize them. Fortunately, there are quite a few energy efficiency metrics available including the commonly used power usage effectiveness (PUE). Nevertheless, some of them are either too simple to reflect the real situations of a server or too complicated to be applied in practical use. So, we believe it is very necessary to single out some metrics concerning server energy efficiency that could be useful in a typical data center.

In practice, we often use floating-point operations per second (FLOPS)/Watt to measure a server's efficiency in energy by putting its performance over its power consumption. The metric is simple but quite useful in showing dynamics of a server provided that it operates at constantly changing levels of workload.

$$\text{Server_power_efficiency} = \frac{\text{performance}}{\text{power consumption}}$$

In case metering devices are not always available (which is realistic as attaching every server with a meter is prohibitive), both the performance and the power consumption can be modeled as quadratic functions by fitting data composed of server power consumption and performance [1]. Therefore, the formula of server power efficiency metric can be further reformulated as the following:

$$\text{Server_power_efficiency} = \frac{c_0 + c_1 * \text{u} + c_2 * \text{u}^2}{d_0 + d_1 * \text{u} + d_2 * \text{u}^2}$$

$$\text{u} = \text{CPU_utilization}$$

In the formula, the parameters c_0, c_1, c_2, d_0, d_1, and d_2 are obtained by data fitting based on the data collected from history run traces where an adequate number of power and performance samples are recorded.

We can also exploit the utilization of various resources combined with the server energy efficiency to provide a comparatively more accurate prospective into our monitoring system. One of the widely used representatives is the server Energy-Efficient Utilization Indicator (EEUI) [2], which is defined as follows:

$$\text{Server}_{\text{EEUI}} = \text{Server}_{\text{EEUI}}^{\text{CPU}} + \text{Server}_{\text{EEUI}}^{\text{Memory}} + \text{Server}_{\text{EEUI}}^{\text{Network}} + \text{Server}_{\text{EEUI}}^{\text{Disk}}$$

where $\text{Server}_{\text{EEUI}}^{\text{X}}$ denotes the EEUI of component $X \in \{\text{CPU, Mem, Net, Disk}\}$. Typically, we calculate the EEUI of CPU in the following way:

$$\text{Server}_{\text{EEUI}}^{\text{CPU}} = \frac{\text{EE}_{\text{CPU}}}{\text{EE}_{\text{CPU_max}}} \times \frac{P_{\text{CPU}}}{P}$$

where EE_{CPU} and $\text{EE}_{\text{CPU_max}}$ stand for the associated energy efficiency level of CPU mapped from CPU utilization and the maximum energy efficiency that CPU can operate, respectively. P_{CPU} and P denote the power consumed by the CPU and the server-wise total consumption, respectively. For memory, network component, and disk, we use similar formulations to determine their EEUI, and thus we get the server's EEUI as a sum of them. It is worth noting that $\text{Server}_{\text{EEUI}}^{\text{X}}$ is not designed for comparing the energy efficiency between different server hardware or architectures. Instead, it is used to monitor the level of energy efficiency when the servers are operating. In summary, EEUI not only considers the utilization levels at which those components are used but also takes into account their energy efficiency and energy proportionality.

19.2.2 Modeling Methods of Servers' Energy Consumption

With rapid development of hardware, new servers from different vendors have supported accessing their real-time power consumption, but solely using physical measurement is never the most compatible and scalable solution, nor can it

predict the future energy demand of the servers. Therefore, establishing an accurate and generic energy consumption model becomes a cornerstone in realizing a mega-scale, real-time, low-cost monitoring system for the sake of optimizing the energy efficiency of the cloud servers.

Based on the methodologies of modeling, we classify the energy consumption models of cloud servers into two categories, namely Performance-Monitor-Counter-based models and resource-utilization-based models. We also introduce the general procedure of modeling briefly.

19.2.2.1 Performance-Monitor-Counter (PMC)-Based Models

Intel has introduced a set of model-specific registers (MSRs) into their performance monitoring systems of processors since Pentium series. The so-called PMC-based model refers to establishing the energy consumption estimation model by analyzing the relationship between server energy consumption and the performance events provided by hardware.

Constructing a PMC-based model consists of three steps:

1. Keep listening for PMC events from subcomponents of servers (e.g., CPU, memory, disk, and Network Interface Controller);
2. Analyze the relationship between specific performance counters and system energy consumption;
3. Establish the relationship model between PMC events and system energy consumption.

The PMC-based model warrants a high accuracy in general, but it may take lots of effort to study which events should be selected because too many events involved could increase the complexity and the risk of over-fitting as well. A defect of PMC-based models is that the model may be invalid as the hardware architecture changes.

19.2.2.2 Resource-Utilization-Based Models

The principle of resource-utilization-based model is to find the correlation between power/energy and resource utilization, which are coarse-grained but easily accessible at the OS level. The measurement of resource utilization can be easily established using existing OS monitoring tools. The most classical form is linear regression model with CPU utilization as the only variable and two parameters of which C_0 is a constant and C_1 is a factor:

$$P = C_0 + C_1 \times u_{\text{CPU}}$$

This type of model is flexible and reliable under a single type of load. However, it may be poor in fitting the power consumption curve when the types of tasks are various and the workload fluctuates frequently and significantly.

19.2.2.3 General Modeling Process of Server Energy Consumption

Generally, the modeling process can be summarized as the following steps (shown as Fig. 19.1): data sampling, data processing, model building and model evaluation.

- Data Sampling

  Data sampling takes two parts of work simultaneously: power consumption sampling and system performance sampling. As Figure 19.2 shows, built-in sensors (e.g., IPMI and RAPL) or external devices can acquire the server power consumption. Monitor tools, like Perf and OProfile, can acquire the system performance data.
- Data Processing

  Missing value processing, denoising techniques, and data normalization are commonly used to preprocess the raw data and analyze the characteristic and potential relationship between the performance and energy consumption.
- Model Building

  Regression methods (e.g., linear/nonlinear regression) or more complex methods (e.g., SVR and neural network) can be used to model the relationship between the input features and server power consumption. After having the basic form of the model, parameters optimization and error correction should be considered to ensure the accuracy.
- Model Evaluation

  After producing a trained model, we should evaluate its accuracy, overhead and other indicators. The model should be tested in the production environment or use a set of benchmarks to simulate the specific task scenario to evaluate the model. Some metrics are widely used to improve the model, such as MSE (mean squared error) and MAPE (mean absolute percentage error).

19.2.3 Power Models of Servers in Data Centers

Mathematically, a power model can be defined as a function that maps the variables related to the system state to the system power consumption, which takes one or more system indicators (e.g., CPU, memory, network adapter and disk

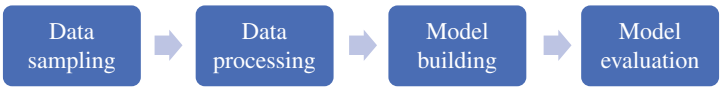


FIGURE 19.1 The workflow of cloud server energy consumption modeling.

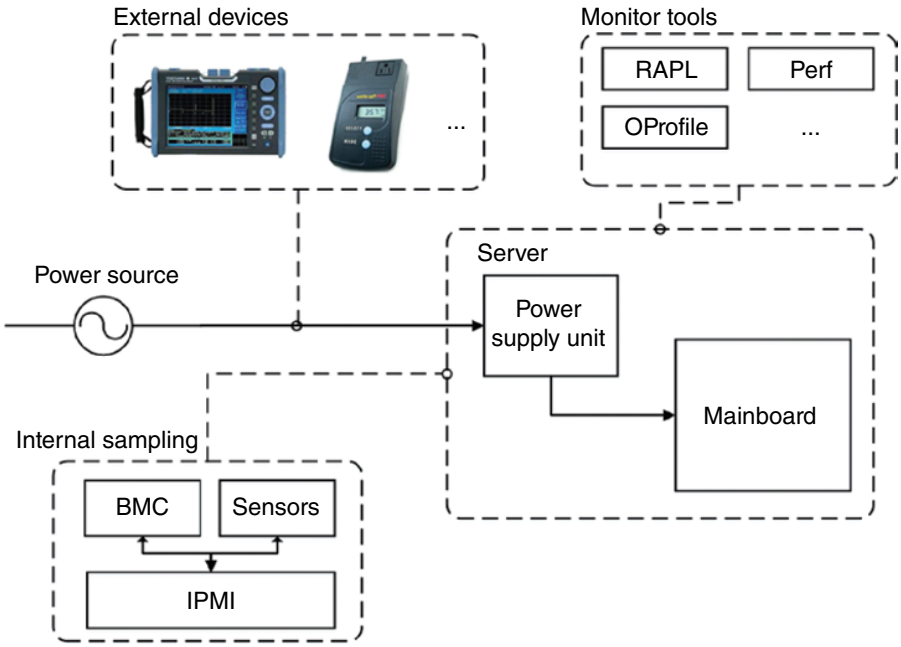


FIGURE 19.2 The data sampling framework of server energy consumption modeling.

utilization) as the function's independent variables, and takes the instantaneous power or the cumulative energy within a period of time as the function output. According to the types of server instances, power models of servers can be roughly divided into three categories: power models of the physical server, virtual machine (VM) power models, and container power models (as shown in Fig. 19.3).

No matter what process management or virtualization technology is used, power consumption is ultimately reflected in the fluctuation of workload. At the physical server level, we summarize existing power models into two categories: coarse-grained power models and fine-grained power models. For coarse-grained models, the essence is to screen the underlying complexity and hierarchical structure of physical server and to model the power consumption at the highest level. Namely, it mainly focuses on the power-consuming entities that can perform work independently (such as running operations and cooling). For example, Fan

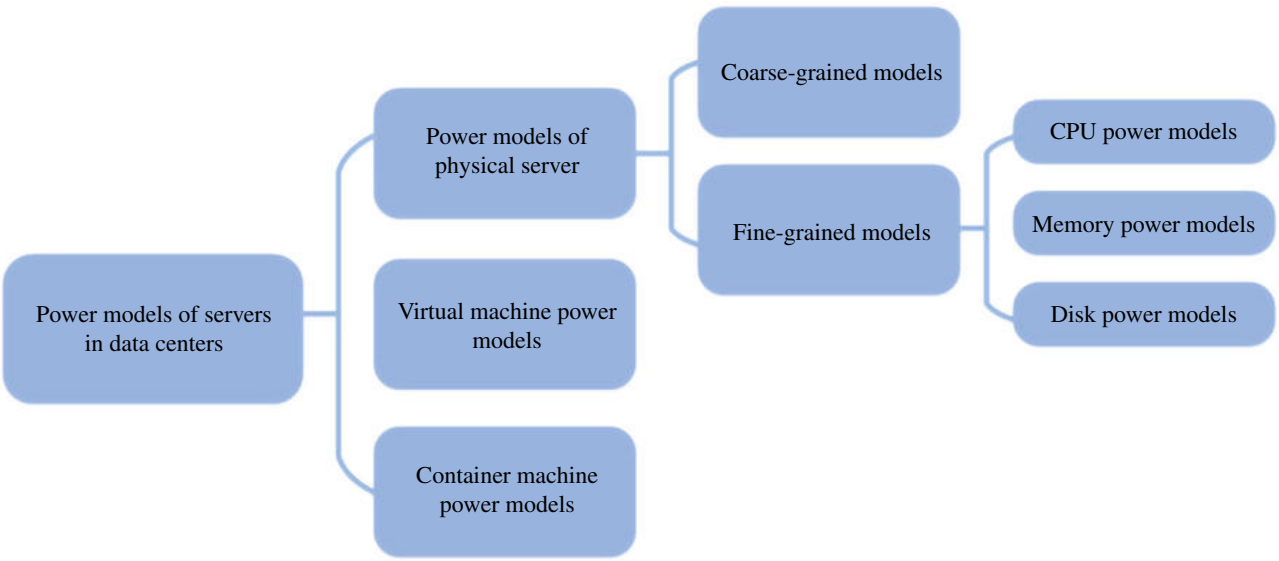


FIGURE 19.3 The power models' specific classification of servers in data center.

et al. [3] proposed a power model for estimating the whole physical server power based on CPU utilization:

$$P_{\text{server}}(t) = c_0 + c_1 u(t)$$

where both c_0 and c_1 are model parameters, $u(t)$ represents the CPU utilization. For fine-grained models, we need to consider the energy-consuming entities that are not working independently and model each of them separately. Using linear forms again, we can formulate the power of major components such as CPU [4], memory [5], and disk [6] as below:

$$P_{\text{cpu}} = P_{\text{cpu}_{\text{idle}}} + \left(P_{\text{cpu}_{\max}} - P_{\text{cpu}_{\text{idle}}}\right) u_{\text{cpu}}$$

$$P_{\text{mem}} = P_{\text{mem}_{\text{idle}}} + C \cdot u_{\text{mem}}$$

$$P_{\text{disk}} = P_{\text{disk_idle}} + C_r m_{\text{read}} + C_w m_{\text{write}}$$

For VM power models, since the management of virtual resources and the physical resources are separate from each other in the virtualization layer, we cannot directly apply the power modeling methods for the traditional hardware layer to VMs. In current researches on VM energy consumption, there are two main methods applicable to modeling VM's power, namely white box methods [7] and black box methods [6]. The main difference is that the former is to embed the monitoring agent into VM and obtain information from the internal, while the latter is to obtain information from the host for modeling and monitoring.

For container power models, since container encapsulates processes and packages and all the resources required for the software to run in an isolated container, and the application process runs directly sharing the host's kernel, it is less intuitive to model the containers' power. Currently, there are only a few studies working on this topic. As a possible solution, some existing work propose to model containers' power at process-level and include machine learning techniques. For example, David et al. [8] introduced a process-level power model, in which the container is treated as a process on its host (a VM or a server). Kang et al. [9] proposed a container power model based on k-medoid clustering, which makes use of the characteristics of both server and container as features.

19.3 ENERGY-SAVING TECHNOLOGIES OF SERVERS

A wide range of technologies and techniques are ready for use if you believe the servers are not running in the optimal style and decided to make some changes. The primary purpose of this part is to provide information and some guidance in terms of how to utilize existing technologies for improving server-wise energy efficiency. Some of the presented ones are cutting-edge from the latest researches and some have been widely adopted in the industry like a rule of thumb or best practice. Anyway, these methods, schemes, and techniques to be introduced have more or less proved to be effective or at least helpful when you are looking for solutions that can make your servers run in a more power-efficient manner.

It is a complex problem to characterize what a server is doing with every joule of energy given to it. The power models detailed in the previous sections could be helpful for understanding each factor that contributes to the total consumption of a server when we break it down into components. From that some may think that energy could be saved as long as we remove some of the components, and the industry knows it well—"You had the opportunity to strip things down to just what you need, and make it specific to your application," says Bill Carter, chief technical officer at the Open Compute Project. However, people do not truly lose weight by not wearing clothes, so do the servers. The most effective way to make servers less hungry to energy is to manage them wisely—which is exactly the topic of this section. With the advances of both research and practice, existing energy-saving techniques for servers have covered a broad range of applications including hardware-based, software-based, and more. Actually, some of the giant companies, like Facebook, have deployed them comprehensively in their data centers around the globe. To make it clear and organized, these techniques will be introduced as they are summarized in the following categories:

- Dynamic Server Management: includes a handful of both relatively conventional power-saving techniques that may need certain hardware function and state-of-the-art algorithms that make use of artificial intelligence (AI) to manage servers in an automatic fashion.
- Task Scheduling: about software-based high-level optimization of server workload by rearranging a group of tasks in given workflows for achieving shorter makespan and less energy consumption.
- VM Allocation and Consolidation: implementation of VM management algorithms that produce the best match between physical servers and VMs and energy-optimal VM migration plans.
- Light-weight Virtualization: a bunch of emerging implementations of resource encapsulation and isolation as lower-cost virtualization technology.
- Load Scheduling with Renewable Energy Provisioning: incorporates scheduling algorithms and frameworks that leverage the dynamics of renewable energy sources to reduce the expensive grid power usage.

19.3.1 Dynamic Server Management

In 2017, Jonathan Koomey, a California-based consultant and leading international expert on IT, surveyed with a

colleague more than 16,000 commercial servers and found that about one-quarter of them were "zombies," gobbling up power without being useful. As a matter of fact, zombie servers may not be rare in large-scale data centers but by many means these power eaters can be avoided.

19.3.1.1 Dynamic Scale-Out/Scale-In

Managing servers is kind of like managing your employees—a popular management technique is to make sure that every one of them runs at full throttle as much of the time as possible, whereas others are turned off/sacked rather than being left idle. Facebook invented a system called *Autoscale* [10] that, as they claim, can reduce the number of servers that need to be on during low-traffic hours, and this led to power savings of about 10–15% in trials.

They implemented a specifically designed load-balancing framework that aggregates the workload on a subset of their entire cluster by rerouting the coming requests. By doing so, there is a big chance that a portion of their servers in the data center will have nothing to do if it is not in peak hours. Once a server turns idle for a certain length of period, it will be shut down or switch to some energy-saving mode—in either way it goes inactive and consumes much less energy. The strategy usually works as the coming workload generally follows some time series patterns that have a lot to do with the habits of users. By setting up rules that prioritizes some of your servers, it is not difficult to get only a group of them to work rather than spreading the load across your cluster stochastically. Nevertheless, the main challenges are how to accurately anticipate workload (e.g., the number of coming requests) so that you don't activate and deactivate servers frequently and how to achieve real-time elasticity by making use of your anticipations.

19.3.1.2 Dynamic Work Mode Switching

Frequently powering on/off servers may cause prohibitive overheads in both energy cost (we often see power peaks in the start-up process) and server lifespan. In terms of energy savings, Facebook manages to scale in their cluster by deactivating servers in a much softer manner - putting inactive servers into power-saving mode. Averagely, with the technique introduced, they are able to achieve >10% power saving over a 24-hour cycle for different web clusters.

It is worth mentioning that mode switching and power-saving tweaks have been well supported by multiple mainstream operating systems including Windows server (supporting six levels of power-saving states from S0 to S5) and many Linux distributions (e.g., the *pm-utils* and component-wise power-saving tweaks for Ubuntu and the *CPUfreq governor* for Red Hat).

19.3.1.3 DVFS and Alternatives

Server power usage has a strong correlation with CPU utilization, making dynamic tuning of CPU state a vital part of server power management. Dynamic Voltage Frequency Scaling (DVFS) is one of the most notable techniques for reducing CPU power consumption and it works based on the following CMOS power principles:

$$P = \alpha C V^2 F$$

where αC is a constant for a specific processor, V and F denote the supply voltage and work frequency, respectively. P is the instant power consumption of CPU, which we want to reduce in case there is not much workload on the server. From the equation, it is pretty clear that DVFS techniques (along with associated techniques such as dynamic voltage scaling (DVS) and adaptive voltage and frequency scaling (AVFS)) are very effective in energy conservation, since lowering the voltage (V) has a squared effect on active power consumption while performance degradation (dependent on F) is basically linear. Specifically, some experiments indicate that with DVFS it is possible to achieve a 3× reduction in power with only 1× reduction in performance.

As an alternative, AVFS is an extension of DVFS. DVFS is usually limited to scaling in a series of fixed discrete steps in terms of the voltage and frequency of the targeted power domains, making it an open-loop system with large margins built in, and therefore the power reduction is not optimal. On the other hand, AVFS deploys closed-loop voltage scaling and is compensated for variations in temperature, process, and IR drop via dedicated circuitry that constantly monitors performance and provides active feedback. Although the control is more complex, the payoff in terms of power reduction is higher.

An obvious side-effect of scaling down CPU voltage/frequency is that it takes you a longer time to finish a task compared to that needed in the max-performance mode, and that in turn lessens the benefit you gain from the reduction of power (see the following equation).

$$E = P \cdot t = \alpha C V^2 F \cdot t$$

Generally speaking, voltage and frequency scaling techniques are very useful when applied together with server mode switching to minimize energy wastage in underutilized or idle states, but its benefits could be marginal in case of continuous peak loads.

19.3.1.4 Proactive Load Control

As effective as DVFS and AVFS are, they are designed to rein in the power of a server in a totally reactive manner. However, when your servers are overwhelmed by workload (e.g., bursts of incoming requests on Black Friday), hardware-based scaling may only make little difference and even lead to system instability.

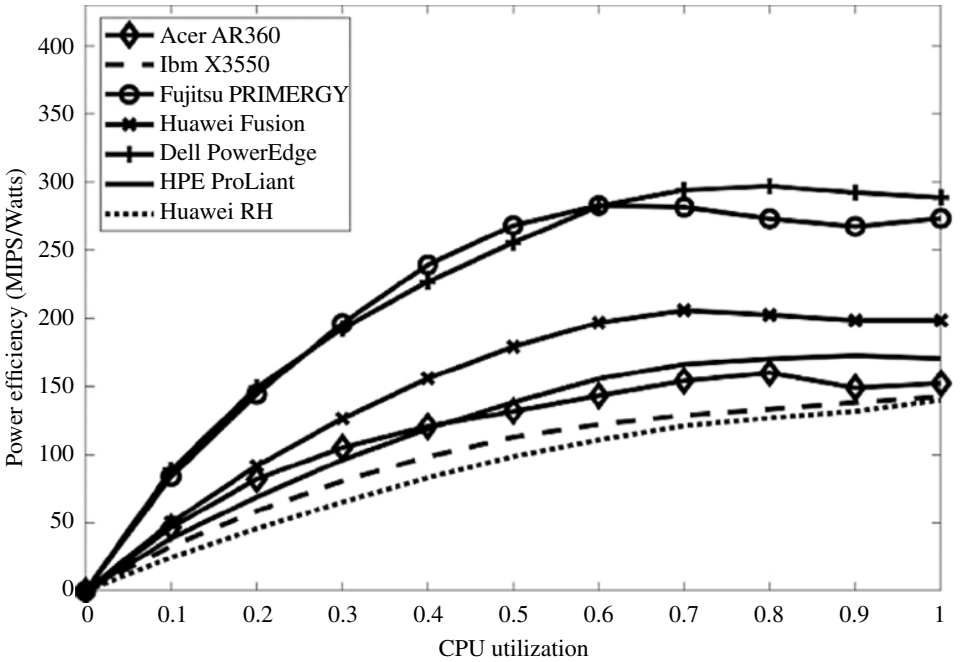


FIGURE 19.4 Power efficiency of seven different servers computed according to the data provided by SPEC [11]. Source: Lin et al. [12]/ with permission of IEEE.

It is always important to keep the load in check no matter from the perspective of service quality or the standpoint of energy conservation. For one thing, the risk that your server gets burned out certainly increases when it is working under constant high load (e.g., CPU usage >95%). For another, research results (see Fig. 19.4) have shown that keeping servers running at their maximum utilization is often not energy efficient—a large number of them attain the optimal performance–power ratio at a load level around 80%.

Proactive load control is very necessary so that we can have enough time and resources to make use of the power characteristics of physical servers. In fact, many studies (e.g., [13, 14]) have suggested managing the load of servers proactively and preventively. This, in many cases, requires the system to be always aware of the workload level on each server, and to redirect or deny some requests once it shows signs of being overloaded. Proactive load control helps to make sure the operators have time to scale up and prevent the jittering workload from affecting server power efficiency.

19.3.1.5 Comprehensive Optimization with AI

In the past few years, empirical analysis and expertise are literally the rules of thumb when it comes to find the optimal configuration and operational policies for a system—server management in data centers was of no exception. But that seems to change recently as the rapid advance of AI has drawn world-wide attention, and AI-driven solutions did come out with a lot of success in a variety of domains that people reckon as collections of complex problems for humans.

As it has been mentioned, dynamic power management of servers is intricate because no one tells you how to find out the optimal configuration of DVFS or what turns out to be the best policy for load management. So the question is: can AI be the leading light of energy optimization of servers? Most people would say yes and maybe more when they see the promising results from some state-of-the-art studies. For example, the authors of paper *Automated cloud provisioning on AWS using deep reinforcement learning* [15] innovatively propose to adopt Deep Reinforcement Learning (RL) to realize automatic cluster scale-out/in. They built a smart cluster controller based on Q-learning—a popular RL method—by modeling *Q-state* as the number of server instances, *action* as the decision of scaling-out or scaling-in, and *reward* as the resulting change of energy consumption (Q-learning, in essence, learns how to take good actions in a given Q-state considering the resulting state and reward). Figure 19.5 shows the typical architecture of using (deep) Q-learning network for automated resource provisioning on cloud infrastructures where the "Environment" represents a circumstance that continuously provides feedbacks ("state change" and "reward") to the model.

After certain rounds of training, the RL-driven controller shows its capability in finding energy-efficient decisions and it is further claimed that it is also able to learn policies that balance performance and energy cost—as long as humans specify what they want. Something even more amazing is

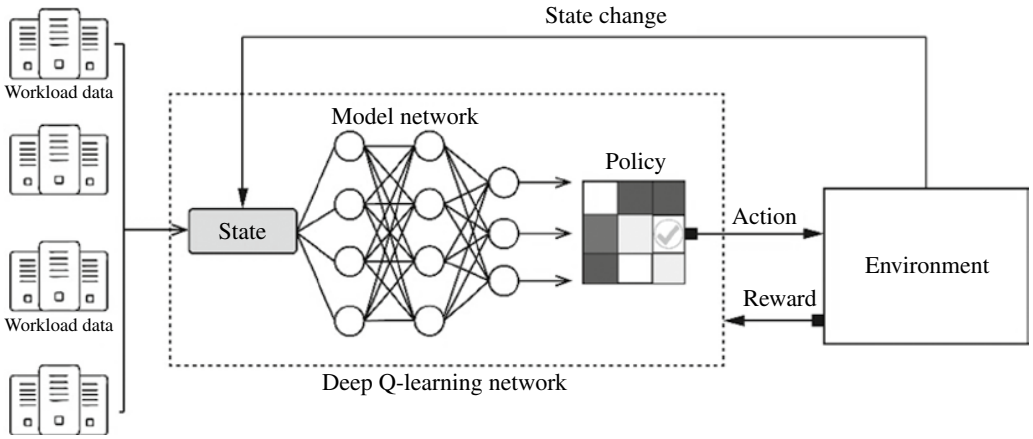


FIGURE 19.5 A schematic diagram showing a general architecture of (deep) Q-learning network where the specific structure of the underlying model network can be fully connected, convolutional, or a combination of both.

that the learned policies are refined over time because RL enables the controller to constantly improve as it works. This is only a small case where one can apply the spell of AI to server power management in modern data centers. Actually, AI and machine learning can do more, and these emerging techniques are promising in taking the place of human experience (at lease, partially) in terms of comprehensive energy usage optimization.

19.3.2 Task Scheduling

Task scheduling is one of the most fundamental problems when it comes to optimizing resource provisioning on servers given a flow of tasks (which in some cases are considered as decomposed user jobs). In general, task schedulers and associated job coordinators are implemented at the software level. Depending on the optimization target, they can have major impact on the servers' productivity and, of course, energy consumption.

Given a batch of tasks, a task scheduler basically needs to decide which task should be assigned to which server and what order should the co-allocated tasks be executed. If you are familiar with the classical bin-packing problem, you can easily find the similarity as well as the complexity in solving it—it is a combinatorial NP-hard problem. Because of this, heuristic approaches (e.g., greedy algorithms and evolutionary algorithms [16]) are commonly used for the problem of task scheduling. For example, *Min-Min* [17] is one of the most famous solutions and it proved its effectiveness in shortening average task makespan (which thereby helps reduce energy usage). However, in cloud data centers, which are usually hyper-scale and virtualized, tasks are most often bound to virtual machines (VMs) to realize resource isolation. The additional role of VMs increases the complexity of task scheduling but many studies have already come up with promising solutions. Lin and Wu [18] introduced an energy-aware task-to-VM scheduling framework (Fig. 19.6) that generates energy-efficient allocation plans by comprehensively considering task demands,

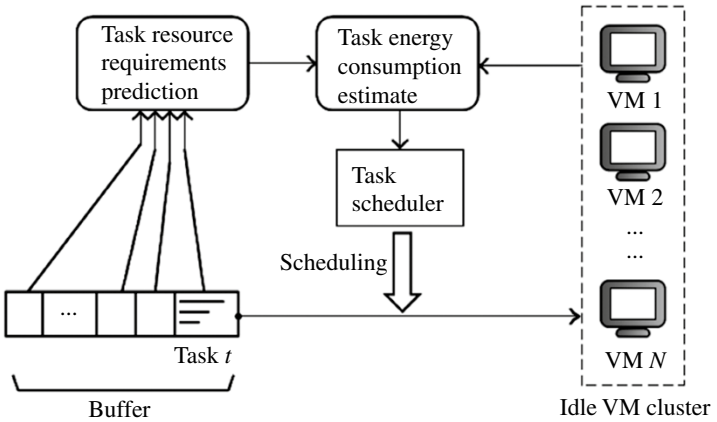


FIGURE 19.6 An overview of an energy-aware scheduling framework that caches arrived tasks in a buffer, estimate task energy consumption, and invoke scheduling functions to allocate tasks to VM instances running on active servers. Source: Lin et al. [18] / with permission of IEEE.

VM's power efficiency as well as server workload. Their algorithm achieves a reduction of over 20% in energy consumption across the cluster.

19.3.3 VM Allocation and Consolidation

Virtualization has been prevailingly adopted in data centers since the last decade, and there are a lot of research and practice focusing on how to maximize the benefits we gain from virtualization through VM orchestration. It is not a one-shot operation as we need to continuously handle newly created instances as well as readjust the location of running VMs for the purposes of both load balancing and energy conservation.

VM allocation refers to a range of strategies that optimize the mapping from VMs to bare-metals and VM consolidation mainly deals with the reallocation of VMs by means of migration. In order to increase server utilization and reduce energy consumption, cloud infrastructure providers pay very much attention to dynamic VM allocation/reallocation since emerging techniques have made live migration operations faster than ever. Compared to allocation, reallocation of VMs requires the algorithms to be a bit more sophisticated—it is a multistep operation that starts from detecting overloaded/underutilized hosts (i.e., physical servers) first, then picking one or more VMs from these hosts for migration, and finally finding a group of candidate servers that are competent as target hosts. Each step could be intricate as a lot of energy-related information and consideration are needed to make the overall process worthwhile with respect to energy reduction. Figure 19.7 displays a schematic framework showing the basic workflow and cooperation between modules that support energy-saving VM allocation and consolidation.

As complex as the framework is, it is always important to keep in mind that the overheads of VM allocation/reallocation must be strictly in check. This may unfortunately restrain you from applying some novel but high time-complexity algorithms into your VM management strategy. Empirically speaking, there is a trade-off between the optimality (i.e., how much energy you can save) and the efficiency (i.e., how long it takes to make a decision) of your strategy, and this should be well considered at the very beginning.

19.3.4 Light-Weight Virtualization

Engineers are always looking for something more efficient and cost-saving and they began to lose their interests in traditional virtualization technology as VMs are so "clumsy" if you look at how many resources are needed to run a hypervisor on your server. Since VMs are too heavy, light-weight virtualization technology earned popularity quickly and the most prevailing branch today is the container (or containerization).

Having its name from the shipping industry, container technology refers to a method for packaging an application so it can be run, with its dependencies, isolated from other processes. The mainstream public cloud computing providers, including Amazon Web Services (AWS), Microsoft

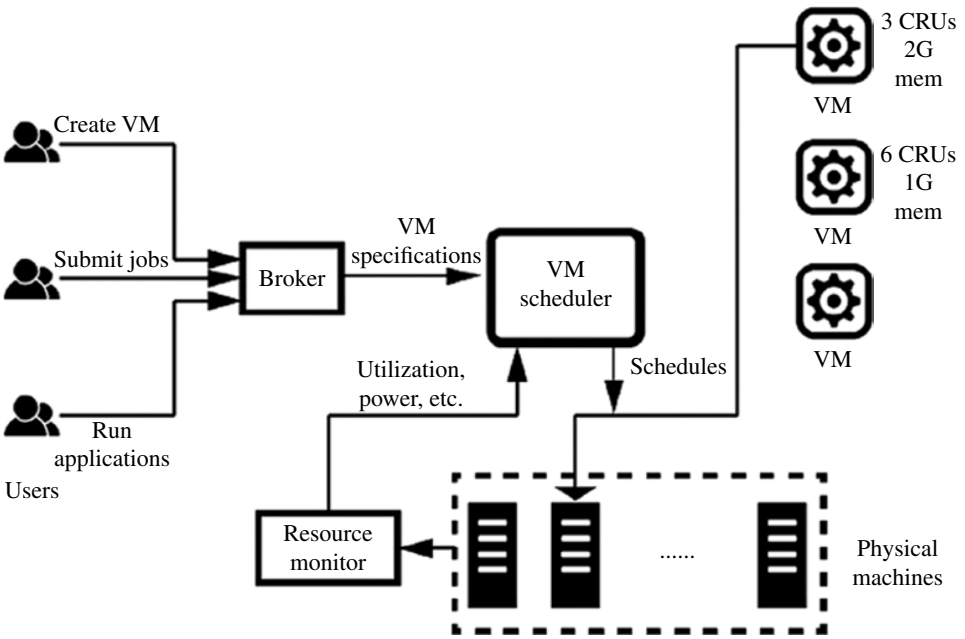


FIGURE 19.7 A virtual machine (re)allocation framework for clouds providing flexible services adopted in the study on energy-efficient cluster management. Source: Lin et al. [12] / with permission of IEEE.

Azure, and Google Cloud Platform, have embraced container technology and widely deployed it in their hyperscale data centers. The main strength of containers, compared to VMs, is that they share the OS kernel and do not require the overhead of associating an operating system within each application. Because of this, containers are far smaller in capacity than a VM and require less start-up time, allowing a lot more instances running on the same server. This drives higher server efficiencies and, in turn, reduces energy cost for running a same number of applications. To date, there are more than a handful of options if one decides to put containerization into practice, here are some of the most popular implementations: *Linux LXC*, *Docker, KataContainer* (by OpenStack), *Shifter* (by IBM), and *FireCracker* (by Amazon).

Another trending technology that shows it promise in further amplifying the benefits of light-weight virtualization is called Unikernel. Also known as container 2.0, the design principle behind Unikernel is a further step into minimalism—provide just enough software to power the desired application and nothing more. Technically speaking, Unikernels rely on specialized compilers to combine application software and supporting OS functions at compile time instead of runtime. This results in a single application image that contains everything the application needs to run. All drivers, all I/O routines, and all supporting library functions normally provided by an operating system are included in the executable, while those unneeded ones will be excluded in the image to keep it as light as possible. For example, *MirageOS* (an established Unikernel project) claimed that they have a working domain name server that compiles into just 449 kB. The project also has a web server that weighs at 674 kB and an *OpenFlow* learning switch that tips the scales at just 393 kB.

The development of virtualization is heading to the direction where resource-efficiency is put on the first place. This is a good news from the perspective of energy conservation—servers theoretically get more power efficient when we keep reducing the overhead of virtualization. But extra benefits always come at a cost, and the cost here is weakened isolation between applications, which may raise some issues concerning both security and resource contention.

19.3.5 Load Scheduling with Renewable Energy Provisioning

In 2011, Facebook made a commitment to using 100% renewable energy. Google (the largest corporate purchaser of renewable energy on the planet so far) and Apple followed in 2012. As of 2017, nearly 20 Internet companies had done the same. Electricity from renewable energy sources was not so advocated by the ICT industry at the beginning of this decade—back in 2010, IT companies were still a negligible contributor to renewable-power purchase agreements with energy suppliers; but by 2015, they accounted for more than half of such agreements [19] (see Fig. 19.8).

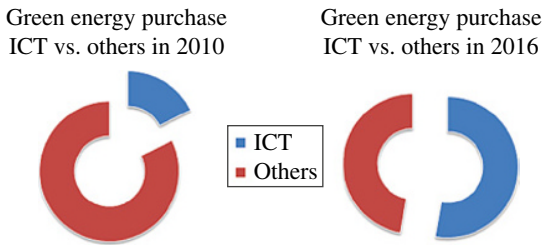


FIGURE 19.8 The contribution to renewable electricity contracts by the ICT industry ramped up between 2010 and 2016, according to the report by International Energy Agency (IEA) [20].

The reason why ICT giants show tremendous interest in introducing renewable power provisioning to their data centers are multifold. First, the government keeps forcing enterprises to reduce their carbon footprint, and apparently replacing brown energy sources with green ones is one of the best options. Besides, the prices of renewable electricity are expected to decrease as people believe it is very hopeful to embrace new technologies that make renewable power generation (e.g., wind and solar power) more efficient than ever.

Data center owners can benefit more from what renewable energy sources already brought to us. Related studies have shown the great potential in further reducing server energy cost through the refinement on workload management by leveraging the characteristics of green power supply. For example, researchers integrated a renewable energy supply prediction model into task scheduling algorithms to rearrange the execution order of tasks so as to maximize the utilization of renewable electricity [21] (see Fig. 19.9 where the "green energy" is used whenever available, while electricity supply from traditional power grid accounts for the rest), and a novel technique called Battery Assisted Green Shifting is introduced to increase the flexibility in the way renewable energy is used for the execution of jobs on a server.

These cutting-edge studies provide very useful insights into what we can do in the operation of data centers when we have the chance to power our servers with renewable energy. We can foresee that there will be more opportunities as well as challenges in the (near) future. With the evolutionary force of green power, maybe we can raise the energy efficiency of servers and data centers to a level higher than ever, or maybe data centers can totally get rid of traditional power grids, or at least cloud services could be repriced based on how much green energy is used. Only time will tell.

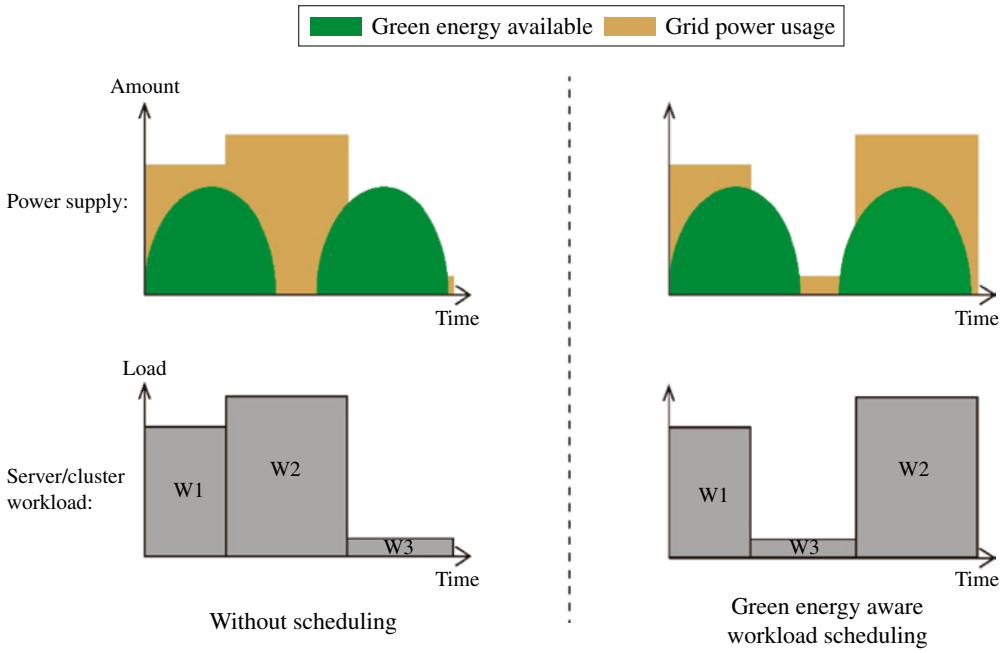


FIGURE 19.9 A schematic showing how to reschedule cloud workload based on the dynamic of power supply to maximize the utilization of green/renewable energy.

19.4 CONCLUSIONS

Considering the significance of energy consumption, this chapter conducted a systematic, in-depth study about energy consumption and energy-saving technologies of servers in data centers. To better understand, we first introduced the modeling energy consumption of servers by presenting the modeling methods and general modeling process of server energy consumption. Next, the power models of servers in data centers are introduced with a hierarchy, focusing on the power models of the physical server, VM, and container. And then, the energy efficiency of servers in data centers was presented to evaluate the power models of servers, which mainly includes the energy efficiency metrics and the examples of server energy efficiency. Moreover, to better provide information and some guidance on how to utilize existing technologies for improving server-wise energy efficiency, the energy-saving technologies of servers including the dynamic server management, task scheduling, VM allocation and consolidation, light-weight virtualization, and load scheduling with renewable energy provisioning are also presented in details.

Simultaneously, we observed that there have been a large number of studies conducted on the energy consumption and energy-saving technologies at the lower levels of the data center, but much less work has been done at the higher levels. And this is the most prominent problem in the current research on energy consumption modeling and energy-saving technology for data centers. Therefore, we need to further explore higher-level issues (such as orchestration of containers) in future research.

ACKNOWLEDGMENTS

This chapter is partially supported by the National Natural Science Foundation of China (Grant Nos. 62072187, 61872084) and Major Program and of Guangdong Basic and Applied Research (2019B030302002). We would like to thank Fang Shi (SCUT, China), Gangxing Wu (SCUT, China), Tianhao Yu (SCUT, China), Gaofeng Peng (SCUT, China), Chennian Xiong (SCUT, China), and Hongping Zhan (SCUT, China) for their editing and data collection of this chapter.

REFERENCES

[1] Lin W, et al. A heuristic task scheduling algorithm based on server power efficiency model in cloud environments. Sustain Comput Inf Syst 2018;20:56–65.

[2] Abaunza F, Hameri AP, Niemi T. EEUI: a new measure to monitor and manage energy efficiency in data centers. Int J Product Perform Manag 2018;67(1):111–127.

[3] Fan X, Weber WD, Barroso LA. Power provisioning for a warehouse-sized computer. ACM SIGARCH Comput Archit News 2007;35(2). ACM.

[4] Hsu CH, Poole SW. Power signature analysis of the SPECpower_ssj2008 benchmark. (IEEE ISPASS) IEEE International Symposium on Performance Analysis of Systems and Software; April 10–12; Austin, TX, USA 2011. IEEE.

[5] Lin W, et al. A cloud server energy consumption measurement system for heterogeneous cloud environments. Inf Sci 2018;468:47–62.

[6] Kansal A, et al. Virtual machine power metering and provisioning. Proceedings of the 1st ACM symposium on Cloud computing; June 10–11; Indianapolis, Indiana, USA 2010. ACM.

[7] Li Y, et al. An online power metering model for cloud environment. 2012 IEEE 11th International Symposium on Network Computing and Applications; August 23–25; Cambridge, MA, USA 2012. IEEE.

[8] Snowdon DC, et al. A platform for os-level power management. The European Professional Society on Computer Systems 2009; April 1–3; Nuremberg, Germany 2009.

[9] Kang DK, et al. Workload-aware resource management for energy efficient heterogeneous docker containers. 2016 IEEE Region 10 Conference (TENCON); November 22–25;Singapore, Singapore 2016. IEEE.

[10] Facebook Engineering. Autoscale. Available at https://engineering.fb.com/production-engineering/making-facebook-s-software-infrastructure-more-energy-efficient-with-autoscale/.

[11] SPEC. Available at http://www.spec.org/power_ssj2008/results/.

[12] Lin W, Wu W, He L, Li K. An on-line virtual machine consolidation strategy for dual improvement in performance and energy conservation of server clusters in cloud data centers. IEEE Trans Serv Comput 2019;2019, Doi: https://doi.org/10.1109/TSC.2019.2961082.

[13] Adams E. Optimizing preventive service of the software products. IBM J Res Dev 1984;28(1):2–14.

[14] Castelli V, Harper RE, Heidelberger P, Hunter SW, Trivedi KS, Vaidyanathan KV, Zeggert WP. Proactive management of software aging. IBM J Res Dev 2001;45(2):311–332.

[15] Wang Z, Gwon C, Oates T, Iezzi A. 2017. Automated cloud provisioning on AWS using deep reinforcement learning. arXiv preprint arXiv:1709.04305.

[16] Pacini E, Mateos C, Garino CG. Balancing throughput and response time in online scientific clouds via ant colony optimization (sp2013/2013/00006). Adv Eng Softw 2015;84:31–47.

[17] Maheswaran M, Ali S, Siegel HJ, Hensgen D, Freund RF. Dynamic mapping of a class of independent tasks onto heterogeneous computing systems. J Parallel Distrib Comput 1999;59(2):107–131.

[18] Lin W, Wu W, Wang JZ. A heuristic task scheduling algorithm for heterogeneous virtual clusters. Sci Program 2016;2016.

[19] Goiri Í, Beauchea R, Le K, Nguyen TD, Haque ME, Guitart J, Torres J, Bianchini R. Greenslot: scheduling energy consumption in green datacenters. SC'11: Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis; November 12–18; Seattle, Washington, USA 2011, pp. 1–11. IEEE.

[20] International Energy Agency (IEA). Digitalization and Energy; 2017.

[21] Niu Z, He B, Liu F. Not all joules are equal: towards energy-efficient and green-aware data processing frameworks. 2016 IEEE International Conference on Cloud Engineering (IC2E); April 4–8; Berlin, Germany 2016. pp. 2–11. IEEE.

20

CYBERSECURITY AND DATA CENTERS

Robert Hunter[1] and Joseph Weiss[2]

[1] *AlphaGuardian, San Ramon, California, United States of America*
[2] *Applied Control Solutions, Cupertino, California, United States of America*

20.1 INTRODUCTION

Data centers occupy an increasingly valuable place in the lives of everyone. This is true because every request for data from a search engine, every piece of music played, and most all of the digital life that each person leads come to us via data centers.

The constant need to access digital information creates the requirement for data centers with the highest uptime possible (Fig. 20.1). Because of this, data centers are designed to have their Information Technology (IT) systems and their supporting Operational Technology (OT) systems operate at 99.999% or higher availability. This means that IT networks and OT electrical, mechanical, and security systems are designed with redundancy to ensure failover options are available in the event of the disruption of a major system. Large sums are spent to ensure redundant electrical and mechanical system paths are available in a data center and, yet, a single click of the mouse from a hacker can bring down an entire data center for hours.

In this light, network engineers have been particularly sensitive to create a cybersecure environment for the IT systems within the data center. Cyberattacks continue to rise and the need to detect and defend against such attacks is a clear and present need. Yet, while large sums of money are spent to protect the cybersecurity of IT systems, a loophole exists with the lack of cybersecurity within the OT networks of data centers. This creates the prospect that even the best electrical and mechanical designs can be compromised by even an inexperienced hacker, let alone a nation-state bent on a destructive attack.

Because there are vast sums of written knowledge on the subject of IT cybersecurity, it is not the purpose of this chapter to try to add further to that body. The purpose of this chapter is to discuss the specifics of the much lesser-understood and vitally important topic of OT cybersecurity. In this chapter, we shall look at the following topics concerning OT cybersecurity:

- Background of OT connectivity in data centers
- Vulnerabilities and threats to OT systems
- Legislation covering OT system security
- Cyber incidents involving data center OT systems
- Cyberattacks targeting data center OT systems
- Conclusions
- References
- Further Reading

20.2 BACKGROUND OF OT CONNECTIVITY IN DATA CENTERS

OT forms the base infrastructure in which all data is created, stored, and transported. All data is created from the electricity fed by critical power systems such as Uninterruptible Power Supplies (UPS's) and Power Distribution Units (PDU's). Data is no more than a series of electronic 0's and 1's created from the critical power systems. Once created, data is no different from any other perishable item. It must be stored and transported in a proper environment to ensure that the 0's and 1's are not destroyed

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

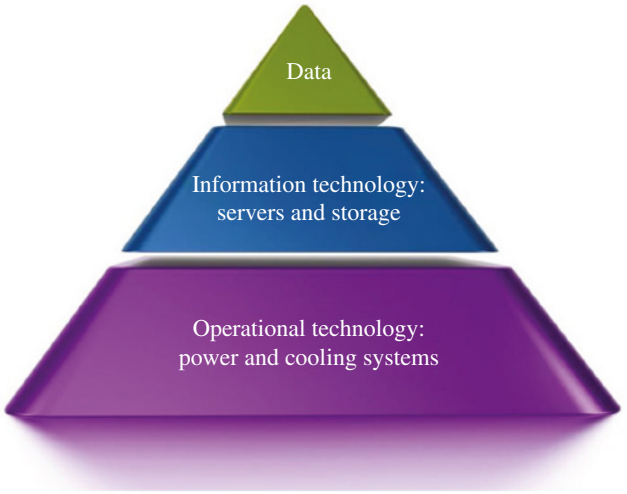


FIGURE 20.1 The foundation of data center operational technology. Source: Courtesy of AlphaGuardian.

due to a thermal breakdown of one of the systems. Finally, data have enormous value and they must be protected, both physically and electronically, in order to keep thieves from accessing, stealing, or destroying any of that data.

A data center can be seen as a hierarchical structure with OT systems at the base to create and sustain IT systems and their data. Any interruption in the OT systems will affect both the IT systems and their data. Any interruption in the IT systems will affect the data. As IT systems standardized on Ethernet network in the 1990s, manufacturers of both IT and OT systems began to work on standardized protocols in order to continuously monitor these systems.

The goal of every manufacturer of data center equipment was to detect faults in their systems as early as possible to avoid disruption of data center operations and the possible destruction of data. Three protocols emerged from this work:

1. Simple Network Management Protocol (SNMP)
2. Modbus/TCP (Transmission Control Protocol)
3. BACnet

These three protocols have become widely distributed throughout the equipment in a data center. SNMP is present in most every IT system and in most all power and cooling systems. Modbus is present in many power and most all cooling systems. BACnet, owing to its creator the American Society of Heating Refrigeration and Air-Conditioning Engineers (ASHRAE), is present in most all cooling systems.

In their most basic iterations, these protocols provide the ability to monitor system health in IT and OT systems through central management servers. Network Management Systems (NMS) provide the bulk of IT management systems and are universally SNMP-based. Building Management Systems (BMS) provide the bulk of OT management systems that are able to support SNMP, Modbus, and BACnet. In addition, it is common for other systems to be in place within data center management structures including Data Center Infrastructure Management (DCIM) systems. These systems query devices within a data center and are used by managers to ensure uptime of operations and the smooth processing of data.

NMS, BMS, and DCIM systems have become universal in data center operations. In fact, they are so pervasive within data centers that there has been considerable debate in the past few years as to whether the volume of SNMP, Modbus, and BACnet traffic is actually causing problems within data centers. Suffice it to say, these systems send and receive massive amounts of data within a data center in order to keep a highly granular view of the present conditions within the data center.

These three common protocols used for data center management were created in an era when data centers and, hence, the amount of OT and IT infrastructure to manage, were much smaller. Because of this, little attention was paid to the network overhead of these protocols at the time of their beginnings. In addition, cyber/physical security was not seen as a great challenge in a time of smaller data centers where physical eyes could see most everything that was taking place and cybersecurity had not become a global threat. In the next section of this chapter, we shall examine each of these three major protocols as to their cyber vulnerabilities and we will also take a look at other security threats that are present within major OT systems.

20.3 VULNERABILITIES AND THREATS TO OT SYSTEMS

The dominant communications protocols used for IT and OT systems were all developed with a great deal of input from vendors and customers. In the case of SNMP and BACnet, formal meetings were held among stakeholders over a period of many years to develop the protocols that are now in use. This section explores the development and implementation of these protocols and examines how cyber vulnerabilities came to exist in each.

20.3.1 SNMP

SNMP is the dominant protocol for management throughout the data center world. Most every piece of IT and OT equipment will directly support SNMP in some form or fashion. This protocol was developed by a private company, SNMP Research of Knoxville, TN. SNMP Version 1 was first published by the International Engineering Task Force (IETF) as a under IETF 1067 in August 1988. Version 1 did not include encryption of data as there was little thought of the need for cybersecurity of SNMP transactions at that time.

A second version, Version 2, was published in April of 1993 and it included some basic encryption and security

features to control access to the SNMP data. The present version, Version 3, was published in 1999 and it included the present generation of security features including encryption and password protection. Unfortunately, security standards that were used 20 years ago cannot easily withstand a present-generation cyberattack scenario. In 2012, an excellent research paper from Dr. Patrick Traynor at the Georgia Institute of Technology showed that SNMPv3 (as Version 3 is commonly known) was no longer up to the security standards of the day. The study concluded the following:

> SNMPv3 fails to provide its advertised security guarantees. . .These vulnerabilities are implementation agnostic and demonstrate a fundamental flaw in the current protocol. [1]

Thus, the most commonly used protocol within data center management systems is fundamentally flawed and does not provide the security necessary to protect the systems which employ it as a management protocol. In fact, Dr. Traynor showed how SNMP can actually be weaponized against the device within which it is deployed. The following examples shown in Table 20.1 below were given in the report as possible outcomes of an SNMP-based attack on various types of devices.

Because SNMP is deployed across IT and OT systems, the chart shown in Table 20.1 demonstrates that attacks can be carried out against IT equipment such as Managed Switches as well as against OT systems, including UPS, PDU, and Air Conditioning Units. Clearly, the ability of a hacker to carry out such attacks necessitates the use of security to mitigate against these scenarios. The specifics of those security mitigation requirements will be discussed in Section 20.7.

20.3.2 Modbus/TCP

Modbus/TCP is sometimes called the grandfather of all present-generation management protocols as its origins date to 1978 when it was a serial-based protocol. The network version, known as Modbus/TCP, was developed in the 1980s and used the same data structure as the serial protocol to keep backward compatibility. As a serial-based protocol, Modbus was developed with no security or encryption whatsoever. That is to say, its messages are passed over a wired or wireless connection in plain text so that anyone with a network sniffer program can read messages and can send messages to a Modbus-based system. Figure 20.2 shows a Modbus transmission captured by a common network sniffer, Wireshark.

As can be seen in the screen shot of the sniffed Modbus/TCP information, the message and all of its contents (which are actually in Hexadecimal form and are translated by Wireshark) can be seen by anyone who is connected to that network.

This inherent insecurity in the Modbus protocol has led to many warnings from government and other organizations. The California Energy Commission specifically noted the danger of using Modbus devices within the State's Demand Response programs due to the potential for damaging the electrical grid and its customers.

> The Modbus protocol has become the de facto industrial communications standard. . .The Modbus protocol lacks the ability to authenticate a user and hence middle man **attacks can <u>easily</u> take place in Modbus.** [2]

Clearly, any Modbus device will require an external security system to protect that device and all the systems that it supports. As with SNMP, this chapter will cover security strategies to enable the safe use of Modbus-based systems in Section 20.7.

20.3.3 BACnet

Just as Modbus is an unsecured, open-text protocol, BACnet is almost universally deployed with no data encryption or security as well. BACnet was designed by ASHRAE in the early 1990s with a very basic-level security and became an ASHRAE standard in 1995. The security that was employed in all present versions of BACnet is considered to be very modest as compared to present cybersecurity hacking capabilities.

The BACnet working group has conceded this fundamental security weakness and has engaged members to work

TABLE 20.1 Examples of OT attacks and their consequences

Device	Capability	Consequences
HVAC	Conceal errors, adjust temperature/humidity, power cycling	Physical damage
Managed switches	Disable/modify authorization, disable/enable ports	DoS/network access
Power distribution unit	Modify voltage/current, low/high power threshold	DoS/physical damage
Perimeter sensors	Door/motion sensors can be disabled or subverted	Conceal physical access
UPS	Modify voltage/current, power thresholds, power cycling	DoS/physical damage

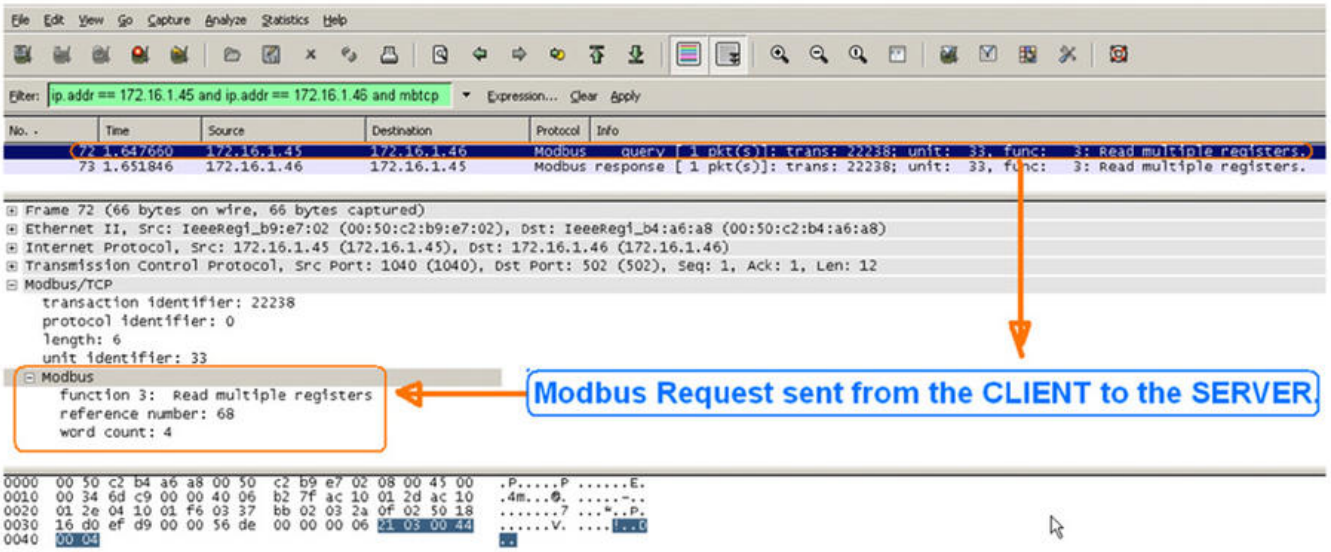


FIGURE 20.2 Modbus transmission captured by Wireshark.

together to produce a next-generation protocol that would meet the needs of securing against present-day attacks. In the working group paper, the group noted that:

> Network security in BACnet is optional. The existing BACnet Network Security architecture defined in Clause 24 is based on the 56-bit DES cryptographic standard and needs to be updated to meet the needs of today's security requirements. [3]

Because only a very modest security option is possible using current versions of BACnet, few manufacturers have chosen to offer security features within their BACnet devices. Consequently, a hacker of even modest means can sniff, find, and then take control of a BACnet device to change set points and manipulate these systems quite easily. As with SNMP and Modbus, this chapter will address security solutions for protecting BACnet devices in Section 20.7.

20.4 LEGISLATION COVERING OT SYSTEM SECURITY

There are a number of pieces of legislation that actively cover cybersecurity. Most are aware of the Health Insurance Portability and Accountability Act (HIPAA), which is a piece of national legislation that provides strict security requirements for healthcare and related facilities. Many are also aware of the General Data Protection Regulations (GDPR), which is the European Union's law that controls the protection of Personally Identifiable Information (PII) of those who live in an EU country. While awareness of these regulations is fairly high among data center professionals, the understanding of the provisions which apply to OT is scant. It's important for the engineering team that designs data centers as well as the facilities team which runs them to understand the key rules within these standards that apply to securing OT systems. This section will examine these two standards as they are related to OT systems and will also look at the recently passed California Law known as SB327 for securing connected devices. SB327 is becoming a template for other proposed local, state, and national legislation with regards to OT security and securing the Internet of Things.

20.4.1 HIPAA

The HIPAA law of 1996 (as revised in 2013) and its companion law, the Health Information Technology for Economic and Clinical Health Act of 2009 (the HITECH Act), form the legal requirements for protecting electronic Personal Health Information (ePHI) for individuals. They also create the penalties for the failure to do so and those penalties include fines in the millions of dollars that are being readily levied against those who break the law. With such enormous fines, it is vital for data center engineers and facilities managers to understand how these laws affect their OT systems and what they must do to protect these systems.

The HIPAA law calls out specific provisions for power systems in a key section. The power provision is contained in Section **164.308(a)(7)(ii)(C):**

> Establish (and implement as needed) procedures to enable continuation of critical business processes for protection of the security of electronic protected health information while operating in emergency mode.

This section was further clarified in the Health and Human Services (HHS) HIPAA Series #2 Bulletin published

in March 2007. In this publication, the above section is repeated with the following statement:

> When a covered entity is operating in emergency mode due to a technical failure or power outage, security processes to protect EPHI must be maintained.

Thus, if there is a power failure, it is a requirement of the law that all medical records must be maintained and protected. As a provision of HIPAA, this means that all physical safeguards and all cyber safeguards must be planned and implemented in order to be compliant with this provision of HIPAA. Enormous sums are spent to engineer multiple failover systems for power in a data center within a medical facility or a data center which is used to process or store medical data. But all such engineering can be rendered mute by a single click on the mouse from a cyber-criminal.

We have seen that power systems rely on SNMP and Modbus as their primary remote management protocols. Because Modbus has no security and SNMP is no longer a significant challenge to hackers, these protocols can be easily used as either a means to shut down power in a facility or to act as a backdoor through which servers and storage systems can be attacked and data theft can be orchestrated. Because of this, all critical power systems within a data center that falls under the terms of HIPAA must have dedicated and specific cybersecurity engineered protection engineered and implemented in their communications systems.

20.4.2 GDPR

The GDPR covers the Personally Identifiable Information of European Citizens, whether that data is located within Europe or not. So, any data center which stores or processes information with respect to European Citizens falls under the coverage of GDPR. The primary provision of the GDPR which affects data center OT is Article 32 which requires in part:

> 1. (b) the ability to ensure the ongoing confidentiality, integrity, availability, and resilience of processing systems and services.
> 2. In assessing the appropriate level of security account shall be taken in particular of the risks that are presented by processing, in particular from accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to personal data transmitted, stored, or otherwise processed.

These provisions are very similar in their goal of maintaining the availability to data under the terms of HIPAA. They also specifically spell out that all processing systems must have ongoing "confidentiality, integrity, availability and resilience." This makes it clear that power systems and cooling systems required by IT operations must be secured against all physical and cyber threats. Like HIPAA now has enormous fines for a data breach where it is shown that the organization failed to follow the provisions of the law, so too GDPR has significant fines for failure to follow the law. In the case of GDPR, a breach or destruction of data can cause a fine to be levied for up to 4% of the annual global revenue of an organization. Without question, GDPR, like HIPAA, is a law with teeth that must be followed.

20.4.3 SB327

California Senate Bill 327 was signed into law in September of 2018 and formally took effect on January 1, 2020. This bill, sometimes referred to as the Internet of Things Security Bill, makes a fundamental shift in data privacy. Rather than simply protecting the personally identifiable information of individuals, it also requires the protection of any and all information that resides within a network-connected device. The logic behind the law is that network-connected devices such as home and office environmental controls, lighting controls, security systems, and others have been successfully attacked and have caused harm to the individuals that those systems were supposed to support. This law does not just cover IoT devices in homes but it covers OT devices in data centers and in all facilities throughout the state.

The specific provisions of the bill are encapsulated by the Legislative Counsel's Digest which states:

> This bill, beginning on January 1, 2020, would require a manufacturer of a connected device, as those terms are defined, to equip the device with a reasonable security feature or features that are appropriate to the nature and function of the device, appropriate to the information it may collect, contain, or transmit, and designed to protect the device and any information contained therein from unauthorized access, destruction, use, modification, or disclosure, as specified.

Because it has been demonstrated that Modbus and BACnet have no such security features and, because SNMP is an openly attackable protocol, it is clear that devices which use these protocols would not be allowed to be purchased in California without an externally equipped security system. There is a great interest in how this law will be enforced and how it will succeed as many other states and federal bodies consider similar legal requirements. It is incumbent upon the data center designer and manager of any California-based California facilities to ensure that their OT systems do come equipped with proper security features beginning in 2020.

20.5 CYBER INCIDENTS INVOLVING DATA CENTER OT SYSTEMS

While cyberattacks make headlines on a regular basis, the most common form of cyber disruption is known simply as

a cyber incident. The U.S. National Institute of Standards and Technology (NIST) defines a cyber incident as follows:

> An occurrence that actually or potentially jeopardizes the confidentiality, integrity, or availability of an information system or the information the system processes, stores, or transmits or that constitutes a violation or imminent threat of violation of security policies, security procedures, or acceptable use policies.

Note that a Cyber Incident may or may not be caused by an intentional act. It is an intentional action that causes an event to be classified as a Cyber Attack as opposed to simply a Cyber Incident. In some cases, as in the case which will be discussed in this section, it may not be determinable as to whether the incident was an intentional event or not. In such cases, the classification of cyber incident is appropriate.

A number of cyber incidents have affected systems inside of data centers. Some of these have become public knowledge and were published in the newspapers and in online publications. One such cyber incident that was widely publicized involved two British Air data center in the UK. On May 27, 2017, British Airways noted that their Boadicea House and Comet House data centers experienced major power outages due to an electrical grid surge. However, their power supplier, the National Grid, advised that there were no problems with its power systems within the vicinity of these data centers. Consequently, any loss of power had to originate from systems within the data center.

In the subsequent investigation, it was discovered that the UPS control at Boadicea House was altered into a configuration that resulted in a hard shutdown. The UPS is supposed to switch to battery in the event it sees power that is outside of a predefined tolerance window. In this case, rather than switch to battery in this alleged power surge, the UPS simply turned-itself-off. Further, because the UPS was in-line with all power connections, the startup of the generator provided no immediate relief to the problem. The net result of the UPS shutting down and remaining down was a significant interruption to data operations that lasted for hours and cost the company enormous sums of money and operations disruptions. While the Comet House data center also failed at, or near the same time as Boadicea, no additional information on its failure has been made available.

How could a UPS fail to perform its normal functions and, rather than switch to battery, simply turn-its-systems-off? The answer lies in the manner in which the SNMP communications card in this unit and in many units used throughout the world operate. One of the features of a UPS card that has been present since the early days of SNMP cards is the option to shut the system down if it experiences a power anomaly. There are a number of ways in which this can be programmed into the SNMP card to create a disaster including:

- Setting the amount of time that the UPS remains on battery before shutting down to 0 seconds
- Setting the battery percentage shutdown trigger to shut down the UPS at 100% battery charge remaining
- Setting the battery time remaining to 0

These settings are presented on many SNMP cards as a legacy feature from the 1990's. During that time, many smaller UPS units were connected to a single server and UPS manufacturers provided server-resident software known as a "graceful shutdown" software to be triggered in the event of the UPS transferring to battery power. In this combination of UPS SNMP communications and server software, the user was allowed to choose to shut down the UPS once the server had gracefully shut down. The UPS SNMP card allowed the user to choose to shut down the UPS via simple timers or by battery time or charge percentage remaining.

Clearly, today's data centers operate with the philosophy that you should never shut down the server, which is the antithesis of the logic presented in these older UPS units. Unfortunately, many companies have failed to remove the shutdown options within their SNMP cards and this has led to a number of incidents such as the British Air disaster.

SNMP cards used in UPS systems are simply an example of features that can be exploited to create a cyber incident or to stage a cyberattack. It is incumbent upon the data center engineer and the facilities manager to understand all of the options that are present within the communication cards of all their mission critical systems. If shutdown features are exposed in these cards, the cards would need to be removed, reengineered, or have their access restricted to a single manager. No system located in a data center that is manned on a 7×24 basis needs to have an option to shut down one of their critical OT systems via a network connection.

20.6 CYBERATTACKS TARGETING OT SYSTEMS

Cyberattacks are events when there is clearly an attacker(s) that have attempted to violate security systems within a location. The NIST formally defines a cyberattack as:

> An attempt to gain unauthorized access to system services, resources, or information, or an attempt to compromise system integrity.

Thus, a cyberattack is an intentional effort to obtain unauthorized access to systems and their data or to destroy the integrity of systems. A number of cyberattacks have taken place and are continuing to take place on OT systems inside data centers. In many cases, the attacks simply use these systems as a backdoor to steal data. In others, systems have been shut down. In most cases where data theft is involved by using OT systems as backdoors, the actions are not recognized until many months after the event. One such attack that used OT systems as a backdoor to steal data was discovered and publicized but, only because the hacker responsible for the attack posted the means which they used to steal the data in an online site.

In this specific instance, a hacker used a Rack Power Distribution Unit (PDU) as a means to enter a server within that same rack. Rack PDU's represent a pernicious threat both because of their proximity to data servers and because they are ubiquitous throughout data center cabinets and racks. The targeted company was Staminus Communications, ironically which is a provider of cybersecurity protection services. In this case, the Staminus data servers were penetrated and data was stolen, all of this using a path through the rack PDU. The data which was stolen were then published to embarrass certain companies and individuals as part of the taunting engaged-in by the perpetrator.

In another case, an international incident evolved from a Russian attack on a Ukrainian power plant which involved attacking a UPS in their small data operations center. In this case, a UPS communications card was penetrated and malware was loaded onto that card. The malware was set to turn-off the UPS at the exact same instant as another attack was shutting down the master breaker which fed power from the plant to the grid. By shutting down the UPS at the same time as grid power was interrupted, this nation-state attack succeeded in creating a blackout in the control room of the plant and left the managers unable to communicate internally or externally.

Power systems are a target of cyber criminals because of their ability to be in proximity to data and because they can be used to shut-off critical power. This makes them prime targets to use as backdoors by data thieves and to use as a means for sabotage for those so inclined.

Many cybersecurity experts look at large scale cyberattacks on infrastructure, such as the attack in the Ukraine, as test cases that are being perfected for larger attacks against government and businesses. Because of this, it should not be a surprise that the SolarWinds cyberattack used that company's SNMP Network Management System (NMS) as the basis for its massive attack on the United States and other countries.

An NMS like the SolarWinds Orion is actually a perfect tool for a cyberattack because it is routinely used to monitor and control both IT systems, such as servers and switches, as well as OT systems like Rack PDUs and even larger UPSs. The key is that these systems both monitor devices using SNMP Get requests and they can also control devices with SNMP Set commands. In the SolarWinds attack, a piece of malicious code was inserted into the software update of the Orion system and this code gave its perpetrators unfettered use of the system's ability to both monitor and control critical IT and OT systems at will.

Two things are clear from this attack. Firstly, that any system which uses SNMP can be used to gain access to IT systems to surveil and steal data from these systems. Second, that SNMP-based systems can be used, much as they were in the Ukraine, to plant code in OT systems to surveil and accomplish other nefarious purposes. The full extent of the damage done in this attack will likely not be known for months as it is believed that many pieces of malware are now in place but are sitting, undetected, until programmed to act.

In order to understand the level of vulnerability that exists in presently installed power systems, the author undertook a study to view power systems located in data centers yet, were directly viewable on the Internet. It was discovered that a significant number of data centers have such systems that were directly connected to the Internet at the time of the viewing. This research turned up well more than 100,000 power systems in data centers that can be discovered. In addition, a large number of environmental control systems were discovered as well.

It is unfortunate that power devices have been so readily connected to the Internet but, by publishing this information it is hoped that the practice can be stopped and reversed for existing systems. In light of this, it should be the goal of the data center engineer and the facilities manager to ensure that all applicable security regulations and laws are being followed and, most importantly, that all basic security practices necessary to ensuring the safe and reliable use of OT equipment are followed. In the following section of this chapter, a roadmap to implementing a successful cybersecurity strategy for data center OT systems will be put forth.

20.7 PROTECTING OT SYSTEMS FROM CYBER COMPROMISE

In this chapter, it has been shown that OT systems within data centers can be compromised and that cyber incidents and attacks have resulted in large losses to those who had used that data center for processing. With a number of pieces of legislation now active and new laws being debated about cybersecurity for OT systems, it is the time for all data centers to work to mitigate the possibility of a cyberattack on their OT systems. In this section, we will explore sound and proven ways to greatly reduce the possibility of a cyberattack on these systems.

As we look to protect OT systems from cyberattacks as well as cyber events, the first thing that must be noted is that there are simply no known devices that will protect every system from all forms of cyberattacks. Because of this, it is necessary to provide a layered system of multiple types of defenses in order to best thwart attacks. Cyber events, whether accidental or not, can be reduced to near zero with proper layering of cyber defenses. Cyberattacks, while never having their possibility reduced to zero, can have their threat greatly reduced by a multilayer defense strategy.

The four layers of defense that every organization should use as a minimum for its OT cybersecurity strategy include the following:

1. Segment the OT network from the IT network
2. Create a perimeter firewall around the OT network
3. Provide Virtual Private Network security to each management console
4. Use 2-Factor Authentication when available

These four defense strategies will be discussed as individual items in the following sections. By employing these four layers within an OT network, a system of speed bumps is created which greatly adds to the complexity that must be navigated in order for an unauthorized individual to gain access to a piece of equipment. While it is also true that these layers do slightly increase the time to login to systems for the local staff, there is always a trade-off between convenience and security. In order to have a reasonable amount of security present in a system, there must be some speed bumps which reduce some measure of convenience.

This is no different from other types of security with which we have all become accustomed. For example, the use of airport screening systems is now a standard as is the use of two-factor authentication to log into critical IT systems. It should be not different for the need to protect OT systems. These four layers of defense will be discussed in order in the following sections.

20.7.1 Segment the OT Network from the IT Network

The facilities department is ultimately responsible for the installation, maintenance, and security of the OT systems under their control. Yet, most often, facilities OT systems are connected to an IT network and any problem with that network will impact the operation of those OT systems. Because OT forms the bedrock of all operations as was shown in Figure 20.1, it is highly preferable to segment that OT network from the IT network to eliminate the possibility of crossover malware and attacks. Further, because authority and responsibility must always lie with the same personnel, it therefore is appropriate that OT system security be placed with the facilities staff.

Segmentation from the IT network can take one of two forms:

1. Complete Isolation
2. Dependent Isolation

In the case of complete isolation from the IT network, the OT network is constructed with its own connection to the Internet, its own dedicated cabling, and its own network switches and systems. This is the preferred method of isolation as there are no dependencies on the IT network for support and no cross-contamination possibilities between the two networks.

The cost to build and maintain a separate network for OT systems is likely to be more than to essentially lease them from the IT department. However, with well-selected staff to monitor and manage the OT network, the long-term costs may match and even be lower than the cost to an organization of subleasing the IT network plant, maintenance, and management. Most importantly, it puts authority and responsibility together in the appropriate department.

For Dependent Isolation, one would take a feed directly from the IT network and then create a separate subnet from this point. The key point to consider when building a dependent network is to ensure that a firewall is installed at the point of connection to the IT network. In this way, the facilities department has the ability to control the communications within its own subnetwork. Specifics of firewalls are discussed in the next section.

20.7.2 Install a Perimeter Firewall System

Firewall systems are a must for any network and certainly when enormous sums of money are spent on mission critical OT systems in a data center, those systems must be protected at the same level as the IT systems which they support. There should never be a consideration that the OT network is less critical than the IT network.

All data center IT networks need to employ a robust perimeter firewall system to keep those that would seek to compromise their data systems at bay. To that end, selecting a firewall should involve the same search for a robust system that can handle the latest cyber threats. The desire by anyone to attempt the use of a firewall system that is intended for a Small Office/Home Office (SOHO) environment is a prescription for disaster when dealing with an OT network. Yet, many of these types of devices are used as the primary firewall for OT system networks.

Whether the facilities engineer and manager decide to employ a completely isolated network or a network that is based on the IT network with isolation added at their connection point, a strong and robust firewall must be chosen. The best firewalls will include the following features:

- Continuous cloud-based updates of the system via a global threat database, DNS threat management and mitigation
- Wide-ranging rule sets for granting and denying access via protocol, IP address, location, date, time, and position
- Easy-to-use and logical Graphical User Interface (GUI)

Facilities engineers and managers are often afraid of firewalls because of the lack of understanding about how they work and what they do. But, like anything one has learned, it merely takes the initiative and desire to understand the basic features of a product. Firewalls are no different than any other newer piece of technology and they should never be shied-away-from. The use of a next generation firewall is a must to protect OT systems in order to keep systems operating.

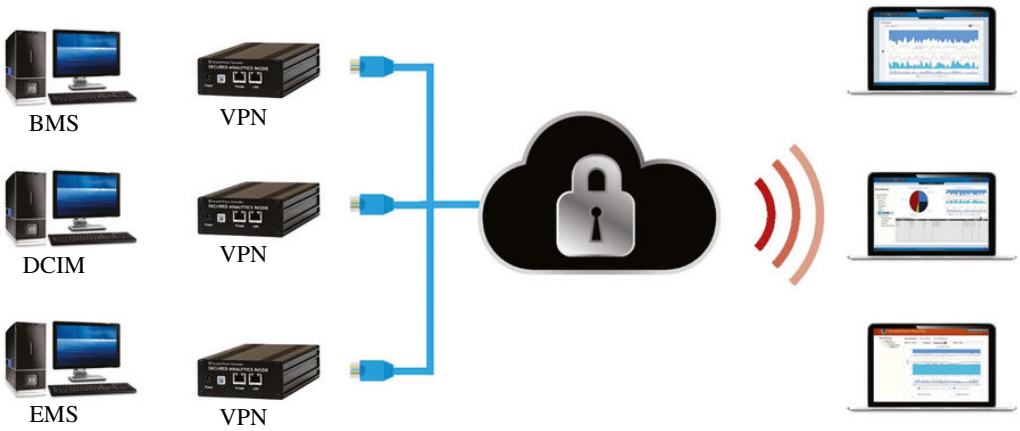


FIGURE 20.3 Using VPN's to secure DCIM systems. Source: Courtesy of AlphaGuardian.

20.7.3 Install a Virtual Private Network for Each Management Console

Most data centers use a variety of management systems including a BMS, DCIM, Energy Monitoring System (EMS), and others. In general, users of these systems have access to them both locally and remotely. Whenever a remote connection to one of these systems is used, there is the possibility of eavesdropping and hacking. Because of this, it is vital that each individual system have a dedicated VPN system.

Through the use of segmentation and the use of a perimeter firewall, two important speed bumps are erected in a network system. By adding a VPN to each console system, as shown in Figure 20.3, you add an addition speed bump by creating a secure, dedicated login system for the individuals who may access one of those systems at any given time.

20.7.4 Use 2-Factor Authentication

Many different types of systems now offer the use of a 2-Factor Authentication (2FA) in order to secure logins and system changes. In some cases, these are offered directly from the vendor but, 2FA software can also be purchased from a number of organizations and then added to console systems such as BMS, DCIM, EMS, etc.

A 2FA system is an added level of security that adds a second "personal touch" layer of authentication to a system login or system change request. Typically, a piece of 2FA software is loaded on the host server and then connected to an external hardware device, such as a finger print scanner or it is virtually connected to a mobile app. Upon logging into that system, the system will either request that the user scan their finger or, will send a message to their mobile phone and await their approval of the login as shown in Figure 20.4 below. The same process can be configured in the event of a major system change request, where a second layer of authentication is needed before the change is implemented.

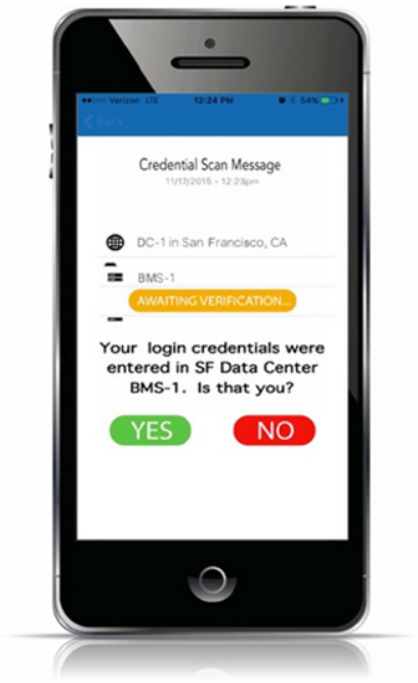


FIGURE 20.4 Using 2-Factor Authentication to secure logins. Source: Courtesy of AlphaGuardian.

2FA systems are standard at most critical government facilities and are now entering many data center IT applications as well. Because the OT network must remain fully operational to support the IT applications, it is a certainty that 2FA must be adopted for all key console systems that monitor or manager OT systems within a data center.

20.8 CONCLUSION

Cyber threats are a real and growing problem for managers of IT data systems and now for managers of OT data center infrastructure as well. Intentional cyberattacks and unintentional

cyber events can be equally destructive to the data that is being processed, transported, and stored in a data center. With a single click of a mouse, a UPS system or other mission critical OT system can be shutdown, offsetting the best electrical engineering and mechanical engineering design efforts.

Fortunately, a wealth of cybersecurity resources are available to the facilities engineer and manager. These resources include the latest generation isolation standards, firewalls, VPN's, and 2FA systems. The use of a layered approach employing all of these options creates a well-fortified defense for the OT network and its systems. This creates the protection needed to ensure that the high uptime goals for the data center can be reached and that they will not be compromised by a cyber event or cyberattack with respect to an OT system.

REFERENCES

[1] The California Energy Commission. Best Practices for Handling Smart Grid Cybersecurity. Final Project Report; May 2014.

[2] BACnet, a Data Communication Protocol for Building Automation and Control Networks. Addendum g to ANSI/ASHRAE Standard 135-204.

[3] Traynor P, Lawrence N. Under New Management, Practical Attacks on SNMP Version 3. Georgia Institute of Technology; 2012.

21

CONSIDERATION OF MICROGRIDS FOR DATA CENTERS

RICHARD T. STUEBI
President, Future Energy Advisors
Institute for Sustainable Energy, Boston University, Boston, Massachusetts, United States of America

21.1 INTRODUCTION

Data centers require large quantities of high-quality electricity. Electricity bills for a data center can easily run in the hundreds of thousands of dollars per year.[1]

Clearly, electricity is a major operational consideration for data center owners and operators. However, electricity plays a much more important role for a data center than merely being a significant line item in the annual budget.

Reliable supplies of electricity are mandatory for data center operators to offer continuous service. Data center customers have minimal tolerance for downtime, whatever the reason may be, so extra provisions are usually taken in the design of data centers to ensure that electricity is always available under any circumstances.

Even so, power outages still remain the most frequent cause of data center downtime, responsible for about one-third of reported incidents [1].

Over the years, power system engineers have accumulated a large set of approaches to assure high-quality/high-reliability electricity to facilities, such as the use of uninterruptible power supply (UPS) systems backed up by (usually diesel-fueled) generators. Ongoing technology advancements in a variety of disciplines continue to expand the toolbox for providing high-quality/high-reliability power to customers with demanding electricity requirements.

One emerging electricity supply alternative that merits its own category or classification is grouped under the general term "microgrid."

[1] See https://www.svlg.org/wp-content/uploads/2014/11/DCES_Arc_11032014_FINAL.pdf, p 11.

Although the term is relatively new, the concept of what is now termed a microgrid dates back to the origins of the electricity industry in the late nineteenth century. Before any common utility grid existed, electricity production and control had to be co-located with electricity consumption, so any household or business that wanted to take advantage of electricity assumed the obligation to build and operate an on-site electricity network.

As the electricity grid became built out, distribution utilities arose to provide universal electricity service at high reliability levels, and the need for most customers to manage their own electricity provision on-site disappeared. However, multiple forces have arisen in recent decades to revive active consideration of microgrids.

21.1.1 Growing Interest in Microgrids

Although microgrids remain rare, the U.S. microgrid market has more than doubled in the past decade [2]. Growing customer interest in microgrids is being propelled by three sets of factors:

- *Greater social/commercial dependence on digital devices and services*. Before the information/computer era, virtually all electricity-consuming equipment was resistance or rotary in nature, for which small disturbances in power quality and short duration outages were not a significant problem, and consequently the analog-based electricity grid was perfectly adequate for the task. As the grid still remains mostly analog, the resulting electricity quality it delivers is less than ideal for digital equipment, which is much less tolerant of any perturbations in electricity supply. Since modern life involves

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

ever-greater reliance on digital rather than analog activities, the limitations of the legacy electricity grid have become increasingly apparent. Moreover, because they are nearly 100% dependent upon digital equipment, data centers are a prime candidate for an alternative to conventional electricity supply from the grid.

- *Increasing likelihood of electricity grid disruption from human and natural causes*. Especially during the last 20 years, the possibility for long-duration regional power outages has become ever-more apparent. The 9/11 attack illustrated the capacity for human malevolence to take down critical infrastructure, and the lurking threat of cyberterrorism looms over electric utility networks worldwide. Meanwhile, natural disasters as disparate as Hurricane Katrina, Superstorm Sandy, the 2011 Japanese earthquake/tsunami, and recurring wildfires in (among other places) California and Australia have demonstrated how nature can produce extended absences of electricity—and may do so more widely and frequently as climate change drives higher incidence of extreme weather events. Not only are data center owners and operators cognizant of these risks, they usually aim to assure their customers that these risks will be mitigated.
- *Improvements in costs and performance of on-site electricity technologies*. Throughout the twentieth century, electricity generation benefitted from economies of scale: bigger was cheaper, so utilities built ever-larger power plants in rural areas and transmission lines to wheel the electricity to urban areas. However, this trend has now reversed, to the extent now that small-scale generation deployable at a customer's site is often cheaper than the larger options formerly selected by utilities on behalf of their customers. The most obvious example of this phenomenon is the advancement of solar photovoltaics (PV), for which costs have declined by several orders of magnitude in recent decades, to the point where PV-based electricity is now the least-cost generation option in many locations around the world—and not just in sunny areas. Lithium-ion batteries that enable electricity storage in large quantities are following a similar cost trajectory. As a result, customer-sited generation of electricity is an increasingly economic alternative to grid-supplied electricity for an expanding set of data centers and other facilities.

Aware of these three factors, data center owners and operators are giving more consideration to adoption of microgrids at their facilities.

21.1.2 Preview of Chapter Contents

The following discussion in this chapter will:

- Describe microgrids, identifying their defining elements/characteristics and how they create value for customers.
- Discuss how data center owners/operators should evaluate microgrids as a possibility during initial facility design/construction.
- Summarize the current U.S. microgrid market, including the microgrid value chain and key players.

21.2 DESCRIPTION OF MICROGRIDS

Before discussing how microgrids can be an attractive electricity supply solution for data centers, it is first important to more clearly understand what a microgrid is.

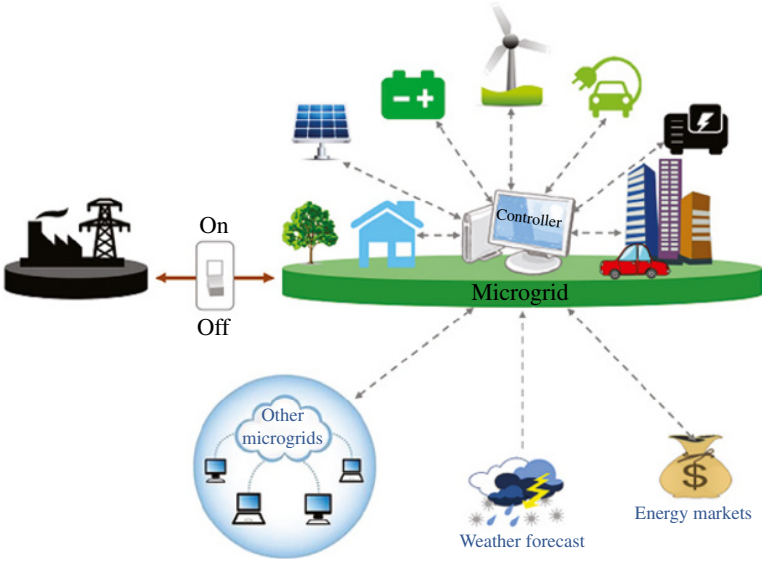


FIGURE 21.1 Overview of a microgrid. Source: U.S. Department of Energy, Microgrids at Berkeley Lab.

21.2.1 Defining Elements/Characteristics of Microgrids

Figure 21.1 presents a schematic depiction of a typical microgrid.

As the word itself connotes, a microgrid is a small electricity grid, containing all of the constitutive elements associated with any electricity grid:

- A set of electricity demands: one or more customers with various devices that consume electricity
- A set of electricity supply sources: one or more devices that have the capacity to generate (or store) electricity
- A network to deliver electricity from the supply sources to the points of consumption
- A control system to manage the interactions between demand points, supply points, and the network—as well as outside information sources critical for optimizing microgrid operations

In remote off-grid or developing world contexts, where there is no electricity available for many miles around, a microgrid is often the most viable option for providing electricity service. By necessity, a microgrid in such a location must be able to operate as a stand-alone system. However, few data centers are built in such areas.

By contrast, for data centers in well-developed areas with ubiquitous access to electricity, a microgrid almost always involves a small grid within the much larger regional utility grid. Most importantly, the microgrid is able to disconnect from the primary grid as needed to operate autonomously and continue providing electricity to its customers when the primary grid is otherwise experiencing an outage.

Most microgrids for data centers in developed economies would normally operate in "parallel" or "synchronized" mode, wherein the data center microgrid remains interconnected with and operates as part of the main utility grid, thereby enabling constant interchange of electricity between the two, with supply provided from whichever is least-cost. However, in the event that the grid is experiencing or might face an emergency, the microgrid can shift into so-called "islanded" mode, during which the microgrid separates itself from the regional grid, thereby effectively creating an "air gap" to secure operation for the customer(s) served by the microgrid.

Islanding a microgrid is therefore the electrical analogue of a fortress pulling up the drawbridges to insulate the local population from external threats.

Because of this design principle, a data center with a robustly designed and well-managed microgrid can expect to have access to high-quality electricity supplies under virtually any circumstance—including prolonged grid outages such as those that follow major natural disasters.

21.2.2 Sources of Value from Implementing Microgrids

Microgrids can create value for customers in three main ways:

- *Electricity supply resilience*. As suggested by the preceding text, a major driver of data center interest in microgrids is the ability to remain operational for prolonged periods of time while the primary power grid is down for some reason. Data centers stand to lose hundreds of thousands of dollars for each hour of downtime, yet conventional UPS systems (even when equipped with backup generators) often are inadequate to prevent data center outages.[2] Therefore, data center owners are increasingly willing to spend significant capital and management attention on a solution such as a microgrid in order to avoid the prospect of multi-hour (or worse, multi-day) grid outages. As might be expected, considering a microgrid to improve electricity supply resilience is especially prudent in locations facing higher likelihood of extended grid service interruptions, as has recently been occurring in California to mitigate the risk of wildfires.[3]
- *Reduction in total expenditures on electricity*. Adoption of a microgrid offers the potential for a data center to reduce total expenditures on electricity—comprising both the costs of electricity purchased from the utility and the costs of owning/operating on-site equipment to supply electricity when the utility grid is down. While still atypical today, this outcome will become more commonplace as the costs of self-generation (especially PV, but also including emerging distributed generation technologies such as fuel cells) and energy storage continue to fall. This is especially the case where electricity prices from the grid are already high or are rising. Indeed, even if electricity prices are only becoming more unpredictable—due, for instance, to local regulatory uncertainties—a microgrid can serve as a hedge against unexpected increases in electricity prices.
- *Ability to ensure desired environmental profile of electricity supplies*. When buying electricity from the grid, a customer is obtaining power that is generated from a wide variety of sources depending upon the region, ranging from zero-carbon solar and wind to carbon-intensive coal generation. Even when a customer purchases renewable energy through a power purchase agreement (PPA), such a commitment is often virtual, involving some reliance upon nonrenewable

[2] https://lifelinedatacenters.com/data-center/data-center-downtime/.
[3] https://www.greenbiz.com/article/microgrids-could-help-california-improve-grid-resilience-face-wildfire-threat.

energy at certain times. In contrast, an islandable microgrid can remove any ambiguity about the source of the electricity: if sufficiently motivated, a customer can design the generation supply side of the microgrid to be 100% renewable, thereby assuring absolutely no contribution of emissions from fossil fuel combustion to support data center operation. A microgrid can thus be an important tool for data center owner/operators in demonstrating commitment to environmental sustainability to both customers and employees.

21.2.3 Microgrids as an Extension of Overall Energy Management Strategy

As noted above, while a microgrid allows a customer to operate independently from the electric utility grid, implementing a microgrid doesn't mandate that the customer go entirely "off-grid."

Because of this optionality, it is important to think of microgrids as an element of a customer's overall energy management strategy, not necessarily the entire energy management strategy. To illustrate, consider the example set by Microsoft at its data center in Cheyenne, WY.

The bulk of the electricity requirements for Microsoft's Cheyenne data center is being supplied by two nearby wind farms, as delivered via the electricity grid owned by the local utility, Black Hills Energy. Meanwhile, when it is economically attractive to do so, Microsoft can sell electricity produced by the data center's backup generators (fueled by natural gas) to Black Hills Energy for the utility's uses, thereby allowing Microsoft to better monetize their otherwise underutilized on-site generating assets. And, if the utility grid goes down, Microsoft can still switch the data center's microgrid into islanded mode and rely upon its own gas-fired generators for constant power supply.[4]

For data centers, a microgrid can thus become the culmination of an integrative energy management optimization that encompasses:

- Maximizing power usage effectiveness (PUE) to minimize energy requirements of the data center.
- Securing power purchase agreements (PPAs) from grid-based electricity suppliers to provide the data center's electricity needs at certain times and/or certain conditions.
- Implementing a microgrid to provide the data center's electricity needs when not utilizing PPA-supplied power.
- Negotiating agreements with the local utility under which the microgrid's assets can optionally serve (and be compensated) as a grid resource.

21.3 CONSIDERING MICROGRIDS FOR DATA CENTERS

As the microgrid market grows, microgrids are becoming of greater interest to data center owners and operators. Although it estimates that data centers have historically accounted for less than 15% of the commercial/industrial (C&I) segment of the global microgrid market, Navigant Research believes that "data centers could represent as much as 40% or more of the total grid-tied worldwide C&I microgrid market" by the mid-2020s [3].

In addition to the Microsoft example discussed above, three other instances of data centers (as of December 2020) that have implemented or are implementing microgrids are:

- Aligned Data Centers has installed a 63 megawatt (MW) microgrid in Phoenix, AZ with the assistance of the local utility, Arizona Public Service.[5]
- EIP Investments is building a $1 billion data center with a microgrid powered by 20 MW of fuel cells in New Britain, CT.[6]
- DataBank is implementing a 1.5 MW microgrid in conjunction with its ATL1 data center in Atlanta, GA." footnote https://www.power-eng.com/on-site-power/microgrids/1-5-mw-georgia-tech-microgrid-project-ready-for-operation-in-atlanta-data-center/

With the foregoing as context, we are now better positioned to discuss the considerations involved in developing and operating a microgrid for a data center.

21.3.1 Development

Implementing a new microgrid for a data center involves the following steps:

Step1-*Determining data center electricity needs*. To create true satisfaction in any market-based activity, it is essential to first deeply understand the customer's needs—in this case, the electricity requirements for the data center. In turn, this requires:

- Estimating total electrical and cooling needs to operate electronics that will satisfy data center volume requirements, maximizing PUE wherever possible.
- Assessing which electrical/cooling loads are "critical" (i.e., essential for 24/7/365 data center operation), and which are more discretionary.
- Identifying which electrical loads are alternating current (ac) and which are direct current (dc), the latter of

[4] https://blogs.microsoft.com/on-the-issues/2016/11/14/latest-energy-deal-microsofts-cheyenne-datacenter-will-now-powered-entirely-wind-energy-keeping-us-course-build-greener-responsible-cloud/#sm.00008uqsgdj0ddgywdv2mh6r0js51.

[5] https://www.utilitydive.com/news/case-study-aps-invests-in-microgrid-as-phoenix-metropolitan-area-grows/503687/.

[6] https://microgridknowledge.com/fuel-cell-microgrid-data-center-connecticut/.

which would require conversion from ac-based electricity sourced from the grid.

- Understanding data center owner's overall strategic objective(s) and willingness for tradeoffs amoung energy supply options: maximizing value-creation, minimizing capital expenditures, attainment of probabilistic operational reliability targets, environmental goals.

Step2-*Developing microgrid that best meets needs*. Since a very large number of microgrid designs can serve the electrical needs of any data center, the challenge is to quickly and efficiently narrow down the microgrid options for a particular data center to a short list of highly attractive candidates, with enough specificity in design so that economics can be effectively assessed in making a final selection. The design process includes:

- Creating bundles of generation sources (optionally including storage), each of which would be adequate to ensure that critical loads are fully supplied at all times with an appropriate level of redundancy.
- Selecting which generation (and storage) technologies to consider as proven and commercially available to include in bundle creation.
- Determining whether or not cogeneration (i.e., generating heat on-site in addition to electricity, sometimes called combined heat and power or CHP) should be considered as an option for the purposes of most efficiently providing the data center's cooling requirements.
- Configuring a distribution network (and circuits as necessary) to distribute electricity from the on-site generation sources and from the utility grid to both critical and noncritical loads.
- Deciding topology of ac–dc conversion (i.e., whether or not to employ centralized vs. distributed inverters).
- Preparing a bill of materials for each microgrid option, comprising both the distribution network and generation bundles to serve data center's needs.
- Estimating the costs (net of any revenues from services sold back to the local utility) associated with installing and operating the microgrid, plus any purchases of power from the utility grid.
- Assessing the relative implementation risks and any non-economic attributes associated with each option.

Step3-*Ensuring microgrid is implemented successfully*. Once the basic parameters of an envisioned microgrid are agreed upon, the following activities are required to convert the selected system design into a working microgrid:

- Retaining an experienced engineering firm to draw up microgrid circuit design and specify required equipment.
- Securing interconnection agreement and negotiate commercial agreements with local utility (working with regulatory authorities as necessary).
- Obtaining any necessary permits and approvals from relevant authorities for microgrid development.
- Procuring required equipment from short-list of qualified vendors.
- Hiring qualified construction firm–as well as commissioning agent–agent to ensure installation has been performed properly and microgrid works as planned.

21.3.2 Operations

Once a microgrid has been installed, the physical equipment—mostly the generation sources (including energy storage, if any)—will require periodic maintenance, not conceptually different from the maintenance needs for physical plant in data centers without microgrids. To the extent that the microgrid is designed for self-generation supply most of the time, the equipment may need more frequent and intensive maintenance than is the case for data centers mostly reliant on the utility grid for electricity supply.

A more interesting issue for ongoing microgrid operation is how to optimize the utilization of the assets to maximize economic benefits to the host data center. This optimization is multidimensional, dynamic, continuous, and stochastic, involving several degrees of freedom that must be simultaneously managed (which also preserving electricity quality):

- Exploiting electricity demand reduction/shifting possibilities
- Generating more from dispatchable generators (i.e., generation assets other than PV)
- Charging or discharging from energy storage devices
- Transacting more (or less) with the utility grid

The microgrid controller—a critical component of the installed microgrid—is responsible for this complex optimization. With ongoing improvements in computational power, artificial intelligence, big data management, and predictive analytics algorithms, microgrid controllers are continually becoming more sophisticated in their ability to better undertake these optimizations.

The microgrid controller is also responsible for another sophisticated procedure: the islanding process, during which the microgrid transitions from being connected to the utility grid to operating in isolated mode.

For the data center to continue operating in the event of a utility grid outage, the islanding transition needs to be done seamlessly, yet instantaneously. In many cases, the transition to islanded mode is undertaken in anticipation of potential grid problems: if/as/when the utility grid indicates (either via an explicit warning communication or by virtue of observed power quality deterioration) that continued availability of stable power will be uncertain for an anticipated duration, the controller first ensures that enough generation (plus storage) is on-line to meet critical loads and then smoothly disassociates the microgrid from the main grid.

Once the utility grid is back operating stably, restoration of the microgrid connection to the grid is also delicately managed by the controller, synchronizing the ac frequencies of the microgrid and the utility grid.

To the extent that the microgrid is frequently interchanging electricity (buying from and/or selling to) the utility grid, it is critical for commercial (i.e., accounting and risk management) purposes for the microgrid controller to also track the flows of electricity at a very granular level of detail. At minimum, this implies stamping the volume and direction of flow between utility and microgrid for each time interval (at least every few minutes, and perhaps ultimately down to the second). More expansively, the desired informational requirements may include the actual source (down to the asset level) for each unit of electricity.

Accordingly, this brings into play the potential for blockchain approaches to provide the necessary transparency and validation that enables multi-asset, multiparty tracking of electricity production, delivery, storage, and use. While this vision is generally not yet commercially available, several start-up companies are actively developing blockchain platforms to enable so-called "transactive energy" markets to emerge, involving peer-to-peer transactions between "prosumers": parties connected to the grid that frequently flip between consumers and producers of electricity.

Data centers that choose to implement a microgrid may thus be eventually compelled to adopt blockchain technology in order to fully capture all the value that the microgrid affords.

21.4 U.S. MICROGRID MARKET

Because the microgrid market is still in its infancy, its growth trajectory remains somewhat erratic, with certain years evidencing rapid growth and other years indicating little growth or even retrenchment. However, growth trends in the microgrid market become more apparent when looking at data over longer periods of time. For instance, in the United States, annual installations of microgrids have more than doubled over the past decade, to roughly 500 MW/year [2].

Recent survey data indicate that the capital expenditures associated with a microgrid range from $1–5 million/MW [4]. At an approximate average of $2.5 million/MW, the size of the U.S. microgrid market is thus on the order of $1 billion/year.

21.4.1 Microgrid Value Chain

In essence, a microgrid is like any real estate project, requiring the full spectrum of development, finance, engineering, procurement, construction, and operational activities. Accordingly, players in the microgrid value chain typically fall into one or more of the following roles:

- *Project developers/integrators*. There is no such thing as an "off-the-shelf" microgrid: each microgrid is a tailored solution for a customer's specific circumstances and needs. A project developer/integrator meets with prospective customers to identify those well-suited for a microgrid application and—for those who commit to pursuing a microgrid—coordinates the design and implementation of the microgrid to its fruition.
- *Vendors*. A microgrid physically consists of a combination of equipment and parts from various suppliers that manufacture the product at a factory, deliver it to a customer site, and provide any needed after-sale support. The bill of materials for a microgrid is extensive, including:
  - Generation: solar, internal combustion engine, fuel cell, microturbine, etc.
  - Storage: batteries of various possible types (lithium ion, lead-acid, flow, etc.)
  - Network: wires/cables, connectors, switchgear, etc.
  - Controls: sensors, analytics, diagnostics, optimization, etc.

  Many vendors aim to package these various subcomponents into standardized elements, which in turn can be combined in multiple ways, with the goal of offering "plug-and-play" compatibility. While a desirable vision to be pursued, this degree of standardization is rarely fully achieved in the marketplace. As a result, and some degree of customized integration is almost always still required.
- *Engineering/construction*. Once the basic design elements of a microgrid are agreed upon with the customer, the developer/integrator will typically (1) hire an electrical engineering firm to architect the microgrid system and define requirements for components in the necessary degree of detail, and (2) an electrical contractor to perform the actual installation and commissioning of the system. Whereas the electrical contractor work is fairly standard and can be performed by many firms (especially those who are qualified in dc power), there is somewhat more specialization involved in microgrid design, and as such the universe of engineering firms expert in microgrids is somewhat smaller.
- *Financial sponsors*. If neither the customer nor the developer/integrator wants (or has the financial wherewithal) to own (i.e., pay up-front for) the microgrid, a growing number of specialized investors are in the market to front the required capital for microgrid projects. The terms of microgrid financing can be a straightforward equipment lease, but more commonly are structured as an "energy-as-a-service" contract encompassing both initial cost and ongoing costs over a defined term (usually more than 5 years, frequently 10–20 years, sometimes up to 50 years) in which the microgrid owner essentially becomes de facto an electric utility to the microgrid customer.

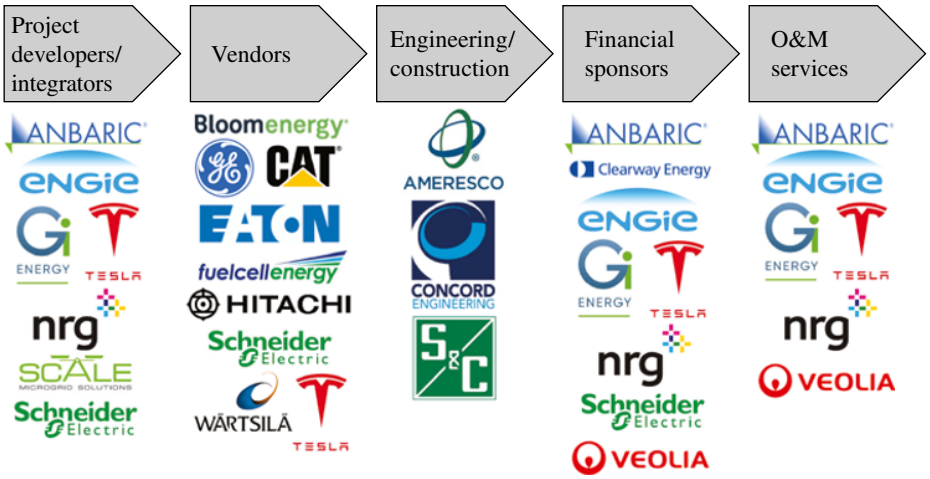


FIGURE 21.2 Microgrid value chain and selected key players. Source: Courtesy of Future Energy Advisors.

- *Operation and maintenance (O&M) services*. Once commissioned, the microgrid needs to function reliably over its lifetime: operated to provide constant high-quality supply of power to the data center at a low cost, maintained and repaired as necessary, and also (to the extent possible) dispatched with respect to the local utility grid to maximally monetize the investment in the system by selling energy surpluses in optimized amounts at optimum times. Although the data center owner/operator could take on these roles, since they are typically noncore to the data center business, these O&M services are likely to be handled by either the microgrid developer/integrator, control system vendor, and third-party providers.

21.4.2 Key Players in Microgrid Market

As demand and interest in microgrids have grown, so too have the number of players active in the market.

Figure 21.2 illustrates some of the leading players in the U.S. microgrid market. Note that, while some companies active in microgrids specialize in only one of the above-listed roles, others participate across most of the value chain, thereby offering a more integrative turnkey microgrid solution to customers.

21.5 CONCLUDING REMARKS

At the forefront of the twenty-first century economy and society, data center owners and operators are ideally positioned to benefit from state-of-the-art electricity service enabled by microgrids, given the ever-advancing technologies that are making microgrids increasingly attractive for economic supply of more resilient electricity.

Compared to conventional electricity service from the utility grid, microgrids leverage a distributed architecture of optimized devices to ensure a constant supply of high-quality electricity essential to data center success.

Those who are seeking to stay at the forefront of the data center sector would be well-advised to monitor advancements in microgrid-related technologies and developments within the microgrid marketplace, as microgrids offer great promise for sustainable competitive advantage in data center design and operation.

REFERENCES

[1] Uptime Institute. Uptime Institute Data Shows Outages Are Common, Costly and Preventable; June 2018. p 3.

[2] Wood Mackenzie. US Microgrid Market Forecast H2 2019: Commercial Customers Lead Market Demand; December 2019.

[3] Navigant Research. Data Centers and Advanced Microgrids: Meeting Resiliency, Efficiency and Sustainability Goals Through Smart and Cleaner Power Infrastructure, 4Q 2017. p 11.

[4] National Renewable Energy Laboratory. Phase I Microgrid Cost Study: Data Collection and Analysis of Microgrid Costs in the United States; October 2018.

FURTHER READING

California Energy Commission. Microgrid Analysis and Case Studies Report; August 2018.

Illinois Institute of Technology. Transactive Energy in Network Microgrids.

Lawrence Berkeley National Laboratory. The CERTS Microgrid Concept, as Demonstrated at the CERTS/AEP Microgrid Test Bed; September 2018.

Microgrid Knowledge. The Evolution of Distributed Energy Resources; 2018.

Microgrid Knowledge. How to Get Your Microgrid Projects Financed; 2019.

Microgrid Knowledge. Microgrids for the Retail Sector; 2019.

Navigant Research. Global Microgrid Trends; March 2019.

North Carolina State University. RIAPS: An Open Source Microgrid Operating System; March 2019.

Smart Electric Power Alliance. The Role of Microgrids in the Regulatory Compact; 2019.

Data Center Frontier. Microgrids and Data Centers: A More Holistic Approach to Power Security; June 2020.

Data Center Frontier: Data Microgrids on the Rise as the Colo Industry Deploys More Resilient Power Systems; June 2020.

Microgrid Knowledge. Data Center Microgrids: A Lot More Efficient and Effortless to Run; 2020.

Microgrid Knowledge. Educating Data Centers About Microgrid Benefits: More Than Backup Generation; 2020.

Mission Critical. Data Centers of the Future Require Microgrids; September 2020.

Schneider Electric. Microgrids for Data Centers: Enhancing Uptime While Reducing Costs and Carbon; August 2020.

PART III

DATA CENTER DESIGN & CONSTRUCTION

22

DATA CENTER SITE SEARCH AND SELECTION

KEN BAUDRY
K.J. Baudry, Inc., Atlanta, Georgia, United States of America

22.1 INTRODUCTION

Almost all data center disasters can be traced back to poor decisions in the selection, design, construction, or maintenance of the facility. This chapter will help you find the right combination and eliminate poor site selection as a cause of failure. It begins with setting objectives and building a team, examining the process, and selection considerations. The chapter concludes with a look at industry trends and how they may affect site selection.

Site selection is the process of identification, evaluation, and, ultimately, selection of a single site. In this context, a "site" is a community, city, or other populated area with a base of infrastructure (streets and utilities) and core services such as police, fire, safety, education, and parks and recreation. This definition is not meant to eliminate a location in the middle of a nowhere, the proverbial "corn field" location. However, experience indicates that unless an organization is really set on that idea, there are few such locations that have the key utilities (power and fiber) and core required to support a data center. Most organizations back away from this idea as they estimate the cost and logistics of operating such a facility.

Site search and selection can be as comprehensive or as simple as necessary to meet the goals of the organization. In practice, it consists of asking a lot of questions, gathering information, visiting potential communities, and perhaps entertainment by local and economic development officials.

The motivation behind conducting an extensive site search tends to be economic (i.e., find the site with the least total cost of ownership). While the site search and selection process attempts to establish an objective basis for the decision, it relies on estimates and assumptions about future conditions and often includes criteria that have an economic value but are not easily fit into an economic model. These criteria include marketing and political aspects; social responsibility, as in giving back to the community; and quality-of-life issues. These issues tend to be subjective and are often weighted heavier in the decision matrix than the economics suggests. The end result is that the site selection process tends to be a process of site elimination based on economic criteria until the list is pared down to a few sites with similar characteristics. The final decision is often subjective.

There is no such thing as a single "best site," and the goal of a site search is to select the site that meets the requirements, does not violate any constraints, and is a reasonable fit against the selection criteria.

In general, the goal in site search and selection is to ensure that there are a sufficient number of development opportunities in the selected community. However, there are good reasons to search for specific opportunities (existing data centers that are suitable or that can be refurbished, buildings that might be converted, pad sites, or raw land) as part of the site search and selection process. There are often negotiating advantages prior to the site selection announcement, and the identification of a specific facility or property might become the deciding factor between the short-listed sites.

22.2 SITE SEARCHES VERSUS FACILITY SEARCHES

Most of the discussion contained in this chapter is based on selection of a site for development of a data center from the ground up a.k.a. "a brown field site." However, there are other viable data center acquisition strategies such as buying

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

or leasing an existing single-tenant data center or leasing a co-location center. With co-location and existing facilities, much of the investigative work should have already been done for you. You will be able to ask the prospective landlords to provide answers to your questions. It is likely that some aspects, such as power company rates and tax incentives, have already been negotiated by the developer.

You will still need to understand what you want to achieve and what your requirements are, and it would be a mistake to assume that the developer had the uncanny ability to anticipate your requirements and has taken them all into consideration. Data centers are not a one-size-fits-all proposition, and existing centers may have been located where they are for reasons that aren't related to the data center business at all. It could be that the original tenant selected the location because it is where someone's grandchildren live.

You will still need to compare the information you are given against your requirements and analyze the total occupancy cost for each prospective facility. In many cases, the lowest cost alternative may not be the one with the best lease rate. It might be the one with the best power company rate and the most opportunities for energy savings driven by the local climate.

22.3 GLOBALIZATION AND THE SPEED OF LIGHT

Globalization has been around since Christopher Columbus's voyage, when a group of adventurers left one area to seek better fortunes in another area. Today, it is largely driven by economic forces as organizations seek out competitive sources of raw materials and labor and new markets for their products. Globalization and technology, in particular air travel and voice and data communications, have made it possible for organizations, more than ever before, to consider sites outside of their country of origin.

Global strategic location plan can be divided in macro, mid and micro levels (see chapter 1). Amazon Web Services have built a hierarchy of "Regions" located around the world. Within a "Region", there are "Available Zones" (AZ) and each AZ has a cluster of data centers.

Historically, organizations have located facilities overseas to secure raw materials or inexpensive labor and to avoid taxes and safety and regulatory policies of their country of origin. The question for us is: "Are there economic benefits to locating a data center outside of one's country of origin?"

The motivation for locating overseas, today, is very different from the raw materials and cheap labor considerations of the past. Data centers don't need a large pool of unskilled labor like manufacturing and call centers, and they are not tied to location by raw materials. Data centers can be located almost anywhere that can provide significant amounts of power and connectivity to high speed national and international communication networks.

Including sites in foreign countries adds a layer of complexity as the differences in tax structure, laws, ownership of real property, security of data, political unrest, etc. need to be considered. There is, however, one difference that is not easily eliminated by technology or money: communication signals cannot propagate any faster than the speed of light.

As an example of how this figures into site selection constraints, let us consider a route from Bombay, India, to New York City, New York, United States. The speed of light in free space is 299,792 km/s or approximately 186,282 mi/s. It is slightly slower in optical fiber, but we are going to ignore that for simplicity. The distance is about 7,800 miles, and assuming that a fiber route would be 50% longer than a direct path, it takes about 70 ms (1 ms is equal to 0.001 s) for the signal to propagate one way. This figure does not include any latency for network interfaces, transmission gear, etc. So unless Einstein was wrong, 70 ms is the best possible signal latency.

Today, the expected latency between the United States and India is between 215 and 305 ms. Network improvements will likely drop this figure over time to about 200 ms. How this limits site selection depends largely on the applications that you use.

One of the most common latency-sensitive applications is storage array replication. In data replication, an application will write data to a disk or storage array, and the array will replicate the same to a remote array. When confirmation comes back that the data has been successfully written, the transaction is complete. If the latency is high, then the performance is seriously degraded. Other latency-sensitive applications include transaction-oriented applications like banking and retail sales where, due to a very high transaction rate, the latency must be very low and burst capabilities are required to meet high demand.

Without a doubt, some organizations will realize significant benefits from selecting an overseas location. But there are challenges, and it will not be a suitable option for everyone.

22.3.1 The Site Selection Team

Like any process with many moving parts, site search and selection requires diligence, clear expectations, schedules with defined deliverables and due dates, and good communication between all stakeholders. This doesn't happen by itself.

Project management is the art of managing a process from beginning to end. It concerns itself with reaching the end state by defining and organizing the work effort, communicating and leading stakeholders, and driving decision making consistent with requirements, within constraints, and in a timely manner.

Project management sets metrics for constant feedback on performance and adjusts accordingly to keep on track. It is a profession, and the role of the project manager cannot be overstated. If you do not have someone in-house who has successfully demonstrated their competence, you should look outside of the organization.

Some key reasons why projects fail are similar across almost all endeavors are as follows:

- Lack of user involvement
- Unrealistic expectations
- Incomplete requirements and unsupportable criteria
- Lack of planning
- Lack of executive support

Before your organization can make good progress with the site search, it will be important to identify potential team members and start building the team. The timing of "when" you bring your team on board is as important to your success as "who" you bring on board. Assumptions about cost and feasibility are often made early in the program, before subject matter experts (SMEs) are traditionally on board, and often spell doom for a project when it is determined that they are not realistic and, more often than not, need to be adjusted upward. You will need your team on board early in the process to avoid this sort of pitfall and to create plausible criteria and constraint list.

As you build your site selection team, you should talk with key executives and ask for recommendations. By involving them early, reporting progress in a concise and effective manner, you will gain their support. If you are the key executive, then you should involve your board, investment committee, and your key subordinates as you will need their support as you push your search forward. As a minimum, this effort should lead to a better understanding of the project and identification of who will cooperatively back the project and who won't.

One of the key consultants will be a site search consultant. A good site search consultant will not only be an SME but also be an advisor. He will not only guide you through the site selection process but also will know what to expect from economic development agencies. He will know where to get answers, know other professional resources in the industry, and understand how to evaluate proposals. He will understand when it is time to bring the search to a close and how to bring it to a close. He will render advice and help you make good decisions. In many cases, he will be able to save you time and money, by eliminating some sites, based on recent relevant experiences prior to beginning a search.

Your team will need to include SMEs such as data center planners, data center consultants, architects, engineers, lawyers, tax accountants, and real estate brokers. Any specialty that can vary in cost between sites will need some level of representation on your team. The greater the expected cost variance between sites, the greater the need for the SME.

An architect might only be needed to create an initial estimate of the space required and nothing more if construction costs are believed to be the same regardless of site selection. An engineer might only be needed to create a power consumption profile and energy consumption and demand for the proposed facility, while a utility expert might be retained to evaluate energy costs for each prospective site.

When building a team, there is often a distinction between consultants and vendors. Consultants typically charge a fee for their services. Vendors typically provide preliminary support for free or for a minimal fee in hopes of making the big sale later on in the project. This includes real estate brokerage firms that will assist with site selection, general contractors that will perform preliminary design and develop budgets, and others. While you should surround yourself with the best resources that you can afford, the key is to surround yourself with resources that have the experience, competence, and that you can trust to act in your best interest. All three are important!

22.3.2 The Nature of the Site Search and Selection Process

There are a few characteristics of site search that require special mentioning; it's generally performed in secrecy, and it's not a search process but an elimination process, and there comes a time when continuing the process will no longer produce further value.

22.3.2.1 Secrecy

More often than not, a site search is conducted in secret. There are many reasons including the following:

- Site searches don't always result in moves or in the development of new facilities. Announcing a site search and then not following through may be perceived as a failure.
- News of a site search may cause concerns among employees over potential facility closures.
- Most businesses don't find it competitive to telegraph future plans to their competition.
- Once the word is out, you will be inundated by vendors wanting to get a piece of the action.

Regardless of your reasons, it is likely that management will expect that the search be conducted in secret. Most site search consultants are going to be aware that this will be the case. As the size of your team grows and many aspects of the work will require team members to consult with resources outside of the team, such as equipment vendors, it is likely that by the end of a lengthy search, there the number of individuals aware of your search will be quite large.

If secrecy is important to your organization, then you will need to take precautions. The first step should be execution of simple confidentiality and nondisclosure agreements with every team member. Many of the consultants will be looking

to perform additional services once the site is selected, and well-timed reminders about disclosure can be very effective.

Some companies give their site selection projects a code name. This works well as it gives everyone a name to reference and serves as a constant reminder that the project is secretive.

When it comes to secrecy, we are often our own worst enemy, and many individuals unknowingly let out the secret by wearing their company access card with picture ID, wearing clothing or carrying pens or clipboards with a corporate logo, handing out business cards, providing phone numbers that can be easily traced back to a company, etc. Even discussing your travel arrangements and where home is can provide significant clues to the curious. While all of these things are obvious, it's the most obvious things that we forget to deal with. If you really want secrecy, you will leave nothing to chance, and let your site search consultant handle all of the communications.

22.3.2.2 Process of Elimination

Site search and selection is a bit of a misnomer as it is more a process of elimination than selection. It starts with a broad universe of possibilities and moves through rounds of elimination (Fig. 22.1). The broadest criteria are applied in the first round to eliminate the most number of sites and narrow the search. The broadest criteria are usually geographic. For example, North America, Europe, cities of five million or more in population, within my service region, etc.

Each successive round uses more selective criteria until all but one site is eliminated. This may sound a bit daunting, but in practice, the number of possibilities falls rapidly and narrows to a dozen or so very quickly.

By the time it gets down to two or three of sites, the remaining sites are all very similar in terms of economics, and the search becomes as subjective as it is objective. At this stage, organizations may eliminate a prospective site based on perception, and previously unknown criteria may spring up as a way to differentiate between the remaining sites.

One thing is certain, once you make your announcement as to final selection, the game is over. There are no more concessions to be won and no more dealing to be done. It is important to keep two or three candidates actively involved until the very end.

22.3.2.3 Relative Costs and the Law of Diminishing Returns

Estimating all of the cost may be important to the funding decision and may be developed as part of the overall project strategy, but as long as there isn't a difference between sites, it's not critical to the decision. For the purpose of site selection, it is only necessary to compare costs that differ between sites. For example, if license fees for operating systems and applications are the same regardless of where deployed, then this is not a cost that needs to be evaluated as part of the site selection process.

Site search and selection involves many forecasts of future events: how much power you will need, how fast you will grow, how many people you will employ, how the political and economic climates will change over time, etc. There is a cost to the site search and selection process, and at some point the cost of obtaining more information, performing more analysis, and perfecting the estimates starts to exceed any benefits that may be derived from the effort.

In the end, worrying about costs that aren't significant or key to making the site selection decision and overanalyzing the costs tend to only delay the decision. There may be a clear choice, but more often than not, there will be two or three sites that are all well suited, and if you have done a good job, all very similar. At this point, additional analysis is not likely to improve your decision, and it is time to bring the process to a close.

22.4 THE SITE SELECTION PROCESS

Site search and selection can be boiled down to several key activities as follows (Fig. 1):

- Develop business requirements and constraints.
- Define geographic search area.
- Define site selection criteria.
- Screen opportunities.
- Evaluate "the short list."
- Close the deal.

22.4.1 Develop Business Requirements and Constraints

The first activity of a site search is to define the business requirements. Typical motivations for conducting a site search may include supporting growth, reducing cost, expanding into new markets, etc. Requirements are the items that we believe are necessary to conduct business including space, power, cooling, communications, etc. Some might add "in a profitable manner," while others might argue that profitability is not part of a requirement.

The industry has created its own rules and vernacular that include requirements, constraints, criteria, etc. Don't get hung up on the subtle difference between these terms. What you need to end up with is a set of statements that define what results are wanted and by whom segregated into "must have" (deal breakers) and "must have if available and affordable" (negotiable), a ranking by priority, and finally rough expectations of cost and practicality.

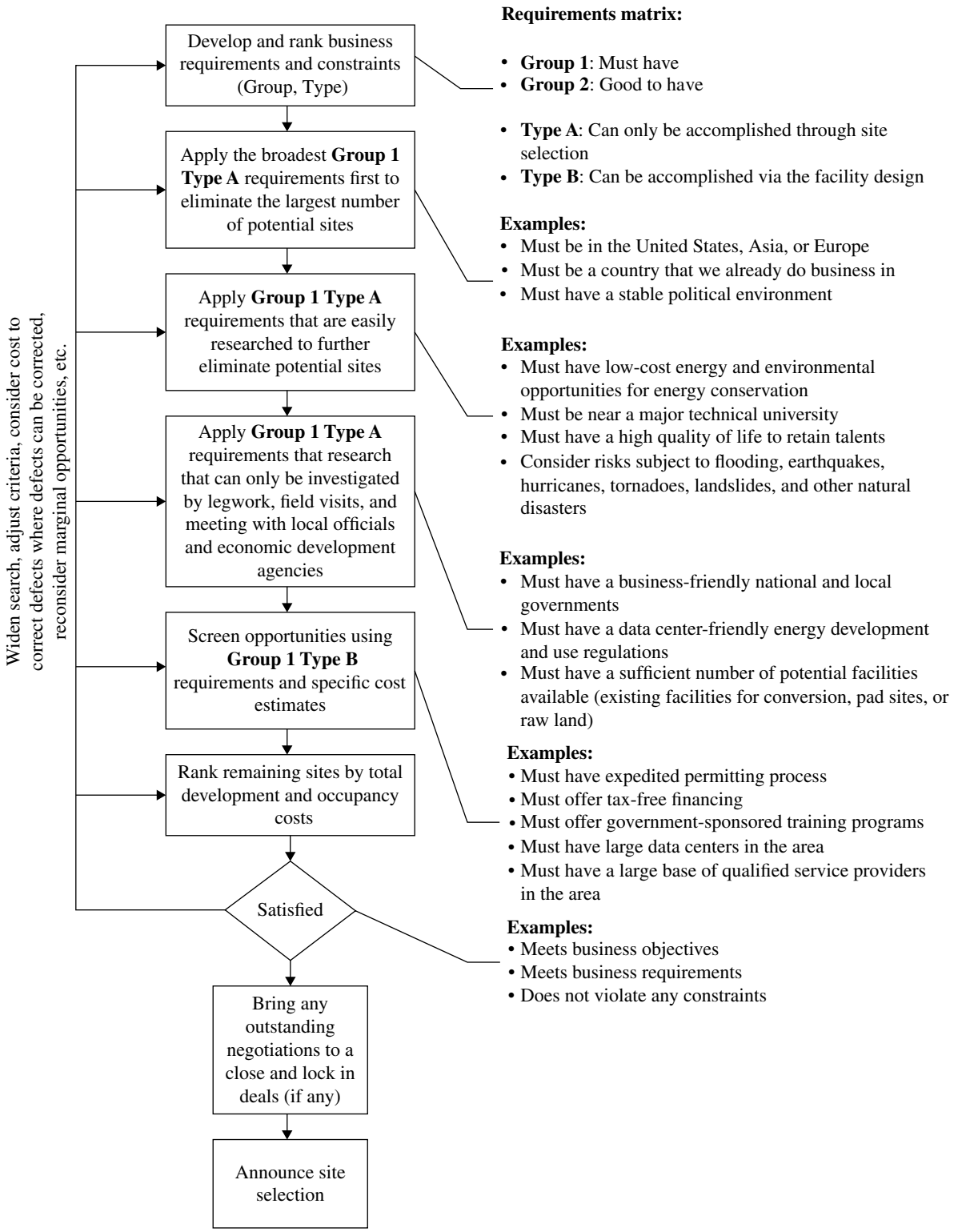


FIGURE 22.1 Site search and selection process. Source: Courtesy of K.J. Baudry, Inc.

By definition, you would think that a "requirement" must be satisfied. But reality is that requirements come from a collection of stakeholders with different motivations, fears, and concerns. Some are more perceived than real. They often conflict with one another, and some are more important than others. As part of cataloging and ranking the requirements, you will need to resolve the conflicts.

You can catalog and track this kind of information in a spreadsheet, database, or an intranet Web-based collaboration tool. There are some specialized "requirement management" type products available as well. How you do this is not important, but it is important that it is well done and documented.

In all probability, by the time an organization funds a site search, they have a pretty good idea of what they want to accomplish and have something akin to a requirement list. Key to being successful will be flushing out this list, eliminating requirements that are not supportable by the business, adding requirements that are missing, and building a consensus among the decision makers.

A good place to start is to look at why you are searching for a new site and establishing if expectations of the benefits that will be gained from the endeavor are reasonable. Even if the reasons for the search seem obvious, there are always alternatives, and all too often, when the final cost estimate is put together, the program changes. An organization may elect to refurbish an existing data center rather than relocate, may decide that a new data center cannot be supported by the customer base (sales revenues), or may find that despite tall tales of cheap electrical power deals, cheap power is a fleeting opportunity at best.

This first high-level pass should be based on what we know today and on any additional information that can be gathered in a timely and inexpensive manner. If the key business requirement is to reduce operating expenses and if the belief is that cheaper power is available elsewhere, then a purchased power survey/report and industry average construction costs might be enough for a first-round estimate of the potential annual energy savings and a potential payback. This effort might confirm, change, or disprove the savings expectation and reason for the site search, saving the organization countless hours and expense.

Each requirement and expectation should be challenged in this manner, as well as assumptions about executive and business unit support. Don't assume that what is in the interest of one business unit is in the best interest of all business units. Spend some time and effort confirming the initial project program with the stakeholders (i.e., anyone who has a credible interest in the outcome). This may include senior management, business units that will use the facility, customers, consultants, etc. The goal is not to kill the site search before it starts, but a search that starts badly usually ends up badly.

It is surprising, and all too frequent, that halfway through the site search selection process, the question of what the organization really needs or wants gets raised as it was never suitably addressed in the beginning. This kind of questioning down the home stretch will lead to delays and missed opportunities. It can also cause confusion among your stakeholders and with the local and economic development leaders that are after your business.

The key here is "plausible." Beginning the search with too specific or unrealistic criteria risks eliminating potential sites and tends to get the search bogged down with detail that often isn't available until further along in the process. On the other hand, beginning the search with too few or incomplete criteria wastes a lot of everyone's time. When you don't sound like you know what you are doing, people are less responsive and don't offer their best.

No matter how hard you try to build a great team (Table 22.1), you will have some team members that are not

TABLE 22.1 Typical search team members and criteria

Typical site selection team members:

- **Program manager**
- **Stakeholders:**
  - Customers
  - Employees
  - Stockholders
  - Business unit manager
  - C-level sponsors
- **Consultants and vendors:**
  - Site selection consultant
  - Architect and engineer(s)
  - Power and energy procurement broker
  - Network and communication specialist
  - Analyst-cost accountant
  - Tax accountant
  - Attorney
  - Construction contractor
  - Data center move consultant
  - Human resource specialist

Typical site selection considerations:

- **Geopolitical:**
  - Stable government and economy
  - Business-friendly government
  - Favorable tax rates and policies
  - Favorable energy production and use regulations
- **Infrastructure:**
  - Fiber networks to global telecom connections [1]
  - Power capacity and reliability
  - Water and sewage systems
  - Ease of travel (roads and airports)
  - Municipal services such as police and security, fire protection, medical, and health care
  - Availability of business and support services such as construction contractors and preventative maintenance vendors

TABLE 22.1 (Continued)

- **Operating expenses:**
  - Low energy cost
  - Environmental opportunities for energy savings
  - Low property and income taxes
  - Government-funded training programs
- **Low risk of local and regional disaster:**
  - Not subject to natural disasters: hurricane, tornado, monsoons, flooding, earthquakes, landslides, wild fires, etc.
  - Proximity to transportation arteries (rail, highway, and air)
  - Proximity to petrochemical plants
  - Proximity to nuclear plants
- **Quality of life:**
  - Low cost of living and home prices
  - Short commute times
  - Employment and educational opportunities for spouses and children
  - A climate that provides residents a year-round playground with plenty to do—mountains, lakes, rivers, and beaches
  - Entertainment including major league sports and theater
  - World-class cultural attractions
  - Exciting nightlife and convenience of travel

Source: Courtesy of K.J. Baudry, Inc.

of your choosing, some of whom will be uncooperative and will present site selection criteria that ensure the search will not be successful. It is important that inappropriate criteria do not make it to your site selection criteria list. Also, once you begin the search process, it is not uncommon to find that a certain requirement is not easily met or that it is prohibitively expensive to meet, and there may be other reasons to redress the requirements.

Be cautious and exercise due diligence when a change is imitated, but keep in mind that the goal is not to adhere to a statement of requirements at all cost but to make the right decision for the business. Business conditions do change, and the original requirements may have been misdirected in the first place or are perhaps no longer appropriate.

22.4.2 Round 1: Define Geographic Search Area

Geographic area is generally the first requirement, or a constraint, that the search team will use to narrow the universe of possibilities. Why? First, because the process is one of elimination and geographic area generally eliminates more potential sites than any other constraints. Second, the research and effort required are practically nil. Connecting these two points, there is simply more bang for the buck with this approach.

The geographical search area can be anything reasonable. Keep in mind that there are often legitimate requirements or constraints that may not be directly cost related and aren't easily estimated or accounted for:

- A power company or other regulated utility might define their search area as their service territory because locating outside their territory might be considered both politically incorrect and akin to suicide with the public service commission or other regulators.
- A bank with sensitive customer credit data might constrain their search to the locations in their home country based on regulatory requirements that certain data be maintained in-country.
- An international business such as Coca-Cola or Home Depot might look to any country where sales are substantial and contribute to their annual income, in an effort to give something back to the community.
- A regional organization that will relocate a number of employees to the new site may limit the search to cities in their region with a population between 1,000,000 and 5,000,000, in an effort to make the relocation more palatable to their existing employees.
- A business in a very competitive environment where being the lowest cost provider is key to winning business might not have any geographic limitation. It simply comes down to the least expensive site available.

Geographic constraints carry a lot of risk if the impact is not fully thought out. Many organizations set geographic constraints based on proximity to existing facilities, the assumption being that there are practical benefits (cost savings) to be gained. However, there is little basis for this assumption. Any operational support or convenience that might be available from the headquarters, office, warehouse, or other facilities in the immediate vicinity is often insignificant when compared with other savings such as energy cost and tax incentives that might be available outside the immediate vicinity. There also may be existing facilities in locations outside of the immediate vicinity that are available because of an unrelated merger or acquisition, bankruptcy, etc. that might be leased or purchased at substantial discount to the cost of new construction.

Once you have the search area, you will want to develop a list of potential communities. There aren't a magical number of prospective communities. Given the costs involved, more than 20 opportunities are probably excessive. Your site search consultant may suggest eliminating some communities based on relevant experience in dealing with them. With fewer than a dozen, you might not end up with any acceptable candidates as you cut the list down, applying more detailed criteria.

22.4.3 Round 2: Site Selection Criteria

We have opted to present this information as taking place in a serial manner for ease of presentation. Depending on your schedule and how committed (willing to spend money) the organization is, there is some benefit to developing the statement of requirements and selection criteria as a contiguous event in advance of starting the search. With this in mind, selection criteria that require little research might be incorporated as part of round one. The idea being to eliminate as many possibilities before the in-depth (expensive) research is required.

The following sections identify concerns that are common to many data center users when searching for a site. The list is not comprehensive but covers the major areas. Some of the items may not be applicable, depending on the type of organization and industry which you operate in.

The idea is that for each of these items you assess how they affect your organization and if they are potentially impacting identify how they can best be dealt with.

22.4.3.1 Political Environment

Most national governments, and especially emerging economies, have reduced barriers to market entry, property ownership, and deregulated privatized industries and encourage capital investment. When looking overseas, the chances are that you will receive a warm welcome. However, there may be significant challenges in the following areas:

- Security
- Laws
- Regulatory
- Employment
- Property ownership and investment
- Censorship

22.4.3.2 Quality of Life for Employees

Data centers can operate with a few employees. In many cases, quality-of-life issues might not be important, as employees will be hired locally. However, if an organization plans on relocating employees, quality-of-life issues will become important in retaining employees. What are people looking for? Some combination of the following:

- Low home price, taxes, and energy cost
- Short commute times
- Employment and educational opportunities for spouses and children
- An environment that provides residents a year-round playground with plenty to do—mountains, lakes, rivers, and beaches
- Entertainment including major league sports and theaters
- World-class cultural attractions
- Exciting nightlife
- Convenience of travel etc.

There are numerous publications such as *Forbes*, *Business Week*, and *Time Magazine* that create "top ten" lists. Best cities for college graduates, best cities for technology, best cities for entrepreneurs, most high-tech employment, etc. make good sources for this kind of information.

22.4.3.3 Business Environment

Site selection is largely about taxes and other costs that are a large part of operating expenses. Taxes come in all shapes and sizes, vary in how they are calculated, and are applied from state to state. Historically, some communities have looked at the economic benefits that data centers provide in terms of direct and indirect jobs created (payroll), investment, and taxes and have identified data centers as good sources of tax revenue. Data centers pay a lot in taxes and demand very little in terms of publicly provided services. They do not use significant amounts of sewage capacity, generate trash, fill up the local schools, and don't require new roads or extra police services. For a politician, it's a significant source of new "unencumbered" income.

The largest single category of taxes tends to be property taxes. In many communities, both the value of the real property, land, and buildings and the value of the fixtures, furnishings, and equipment are taxed. When you consider the cost of the facility is often well over $1,000/sf and that the IT equipment installed can easily exceed this figure, even a low tax rate can result in a significant annual outlay.

Local communities typically have offered incentives in terms of tax abatement on property taxes and reduced sales taxes on equipment purchased for installation in the facility. Programs will vary, but almost all communities phase the incentives out over a period of years. Other state and local incentives may include an expedited permit process, land grants, improved road access, extension of utilities, tax rebates, and financing through industrial revenue bonds. There may also be community development zones and associated development grants. Many states offer job credits and training programs through local community colleges.

You will need to calculate the economic benefit of the incentive package and include it in your overall analysis. A word of caution: a significant percentage of incentives, well over 50%, are never collected due to failure on the part of the organization to follow through after the facility is opened or due to changed economic conditions, overenthusiastic rates of growth, etc. For example, a delay in growth that pushes large

purchases of equipment beyond the first couple of years could result in significant reduction in tax savings if the incentive was highest in the first year and fell off over successive years.

When considering overseas locations, there are differences in the way taxes are levied, and it will affect your cost structure. This is one of the reasons for having an accountant on your team who can evaluate the implications for your economic modeling.

Politicians, local officials, and economic development groups love to make a splash in the news headlines by announcing big deals, and data centers tend to be big deals in terms of dollars invested. But it is a two-way street, and it will be up to you to show them the value that you bring to the community. In short, the better you sell yourself, the more successful you will be at winning incentives.

22.4.3.4 Infrastructure and Services

While airports, roads, water, sewage, and other utilities are all important, two utilities are showstoppers. A data center must have electrical power and telecommunications.

Telecommunications engineers often represent networks as a cloud with an access circuit leading in/out of the cloud at the "A end" and in/out of the cloud at the "Z end." It's a bit of a simplification but is a good representation. The in/out circuits are called end or tail circuits and are provided by a local exchange carrier (LEC). They typically run to an exchange where traffic can be passed to long-haul carriers.

Depending on how you purchase bandwidth, you may make a single purchase and get a single bill, but it is very likely that your traffic is carried over circuits owned by several different carriers. Having more than one carrier available means that there is always a competitive alternate carrier who wants your business. At one time, network connections were priced on bandwidth miles, but today, supply and demand, number and strength of competitors in the local market, and available capacity all factor into the cost of bandwidth. The only way to compare bandwidth cost between site options is to solicit proposals.

While small users may be able to use copper circuits, T1s and T3s, many organizations will require fiber medium services, or the latest carrier grade Ethernet. LECs typically build fiber networks using self-healing networks, such as synchronous digital hierarchy (SDH) or synchronous optical network (SONET) rings. These are very reliable. Typically arrangements can be made to exchange traffic to long-haul networks at more than one exchange making the system very reliable and resilient.

Many organizations require that there be two LECs available. Reasons for this include reliability, pricing, and perhaps a lack of understanding about networks. In general, site search and selection tends to be initiated and led by the financial side of the organization, and most organizations do not involve their network engineers in the site search and selection process. However, network connectivity is one of the fundamental requirements of a data center. It is a key cost and performance issue, and a network engineer should be considered a critical part of one's team.

Electrical power is the other big piece of infrastructure that is a must have. Smaller data centers with less than 5 MW of load can generally be accommodated in most large industrial and office parks where three-phase service exists. Larger data centers require significant amount of power and may require planning with the power company and the upgrading of distribution lines and substations and, in some cases, construction of dedicated substations. All of this must be addressed with the power company prior to the site selection.

Energy can make up as much as a third of the total occupancy cost and more if the rates are high. Almost all power companies use the same basic formula for setting rates: recover the capital cost of serving the customer, recover the cost to produce the energy, and make a reasonable return for the stockholders. Understanding this is important to negotiating the best rates. Yes, most utilities can negotiate rates even if they are subject to public service commission regulation, and regulated utilities can be every bit as competitive as nonregulated utilities!

If you want to get the best rate, you need to know your load profile, and you need to share it with the power company. Your design may be 200 W/ft^2, and that is great, but that's not your load profile. Your load profile has to do with how much energy you will actually consume and the rate at which you will consume it. Retaining a consultant who understands rate tariffs and can analyze your load is important to negotiating the best rates.

Negotiating rates is best done in a cooperative manner, sharing information as opposed to the more traditional negotiation stance of sharing little, demanding a lot, and constantly threatening to take your business elsewhere. Demanding oversized service only results in the utility spending more money to serve you, and more money has to be recovered from you in upfront capital contribution or through higher rates; therefore, properly sizing the service results in the most competitive rates. Further, most utility companies, either because of regulation or because of board governance, will not knowingly provide you with below cost rates and then make it up by charging another customer higher rates.

Climate is not part of the infrastructure but makes the checklist twice: first, as a factor that affects your energy cost and second as a quality-of-life issue. Your energy cost is dependent on the rate, but it is also dependent on how much energy you use. Recent thinking has led air-conditioning engineers to consider ways of reducing cooling costs such as air-side and water-side economization. The potential savings are greatly increased when the outside air is cool and dry for substantial parts of the year.

However, there is more that should be considered than just cost; the capability of the utility and responsiveness when and if an emergency should occur is also important. A short

example is probably worth more than a dozen paragraphs. In September 2005, Hurricane Katrina devastated the Mississippi coast. Mississippi Power, a small operating unit of Southern Company with 1,250 employees, restored power to almost 90% of their customers within 11 days (10% were too decimated to receive power). They rebuilt 300 transmission towers, over 8,000 poles, and 1,000 miles of overhead lines against all odds by bringing in a workforce of over 10,000, providing temporary housing in large circus tents, food, and water; over 8,000 tetanus shots; and burning 140,000 gallons of diesel fuel a day. A small municipal utility, affiliated with a regional utility, may not be able to come through when the going gets tough. The preparation and logistics necessary to maintain operations during a disaster get exponentially more difficult as the duration becomes longer.

There are three concepts in business continuity that need to be considered in the site selection process: walk to, drive to, and fly to. In major regional outages such as earthquakes and hurricanes and after 9/11, it became apparent very quickly that accessing to a primary or fall back site can be challenging if not impossible. Roads become blocked with traffic or debris, and even air travel may become curtailed. If your business continuity planning requires that your data center continues operating in an emergency, then it becomes important to have multiple means of transportation.

It is a bit unusual to think of maintenance vendors as part of the local infrastructure. However, the role of preventive maintenance is of as much importance as it is to have redundant systems, and perhaps even more important. While it is important that a vendor know everything there is to know about their equipment, they must also be familiar with and have the discipline to work in data center environments. If you are the only data center in the region, you may not find suitable maintenance vendors locally. If qualified maintenance vendors are not available, you will need to consider how this will affect your normal maintenance program as well as your ability to recover from a failure in a timely manner. Operator errors, including errors made by maintenance vendors, account for a significant percentage of all failures. Having other significant data center operations in the area is a good indication that maintenance vendors are available.

Finally, if there are other data centers in the area, ask the local economic development authority to introduce you. The odds are that they played some role in their site selection decision and will already have connections within the organization. Buying one of these contact's lunch is probably the best investment you can make in the search process.

22.4.3.5 Real Estate and Construction Opportunities

Depending on your requirements, you may be looking for an existing facility to purchase or lease, for pad sites within a development, or for raw land. Regardless of the need, you will want to make sure that there are a sufficient number of opportunities.

During the dot-com boom, a new breed of business, co-location, was created. Aiming at the outsourcing of information technology needs by corporate America, and armed with easy money, these companies built large, state-of-the-art facilities across the country, some at a cost of $125 M or more and as large as 300,000 ft^2. Many of these companies failed, some because of a poor business plans and others because they were simply ahead of their time. These facilities were placed on the market at substantial discounts to construction cost. Some were practically given away.

Today, these types of opportunities are rare, but occasionally companies merge or change computing strategies to find that they now have excess data center facilities. Other facilities become available because of lease expirations, growth, changing technologies, etc. While these may be great opportunities for some companies, they are often based on outdated standards and were built prior to the current focus on energy costs. It may not be necessary that a facility meet "current standards," but it must meet your standards and needs. Great care should be taken to properly assess the opportunity and cost to upgrade the facility if necessary.

It is not uncommon to find industrial or mixed-use parks advertising themselves as a "data center opportunity." This may mean that the developer has already worked with the local power utility, brought multiple fiber carriers to the park, negotiated for tax abatement, and taken other steps to prepare the site for a new data center. However, it often doesn't signify anything more than someone could build a data center on this site if they choose to.

When selecting a community, it is important that there are multiple sites available. These sites should be competitive (owed by different investors) and suitable for constructing a data center. It is important that your team verify any claims made by the owner, economic development agency, or other organizations trying to attract your business.

If there is a single existing facility or only one site available, then the negotiations with the various parties involved need to be tied together for a simultaneous close or with dependencies written into any purchase agreements so that the failure of any one part invalidates any other agreements.

Construction costs tend to be relatively uniform, and there are construction indexes readily available that can be used to estimate differences between communities or geographic regions. Development costs, permitting process, and construction requirements may vary between communities and, in some cases, might represent a significant impediment to your project, especially in terms of schedule. If you have a target date in mind, then finding a community that is willing to expedite the permit process may be important.

22.4.3.6 Geography, Forces of Nature, and Climate

Avoiding forces of nature is one area where there is often a discrepancy between what people say and do. This is due in part to a checklist mentality: the idea that by checking off a generic list of standard issues, we can avoid during our own legwork.

The general recommendation is to avoid risk due to forces of nature. Yet, California is home to a large number of data centers and continues to capture its share of growth despite an elevated risk of seismic activity. Almost half of the continental United States falls within a 200 mph or greater wind speed rating and is subject to tornado activity. Yet these areas continue to see investment in new data center facilities. Finally, East Coast areas with elevated risk of hurricane activity such as around New York and Washington, DC, continue to be desirable.

There are a couple of reasons why this happens. First, if there are compelling reasons to be located in a specific area, the risk may be substantially mitigated through good design practices and construction. Second, many individuals in key decision-making positions cling to the idea that they need to be able to see and touch their IT equipment. For these people, sometimes referred to as "server huggers," the idea of a remote lights-out operation entails more risk than the local forces of nature.

Regardless of the reason, it is important to assess the level of risk, identify steps to mitigate the risk, estimate the cost of risk mitigation, and include that cost in your financial modeling. Depending on whom you ask, natural disasters account for anywhere from 1% to almost 50% of all major data center outages. Another report puts power-related causes at 31%, weather and flooding (including broken water lines) at 36%, fire at 9%, and earthquake at 7%. A lot depends on the definition of a disaster, the size of data center under consideration, and what product or service is being promoted. The fact is that many data center outages are avoidable through good site selection, proper planning, design, and maintenance.

Design and construction techniques that effectively mitigate risk of damage from earthquakes, tornados, hurricanes, and flooding are well known and, while expensive, can be economically feasible when considered as part of the total cost.

While protecting your facility may be feasible, it is generally not possible to change the local utility systems and infrastructure. Many seemingly well-prepared organizations have found out the hard way that it is easy to end up being an island after a region-wide event such as a hurricane. An operating data center is worthless without outside communications and without its resupply chain in place (fuel, food, and water).

22.4.3.7 Man-made Risks

Avoiding man-made risks ranks up there with avoiding natural disaster, and the variance between what one asks for and what one does is even greater with man-made risks. Requirements that we often find on checklists include the following:

- Two miles from an airport
- Two miles from a broadcast facility
- Four miles from a major state highway or interstate highway
- Five miles from a railroad
- Ten miles from a chemical plant
- One hundred miles from a nuclear facility

Many of the items on this list appear obvious and reasonable on first review. However, trains and tankers full of hazardous material traverse our railroad tracks and highways every day of the week and at all hours of the day and night. So being a minimum distance from potential accidents makes sense. But most accidents are not contained within the highway right of way, and when you consider that released toxic chemicals can be transported by wind for miles, you realize that one-half mile, two miles, or four miles do not substantially change the risk.

Now consider where fiber and power are readily available. Utility providers built their networks where they expect customers to be, in industrial, mixed-use, and industrial parks (i.e., where there is commerce), and where there is commerce, there is man-made hazards.

It is important that all risks be considered and they should be considered from both short- and long-term points of view. It is even more important that the nature of the risk, relationship between the risk and distance, and the potential impact be understood.

22.4.4 The Short List: Analyzing and Eliminating Opportunities

The site search and selection process is built around elimination, and, to be most effective, we apply broad brush strokes in order to eliminate as many sites as possible. In doing this, we take a risk that we might eliminate some good opportunities. We eliminate some communities because they have significant shortcomings and others because they might have simply lacked the preparation and coordinated effort to provide a credible response to our request for information.

We are at a critical point in the process; we have eliminated most of the sites and are down to the short list. Each remaining community will need to be visited, the promotional claims and statements made during the initial rounds researched and supported, specific proposals requested, high-level costs estimates brought down to specific costs, and agreements negotiated. If there is a specific facility that figures into the site selection decision, you may want to have an architectural and engineering team evaluate the facility, create a test fit, identify the scope of work, and prepare a construction cost estimate (Table 22.2). This is a considerable effort for each site, and most organizations will want to limit the number of sites to three, the two front-runners and a replacement.

TABLE 22.2 Simple visual presentation of sites

Comparison of sites							
			Criteria ratings				
	Selection criteria	Weighting	Option A	Option B	Option C	Option D	Option E
1	Ability to meet schedule	4	2	5	2	4	1
2	Availability of property	3	3	5	3	3	1
3	Cost of property	3	3	4	4	3	1
4	Business environment—economic incentives	4	4	1	4	3	1
5	Availability of low-cost financing	3	4	1	4	3	1
6	Availability of power	5	4	3	5	4	4
7	Reliability of power system	4	4	4	5	4	4
8	Low-cost energy	5	4	3	5	4	4
9	Environmental energy savings opportunity	3	2	5	3	4	5
10	Availability of fiber networks	5	3	4	2	4	2
11	Availability of skilled vendors	3	2	4	1	4	2
12	Availability of IT labor	3	2	4	1	4	2
13	Easy to "fly to"	4	1	3	4	4	2
14	Easy to "drive to"	4	1	3	4	4	1
15	Proximity to major technical university	3	1	4	3	4	1
16	Local job opportunities for family members	3	1	3	3	4	0
17	Quality of life	4	1	4	3	4	1
18		0	0	0	0	0	0
19		0	0	0	0	0	0
20		0	0	0	0	0	0

Weighting—The default weighting is 2.5 and is the mean rating.
Rating—Ratings are high (5) to low (1).
Score—Score is calculated as "weighting × rating." For example, a weighting of 3 and a rating of 2, produces a score of 6 (3 × 2). The resulting score is an indication of how well a vendor meets the weighted selection criteria. The higher the score, the better the vendor meets the criteria.
Source: Courtesy of K.J. Baudry, Inc.

At this stage, it is inappropriate to eliminate a site without identifying a significant shortcoming. Earlier in the chapter, we stated that there was no such thing as the best site, just the best combination of site characteristics. In light of that idea, sites can be flawed but fixable or flawed and unfixable. In most cases, the reason for conducting an extensive site search is to minimize costs, and a site that has a fixable flaw, even if it is expensive to fix, might be the least expensive opportunity due to other line items. Consider a community without a competitive LEC. Could another carrier be attracted to the area because of your business? Could you negotiate a long-term agreement with the existing carrier to assure good service and pricing?

If you identify an unfixable flaw early enough in the process, it may be worthwhile eliminating the opportunity from the shortlist and moving one of the lower-ranked sites up to

the list. It may be that a promise or claim made by the development agency was not correct or turns out to be incomplete. This can and does occur because of deception, but for legitimate reasons as well. The community may have expected a bond referendum to pass that didn't; another company may step in front of you and contract for the available power; etc.

If a specific facility is part of the selection decision, some of the business requirements will need to be translated into technical or construction requirements. This can be a challenge. The bulk of our efforts will have been financial. This does not change with technical requirements but does take on a new complexion. It is not uncommon for the engineering members of your team to designate specific properties as unacceptable, especially when dealing with upgrading an existing data center or retrofitting an existing structure. In many cases, this will mean that the engineer perceives one or more aspects as too difficult or expensive.

It is important that questions of suitability, especially when they potentially remove a site from the list, be questioned and thoroughly reviewed. Often the cost, when taken into perspective, is not significant, or that the many cost savings features of the site easily offset the added cost. Not every member of your team will be fully aware of the overall cost and potential benefits that a community or facility brings to the program.

22.4.5 Closing the Deal

The process is complete when the site selection is announced. The site selection choice may be made long before it is announced and long before the process is complete, and it is important that at least two communities be kept in the running until all concession and agreements are in place and the deals are executed. The reason is very simple; you lose all negotiating power once all the other sites are eliminated. So despite the effort involved, it is important to continue to pursue your number one and two choices until the very end. Just like secrecy, the team members are the ones most likely to let out the decision through innocent conversations with power company employees, local consultants, permitting agencies, and other entities that are involved in your due diligence.

22.5 INDUSTRY TRENDS AFFECTING SITE SELECTION

22.5.1 Globalization and Political and Economic Reforms

Globalization has been a consistent force since the beginning of mankind. It has moved slowly at times and with great speed at other times. The past 30 years have seen great advancements. Trade barriers have fallen, and governments have sought out investment from other countries. We may be on the verge of a temporary slowdown as governments move to protect their turf during economic downturns.

Globalization will continue as long as there are opportunities to sell more products and services and as long as there are differences in the cost of key resources such as utilities, differences in energy policies, and taxes.

22.5.2 Global Strategic Locations

While the world keeps getting smaller and smaller, it is still a pretty big place. Every search needs a starting point. Traditional economic criteria such as per capita income, cost of living, and population might be applicable. Perhaps, the location where your competitor already has facilities. Or maybe the location already has major international IT companies such as Facebook, Google, HP, and IBM located.

Google has facilities in many U.S. cities as well as cities outside the United States. It is important to note that these locations are not all major data centers. Some may simply be peering points. For more information on Google Data Centers, see http://www.google.com/about/datacenters/locations/index.html#.

In addition to traditional criteria, we offer one more that might be worthwhile considering volume or other measures of IP exchange traffic.

The largest exchanges in the United States are the New York Area, Northern Virginia, Chicago, San Francisco Bay Area, Los Angeles Area, Dallas, Atlanta, and Miami.

The largest exchanges outside the United States include Frankfurt, Amsterdam, London, Moscow, and Tokyo. A full list of exchanges can be found on Wikipedia by searching on "peering" or "list of IP exchanges."

22.5.3 Future Data Centers

Today, we design and build data centers from a facility-centric approach. They are conceived, designed, and operated by facility groups, not by the real customer IT groups. Facilities design has had hard time keeping up with the rapidly changing IT environment. Our facilities designs have changed little in the past 20 years. Sure, we have adapted to the need for more power and cooling, and there have been incremental improvements. Energy efficiency has become important. But by and large, we have simply improved old designs incorporating time-proven techniques used in other types of facilities such as industrial buildings, schools, hospitals, and offices.

Today, every data center is a unique, one-off design to match an organization's unique requirements and local conditions. Tomorrow's data center will need to be designed with a holistic approach, one that marries advancements in IT technology and management with facilities design and life cycle economics.

One approach that we expect in the future will be a throwaway data center with minimal redundancy in the mechanical and electrical systems. A large appliance inexpensive enough

to be deployed at sites around the world is selected based on business-friendly governments and inexpensive power.

Equipped with self-healing management systems, meshed into a seamless operation, and populated with IT systems prior to deployment, these systems would serve their useful life, and the data center would be decommissioned.

With sufficient redundancy and diversity of sites, any one site could be off-line due to planned or unplanned maintenance. To take an existing data center off line and redeployed it can base on changing economic needs such as customer demand, short-term energy shortage, or fiber transmission services. At the end of life, it could be refurbished in the field, shipped back to a central depot for refurbishment, or sold to a low-end noncompeting organization.

The combination of the reduction in initial facility development cost, mass production to a consistent standard and performance level, speed of deployment, and the flexibility to meet changing economic environments is a very attractive package. The ability to implement this approach largely already exists for the most part, considering the following:

- Lights-out operations already exist. Management systems have progressed to a point where most operations can be performed remotely. They allow applications to exist at multiple locations for production and backup purposes, for co-production, and for load balancing or to meet performance objectives.
- IT platforms are getting smaller and smaller. Servers have progressed from several rack units (RUs) to one RU and blades. You can fit as many as 70 blades servers in a single rack. Processing capacity is almost doubling every 3 years, and virtualization will greatly increase the utilization of the processing capacity.
- While the amount of data we store today has increased tremendously, the density of storage systems has increased tremendously and will continue to do so. In 1956, IBM introduced the random access method of accounting and control (RAMAC) with a density of 2,000 bits/in^2 Today's latest technology records at 1.34 Tbits/in^2 (Seagate) more than10^{12} in that's in 60 years.

A containerized data system could be configured to provide the same level of computing that is provided in today's average data center of perhaps 10–20 times its size.

There are advantages from the facilities perspective as well. Data centers are built for a specific purpose. We struggle to increase the availability of the physical plant, and every improvement comes at an ever-increasing expense (the law of diminishing returns). No matter how much redundancy we design into the facilities plant, we cannot make one 100% available. By comparison, inexpensive data center appliances with dynamic load balancing and redundant capacity deployed throughout the world would not need five nines of availability to be successful.

According to the Uptime Institute, the cost of a Tier III facility is generally believed to be $200/sf and $10,000/kW or higher. While the building does depreciate, it's the physical plant (the $10,000/kW) that becomes obsolete over time. In a manner of thinking, we are already building throwaway data centers, at a great expense and without a good exit strategy!

There has always been a mismatch between the investment term for the facility and IT systems that the facility houses. IT refresh cycles are typically 3 years (probably more likely 5 years for most organizations), yet facilities are expected to last 15 years or more. This mismatch between investment terms means that data centers have to be designed to accommodate an unknown future. This is an expensive approach.

While such a concept will not be suitable for companies that are the largest of centers, it could be very beneficial for smaller operators, the very clients that collocation centers are seeking to put in their centers today.

ACKNOWLEDGMENT

This chapter is adapted from the first edition of Data Center Handbook and has been updated by its Technical Advisory Board. Sincere thanks are extended to Mr. Ken Baudry and the members of the Technical Advisory Board who spent invaluable time in sharing their in-depth knowledge and experience in preparing this updated version.

REFERENCE

[1] Submarine cable map, TeleGeography. https://blog.telegeography.com/2020-submarine-cable-map (accessed 10/22/2020)

FURTHER READING

Best Practices Guide for Energy Efficient Data Center Design, Office of Energy Efficiency and Renewable Energy (EERE), Department of Energy. https://www.energy.gov/eere/about-office-energy-efficiency-and-renewable-energy. Accessed 10/22/2020.

National Oceanic and Atmospheric Administration. Available at http://www.nhc.noaa.gov/. Accessed on October 22, 2020.

National Weather Service. Available at http://www.weather.gov/. Accessed on October 22, 2020.

Rath J. Data Center Site Selection, Rath Consulting. Available at http://rath-family.com/rc/DC_Site_Selection.pdf. Accessed 10/22/2020.

The Uptime Institute. Available at http://uptimeinstitute.com/. Accessed on October 22, 2020.

U.S. Geological Survey. Available at http://www.usgs.gov/. Accessed on October 22, 2020.

23

ARCHITECTURE: DATA CENTER RACK FLOOR PLAN AND FACILITY LAYOUT DESIGN

PHIL ISAAK
Isaak Technologies Inc., Minneapolis, Minnesota, United States of America

23.1 INTRODUCTION

The success of the data center floor plan design process is dependent on the participation of the facilities design team and the information technology (IT) design team. The facilities design team will consist of the building architect, electrical, mechanical, and structural engineers. The IT design team will constist of those responsible for the network topology, server and storage platform architecture, and network cable plant design.

The design of the data center layout must be developed with input from all the key stakeholders from the facility and IT groups within the organization. This integrated design approach will help to ensure that the data center will function and perform throughout the facility life cycle, providing operational efficiency to support the many technology life cycles that the data center facility will see.

23.2 FIBER OPTIC NETWORK DESIGN

23.2.1 Fiber Optic Network Interfaces

Legacy fiber optic network interfaces of 10 GB or less have been based on serial optical interfaces, requiring two fiber strands for each interface port. The first 40 GB fiber optic interfaces available were based on parallel optics that teamed four 10 GB optics into a single interface that required eight active fiber optic strands. This was the lowest cost solution to provide the 40 GB interface (at least as far as the cost of the fiber optical interfaces was concerned). As the industry has evolved with faster 100, 200, and 400 GB network interfaces, the number of strands required to support the newer network interfaces has also evolved between parallel and serial interfaces. This constantly changing (or at least it can seem like it's constantly changing) number of strands to support any given throughput places challenges on the data center designer. In addition to the total quantity of strands that fiber optic trunk cables must support, these changing requirements also need to be accommodated by the following:

- Total rack units (RU) available for fiber optic shelves terminating the fiber strands (parallel optics with increased quantity of fiber strands requires more RU's).
- Total capacity available within the fiber optic pathways (parallel optics with increased quantity of fiber strands requires more cross-sectional area).

The number of fiber optic strands per port is not only simply a function of the interface throughput (i.e., 40G, 100G, 200G, or 400G) but also a function of the specific interface technology used within the network equipment (Tables 23.1 and 23.2).

As network speeds continue to evolve to provide faster throughput, the number of strands per port may continue to cycle between increased quantity of strands to support parallel optics and reduced quantity of strands upon technology advancements. The initial solution for increased network throughput is to implement parallel optics (typically lowest initial cost with respect to latest fiber optic interfaces), requiring additional fiber optic strands. This is followed by cost-effective technology advancements (i.e., coding schemes, multiplexing) that provide the increased throughput with fewer strands. This will likely continue to be a recurring cycle with respect to the quantity of fiber strands required to

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

TABLE 23.1 Multimode fiber interfaces

Multimode optical interface	Fiber strands
10GBASE-S	2
40GBASE-SR4	8
40G BiDi	2
100GBASE-SR10	20
100GBASE-SR4	8
100G BiDi	2
200GBASE-SR4	8
400GBASE-SR16	32
400GBASE-SR8	16
400GBASE-SR4.2	8

Source: Courtesy of Isaak Technologies.

TABLE 23.2 Single-mode fiber interfaces

Single-mode optical interface	Fiber strands
50GBASE-FR	2
50GBASE-LR	2
100GBASE-DR	2
200GBASE-FR4	2
200GBASE-LR4	2
200GBASE-DR4	8
400GBASE-FR8	2
400GBASE-LR8	2
400GBASE-DR4	8

Source: Courtesy of Isaak Technologies.

support a single fiber optic port connecting devices, depending on the network throughput required and at what point the optical interfaces are procured in the life cycle of the technology for the specific throughput.

So how does this affect IT layouts and pathway design? A fiber optic cabling solution should be able to extend beyond a single technology life cycle of the fiber interfaces it is to support. That is to say, a 40 GB fiber optic cabling solution should certainly have been designed so that it can also migrate to support 100 GB (or higher) technology. The fiber optic pathways should be designed to support multiple life cycles of fiber optic cabling solutions. That is to say, a pathway that was designed for an OM4 fiber optic cabling solution should be able to support a migration to OM5 cabling.

The pathways that support fiber optic cabling should be sized to support the quantity of ports required to interconnect the maximum number of anticipated devices, and the number of fiber optic strands per port should be able to support between 2 and 64 fiber optic strands per port. The space available within the racks and cabinets to support fiber optic shelves should also be designed to accommodate varying number of strands per port throughout the life cycle of the facility. When identifying how much rack and cabinet space to allocate for fiber optic cabling network infrastructure, the designer should consider serial and high-density parallel fiber optic ports.

23.2.2 Very Large Computer Rooms

Very large data centers with computer room space totaling 9,300 m^2 (100,000 ft^2) or more may have unique challenges with respect to interconnecting devices without exceeding distance limitations. There are many ways to implement the compute, storage, and network systems within these very large data centers. There is no "one right way" to configure all the IT systems.

To illustrate this, we will look at two (of many) different compute, storage, and network configurations within a very large data center. The two configurations will implement a spine and leaf network architecture utilizing middle-of-row access switches as shown in Figure 23.1.

The first scenario will incorporate "pods" of infrastructure that contains all the necessary compute, storage, and network within each pod to support all the IT services that the pod is delivering to the customer(s). This is one method that is common within hyperscale data centers, one type of very large data center. As additional capacity is needed, additional pods are deployed. This method can simplify the coordination and building out of all the compute, storage, and network infrastructure over the very large data center computer room. The schematic shown in Figure 23.2 illustrates three data center computer rooms that are connected to two entrance rooms. Within each computer room, the data center pods are built out in five rows with compute, storage, and network contained within each pod.

The cabling infrastructure for this example includes fiber connectivity from the middle-of-row switch back to the core network for Ethernet traffic and to the SAN director for the Fibre Channel storage traffic. As shown in Figure 23.3, the channel with the most fiber optic connectors is between the middle-of-row switch and the spine switch (Ethernet Channel 2), with a total of six connectors. This meets the design limitations using the standards-based component channel method.

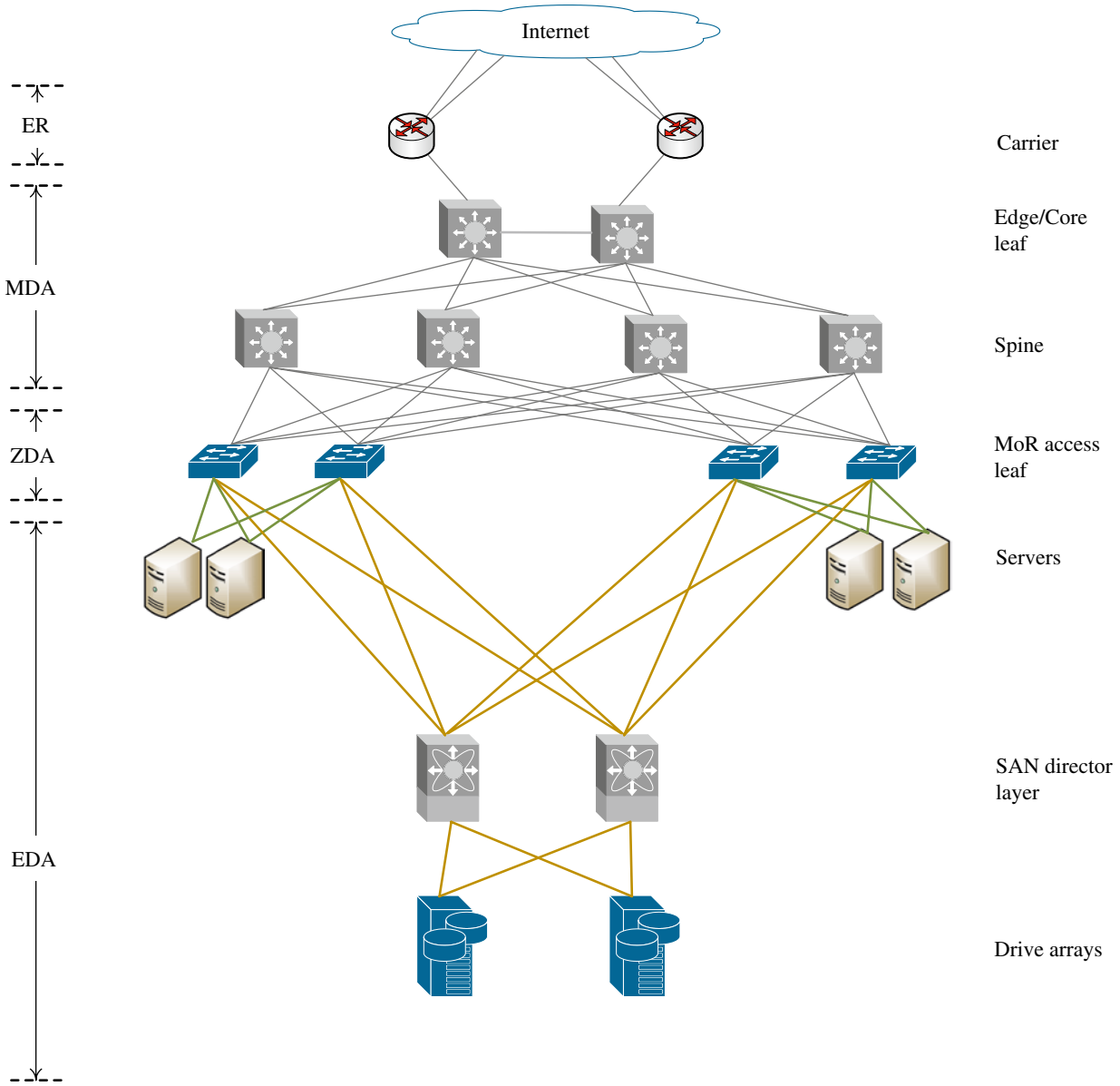


FIGURE 23.1 Example of network architecture. Source: Courtesy of Isaak Technologies.

The second example will incorporate middle-of-row switches with a centralized enterprise storage drive array for the entire data center. This method decouples the scalability of the compute from the scalability of the storage environments. The schematic shown in Figure 23.4 illustrates three data center computer rooms that are connected to two entrance rooms. The storage arrays are all located on the third level. The fiber optic infrastructure must provide connectivity from any compute system in any of the three computer rooms back to the SAN directors within the storage environment on the third level.

The cabling infrastructure for this example includes fiber connectivity from the middle-of-row switch back to the core network for Ethernet traffic and from each middle-of-row switch to the SAN directors located on the third level for the Fibre Channel storage traffic. As shown in Figure 23.5, the channel with the most fiber optic connectors is between the middle-of-row switch and the SAN director (Fibre Channel 3), with a total of eight connectors. This exceeds the design limitations using the standards-based component channel method. This example must ensure the design meets the limitations using the application-based method and calculating fiber loss budgets, which requires a detailed understanding of the applications, storage, and network configurations.

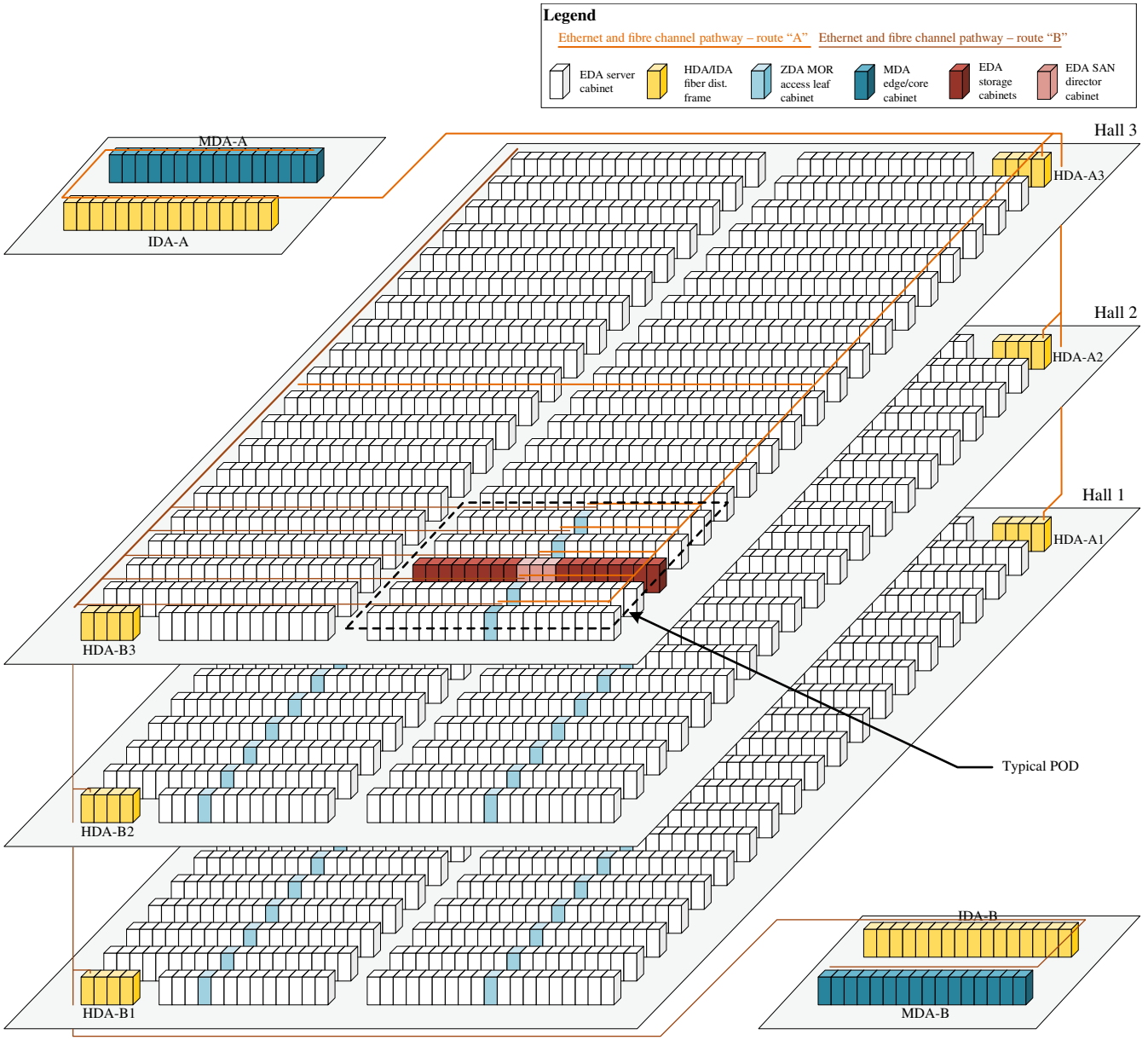


FIGURE 23.2 Example of very large data center computer room with compute, storage, and network configured in PODS. Source: Courtesy of Isaak Technologies. Source: Copyright Isaak Technologies, Inc.

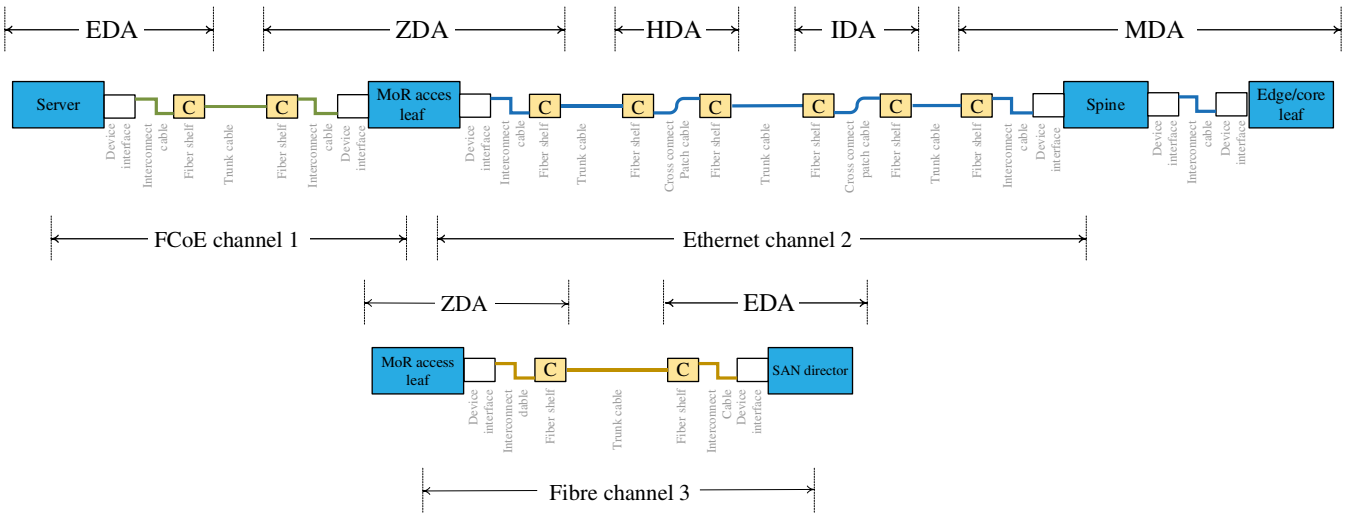


FIGURE 23.3 Example of fiber optic connectivity required within Ethernet and Fibre Channel connectivity for the very large data center implementing PODS. Source: Courtesy of Isaak Technologies. Source: Copyright Isaak Technologies, Inc.

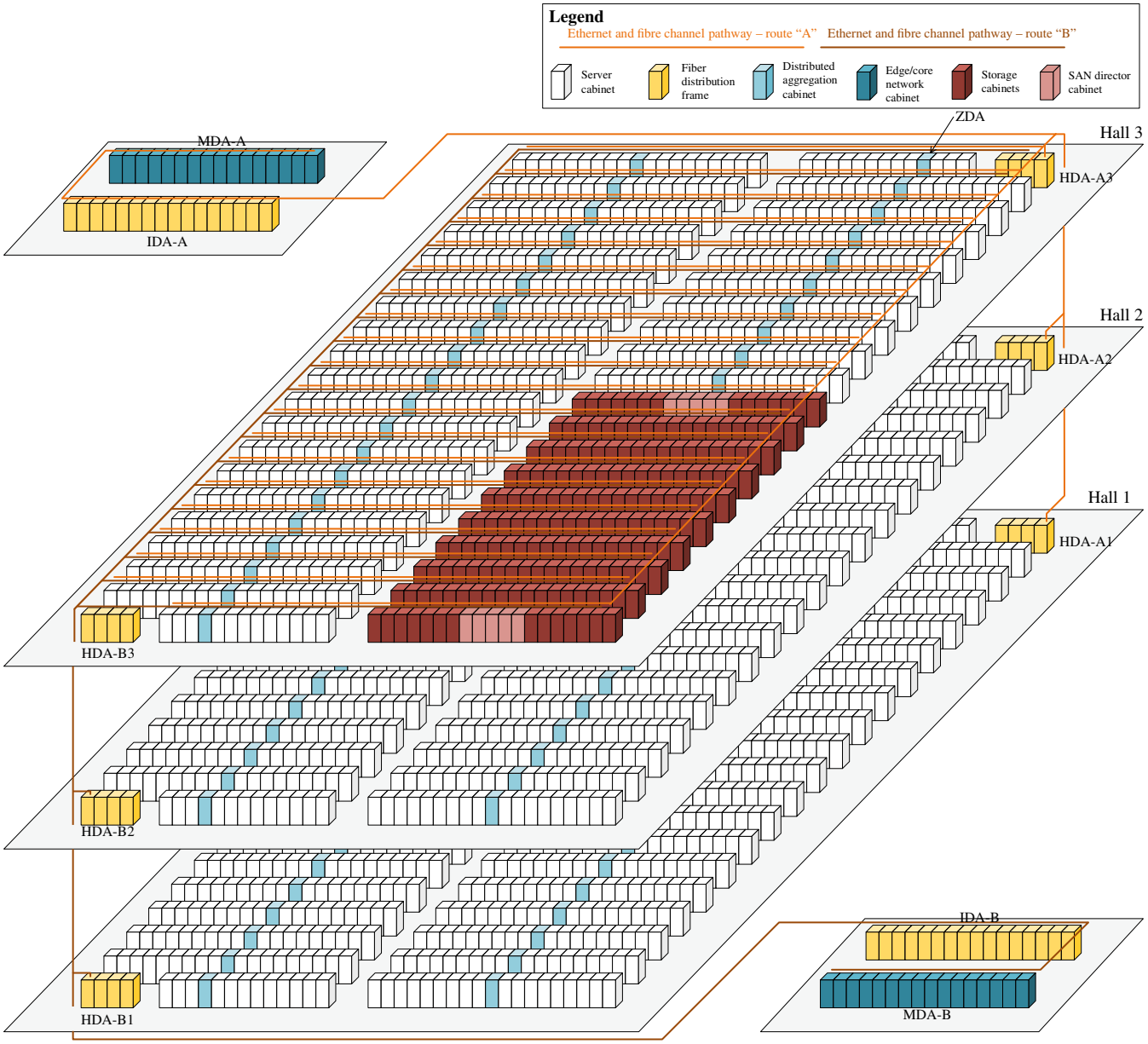


FIGURE 23.4 Example of very large data center computer room with centralized enterprise storage drive array on third level. Source: Courtesy of Isaak Technologies. Source: Copyright Isaak Technologies, Inc.

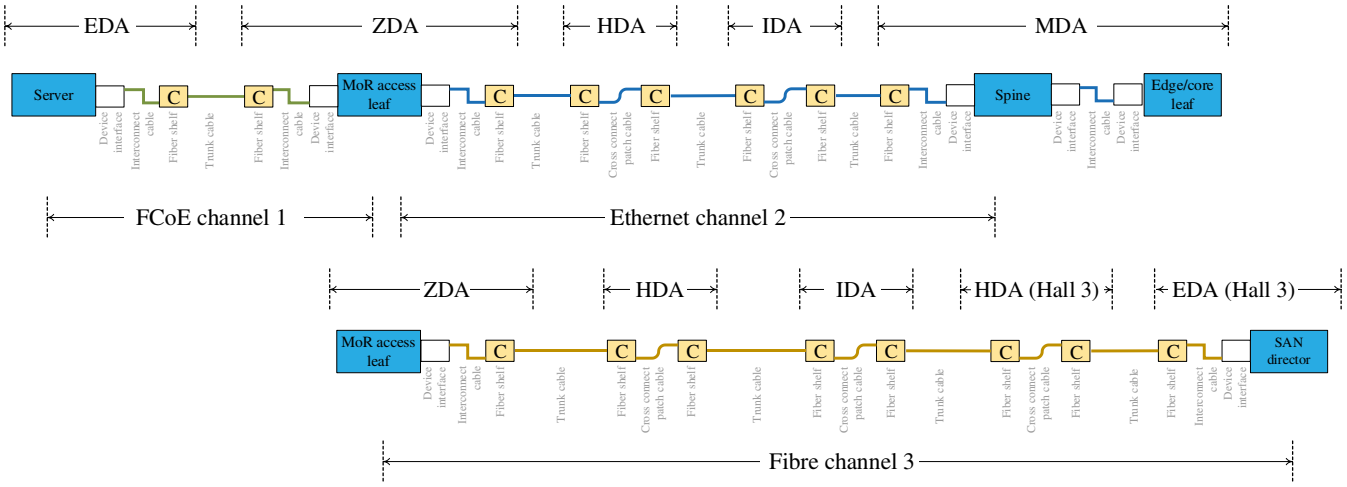


FIGURE 23.5 Example of fiber optic connectivity required within Ethernet and Fibre Channel connectivity for the very large data center implementing centralized enterprise storage drive array. Source: Courtesy of Isaak Technologies. Source: Copyright Isaak Technologies, Inc.

These are two of many configurations that are possible. These two examples are implemented within the same data center computer room, the same number of cabinets, the same spine and leaf architecture with middle-of-row switches, but end up with very different fiber optic cabling solutions. This is a good example of illustrating that the data center network infrastructure cannot be designed without thoroughly understanding the:

- Network architecture
- Network topology
- Compute architecture
- Storage architecture
- Cabling topology
- Floor plan layout of the compute and storage system

Each of the design considerations noted above needs to be analyzed and coordinated as a complete solution in order to identify the solution that will be cost effective and scalable and meet the redundancy requirements.

23.3 OVERVIEW OF RACK AND CABINET DESIGN

23.3.1 EIA/CEA-310-E Two- and Four-Post Racks

Two- and four-post racks are open frames with rail spacing that meet industry standards and are manufactured to EIA/CEA-310-E standards. Some data centers also utilize racks that are manufactured to the Open Compute Project requirements, which are discussed further in Section 23.3.3.

The mounting rail RU (1 RU = 1.75 in) should be clearly marked on all rails. The RU markings typically start at 1 on the bottom. However, there are some manufactures with products that start RU number 1 at the top. The RU designations should be consistently applied throughout the data center on two- and four-post racks as well as cabinets so there is no confusion for the technicians or the ability to integrate the ITE mounting positions with a data center information management (DCIM) application.

Two- and four-post racks should be provided with vertical cable management on either side of each rack, with sufficient capacity to accommodate the maximum number of patch cords or cabling infrastructure that is anticipated within the rack. The vertical cable management should have fingers at 1 RU increments to align with the ITE mounted within the rack. The depth of the vertical cable managers mounted on either side of the rack and the depth of the horizontal managers mounted within the rack should be coordinated so that they are in alignment, providing an even smooth consistent pathway throughout the cabling pathway provided for the rack.

Historically, two- or four-post racks were challenged with the ability to mount zero-U power outlet strips to the rack rails or frame. For this reason, horizontal power outlet strips are sometimes used for two- and four-post racks. Mounting brackets have become available from most rack manufacturers to fasten zero-U vertical power outlet strips to the two- or four-post rack.

The position of a row of racks first needs to be coordinated with the floor grid (if on a raised floor), any adjacent cabinets, and any overhead pathways. Refer to Section 23.4 for further guidance on coordination with pathways. Refer to Section 23.5 for further guidance on coordination with other systems.

The racks should be provided with a bonding point for bonding the rack to the data center grounding system. The bonding point should provide metal-to-metal contact without any paint or powder coating inhibiting the effectiveness of the bond. The resistance to true earth shall be either 5, 3, or 1 ohm measured by the fall of potential method (ANSI/IEEE Std 81) depending on the Class of data center per ANSI/BICSI 002.

The recommended methods of grounding racks and cabinets exceed the minimum requirements to meet the building codes. While the grounding requirements within building codes are provided for life safety, the grounding requirements in standards such as the ANSI/BICSI 002 Data Center Design and Implementation Best Practices standard and IEEE 1100 Recommended Practice for Powering and Grounding Electronic Equipment provide guidance for safety, noise control, and protection of sensitive electronic equipment.

23.3.1.1 Two-Post Racks

A two-post rack provides a single rail to which the ITE is mounted on. It is recommended that the ITE mounting brackets be set back from the front of the chassis so that the center of mass is positioned at the point the ITE brackets are installed. Large chassis ITE (either in RU or depth) may require a special shelf to adequately support the equipment.

Two-post racks have typically been used in network IDF/MDF closets where there is a single row of equipment. The two-post rack should only be used in the data center where space constraints limit the use of four-post racks or cabinets. ITE that are mounted in two-post racks are more susceptible to physical damage as the ITE is exposed beyond the rack frame.

23.3.1.2 Four-Post Racks

A four-post rack provides a front and back rail to which the ITE is mounted on. Manufacturers provide four-post rack models that offer fixed position front and back rails or for a fixed position front rail with a variable position back rail if required. The variable position back rail typically is adjustable as a single system from top to bottom, so all ITE mounted in the rack must accommodate the rail position selected.

Four-post racks are typically the preferred open frame solution as they provide greater physical protection for the ITE mounted within them than a two-post rack offers. For example, fiber enclosures mounted in a four-post rack will have the back of the enclosure within the footprint of the four-post frame. The fiber typically enters the fiber enclosure at the back. If the enclosure was installed in a two-post rack, the fiber entering the enclosure would be physically exposed, but in a four-post rack, it is within the four-post frame footprint.

23.3.2 EIA/CEA-310-E Cabinets

Cabinet are closed frames with rail spacing that should meet EIA/CEA-310-E manufacturing standards.

The mounting rail rack units (1 RU = 1.75 in) should be clearly marked on all rails.

The design and selection of cabinets often overlook the details that are required to ensure a suitable solution. The cabinet selection is based not only on the ability to mount ITE to the rails but also the ability to support the network cable plant entry from overhead or underfloor pathways, power outlet strip implementation, and airflow management becoming increasingly important. The cabinet selection requires the designer to be knowledgeable in hardware platforms, network cable plant, power distribution, and cooling airflow to be able to recommend the appropriate solution.

The cabinets should be provided with a bonding point for bonding the rails to the data center grounding system. The bonding point should provide metal-to-metal contact without any paint or powder coating inhibiting the effectiveness of the bond. The resistance to true earth shall be either 5, 3, or 1 ohm measured by the fall of potential method (ANSI/IEEE Std 81) depending on the Class of data center per ANSI/BICSI 002.

23.3.3 Open Compute Project Racks and Cabinets

The Open Compute Project developed a set of requirements for network, compute, and storage IT equipment along with a set of requirements for the racks and cabinets that the Open Compute Project IT equipment are mounted in. The Open Compute Project developed the requirements for the IT equipment to overcome specific challenges of traditional commercially available products based on a specific IT equipment deployment strategy.

One of the significant benefits of IT equipment manufactured per the Open Compute Project requirements is that the IT equipment have the power and network connections on the cold aisle side of the equipment. This is not always the case with traditional commercially available IT equipment, as they may have power and network connections on opposite sides of the equipment. The benefit that the Open Compute Project requirements provide is that in a hot aisle containment configuration, all of the supporting infrastructure including power and network cabling and connections are located in the cold aisle space. There is essentially no need for a technician to work within the hot aisle space for any day-to-day tasks in support of the IT equipment.

Open Compute Project racks or cabinets are required only if the data center will implement IT equipment based on the Open Compute Project requirements. The dimension between the mounting rails and the spacing of the mounting threads on the rails are unique to the Open Compute Project requirements.

23.3.4 Network

Two-post open frame racks are typically not used in the data center for the core network equipment due to the physical size and weight of the chassis. Four-post open frame racks may be used as they are able to support the size and weight and provide suitable cable management.

Open frame racks can be incorporated into a hot or cold aisle configuration; however, the specific containment solution must be coordinated with the IT equipment within the rack and the rack selection to ensure the solution will be effective in isolating the hot and cold air streams. Not all containment solutions work with open racks, so this must be part of the vendor selection process when evaluating containment solutions.

Manufacturers provide cabinet models that are purpose built to support server platforms or purpose built to support network platforms. It is not recommended to implement core network platforms within cabinets that have been purpose built for server platforms. Purpose-built network cabinets have options to support core network equipment with front-to-back airflow or side-to-side airflow within the same cabinet.

The physical requirement nuances between manufacturers and models of network equipment place an increased burden on the designer to ensure that the appropriate mounting frame is incorporated into the data center. Core network equipment will often consist of equipment with front-to-back and side-to-side airflow within the same cabinet. It is important to identify the appropriate airflow management solutions for the specific network platforms prior to finalizing the equipment elevation design and cabinet selection. When coordinating network equipment requirements with cabinet manufacturer options, the airflow management may require additional RU space above and/or below specific chassis to provide adequate airflow (i.e., Cisco 7009 platform), while others require external side cars fastened to the cabinet, increasing the width of the total cabinet solution, to provide adequate intake and exhaust airflow capacity (i.e., Cisco 7018).

The specific airflow path through a chassis must be validated prior to finalizing equipment layouts and cabinet selection. The following are some examples of various airflow management solutions, each requiring different approaches in cabinet design and equipment layouts (the port side of network switches is referred to as the front):

- Side-to-side/right-to-left airflow and front to back (Fig. 23.6)
  - Cisco 7004, 7009, 7018
- Front to back (Fig. 23.7)
  - Cisco 7010, 7700 series, 9500 series
  - HPE 12900E series
- Front to back or back to front
  - Cisco 9200 series, 9300 series
  - HPE 5700 series, 5900 series, 7900 series

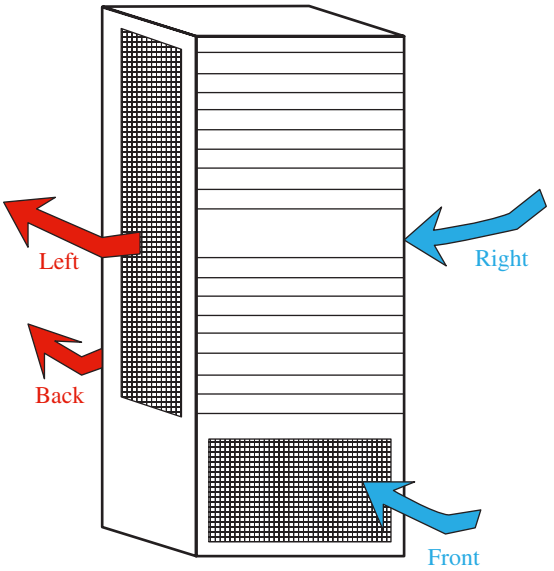


FIGURE 23.6 Network chassis with front-to-back and right-to-left airflow. Source: Copyright of Isaak Technologies.

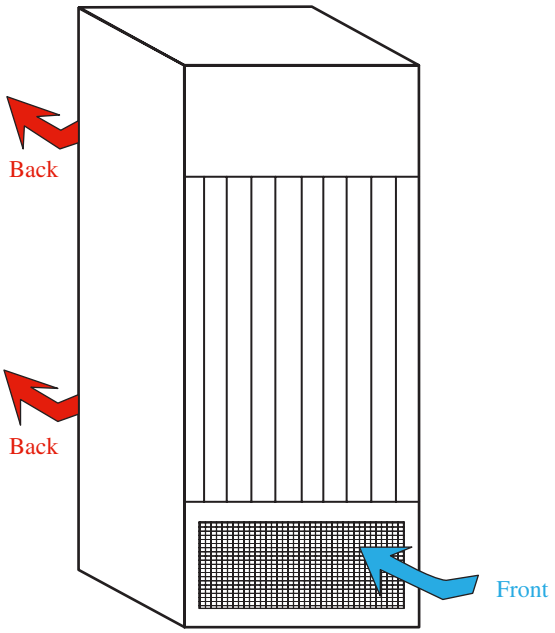


FIGURE 23.7 Network chassis with front-to-back airflow. Source: Courtesy of Isaak Technologies.

Cabinet manufacturers design their network cabinets to support front-to-back with options to support side-to-side airflow in the right to left direction with the port side of the network switch installed on the front rail (cold aisle). If network equipment has airflow patterns that differ from front-to-back or right-to-left, the selection, configuration, and layout of the network cabinets will need extra coordination to ensure the network chassis airflow pattern aligns with the hot and cold aisle layout within the room.

All associated fiber enclosures or copper patch panels providing interconnection to the network equipment should be installed so the port side aligns with the network equipment installation.

Network switches that are classified as top-of-rack or end-of-row switches often have the airflow through the switch from back to front. This is done to accommodate the port side of the switch (front) to face the same direction as the ports on the server, which are positioned on the back of the server. This enables airflow through the cabinet from the cold aisle to the hot aisle to support both the servers and the top-of-rack or end-of-row switches.

Network cabinets typically have more network connections for IT equipment than server or storage cabinets. Therefore network cable management is one critical design criteria for the network cabinet design. Network connectivity may consist of copper UTP, fiber, and proprietary interconnections between redundant chassis or between multiple chassis to make a single virtual switch stack. Adequate space with suitable bend radius should be provided within the cabinet for all network cablings, preferably with physical separation between the copper and fiber cables.

The management of power distribution cables is also a factor when selecting the appropriate cabinet manufacturer and accessories. It is recommended to have the power cables routed vertically in the cabinet in the opposite corner from the copper network cabling. Large network chassis typically will have multiple 20 A, or higher, power cords. The power cords should be provided with cable management solutions that are sized to support all of the cordage without exceeding the pathway capacity limits.

23.3.5 Server and Storage

Rack mountable server and storage platforms are mounted within ITE cabinets. Some of the main considerations when selecting ITE cabinets for server or storage platforms include:

- Top or bottom entry of network cabling
- Top or bottom entry of power cables
- Width and depth of standard cabinet to be used
- Will a vertical exhaust duct be incorporated into the overall cooling solution
- Is physical security required for each or specific cabinets, if so, manual locks sufficient or is electronic locking to provide entry logs required

Most cabinet manufacturers have options to provide electronic locks on their cabinet doors. If the cabinet doors are required to utilize a specific vendor's electronic lock system to coordinate with a security vendor's electronic locks used throughout the data center, the available cabinet vendors may be limited. Electronic lock systems from security vendors may not work within cabinets or within all cabinet vendor solutions.

Cabinets that have bottom entry of either power or network cables will need to have floor tile cutouts positioned within the footprint of the cabinet. In order to provide flexibility with respect to where the floor cutouts are positioned, the cabinet frame should have minimal obstructions at the base of the cabinet. Some cabinet manufacturer solutions have flat plates that provide structural support between the inner rails and the outer cabinet panels. These flat plates may impede where the floor tile cutouts can be positioned.

The cabinet width and depth should provide sufficient space to mount at least two vertical power strips in one corner and route network cabling in the opposite corner.

Server and storage manufacturers with rack mountable solutions may provide a swing arm accessory to manage power and network cable management to each device. The swing arms can significantly impede the airflow out the back of the server or storage platform, increasing the temperature within the cabinet. The benefit of the swing arm is that a device can be "racked-out" without disconnecting the power or network connections. It is prudent to ask if standard operating procedures include powering down a chassis before any hardware upgrades or modifications are made. If this is a standard operating procedure, which is typical, then there is no need to have the swing arms on the back of the equipment.

ITE cabinets have historically been black in color. The design may want to consider implementing white colored ITE cabinets. White cabinets help to reduce the amount of energy required to provide the recommended lighting levels within the computer room. Ladder racks for overhead cabling pathways are also available in a white to match the cabinet color.

23.3.6 Large Frame Platforms

Large frame platforms are systems that do not fit within a standard 2,100 mm (7 ft) high cabinet with standard EIA/CEA-310-E mounting rails. These systems often include large disk arrays for enterprise storage, mainframes, HPC systems, supercomputing systems, or tape libraries. The cabinets that are used to support these large frame systems are often wider and deeper than typical server cabinets.

If large frame platforms are used within a data center, the layout of these systems on the computer room floor must be planned early to ensure that the critical building systems are design appropriately. Power distribution, cooling methodologies, lighting layouts, fire detection, and suppression system layouts must all be coordinated with the IT equipment layout.

It is also common for large frame platforms to have the bottom entry power and network connections. In a non-raised floor computer room, an appropriate method of routing the power and network cabling must be identified.

23.4 SPACE AND POWER DESIGN CRITERIA

Power demand density (w/sf or w/rack) has often been used as the criteria to establish power and cooling capacity requirements. This approach may lead to inappropriate capacity projections but is often used by facility engineers who either don't understand how the various ITE platforms impact capacity planning or do not have access to the necessary information.

Proper capacity planning does not simply identify the existing power density, say 1,000 w/m^2 (92 w/ft^2), and then apply some multiplier, say ×2, and define the future power, and cooling requirements should provide 2,000 w/m^2 (184 w/ft^2).

23.4.1 Platform Dependent

The recommended approach to develop and define future space, power, and cooling capacity requirements is to analyze each hardware platform and the supported applications. ITE platforms are a subset of equipment within each ITE system (network, compute, storage) to differentiate between various types of equipment function, form factor, class, generation, etc. This exercise is not a facility engineering analysis, but an enterprise architecture and IT analysis driven by types of applications used to support the business objectives.

Identify the hardware platforms and review historic growth by platform if available. Review the supported applications and identify impact of future requirements on hardware capacity planning. The hardware platforms are typically compartmentalized into the following categories: (i) network, (ii) server appliance (non-blade), (iii) blade server, (iv) large frame processing (mainframe, HPC, etc.), (v) large frame disk arrays, and (vi) rack-mounted disk arrays.

23.4.2 Refresh

Refresh capacity is required when applications or data are being migrated from legacy systems to new systems. Refresh also can have a significant impact on capacity planning. If an organization utilizes rack-mounted appliance servers and blade servers with very little storage requirements, refresh may not be that significant. These servers are typically refreshed individually, not as an entire platform. In this scenario the refresh

capacity required may be less than 5% (space, power, and cooling) of the total capacity supporting server platforms.

If an organization has implemented large frame disk array platforms within their storage architecture and the disk arrays require a significant amount of the space, power, and cooling capacity in comparison with the processing capacity requirements, the capacity planning results will differ significantly from the previous example. When disk arrays go through a technology refresh, they are typically refreshed as an entire platform. If the entire disk array consists of two 9-frame systems, the refresh migration will require the space, power, and cooling capacity to stand up an entire new system to replace the two legacy 9-frame systems. The legacy system and the new system will need to function alongside each other for a period of time (months). The new system will need to be powered up, configured, and tested before the data will be migrated from the legacy system. After the new system is in full production, the legacy system can be decommissioned and removed from the data center. This scenario results in the capacity plan requiring double the space, power, and cooling requirements for the anticipated disk array platforms to facilitate technology refresh.

If a data center's computer room that is supporting multiple platforms (Fig. 23.8) has more than 80% of the computer room space, power, and cooling capacity consumed, there may be little or no room for growth but only excess capacity available to support the refresh of the current platforms.

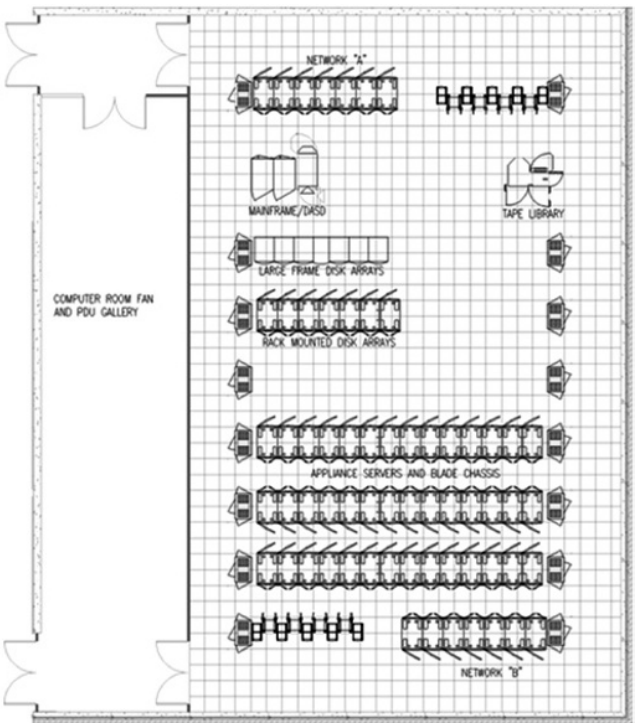


FIGURE 23.8 Example of multiplatform computer room layout. Source: Courtesy of Isaak Technologies.

23.4.3 Power Density

The power density of the computer room is an outcome of analyzing the space and power capacity planning exercise. It is not the recommended starting reference point for establishing power and cooling capacities. Once the growth and refresh requirements have been established as noted in the previous sections, the power density can be expressed.

23.5 PATHWAYS

23.5.1 Entrance Network Pathway

It is recommended that customer-owned maintenance holes be installed at the property line. Customer-owned conduits are to be installed between the customer-owned maintenance holes and the data center. This ensures that the customer manages and controls which network service providers have access to their data center and how they physically provision the fiber to the data center.

The elevation of each maintenance hole cover must be lower than the elevation of the entrance conduits terminated inside the data center. This will ensure that moisture does not migrate into the data center from a flooded maintenance hole.

The minimum recommended conduit size and quantity for entrance pathways are four 100 mm (4 in) conduits to support up to three network access providers. The conduits will have either hard-walled or fabric inner duct placed inside the conduits to enable future fiber pulls without damaging the initial pull. When more than three network access providers are anticipated to serve the data center, one additional conduit is recommended for each additional provider. The conduits are recommended to be concrete encased from the maintenance hole to the facility, with a minimum of 1.2 m (4 ft) separation from any other utility.

The routing of the entrance pathway from the property line to the entrance room in the data center should meet the following standards:

- BICSI Class 1: One route with at least four conduits to the entrance room.
- BICSI Class 2: Two diverse routes, each with at least four conduits to the entrance room.
- BICSI Class 3 and 4: Two diverse routes, each with at least four conduits to each entrance room.

For BICSI Class 3 and 4, where two entrance rooms are recommended, it is recommended to install 100 mm (4 in) conduits between the entrance rooms as well. The quantity of the conduits between the entrance rooms should be the same as the quantity entering the building. These conduits are provided to give network access providers flexibility in how they route their ringed fiber topology to the data center, either in and out the same entrance room or in one entrance

room and out the other. These conduits should not be used for any other function than to support network access provider's fiber infrastructure.

23.5.2 Computer Room Pathway

Power or network pathways can be routed overhead or underfloor in a data center with a raised floor. There are viable solutions for either scenario. The decision as to which method to use is often designer or owner preference over technical merits.

Whether overhead or underfloor is selected, the fiber optic pathway design must include sufficient spare capacity for additional future fiber optic cables. Future fiber strands may be required to support the following:

- Additional networked devices.
- Additional ports on the existing devices for either redundancy or network throughput.
- Migrating ports based on serial optical interfaces to ports based on parallel optical interfaces.

23.5.2.1 Overhead

Power

Overhead power distribution methods consist of running power whips, or wire in conduit, from a distribution frame to receptacles above each cabinet, or a wire busway with hot-swappable plug units.

The wire busway provides greater flexibility and reconfiguration as the ITE changes over the life of the data center.

Overhead power alongside overhead network requires greater design coordination to ensure that adequate physical separation exists between power circuits and any unshielded copper network links. Minimum separation between EMI sources and network cabling is dependent on the type of EMI source such as motors, transformers, and shielded and unshielded power circuits (refer to TIA-569, ISO/IEC 14763-2, CENELEC EN 50174-2 standards).

In the past, large frame systems such as large drive arrays and mainframes had their power and network connections only accessible from below the systems. With the industry moving more toward overhead power and network distribution throughout the data center, these vendors have created options for their large frame systems to have the power and network connections fed from above. For high ampacity (50 amps or higher) power connections, this may add additional clearance required above the system frames as the power connectors are larger.

Network

Overhead network distribution is generally the preferred method used. Overhead copper network pathway options include top-of-cabinet trough systems, basket trays, and ladder rack. For overhead fiber optic distribution, it is recommended to use fiber optic troughs designed specifically for fiber optic infrastructure. Cable trays, typically used to distribute power circuits, are not recommended for network distribution within the computer room of the data center.

The top-of-cabinet trough systems are often used in smaller data centers with one or two rows of ITE. The trough system requires less coordination, minimizes ceiling height requirements, and is a cost-efficient solution. The trough system impedes moving or replacing a single cabinet with the row in the future as the cabinet itself supports the pathway.

Basket trays are very common as they are a cost-effective solution. Basket trays ease the installation of the pathway in applications where there are numerous elevation changes.

Ladder racks are also very common as they provide the most options to assist in the transition of the copper cabling from the ladder rack down to the racks or cabinets. Top rung ladder racks allow the copper cables to transition off the side or through the rungs, either method using water fall accessories to limit the bend radius. The rigidity of ladder racks also provides a good structural system that can be used to directly attach fiber optic troughs if used.

Fiber optic troughs designed specifically for fiber optic infrastructure provide a common method of distributing fiber throughout the data center's computer room. The fiber optic trough can ensure that the minimum bend radius is maintained throughout the pathway and at all transitions from the trough down to racks and cabinets. The designer will need to coordinate the fiber trunk strand counts and outside diameter of the fiber optic cable sheath with the bend radius of the fiber optic trough. Most fiber troughs have bends with a 50 mm (2 in) radius. Fiber optic trunks with strand counts greater than 48 will typically require bend radius greater than 50 mm (2 in). Split corrugated tubing should be used in the transition from the trough to the cabinet or rack to provide physical protection for the fiber optic cables.

23.5.2.2 Underfloor

One simple and low-cost system that is often overlooked in new data center designs is to include lighting within the underfloor space. Including light fixtures below the raised floor will help provide a safe working space for any technicians that need to maintain systems within the underfloor space.

Power

The routing of power cabling under the floor is often incorporated into current data center designs. The typical solution uses a liquid-tight flexible metal conduit that lies on the floor below the raised floor system. Some jurisdictions require wire in conduit.

Network

Routing network cabling under a raised floor requires coordination with all of the other building systems within the underfloor space. This includes power cables and conduits, chilled water piping, grounding systems, fire detection, and suppression systems.

The industry is trending away from routing network cabling in the underfloor space unless it is specifically for large frame systems that require cable entry from below.

For copper UTP cabling, the most common method is to use a wire basket solution to distribute the network cabling. The wire baskets are either supported by stanchions independent from the raised floor pedestals or supported by the raised floor pedestals themselves. When being supported by the raised floor pedestals, the designer must ensure that the pedestals will support the cable pathway system in addition to the raised floor and the equipment on the floor. This also requires coordination between the floor pedestal and underfloor pathway vendors to ensure manufacturer's warranties are not voided.

For fiber optic cabling, there are purpose-built fiber troughs available to distribute the fiber cable. If the underfloor space is used as an air distribution plenum, the pathway must be rated to be used within a plenum space. The only products available to meet this requirement are metal troughs. Metal troughs are often not desired as they are typically made out of light gauge sheet metal that easily cut when being handled in a confined space.

Another method to ensure physical protection for fiber optic cable that is distributed in an underfloor space is to use armored fiber optic cable. This method does require the designer to coordinate the quantity of fiber sheaths that will be terminated in any rack-mounted fiber shelves above the floor.

23.6 COORDINATION WITH OTHER SYSTEMS

There are many benefits to locating non-IT systems outside of the data center computer room. Non-IT systems such as CRAC/CRAH units, power distribution units (PDU), and uninterruptable power supply (UPS) systems that are sometimes located in the computer room for smaller-sized data centers. Benefits include minimizing the activity of non-IT personnel within the IT space, minimizing the activity of IT personnel within facility spaces, removing from the computer room heat generated by the facility systems, and simplifying the coordination of the IT systems placement within the computer room.

There are various design solutions that result in non-IT systems located outside of the computer room. One method is to configure galleries (Figs. 23.9 and 23.10) adjacent to the computer room where cooling and power distribution equipment is located. Another method is a multistory facility (Fig. 23.11) with the computer room above or below the level the cooling and power distribution equipment is located.

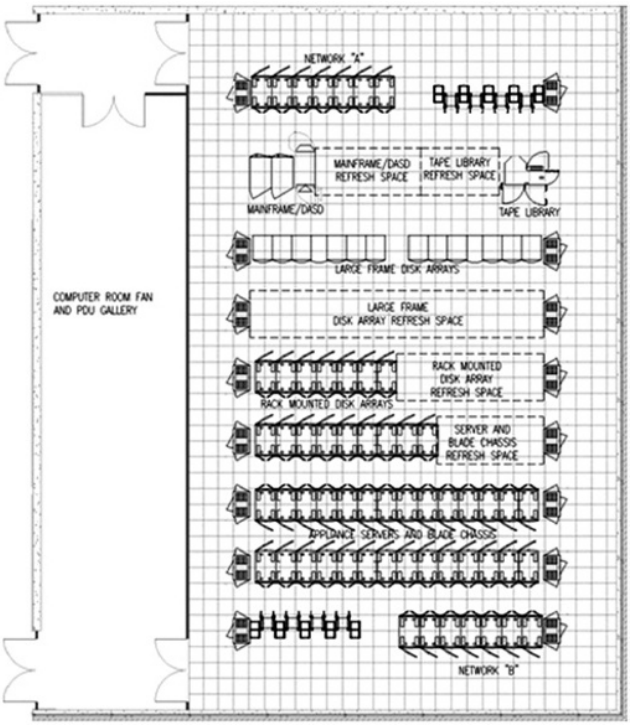


FIGURE 23.9 Example of multiplatform computer room with 18% space required for technology refresh. Source: Courtesy of Isaak Technologies.

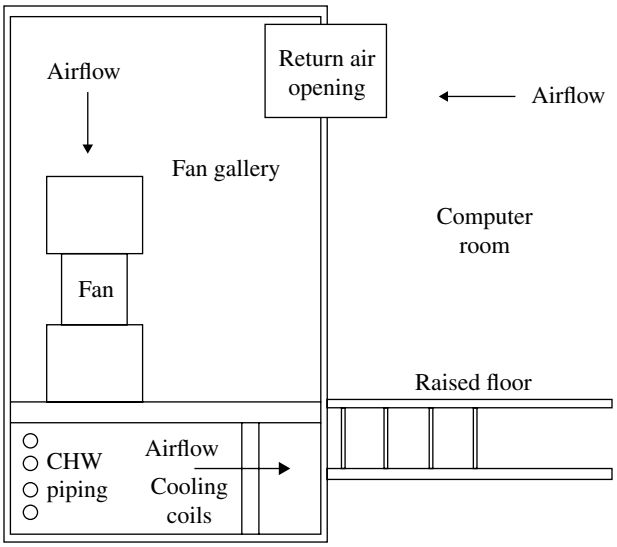


FIGURE 23.10 Example of fan gallery (section view). Source: Courtesy of Isaak Technologies.

The design considerations noted below assume the facility support systems are located within the computer room.

It should be noted that the floor grid measurements of 1.2 m (4 ft) is not a nominal equivalent typically used when expressing dimensions in both metric, imperial, and US units. In the United States the floor grid is based on

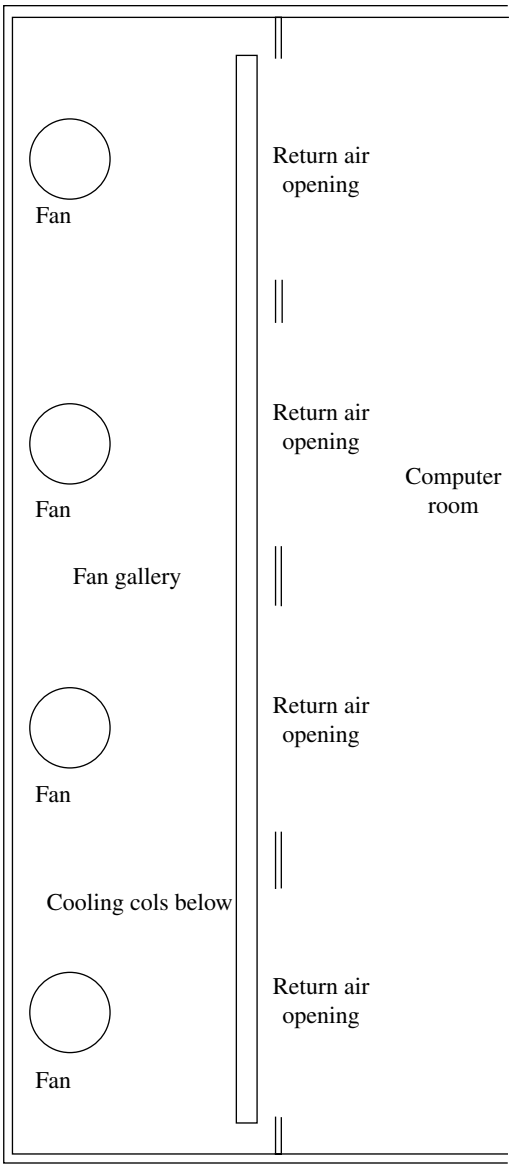


FIGURE 23.11 Example of fan gallery (plan view). Source: Courtesy of Isaak Technologies.

2 ft square floor tiles that are 609.6 mm square. In countries that have incorporated the metric system (almost everywhere other than the United States), the floor grid is based on 600 mm square floor tiles that are 23.622 inches square.

23.6.1 CRAC/CRAH

CRAC/CRAH units are typically placed along the longest wall of the computer room, with the rows of ITE perpendicular to the longest wall. The CRAC/CRAH units are recommended to be aligned with the hot aisle to maximize the distance from each CRAC/CRAH unit to the closest perforated floor tile. This also simplifies the return airflow stream from the hot aisle directly to the CRAC/CRAH unit return air inlet, which is more critical for computer rooms with open room return and low ceilings.

The minimum recommended distance measured from the corner of a CRAC/CRAH unit to the closest perforated floor tile should be 2.4 m (8 ft). If less, the pressure above the tile may be more than the pressure below the tile, with the air flowing from above the floor to the underfloor space. Perforated floor tiles with air flowing from above the floor to the underfloor space could be replaced by a solid floor tile as they are likely providing no cooling value to the IT equipment above the floor.

A minimum of 1.2 m (4 ft) is recommended for cold aisle spacing. A minimum of 900 mm (3 ft) is recommended for hot aisle spacing, with 1.2 m (4 ft) preferred. The exact aisle spacing should be coordinated between the IT design and the cooling system design to ensure that adequate air floor can be delivered from the cold aisle and returned to the CRAC/CRAH unit from the hot aisle.

There should be a minimum 1.2 m (4 ft) aisle space between the end of the ITE row and any CRAC/CRAH unit, resulting in a 1.2 m (4 ft) aisle around the perimeter of the computer room. It is also recommended to consider providing 1.8 m (6 ft) aisle clearance along one or more of the perimeter walls to move CRAC/CRAH or large frame ITE in or out of the computer room, providing more clearance along the route to the entry/exit doorway.

23.6.2 Power Distribution

PDU is the electrical components that transform the voltage from the building distribution voltage levels (480, 600 V or medium voltage) down to the ITE voltage level (208, 240 V). It is recommended that the PDUs be located outside of the computer room.

Remote power panels (RPPs) are the electrical components that provide higher quantity of power circuit pole positions in a high-density frame compared with standard wall-mounted panels. The RPPs are downstream from the PDU and feed the power outlet strips within the IT equipment racks and cabinets. The RPPs are recommended to be placed at one or both ends of the rows of IT equipment, depending on the level of reliability required.

RPP is typically made up of four 42-pole panels. The entire panel can be fed from one upstream breaker, an individual breaker feeding each panel, or any combination in between. It is possible to feed two of the panels from power source “A” and two panels from power source “B.”

For data centers designed to meet ANSI/BICSI Class F2 or lower one RPP at one end of a row of equipment will meet the design standards. For Class F3 and F4 in order to meet the minimum design standards, all panels within the RPP will be feed from one power source upstream. This results in each row of ITE being fed from an “A” RPP and a “B” RPP, which is typically accomplished by placing an RPP at both ends of the ITE row.

Placing the RPPs at the ends of the rows reduces coordination with other systems compared with placing them against the outside wall. When they are placed at the ends of the row, the power whips feeding the racks or cabinets within the row are contained in the floor grid directly under the racks or cabinets.

When the RPPs are placed against the outside wall, the power whips have to transition across the perimeter aisle to the row of racks and cabinets. If this method is used, the power whip installation needs to be coordinated with, when used, other underfloor pathways running perpendicular to the ITE rows, such as chilled water or refrigerant lines.

Overhead power busses may also be used to distribute power to the ITE racks and cabinets. Typically, RPPs are not used when overhead busses are implemented. The overhead buss design provides flexibility for the IT designer in that the exact power circuit feeding each rack or cabinet can be easily and quickly changed by inserting a plug-in unit that has the correct breaker and receptacle configuration required. Overhead busses need to be coordinated with all other overhead systems such as sprinkler head locations, network cabling pathways, and lighting. If there are two overhead busses providing an "A" and a "B" power source, the position in either the horizontal or vertical plane needs to provide sufficient separation to be able to insert the plug-in units without conflicting with the other buss or plug-in units.

Power cables entering cabinets, either overhead or through a floor tile below, should be sealed with an appropriate grommet. The grommets need to be sized to provide sufficient space to pass the power outlet strip cords through, which is the diameter of one power outlet strip cord plus the diameter of one power outlet strip cord end cap.

Power outlet strips are available in either horizontal or vertically mounted models. Vertical, referred to as zero-U, models are typically used in server cabinets. Vertically mounted power outlet strips may not be the preferred model for network cabinets as the air dam kits required to support ITE with side-to-side airflow may restrict the placement of the power outlet strips on the exhaust side of the cabinet. This is cabinet manufacturer and model dependent. For this reason, horizontal power outlet strips are sometimes used for all network cabinets. Horizontal power outlet strips have often been used for open two- or four-post racks as previously there were no good options to mount vertical power outlet strips to open racks. However, manufactures have developed mounting brackets that enable vertical power outlet strips to be mounted to open racks.

23.6.3 Sprinkler and Fire Protection Systems

The local authority having jurisdiction will define if sprinkler or fire protection systems are required below the raised floor or above a suspended ceiling in addition to the computer room space.

Sprinkler and fire protection systems should be the system that is mounted the highest when required under a raised floor or above a suspended ceiling, with all other pathways and systems mounted at a lower level. Fire detection devices can be mounted vertically if specifically designed and approved for vertical mounting applications.

Sprinkler head placement may be a critical coordination issue for computer rooms with less than 3 m (10 ft) ceiling height, especially when overhead power or network cabling is used. Sprinkler heads typically require 450 mm (18 in) clearance below the sprinkler head; local Authorities Having Jurisdiction (AHJ) may have greater restrictions.

The sprinkler and fire protection systems coordination challenges are often eliminated if ceiling heights of 4.2 m (14 ft) or higher are implemented within the computer room.

23.6.4 Lighting Fixtures

When a suspended ceiling is used, lighting fixtures are generally inserted in the ceiling grid system. This method requires close coordination with the ITE rows, overhead pathways, sprinkler, and fire protection devices.

Computer rooms with higher ceilings may implement indirect lighting by using suspended light fixtures with most of the light directed up and reflected off the ceiling (painted white) to provide sufficient light within the room. This method provides a more even distribution of light throughout the room with less shadowing compared with light fixtures inserted in the ceiling grid. When using an indirect lighting system, it is recommended to have the suspended light fixtures installed above any other suspended systems such as power or network cabling pathways. This will minimize the risk of lamps breaking when technicians work on systems above the light fixtures.

23.6.5 Raised Floor vs. Non-Raised Floor

Building constraints in floor to deck heights may restrict the design from incorporating a raised floor. Incorporating a data center within an existing building may also restrict the ability to have a sunken slab for the computer room space. It is always desirable to have the computer room floor at the same elevation the entire route from the adjacent corridor space to the loading dock. Ramps to accommodate a change in floor elevation between the computer room and the adjacent corridor are not only a functional annoyance; they require additional foot print within the computer room. The maximum recommended slope for the ramp is 4.8°, a rise of 1 : 12. For a 600 mm (24 in) raised floor, this would result in a 7.2 m (24 ft) long ramp. The ramp is also required to be 900 mm (36 in) wide with a 1.5 m (60 in) clear landing area at the top and bottom of the ramp to meet building codes in support of disabled persons. This results in an additional 9.18 m^2 (102 ft^2) of space required to accommodate the ramp.

For a large computer room, this will not be significant, but for a small data center, this may significantly reduce the space available for ITE.

It has often been stated that if the data center cooling system does not use the underfloor space for air distribution, then a raised floor is not required. This is only considering one aspect of the raised floor when others exist. A raised floor environment provides flexibility to accommodate future ITE technology requirements that are not within the initial design criteria. This could include liquid-cooled ITE using liquid-to-chip or immersion technology, where the fluid lines would be preferred to be routed below the ITE vs. overhead.

If a raised floor is not incorporated into the design, coordination issues are introduced even when using traditional CRAC/CRAH cooling technology. The installation and placement of drains for CRAC/CRAH condensate lines must be provided, along with sufficient floor slope to keep any moisture accumulation away from ITE. Proper overhead ITE grounding system as an underfloor supplemental bonding grid will not be available.

23.6.6 Aisle Containment

The design of aisle containment systems should always be reviewed with the local AHJ over fire protection systems. Containment systems either have fire detection and suppression systems integrated within the contained space or have top panels that are designed to automatically be retracted so they do not impede the suppression system coverage pattern from the nozzles above the top panels or obstruct the egress path for any personnel within the contained space. Legacy systems consisting simply of drop down panels in the event the contained space exceeds a specified temperature do not meet the requirements of NFPA 75. The local AHJ may place certain constraints on how a containment system is integrated into the overall fire protection plan.

Lighting fixture type and placement will need to be coordinated with the aisle containment system to ensure sufficient light levels are provided within the contained spaces.

Overhead or underfloor power and network cabling pathways can be easily incorporated into either a hot aisle or cold aisle containment system. The containment system itself does not introduce any new coordination challenges.

Vertical heat collars (also known as chimneys) are technically not aisle containment, but rather a contained vertical exhaust duct. Vertical heat collars do introduce additional coordination challenges in that up to half of the space available over the cabinets is consumed by the vertical duct, and therefore no longer available for routing the power or network cabling pathways. Additional coordination is required when incorporating overhead network cabling pathways and overhead power distribution busses together with vertical heat collars. All of these systems need to fit within the limited space over the cabinets.

When implementing containment, decisions need to be made with respect to placement of overhead power and network pathways. Figure 23.12 shows one method of containment, a hot aisle containment with the hot air return path utilizing the plenum space above a suspended ceiling.

No matter if hot aisle, cold aisle, or vertical exhaust ducts are used for containment, significant coordination is required with the power and network pathways, lighting, fire detection, fire suppression, and the containment system. There are several ways to configure the power and network over the IT cabinets; however the following are examples of design considerations that should be analyzed by the owner and designer:

- Are the power and network pathways to be located in the cold or hot space? This can dictate if the containment panels are positioned at the front or back of the cabinets.
- Will cabinets and open racks be placed adjacent to each other within the same row? This may require custom containment panels to align between the cabinets and open racks.
- How will fire detection and fire suppression systems be implemented within the contained spaces?
- How will lighting be implemented within the contained spaces?
- If implementing overhead power bus that is positioned over top of the cabinets, it is important to review with the local AHJ early in the design process to ensure it does not violate local safe work space codes.

23.7 COMPUTER ROOM DESIGN

23.7.1 By Size

For the purposes of our discussion, we will define small computer rooms as less than 280 m^2 (3,000 ft^2), medium computer rooms as less than 930 m^2 (10,000 ft^2), and computer room with more space as large. These parameters are certainly not a standard within the industry but are simply used to guide the discussion on the design nuances between different sizes.

23.7.1.1 Large

Large data centers will require columns to support the roof structure or have the computer room space compartmentalized into multiple spaces with structural members to support the roof structure. The location of the columns or structural members should be coordinated between the structural engineer and the IT designer to minimize the interference

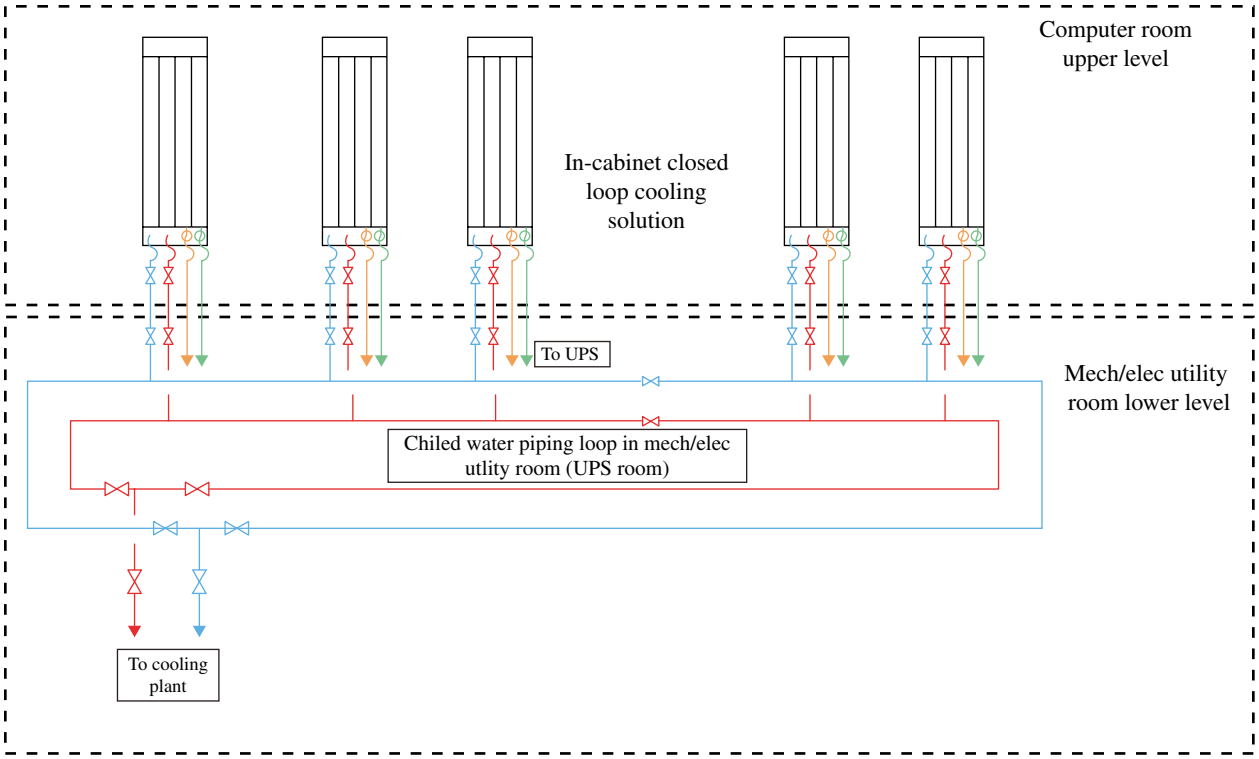


FIGURE 23.12 Example of multistory data center. Source: Courtesy of Isaak Technologies.

with the ITE layout. The objective is to provide as open as a space as possible to accommodate various and changing IT equipment layouts.

Large data centers may also require additional network frames distributed throughout the computer room space to support distribution network switches. This will be dependent on the network architecture and topology that is deployed in the data center.

23.7.1.2 Medium

Medium data center may require columns to support the roof structure. However, the designer should identify solutions that are available to avoid or minimize the quantity of columns. Columns are a coordination challenge with the initial ITE layout and all future technology refreshes.

Medium data centers may be able to centralize the network distribution with top-of-rack access switches uplinked to a centralized core network. Space to support additional network distribution frames throughout the computer room will likely not be required.

23.7.1.3 Small

Small data center can be the most challenging to coordinate the ITE with all the other systems. They often push the design to high-density solutions as owners are trying to compress as much processing capability within a small defined footprint.

Columns within the computer room should be avoided.

Small data centers will typically have a centralized network core.

23.7.2 By Type

23.7.2.1 In-House Single Platform Data Center

Organizations that own and manage their own data centers with single or minimal platform variations required to be supported can have a consistent repeatable ITE layout. Organizations that have all compute processing on rack mountable appliance or blade servers, and all storage on rack mountable drive arrays, have the capability to plan their ITE layout using consistent zones of cabinets.

The computer zones can be based on the standard cabinet width and depth and standard aisle spacing. As an example, if the standard cabinet is 800 mm (31 in) wide and 1.2 m (48 in) deep, the computer room layout will consist of repeatable rows with 1.2 m (4 ft) hot and cold aisles between the ITE cabinets. Since all platforms are rack mountable in standard ITE cabinets, it is not necessary to know exactly where each system, whether network, appliance server, blade server, or storage drive array, when developing the initial ITE floor plan layout. If the power distribution from the UPS downstream to the ITE cabinets is designed

according to the ANSI/BICSI 002 standard, there should be sufficient flexibility in the capacity of the PDUs and RPPs to support each zone independent of the specific platform within each zone.

Since all the ITE systems are consistently installed in standard cabinets (Fig. 23.13), the power and network cabling pathway design can also be consistently applied across the computer room space.

23.7.2.2 In-House Multiplatform Data Center

Organizations that own and manage their own data centers but have numerous platforms may require unique zones to support each platform type (Fig. 23.14). Unique platform types may consist of rack mountable servers (appliance or blade), large frame compute processing (mainframes, HPC, supercomputers), large frame drive arrays, or carrier class network platforms (580 mm (23 in) rails).

The computer room ITE layout will need to identify each unique zone and the size of each zone. The placement of the power distribution (RPPs), power pathways, and network pathways will all need to be coordinated with the unique requirements of each zone (platform). Each zone will need to have sufficient spare capacity to accommodate the anticipated growth and the technology refresh requirements.

Large frame systems can have ITE with depths up to 3.6 m (72 in) and varying frame heights and widths. Systems such as tape libraries will have even larger footprints. Large frame systems have various airflow patterns through the equipment for cooling such as front to back, side to side, and bottom to top that will be unique to the platform, manufacturer, and model of platform. Large frame systems have various power or network cable entry points, sometimes accessible only from the bottom of the ITE. Bottom cable entry points may have very little tolerance as to the placement of a floor tile grommet below the specific frame. These unique characteristics of large frame systems require extra attention to the design details and coordination with the supporting systems.

Large frame systems with power and network cable entry points at the bottom of the systems typically route the power and network cabling supporting these systems under the raised floor. It is common to have overhead pathways for rack mountable systems in standard cabinets and underfloor pathways for the large frame systems within the same computer room.

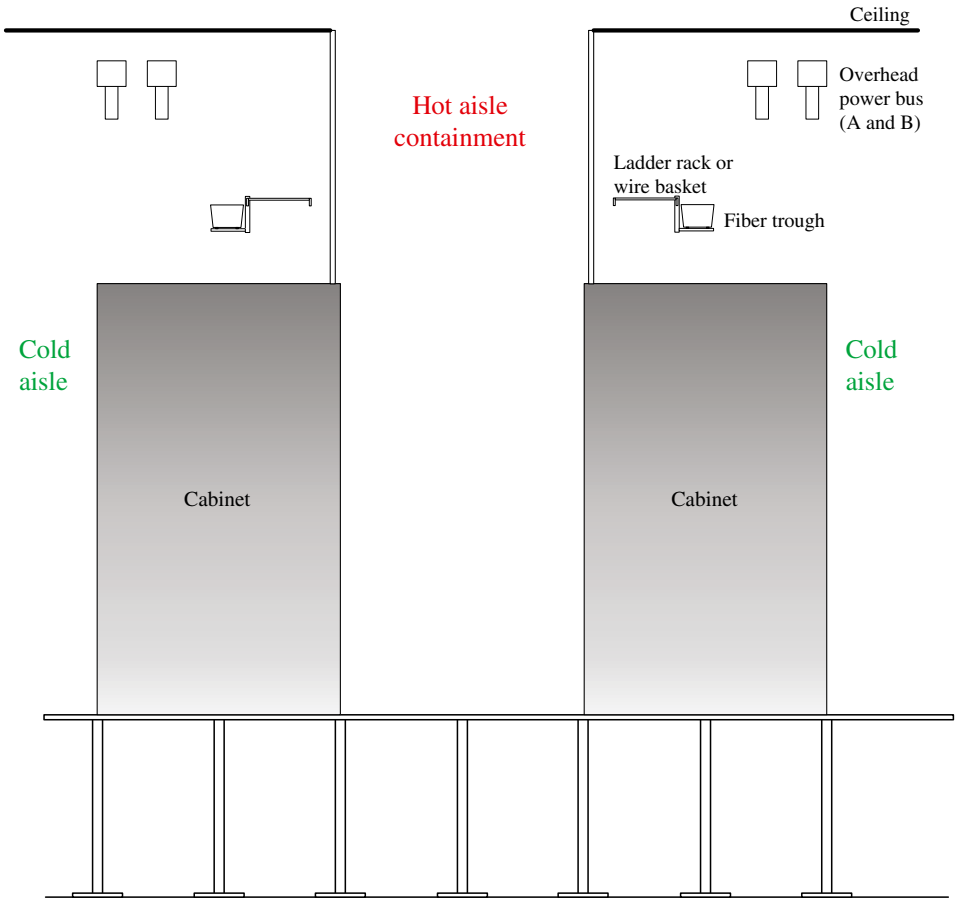


FIGURE 23.13 Example of hot aisle containment. Source: Courtesy of Isaak Technologies.

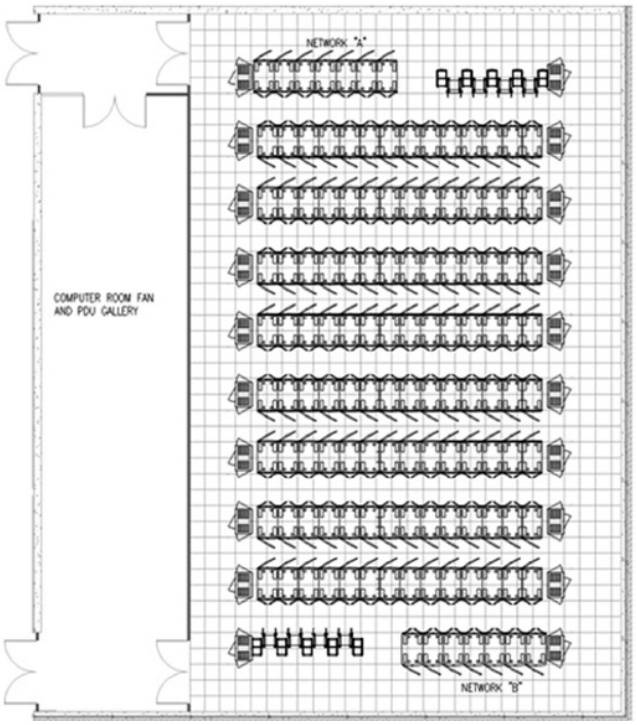


FIGURE 23.14 Example of computer room layout with all IT platforms mounted in standard cabinet—all equipment in 4 ft zones. Source: Courtesy of Isaak Technologies.

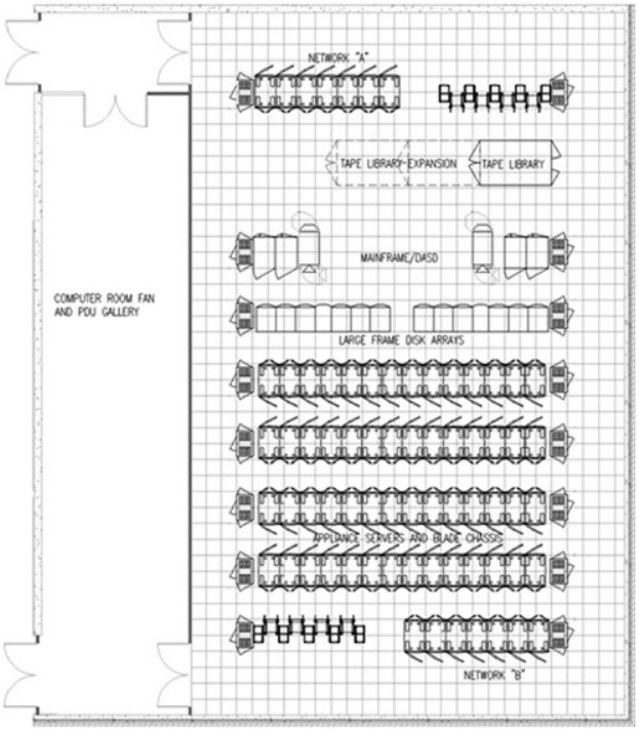


FIGURE 23.15 Example of multiplatform computer room layout. Source: Courtesy of Isaak Technologies.

23.7.2.3 Outsourced Services Data Center

Colocation data centers consist of organizations that own data centers and manage the space, power, and cooling infrastructure to support their customer's platforms placed in either caged spaces or cabinets (Fig. 23.15).

This type of data center requires a different approach to defining the space, power, and cooling capacity requirements. The owner does not know exactly what the ITE layout will look like until customers have committed to their services and defined the systems they will be placing within the colocation data center. This information is not known at the time the colocation owner is planning and designing the data center. Therefore, colocation data center design drivers typically are cost control, flexibility, and scalability.

Cost control is required to ensure the levels of reliability, redundancy, and the services provided by the data center are in alignment with the potential customer's requirements and price point. Flexibility is required as the capacity requirements of each individual caged space or cabinet will vary significantly over the life of the data center, as customer's technology changes or as customers move in and out of the data center. Scalability is required to enable the colocation data center capacity (space, power, cooling) to be built out as customer demand requires.

Colocation data centers typically provide space, power, cooling, and connectivity to network access providers. The colocation owner typically does not manage the network, but rather simply provides a handoff of the service provider's circuit at the entrance room to the customer's equipment located in a caged space or cabinet. There are distance limitations on various circuits provided by the network service providers. For large colocation data centers, there may be a requirement to have multiple entrance rooms so that T-1/E-1 or T-3/E-3 circuits can be extended to the customer's space without exceeding distance limitations. The multiple entrance rooms do not provide any redundant capabilities in this scenario. Requiring multiple entrance rooms can be significant negotiation challenge with the network service providers. They are mandated to provide network services for each street address. For a large colocation data center (Fig. 23.16), there are many customers for the network service provider, but there is only one street address. It is typical for the network service providers to require the colocation owner to pay for all installation costs for any additional entrance rooms, even though they are required to bring services to the customers within the data center (the customers are also the network service provider's customers). The negotiations revolve around cost sharing between the network service providers and the colocation data center owner, but the data center owner would prefer cost avoidance.

23.8 SCALABLE DESIGN

Incorporating a scalable approach to a data center design is generally always applied. The exception to this is small data centers where the ultimate power and cooling capacity is not less than

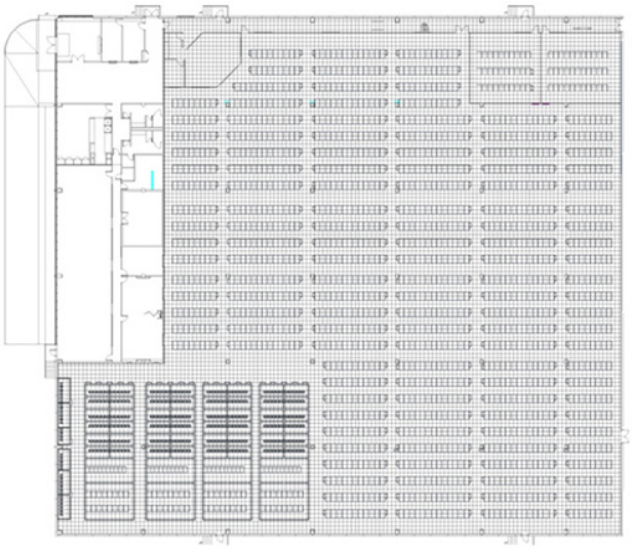


FIGURE 23.16 Example of colocation data center with customer caged space and customer cabinet layout—lease by cabinet or caged space. Source: Courtesy of Isaak Technologies.

40% of a single module or component. Scalable design needs to address space, power, cooling, and network capacity.

A critical consideration when incorporating a scalable approach is that the design must be able to support future expansion without reducing the level of redundancy of the critical systems. This also must be able to be accomplished without disrupting the data center compute processing production.

23.8.1 Computer Room Space

Of all the facility-related aspects that are affected by scalable designs, space is the one that often impacts cost the least. The total cost of data center facilities is generally composed of 30% of the building shell and interior build-out and 70% in the electrical and mechanical systems (land and IT systems not included). This ratio will fluctuate based on the level of redundancy required and the size of the computer room space. Since the building represents the smaller portion of the total facility costs, historically it was not uncommon to build out two or three times the initial required floor space to accommodate future growth. However, with the technical advancements of virtualization and containerization of compute systems, the total compute capacity continues to increase without requiring additional computer room space. This is one reason why a detailed capacity plan and forecast are important planning exercises to complete to provide a detailed profile of future capacity requirements.

It may be necessary to plan for the expansion of space to accommodate future growth. This can be accomplished by constructing additional computer rooms adjacent to the initial building or incorporating knockout panels in the computer room perimeter wall that can be removed in the future.

23.8.2 Power and Cooling Infrastructure

Incorporating a scalable design in the power and cooling systems is a standard approach to use. It is very common for new data centers to have the initial power and cooling capacity less than 50% of the ultimate capacity design.

The initial build-out of power capacity must have the power distribution coordinated with the ITE layout. It is more practical to build out the computer room from one end and systematically expand across the computer room space. This allows the initial electrical distribution (PDUs and RPPs) and cooling equipment to provide capacity to the initial zone of ITE. Future PDUs, RPPs, and CRAC/CRAH units will be added in the adjacent ITE zones as additional capacity is required.

It is critical that the future PDUs, RPPs, and CRAC/CRAH units can be added without disrupting the systems initially installed. It is recommended that the installation of future PDUs not require a shutdown of any upstream distribution that is feeding "hot" PDUs and the installation of future RPPs not require a shutdown of any upstream PDUs that are feeding "hot" RPPs.

23.8.3 Network Capacity

Incorporating a scalable design for the network addresses not only capacity but also the physical location of the entrance room within the data center.

23.8.4 Scalability Versus Reliability

Data center operators often desire to have reliable and scalable solutions; however these are fundamentally opposing criteria. Scalability requires smaller capacity components in greater quantity to make up the ultimate design capacity (i.e., seven 500 kVA UPS modules vs. five 750 kVA UPS modules).

- The 500 kVA UPS module example can scale in five 500 kVA increments from 500 to 3,000 kVA (assuming redundancy is required for the UPS modules). If each module had a reliability value of 80% over a defined period, in an $N+1$ configuration the seven 500 kVA UPS module example would have a system reliability of 85.2%.
- The 750 kVA UPS module example can scale in four 750 kVA increments from 750 to 3,000 kVA (assuming redundancy is required for the UPS modules). If each module had a reliability value of 80% over a defined period, in an $N+1$ configuration the five 750 kVA UPS module example would have a system reliability of 88.2%.

Increasing scalability inherently decreases reliability. The designer of any system, whether it is for power distribution,

cooling distribution, or network architecture, must balance scalability and reliability to ensure that the appropriately sized building blocks are selected for the initial design and for the future incremental increases in capacity.

23.9 CFD MODELING

Computational fluid dynamics (CFD) is a method of modeling the effectiveness of the cooling system and its ability to meet the demand of the ITE being supported. In order to conduct a CFD analysis, the computer room space must be modeled, including the room dimensions, the placement of heat producing equipment within the room, the placement and type of the cooling supply units (CRAC/CRAH), the placement and type of perforated floor tiles, all openings within the floor tile system, and any obstructions to the air flow (pipes, cable trays, etc.).

The output of a CFD analysis will model the temperature and pressure variations throughout the computer room space (three dimensional). This has been proven valuable in data center design as the designer can validate the cooling system design prior to installing the system. It is also beneficial to data center operators as they can model:

- How the placement of future ITE will impact the cooling systems ability to meet the computer room demand
- Simulate various failure scenarios by "turning off" components within the CFD model and analyzing if the remaining cooling system can support the ITE load.

There are a few vendors that have developed the CFD software tools, with varying degrees of accuracy, level of modeling complexity, and cost.

23.10 DATA CENTER SPACE PLANNING

23.10.1 Circulation

The data center must support the replacement of all ITE and power and cooling system components by providing adequate clearances from the loading dock to the computer room, electrical, and mechanical rooms. Corridors should be at least 2.7 m (9 ft) high. Doors should be a minimum of 2.4 m (8 ft) high and 1.1 m (3.67 ft) wide for single doors or 1.8 m (6 ft) wide for a pair of doors. Consideration for corridors with higher ceilings and 2.7 m (9 ft) high doors should be made as a packaged standard 42 RU cabinet on a pallet jack typically does not fit under a 2.4 m (8 ft) high door.

The data center layout should be defined into various access types such as noncritical and critical facilities and critical IT. It is recommended to minimize personnel traffic between these zones as much as possible, keeping facilities personnel out of IT spaces and IT personnel out of facility spaces.

23.10.2 Support Spaces

Any function that is required in supporting the IT systems within the computer room is considered part of the data center. Functions that are not directly required to support the IT systems within the computer room are considered to be outside the data center.

The following critical spaces are required to support the IT systems within the computer room.

23.10.2.1 Entrance Room

The function of the entrance room is to provide a secure point where entering network outside cable plant from access providers can be converted from outdoor cable to indoor cable and to house the access provider-owned equipment such as their demarcation, termination, and provisioning equipment.

The entrance room should be located adjacent to or in close proximity to the computer room. The pathways from the entrance room to the computer room should not have to transition through any nonsecure spaces. The entrance room should also be located in close proximity to the electrical room where the main building ground bus bar is located in order to minimize the length of the bonding conductor for telecommunications.

For data centers with redundancy requirements, a second entrance room is recommended to provide physical separation between redundant access provider services. These entrance rooms should be located at opposite ends of the computer room from each other.

The entrance room often houses multiple network service providers. The configuration of the entrance room should be coordinated with each network service provider to ensure that their requirements are met and that all clearance requirements and special security concerns are understood.

23.10.2.2 Network Operations Room

The network operations room or center (NOC) supports IT operations. The NOC has technicians within this room monitoring the network and IT system operations, typically on a 24/7 basis.

The NOC is typically located adjacent to the computer room with an entry door into the computer room. This can act as another level of security in that everyone that enters the computer room would gain entry through the NOC, enabling the NOC personal to physically see each individual accessing the computer room.

Since the NOC provides 24/7 operations, personnel comfort is a driving design criteria to ensure that technicians

are alert and can easily access the critical information. This influences the type of furniture selected, the multiunit display systems, and possibly some level of natural lighting provided.

Even though the roles of the technicians within the NOC are primarily IT related, it is recommended that the building management systems (BMS) have monitoring capability within the NOC as well. This will enable the technicians to have an understanding of the building system status in real time. The BMS should not have control functionality within the NOC.

23.10.2.3 Entry Way

The entrance into the data center should have a physical security station to monitor and control all access to the facility. Visitors and outside vendors should have to sign in and verify the need for them to gain access to the computer room. No access to critical spaces should be allowed without proper authorization past the main entrance into the data center.

23.10.2.4 Support Staff

Support staff that directly manages the daily operations of the data center will have their offices or work space within the data center space. Data center support staff may consist of:

- Data center manager
- Data center facilities manager
- Data center facility engineers and technicians
- Data center shipping/receiving clerk
- Data center security
- NOC personnel

IT network or system engineers and administrators are not necessarily located within the data center. The IT personnel may be located off-site from the data center with remote access capability.

23.10.2.5 Electrical Room

The electrical rooms should be located adjacent or in close proximity to the computer room to minimize the lengths of copper feeders from the electrical distribution to the ITE within the computer room. There are significant quantities of power circuits feeding the ITE, and minimizing the feeder lengths between the electrical rooms and the ITE within the computer room helps reduce installation costs.

The size of the electrical room is directly related to the ultimate design capacity and the level or redundancy of the electrical distribution. When redundant electrical distribution is required, it is recommended that these rooms be positioned within the data center with as much physical separation as possible to reduce common modes of failure.

23.10.2.6 Battery Room

Battery rooms are recommended so that the battery systems are contained within a dedicated space. Wet cell batteries required dedicated battery rooms with special ventilation requirements to meet building codes. Other battery technologies may also require dedicated battery rooms and/or special ventilation depending on the total quantity of battery acid within the battery system or local building codes.

23.10.2.7 Mechanical Room

The mechanical equipment room requirements vary depending on the type of cooling technology used. The water-based cooling system that incorporates a chiller system requires sufficient space for the chillers, pumps, and piping. The mechanical equipment room should be in close proximity to the computer room to minimize the routing of piping through nonmechanical spaces between the mechanical room and the computer room.

23.10.2.8 Storage Room

Data centers have a need for storage rooms to support two different functions. A storage room is required for facility-related spare parts. The spare parts that should be on hand include belts, filters, and other general maintenance-related items. A storage room is also required for IT systems that include temporary placement of high value equipment prior to deployment in the computer room, such as spare network line cards, interface cards, network modules, optical interfaces, power supplies, and critical components with higher failure rates.

Secure storage may be required for vendor storage. Vendors that are supporting IT platforms within the computer room with defined SLA's may need to store critical components on-site in order to meet the terms of the SLA's. Even though these spare parts are stored on-site, they are still in the vendor's inventory until such time as they are required to be installed in the owner's IT systems. Therefore, the vendor may require a secure storage space to ensure that their high value components are securely stored. This vendor storage may not need to be a dedicated room, but simply a secured space, closet, or shelving within a larger storage area.

23.10.2.9 Loading Dock/Receiving

The loading dock should provide protection from the elements so that the delivery of high value equipment is not exposed to rain or snow during the receiving of a shipment. It is recommended that the loading dock have a secured

entry between the loading dock and the rest of the data center space to ensure that only authorized personnel can gain access from the loading dock to the rest of the data center. The loading dock should be sized so that there is sufficient space to temporarily house all the equipment from the largest anticipated delivery at one time. Once the high value equipment is received and the loading dock overhead door is closed and secure, the equipment should be moved into an adjacent secure staging space.

The staging space is where the packaging will be removed from the equipment. All packaging material should be placed in waste containers, helping to ensure that cardboard dust does not enter the rest of the data center facility.

It is recommended that the route from the loading dock to the staging space, burn-in room, equipment repair room, to the computer room have the floor at the same elevation. There will be several technology refresh occurrences throughout the life of the data center facility. Each refresh requires the delivery of new equipment and legacy equipment being shipped out; therefore it is preferred that no ramps or changes in elevation be required as this introduces risk and increases the difficulty when delivering the high value equipment.

23.10.2.10 Burn-In/Equipment Repair

A burn-in or equipment repair room is recommended so that the IT equipment can be initially powered on and tested prior to being placed inside the computer room. This ensures that the equipment is not defective or will cause a short circuit within the critical computer room space. A separate dedicated UPS system should be considered for the burn-in and equipment repair room to ensure that a burn-in process is not disrupted due to a power utility outage. The UPS circuits for the burn-in or equipment repair room should not be fed from the main computer room UPS.

The burn-in or equipment repair room may be incorporated together with the storage room depending on internal operating procedures. The combined storage, burn-in, and equipment repair room would need to provide sufficient space to support all these functions.

23.10.2.11 Security

The security space requirements include space for the security personnel to monitor and control building access and a space to support the security systems.

The security personnel space should be at the main entrance into the data center to control access into the building.

The security system space can be a dedicated secure room with ITE racks or cabinets housing the security systems such as access control and CCTV monitoring. The security systems are critical to the data center operations, and as such the power circuits should be fed from a UPS source.

Other critical building systems that require rack-mounted systems that are not managed by the IT department may also be placed within the security systems room. Other building systems may include servers supporting the HVAC control systems or the BMS.

23.11 CONCLUSION

The data center is a complex combination of facility systems and IT systems working together to support the critical business applications. These systems do not function in isolation from each other, and should be designed with a methodical coordinated approach. A design or operational change in one system can have a cascading effect on numerous other systems.

A data center project begins with understanding the IT applications and supporting IT platforms. The process continues with coordinating the facility requirements, the IT network architecture and topology, and the computer room layout of IT and non-IT equipment and is completed with the business applications migrating to the new platforms supported by all the critical data center infrastructures.

FURTHER READING

ANSI/BICSI 002-2019. Data Center Design and Implementation Best Practices Standard.

ANSI/NECA/BICSI 607. Telecommunications Bonding and Grounding Planning and Installation Methods for Commercial Buildings; 2011.

ANSI/TIA-942-B. Telecommunications Infrastructure Standard for Data Centers; 2010.

IEEE 1100-2005. (The IEEE Emerald Book), Recommended Practice for Powering and Grounding Electronic Equipment.

ISO/IEC 22237 Series. Information technology—Data centre facilities and infrastructures.

NFPA 75. Standard for the Protection of Information Technology Equipment; 2017.

NFPA 1600. Standard on Continuity, Emergency, and Crisis Management; 2019.

UL 60950-1 2007. Information Technology Equipment: Safety—Part 1: General Requirements.

24

MECHANICAL DESIGN IN DATA CENTERS

Robert McFarlane
Shen Milsom & Wilke, LLC., New York City, New York, United State of America
John Weale
The Integral Group, Oakland, California, United States of America

24.1 INTRODUCTION

Data center mechanical design is not inherently complex, but the requirement for high reliability combined with very obvious (and expensive) failure if it is not met adds a degree of challenge not seen in common mechanical design. Further, the rapid increase in heat density produced by ever more compact and high performance IT equipment (ITE), combined with the drive for energy efficiency, results in a challenging balancing act for the mechanical designer. Against this high-stakes design background, traditional design has leaned heavily on repeating proven legacy designs—often at the expense of innovation that can improve reliability, flexibility, cost, operating efficiency, and other aspects of design quality. The objective of this chapter is to acquaint a mechanical designer with data center design and give them the technical grounding required to move beyond replication of proven, yet often obsolete, designs and into creating optimized solutions that meet the unique requirements of their clients.

The best mechanical designs for data centers show not only skill in system design but also a clear understanding of the fundamental purpose of a data center: to make money. A careful investigation of the design criteria and consideration of their impact on the design help to best serve the actual needs of the client. But, surprisingly, this is often not done. The reliance on reusing old, "proven" designs is often used to justify doing only a cursory investigation of the needs of the current client. Some level of assumption is required to maintain the flexibility to accommodate future, unknown ITE requirements, but the needs of the initial ITE set and operations should be evaluated for each individual project.

The system configurations and equipment used in data center design should be familiar to experienced mechanical engineers, but there are a number of specializations made to adapt them to the needs of data centers. Equipment is configured to provide the high reliability required by the data center, in addition to serving the sensible-only nature of the dominant internal loads. System configurations are designed to accommodate the point source loads of ITE with various approaches used to provide cool air to the intakes while reducing recirculation of the hot exhaust.

How the design process for data centers fits into the traditional design stages and milestones is discussed at length. Consistent communication with the design team and owner is important in any project, but the high costs and critical nature of mechanical systems for data centers increase the need for clear and direct communication.

With a strong grounding in the equipment, system configuration, and design process used for data centers, a discussion of current best practices offers a springboard into the final subject—future trends—that is not a conclusion to the chapter but rather where the dialogue is handed off to the dynamic world of practice.

24.2 KEY DESIGN CRITERIA

There is no single type of data center. While it is feasible to design a generic "standard" data center based on common assumptions alone, the best balance of flexibility and first

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

cost requires careful collection and evaluation of specific client requirements including the following.

24.2.1 Reliability

High reliability is a, and often the, critical criterion of data center design. As it has a wide impact on the mechanical design, the level of reliability is defined early in the process. It is common that a redundancy level such as $N+1$ (no single equipment failure will result in any loss of capacity) will be an explicit part of the design criteria. Redundancy is also important to enabling critical cooling units to be turned off for maintenance, without which failures are inevitable. Sometimes, a standard or guideline will be referenced to define the reliability and redundancy requirements, or there will be an insurance company requirement document or internal client standard that must be met.

It is important for the mechanical engineer to fully understand the reliability requirement and explicitly state how it will be met—particularly if cost is driving an aggressive interpretation. For example, a common area of interpretation and compromise is on how an $N+1$ reliability requirement impacts chilled water piping design. Some clients will require two independent chilled water loops for the absolute highest redundancy (a burst pipe does not bring down the cooling system), while others accept a single piping system with parallel paths and valves to allow any segment to be bypassed (planned piping work for repair of slow leaks and planned maintenance can be performed without interrupting the system operation). These two approaches offer different operational capabilities and first costs.

This question of whether chilled water piping requires redundancy or not illustrates where the standard data center approach to reliability—requiring a constant redundancy for all components—is a gross simplification: the probability of a pipe failing is far, far less than the probability of a complex device such as a chiller failing, yet it may be given the exact same redundancy design requirement ($N+1$, $2N$, etc.). Yet the alternative, a detailed analysis of the actual probability of failure based upon the component probability of failures, is simply not done as a normal part of mechanical design.

The specific reliability requirements have a large impact on system cost and determine if the final product meets the client's needs or is a catastrophic failure. Fully understanding all the details of the redundancy requirement and communicating all of the implications of it to the client in a clear documented manner is an important task that spans all design phases. It is usually not a difficult-to-define issue, but due to its wide-reaching impact, it is not an aspect that should be rushed through with any assumptions.

The project location can have a significant impact on design for reliability. Designing for tornado or hurricane resistance can have a large impact on mechanical design, often leading to hardening approaches ranging from bunkered dry coolers to cooling towers protected behind (very low free area) armored louvers. Client concerns over local forest fires or even objectionable odors from industrial neighbors can limit the value of air-side economization. Local water consumption restrictions during drought years or code inspectors' proclivity to shut down equipment feared to harbor excessive *Legionella* (particularly in Europe) may present nonnegligible failure modes that influence design.

The importance of the *appearance* of reliability can be missed by mechanical engineers. Data centers that serve external clients (colocation, or colo, facilities) place a high value on a marketable design. Due to the high demand for reliable and abundant cooling, the mechanical system often is a key part of the sales pitch. Even for owner-occupied data centers, a failure can prove very costly for a company, and nontechnical executives are often quite interested in the reliability. These critical clients are not trained engineers; the system design needs to not only be highly reliable in fact but also easily assure all the nonengineering stakeholders—from the marketing manager to the chief financial officer—that it is highly reliable. This can be a powerful driver toward using legacy design approaches, even when such legacy designs are not the technically most reliable option. The best designs appreciate the importance of appearances without compromising on providing optimized design. Close coordination with the architect and client is needed to ensure this common soft, but critical, design requirement is defined and met.

24.2.2 Security

Another common characteristic of data centers is security. As a visible aspect of reliability, security measures can become closely intertwined with marketing. Security requirements are usually relatively simple design parameters to accommodate as long as they are identified during the appropriate phase of schematic design. Adding the equivalent of security bars to an exterior air economizer's louver bank during construction can be an expensive and embarrassing consequence of neglecting security concerns. Space pressurization control, particularly when air-side economizer is implemented, can also be a major security issue if overpressurization results in doors not closing properly.

It is not unusual that large data centers desire an anonymous curb presence that does not advertise the nature of the facility's use. Architecturally, this means an exterior treatment without signage. Typically, this only impacts the architect and signing scopes, but in some cases, it may dictate the placement and screening of exterior equipment.

24.2.3 Safety

Unlike some critical facility designs, such as chemical laboratories, there are few opportunities for the mechanical engineer to kill people by the particulars of their data center design. Fire control systems, including dry extinguishing systems that use

a gas to smother fires, are one area with serious life safety implications. Exiting requirements can also trip up some air management schemes that may drop curtains or have other obstructions in exit paths during a fire event. Close attention should be paid to the fire code requirements,[1] which are continuing to evolve to catch up to the current state of data center design.

Worker productivity can suffer as high-density data center facilities come to resemble industrial facilities. Very effective air management designs can result in some portions of the data center operating at hot exhaust air temperatures in excess of 95°F (35°C). Often, the high-temperature heat exhaust paths correspond to spaces that are occasionally occupied by workers installing cabling or performing other tasks. Design consideration needs to be given to accommodating these workers (and meeting applicable OSHA codes). Noise limits can also be a concern, although the operational accommodation for a high noise space is simply ear plugs, while an excessively hot space may require operationally intrusive frequent mandatory worker breaks.

24.2.4 Aesthetics

As mechanical design approaches evolve, the interior appearance of data centers can vary dramatically from the traditional. This is usually a problem for clients. Data centers that need to attract rent-paying occupants want tours of the facility to immediately project an image of a traditional—read highly reliable—data center. Internal data centers, being a large investment with banks of high-tech equipment and intriguingly blinking lights, are also popular tour material for executives wishing to show off their organization's technical prowess. While these groups are rarely at the design table, their desires are ignored at the designer's peril.

The client's expectation for the space appearance can be difficult to define for several reasons. A highly placed executive rarely sits at the table during schematic design but can force an 11th-hour design change by a brusque condemnation of the appearance of hanging curtains for containment after seeing a design rendering. Or a concerned question about how it can be a data center if it does not have a raised floor from a trusted facility manager can exert pressure to completely redesign the airflow system. Sometimes, clients delay raising concerns due to some embarrassment about bringing up appearances during the early-stage design discussions filled with concerns about important technical issues like kW capacity, tons of cooling, redundancy, and tiers—but that early design stage is exactly where concerns about appearance should be raised. Good visual communication, ranging from renderings to rough sketches, dealing specifically with the system's appearance during initial selection is important to keep aesthetics from upsetting the design process. It would be wise to insist on Revit or a similar 3D design program, rather than conventional CAD, for the production of design drawings. This can provide a clearer picture to non-engineers, but also ensures coordination of the large and complex mechanical systems with the electrical, cabling, fire protection and lighting systems that often need to occupy the same space.

24.2.5 Flexibility

The primary data center load, that is, the ITE the system is supporting, typically passes into obsolescence and is replaced in whole or piecemeal on a 3–5-year cycle. Beyond the standard need for the mechanical system to be able to support changes in load size and physical location in the space, changes in infrastructure requirements ranging from the need for liquid cooling to the equipment airflow requirements may need to be considered. These future changes typically need to be accommodated while the balance of the data center is in full operation, increasing the need for the design to provide appropriate access to installed components and consideration of system expansion in future.

The need for flexibility can be a challenge to distribution design. In an air-based system, having excess ducting capacity provides for future flexibility and can yield day-one efficiency benefits through lower pressure drop if the fans are designed to turn down. Likewise, oversized piping is also prudent. It is not unusual for air-cooled spaces to provide for the future installation of chilled water as a hedge against future cooling requirements. Flexibility also frequently justifies the addition of valved stubbed-out connection points to allow for future expansion to be possible without any system downtime or costly work on operating hydronic loops (hot taps, freeze plugs, etc.).

24.2.6 Waste Heat Reuse

An unusual aspect of data centers is that they are very reliable, constant sources of large quantities of low-quality heat. Capturing and reusing the waste heat stream may be a profitable design goal, providing free heating and often good publicity. It may also help satisfy the ASHRAE 90.1 and 90.4 Energy Efficiency Standards. The waste heat is low quality (relatively low temperature) but can be a tremendous asset if there are adjacent spaces that require heat. There is a particular synergy between laboratory facilities with constant outdoor air requirements and data centers; projects incorporating both are a treat for the designer who treasures elegant design. Heat recovery chillers or heat pump-based systems can boost the heat quality, at some electrical consumption cost, to even feed a campus loop.

Computer chips themselves commonly have safe operating temperatures over 150°F (66°C), but maintaining that chip temperature requires much lower air temperatures with

[1] Any curtain system that interacts with the distribution of fire suppression systems, in particular by dropping upon alarm to remove barriers to agent distribution, should comply with the current NFPA 75 requirements regarding not blocking exit paths in addition to other applicable NFPA standards and local codes.

traditional air-cooled heat sink design. The low quality of heat available for recovery from ITE is often not a function of the chip requirement itself, but of the heat sink and casing design—changes in either area could make waste heat harvesting more practical.

24.2.7 Profitability

Every member of the design team knows but rarely bothers to say that the primary reason for a data center to exist is to make money. This impacts the mechanical design in countless ways. There are the obvious construction budgeting exercises the mechanical engineer often supports, ranging from construction cost estimating to life cycle maintenance and energy cost. There are also less explicit aspects, such as providing a high enough level of reliability and adequate flexibility to allow for economical future expansion or providing a system that is attractive to potential tenants. A tight focus on the technical and first-cost challenges of the design is natural, but stepping back and regularly considering this larger picture, which can be more heavily influenced by long-term maintainability, flexibility, and efficiency aspects of the design, can help ensure the best ultimate design for the client.

24.2.8 Efficiency

While often overshadowed by reliability and schedule demands, efficiency is relevant in the long term since data centers make money by running computing equipment, not cooling equipment. Increasing the efficiency of the supporting mechanical system allows for more power to be dedicated to supporting the profitable equipment.

As data center design matured past the building boom of the Internet bubble, attention has turned to the electrical costs of these facilities. Dedicated data center operators may view higher efficiency as being a key competitive advantage, while a data center supporting a corporate office may look to data center efficiency improvements to meet carbon emission reduction goals. Large data centers may find their growth limited by the amount of power available from the utility and look to mechanical efficiency to free up capacity for ITE expansion.

A common metric used to help evaluate the efficiency of a data center is the power usage effectiveness, typically referred to as the PUE. This metric was developed by The Green Grid (TGG), and has become almost universally accepted as the way to monitor the energy efficiency of an operating data center. It is not, however, a design metric. In the design stage, the ASHRAE 90.4 Energy Efficiency Standard for Data Centers should be used (see Chapter 11.) Designers should be well acquainted with the PUE, however, and the design team should ensure that monitoring capability is included that enables owner/operators to, at the very least, monitor their PUE. More sophisticated monitoring is also available in Data Center Infrastructure Management (DCIM) systems (see Chapter 33) that may require remotely readable sensors for temperature, flow rate, humidity and the like (see Chapter 10.) The PUE is defined as:

PUE = (Total power used by data center) / (Total power consumed by IT equipment).

By definition, the theoretical best possible PUE is 1.0. There are occasional debates as to whether a PUE of less than 1.0 is possible by taking credit for recovered heat, but credit for that type of approach is captured through use of a different metric. The basic PUE is the total facility electricity usage as measured by the utility meter divided by the uninterruptible power supply (UPS) system delivered electricity, but the precise application of the PUE calculation still varies in practice.

It should be noted that the PUE metric is meant only to monitor efficiency in a given facility. It was never meant to compare the efficiencies of different facilities. Although when the PUE metric was new, it was often used quite loosely, occasionally quite incorrectly, and remains an ambiguous term in design practice. There was always question as to whether the annual average PUE or peak day PUE is being discussed, and local climate conditions can result in comparisons of facilities being very misleading. The exact same data center design in cool Oregon will have a better PUE than if it were in warm Florida. There were also difficult questions in where to draw the line between the ITE and the supporting mechanical equipment, particularly in innovative system approaches. For example, if the UPS is incorporated directly into the IT power supply in the IT server box itself, does that move the UPS loss from the numerator to the denominator? What about the opposite, a system that breaks the power supply out of its traditional place in the IT server box to consolidate in a large, potentially more efficient, centralized power supply? Small fans within the IT are considered part of the IT load, yet a design that measurably increases their "computing equipment" power consumption results in an "improvement" in PUE but a net increase in facility power usage. Cabinet–mounted cooling fans that operate on the UPS also become part of the IT load in the equation using the basic method of UPS output power.

In order to reduce ambiguities, TGG revised the PUE metric to include four different measurement methods; PUE_0, PUE_1, PUE_2 and PUE_3. In essence, PUE_0 is the original PUE – namely an instantaneous power measurement made at the utility meter and UPS output, or equivalent. PUE_1, PUE_2 and PUE_3 are long-term, total energy measurements, each made with increasing accuracy, with PUE_3 requiring power usage data to be delivered directly from the IT hardware. Today, when stating a PUE number, the PUE subscript should always be used to clarify how the measurement was made. (See Chapter 11 for a discussion of the joint ASHRAE TC 9.9/TGG Datacom Series book "*PUE: A Comprehensive Examination of the Metric*".)

In short, PUE is the accepted metric for data center owners to evaluate their facility efficiency. It is the responsibility of the mechanical. It is the responsibility of the mechanical

designer to ensure their client understands the ramifications of design decisions on not only the PUE but also the underlying efficiency of the facility.

24.2.9 Design Standards and Guidelines

Clients often have guidelines or standards that they wish to meet, such as conditioning to meet an ASHRAE TC9.9 Class rating, meeting an Uptime Institute Tier rating, or a legacy internal design criteria document. It is important that the design engineer understands both, the standard and *why* it is being sought. In some cases, meeting a standard may be a matter of precedent—"that's how we did it last time"—while in other cases it is a hard insurance requirement that will be fully audited prior to profitable occupancy. Clearly defining the driving requirement with the client at the start of project can help focus design effort and resources toward meeting the underlying objectives of the client.

24.3 MECHANICAL DESIGN PROCESS

Three common design styles can be summarized as implementation, optimization, and revolution. They all follow the same process but influence it at every step. ASHRAE Standards 90.1 & 90.4, which are or may be adopted as codes in many jurisdictions, may also influence the design process and choices.

An implementation design relies heavily upon using off-the-shelf data center systems and very mature configurations. For example, locating a handful of computer room air-conditioning (CRAC) units around the perimeter of a data center (server room) with an underfloor supply plenum and through-space return air path is an implementation design. This approach allows for quick design of a high-reliability space but often falls prey to the colloquialism that you can have it fast, cheap, or efficient—pick two. For a small or temporary data center, the small investment in design and integrated controls can make this an attractive option. In some cases, the entirety of the system design can be adequately, or at least at low cost (design cost, not equipment cost) and quickly, provided by the equipment vendors.

An optimization approach evaluates several different options and requires significant engineering calculation to implement. The system type and equipment will ultimately copy an existing design but be tweaked and optimized to best fit the current client requirements. Chilled water or glycol water systems often fall into this category, as do systems with air distribution more complex than underfloor with through-space return. Central control systems and evaluation of several different system types during the schematic design phase would be expected. This is what most mechanical engineers assume as desired.

The final design style of revolution seeks an optimal design but allows it to differ from precedent. Due to the top priority of reliability, this approach is challenging but can be critical to meet an impossibly low construction budget or an impossibly high efficiency, density, or other program requirements. Common hallmarks include using systems not marketed "off the shelf" to data centers, unusual distribution systems, integration of typically independent design aspects (e.g., HVAC components designed into the racks, heat recovery for other uses on-site, custom requirements on the ITE operating envelope, etc.), and a sophisticated client. A revolutionary design requires significantly more design work, a closely coordinated design team, and a technically proficient client. It is not appropriate for every project, but as it is open to embracing the best solutions, it is the theoretical ideal.

The standard data center design path is now well worn, and while it has an unforgiving cliff next to it (there is no margin for error), its challenges lie in smoothly and economically executing the design more than creating it. There are always different components to consider, but short of the inevitable handful of unique project requirements (a tape drive room, a glass tour hallway, no roof penetrations), the design is an implementation and optimization exercise. Efficient design follows the same path but with a more deliberate effort to question assumptions and quantify the efficiency of design options. This requires more engineering time and/or a design team with a wide variety of experience. It is the questioning of assumptions—whether driven by a desire for a more efficient facility, first-cost constraints, unique site opportunities, etc.—that can combine an acceptance of risk, by design effort an acceptably minute additional risk, which can lead to a revolutionary design approach.

The system type is generally selected early and is often the single most critical choice driving the ultimate data center efficiency. Loads are based primarily on the program, how many kilowatts or megawatts of computing equipment desired, and plus the overhead of the mechanical equipment supplying cooling. Typically, program loads are known at the start of the project, although in some cases they will be refined and adjusted as further information about available electrical capacity and mechanical system loads is developed. Although total load is obviously important, the wide variation in ITE types and capacities used in modern data centers makes it important to also define "per cabinet" or "zone" loads as early in the project as possible.

After system-type selection and determination of the internal load, the design process continues with equipment selection and layout. As with any project, it is common to have some iterations of design revisions driven by the need to reduce the construction budget. More specific to data centers, there is a tendency to have design revisions to increase the cooling capacity of the mechanical system as the design team identifies opportunities to add more kilowatt of usable power capacity for ITE (internal load to the mechanical designer) into the budget.

Drawing production and construction administration phases can vary somewhat based upon the construction model

in use, ranging from design–bid–build to design–build. Beyond the delivery model, the scope and schedule of the project can also impact the phases of design, in some cases condensing phases or on the opposite end of the spectrum requiring a phase to have multiple iterations. While there is no one universal design process, the most standard is some variation of the following design phase progression: predesign, schematic design, detailed design, construction documents, construction administration, and postdesign support.

24.3.1 Predesign

There are several different ways a data center design project is initiated that can impact the mechanical design challenge. While not required, input from the mechanical engineer at the earliest stages can ensure the most efficient facility. The type of project will inform the type of mechanical system insight needed.

24.3.1.1 Greenfield Development

The least-defined designs begin with only a desired facility size and capacity goal (stated in kilowatt or megawatt of ITE capacity). If the mechanical system's power requirement significantly influences site selection or other key activities common in the predesign phase, the mechanical engineer should provide the design team with estimated design condition system efficiency as a design parameter. System efficiency varies widely, so the estimate will be only approximate but can be adopted as a driving parameter that must be met as the design progresses.

The key mechanical efficiency needed in this case is not an annual average efficiency, but rather the highest mechanical power requirement in the worst-case, peak cooling load, extreme design condition. It is this highest demand that will dictate the amount of the electrical feed that will have to be diverted from supplying ITE—also known as the reason the entire facility is being built—to supporting the mechanical system.

Data centers produce an enormous quantity of low-quality (low temperature, ranging from 72 to 100°F (22–38°C), dependent upon the design) waste heat. In cases where there is flexibility in the location of the facility, such as on a corporate campus, demand for the waste heat can play a role in the site selection.

Sometimes, potential data center sites can differ by tens or hundreds of miles, bringing an evaluation of the climate zone into the siting question. The most extreme case is where the data center can be sited in a climate zone that requires little cooling. With less geographical span, the data center may still have varying access to a large body of water that can be used for cooling, such as a river, large lake, or even a large wastewater treatment facility. A small investment of mechanical designer input can catch when such rare, but highly valuable, opportunities appear during site selection.

24.3.1.2 Converting an Existing Building into a Data Center

Data centers are surprisingly flexible in the kinds of buildings they occupy. A renovation situation can offer mechanical options and limitations that would never occur in a purpose-built facility. For example, an extraordinary amount of floor-to-floor height may be available for a data center sited in an unused warehouse, offering great opportunities for efficient airflow—or a drop ceiling with a wasted 18 in high space above it. Or the data center may be crammed in an existing office building's windowless basement with no exterior walls and only an 8 in floor-to-floor height. Large numbers of closely spaced columns are one of the most common drawbacks to adapting existing buildings because they cause highly inefficient cabinet layouts. Likewise, buildings with electrical service in a basement subject to flooding is probably a poor choice. Any of these example situations might still be adapted to successfully house a data center, but the client should be informed of the impact they will have on the mechanical system options that would be available, as well as the costs of inefficiencies or disaster mitigations. Either of these example situations can successfully house a data center, but the design team should be informed of the impact they will have on the type of mechanical system options that would be available.

24.3.1.3 Expansion

It is quite common to expand an existing data center. The mechanical engineer should assist in evaluating the existing systems to determine if there is usable capacity to dedicate to the expansion. If the operator has had a negative experience with the existing system, that can drive a change in system type for the expansion. One common opportunity with expansions, particularly with chilled water systems, is to reduce the cost of redundancy by intertying the expansion system with the existing mechanical system. There may also be an opportunity to improve the system efficiency of the existing system by integrating it with a newer, more efficient expansion system. An example of this would be a chilled water plant expansion to support new data center space that installed new, high-efficiency chillers to serve the new load and retired an existing, low-efficiency chiller system to standby backup operation.

24.3.1.4 Remodel of Existing Data Center

Remodeling of an existing facility without an expansion of footprint or cooling capacity is not very common. When reliability is the top priority, it is rare to modify anything that is not broken. The mechanical engineer should use any predesign phase to identify explicitly the driving motivation for the remodeling: solving hot spot issues, reducing energy use, capitalizing on energy reduction incentives, achieving a corporate carbon reduction goal, meeting a potential client's redundancy requirements, etc. Often, the motivations for a remodel project are typically well known, and these predesign

tasks are condensed into a single effort that generates a request for proposal document or a preliminary meeting with contractors to request a quote for services.

24.3.2 Schematic Design

There are several common objectives in the SD phase, with the priority placed on each, varying by client. The typical SD process begins by identifying the key design requirements. Creating an estimate of the load and selecting the system type can proceed in parallel to some extent, with development of initial equipment lists and/or schematic drawings of the ultimate deliverable, which will vary depending on the project schedule and objectives. When design time or budget is very short, schematic design may be combined with design development.

24.3.2.1 Objective

Identifying the objective may entail no more than a careful reading of the detailed scope document or request for proposal provided when the design team was selected. Bear in mind that most design requirements appear obvious—all clients want low cost, effective space control, redundancy, etc. *But* the relative priority and underlying operational requirements need effort to clearly define and understand. In most cases, trade-offs between design priorities must be made. A good understanding of what motivates each design requirement allows the best trade-offs to be evaluated and offered.

The objectives will drive the required deliverables. Some projects may only require an initial equipment list and material estimate to aid in a schematic cost estimate required to assess the business plan. In this case, while schematic diagrams and block equipment layouts are likely to be performed to some level to ensure an accurate equipment list, the traditional schematic drawings might be omitted entirely from the final schematic design deliverable in lieu of a text narrative description. Or data center expansion may place speed over early cost estimation and prioritize advanced drawings over equipment sizing and listing. Understanding the specific needs of the project allows the mechanical engineer to most efficiently allocate their time.

24.3.2.2 Define Space Requirements

Historically, data centers were kept quite cool – as low as 55°F (12.8°C) in early mainframe computer rooms. Sweaters and jackets were standard dress. Relative humidity was kept at 45% – 50% to keep the punch card feeders operating smoothly. By the 1980's, most paper was gone from data centers, but tape still required fairly strict operating environments. Manufacturers allowed temperatures as high as 72°F (22.2°C), but most data centers were still kept much cooler out of concern that higher temperatures might increase equipment failures. Further, almost all cooling was underfloor, raised floors were generally only 12" (30 cm) or less in height, and were often crowded with cables, making it difficult to deliver uniform cooling. Cooling equipment also had unsophisticated controls, so one unit might be cooling and dehumidifying while a unit close by was heating and humidifying.

Low temperatures and high humidity continued as "de facto" standards for years, partly because there was no industry-accepted standard, but also because designers often copied prior data center design documents. Operators were often also more comfortable with "what has worked for years". Unfortunately, many facilities still use legacy designs, but the 2007 report by the United States Environmental Protection Agency (EPA) on data center power consumption started a dramatic change. It said US data centers were using more than 2% of the total energy consumption of the country, and could easily double that in five years, outstripping both power grids and generating capacity. The industry stepped-up, improved equipment, encouraged better engineering, and the dire predictions were avoided. A major contribution came from one of the largest professional societies in the world - ASHRAE, the American Society of Heating, Refrigeration and Air Conditioning Engineers.

ASHRAE Technical Committee TC 9.9, for the first time in history, established operating conditions for computing equipment that are accepted by every major manufacturer (see Chapter 11 for an in-depth review of ASHRAE contributions to the data center industry.) Based on their work, a maximum normal operating temperature of 80.6°F (27°C) is recommended as the maximum allowable inlet air temperature to computing hardware. This made a significant reduction in mechanical cooling energy, but has also enabled many more hours of "free cooling" (cooling with outside air) in many climates, which was actually the main purpose for establishing the higher temperature norms. ASHRAE also recommended controlling humidity on Dew Point (DP) temperature rather than Relative Humidity (RH) which enables much more uniformity, further reducing the enormous energy needed to humidify and dehumidify air. Their more recent work has also shown that relative humidity levels as low as 8% are acceptable in modern data centers where good grounding practices are followed and tape media are no longer used. This is also covered in Chapter 11.

In existing data centers, a common practical reason for continuing to use lower space temperatures is to compensate for a failure of airflow management. With poor or no airflow management, which is a common situation in older data centers, at some point in the room, the hot air exhaust from an IT rack is recirculated back into the intakes of other, or even the same, ITE. This recirculation results in a localized hot spot that could eventually cause equipment damage. Another problem with poor airflow management is air bypass, where cooling air moves through open spaces in and

between cabinets without passing through ITE. This is not only wasteful of expensive cooling air, but also lowers the temperature of return air, which further upsets air conditioner control. Simply lowering the temperature set point of a data center is a common reaction to hot spots, and it does help, but it is terribly wasteful of energy.

Another reason sometimes given for maintaining a low space set point is to provide a reservoir of cooling to the space in case of equipment failure. But the hope for a cooling such as buffer provided is far less than one expects. When the amount of "stored" cooling is calculated, it is found to offer a negligible safety buffer to all but the most lightly loaded facilities.[2]

There are other rationales for designing and operating a data center at a low-temperature set point, including client expectations, but a proper assessment of the actual set point needed can often yield significant first-cost savings, higher space capacity, and lower operating costs for an educated client.

The humidity set point is another area where the requirement is often based on custom and an assumption that "tighter is better," Too aggressive a humidity control band can actually harm data center's reliability. Humidifiers are a potential source of catastrophic failure due to their water supply. They should be minimized or even eliminated if possible, particularly considering the latest ASHRAE recommendations on humidity levels. They also carry a significant operational cost, including maintenance and power consumption, which in extreme overdesign conditions can even impact emergency generator sizing. The true need for humidity control should be carefully evaluated, and modern guidance on static control considered (i.e., that humidification is not so necessary for protecting components from electrostatic discharge as was once thought.)

24.3.2.3 Cooling Loads

One characteristic aspect of data centers is that their cooling load is almost entirely internal space load, the heat generated by the ITE inside. The space load is defined by the amount of ITE that the client wishes to house. Historically, this was discussed in terms of watts per square foot or per square meter by the mechanical designer, and that's fine for total equipment sizing. But today's loads are no where near being uniform, so the industry norm is now to estimate on watts per cabinet or per zone, then confirm realities with a watts per unit area average.

[2] For a space with a 9 ft high ceiling and 20 W/sf ITE load, based on the thermal mass of air, the temperature would rise 5–10°F/min in the absence of cooling; consideration of thermal mass such as the floor gains little due to a low rate of heat exchange. In practice, unless loads are very low, the benefit of overcooling a data center to gain some margin for error in a total failure situation is illusionary.

Unlike an office building or other common commercial designs, data centers have an industrial load profile—almost flat, 24 hours a day. "No windows" is a common design requirement for data center spaces, removing the largest source of shell cooling, which is usually already a negligible fraction of the load. This also eliminates the possibility of moisture condensation on glass in winter climates. Likewise, very little outdoor air is supplied for these transient occupancy spaces, with just enough for pressurization if code justification can be made to deem the space normally unoccupied. Typically, the heat generated inside by the ITE is an order of magnitude or higher than even the peak heat gain through the shell, so there is negligible impact from the shell. All these result in a very flat load profile.

Shell loads can be calculated using traditional load analysis methods but are really only explicitly determined out of the abundance of caution that underlies all critical facility design; in a typical data center, the peak shell load is negligible (far less than the sizing safety factor) and could be assumed to be zero with little impact on the design. For designers less familiar with data centers, understanding the nature of this load can have a surprisingly wide impact on the design. Not only are load calculations radically different, but also the stable and efficient part-load performance of the system takes on a higher priority. Unlike an office building, most data centers have a minimal increase in cooling load on even the hottest summer day, and they are also typically designed to have redundant capacity at all times; these two characteristics combined result in them very rarely (never by typical design) operating cooling equipment at full load.

The design load assumptions should be regularly checked. It is not uncommon for them to change significantly as business plans change or additional information about utility power availability is discovered. It is often convenient to coordinate with the electrical designer who often is the first informed of internal load changes, since they have a direct impact on the sizing of the facility's extensive infrastructure.

In some large data center projects, there is a limited electrical capacity economically available to the site, and this limited feed capacity makes the assumed efficiency of the mechanical cooling system a critical factor in determining the power available to run ITE—every watt in cooling equipment is a watt less to run the profit-generating ITE. In this case, the mechanical system efficiency can become a critical design parameter that must be met to maintain the integrity of the business plan, and the designer needs to regularly calculate it and defend it (typically from cost-cutting exercises) accordingly. An economically efficient design will also be likely to meet the energy efficiency requirements of ASHRAE Standards 90.1 and 90.4.

A final aspect of the design is the exterior design conditions. While they typically have little impact on the cooling load that has to be delivered (which is overwhelmed by the large ITE internal cooling load), external conditions do sig-

nificantly impact the capacity of the heat rejection plant. As a critical facility with 8,760 hours operation, extreme outdoor climate design conditions are often used rather than the more typical 1% or even 0.5% conditions. These can be significantly higher than the standard conditions used for mechanical design and will impact the sizing (and cost) of heat rejection systems. The outdoor design condition needs to be appropriate for the project and clearly documented for communication to the client.

24.3.2.4 System-Type Evaluation

The mechanical system type may not be completely set during the schematic, but a preferred approach is often selected. The key parameters of the system type that should be evaluated include the cooling medium, delivery path, heat rejection method, and airflow management. The objective to selecting a design basis is primarily to assist in cost estimation, define the footprint requirement, and evaluate the efficiency. The very high-level selection of these system parameters can set the efficiency of the final product and have a major impact on operating energy costs. The system-type selection can also impact architectural parameters including ceiling height, external equipment space, and interior layout.

The selection of system type has an enormous impact on the mechanical design and the capabilities of the final product. During schematic design, different system types should be assessed for the ability to meet the design objectives. Beyond the main requirements, details like lead time requirements in a fast-track project and cost impacts on other disciplines such as the requirement for a raised floor or a larger emergency generator system should be noted and considered. A thorough high-level coordination need not take a lot of time, but can often be skipped if not made an explicit part of the process.

Some sophisticated clients may require that the data center meet a very specific efficiency metric or require a formal value analysis of multiple system options. The choice of system type and options will heavily influence the ultimate efficiency, so the relative efficiency of different options can be a key deciding parameter in scoring what system is the best option for the site. Even when the client does not require it, in light of the magnitude of energy consumption over the life of a data center mechanical system, the relative efficiency of differing system types should be considered.

24.3.2.5 Footprint Evaluation

Data center mechanical systems have a significant impact on the architectural program, layout, and costs. At this stage of the design, a full layout is impractical due to the fluidity of the design, but rough estimates of the footprint of the major pieces of equipment, general paths for piping and ducting, distribution concepts, and machine room spaces are needed to coordinate the program requirements with the architect. The largest pieces of the system (including air supply mains if applicable) can be represented as rough rectangular blocks to quickly generate layout estimates. Any equipment located on the data center floor is of particular concern as it subtracts from the program space available to house ITE. The space required for airflow management and ducting is another large element of data center mechanical systems. Particular attention should be given to routes for under-floor liquid piping since large, insulated pipes can be result in significant airflow blockages if not carefully coordinated with the total infrastructure.

Significant cost and efficiency benefits can be realized by closely coordinating the architectural design with the mechanical system. The method of coordination varies greatly, from three-dimensional (3D) computer models to hand sketches on tracing paper, but regardless of the method, they all serve to allow the mechanical engineer to communicate to the architect the size of the system, the ideal layout, and the compromises that are implicit in the proposed actual layouts. All designs have compromises, and it is important to consciously identify them and use the design team's combined expertise to quantify them as much as schedule and budget allow.

Savings from architectural integration tend to be most significant in large, dedicated data center spaces. Beyond the traditional use of a raised floor, there can be opportunities to optimize and distribute airflow using architectural elements such as a ceiling plenum or partitioning walls. Cost savings may be realized by placing the mechanical plant on a subfloor below the data center or by using exterior rooftop-mounted air handlers to reduce the conditioned space that must be built (in temperate climates where maintenance would not be hindered). Most designs can benefit from centrally locating utilities to shorten the lengths of the largest mains and offer opportunities to reduce the power requirements from the fans and pumps. Some system solutions, such as air-side economization, are heavily dependent on the architectural configuration to produce a functional system. Some products allow for integration of air handlers into the building structure, for example, by replacing an exterior wall with built-up air handlers with easy access to exterior air for economization.

Smaller data centers can also benefit from close integration with the architecture. A common potential benefit is the harvest of low-quality heat from a data center housed in an office building to warm adjacent office space during winter. There can also be low-cost-efficiency opportunities that can be realized by utilizing an adjacent office mechanical system during nonbusiness hours to provide air-side economization to a data center or cost savings from using the office HVAC system as a redundant cooling source (with the clear communication that the office space will sacrifice cooling to support the data center when necessary). Opportunities in these cases are typically limited by the design cost and a match of

the humidification requirements between the office and the data center, with significant custom engineering to realize a workable interplay between the spaces.

24.3.2.6 Code Evaluation

As with any project, an overlooked code requirement can become a late-in-the-design land mine. Code review should be part of every phase of design. Different localities will face different code challenges and inspector expertise in the area of data centers. An open evaporative cooling tower may be a good standard design solution in California but wrought with code implications in the United Kingdom where a Legionnaire's disease scare can result in shutdown orders for all cooling towers in miles. Major code impacts like this are rare and should be familiar to the design team based on past experience; explicitly identifying and documenting code concerns are important parts of schematic design.

Specialized fire control systems that use a gaseous fire suppression agent or dry pipe preaction systems are common. While the fire suppression system is typically designed by a fire protection engineer, purge fan requirements, isolation dampers, and distribution piping often require coordination and assistance from the mechanical designer. Management of airflow is a critical task for high-density data centers, and associated partitions may impact the fire system design. As a longer-term concern, the future flexibility of a design should be evaluated in light of the fire control code requirements. For example, the use of flexible curtains to control the airflow of hot spent air is currently a common air management approach to allow ease of reconfiguration, but the curtains can interfere with the dispersal of fire extinguishing agents and require integration with the fire control system.

In some areas of the country, typically those with stringent energy efficiency written into the local codes, utilities offer incentive money to encourage more efficient design. This opportunity is only available in a limited number of areas, but it is worth checking with the local utility as early in the schematic as possible to identify any incentive money that may be available to invest in more efficient systems and protect them from deletion during the inevitable efforts to make budget later in the design process.

24.3.2.7 Prepare Deliverables

Deliverables for the schematic design phase will vary depending upon the client and design team, but at a minimum should serve to document the design assumptions, compromises, and recommendations developed during the schematic design phase. The most common deliverables are the same as any other design project: a design narrative and a set of schematic drawings. But deviations from these common deliverables can be called for in some cases. Due to the complexities and costs of data centers, it is not unusual to see "schematic" phase documentation that looks more like "design development" phase work.

In a cost estimate-driven project, drawings may be omitted entirely in favor of a more detailed narrative with an equipment list. The reasoning behind this is that when the primary objective of the schematic design is to develop a construction cost estimate, traditional schematic design deliverables like system-level single line drawings or block layouts of the main mechanical spaces are of little value; the design budget can be better spent developing a more detailed list of the basis of design equipment, area required, feet of piping, and pounds of duct. For the most efficient design process, the cost estimator will be accessible throughout the schematic design to make clear what they need for the estimation exercise and to highlight key design areas with a high cost sensitivity. However, the importance of realistic budgeting in any size or type of data center project cannot be over-emphasized. Data centers can easily be more than ten times the cost per unit area of conventional office space, which is often disbelieved until too late in the process, resulting in compromises to meet budget that can negate the reliability goals for which the data center was supposed to be built.

A more speculative developer-driven project may focus on generating a marketing piece out of the schematic. Developers may require a deliverable with attractive cartoon-style sketches of the proposed design, a layperson-level narrative of its advantages, and little focus on equipment lists and sizes. While a properly sized pump with the perfect breakwater distance balancing efficiency with longevity is a beautiful thing, few nonmechanical engineers care; a nice 3D rendering of the space in full color is more important if the objective is to attract client deposits or sell a building owner on a project.

Because the expected SD deliverables can vary, it is important that the mechanical engineer communicates with the design team and (often indirectly via the architect) the client to ensure the correct materials are developed. Regardless of the primary materials required for delivery, a document that clearly states the design assumptions and limitations must be generated. While usually part of the design narrative, it could be a separate memo to the design team lead outlining parameters including the design load, space temperature requirements, and system requirements, such as the need for water piping on the data center floor or the exterior space required by dozens of independent air-cooled CRAC condensers.

24.3.3 Design Development

In this phase, it is expected that the system-type selection is finalized and equipment sizing begins. Layouts of the mechanical spaces and distribution are made and coordinated with the architect. The ducting and piping mains are defined and documented in drawings to allow for clear coordination. Controls should be considered, although it is

common (if often unwise) to do so only to a cursory level. Code authorities may be directly contacted to test any code interpretations and, in jurisdictions unfamiliar with data centers, begin an education process. Cost estimating, and the closely associated efforts to reduce construction costs to make the construction budget, often starts in earnest during design development. As noted above, data center schematic design (SD) documents can often be closer to the level of conventional design development (DD) products, and DD documents can often contain detail normal to construction document (CD) level materials. This is partly driven by the need to resolve complex issues early, but also because the CDs must often be delivered early so the data center can be finished weeks ahead of the rest of the building, enabling IT to install and test equipment and be ready for building occupancy. In these cases, dirt and contamination control are major concerns that designs must consider.

24.3.3.1 Finalize System-Type Selection

The system type that can vary from air-based cooling of the entire room all the way to cool water piped directly to the computing equipment has wide impacts on the mechanical design. Making a firm system-type selection early is good for controlling budget by keeping to a tight schedule, but there can be tension to keep the system type flexible to accommodate changes in the layout, incoming cost information, client preferences, and other concerns. The design budget and schedule will dictate how critical it is to end system-type comparisons. The mechanical engineer should be sensitive to the needs of the client and architect but be clear on when aspects of the system type to be used need to be decided to maintain schedule versus aspects that can change later to accommodate additional information. And regardless of when the system-type selection is finalized, be aware that a high cost estimate will almost always lead to a reopening of the discussion, so some amount of design rework should be assumed if the cost estimation or "value engineering" exercise is planned late in the phase.

Once made, the system-type selection should be explicitly documented by an email, a memo, or incorporated in a progress drawing set sent to the entire design team to help in coordination. There is little that hurts a designer's budget as much as a late change in system type, for example, from air-cooled CRAC units distributing via a raised floor to water-cooled built-up air handlers using overhead ducting and plenum space. When a base system selection is made, declare quite explicitly to the team that it is a foundation assumption and changing it could result in additional cost and delay. The mechanical system also impacts most aspects of the data center design. Clear coordination of the selected type is important enough to warrant the redundancy of documenting the final decision even if all design fields were directly involved in it.

24.3.3.2 Value Engineering

As with any project, there is a need for the final design to be constructible with the budget available. Commonly referred to as value engineering, this exercise of cutting construction budget from the design is becoming more common in data center projects as they become more of a common commodity space. The large size and expense of the systems within the scope of the mechanical engineer typically require their significant participation in value engineering.

When investigating lower-cost design options, it is important for the mechanical engineer to coordinate with the electrical engineer to ensure the client understands that an extra kilowatt used on HVAC equipment, perhaps due to the use of lower-cost mechanical equipment, is a kilowatt of generator and utility capacity not available to make money. The assessment of an alternative mechanical system or equipment option needs to take into account not only a potential reduction in the installation cost of that mechanical component but also any *increased* costs that may be incurred on the electrical system by the alternative. Impacts on redundancy, space flexibility, and expandability must be clearly defined and communicated to the client to ensure that an accurate assessment of cost-saving measures is made. Good value engineering can reduce the cost of the whole project without harming performance, but a design team myopically focused on only their own discipline's line item costs can reduce the final space's utility and actually increase the whole project cost.

24.3.3.3 Revise Load Estimate

The key component of the load estimate is the power of computing equipment that will be supported. As the design process progresses, this critical design parameter can abruptly shift. Regular communication with the design team should ensure that the mechanical designer is aware of any relevant revisions. The mechanical engineer should also keep the electrical engineer updated of any changes in the need for power to support the mechanical system, with a keen awareness that decreases in the mechanical system's efficiency can cascade into a nonnegligible need for more generator and transformer capacity.

24.3.3.4 Preliminary Layouts

Floor plans and data center layouts take shape during design development. Targeted and succinct input from the mechanical designer can ensure that mechanical concerns are met and issues such as minimizing distribution length (a cost and energy efficiency driver), providing enough space for appropriate maintenance access, airflow management, and planning for future capacity expansion are well handled.

The mechanical room layout has a significant impact on the system efficiency and operational requirements. There

are often numerous trade-offs, such as desiring a very compact footprint but needing space to allow for maintainability or minimizing first cost by downsizing mains sizing at the cost of hurting future flexibility and efficiency. Mechanical layouts should be generated as early as possible. It is easy enough to make a high-velocity air system with a very small footprint, but the future flexibility, expandability, and operational energy cost implications of such an approach are grim. Optimization of the mechanical system layout is critical.

Where high efficiency is the primary goal of the system, mechanical equipment should be laid out accordingly. Airflows and water flows inherently waste energy when they make sharp right angle turns. Recognition of this can often result in a mechanical room where equipment is located at an angle to the walls, piping is kept near the floor rather than routed over a rigid grid of aisle ways, and long radius turns and 45° laterals are common. One pipe fitter compared a particularly efficient plant layout to a sanitary sewer system—an apt comparison, since gravity-driven sanitary sewer systems are forced to adhere to low pressure drop layouts. While piping layouts to this level of detail are not appropriate in design development, the modest extra effort and (sometimes) floor space required for efficient layouts should be acknowledged and planned for if efficiency is a high priority. Air handler sizes should be optimized for 8,760 hours of power-consuming operation per year, rather than by office-based rules of thumb such as a 500 fpm (2.5 m/s) coil face velocity.

Rejecting heat from the data center to the outside is the primary task of a mechanical system. The size, type, and location of the exterior heat rejection components, whether they are a cooling tower or a louvered wall, should be identified during the design development phase and any limitations identified. For example, a ban on rooftop equipment for security or leak concerns, hardening against tornado and hurricanes, or other uncommon but critical specific project requirements need to be determined and accommodated in the system selection and layout. Aesthetic and acoustical concerns can also be factors if the facility is located in a residential area or within the line of sight of a residential area; expensive houses on the hill with a direct view of the best place for a noisy cooling tower yard anecdotally tend to house local politicians and code inspectors with sensitive hearing. In any design, the potential for disturbing noise and vibration transmission to building offices cannot be overlooked either. The data center is important, but it is there to support the business, and if it negatively impacts business efficiency, it is not serving its full purpose.

Future expansion also plays a role in determining how much space is required, both on the interior and exterior. If a future expansion path is desired, it should be an explicit project goal and be directly incorporated in design development by considering and documenting where future equipment and distribution would go to support additional load. It often is cost effective to provide some infrastructure to support future equipment, such as extending a tower structural support platform to fit more cells in the future, oversizing piping to provide future capacity, and adding empty electrical conduits when casting foundations.

24.3.3.5 Equipment Selection

The selection of equipment is an important step in ensuring that equipment exists that can provide the desired performance within the rapidly solidifying space, cost, and energy budget available.

After the system type is finalized, and in parallel with developing layouts, preliminary basis of design equipment selection should begin by calculating equipment capacities and sizes. At the beginning of design development, a detailed equipment list should be started, and the equipment schedule drawings begun. The most expensive equipment should be sized first, followed by the physically largest equipment and finally the auxiliary equipment, with the overall goal being ensuring that equipment is available that can provide the desired performance within the rapidly solidifying space, cost, and energy budget available. Items like pumps and fans can usually be approximated by calculation based on estimates of pressure drop requirements, while larger equipment such as chillers, CRACs, air handlers, cooling towers, and other similar items should have preliminary selections made to better define size, cost, and efficiencies.

The nature of the cooling load presented by a data center differs in several ways from a typical commercial office building load. The selected system equipment must be able to stably carry the design load even during design heating (lowest outdoor air temperature) conditions, at a time that office cooling plants are often shut off. The cooling system also must switch seamlessly and stably between any economization mode and mechanical cooling. In air-based systems, airflows are sized to accommodate the sensible-only load. Reheat of IT space is unnecessary.

Projects with an energy efficiency requirement to meet use the preliminary equipment selections to calculate the predicted system efficiency to ensure contract or design requirement compliance. While there are many energy modeling programs available for buildings, they are not generally useful for data center energy modeling because data center design parameters are so different from those used for offices. Attempting to insert those parameters in commonly used modeling programs will usually result in error notices." Due to the relatively simple nature of data center load (approximately flat, 8,760 hours a year), a spreadsheet calculation that uses hourly typical meteorological year data available from a number of sources or bin weather data can often be successfully used to streamline this task. System interactions should be considered throughout the design. For example, a successful airflow management system that collects heat exhaust from the ITE can increase the air-side delta T and allow for smaller

air handlers, paying for some of the first cost of the airflow management elements. Using low pressure drop plenums for air movement instead of ducting and allowing a higher temperature and humidity range in a portion and all of the data center are other system design decisions that can have far-reaching impacts on the mechanical system.

24.3.3.6 Size and Locate Distribution

The data center mechanical system exists to move heat out of the data center. Regardless of the medium it uses to do this (air, water, glycol, refrigerant), there will be a significant distribution system (ducts, pipes, or both) to move the heat around.

An air-based system will require large ducts or plenums to allow for the volume of airflow required. Within the data center footprint, plenums formed by a raised floor and/or a false ceiling are typically the most efficient and flexible method of air distribution. The space itself is often used as a plenum to move the large volumes of air needed to cool the equipment. Ducting can offer a more controlled distribution system that can avoid some code requirements regarding wiring through space used as an air path plenum, but it is often less efficient. The choice of air system can significantly impact the cost of the fire suppression system by increasing or decreasing the active volume.

Raised floors are often used to create a supply air plenum. This is a simple design approach but can run into limitations at high load densities as floor heights become economically unattractive (particularly in zones with extensive seismic requirements). If the underfloor space is shared with any other utilities, such as electrical distribution systems or data cabling, it can become surprisingly congested, resulting in inadequate airflow to portions of the spaces—close and consistent coordination with other trades is required, starting from when the floor height is initially estimated and continuing throughout design.

Raised floors are rarely used as a return air path; while having a floor plenum that serves as a return path is theoretically feasible (the buoyance effect of hot air is negligible at the air velocities seen in all but the most lightly loaded or specially designed data centers), current design practice and commercial products available only support use of raised floors for supply air.

Overhead plenums are often used for return air. In the common legacy design of CRAC located on the ITE floor using a raised floor for supply air distribution and through-space return, converting dead space above the ceiling into a return plenum is a common method of reducing mixing of supply and hot exhaust to eliminate hot spot problems, improve capacity,[3] and increase system efficiency. Code requirements on the type of electrical supply wiring and equipment allowed in the plenum space need to be considered, particularly when considering retrofit of an existing facility, along with any impacts on the fire suppression system from the added active volume.

Overhead plenums are rarely used for air supply, with ducting preferred for overhead supply. A return plenum can be combined with supply ducting to offer a hybrid plenum/ducted air management solution that does not require a raised floor.

24.3.3.7 Investigate Airflow Management

Airflow management is a critical aspect of avoiding potentially damaging hot spots in high-load-density data center design that relies on air for cooling (as opposed to cooling water to a rack-level system). The airflow management approach needs to be considered early in the design phase as it has extensive impacts on most areas of the mechanical design, including cost, effectiveness, efficiency, and system sizing. Architecture may also be significantly impacted.

The ITE housed in most data centers draws cooling air in one side and then ejects a high-temperature exhaust out the opposite side, ideally drawing air from the front and exhausting hot air out the back. While front-to-back: airflow is now the most commonly provided, some ITE most usually large network switches, are designed with side-to-side or some other airflow pattern. It is important to coordinate with IT to ensure that special racks and cabinets are provided for this hardware so that airflow is changed from front-to-rear. If such cabinets are not provided, the design of this part of the room becomes a very different challenge for the mechanical engineer.

Airflow management can take many forms, but the objective of all of them is the same: capture the hot exhaust and cool it before it is pulled into the cooling airstream of another (or the same) piece of equipment. Discussed in greater length elsewhere in this chapter, airflow management can take the form of anything from hung plastic curtains partitioning off the intake side of racks from the exhaust side of racks to distributed floor tiles with integrated and independent variable speed supply fans that vary the volume of cool air supplied from an underfloor plenum on a per-rack basis. In highly customized cases, the airflow management will likely dictate the architecture by dictating the space height or layout of spaces relative to the exterior wall.

In design development, the main priority is to determine the kind of airflow management system design that best fits the program and communicate to the other trades the impact it has on their design work.

24.3.3.8 Drawings

While drawings may be skipped in favor of a costing-targeted narrative in the schematic phase, it is rare that the design development phase does not produce drawings.

[3] The cooling of most CRAC is a function of the temperature difference between supply and return air. Improved airflow management can increase this temperature differential and increase the usable capacity of currently installed equipment.

For larger projects, there are often one or two progress sets compiled during the design development phase to assist with interteam coordination.

Drawings are developed to support costing exercises, document design progress, and aid coordination in this phase. Any coordination agreements between the disciplines, ranging from the location of mechanical rooms to the electrical capacity (or, more crudely, the motor horsepower) required for mechanical, should be clearly documented as a valuable product of this phase that could be lost if left in notebooks or buried in an email chain. Common drawings required in this phase include an equipment schedule, air-side diagram, and water-side diagram that serve to record the current state of load calculations and system selection. Layouts of plant rooms and equipment yards are also typically provided, although they are subject to adjustment during the construction document phase.

Preliminary layouts of mechanical equipment and distribution are an important coordination tool between the architect, mechanical engineer, and electrical engineer. They also serve to inform more experienced clients of the scope and type of systems they will need to support with operations staff.

Detailed drawings are primarily a task for the next design phase, construction documents, but when significant detailed design work was done in the design development phase, it is appropriate to document it. This most often occurs when an unusual system or design approach is being considered, and it must be designed to a significant level simply to verify it is a feasible option. Such design aspects tend to be defined by their unpredictability, but they could include features ranging from the suspension hardware configuration of a hung curtain air management partition to the construction details of a built-up air handler with direct evaporative cooling/humidification incorporated into a structural wall. Beyond the case of subsystems that are developed to unusual detail to prove feasibility, a significant number of generic details will be included in this phase as they are available from the designers' standard work to "cut and paste" into the project; while not necessary, this can help coordination for projects with short schedules, little communication between the design disciplines, or design team members who are unfamiliar with the proposed data center systems.

24.3.3.9 Code Investigation

Any outstanding code questions that impact the design should be settled in this phase. They may be settled in a variety of ways, ranging from verbal or email confirmation from the authority having jurisdiction to an agreement with the client representative about the interpretation being used and the worst-case cost of the interpretation being rejected. At this stage, the risk is typically design rework and the associated costs and possible delays. The implications of worker safety codes on operation should also be determined and communicated to the client. For example, as data centers move to creating hot aisles that operate at high temperatures, operators may be legally obligated to limit worker time in those areas—which can be a problem if extensive rack wiring and hookup need to be regularly performed from the hot aisle side of the IT rack.

24.3.3.10 Cost Estimating and "Value Engineering"

Supporting cost estimating efforts and investigating opportunities to reduce system first cost are often a high priority throughout the design process. Any deviations from very traditional standard design should be clearly documented for the cost estimator and reviewed closely by the engineer. The mechanical engineer should review the cost estimate for equipment type, sizing, pounds of ductwork, and other key cost elements. Cost estimating is often done by a contractor, who in the process of a cost estimate can often offer a useful viewpoint on the design's apparent constructability and the clarity of documents.

If a design change being considered to reduce cost will impact the cost of other trades, the mechanical engineer should inform the cost estimator and review the ultimate cost estimates to ensure it was accurately captured. For example, utilizing smaller, high-velocity air handlers may reduce air handler cost but need more fan power and increase the cost of electrical support systems ranging from panels to the building transformers. Whole-building impacts of this type are often overlooked in the early design cost estimates, which can lead to poor value engineering decisions being made. Some savvy clients may also request a net present value analysis to capture the operating cost impact of changes.

24.3.3.11 Controls

Control design is often left to the construction document stage. This is a reasonable strategy to avoid rework, but research should be completed and documented by the end of design development to identify the type of control system desired to ensure cost estimates are accurate and assume an adequate level of control investment to support the proposed system. Common types of control include central direct digital, independent integrated CRAC unit controls, or some combination. It is not unusual for smaller data centers to have the control system consist of the onboard controls of CRAC units—which have a very different capability and cost profile than a central direct digital control (DDC) system. The intended type of control should be clearly defined and communicated to ensure that it is captured in the cost estimate, electrical and architectural coordination issues are identified, and owner expectations are appropriate.

Any unique or complex control approaches should be described and detailed as far as necessary to verify feasibility. Water-side or air-side economization features in data centers,

which offer huge operating cost savings, often require control approaches that differ significantly from the standard controls used when these common systems are applied to office space. Extensive system monitoring with robust Data Center Infrastructure Managment (DCIM) systems will also require the inclusion of sensors that might not normally be provided. The mechanical designer should obtain information from IT as to what they intend in the way of monitoring and management systems, and provide the necessary sensors accordingly.

24.3.3.12 Prepare Deliverables

Deliverables for design development phase will vary depending on the client and design team. Again, early coordination with the client and/or architect that clearly defines what the mechanical designer *will* deliver as opposed to what they *can* deliver is critical to providing a high-quality and complete deliverable. Most, if not all, design development deliverables represent a preliminary version of a construction document set deliverable. Typical deliverables include a design narrative summarizing the system design, initial specifications, and drawings that illustrate the location and size of major pieces of equipment, air distribution paths, required active plenum spaces, main piping distribution paths, and preliminary sizing and power requirements (for electrical coordination) of major pieces of equipment. Significantly more detailed information may be required in some cases, for example, if the project delivery model includes some form of bid and award at the end of design development to bring in a contractor, developer, or another external entity to take the project through construction.

The design narrative will typically be an update of the schematic narrative deliverable. While the schematic deliverable will often discuss and compare different options, the design development deliverable focuses on the single selected system approach. Space load assumptions in terms of the computer equipment power consumption in the data center are clearly defined, ideally in both a watts per square foot capability for each program space and a total system kilowatt for the entire building. Where basis of design equipment selection have been made, it is appropriate to include preliminary submittal data as an appendix.

Specifications should be focused on defining the equipment requirements and any expensive execution requirements, such as requiring all welded piping or high-efficiency axial vane fans. While ideally a full set of draft specifications are collected, they may be very preliminary with minimal placeholders used for typical areas. Not all projects will require preliminary specifications in the design development phase, but even if not required for the submittal, it is often a design efficiency to begin tailoring them as the equipment selection tasks of design development are completed.

Drawings allow for more detailed coordination between the disciplines and should provide enough data for peer review, be it external or internal to the design team. Drawings should contain as much information as available on calculated loads, equipment sizing, distribution duct and piping sizes, system configuration, and layout. Avoid adding "filler" information hastily cut and pasted in merely to make the drawings look more complete to avoid problems arising from the unpredictable use of the design development drawings. A commissioning plan may be developed from this design deliverable, or an energy model created, or additional cost estimation, or other tasks that require information on the mechanical system configuration. It is better that incomplete areas are left undefined rather than a hastily added filler misleading other works and ultimately resulting in wasted time.

24.3.4 Construction Documents

The construction document phase is the completion of all design tasks required to allow for the permitting, bid, and construction of the facility. It is often the most costly design phase, but at the same time, the majority of big design decisions impacting system capacity, flexibility, and efficiency have been completed at the outset of this phase. The designer should recognize that construction documents for a data center are necessarily more thorough and detailed than for general office space. This is due to both the complexity and special nature of the data center systems, and the lack of familiarity most contractors have with these systems. Small differences in physical locations of equipment and piping in the field can result in reduced cabinet counts and/or impaired cooling performance that the contractor is unlikely to realize unless the design documents are very specific.

24.3.4.1 Finalize Equipment Selections

Load calculations are finalized, and the final equipment selections are made during this phase. Depending on the construction schedule anticipated, lead time of major pieces of equipment may be a factor in the final equipment selections. It is good standard practice to ensure that there are multiple providers of equipment that can meet the specifications—often a requirement with large or government clients. Beyond the typical savings advantage of a competitive bid to supply equipment to the project, verifying multiple suppliers of equal equipment ensures that the project will not be disrupted by a single supplier withdrawing the basis of design equipment from the market or, at a minimum, clearly highlight where equipment substitution may require design changes. Such postbid design changes tend to be costly, be it merely a forced increase in mechanical room size because an alternate air handler has a larger footprint or a full redesign to an entirely different air management solution.

If initially performed by the design team using software, website, or catalog procedures, key basis of design equipment selections should be verified with a manufacturer

representative to ensure accuracy. All details of the equipment selection need to be defined, verified, and recorded in the design documents. The number of details that need to be verified are as varied as the types of equipment that may be applied to a data center. Care must be taken to properly specify the right options, particularly in the area of controls and low outdoor temperature operation (unlike office buildings, data centers will need to generate cooling even on the coldest days).

The redundancy strategy used, such as $2N$ or $N+1$, should be included in equipment schedule notes to record the basis of design and aid commissioning efforts. Equipment should be selected with consideration of the reliability and maintainability required for data center operation.

24.3.4.2 Clearance and Interference Issues

The general equipment layout and distribution paths should be defined by this design phase. The final coordination of equipment layout with all other trades should ensure that there will be no interference or conflicts between trades. It's a risky game to count on contractors in the field to solve interference issues during construction, even if the job utilizes a design–build delivery model. Pipe sizes need to be laid out with allowance for insulation thickness, ducts fitted with consideration for the size of flanges, and equipment placed with the required code and desired maintenance clearances around them.

When coordination is done primarily by two-dimensional (2D) plan layouts and sections, piping and ducting need to be shown with thickness (double line) on the drawings. In congested areas, sections need to be provided to verify that the systems fit. Sometimes, elevation levels are assigned for different equipment, for example, defining the ceiling and lights as being in the band 9 ft 0 in to 9 ft 10 in above finished floor (AFF), mechanical piping and hangers at 9 ft 11 in to 12 ft 11 in AFF, and fire and electrical distribution at 13–15 ft AFF. This method of assigning elevations can be effective but may require more height than is absolutely required and additional coordination to ensure that vertical elements, typically hangers and seismic bracing to the aforementioned, are accommodated. Equipment should show clearly the clearance and service space required around it, including code clearance requirements in front of electrical panels.

3D modeling is becoming more common and can be a valuable tool to solve interference problems before they cause trouble on the construction site. 3D modeling significantly impacts the construction document process. Designer time and budget are shifted out of the construction administration phase, where the final coordination was often in practice completed, and into the construction document phase. Budget and staffing hours need to be shifted accordingly. The objective of this design investment is a better coordinated design that minimizes construction delays and change orders—ideally saving time and change order costs that more than offset the additional construction document time.

An often-overlooked coordination issue is the location and airflow around exterior heat rejection equipment. Data centers are designed for continuous control of the space, including during hours of extreme high temperatures. This will highlight any problems such as cooling towers that suffer from recirculation due to the placement of a screening wall or dry coolers bunched together in the middle of a black roof heat island with local air temperatures a dozen degrees higher than ambient. The common presence of redundant capacity that can be used during extreme heat periods provides some leeway but only a small amount since failures often occur on the extreme hottest days (not due just to bad luck, but rather the highest cooling loads correspond with the worst operating conditions for bearings and windings). Extreme hot exterior conditions will expose poor heat rejection airflow design on a fully loaded data center. Lawn sprinklers wetting overtaxed dry coolers on the roof of a data center are a depressingly common dunce cap placed on inadequate heat rejection designs.

24.3.4.3 Controls

The building controls are a critical element of a successful system yet are often left until late in the design process to be designed. To some extent, they are delayed simply because there is not a pressing coordination need for them to be defined earlier. Beyond defining a few locations where electrical power is required or wall space is needed to hang the control boxes, control coordination occurs entirely within the mechanical design.

Coordination of the control design with the equipment selections and specifications is critical. While a small data center facility may require no more than onboard controls that are integrated into the air-conditioner units installed on the data center floor, many system types used for larger facilities will require a networked system with external sensors or the flexibility of a more customized central DDC system. A number of control aspects require definition. Each piece of equipment must have the proper interface type defined and control input capabilities. Commissioning, an important testing aspect for a critical facility, may also require control features such as trending or remote Internet access (a very helpful monitoring and diagnostic tool, albeit one that carries a security requirement).

The control sequence is the logic that defines the system operation. The best control approaches will be simple enough to ensure reliability but complex enough to provide flexible and efficient control. As energy costs increase, the demand for controls to minimize system power consumption also increases. While reliability and robustness are the primary design concerns, a good control design will implement common efficiency best practices, such as varying airflow, controlling the temperature of the air supplied into ITE rather than returned, and efficiently adjusting system operation to most efficiently match part load conditions.

The most traditional control strategies tend to be reliable but very inefficient. For example, maintaining a return air temperature set point equal to the desired space temperature is simple (and simple is reliable), but since the return air temperature should be higher than the temperature of air supplied into the ITE intakes (the point where temperature control is required), this approach will chronically overcool the space. It will not directly control the parameter of concern, that is, the air temperature supplied into the ITE intakes. As overcooling is not typically viewed as a failure in data center control—the expectation of a computer room as being almost refrigerator cold is unfortunately, still common—traditional control sequences are often biased toward inefficient and even uncontrolled overcooling. Most CRAC manufacturers now to offer more efficient control options that utilize supply air temperature sensors, remote sensors located near the ITE intakes in the space, and variable speed air supply fans to offer improved efficiency. DDC systems have offered the flexibility to provide this kind of control for many years but at the cost of increased complexity and design effort.

Every effort should be made to minimize control complexity to the extent that doing so does not harm control capability and efficiency. Complexity tends to introduce delay in system start-up as problems are identified and corrected and introduce more points of failure that can reduce the reliability of the system. Some complexity is a requirement to provide the best space control—with the exception of lightly loaded and expensively overdesigned data centers, simply turning the mechanical system to full on is not an acceptable modern design. A good control system has the ability to match cooling output to the actual load, prevents overcooling of the space upon sensor or actuator failure, provides efficient space control, and can be fully understood—and therefore maintained—by the future system operator, not just the design engineers.

Humidity control in particular can be a problem in data centers. The humidity control set points should be relaxed to properly match the actual needs of the housed ITE to ease the control problem. The latest recommendations from ASHRAE TC 9.9 should be used, both to control humidity via dew point measurement, and to minimize energy use by allowing a much wider humidity range than has historically been specified. (See Chapter 11.) The control approach needs to acknowledge and accommodate expected sensor drift over time, since humidity sensors are significantly less reliable than temperature sensors. The design should also take pains to avoid a situation where sensor error over time can result in independent systems serving the same space fighting, a situation commonly seen with CRACs using independent humidity sensors where due to sensor error one is humidifying, while another serving the same space is dehumidifying.

24.3.4.4 Coordinate with Electrical

All disciplines must coordinate and integrate their designs during this phase. Coordination with electrical is sometimes relegated to "throwing drawings over the wall," but significant system savings and optimization may be achieved through more frequent coordination. The design capacity of the UPS system typically dictates the IT load that the mechanical system must be sized to support, so this design capacity should be verified regularly to catch any last minute changes that could greatly impact mechanical sizing. Impacts run from mechanical to electrical too; for example, the size of the emergency generator is significantly driven by the mechanical system efficiency. If the generator is near a size break point where a minor reduction in load could allow the use of a smaller unit, the cost-benefit assessment of the value of buying more efficient equipment to reduce the peak mechanical system kilowatt can change radically, perhaps to the point that a first-cost reduction and operating cost reductions can be achieved from taking the whole-building assessment approach. Similar effects may occur all the way back to the building transformer level. Capturing the extensive interaction between mechanical efficiency and electrical first cost during life cycle cost estimation and evaluation should have occurred in design development, and it should continue with the greater design resolution available during the finalization of the design in construction documents.

24.3.4.5 Coordinate with Fire Protection

Fire protection systems are highly specialized and jurisdiction dependent, so their final design drawings are typically produced by a fire control specialist—introducing another discipline that requires coordination. Airflow management design approaches often interact heavily with the fire protection scheme by introducing partitions in the space. The fire protection design must accommodate the partition scheme and any active plenums to ensure code compliance and proper protection of the space. The mechanical engineer should also capture the fire behavior required of the mechanical system during an alarm condition. While office space air handlers are commonly shut off during a fire alarm event, the critical nature of data center cooling often calls for a fire control scheme that keeps the cooling system, including air handlers, operating during a fire alarm. Another coordination issue is meeting any space exhaust requirements associated with the fire protection system if a dry gas-based system is utilized, including exhaust fans, relief dampers, and the associated control integration. In the United States, Article 645 of the National Electrical Code (NFPA 70) must be given careful consideration in that it can require Emergency Power Off (EPO) switches at each exit door that instantaneously "crash" all systems. There are ways to design without this requirement, and the 2011 version of the Code allowed

an important design option, but coordination among all design disciplines, as well as with IT, is necessary to produce a complient design.

24.3.4.6 Coordinate Equipment Layout with Architectural and Electrical

A minimum level of coordination with electrical is achieved by accurately showing mechanical equipment locations and requirements on the coordination drawing sets that are generated regularly through this phase. It is also important to ensure that the electrical parameters for equipment shown on the schedule—the phases, voltage, and design amperage—are accurate. If control panels require UPS power to avoid unacceptable reboot delays upon a loss of power, or to maintain circulating pumps for ride-through cooling of critical equipment then that should be clearly communicated along with the locations of all control panels that will be powered. The equipment that requires emergency generator backup also needs to be clearly defined, with any equipment that does not need backup clearly identified.

24.3.4.7 Coordinate IT Layout

The layout of ITE is often defined to some extent by the mechanical system airflow design. High equipment loads require an airflow management design that prevents hot exhaust streams from overheating an adjacent piece of ITE. Most airflow management designs enforce some limitation on where ITE will intake cool air and exhaust hot air. A common requirement is that ITE will be placed into standard-size racks and arranged in rows with cool air pulled in from the front "cold aisle" and hot air be ejected out the back into the "hot aisle." Most (but not all) ITE follows this airflow arrangement; if it is required for proper space control, the mechanical designer should clearly state that limitation of the system and coordinate with the client to ensure that the design will meet their needs. And if it does not, the first-cost and operating cost penalties of incorporating more flexibility should be summarized and communicated before the design is modified. Designing to allow for random rack layouts, as may be required for some applications where space is rented out to multiple different clients (often referred to as colocation facilities), is more expensive and cannot handle high density loads unless rack-based cooling (which solves the air management problem by placing the cooling coil or water-cooled heat sinks literally within inches of the heat load) is used. IT systems may also use pre-configured cabinets of unusual sizes and shapes that do not fit into standard rack rows and cabinet footprints. They may also require oddly shaped and located service clearance areas. It is sometimes necessary to design a specific zone for these special systems, with a different cooling arrangement than is used in the rest of the space.

24.3.4.8 Complete Distribution Design and Calculations

As with any design, pumps and fan sizing is customized for the project's specific distribution layout. The final sizing is done with consideration of the data center operation profile. Data centers operate 8,760 hours a year without downtime available for reconfiguration work; flexibility must be designed in with features such as "oversized" distribution sizing to allow future expansions or rezoning of loads. Such oversizing can also reap significant energy savings if the system is designed to capitalize on it by turning down efficiently.

24.3.4.9 Complete Specifications

Design specifications fully define the required equipment, components, and installation methods for the mechanical system. While the outline of the specifications is produced in design development, significant work occurs in construction documents to complete the specification book. For data center designs, particular attention should be paid to the allowable equipment substitutions. Commercial air-conditioning equipment may be significantly cheaper than data-center-specific equipment (commonly known as "precision air conditioning" and defined under ASHRAE Std. 127), but wholly unsuitable to serve the primarily sensible load and 24 hour reliability needs of a data center. The specifications need to tightly define all aspects of the equipment, in particular the redundancy, reliability, part load efficiency, and control components that tend to be significantly different from and more expensive than in commercial equipment.

Specifications are developed from a number of sources. The starting point is often a library of standard specifications either produced over time by the designer or licensed from a specialist source. Equipment manufacturers often provide guideline specifications that are useful once unimportant aspects ranging from trademarked coil treatments to the color of primer coat are trimmed out to allow reasonable substitutions. Regardless of their initial source, specifications must be fully reviewed and revised to meet the reliability and critical facility nature of the data center. Using specifications produced for a prior successful data center is an acceptable starting point, but full review by the designer is a (tedious and time intensive) must. Submitting a specifications set that carefully defines by name the equipment that does not exist on this job is embarrassing but is only an inkling of the expensive grief that can occur. Erroneous specifications combined with contract language on allowable substitutions can make denying the substitution of unacceptable (but relatively cheap) commercial air handlers in lieu of purpose-built CRACs an expensive change order.

The basis of design space conditions, loads, and design weather conditions should be included in the specifications if they are not stated in the drawings. These critical parameters

are defined and approved in the prior design phases' narratives. Including this information clearly in the construction documents, which will usually become part of the building operator's documentation while design narratives do not, is of significant value as the data center is loaded up and potentially remodeled in future.

24.3.4.10 Generate Coordination Drawing Sets

The final construction drawing set is developed in this phase. Multiple drawing sets are submitted to aid in design team coordination, usually including a 30, 60, 90%, permit, and final CD set. A small data center may combine some of the coordination sets into a single review, and a larger data center may rely on bimonthly meetings to review a common 3D model. The number of drawing sets should be clearly defined in the contract scope and verified during normal design coordination communication between the mechanical designer and design team lead (architect, client, design–build general contractor, etc.). The exact scope of each coordination set varies with the project; the following discussion is a general guide.

Care should be taken to ensure that estimated data placed on the drawing to allow for early coordination is clearly tracked and replaced with the correct calculated data as it becomes available—round numbers for design parameters such as 100.0 ft water gauge for all pump heads or 10 horsepower for every fan motor are common telltales of placeholder data that has escaped proper update with the final calculated sizing. Small oversights such as not updating the estimated pump size to match the final calculated pipe loop pressure drop (plus safety factor) can prove costly.

The 30% drawing set provides information on the proposed equipment layout and types, with a focus on aspects that require coordination with the other design disciplines. An early priority to support the electrical design is to set the location of major pieces of equipment and define their electrical demands. To support architectural integration, the location of all outdoor air intakes and exhausts, major ducting, external equipment, and piping are early priorities. This data should be clearly presented in the 30% drawing set and is therefore often the subject of the earliest coordination meetings. In a tight layout situation, problems with architects are traditionally solved with drawings and sketches on overlaid tracing paper. 3D modeling software is a rising alternative for coordination.

All coordination concerns raised by the 30% set should be resolved by the issuance of the 60% set, with additional information added. Additional coordination problems may arise as detail is added to distribution routing, and duct and pipe sizes are fully defined. The 60% set has most or all equipment selection finalized and distribution pathways clearly shown. Final pump and fan sizing calculations that are based on the final routing and sizes of ductwork and piping have at a minimum calculated estimates completed and in the equipment schedule.

Controls are defined for all equipment with preliminary sequences shown to illustrate the intended operation in the 60% set. If the controls are integrated into the specified equipment, the required options and intended settings are defined by clear notes on the equipment schedules (defining options only in the specifications or in control details is technically acceptable but in practice is more prone to being missed by contractors, causing trouble during construction). Integrated control capabilities can vary significantly between equipment suppliers, so it is important to define the control requirements fully, assess their impact on availability of alternate equipment, and ensure any bid or construction alternates offered provide equal control capability.

All drawing sheets that will be present in the final design package should be represented in the 60% set. The 60% set should include drafts of all drawings, including controls, permit sheets, plumbing, and fire suppression. Requiring that the responsible team members provide drafts of these drawings for this set ensures design team members are fully aware of their scope. While there should be no scope confusion this late in the design, if the mechanical design is assuming that an external consultant will provide a fire protection design and permit documents while the client is expecting fire protection to be integrated into and provided with the mechanical set, the 60% coordination set can be a painful but not catastrophically late point to recognize and correct the scope confusion. The mechanical engineer should check and verify that all expected sheets are in the coordination set and follow up to verify if any are missing. It is also important to verify that electrical is supporting all the mechanical equipment, including any accommodations for future equipment.

The 60% level is often the point where all disciplines have provided drawings (and/or electronic 3D models) with the level of detail and accuracy suitable for identifying interferences and other conflicts. Regular meetings, either in person or by voice conferencing with Internet screen sharing, are often begun to resolve interference issues as the final design package is completed.

With all drawings represented in the 60% drawing set, the 90% drawing set is simply a completed version of the 60% set. While rarely attained, the objective of the mechanical designer is for the 90% set to be the final design and require only cosmetic title block updates prior to release for bid and construction. Equipment sizing is completed and based upon final design calculations of load, pressure drop, and layout-specific parameters. Equipment layout is completed, including maintenance access and verification of installation/removal corridors and door heights. All distribution requirements, including the minor but critical plumbing associated with humidifiers and condensate control, are fully defined, sized, and shown. Plumbing is captured, and airflow management is shown in

the set and integrated with the fire suppression system and equipment layout as shown on the architectural backgrounds.

Controls are defined for all equipment in the 90% set, including completed point lists and sequences. Coordination with electrical should be completed for all aspects of the design, be it circuiting support for high-amp electric humidifiers located in nonmechanical spaces, control panels that require a UPS power circuit, or control of a restroom fan by a wall switch that will be installed under the electrical contractor's scope. When custom control sequences are defined, care should be taken to carefully check them and ensure they are complete, correct, and simple enough to be properly implemented. In a critical environment, the limiting factor on control logic should not be what the specified system can do, but rather what is the minimum it must do to provide the required reliability, control, and efficiency. As the last portion of the construction completed, flaws and errors in the control sequence can lead to delays and costs late in the critical final stages of the construction calendar when both time and contingency funds are often exhausted.

The production of details is a major task to complete the 90% set. Details show the exact construction and installation designs for equipment and mechanical components. Where details are pulled from the designer's predeveloped library, care should be taken to ensure they are applicable to the design. It can be confusing to include details for how to hang ducting from a concrete slab when the building is a single-story structure with steel roof trusses and inappropriate to include steam trap details for a building with no heating requirements at all. If distribution or mechanical room layouts are tight with significant coordination and interaction with other trades, sections and room layouts are appropriate. While piping layouts can be fully defined in plan view by noting bottom of pipe elevations or heights AFF, carefully selected section details tend to reduce field confusion and catch more interference problems in the drawings rather than in the field. Section details can also be valuable to ensure that air distribution plenums are not being clogged by mechanical, fire, electrical, and architectural elements. It is worth again emphasizing that the complexity and special nature of data center designs and systems requires an abnormal of detail to achieve trouble-free construction and a successful outcome.

24.3.4.11 Permit Drawings

Drawings submitted for building permit should be as complete as possible. Depending on the jurisdiction, changes between the permit drawings and final construction drawings may need to be noted by revision bubbles on the set—cumbersome bookkeeping if there are extensive changes. Depending on the schedule, the 90% drawing set may be used as the permit drawing set. If a separate permit set is produced, it usually differs from the 90% set by including permit-specific forms (sometimes inserted into drawing sheets). Time requirements often also result in it being less complete, with control sheets often neglected since they tend to have few code requirements. Details such as seismic bracing of ductwork, fire smoke dampers, fire alarms for air handlers (or the justification for omission of them), outdoor air ventilation rates (as low as possibly allowable), and other code-related design aspects need to be included and complete. Sections and large-scale room layouts dimensioned for construction layout are less important in the permit set.

The permit set is the completion of the code compliance research and design that occurred throughout the entire design process. Notes and narratives from the schematic and detailed design phases should be referenced, and any code concerns that were raised should have their resolution clearly shown, noted, and specified in the permit drawing set. Code concerns that were not raised in communication with officials may benefit from not being highlighted in the interest of keeping the set clear and concise for review.

24.3.4.12 Bid Package: Construction Drawings and Specifications

The bid drawings include the final design drawings and specifications. There should be few changes from the 90% set, and any signification changes should be explicitly coordinated by phone, email, and/or meeting with all affected disciplines. A final review of the specifications to ensure they are complete and applicable may result in additional changes. To ensure completeness, at a minimum, all equipment noted on the schedule will be represented in the specifications, all distribution systems will have installation and accessory information in the specifications, and every control point type will be fully described in the specifications. Changes that occur after the bid set is released can be costly; while it is inevitably difficult to find time and budget in these late stages, a final quality control review at least 3 weeks prior to the release of the bid set is a must for all but the smallest projects or the largest contingency budgets (and most understanding—often incredibly rushed—client).

24.3.4.13 Bid Support

During the bid period, requests for clarifications from contractors may be submitted. Following the protocol set by the client, the mechanical designer should be prepared to promptly offer written responses as required. Bidders will be looking to assess the lowest cost options to satisfy the design—which is in the interest of the owner, as long as the cost savings do not reduce the reliability, redundancy, and operational capabilities desired for the data center.

In some cases, due to time constraints, the bid package is released incomplete with additional addendum packages

released prior to the bid due date to complete the design documentation. If any mechanical scope needs to be included in the addendum, it is critical that the design team leader (typically the architect) knows what to expect and incorporates the mechanical materials in the addendum. Addendums may also be used to respond to bidder questions that reveal ambiguity in the design documentation or last minute cost reduction opportunities.

24.3.5 Construction Administration (CA)

The design job does not end until the data center is properly operating. Construction administration is a more significant time demand than for normal construction and critical to a successful project. The mechanical designer provides submittal review of equipment selections, site inspections to determine if installation requirements are being met, interpretation of the design documents when questions arise, quick correction of design ambiguities (or outright errors), solutions to interference problems, support for commissioning, and final inspection to ensure correct installation. None of these tasks differ significantly from any other mechanical design, other than the abnormal complexity and special nature of the systems, and the high-reliability demand of the data center that increases the importance of delivering a fully debugged system on day one. Design budgets should anticipate at least 50% more CA time for the data center than for the rest of the project.

24.3.5.1 Submittal Review

Submittal review ensures that all equipment meet the design requirements. A methodical approach should be taken to check the submitted equipment against the drawing schedule information and the specifications. When the submitted equipment matches the basis of design, the submittal review is primarily limited to verifying that the correct configuration and options are specified. Substitutions require more in-depth investigation to ensure they meet the letter of the design documents—as well as any design requirements that were not explicitly included in the design documents but assumed as a standard equipment feature that was included in the basis of design equipment selection.

24.3.5.2 Site Inspections

Regular site visits should focus on verifying that equipment and distribution are being properly installed. Contractors sometimes will install equipment as they have done in the past rather than as dictated on the design drawings. This can be an advantage in some cases, where an experienced contractor can compensate for design documents with weak specifications or drawing detail. But it can cause significant trouble if the contractor is not familiar with data centers and begins to make incorrect design interpretations, perhaps removing redundancy to save money or placing dry coolers closer together to reduce the size of the mechanical yard but at the cost of harming extreme day cooling performance. One of the most critical details to watch is piping locations to avoid airflow obstructions in under-floor cooling designs. And as with any project, there is always the need to ensure that the quality of installation meets the design requirements. The sooner an incorrect installation technique is caught and corrected, the lower the potential for adverse schedule impacts and costly and/or time consuming changes.

24.3.5.3 Design Interpretation

The design documents will fully describe the mechanical system and how to install it. But for more complex mechanical systems or ones that have unusual systems, there can be value in the mechanical designer discussing the design intent directly with the installation contractors. Caution is required to ensure that all parties understand that nothing in the discussion represents any approval for deviation from or addition to the contract document's scope. Casual discussions during site visits risk misinterpretation as being changes made (with cost implications) to the contract documents. Set meetings with documented meeting notes that clearly state no design changes were implied or approved at the meeting can be useful. For example, if an unusual piping design is used to provide piping redundancy, a 30 min meeting at the job site with the pipe fitters to describe the intent can ensure it is installed as shown. Control sequences are another area where direct discussion between the contractor and design engineer typically saves more time than it consumes; a half day spent reading through the control sequence and ensuring the actual programmer understands the intent can save considerable time versus correcting erroneous assumptions when they show up as failures during commissioning.

24.3.5.4 Design Modification

The installation offers the final word on whether all interference and coordination issues were properly resolved during the design process. There is no ignoring a column that is directly in the way of a pipe-run or a gravity-driven condensate drain line that hits the drop ceiling as it "slopes down to drain." Interference issues are often corrected by on-site meetings, but more significant problems (that often carry cost impacts) require written responses to requests for information. (RFI's) Promptly correcting any problems found is critical to minimizing the disruption and potential delay. Lead time issues may also come up during construction, which may require the mechanical designer to provide

substitute equipment options. When a tight construction schedule is expected, lead time concerns are another key reason to try to ensure that multiple vendors are available for all critical pieces of equipment.

24.3.5.5 Commissioning

Commissioning is a methodical testing of the installed systems. Theoretically, commissioning should be unnecessary: if the design is installed exactly per design intent in all respects and all equipment functions perfectly, it is not needed. In practice, commissioning is a very important process to ensure that the design is properly installed, that the design meets the requirements, and that the system will operate without failure in all anticipated conditions. Commissioning should actually begin no later than the DD phase of a data center project. Document review and comment by a knowledgeable commissioning agent can avoid costly time and changes when final commissioning is done. Support for commissioning is often integrated deeply into the fundamental design of the system. Test ports provided into piping, duct access doors at key points to allow inspection of dampers or turning vanes, and the requirement for extensive trending capability in the central control system are all common accommodations made for both commissioning and ongoing system maintenance.

The mechanical designer will need to ensure that the commissioning agent is fully informed of the design intent, in particular the control sequences, interior design conditions, and outdoor design conditions. Commissioning a high availability data center is very different than commissioning an office building, and a commissioning agent specializing in this work should be considered for mission critical projects. If no commissioning is planned or budgeted for by the overall project, the prudent mechanical designer (and experienced data center contractor team) allows time and budget to perform a targeted commissioning of their own. Without the active testing of the system provided by commissioning, there is a risk of system failure well after system start-up as internal loads change and external weather varies.

Commissioning is ideally performed only after all systems are completed and operating. Trouble can arise if it is performed prior to system completion; for example, use of load banks to test the capacity of a UPS system before the UPS room cooling system is operational can lead to such extreme room overheating that it can pop sprinkler heads.

24.3.5.6 Final Approval

The final inspection generates a punch list or list of problems that need to be corrected before the system installation is deemed completed per design documents. It is critical to note that the common punch list is no substitute for commissioning. Commissioning applies in-depth active testing to ensure that all systems meet design intent under all expected operating conditions. A punch list is generally based on a passive inspection of the installation and balance reports to verify that everything looks like it is installed per requirements.

Any changes to the design are collected and applied to the construction drawing set to provide a final and accurate as-built set of design documentation to the owner. The as-built documentation is critical to support the ongoing operation of the system as well as any future modifications or build-out. As-built documentation collects information from the field into the design documentation and needs to be completed before the construction team dissolves.

After all items on the punch list have been corrected and the as-built has been delivered, the mechanical designer signs off that the contractor has met their contractual obligations and their role in the construction project is completed.

24.3.5.7 Post-construction Support

The traditional designer's scope ends after project completion, but there are often continuing services that would benefit the owner. The designer is in the best position to offer recommendations for how to build out and operate the system to the highest possible efficiency. Questions regarding changes in the intended loading, slowed or accelerated build-out approaches, and operational optimization can all benefit from direct designer input. There is also often opportunity to improve the system efficiency by tailoring operation to the actual IT load and installation; once the system is constructed and operating, it can be optimized to the actual conditions rather than design assumptions.

A well-trained site staff can operate a properly designed system well, but the ultimate expert in the system will be the designer of record. Keeping them involved at a low level of review and comment can significantly improve operation. Enabling the remote monitoring of the system can help make this involvement an economical option.

24.4 DATA CENTER CONSIDERATIONS IN SELECTING KEY COMPONENTS

Most data center cooling systems rely on a number of common cooling components. While data center cooling applications have strict reliability requirements and a unique load profile, these can often be met by the proper application of common commodity mechanical equipment. When selecting components for data center use, all the standard selection concerns and approaches apply with additional considerations such as the following.

24.4.1 CRACs (Computer Room Air Conditioners) and CRAHs (Computer Room Air Handlers)[4]

CRACs and CRAHs, as indicated by their names, are specifically designed to provide cooling to data centers. Integrated controls are typically provided, the system is designed to carry the sensible load typical of a data center (i.e., a far higher airflow per ton of cooling provided than typical), and reliability is a primary design concern. Although there is a significant difference between CRACs and CRAHs, both types are often called CRACs as a generic term.

Many CRACs offer the option of a reheat stage. Reheat is typically used to prevent overcooling a space when a system is in dehumidification. As most data centers are not concerned about overcooling, reheat is usually an option best eliminated; it can be a surprisingly expensive energy consumer and offers little benefit. It is exceedingly common to find in operating data centers that reheat has been disabled by data center operators due to its expense and very justified operator confusion at the purpose of electric heaters operating in their high cooling demand facility in the midst of hot, humid summer weather.

Humidifiers incorporated in CRACs are another area of operational concern. While often necessary to meet client humidity control requirements, they tend to be high-maintenance components. Humidity sensor drift also commonly results in facilities with multiple units "fighting," that is, one unit may be humidifying, while a literally adjacent unit is dehumidifying—a significant adder to maintenance requirements and energy waste. Both reheat and humidifiers have become even less necessary under the newest ASHRAE humidity guidelines, which eliminate the need for humidification in the vast majority of designs. If humidifiers are deemed necessary, a shared control system for all humidifiers in a space is often appropriate to prevent this situation and reduce the number of sensors that inevitably require frequent service (calibration or replacement).

System efficiency at part load, which is not always publicized by manufacturers, is the critical parameter in determining the operating cost of a piece of equipment in a data center system design. A data center with all units running at full load is very rare due to the design of redundancy and surplus capacity for the future. Since the systems will operate at part load capacity in the majority (if not all) of the time, if the selection of CRAC is to consider the operating cost, carbon footprint, or other common efficiency metrics, then the part load system efficiency must be defined for an accurate analysis. With the advent and now common use of variable frequency drives (VFDs) to control fan speed, particularly when driving electronically commutated (EC) motors and sometimes variable speed compressors as well, part load operation can be much more efficient than full load. Recommended practice, therefore, is to run all cooling units, both "base" and "redundant", simultaneously at reduced speed and capacity. This not only saves energy, but ensures that all units are operating properly, and that there is no delay in bringing cooling back to full capacity when a unit fails or is taken out of service for maintenance.

[4] CRAC equips with a condenser and compressor but a CRAH does not.

24.4.1.1 Chiller Plant

The chiller plant efficiency can be significantly improved by recognizing and designing specifically to meet the sensible nature of the data center load. It can be a major operating cost benefit to recognize that the vast majority of the cooling load from a data center is sensible only—no latent load requiring dehumidification is generated by the ITE and ventilation rates are negligible—and to serve that load with a plant optimized specifically for that regime, A medium temperature chilled water plant operating at 55–60°F (13–16°C) is now common, but the newest facilities are delivering water at 70–75°F or even higher. Mechanical cooling equipment operates significantly more efficiently when the temperature difference between the evaporator temperature and condenser temperature (dry coolers, condensing coil, or cooling towers) is reduced, typically reaping far more energy savings from reduced compressor load than the fan or pumping cost associated with using a higher-temperature and lower-temperature delta cooling medium loop. A small conventional air-cooled system can be dedicated and optimized to provide the small amount of dehumidified outdoor air required to maintain space pressurization. Modern chiller plants can also reap significant energy benefits from the use of VFD drives, as can the associated pumps.

24.4.1.2 Air-Side Economizer

Use of an air-side economization cycle, which is bringing in outdoor air directly when the outdoor air is cooler than the space requiring cooling, offers tremendous energy savings for data centers, has a few well-known but specialized design requirements, and when properly designed improves reliability. Often initially considered for energy savings, the challenge of properly controlling the system and fear of potential contaminates from outside air are common hurdles that successful implementations overcome. A common method of separating outside from inside air while transferring heat is a heat wheel. Versions exist designed specifically for data center cooling. Likewise, adiabatic cooling systems are available specifically designed for data center applications. Economizer systems tend to increase complexity simply by their existence adding additional controls that can fail, but they also offer an excellent source of cooling redundancy during much of the year and can be designed in a fail-safe manner.

Data centers are a perfect candidate for economization since they have significant cooling loads 24 hour a day even

when cool outside. In many climates, an air-side economizer can cut annual mechanical system power usage in half. There are also benefits from reduced maintenance of mechanical cooling equipment; for example, compressors and dry cooler fans will see significantly reduced run hours and have long periods of time when they are not required to operate.

Economization savings are maximized when combined with an effective airflow management regime and modern cooling set points. Current standards for maintaining a supply air temperature of up to 80°F to ITE inlets result in an exhaust airstream around 100°F, theoretically allowing savings from air-side economization whenever outdoor air temperatures are lower than 100°F. In practice, control sensor accuracy and humidity control concerns do reduce this opportunity but only slightly in most climates.

As with all aspects of data center design, air-side economization must be carefully engineered to provide improved reliability and control of the space. Contamination from the outdoor air is a common concern, although there is little research to support this concern. Proper filtration has been found to provide appropriate protection. Consideration of unique local conditions, such as a location directly adjacent to saltwater, intakes near diesel exhaust or other hazardous-to-health fumes, odor problems that raise comfort issues, or unusual ITE requirements, must be understood and accommodated in the design. The large loads of data centers require large volumes of outdoor air to remove; the significant size of filtration and associated maintenance cost and accessibility should be considered during the evaluation of this design option. Humidity control is the other design concern with air-side economization. Redundant control sensors and an adiabatic humidifier system that utilizes the data center waste heat are the common approaches to ensure that an air-side economizer does not create a wasteful false humidification and dehumidification load.

Reliability of the data center as a whole can be improved by the second source of cooling provided by economization. The majority of ITE can operate for a period at temperatures significantly above the design temperature of a data center—temperatures that can often be maintained through just the operation of an economizer system if the primary source of cooling failed. This additional level of redundancy offered is often overlooked since any temperature excursion above the design temperature is unacceptable. However, it should be noted that an economizer system may be able to prevent downtime even if it cannot maintain the design temperature—a safety factor no data center design should ever depend on, but it does add value to the owner. The benefits of an economizer system are organically understood by most operators, with "open the doors" a common last-ditch maneuver in response to cooling equipment failure in a computer room.

The economizer is fully backed up by mechanical cooling and therefore does not require redundancy to protect reliability, but reliability can be hurt if the economizer controls are not designed to be fail-safe. The system should be designed such that no single failure, such as a temperature sensor giving an erroneously low reading, can result in *hot* outdoor air being brought in and overwhelming the mechanical cooling system during periods when the economizer should be inactive. Damper actuators controlling the economizer outdoor air intake should fail to be closed and be monitored by end switches with appropriate alarms. Redundant sensors should be used to sense outdoor air temperature and humidity, and a regular sensor maintenance regime (replacement or calibration) followed.

The relative benefits and disadvantages of air-side economization versus water-side economization are discussed in the following section.

24.4.1.3 Water-Side Economization

ASHRAE 90.1-2010 required the use of economizers on virtually all data center cooling systems. The 2020 version now reverts to ASHRAE 90.4 to meet energy efficiency requirements where economizers are encouraged, but not required. (See Chapter 11 for more detail.) Use of a water-side economization system, that is, bypassing the energy-intensive mechanical compressor equipment to create chilled water through evaporative cooling alone, offers tremendous energy savings for data centers, has specialized design requirements, and has varied impacts on reliability. When properly designed economizers can even improve reliability, but can be highly detrimental if treated as a standard office building component instead of part of a mission critical system. The greatest savings are seen in climates with a significant wet-bulb depression, but most climates offer good opportunity. Design concerns focus primarily upon ensuring reliable chiller plant staging. When properly implemented, water-side economization offers an additional source of cooling in case of chiller failure during cool weather.

Data centers are a perfect candidate for economization since they have significant cooling loads 24 hour a day even when cool outside. In many climates, a water-side economizer can cut mechanical power usage in half and in some places can eliminate the need for mechanical cooling entirely when ASHRAE guidelines are followed to the fullest. There are also benefits from reduced maintenance of mechanical cooling equipment; for example, chillers will see significantly reduced run hours and have long periods of time when they are not required to operate (or be rushed back into service to provide redundant standby capacity) and can have preventive maintenance performed on them.

The primary design challenge is to ensure that the operation of the water-side economizer will not result in loss of cooling from the plant when the plant is switching from chiller operation to water-side economization operation, and vice versa. The system must be designed to allow a seamless

transition since, unlike office buildings, loss of cooling for even 10 min is not acceptable.

To ensure stability, the chillers must be able to start while the water-side economizer system is in operation—a challenge since in free cooling operation, the towers may be full of water at 45°F (7°C), while many chillers cannot operate stably until towers provide them with water at 60°F (16°C) or even higher. There are several possible methods to ensure chillers can start even when the towers are cool and operating at free cooling temperatures. One is the use of some form of head pressure control within the chiller to allow it to start up using the same cold condenser water as the free cooling system is using. Another common approach is to create an independent water-side economizer loop (often by temporarily isolating the redundant cooling tower capacity and providing a dedicated water-side economizer condenser water supply line) to ensure that the main chiller condenser water loop temperature can be quickly raised high enough for stable operation. Alternatively, some form of mixing loop can be configured to ensure the chillers can be provided with an acceptably high condenser water supply temperature for start-up even when the cooling tower sumps are at a free cooling temperature. Regardless of the design method selected, the designer must allow for reliable start-up of the chillers simultaneously with operation of the water-side economizer system.

Retrofitting a water-side economizer system often offers attractive paybacks but carries its own design challenges. The retrofit typically must be done with no system downtime, requiring luck in the location of access valves or, more often, use of techniques ranging from hot taps to freeze plugs to do the work while the system operates. A careful evaluation of the system should also be made to identify any existing problems that may assert themselves when the free cooling system is in operation, such as chiller problems that would prevent low load operation or exterior piping with failed (or missing) heat trace—problems that could be unmasked by the new operational profile introduced by water-side economization.

If properly designed for a data center application, water-side economization offers a backup source of cooling for much of the cooling season. It is often compared with air-side economization. Compared with a direct (non heat wheel) air-side approach, water side economization isolates the interior environment more since outdoor air is not brought in. But when evaluating reliability, an air-side economizer often offers backup to more failure modes; for example, a burst condenser water pipe during extreme cold weather could shut down the entire chiller plant—including water-side economizer—but a data center with an air-side economizer could remain fully operational while its entire chiller plant was down and likely remain up until repairs could be made. Which approach offers better energy savings depends on the local environment, specifically if the average wet-bulb depression is enough to overcome the lower temperature required for water-side economization due to the approaches of the cooling tower, flat plate heat exchanger, and air-side coils. Ultimately, the choice of system type requires a broad system evaluation by the mechanical designer and discussion with the client.

24.4.1.4 Humidification

Humidifiers introduce a significant source of catastrophic failure to a data center. The relatively high-pressure domestic water lines are a source of flooding concern. A leak detection system and supply shutoff is recommended, along with an appropriate maintenance and testing schedule to ensure the proper operation of the system. Domestic water piping can supply an almost infinite volume of water, making it a higher risk than a chilled water piping system that typically has a very limited volume of water.

The need for humidification in data centers is an evolving. If humidification were not a traditional standard started in the age of punch card feeders, it is unlikely it could be justified for use in most critical facilities today—precedence is its primary remaining justification. The vast majority of ITE is protected from static discharge by chassis design. And if the static discharge protection designed into the equipment casing system is bypassed for internal work, then humidification alone does not provide an acceptable level of protection. However, precedence carries significant weight in data center design, and many clients still require humidifiers in their data centers. Despite precedence, from an engineering viewpoint, it is peculiar how wedded to humidification operators tend to be since it is far more common to hear a verifiable data center horror story involving knee-deep water under a raised floor caused by humidifier system piping or valve failure than a horror story involving static discharge.

The humidification requirement for a data center is theoretically low since there is little outside air brought in; however, operation and control problems regularly result in excessive humidification. This is particularly problematic in regions with high levels of gaseous contaminates which can turn into corrosive acids at relative humidity levels above 60%. Uncontrolled dehumidification is an expensive energy waste that is also common, particularly in direct expansion (DX) refrigerant cooling coil-based systems that often have simultaneous dehumidification and humidification due to an unnecessarily high space relative humidity set point, low dry-bulb temperature set point, and a tendency for portions of DX cooling coils to run cold. In systems with multiple independently controlled CRAC serving the same space, sensor drift over time will often result in adjacent systems "fighting," that is, one in humidification, while the other is in dehumidification; this problem is also exacerbated by tight set points, with the humidity deadband frequently far smaller than required.

Standard humidification systems are a significant operating cost in both energy and maintenance, with the partial

exception of adiabatic systems. The most common humidifiers use electricity to vaporize water, a very energy-intensive task. If a low-humidity set point is used and there is minimal outdoor air (and no air-side economizer), the inefficiency of an electric humidifier may be of little net annual cost due to infrequent use.

An adiabatic humidifier that uses the waste heat of the data center itself to vaporize water offers an energy benefit by providing free direct evaporative cooling and reducing electrical demand on the generator. Adiabatic humidifiers raise maintenance and operational cost concerns in the larger sizes, where many atomizing types (ultrasonic, high-pressure water nozzles, compressed air nozzles) require significant water purification. While atomizing nozzles are a technology of choice in many critical environments that need to condition large quantities of outdoor air—such as a data center taking advantage of air-side economization—the associated water treatment plant can rapidly grow to expensive proportions and lead to a reconsideration of the simpler but more fouling-prone wetted media approaches.

24.4.1.5 Dehumidification

Uncontrolled dehumidification in a data center can be a significant overlooked design challenge. A data center run with outdated but common set points, such as 70°F (21°C) dry bulb and a minimum humidity of 45%, has a dew point of 48°F (9°C). If the space cooling coil is operating below that temperature, it is possible for condensation, that is, uncontrolled dehumidification, to occur on some portion of the coil—a situation that reduces the capacity available to cool the space and can significantly increase operating costs and energy consumption. The most effective protection against uncontrolled dehumidification is to separate the sensible cooling system from the dehumidification system entirely and design the sensible cooling system such that the working fluid temperature never drops below the desired space dew point. An example would be using a 52°F (11°C) chilled water supply temperature to supply air handlers cooling the data center space and a small dedicated outdoor air system[5] with a stand-alone DX coil to provide dry air for pressurization.

24.4.1.6 Fans

The use of variable speed drives for fans in data centers has become popular and offers the most significant energy savings opportunities to air-based systems if properly controlled. Variable speed fan systems take advantage of redundant capacity to allow all fans to operate at lower speed even when the data center is at design capacity. The savings accrue quickly due to the cube law nature of fan power—turning fan speed down by only 15% reduces fan power consumption by almost 40%. The fan speed needs to be controlled in series with the cooling coil to ensure the fan speed is reduced, with common algorithms including controlling the fan speed to temperature sensors in the space while the coil is controlled to maintain a constant exit air temperature or sequencing the fan speed and coil output in series (turning down fan speed and then reducing coil output). Quite sophisticated control systems exist today that can even monitor and adjust different types of cooling unit fans and cooling capacities to maximize energy savings while maintaining proper cooling. The key is simply to ensure that the fan speed turns down and that the coil and fan speed control loops do not interact in an unstable manner.

Traditional fan optimization techniques can also yield significant savings. For example, the common raised floor data center configuration can improve fan efficiency with features such as "PLUG" type plenum fans lowered under the floor to directly pressurize the underfloor plenum. If a built-up air handler is used, there are greater opportunities ranging from large and high-efficiency vane axial fan systems to the use of many smaller fans in parallel configured to create a wall.

Minimizing fan power installed helps reduce generator and electrical sizing, but it also has a major impact on operating costs. Data center fans operate 8,760 h/year at a near constant load, which justifies a significantly higher first-cost investment in larger ducts and air handlers to reduce fan power operating costs than would be considered for a 2,600 hours a year variable air volume office system; applying standard rules of thumb that have evolved from office system design to duct sizing or selecting coil face velocity in data centers will result in a working system but miss many opportunities to optimize operating costs and energy usage over the life of the facility.

24.4.1.7 Cogeneration

Cogeneration is typically in the realm of the electrical designer, but it crosses into the mechanical design when the waste heat is used to generate cooling (sometimes referred to as trigeneration, even if there is no creation of heating from the plant). The use of an on-site cogeneration plant to power the data center and drive a cooling plant can offer a compelling story, but the business case can be difficult. Local incentives, the marketing aspect of a "power plant on-site," and the specifics of how it aids redundancy and reliability are all key factors that the mechanical engineer can assist in defining. Fuel cells are another emerging power approach. The significant heat generated by these systems does present opportunities for waste energy usage.

[5] In some cases where first cost is a driving factor, humidity control and positive building pressurization are not deemed a critical system that requires full $N+1$ design, and only a single outdoor air unit is provided.

24.5 PRIMARY DESIGN OPTIONS

We will discuss the most common system approaches to four key areas of data center design: the cooling medium, heat rejection method, air delivery path, and air management. The selection of cooling medium defines whether the system will be moving heat primarily using airflows or water flows, which has profound design implications all the way down to the client's ITE selection in some cases. The heat rejection approach influences the equipment options for the mechanical system. The last two areas apply primarily to air-based cooling system: the delivery path used for supply and return and the air management system used to avoid hot spots and efficiently collect waste heat.

These are only the most common current approaches; there are many other configuration options for an analysis-driven design to assess and potentially pursue. Designs that represent evolutionary changes, the development of fundamentally new system approaches, are strongly incentivized by the magnitude of operating cost savings possible from more energy-efficient design approaches (and, at the moment, a few Internet behemoths with the savvy and design budgets to vigorously pursue those cost savings). It should be noted that ASHRAE Standard 90.4 was developed specifically as a performance-based standard with the express intent of encouraging innovative approaches to cooling data centers in an energy efficient manner.

24.5.1 Cooling Medium

24.5.1.1 Air from CRACs and CRAHs

CRACs and CRAHs are specialized air handler units available from a number of manufacturers. Placed directly on the data center floor, they are an imposition on the finished data center floor space. They often include integrated controls and options for humidification and dehumidification that are directly tailored to the data center environment. With capacities typically ranging from 5 to 50 tons each, they offer the simplest design option—essentially requiring little more than following the catalog selection process although selecting and positioning units to achieve efficient airflow to cabinets can be very challenging, often requiring computerized fluid dynamics modeling (CFD) to resolve. While the use of computer room air conditioners offers a predesigned system in many ways, there are additional mechanical design details required including pipe routing for the cooling fluid (be it refrigerant, a glycol/water mix to and from external dry coolers, or chilled water from a central plant), humidification piping or system, condensate removal drain system, any associated ducting, service and support areas, layout of interior and exterior (heat rejection) units, correct specification of control options, and fire system integration (including any required purge ventilation). A traditional design favors for a relatively simple but with redundancy of having multiple units of a system. For large facilities, maintenance can become costly due to having many small distributed fans, condensers, compressors, and other components.

A recent evolution of CRAC is the in-row unit. An in-row unit is designed in the same form factor as a standard ITE rack and is made to be installed directly in line with IT racks. This can significantly simplify the air management design and often offers variable speed fan capabilities that further improve the efficiency. Currently popular for small data centers and as a retrofit to address hot spots in larger facilities, it is an interesting design evolution in the area of CRAC.

24.5.1.2 Air from Central Air Handler

A central air handler system can offer the opportunity for a more customized and lower-cost system but requires more design effort. The air handler itself must be carefully specified; in particular, it must be sized for the almost all-sensible load expected from a data center. Extensive control system design is needed to provide appropriate data center control and robustness (no single point of failure, be it a single controller or a single control point such as a common supply air temperature sensor). Layout is typically more complex, but the data center floor is kept clear of most if not all mechanical equipment. For larger facilities, the flexibility, efficiency, and cost-saving potential of using central air handlers rather than the prepackaged CRAC options often outweigh the significantly greater design effort required.

Systems integrated into the building, where entire plenums may be combined with wall-sized coils and fan systems to essentially create massive built-up air handlers, have resulted in some elegant and highly efficient central air handler designs.

24.5.1.3 Liquid to Racks

The most efficient method of moving heat is using a liquid like water, which has a volumetric heat capacity over 3,000 times greater than air. Some data center equipment takes advantage of this to cool equipment directly with water, either through a plumbed heat sink system or (more commonly at this time) a cooling coil integrated directly into the equipment rack that cools hot exhaust air directly as it is exhausted. This approach requires piping in the data center footprint, raising concerns for some clients who are very concerned about liquid leaks onto ITE although leak problems from chilled water supply systems designed for this kind of industrial environment are very rare (liquid process cooling loops are the norm in many critical facilities, such as semiconductor clean rooms or pharmaceutical laboratories). Locating the distribution piping in a manner that allows free rolling of equipment around the data center floor, provides for future addition and repositioning of equipment, and meets redun-

dancy requirements is a key consideration when selecting equipment. The use of liquid cooling offers tremendous potential for efficiency improvements and can allow high-power densities easily, even in spaces with very little height and space to move air. The greatest efficiencies are achieved by closely integrating a water-side economizer system and minimizing heat exchanger steps between the heat rejection and the coil itself to minimize the resistance to heat rejection (the combined approach of all the components between the heat sink/outdoor temperature and the ITE hot airstream).

With most liquid cooling applications, there is still a need for a small air-side system to carry the minor heat load picked up from convective and radiative heat transfer through the sides and tops of the racks, as well as lighting and any shell loads. A humidity control system must also be provided since liquid-cooled racks cannot provide dehumidification at the rack level. Depending on the liquid cooling system cost per rack, it may not make sense to provide low-power racks, usually under 2 or 3 kW, with water cooling—their load is another that usually is handled by an (much smaller than typical) air-based system.

24.5.1.4 Others

Some manufacturers offer systems that use refrigerant loops to coils in the rack. With most of the same design concerns (piping layout to racks, flexibility, ambient system for humidity control), these systems are essentially liquid to racks. They do address concerns about liquid piping on the data center floor since any leak would flash to a gas without harming any data center operation but often at a significant cost and efficiency penalty.

Research is always continuing into innovative methods of transferring heat out of racks, ranging from bathing the servers directly in a cooled dielectric fluid to using very long heat pipes or solid heat sinks to produce entirely passively cooled systems. The availabilities of these less conventional approaches varies, but the mechanical engineer should remain vigilant for unexpected changes in both the form and temperature of the load itself and mechanical equipment options to move it.

24.5.2 Heat Rejection

24.5.2.1 Dry Cooler

A common method of heat rejection to the outdoors is a dry cooler, which consists of a coil and fan unit placed outside (similar to a car's misleadingly named radiator, which works almost entirely through forced convection). Dry coolers can be used to cool a liquid condenser loop or to condense refrigerant directly. As dry heat rejection systems, they are theoretically limited to cooling to a temperature above the outdoor dry-bulb temperature and in practice tend to be the lowest efficiency heat rejection option. They do offer a low-profile solution and do not require any significant operation maintenance. Design can also be very simple, with everything but choosing options completed by the provider of an associated CRAC, or, in the simplest possible solution, the dry cooler is integrated into a traditional packaged rooftop air handler system (properly selected to deal with the sensible nature of the load).

24.5.2.2 Open Cooling Towers

The most effective form of heat rejection is typically an open cooling tower. Open cooling towers have the advantage of using evaporation to increase their capacity in all but the most humid climates, providing a more compact footprint than dry cooling-based systems. They are dependent on water, so for reliability reasons, backup makeup water is stored on-site, sometimes in the form of a sump that holds several days of water; at a minimum, enough makeup water to account for evaporation during design load for the same length of time as on-site diesel storage will run the generators is required. Freeze protection is a significant concern during build-out and operation. During system start-up, the level of waste heat from the facility may be vastly lower than at design, creating a period where the towers freeze due to the lack of waste heat from the data center load. During operation, any piping that is allowed to sit without flow, such as redundant piping or piping bypassed during free cooling operation, may become a potential single point of failure for the whole system and should be freeze protected and possibly isolated as necessary to ensure system reliability design goals are met.

24.5.2.3 Air-Side Economizer

An air-side economizer system rejects heat by exhausting hot air and bringing in cooler outdoor air. Air-side economizer is a tremendous opportunity to improve many data center designs as long as it is carefully designed for data center use. Having the ability to condition the space with outdoor air provides a degree of redundancy: if the compressor-based cooling system fails, the data center temperature can at least be maintained at around the same temperature as the outdoors. In most cases, this will allow the equipment to remain operational during emergency repairs even if temperatures may exceed the recommended operational temperature envelope. The ASHRAE "Allowable Temperature Envelope" was developed specifically for this reason. It confirms that ITE inlet temperatures can rise above the Recommended 80.6°F (27°C) maximum for as long as three days without significant equipment impairment.

Concerns of contamination by outdoor air are addressed by proper filtration, designed for ease of regular filter replacement unless a heat wheel or other form of high efficiency air-to-air heat exchanger is used. Recent studies

have indicated that moderate filtration is more than adequate to eliminate particulate concerns. Locations with high levels of vehicular or industrial pollutants must be particularly concerned if they also have high humidity levels. Moisture content above 60% RH can turn these gasses into acids that eat away circuit board lands and connector terminals. Some sites may also have gaseous have unique pollutant concerns such as ammonia gases near agricultural land or forest fire smoke. However, reality can take second place to client perception, so any implementation of an air-side economizer should be carefully vetted and approved early in the process to avoid wasted design effort on a system the client just does not want.

Humidity control is a far less concern with economizer designs than it was once thought to be as a result of the ASHRAE research into humidity and static discharge in data centers. (See Chapter 11.) High humidity must be controlled as noted above, but it should no longer be necessary to add moisture in any but the very lowest humidity winter environments. Relative humidity levels down to 8% have been shown to be acceptable, but engineers often design for a minimum of 15% RH to feel safe. If humidification is considered necessary, an adiabatic humidification system is strongly recommended to minimize operational costs. Adiabatic systems can be configured to use the heat energy in the hot return airstream from the data center itself to evaporate the water needed for humidification. There are many adiabatic technologies available, including wetted evaporative media pads, atomizing sprayers, and ultrasonic. The ultimate system selection should balance operational cost (including any water treatment requirements and auxiliary pump or compressor power), maintenance requirements, and reliability. Electric steam humidification systems will have comparatively high energy costs at the volumes of air that economization introduces into the space and may negate the energy savings of economization entirely.

Another aspect of air-side economization is ensuring that pressurization control is maintained. The large volumes of outdoor air introduced into the data center must have a proper exit path provided. Poor pressurization control can lead to doors that do not close properly—an inconvenience in commercial construction but an unacceptable security risk to most data center customers. However, best practice is to always maintain some level of positive pressure in a data center in order to minimize the entrance of particulate contaminates when doors are opened.

As an efficiency feature, economization systems typically do not require redundant design. However, they do require proper fail-safe design. Outdoor air economizer dampers should fail to a closed position. At a minimum, including redundant outdoor air sensors is needed to ensure a sensor failure does not result in hot outdoor air being allowed in during the summer. Also advisable are end switch feedback from damper actuators to alarm stuck dampers and control logic that is biased toward locking out economizer operation when its benefit is in doubt.

24.5.2.4 Others

There are other less common methods of heat rejection. A geothermal system that rejects heat to a large ground closed-loop heat exchanger is theoretically an option; however, it is usually quite expensive in the scale required by a data center and poorly suited to the year-round, cooling-only operation of a data center. A geothermal system that rejects heat to surface water can be a very good option but requires excellent siting adjacent to an appropriate body of water. Geothermal can be an excellent consideration in hurricane or tornado regions where above-ground heat exchangers are particularly vulnerable to damage. Integrating the data center into a larger facility that requires heating so the data center's waste heat can be rejected to a building that requires heat or even a district heating loop can be a very successful and efficient approach. Note that the temperature of heat coming out of a data center tends to be quite low. Air or water from the data center floor can still be as low as 75–85°F (24–29°C) but return air and water temperatures of 100°F (38°C) are becoming more common. In any case, a heat pump or other compressor-based system is often used to boost it to make it more useful for reuse. There are also options that combine evaporative cooling with a dry cooler, such as closed-loop cooling towers that can be appropriate in some situations.

24.5.3 Air Delivery Path

24.5.3.1 Underfloor

The most common traditional data center air delivery path is through a raised access floor. Several manufacturers provide raised floor systems consisting of 2 ft × 2 ft or 600 × 600 mm square tiles placed on pedestals. The underfloor plenum is pressurized with conditioned air, and airflow tiles are placed where air supply is required. The tiles are available in perforated and grate designs, with and without sliding dampers, to enable airflow to be correctly established in front of each cabinet", and are easily picked up can be easily picked up and moved, providing a great deal of layout flexibility.

The height of the raised floor is usually specified based upon the amount of airflow it must accommodate, which defines the free area needed under the floor (free area height is usually a smaller number than the nominal floor height due to the height consumed by the tile itself). It is critical that the mechanical engineer carefully consider the amount of underfloor free area height that will be blocked by electrical conduit, data cabling, and any other infrastructure intended to be routed underfloor. It is quite common to find data centers suffering from poor air distribution (manifesting

as localized overheated "hot spots") due to the underfloor airflow being blocked by bundles of wiring, electrical raceways, and mechanical system piping. Underfloor plenums are not magic; adequate free area must be provided and maintained to allow for design airflow. In fact, the much higher air volumes required to cool today's high density cabinets make uniform under-floor air delivery a significant challenge. Computational Fluid Dynamics (CFD) analysis by a competent data center modeler is highly recommended, both in the design stage, and as part of commissioning, to ensure proper air delivery.

While the common approach to provide underfloor air delivery is to use a pedestal and tile system, it is not unheard of to take a more industrial approach and provide air supply up from a mechanical subfloor (sometimes shared with electrical infrastructure) through the structural floor of the data center. This approach can provide very low pressure drop and controllable air distribution and be a good fit if a retrofit space or clever architecture (and/or high pedestal and tile system costs) provides an economical way of constructing the subfloor. It also can offer a great deal of operational flexibility, with the ability to add water supply piping and ducting, add cooling equipment, or make other significant modifications on the mechanical subfloor level while the data center is in operation.

24.5.3.2 Overhead Ducted

Delivering air overhead can be a quite effective approach. With the volumes of air motion required by most data center loads, the impact of natural convection flows is almost irrelevant, so it matters little if air is blown down from overhead versus up from a raised floor. Overhead ducting can integrate well with central air handling systems. Note again that data centers typically have high loads and large volumes of air in motion, which can result in the need for very large ducting, ducting that must also coordinate with cabling, lighting, and fire suppression. Care should be taken to optimize the ducting size for operational costs and future flexibility; undersizing ducting trades first-cost savings (which can be small if you must upsize fans and electrical infrastructure to support higher duct pressure drop) for a lifetime of higher operating costs. As data center loads rise, ducting can become prohibitively large and blur the line between ducting and plenums.

A ducted approach can carry the penalty of limiting data center flexibility. The locations of air supply need to be fixed during the initial construction and cannot be moved easily during future operation if the locations or footprints of the racks change. The initial ducting design can provide some allowance for future flexibility by sizing and laying out ducting in a manner that allows large increases in airflow with minimal pressure drop cost as required to meet future movement of load in the space. There are also creative approaches that can be used to stretch the flexibility of a ducted approach, including easily reconfigured fabric ducting or a methodical inclusion of taps and distribution paths to add future duct runs. It should be noted that when conventional perimeter CRACs are used with overhead ducted designs, the CRAC return air intakes are at the bottom since discharge is at the top. This makes return air circulation problematic, since it defies conventional thermodynamics. This can be addressed in two ways. First, CRACS can be ordered with rear returns so that ceiling plenum return air can be coupled down to the return air intakes. This obviously requires that the CRACs be located away from the walls, consuming valuable space, or that the return ducts be installed within the walls or through the room behind. Alternatively, the CRACs can be installed in a "cooling gallery" with a demising wall separating them from the computing space. The mechanical room becomes an entire return air plenum, so normal front return CRACs can be used. However, fire codes may require very large dampers in the return air path. This also requires additional floor area, but an added advantage to this approach is that it isolates both machine noise and workman from the IT space.

24.5.3.3 Overhead Plenum

An overhead plenum is another option for distributing air to the space. The most common approach is to use a standard drop ceiling to create a return air plenum. This can be a very effective and low-cost approach that integrates well with many air management design approaches. Return grills are located over hot aisles, and the air handlers are configured to draw return air from the plenum space. Exhaust fans and adiabatic humidification systems can be added to the plenum as part of an air-side economizer system (and serve double duty as fire suppression gas exhaust fans if necessary).

Without the structural need to carry the weight of the racks and provide a floor surface, overhead plenums can usually be much more generously sized than underfloor plenums. The overhead plenum can even grow to become a de facto mechanical interstitial floor that houses mechanical equipment and provides access for maintenance without intruding into the data center floor.

24.5.3.4 Through the Space

It is difficult to beat the simplicity of supplying cooling air by simply blowing it through the space toward the racks. Oddly, this design can be a very poor approach with hot spot problems and low efficiency, or very well performing with excellent control and efficiency—it all depends on the airflow management separating supply and return.

With fully mixed airflow management (also known as no airflow management), air delivery through the space typically reduces usable cooling capacity by haphazardly

mixing hot exhaust air with the supply air. But it is extremely low cost and easy to implement in nearly any space, requiring little more than placing a CRAC unit on the floor and hooking it up. If the expected equipment loads are low, then the simplicity and low cost of this approach can be compelling. However, since these designs almost always use CRACs with overhead air discharge (also known as "top blow"), return air becomes an additional problem since return grills are at the bottom front, further reducing cooling efficiency. It is virtually certain that this approach will not meet the requirements of either ASHRAE 90.1 or 90.4.

With good airflow management, a through-space approach can be extraordinarily elegant. If all hot exhaust air is collected as part of a heat exhaust, the access walkways provided for moving equipment and operators through the space can serve as very low pressure drop supply air ducting. While a simple concept, this optimized airflow management integration can sometimes require significant analysis to ensure that it will provide reliable distribution during real operational conditions such as when racks are moved in or out. At the volumes of air movement required, computational fluid dynamics analysis may be recommended. While good airflow management systems usually pair through-space supply with a hot air return plenum, through-space air delivery can be paired with creative air management containment systems to provide a comprehensive single-floor air supply and return solution by partitioning the space into strips of hot through-space return aisles and cool through-space supply aisles.

24.5.4 Airflow Management

24.5.4.1 Fully Mixed

Low-load data centers may have airflow similar to the standard design used for office spaces. A standard conditioned office space throws conditioned air supply throughout the space, mixing the cool air throughout to dilute the heat being generated in the space. In data centers, it is typically done through the use of CRAC units that throw air out from the top of the unit and draw it back in from the lower face of the unit—with little to no ducting required.

This approach typically has poor performance but is very easy to implement. However, while simple, a fully mixed approach is limited in the capacity it can serve due to its inability to address localized hot spots. A prohibitive amount of airflow (equipment size and power usage) is required as loads increase. It is simply less efficient and effective to dilute hot air exhaust from ITE than to remove it directly. The dilution approach also impacts the effective capacity of the cooling units since the return air temperature is usually only as high as the desired space temperature set point, limiting the temperature differential possible across the coil (Fig. 24.1). In short, it is unlikely that this approach will meet the energy efficiency requirements of either ASHRAE 90.1 or 90.4, and it it totally inapplicable to high load data centers.

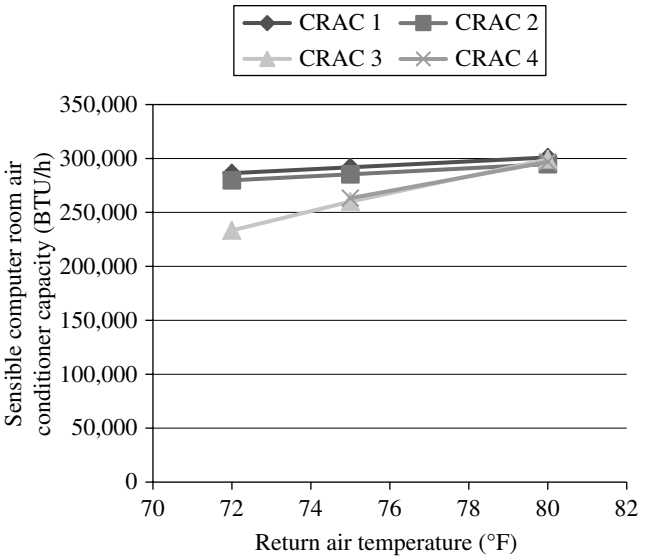


FIGURE 24.1 Impact of return air temperature on computer room air conditioner cooling capacity varies by specific model but can be significant.

24.5.4.2 Balanced Distribution

Recognizing how ITE uses air for cooling, the distribution system can be designed to supply cooling air directly to the ITE intakes in approximately the quantity required. The return can then be configured to collect the hot exhaust air. For example, a design that uses a raised floor as a supply distribution plenum places the air-flow supply tiles at the front of ITE, where the cooling air is drawn into the rack and return is drawn through the space hot aisles or a ceiling plenum to the air handler. Another approach would be to use in-row air handler units that monitor the inlet temperature of adjacent racks and modulate cooling supply airflow to attempt to match the volume of air the ITE requires.

This approach is effective at reducing hot spots and improving system air temperature difference. The main limitation is the difficulty in balancing supply volumes to match the ITE air demand on a localized basis. If too much air is supplied from the tile in front of an IT rack, then some will bypass to the return and be wasted airflow that provided no useful cooling. But if not enough air is supplied, then hot air could be recirculated and sucked in from the IT exhaust area, usually into the pieces of equipment located at the top of the rack. In practice, it is very difficult to get and maintain this balance, but this limitation is primarily one of cost—operating cost and equipment cost due to capacity loss to bypass. Fan-boosted floor tiles are now available that can adjust air delivery via rack inlet temperature sensors, resulting in very

good airflow matching and management from raised floor systems. However, the tiles are expensive, fans use energy which offsets some of the efficiency savings, and since fans take the air they want, they can air-starve other cabinets supplied by conventional tiles. A balanced distribution system can carry high loads if it is sized properly.

24.5.4.3 Containment

Airflow management (also see chapter 34) has become so critically important to data center operational reliability and energy efficiency that virtually all new facilities are built with some form of aisle containment. The improvements are so great that many existing data centers are being retrofit with containment solutions as well. Containment is merely an extension of the decades-old "hot aisle / cold aisle" cabinet arrangement where equipment racks are placed front-to-front and back-to-back to minimize the amount of hot exhaust air that can enter the intakes of equipment in the next row. This concept was a giant leap forward when it was introduced, and served well until cabinet heat densities rose astronomically and the resulting need for much higher airflows made air control more challenging. Containment is simply the addition of air barriers above cabinets, blanking panels between empty U's for ITE and at the ends of aisles to enclose either the air delivery (cold aisle) or the air exhaust (hot aisle) spaces. The approaches are known as "cold aisle containment" or "hot aisle containment" (see figure 24.2). Only one or the other is used at a time, and each has its advantages and disadvantages, as well as its adherents and detractors.

Further, "containment" can be either "full" or "partial". Full containment, as the name implies, requires the installation of air barriers above the cabinets as well as at the ends of rows so the hot or cold air space is fully enclosed. Partial containment generally encloses only the ends of the aisles, and has been shown in actual practice to be as much as 80% as effective as full containment in reducing air mixing. Partial containment is more often seen in retrofit installations where full containment would require a prohibitively expensive and disruptive renovation of the fire protection system. The actual level of effectiveness of a partial containment solution depends on numbers of other factors, such as air velocities and ceiling conditions, but the important point is that partial containment is still better than no containment at all.

There are many containment solutions on the market, ranging from relatively simple plastic curtains that can be easily field cut to size and hung in place, to complete systems integrated with the equipment cabinets. The best full containment systems use plastic panels fit into place and sealed above cabinets, plus end panels with doors that auto-close with air-tight seals. But whichever solution is chosen, it is important to confirm that it meets the NFPA-75 standard for fire protection of information technology equipment (ITE), whether or not that standard is adopted as "code" in the jurisdiction. NFPA-75 includes very specific requirements for the fire ratings of containment materials, but perhaps even more important are the requirements in the event of a fire. As noted above, partial containment solutions are often implemented in existing facilities to avoid a costly and disruptive revamp of the fire protection system, which may have heads in every other aisle, or over cabinets so they span two aisles. In order to provide full containment solutions without requiring massive fire protection system changes, manufacturers developed several solutions for dropping containment curtains or panels out of the way in the event of a fire. Most of these were later rendered unacceptable by NFPA-75 for two reasons. Those that relied on fusible links were useless with gas-based fire suppression systems because they required a raging fire to melt the links – exactly what gaseous systems are meant to

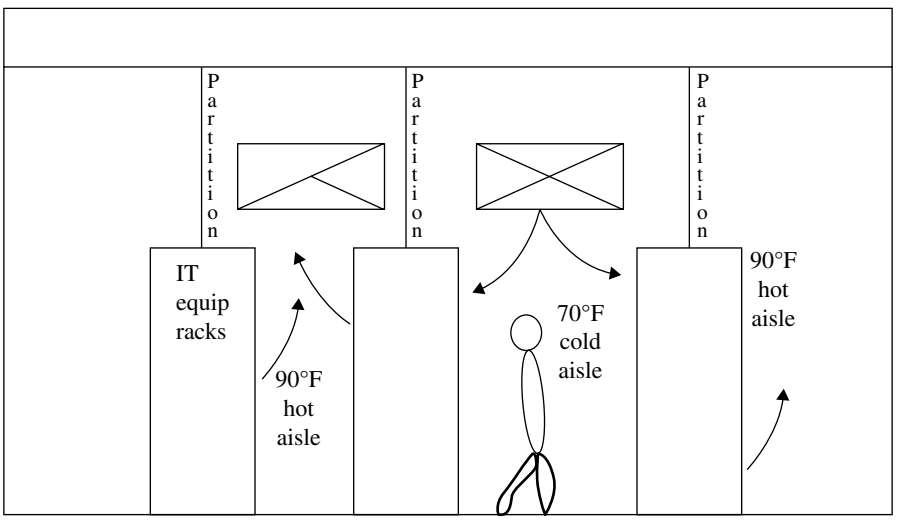


FIGURE 24.2 Hot aisle–cold aisle.

avoid. And those that use other means, such as heat shrinking, to drop curtains and ceiling panels, introduced obstructions into the fire egress path – a clear safety hazard. In short, it is imperative that the cooling system be designed in combination with the cabinet layout, containment solution and fire protection system, and that both the design engineers and the authority having jurisdiction (AHJ) agree that they comply with NFPA-75. Some manufacturer's claims of compliance are highly questionable.

The decision as to whether to use hot-aisle or cold-aisle containment involves several considerations, and is obviously critical since it becomes an integral part of the cooling system design. It is generally accepted that hot-aisle containment is slightly more energy efficient than cold-aisle containment, as well as easier to implement successfully. But cold-aisle containment is more common in retrofit installations unless a return air plenum ceiling already exists. The reasons are rather straight-forward, and are directly related to the air balancing discussions in previous paragraphs. With hot aisle containment, the only requirement is that sufficient air be delivered to the equipment. Assuming there is enough air volume available from the cooling units to satisfy the total needs of the space, if the floor tiles or overhead ducts in one aisle are delivering slightly less air than the IT hardware requires, the ITE will pull additional air from other parts of the room, albeit at in increased fan energy cost. Further, it will pull only cold air since the hot aisles are "contained" so their air becomes unavailable. And since a full hot aisle containment system, which includes unused rack spaces being properly sealed, also prevents cold air from bypassing cabinets and entering the hot aisle return air stream, the air conditioners receive the highest temperature return air the room can deliver this maximizes coil delta T, cooling efficiency and cooling unit capacity.

With full cold-aisle containment, the ITE can only use the air that is delivered to the aisle, making it critically important to correctly manage airflow. Pressure differential sensors are often used in addition to temperature sensors to ensure this, which make a more complex control system and also relies on sensor types that can go out of calibration over time so must be checked periodically. With partial cold-aisle containment, of course, the ITE can still pull air from wherever it is available, but that might include hot exhaust air which reduces the effectiveness of the containment solution.

The other consideration in making the decision between the two approaches is human comfort. With the higher inlet temperatures now recommended by ASHRAE, even a 20°F (11°C) temperature differential (TD) through the ITE can result in hot aisle temperatures above 100°F (38°C). And since most equipment now runs with TD's of 25°F (13.5°C) or more, hot aisles can be very uncomfortable, particularly when "contained". OSHA heat exposure levels of concern have not yet been met in most data centers, but they are getting close, and technician health and comfort should also be a consideration, along with the effect excessive heat might have on work accuracy in critical operations, regardless of legal requirements. With hot aisle containment, the rest of the space is at cold aisle temperature, which can still be warm by legacy standards, but not intolerable over a full day. Cold-aisle containment, on the other hand, puts the rest of the room, including communicating aisle ways, at the hot-aisle temperature. Depending on how people generally work in the facility, this can be a consideration if they can't regularly leave the high temperature space.

Regardless of the chosen approach, modern designs using air cooling should always include some form of air containment in order to maximize both cooling effectiveness and energy efficiency.

24.5.4.4 Rack Level

Rack-level management of the hot exhaust can contain the hot air exhaust within the IT rack itself. This can take several forms, from incorporating a cooling coil directly in the rear door of the rack to a small traditional air conditioner integrated entirely in the rack itself, to a ducting system that collects hot air and directs it up a "chimney," and to an overhead return. These systems can carry very high loads without hot spots overheating adjacent systems. Any power consumption associated with non-passive systems, in the form of internal fans, pumping, compressors, or other associated equipment, should be considered in the electrical system design and operational maintenance and power costs.

24.5.4.5 Hybrid

Using a single airflow management design often has long-term operational benefits from having a common operational and maintenance profile, but it is not a requirement. Mixing airflow design approaches is often the best solution when a data center is expected to carry ITE with significantly different load concentrations or airflow characteristics. Many of the aforementioned airflow methods may be combined within the same data center to achieve the best balance of system flexibility, reliability, first cost, and efficiency. A partitioned data center may utilize rack-level cooling within contained aisles to carry a few unusually high-load ITE racks. Legacy equipment such as a low-load old tape library system may be conditioned in a fully mixed airflow portion of data center floor, while high-load modern IT with standardized front-to-back airflow uses a hot/cold aisle arrangement with full partitions containment. Active fan tiles can offer short-term "fixes" to existing data centers with underfloor air supply plenums and hot spots while a severely congested underfloor plenum issue is addressed.

24.5.4.6 Future

The airflow management design configurations discussed here range from common legacy to current leading-edge approaches. However, there is little doubt that new airflow management variations will appear in the future. New approaches should be evaluated on the basis of their reliability, flexibility, and efficiency, The served ITE needs should also be continuously evaluated; beyond efforts to push the maximum operational temperatures up, some new server designs incorporate integrated convective spaces, heat pipes to a common backplane heat sink, and other unusual features that radically change the airflow management requirements and opportunities. While the principle of isolating the waste heat exhaust from the cooling air intake is as fundamental as drawing drinking water from a river upstream of where raw sewage is dumped, the details of the various airflow management techniques to achieve this are still maturing and solutions like rear-door rack cooling doors actually require air re-circulation to work..

24.6 CURRENT BEST PRACTICES

Data center design continues to evolve and is ultimately dictated by the needs of the client, but it is possible to identify current best practices. Good data center design must meet the specific needs of the location and the client and may not achieve all of these best practices, but it should consider all of these approaches.

24.6.1 Redundancy

Redundancy is a defining feature of data center design. Providing $N+1$ redundancy for all critical components is standard best practice for the vast majority of facilities. It provides cooling continuity in the even of a cooling unit failure, but more important, it provides concurrent maintainability, enabling any one unit to be taken out of service for preventive maintenance without under-cooling the critical load.. However, a best practice design will fully define and document the owner's needs through the design process. Opportunities to reduce the redundancy are sometimes available and appropriate to reduce construction and operation costs. Design decisions such as not providing fully redundant chilled water piping, designating a portion of the data center as nonredundant (most easily verified by the omission of UPS to the ITE and depending on temporary rental equipment to provide servicing redundancy are examples of redundancy reductions that may occur through a best practice design process. Adding an air-side economization system to a data center or a cogeneration system with on-site fuel storage in addition to emergency generators is an example of design decisions that may be made to add redundancy beyond the generally recommended standard $N+1$.

24.6.2 Reliability

The reliability impact of all system design decisions should be fully evaluated, rather than depending solely on assuming replication of past practice is adequate. The reliability benefit of free cooling can be significant and should be acknowledged in design evaluations. The reliability risk of a humidifier (with domestic water supply piping) in the space likewise should be evaluated. All design features should be similarly assessed for reliability impacts.

24.6.3 Layout and Air Management: Hot Aisle–Cold Aisle

In a high-density data center, the layout of the IT equipment is an integral part of best practice mechanical design. The layout must prevent hot air from being exhausted from one piece of equipment into the intake of another piece of equipment. Installing the ITE so that the hot air is exhausted to the same area as adjacent and opposite ITE is commonly referred to as creating a hot aisle–cold aisle configuration. For equipment with front intake and rear exhaust, this takes the form of arranging rows so that the hot air is exhausted into dedicated hot aisles, while the intake side of the rack is served from dedicated cool aisles: a hot aisle–cold aisle arrangement. Equipment with side exhaust may employ shrouds to direct the exhaust to the hot aisle or incorporate vertical chimneys to exhaust it to a common overhead "hot" plenum. Partitions can be used to create a small hot aisle enclosure, or the cold aisle enclosure could be smaller, but regardless of the specific form it takes, the ultimate function of aisle containment is to prevent recirculation causing hot spots that could damage equipment. In a best practice design, the hot aisle is also capitalized on to provide a return hot air stream that is 20°F or higher than the room set point temperature (Fig. 24.2).

Creating a hot aisle heat exhaust improves a number of free cooling design options that can significantly improve reliability. The capacity of the cooling equipment is typically increased by the higher-temperature return, allowing it to support a higher IT load per air handler than lower-temperature return designs. Rejection of heat can also be done more efficiently at higher-temperature differences—the higher the temperature of the waste heat coming from the data center, the more easily it can be rejected to the heat sink (usually the outdoor ambient environment), allowing for higher compressor cycle efficiency. A high-temperature waste heat stream also can allow a higher-temperature differential between the supply and return streams (water or air), reducing flow requirements and energy consumption. In some cases, the high-temperature return airstream not only increases efficiency and system capacity but also offers a practical source of free heat for adjacent spaces.

24.6.4 Liquid Cooling

Moving heat with liquid is far more efficient than moving it with air, with a small pump system having a heat moving capacity an order of magnitude bigger than an equivalent fan system. Typically, the closer to the IT rack that heat can be transferred to a liquid loop, the greater the efficiency. Liquid cooling can also offer more flexibility since a great deal of future capacity can be added by increasing a pipe diameter an inch or two; adding an equivalent amount of future capacity to an air-based system would entail adding a foot or two to the ducting dimensions.

Not all facilities are ultimately appropriate for a liquid cooling system, but it should be considered. ASHRAE TC 9.9 publishes thermal guidelines for various classes of liquid cooled equipment, similar to what they publish for air cooled systems. Liquid cooling is also thoroughly covered in Book #4 of the Datacom series (see Chapter 11).

24.6.5 Optimize Space Conditions

With the notable exception of a very few pieces of legacy equipment and some tape drives, modern ITE can operate reliably, and without warranty or degradation issues, at inlet temperatures significantly higher than legacy designs. The ASHRAE Thermal Guidelines and Thermal Envelope define six categories of air-cooled ITE, with the most stringent being A1. This category covers most data center hardware, and allows a constant, year-around intake air temperature of 27°C (80.6°F). This temperature was selected, not because higher temperatures will harm equipment, but because it is the temperature at which server fans begin to seriously ramp-up speed. As noted previously, fan power increases as the cube of the speed, so significant speed increases in server fans (which may be numerous in some hardware), can quickly offset any energy savings gained by reducing cooling demand. When the guidelines were published, they were the result of agreements among all the major IT hardware manufacturers, who revealed their internal test data to each other under strict non-disclosure. They acknowledged that higher temperature operation would not affect the vast majority of legacy hardware, and also that operating at the higher temperatures would not void warranties, even though owner's manuals usually specified lower temperature conditions. The manufacturers further agreed that equipment could operate at temperatures as high as 32°C (89.6°F) for as long as three days without degradation. This "allowable envelope" enables more hours and days of free cooling, even when outside air temperatures rise a few degrees above the "recommended" level, and reduces opportunities for cooling failures as systems are switched between mechanical and outside air cooling. It further relieves data center operators from concerns about increased temperatures in the event a partial cooling failure takes more than a day or two to repair.

The thermal guidelines also specify low temperatures and thermal gradients (rates of thermal rise), and the newest edition includes data charts that can be used to predict the possible reduction in ITE longevity if equipment is operated continuously at temperatures even higher than the ASHRAE recommended maximums. Some operations find that, despite the increased fan speeds, they can save a considerable amount of energy by operating at elevated temperature. These operations are also unconcerned about reduced service life because they have a high rate of equipment "churn", replacing the entirety of their processors every two to three years anyway, or even more often.

ASHRAE emphasizes, however, that the highest allowable temperature recommendation is not meant to encourage the uninitiated to simply turn up CRAC set points. Doing so without complete knowledge of the airflow characteristics in the room could result in cabinets receiving intake air at temperatures even higher than the "allowable" maximum. In rooms with less than ideal airflow management, it may well be necessary to operate with CRAC discharge temperatures of 70°F (21°C) or lower in order to maintain anything close to 80°F or 27°C at the worst-case racks. While a perfect air delivery system would discharge at the same temperature as was supplied to it, there is no such thing as a perfect system. Heat gains occur along the entire path, and since even the best containment system is not hermetically sealed, there will always be some air leakage and recirculation. Best practice is often for the design engineer to base calculations on a CRAC discharge temperature of 75°F (24°C), and design for a maximum 5°F (2.8°C) temperature rise to the worst case locations.

24.6.6 Economization

Any best practice design fully investigates a careful implementation of economization to increase redundancy and allow for very low power cost cooling when it is cool outside—remember that even in a snowstorm the data center needs cooling. The normal operational benefits of economization are significant energy savings, sometimes reducing total data center power use by 25% or more, but the reliability benefits are potentially more valuable. It is best practice for even the most fail-safe design to consider the worst-case scenario. If the mechanical cooling system fails beyond the ability of redundancy to cover, a parallel economization system may still be functional and able to keep the space operating until a fix can be applied. Note that even if it is a summer day and economization can only hold a facility to 95°F (35°C), it may be enough to keep the housed ITE from dropping off-line—which is a vastly preferable failure mode than the alternative of ITE hitting the temperature at which it automatically shuts down due to runaway overheating.

The energy benefits of economization are significant in almost every climate zone, particularly if air management offers the potential for a heat exhaust at over 90°F (32°C).

24.6.7 Cooling

For large data centers, over 1 MW, a centralized evaporatively cooled central plant is best practice—although climate and local regulations that may result in cooling tower shutdowns can dictate an air-cooled design. Smaller data centers often utilize more modular air-cooled equipment that offers cost and control benefits but at higher operating costs. The ideal, which requires a rare combination of ITE, client, and climate, is elimination of mechanical cooling entirely through the use of economization and evaporative cooling. The ASHRAE Thermal Guidelines are specifically designed to encourage achieving as close to this scenario as possible.

Central control is best practice, whether in the form of a building DDC system or networked integrated controls, to avoid the common problem of adjacent cooling systems simultaneously humidifying and dehumidifying due to sensor drift over time. The system should provide comprehensive alarms to identify failures. Integrated power monitoring can pay for itself by identifying operational opportunities to reduce power bills and predicting equipment failures by alarming the increase in power consumption that often precedes them. Comprehensive DCIM (Data Center Infrastructure Management) systems, which are using more AI analysis every year, can pay for themselves in energy savings from thorough power and cooling analysis, as well as by alerting to potential equipment failures before they occur.

Best practice includes some form of economization, be it a properly filtered air-side economizer, a water-side economizer system configured to allow stable transition between chillers and water-side economizer, a dry cooler-supplied glycol coil, or a pumped refrigerant-based system integrated into some newer compressor-based systems. The added redundancy benefits of a second source of cooling (even if intermittent) combined with a constant internal cooling load 24 hours a day during all weather conditions make a free cooling system best practice in all but the most severe and humid climates.

24.6.8 Humidity Control

Best practice is to minimize the humidity control in data centers to match the actual requirements of the ITE. ASHRAE TC9.9 guidelines offer a best practice range, to be adjusted to client requirements as necessary. A legacy control band of 45 ± 5% is often requested, but is no longer necessary – at least not to control static discharge which has historically been the reason for maintaining relatively high humidity levels. Humidification also incurs significant operational and first costs, and is terribly energy inefficient.

Humidity should not exceed 60% RH to avoid the possibility of gaseous contaminants becoming corrosive acids, but RH levels as low as 8% are also acceptable, greatly reducing the need for humidification, and even eliminating its need in most locations. When humidification is required, however the large amount of waste heat continuously available makes adiabatic humidification the ideal approach. Adiabatic humidification is the only solution appropriate for large humidification, such as may be seen with an air-side economizer system. Ultrasonic adiabatic humidifiers offer high precision and several market options, but an atomizing nozzle or media-based humidifier will typically be a lower-cost solution.

24.6.9 Efficiency

Efficient design is no longer an option. ASHRAE 90.1-2020, which references the ASHRAE 90.4 standard as the compliance path for data centers (facilities with IT loads greater than 10 kW and 20 Watts/ft^2 or 215 Watts/m^2) will be adopted as code in most every federal, state and local jurisdictions by 2023 if not before, as well as in many places abroad. The energy demands of modern data centers are such that operations can no longer ignore energy efficiency and the accompanying economics. ASHRAE 90.4, however, is a minimum performance standard. Efficiencies greatly exceeding the standard minimums are quite achievable, and the annual energy savings realized can often justify additional first costs.

24.7 FUTURE TRENDS

The changes in data center design and efficiency between 2010 and 2020 have been so extensive they have been difficult to keep up with. With the rate at which IT systems continue to evolve, it is likely that changes in the next decade will be just as dramatic, if not more so, and that the infrastructure that supports them will need to change significantly as well. The industry has been far ahead of government mandates in dramatically improving energy efficiency, in both the IT hardware and the cooling systems that support it. That trend is expected to continue as the demands for automation, machine learning and social media continue to grow, and the power demands to support these technologies become more strained. The rapid adoption of cloud computing, software as a service (SaaS) and other remote, hosted, and virtual approaches have made data center designs increasingly challenging, as planning must now include preparations for both expansion and contraction. In fact, the financial term “fungible” might well be applied to the requirements of modern data center design as enterprises experience three factors that can change power and cooling demands virtually over-

night: mergers and acquisitions that suddenly explode data center size; off-loading of significant parts of computing operations to the cloud or to hosting sites that reduce energy demand; and the return of critical computing to the enterprise data center when remote operations are found to be unreliable, insecure, overly expensive, or insufficiently responsive to immediate needs, increasing energy demand again. Codes and standards will impose increasingly stringent levels of energy efficiency, but only in the planning and permitting stages of design. It will be the responsibility of the designer to configure infrastructure in ways that can respond to changes in size in both directions while maintaining levels of energy efficiency that continue to meet, or preferably significantly exceed, legal design requirements.

24.7.1 Water in the Data Center

Some data center operators do not want any chilled water or condenser water piping within the data center footprint out of concerns of a leak flooding the data center and causing a catastrophic failure. With many of the most efficient system design approaches based on a central chilled water plant approach, this can be a significant limitation on the design. Actual failure of operating chilled or condenser water piping is very rare, and many operating data centers do utilize chilled water piping across the data center floor.

Design approaches can reduce the risk inherent with water in a data center. Common techniques include placing the piping underfloor, using water sensors for alarm, and alarming any automatic water makeup system on the chilled water loop. Automatic shutoff and isolation of piping based on water sensors to automatically isolate leaks is feasible but should be approached with great caution since a false alarm may be more likely than an actual piping failure and may even become the cause of a catastrophic shutdown; even with a real leak, a data center may remain up and running with several inches of water in the underfloor but overheat and shut down in minutes if cooling is lost. ITE heat densities continue to increase, with liquid cooling becoming increasingly necessary in more than just "supercomputer" systems. This makes it imperative to consider infrastructure for liquid cooling in every data center design, and for designers to become aware of the ways in which it can be done with minimal risk. ASHRAE Book #4 in the Datacom series delves deeply into this subject.

24.7.2 Hot Data Centers

Early data centers were kept at temperatures much cooler than office spaces because that's what the technology of the time required to maximize component reliability and longevity. As technology improved, cooling continued to be over-designed because computer manufacturers required it, as well as out of an abundance of caution on the part of designers. But that changed in 2004 when ASHRAE Technical Committee TC 9.9 got all the major manufacturers to reveal their actual test data. It became apparent that ITE manufactured as long ago as 1990 could operate with full reliability at much higher temperatures than were being specified. The critical need to reduce energy consumption led ASHRAE to publish the Thermal Envelope which has an upper temperature limit of 80.6°F (27°C). ITE can operate reliably at even higher temperatures, but the reasons for choosing this particular temperature have been discussed in previous sections as well as in Chapter 11. Also previously discussed in this chapter are the cautions and guidelines regarding airflow, temperature and humidity sensing, working environments for personnel in contained aisles, and the establishment of cooling unit set points. The important thing for designers to remember is that modern data centers should be operating with the least mechanical cooling possible for the most hours and days possible, and at the highest temperatures the overall space will allow, in order to minimize the use of electrical or any other form of energy. Designers should also be aware of the full spectrum of ASHRAE guidelines which include even higher temperature normal operating conditions for more robust ITE that is either already here, or that will almost certainly be coming in the future. The day could very well be close when mechanical cooling is no longer necessary in even the hottest parts of the world, and the mechanical engineer is almost entirely concerned with air or liquid flows. Exceptions certainly do, and will likely continue, to exist, with mechanical storage media (disks and tapes) more sensitive to temperature than processors or memory chips, but that too could change is ever more efficient storage systems are developed.

24.7.3 Questioning of Assumptions

Perhaps the safest prediction also serves as an overall final approach for all good quality design: traditional design assumptions will be challenged more frequently and overturned more often as data center design continues to evolve. The current maturity of the data center market has introduced more competition and a focus on finding cost reductions everywhere, including in the mechanical system design. The traditional data center configurations, such as stand-alone CRAC controlling to return air temperature and using raised floor distribution, are revered as "tried and true"—safe options in a field where reliability is job one. But the long-established legacy approaches are often not the most cost effective or even the most reliable design for the current and future data center. A proven design that has worked in the past is may be a reliable option; since it has been shown to work! However, the proven designs were proven on the ITE loads of years ago

and often have implicit assumptions about ITE and mechanical equipment capabilities underpinning them that are no longer valid.

Often, even user functional requirements (such as the need for cabling space or control of electrostatic discharges) are incorrectly presented as system requirements (a raised floor or humidifier system). Good communication between the design team and the user group to separate the functional requirements from traditional expectations helps the mechanical designer identify the best system to meet the specific demand.

Now and in the future, the best mechanical design will be recognized for how it excelled at meeting the needs of the client, not the assumptions of the design team.

ACKNOWLEDGMENT

This chapter is adapted from the first edition of Data Center Handbook and has been updated by its Technical Advisory Board. Sincere thanks are extended to Mr. John Weale and the members of the Technical Advisory Board who spent invaluable time in sharing their in-depth knowledge and experience in preparing this updated version.

REFERENCE

[1] ASHRAE Technical Committee 9.9. Mission critical facilities, data centers, technology spaces and electronic equipment. http://tc0909.ashraetcs.org/about.php. (Accessed 9/28/20)

FURTHER READING

ASHRAE. Datacom Series (14), ASHRAE Technical Committee 9.9, https://www.ashrae.org/technical-resources/bookstore/datacom-series. Accessed 9/28/2020.

ASHRAE, Technical Committee 9.9. Data center networking equipment—issues and best practices; https://tc0909.ashraetcs.org/documents/ASHRAE%20Networking%20Thermal%20Guidelines.pdf. Accessed 9/28/2020.

Open Compute Project. https://www.opencompute.org/. access 9/28/2020.

Eubank H et al. Design Recommendations for High Performance Data Centers. Snowmass: Rocky Mountain Institute; January 2003.

Lawrence Berkeley National Laboratory. CENTER OF EXPERTISE for Energy Efficiency in Data Centers. https://datacenters.lbl.gov/. Accessed 9/28/2020.

25

DATA CENTER ELECTRICAL DESIGN

Malik Megdiche[1], Jay Park[2] and Sarah Hanna[2]
[1] *Schneider Electric, Eybens, France*
[2] *Facebook, Inc., Fremont, CA, USA*

25.1 INTRODUCTION

In order to design an optimal data center, one must go through the process of determining its specific business needs. Planning and listing the priorities and the required functionality will help determine the best topology for the data center. Outlining the key ideas and concepts will help structure a focused and effective document.

To adequately define the basic functionality requirements, business needs, and desired operations of the data center, consider the following criteria:

- The facility's uptime
- The electrical equipment to be deployed
- The electrical design strategy

The basic requirements, business needs, and desired operations are collectively known as the *backbone requirements*. Based on these requirements, the designer of the power system will first need to answer the several questions before going in the classical steps of an electrical design as shown in Figure 25.1.

25.2 DESIGN INPUTS

25.2.1 Uptime Level

First, determine the required uptime of the facility. Can the system allow some downtime?

If it can, you must address how much downtime can occur without affecting business operations. Due to the criticality of their businesses, financial institutions, colocation facilities, or institutions directly related to revenue generation require the highest levels of uptime. Less mission-critical organizations have the flexibility to lower their uptime requirements significantly.

To determine the criticality of the information technology (IT) process, several questions need to be asked:

- Can the IT process tolerate some interruptions and how many interruptions per year?
- Is the IT process providing IT redundancy so that a part of the IT loads can be shut down without IT process loss of continuity?

Several types of power system failures can be defined as follows:

- Grid short disturbances such as severe voltage sags (10–30 times/year) and short interruptions (0.5–10 times/year)
- Grid long interruptions (0.1–1 time/year)
- Site electrical distribution equipment failures (0.01–0.001 failure/year for a switchboard) and planned maintenances (each approximately 3–5 years)

Depending on the criticality of the business, the data center facility needs backup equipment to face or not one or several of the power system interruption. Several levels of service continuity can then be defined:

- Level 1: No redundancy
- Level 2: Backup in case of grid short perturbations
- Level 3: Backup in case of grid short and long interruptions

Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center, Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

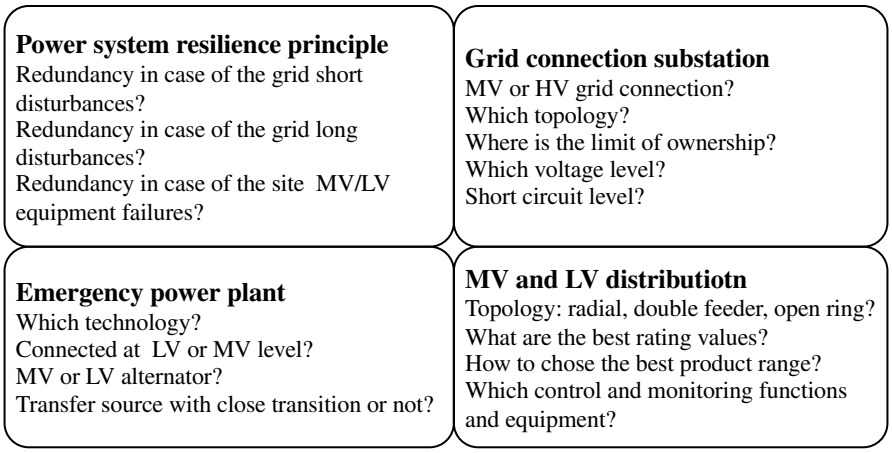


FIGURE 25.1 Questions related to the data center power system. Source: © Schneider Electric.

- Level 4: Backup in any case (grid or on-site distribution equipment failures) by achieving full redundancy on the whole data center facility

Reliability performances in between level 3 and 4 are also common when considering an architecture that is redundant in most of the cases but still one or several rare failures that could lead to the loss of servers (single failure points).

25.2.2 IT Load

A server rack is characterized in an electrical point of view by:

- The rated voltage and power,
- The diversity factor to take into account the real power usage of the racks,
- The power factor,
- Single phase or three phases,
- Single or dual inputs,
- Dual power supply (that takes 50/50% on each cord) or single power supply with embedded static transfer switch (STS) (that takes 100% one side and 0% on the other side),
- Its ability to withstand voltage drop (normally in accordance with the ITI curve),
- The maximum residual current leakage,
- The maximum inrush current,
- The maximum total harmonic distortion of current (THDi) at full load,
- The configuration of the IT room determined by the number of racks per row, the number of rows, and the number or row per IT room.

Classically design institute takes a power factor of 0.9, but nowadays new server power supply shows power factor above 0.95 (Fig. 25.2). The THDi are also now below 20% at full load.

During operation, it is quite common that the data center power systems are underloaded (about 50% of the design capacity) due to the gap between the installed capacity and the actual load because there is an uncertainty on the actual load (server load factor and IT load planning). A way to improve the cost effectiveness of the power system is to provide a power system that is modular and scalable to match the IT growth plan and also to use some diversity factors (also named as "overbooking") to take into account the real power usage of the servers while keeping a safety margin to ensure no risk of overloading the distribution equipment (Fig. 25.3).

25.2.3 Cooling Loads

Depending on the cooling technology selected for the data center, for each mechanical load [chillers, pumps, CRAH (computer room air handling) units, fans, direct expansion (DX) units, etc.], consider the following:

- Rated power
- Rated voltage
- Maximum load consumptions according to extreme conditions
- Starting currents
- Sensitivity to power supply interruption (loads that are not supplied through UPS (uninterruptible power supply) need to be able to restart after the generator backup has reenergized the system without any disruption of the servers)

25.2.4 Other Loads

Building loads (lighting, security, access control, fire protection, control rooms, offices, storage rooms, etc.) and critical auxiliaries (for backup generators and medium voltage (MV) and low voltage (LV) switchboards) that not only represent a small part of the power but also are important for the data center operation.

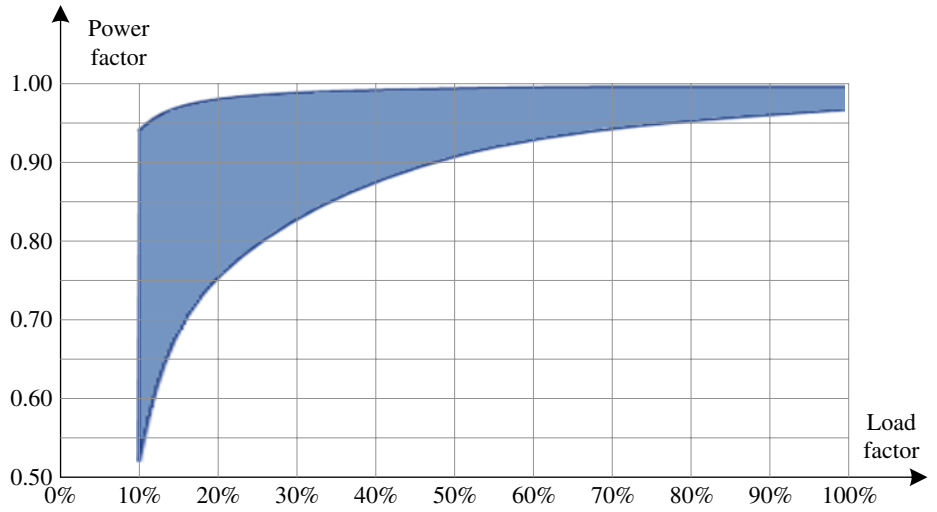


FIGURE 25.2 Measurements of actual power supply power factor. Source: Data extracted from APC White Paper 200.

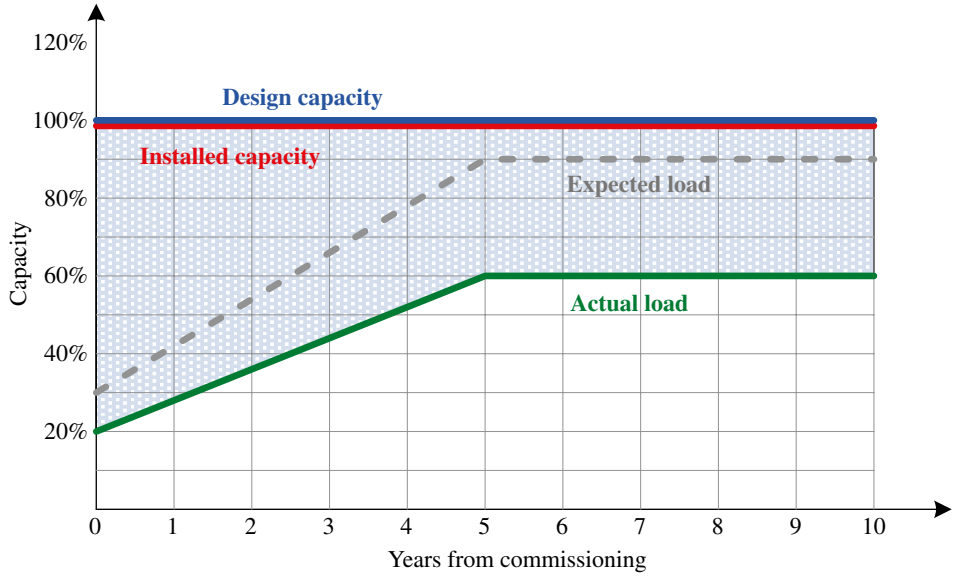


FIGURE 25.3 Design and installed capacity during the life of data center. Source: Data Extracted from APC White Paper 37.

25.2.5 Modularity and Scalability

Depending on its IT growth plan, the data center facility will need a level of modularity and scalability in order to grow up in accordance to the IT growth to optimize the capital expense investment.

25.2.6 Key Performance Indicators

The main key performance indicators of the data center owner such as reliability and availability target, capital expense, operating expense, and/or footprint are valuable information for the data center designers during the optimization of the overall infrastructure.

25.3 ARCHITECTURE RESILIENCE

25.3.1 Grid Supply Reliability Performance

The architecture of the HV/MV grid of a country, shown in Figure 25.4, consists of the following:

- Power generation sites about hundreds of MW interconnected by a meshed HV network called the transmission system.
- The sub-transmission system is used to supply the HV/MV substation using also a meshed topology.
- From the HV/MV substation, MV distribution networks are supplying the MV customers and MV/LV

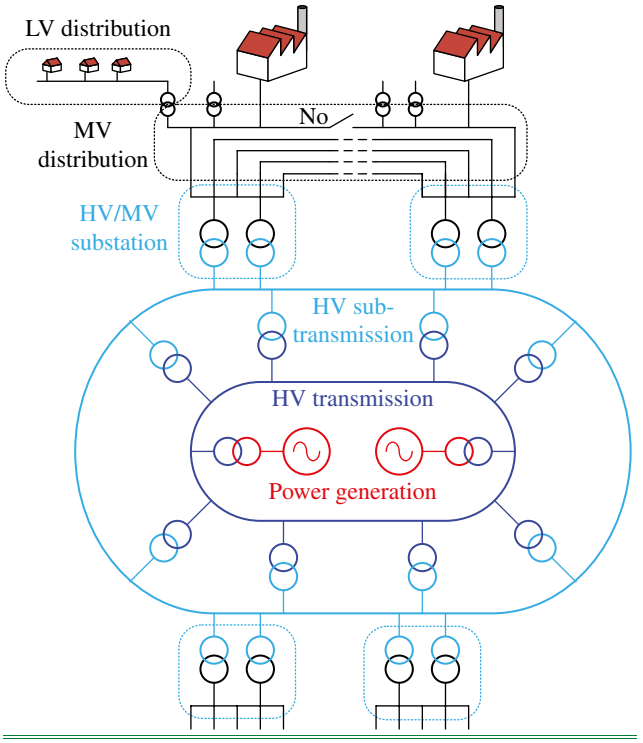


FIGURE 25.4 Topology of the utility grid in a country. Source: © Schneider Electric.

distribution transformers using classically open loops topology to be able to reconfigure the power system in case of a fault on the MV feeders.

- From the MV/LV transformers, the LV networks supply the LV customers using a radial architecture.

At HV level:

- When a fault occurs on the HV lines or substations, the fault is cleared in less than 100–200 ms without losing any customers thanks to a meshed topology equipped with zone protection substations. However during the fault clearing, a voltage drop will affect several substations next to the fault.
- When the HV power system experiences very critical events, several emergency actions can be performed to avoid a complete blackout including shedding low priority MV feeders or HV customers.
- If a blackout occurs at HV level, the HV system follows a restart procedure to recover step by step the different parts of the system and will take few hours.

At MV level:

- When a fault occurs on an MV feeder, the fault is cleared in 150–500 ms by the feeder protection. The protection includes classically several reclosing operations.

TABLE 25.1 Electrical utility reliability indexes

Reliability performance of electrical utility in Europe	Order of magnitude frequency (per year)	
	MV grid connection	HV grid connection
Regional blackout > 4 hours	0.01–0.02	0.01–0.02
Long interruptions > 3 min	0.5–5	0.1
Short interruptions < 3 min	0.5–20	0.5
Severe voltage drops	10–100	5–15

Source: © Schneider Electric.

- If the fault is momentary, the restoration time can be 500 ms to several seconds. If not, a fault location and restoration process will take few minutes to few hours to locate, isolate the faulty section, and restore the healthy sections.

The reliability performance of the grid supply depends on the country and the location in the country. To give a rough idea, Table 25.1 shows an idea of the utility reliability when connected at MV level or HV level.

Depending on the data center site maximum apparent power and according to the grid characteristics, the data center site will be supplied with an MV grid connection or with an HV grid connection with a dedicated HV/MV substation.

25.3.2 Backup of the Grid Supply in Case of Short Interruptions

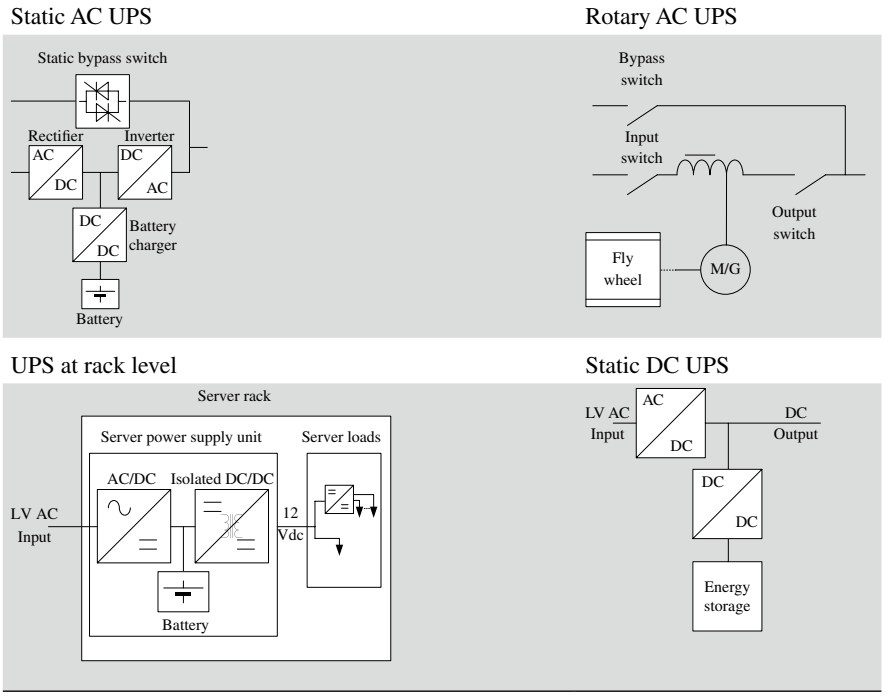
In case of utility short interruptions (below 3-min interruptions), UPS technologies with small storage are needed. Several technologies can be used shown in Table 25.2. The most widely used in data center are the static UPS using lead–acid batteries or lithium batteries. As UPS and its energy storage are important parts of the LV equipment cost and losses, there is a big trend to find new ways to improve both CAPEX and OPEX (capital expenditures and operating expenditures) of the UPS products:

- New static UPS products using new power conversion topology
- UPS function embedded at the server power supply unit (PSU) level
- Investigations in the area of the direct current (DC) UPS where the output is in DC instead of alternative current

The UPS energy storage autonomy can be set according to:

- The generator starting and loading time
- The IT requirements like the IT process emergency shutdown that will need the time to properly shut down the IT to not lose any data before the power blackout.

TABLE 25.2 Different UPS technologies



Source: © Schneider Electric.

25.3.3 Backup of the Grid Supply in Case of Long Interruptions

For a backup in case of long interruption, the most common solution is the use of standby generators with fuel storage associated with a PLC (programmable logic controller) that manages the start and stop sequences and an automatic transfer switch that allows the switch between utility and generator backup as shown in Figure 25.5.

A redundant utility grid incomer and substation can be a solution but, in case of HV blackout, there will be always a risk to lose the two redundant grid substations at the same time. Moreover, depending on the grid characteristics, a redundant grid connection can be more expensive than a standby diesel power plant. Due to this single failure point, data centers use diesel backup generators. However, there are some cases where it could be interesting such as a data center an HV grid connection using redundant HV lines, a redundant HV/MV substation and an IT process that can tolerate a black out each 50 years.

Other technologies are also investigated such as gas generators and fuel cells.

25.3.4 Backup in Case of a Failure in the Data Center Electrical Distribution

There are three main hierarchies of electrical data center design: N, $N+1$, and $2N$. The N design system uses the exact number of equipment or systems without any built-in redundancy. $N+1$ designs have one additional system built in for redundancy, while $2N$ refers to designs that have double the equipment required, which provides maximum redundancy.

25.3.4.1 N *Topology*

This topology has no redundancy except a backup in case of the utility failure using a UPS unit with a standby generator as shown in Figure 25.6. Using this topology, several single failure points will lead to server shutdown including failures and planned maintenance on the LV distribution.

25.3.4.2 2N *Redundant Topology*

In this topology, power flows from the utility through the UPS/power distribution unit (PDU) of two separate systems and connects to the server. A $2N$ configuration, as shown in Figure 25.7, provides redundancy throughout the system, accommodating single- or dual-corded racks.

25.3.4.3 N+*1 Block Redundant Topology*

In this topology, also known as a *catcher system*, power flows from the utility through the UPS/PDU and connects to the server. As shown in Figure 25.8, each set of PDUs has a UPS dedicated to it, with one reserve to provide power in case of an outage. A block redundant topology accommodates single- or dual-corded rack configurations providing redundancy at both the UPS and PDU levels.

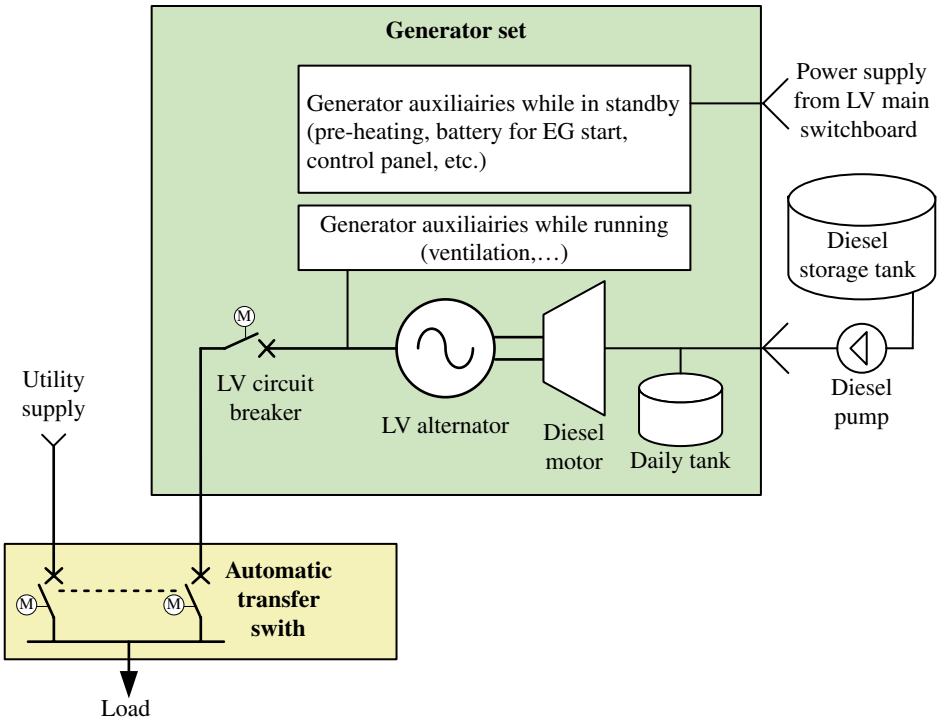


FIGURE 25.5 Diesel generator backup architecture. Source: © Schneider Electric.

In normal condition, each block is loaded at 100% maximum, and the reserve block is off load. If one block fails, the STS will switch the loads on reserve block in less than 10 ms.

25.3.4.4 N+*1 Distributed Redundant Topology*

In this topology, power flows from the utility through the UPS/PDU and connects to the server. The data center load is distributed across the PDUs, leaving enough capacity for the UPS. As shown in Figure 25.9, if there are four systems in the data center, each system should be loaded to 75% maximum in normal conditions; if one system fails, the load is transferred to the remaining live systems.

A distributed redundant topology accommodates single- or dual-corded rack configurations, providing redundancy at the system level.

25.3.4.5 N+*1 Diesel Rotary UPS Unit Topology*

This topology was designed with diesel rotary UPS (DRUPS or Diesel Rotary UPS) technology to achieve a cost-effective design with $N+1$ units. At MV level, $N+1$ DRUPS are suppling several MV/LV blocks (Fig. 25.10). The load is shared between all the DRUPS units using the paralleling bus. If a DRUPS fails, the remaining units will supply the loads without any change. If a short circuit occurs downstream the DRUPS, the fault will be cleared by the zone protections. During the short circuit, the chokes on the common bus are designed to keep the voltage in an acceptable range according to the ITI curve.

25.3.4.6 N+*1 Block Redundancy Using IT Redundancy*

In this topology, there is no redundancy at the block level. The $N+1$ redundancy is achieved at IT level. If a block fails, the loss of the servers supplied by the failed block is backed up by the other servers, and the IT processes do not experience any blackout. This kind of architecture needs a bit more investment on IT infrastructure but gives the advantage to not design any redundancy at LV level (Fig. 25.11).

25.3.5 "Classical" Misunderstanding of What Is a Full Redundant Power System for Data Center

The Tier classification that defines four levels of data center architecture reliability has been widely adopted by the data center business. Here is a definition of the four levels:

- Tier I "Basic capacity": No redundancy required
- Tier II "Redundant capacity components": Redundancy on nonreliable equipment
- Tier III "Concurrently maintainable": Tier II requirements + each equipment can be removed and repaired without a data center blackout
- Tier IV "Fault tolerant": Tier III + fault-tolerant architecture (no single failure point)

However the classification does not consider any utility into the reliability performances of the data center infrastructure. This is not acceptable if the grid performs very poor with several power black-out per week, but this is acceptable if

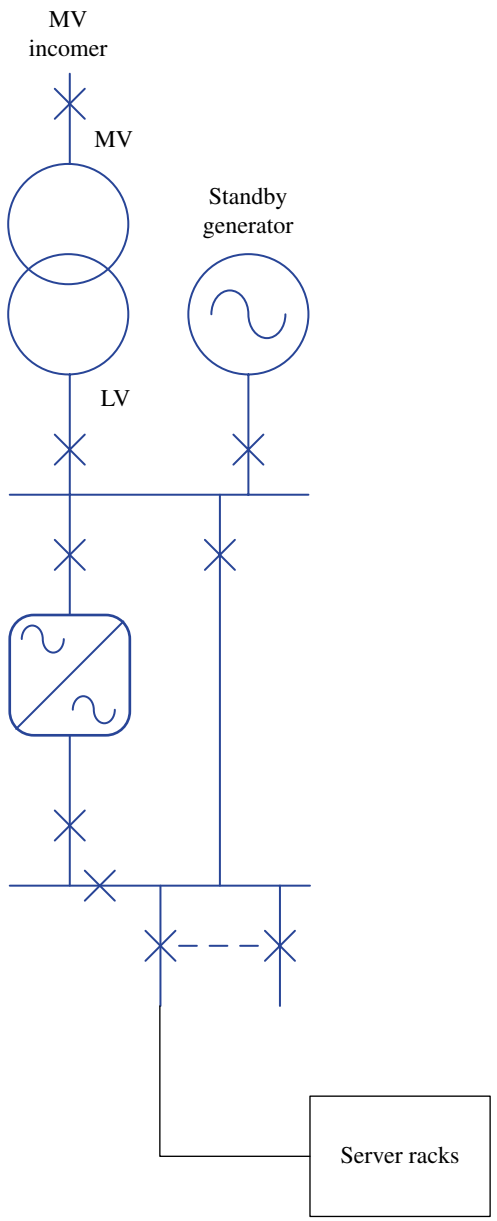


FIGURE 25.6 Topology with *N* redundancy. Source: © Schneider Electric.

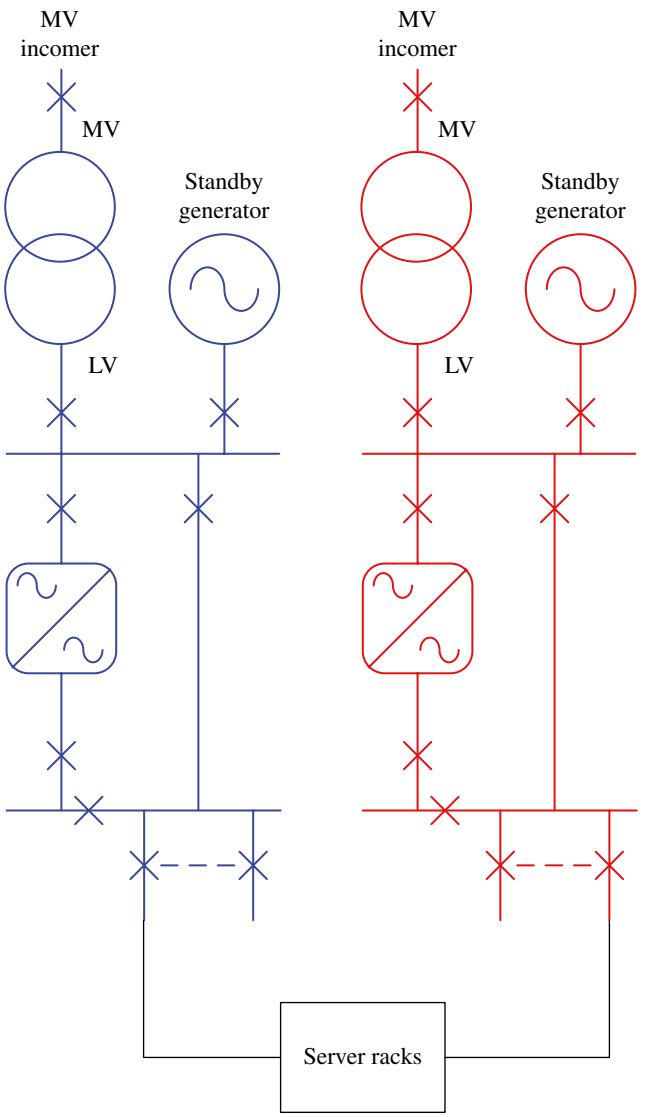


FIGURE 25.7 Topology with 2*N* redundancy. Source: © Schneider Electric.

the grid has an availability above 99%. Due to the tier classification assumption, it is common that people consider that an architecture similar to the one shown in Figure 25.12 is not fault tolerant. This "classical" misunderstanding, especially on end-user side, leads to some useless oversizing of the generator power plant to reach the requirements of Tier III or Tier IV levels.

25.3.6 Shutdown of One Path of Double-Corded Server for Planned Maintenance

The redundancy of IT equipment power supply is designed to provide fault tolerance and also to allow maintenance operation on one path, while the server racks are supplied by the other path.

However, it is quite common that end users ask for no planned shutdown of one path of double-corded server. This is particularly hard to achieve for LV (low voltage) distribution between the main LV switchboard and the servers.

This requirement can lead to two main drawbacks:

- The data center owner will spend more money to respect these requirements by adding more circuit breakers (CB) to make bypasses during planned maintenance activities.
- Or the data center owner will take more risks by deleting some planned tests and maintenance activities like switchboard cleaning and protection tests.

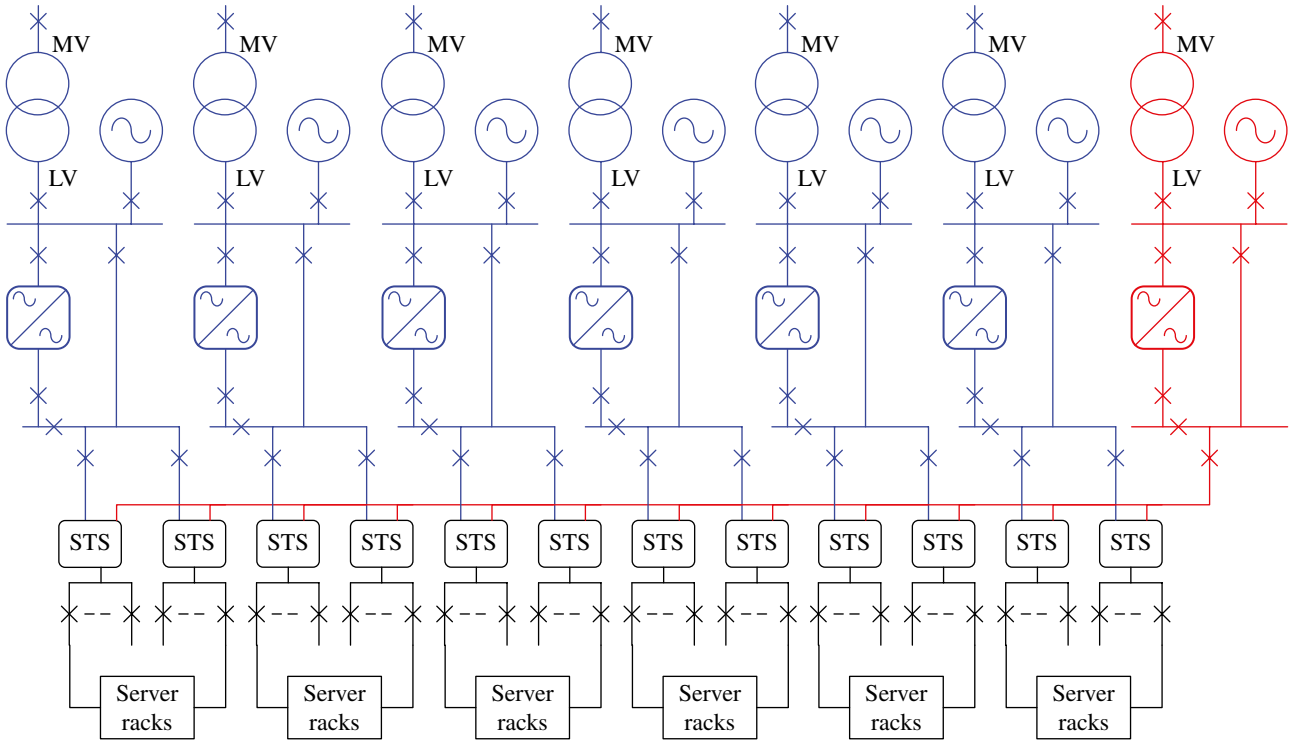


FIGURE 25.8 Topology with 6 + 1 block redundancy. Source: © Schneider Electric.

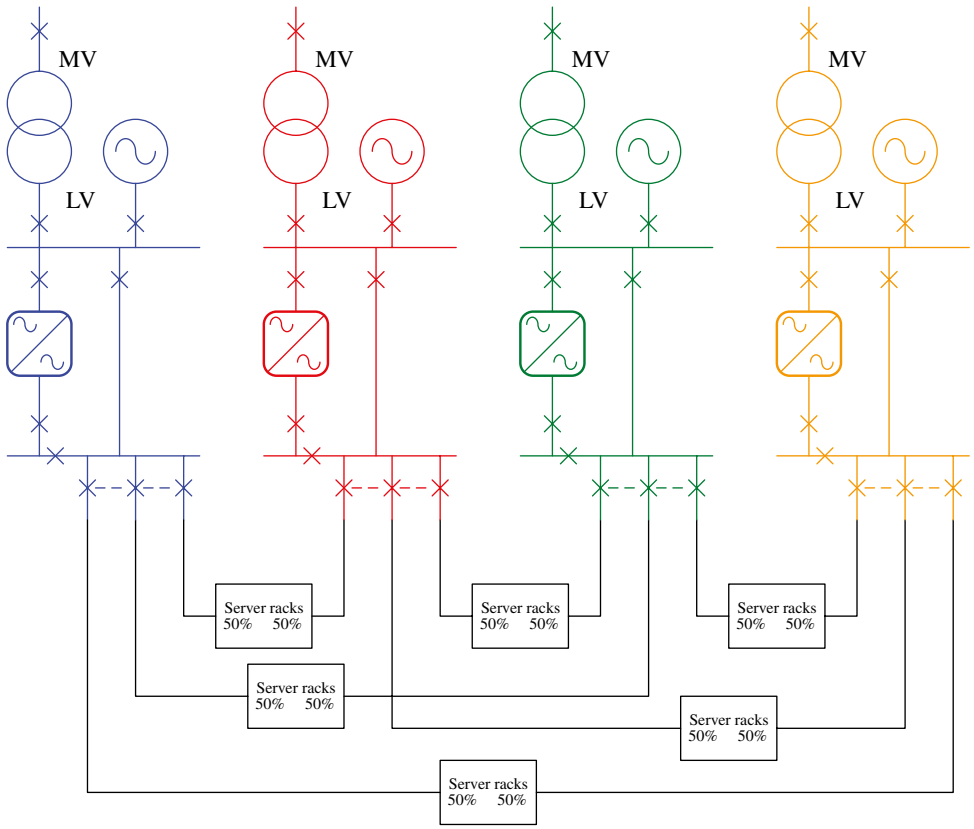


FIGURE 25.9 Topology with 3 + 1 distributed redundancy. Source: © Schneider Electric.

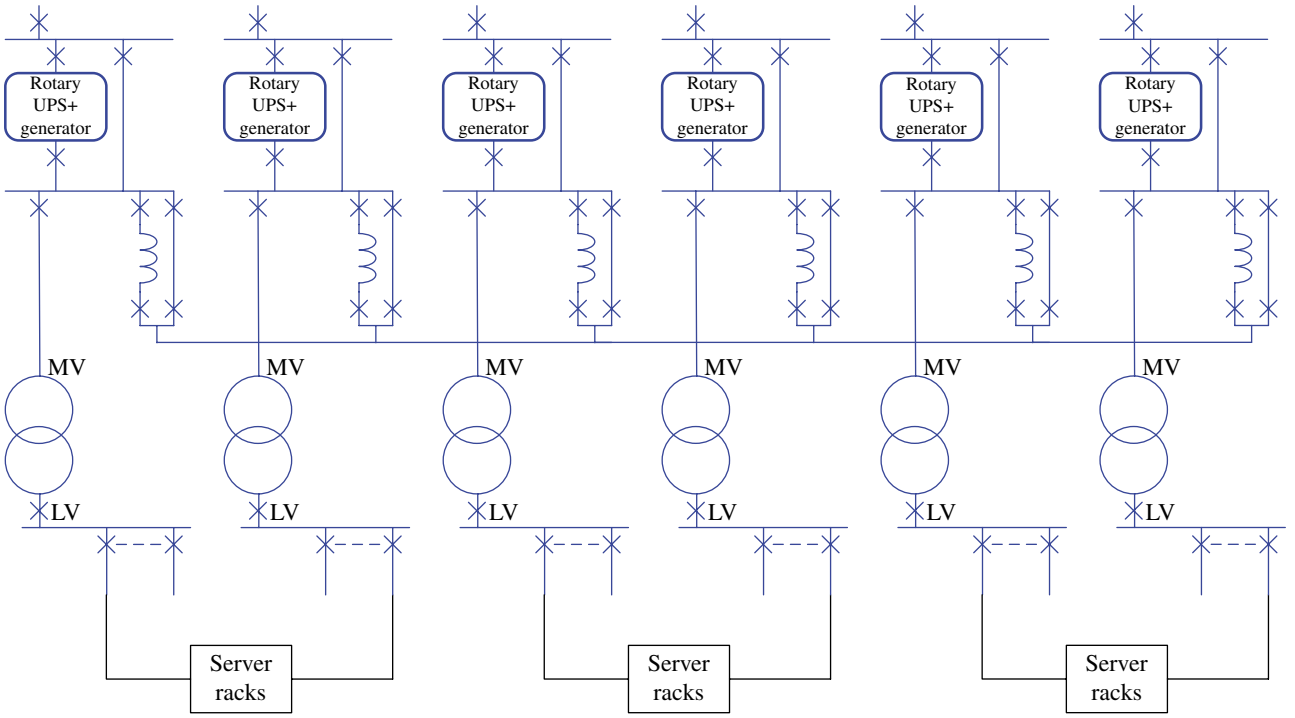


FIGURE 25.10 Topology with iso-paralleling bus with 5 + 1 redundant DRUPS (diesel rotary UPS) at MV level. Source: © Schneider Electric.

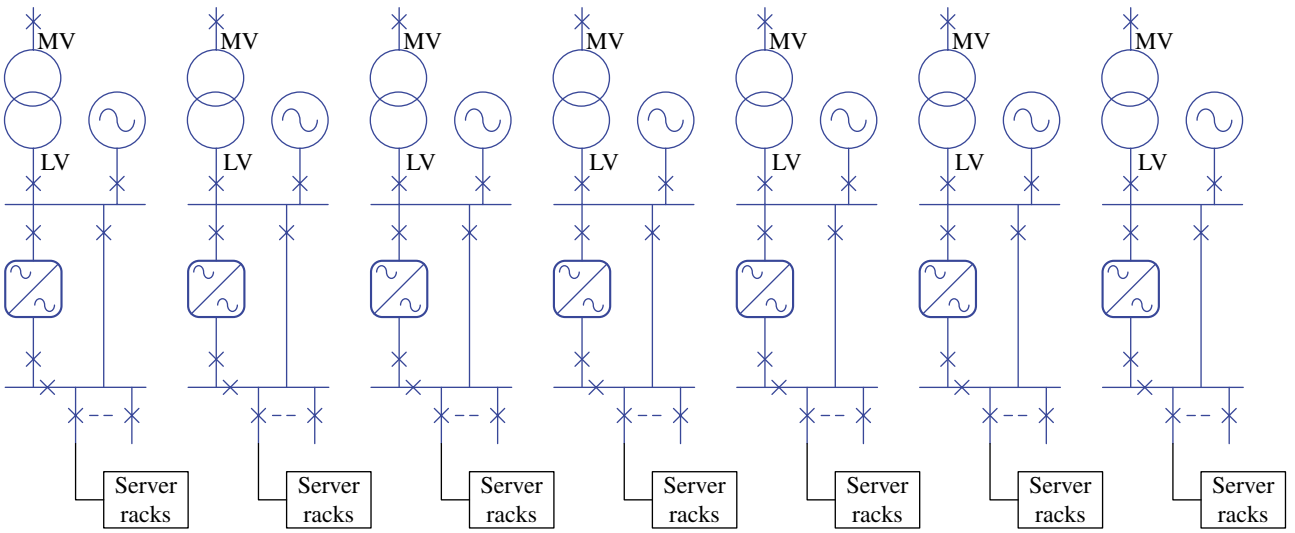


FIGURE 25.11 Topology with 6 + 1 redundant blocks. Source: © Schneider Electric.

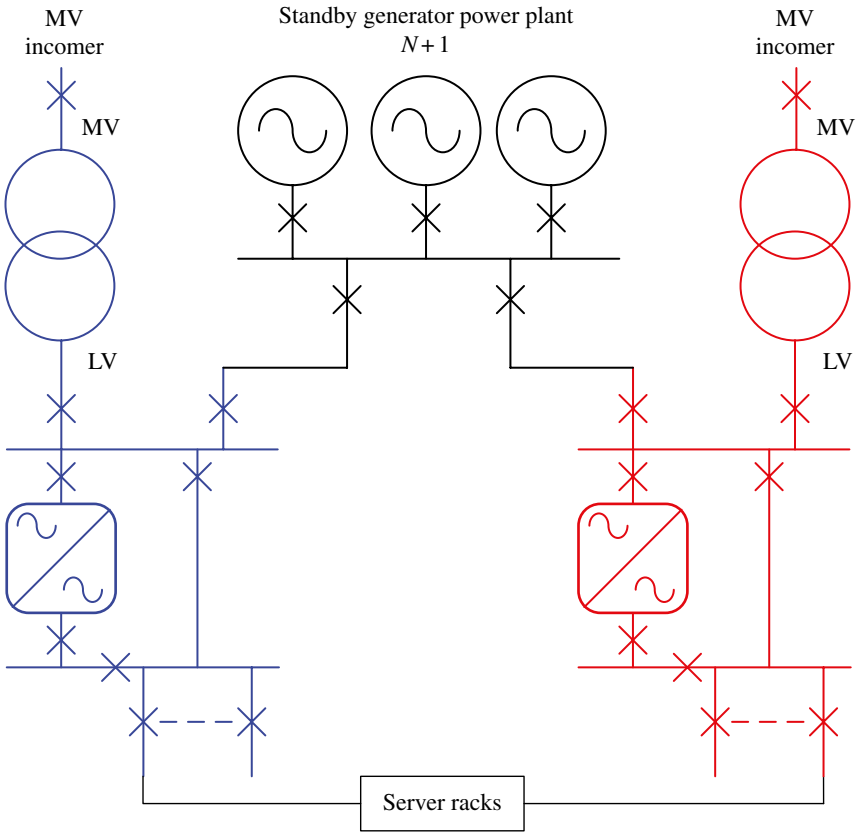


FIGURE 25.12 Example of a fault-tolerant architecture. Source: © Schneider Electric.

25.3.6.1 Topology Comparison

The Table 25.3 outlines the most common data center topologies along with their pros and cons.

25.4 ELECTRICAL DESIGN CHALLENGES

25.4.1 Load Balance Calculation

Before going deeper in the MV and LV design, the first step of the electrical design is to calculate the maximum current and power for each piece of equipment starting from the loads up to the grid connection.

To perform this calculation, the maximum permanent loads need to be known to make the power system load flow and set the correct current ratings. A particular attention needs to be paid to the following:

- The maximum IT load is often overestimated, and diversity should be used in many cases to not oversize the whole data center infrastructure.
- The maximum cooling load consumptions can be also overestimated particularly when cooling design is not ended. Classically, the maximum cooling consumption needs to be calculated by the mechanical engineer according to the environmental characteristics and according to the customer additional safety margin requirements.
- The different operating modes of the electrical installation (site supplied by utility or generator, UPS online mode or on line plus battery charging or bypass mode or in manual bypass mode).

For an accurate load flow, the LV cable losses, the UPS losses, the power factor of the load, the UPS input power factor, the MV/LV transformers losses, and reactive consumption need to be considered.

When specifying the equipment, the capacity of each device in terms of active power (kW), reactive power (KVAR), and maximum current (Amps) needs to be checked.

25.4.2 Grid Connection Substation

25.4.2.1 Connection to the MV Grid

If the maximum apparent power of the data center is below 10–20 MVA, depending on the local grid characteristics, the site will be connected at an MV level. Various architectures of MV grid connection are possible and are also depending

TABLE 25.3 Data center electrical topologies comparison

	N redundant	2*N* redundant	*N*+1 block redundant	*N*+1 distributed redundant	*N*+1 isolated parallel bus	*N*+1 block redundant and IT redundancy
Redundancy level	No Redundancy on the LV distribution	Maximum redundancy, two identical systems	One system capacity worth of redundancy	One system capacity worth of redundancy	One system capacity worth of redundancy	One system capacity worth of redundancy
Pros	• Less electrical equipment required • Lowest cost: initial build and maintenance	• System separation provides true redundancy on every level	• Reserve bus is always available in case of outages and maintenance • Easy load management	• All equipment is utilized • Cost effective solution	• All equipment is utilized • Cost effective solution	• All equipment is utilized • Cost effective solution
Cons	• Outages and failures will bring down server cabinets	• High equipment cost • Increased maintenance cost	• Requires installation of load transfer capability equipment • Low utilization of redundant system leading to decreased efficiency	• Strenuous ongoing load management exercises to ensure adequate distribution	• The data center reliability relies on the discrimination of a complex protection system discrimination of the IP Bus • IP Bus complex to operate	• Suitable only if the data center owner is managing also the IT process

Source: © Schneider Electric.

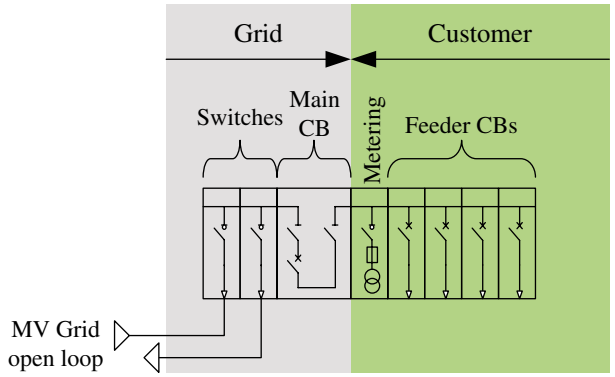


FIGURE 25.13 Example of a single MV grid substation with a single MV incomer (French standard). Source: © Schneider Electric.

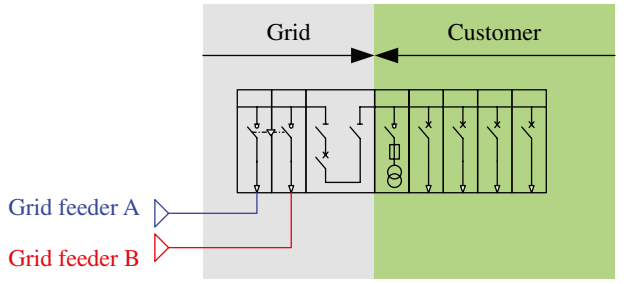


FIGURE 25.14 Example of a single MV grid substation with two redundant MV incomers (French standard). Source: © Schneider Electric.

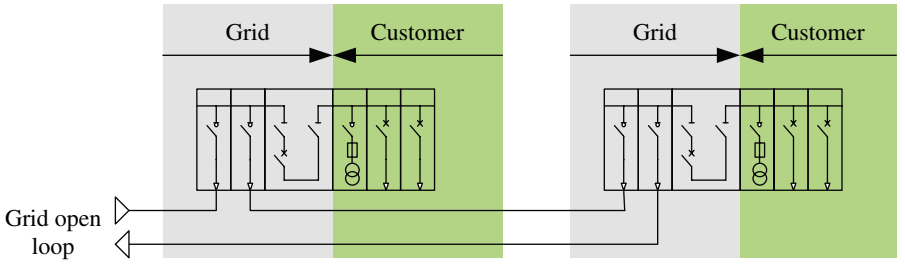


FIGURE 25.15 Example of two redundant MV grid substations with a single MV incomer (French standard). Source: © Schneider Electric.

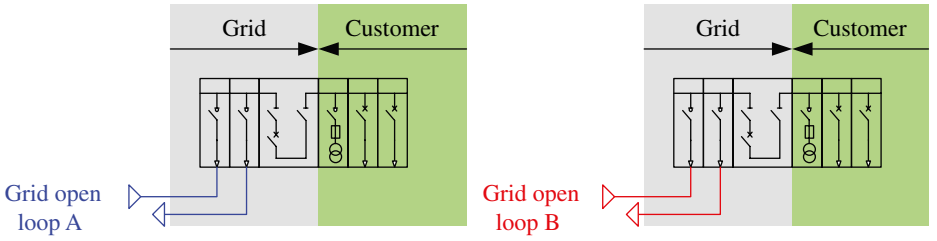


FIGURE 25.16 Example of two redundant MV grid substations with two redundant MV incomers (French standard). Source: © Schneider Electric.

on the local grid standards. Some examples of different alternatives are given in Figures 25.13–25.16.

When connected to an MV grid, the MV protection selectivity is not easy using only time-graded selectivity. As shown in Figure 25.17, the grid MV feeder protection is classically set with 500 ms delay. The setting of the main protection at the grid connection substation is required to be set at 200 ms. This means that the protection selectivity between the different MV protections on the site can only be achieved using overcurrent protection with logic selectivity or other protections like differential protections.

Another challenge is the use of the close transition between the MV grid and the MV generator power plant. Depending on the grid requirements and the size of the generator power plant, the paralleling operation between the power plant and the grid may not be allowed because of the following:

- The grid equipment could not withstand the additional short-circuit current contribution from the generator power plant.
- The generator power plant could perturbate the grid protection system.

25.4.2.2 Connection to the HV Grid

If the site apparent power is above 10 or 20 MVA depending on the local grid characteristics, the site will be connected at HV level with an on-site HV/MV substation.

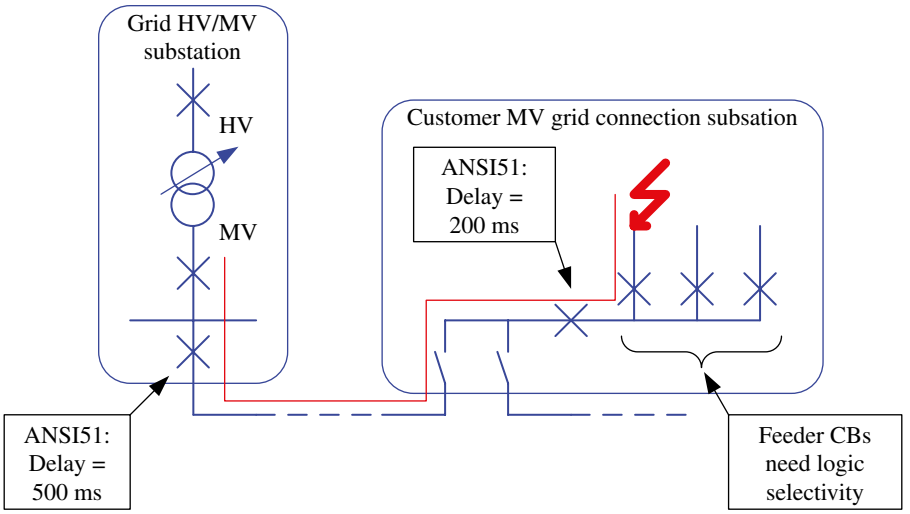


FIGURE 25.17 Protection selectivity for MV grid connection. Source: © Schneider Electric.

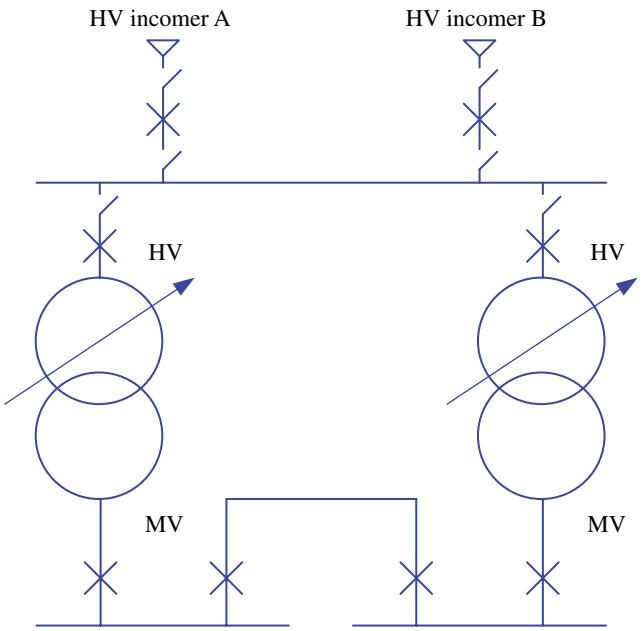


FIGURE 25.18 Example of an HV substation "single bus bar" with redundant incomers and transformers. Source: © Schneider Electric.

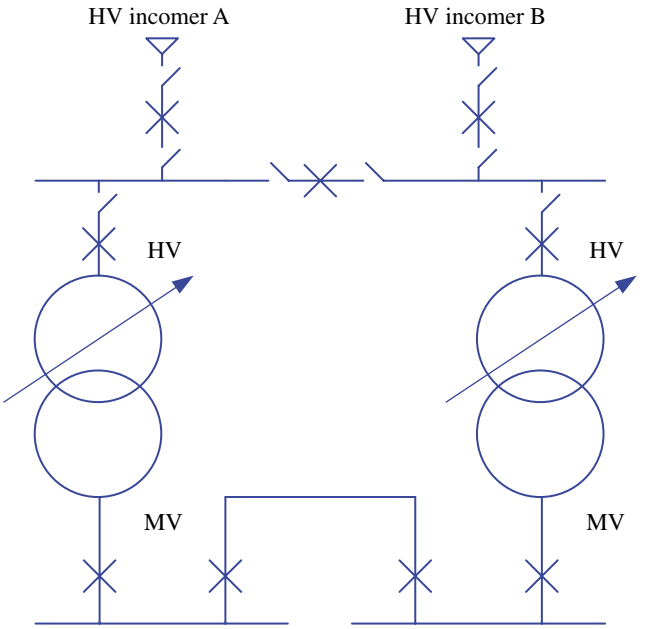


FIGURE 25.19 Example of an HV substation "single bus bar with a tie" with redundant incomers and transformers. Source: © Schneider Electric.

When considering the planned maintenance on the HV/MV transformers, the HV/MV transformers are classically redundant. For the HV substation, several architectures can be designed according to the redundancy and the HV utility standards (Figs. 25.18–25.20).

In terms of operating, various ways to operate the HV/MV substation are possible depending on the utility requirements and the operational cost:

- The incoming lines can be loaded at the same time or used as the main line and the other in standby.
- The incoming lines can be coupled if allowed by the utility to provide a better reliability performance.
- The HV/MV transformers can be used both active or one transformer as the main and the other one as backup and de-energized to avoid the off-load losses.
- The HV/MV transformers are never operating in parallel to minimize the short-circuit current at MV level.

For rated voltage above 50 kV, two kinds of HV substation technology are used depending on the site characteristics:

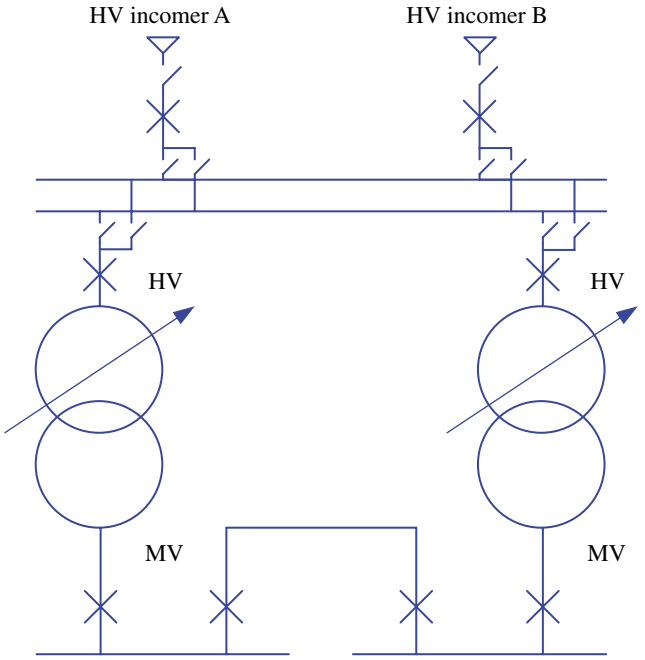


FIGURE 25.20 Example of an HV substation "double bus bar" with redundant incomers and transformers. Source: © Schneider Electric.

- Outdoor air-insulated switchgear (AIS) where the bus bars and the disconnectors are completely air insulated and the CB are using breaking technics in SF_6 gas
- Indoor gas-insulated metal-enclosed switchgear (GIS) where all the components are isolated in SF_6 gas in a metallic enclosure that brings more flexibility to choose the HV substation location, more safety during operation and maintenance, and no sensitivity toward the environment

AIS gives the advantages to use cost-effective equipment cost but has a larger footprint and needs more civil works. GIS is far more compact and generates less constraints to install. Classically, AIS are more used in areas where the land is cheap, whereas GIS is more adapted to urban area.

Having an HV/MV substation on-site gives the opportunity to set the most optimized MV distribution equipment matching the data center needs, which is not the case when the data center is connected to the MV grid as the voltage and the short-circuit level are set by the utility HV/MV substation. When the customer owns the HV/MV substation, the MV voltage level and the short-circuit level can be set thanks to the HV/MV transformer characteristics as shown in Figure 25.21.

The main key parameters of the HV/MV transformer to specify are the primary voltage, the secondary voltage, the rated power, and the short circuit impedance Z_T (%). According to IEC MV voltages, the transformer secondary voltage level can be set to match the best cost-effective MV products. To select the MV equipment rated voltage, a common practice to mitigate the overvoltages is to select the rated voltage of the MV equipment by applying a margin of 10% on the normal operating voltage as shown in Table 25.4.

The transformer short-circuit current at MV level can be decreased by increasing the transformer impedance, decreasing the power ratings (diving the transformer in two smaller ones or using double secondary winding transformer), or increasing the output voltage rating. Depending on the project characteristics, it can be interesting to spend a small over cost on the HV/MV transformers and make savings using cost-effective MV switchboards and smaller cross sections for MV cables.

HV/MV transformers are classically oil type and installed outdoor. The sizing of the transformer apparent power can be done according to its natural cooling capacity (oil natural air natural) or forced cooling capacity (oil natural air forced). Sizing with the forced cooling capacity gives the best cost-effective solution, whereas it can decrease the reliability of the transformer if the real load of the data center reaches a value that requires the transformer forced cooling capacity. As the HV/MV transformers are generally redundant and as there are the generator backup downstream, sizing the transformer using forced cooling capacity can be acceptable. HV/MV transformers are also generally equipped with on-load tap changer (OLTC) to compensate actively the voltage variation on the HV grid and also the voltage drop across the HV/MV transformers due to load variations.

Another key point is that the data center owner needs to start the HV interconnection process in the early phase of the project because the time line required for an HV grid connection substation studies, design, and construction can last for several years:

- The interconnection agreement process requires 1–2 years of studies to make the assessment of the impacts on the transmission system, evaluate the cost of the upgrades on the transmission system, and get the approval from the transmission provider.
- The construction time line can take 2–3 years including the engineering, the equipment procurement, the civil works, the installation, and the commissioning tests.

25.4.3 Backup Generators

The key points in designing the emergency power plant are to ensure the correct level of reliability and to design the best cost-optimized solution because generators represent a significant part of the power system cost of a data center.

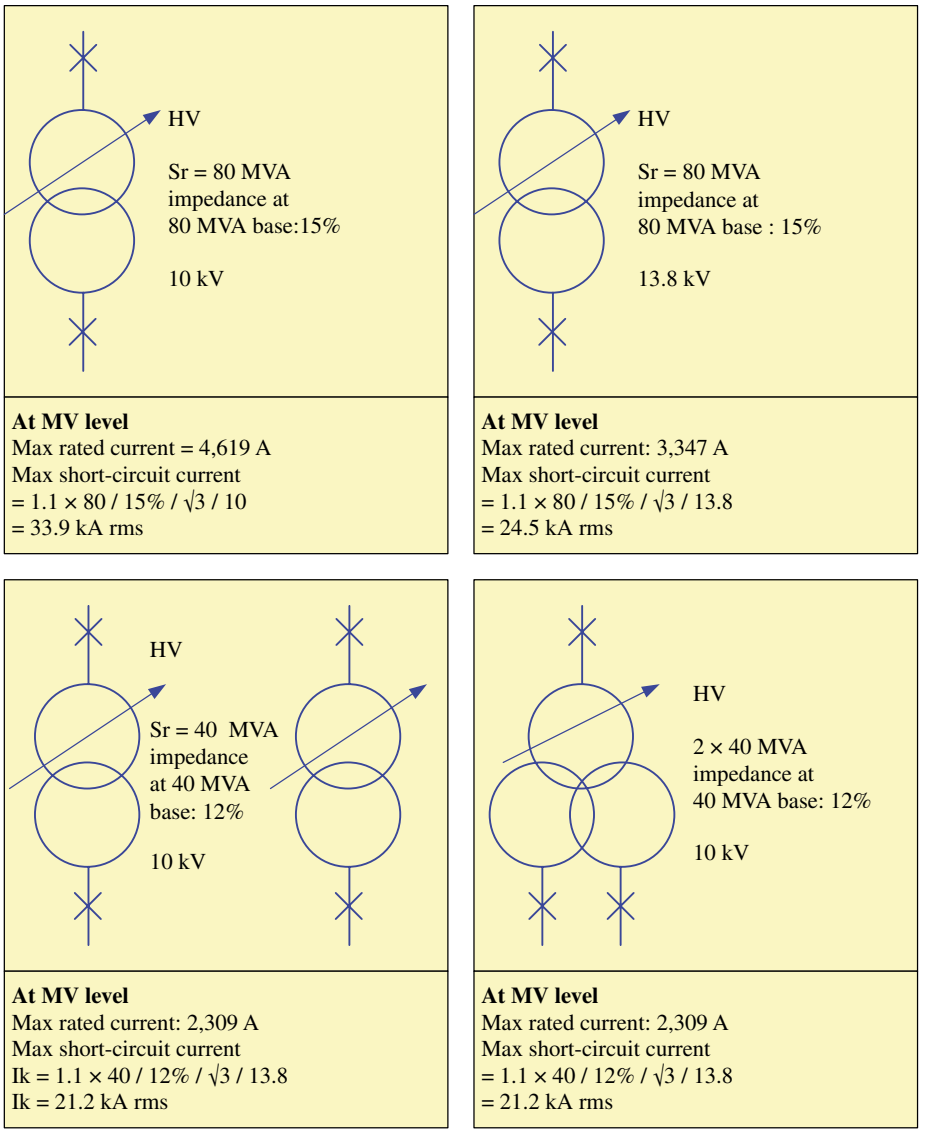


FIGURE 25.21 Different options for sizing the HV/MV transformers. Source: © Schneider Electric.

TABLE 25.4 Standard voltages according to IEC 62271-1

Rated voltage	Rated power-frequency voltage (50 Hz—1 min)	Rated lightning impulse withstand voltage (1.2/5 µs 50 Hz)	Normal operating voltage
kV rms	kV rms	kV peak	**kV rms**
7.2	20	60	**3.3–6.6**
12	28	75	**10–11**
17.5	38	95	**13.8–15**
24	50	125	**20–22**
36	70	170	**25.8–33**

Source: © Schneider Electric.

25.4.3.1 Generator Connection

The emergency backup generators can be connected at LV on the main LV switchboard as an alternative source as shown in Figure 25.22.

The other alternative is to design a power plant connected at MV level to backup several MV/LV transformers as shown in Figure 25.23.

A table giving a comparison between both solutions is shown in Table 25.5.

25.4.3.2 Generator Reliability

For critical applications, regarding the reliability of both utility and generator, the backup generators should be redundant. As shown in Figure 25.24, one MV utility and

one generator give a reliability performance of one data center blackout each 30 years, while having redundant generators gives a blackout each 1,000 years.

To lower the cost, the challenge is to use $N+1$ (or $N+2$ if required) redundancy instead of $2N$ redundancy.

Considering a generator power plant using multiple units in parallel with $N+1$ redundancy, several levels of redundancy can be applied on the other equipment of the power plant depending on the reliability target of the end user:

- Redundancy on the auxiliary power supply and the diesel pump distribution
- Redundancy on the power plant main PLC that manage the starting and shutdown sequences and the number of units in operation
- Redundancy on the power plant electrical distribution (required if the customer wants to comply with Tier III or Tier IV classification or if a very high reliability is targeted)

To maintain a good level of availability, the generators need to be monitored and tested on a regular basis:

- The monitoring includes classically the battery state of charge.

TABLE 25.5 Comparison between generator connected at LV level or at MV level

Generator connected at LV level	Generator connected at MV level
The generator redundancy level is set according to the redundancy level of the MV/LV power trains	The generator redundancy level is set according to the redundancy level of the MV/LV power trains
The solution is less scalable	The number of generators can be set according to the real load providing a good scalability
The size of the generator is designed according to the load of the main LV switchboard	The size of the generator is designed according to the best cost-effective size (depends on the project characteristics)
Not always suitable for close transition option due to issues with high short-circuit current	Suitable for close transition option if the site is supplied by the HV grid (not always suitable if the site is connected to the MV grid)
Generator size: from few kW to 3.2 MW	Generator size: from 2 to 10 MW
Voltage from 380 to 690 V	Voltage from 4.16 to 15 kV with MV alternator No voltage limit if using a step-up transformer

Source: © Schneider Electric.

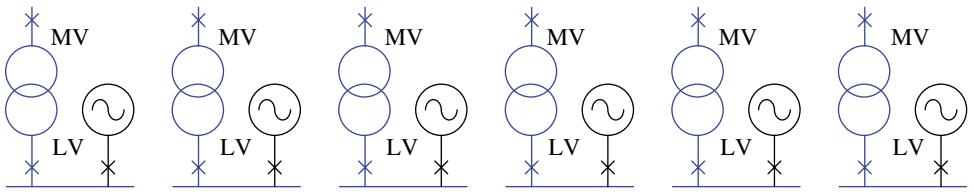


FIGURE 25.22 Generator connected at LV level. Source: © Schneider Electric.

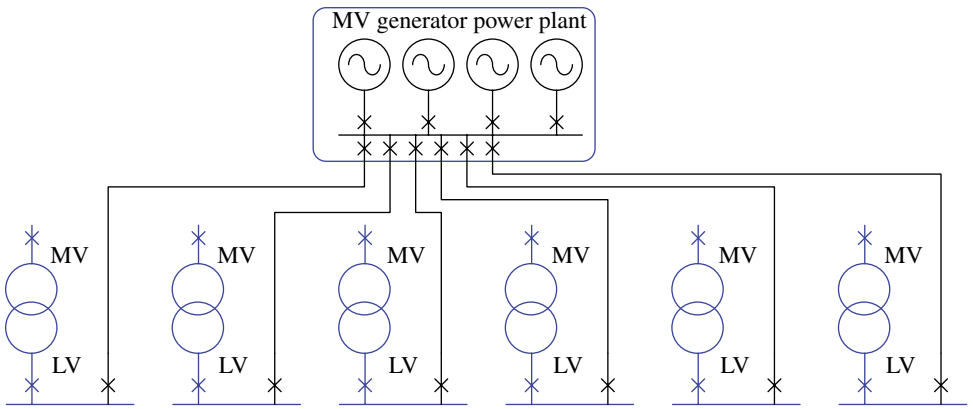
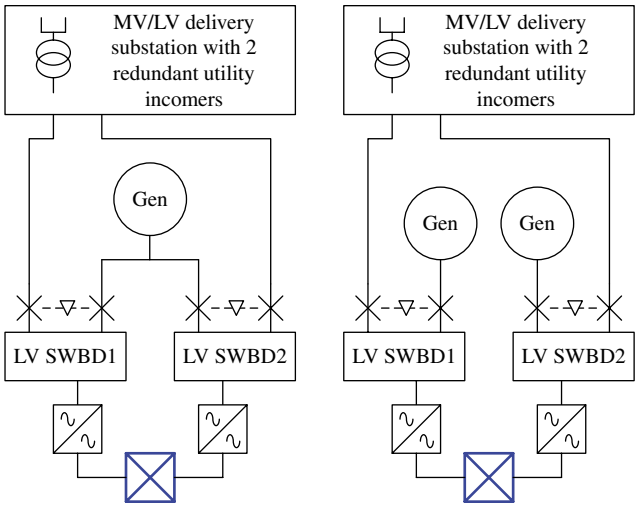


FIGURE 25.23 Generator connected at MV level. Source: © Schneider Electric.



Reliability comparison	Failures/year
Case with no redundancy on the generator	**0.03**
Case with redundancy on the generator	**0.001**

FIGURE 25.24 Reliability improvement considering generator redundancy. Source: © Schneider Electric.

- Off-load test that will detect any failure on the starting system and a part of the generator block itself.
- On-load test that will detect any failures of the generator.

The on-load test can be done in several ways:

- Test on a load bank at full load
- Test on the real load using close transition transfer
- Real backup test by doing a real shutdown of the utility incomer

25.4.3.3 Close Transition Versus Open Transition

When using an automatic transfer switch to switch from a source to another source, there are several ways to make the transfer when both sources are available as shown in Figure 25.25. Close or soft transitions have the advantage to avoid any interruption for the loads when a changeover is required for planned maintenance or test and provide less stress for the UPS batteries and nonsecure loads such as the chillers. However the close transition increases the short-circuit current level due to the generator contribution in case of short circuit during the paralleling operation of both utility and generator. This short-circuit current level can have a significant impact on the cost of the electrical distribution equipment.

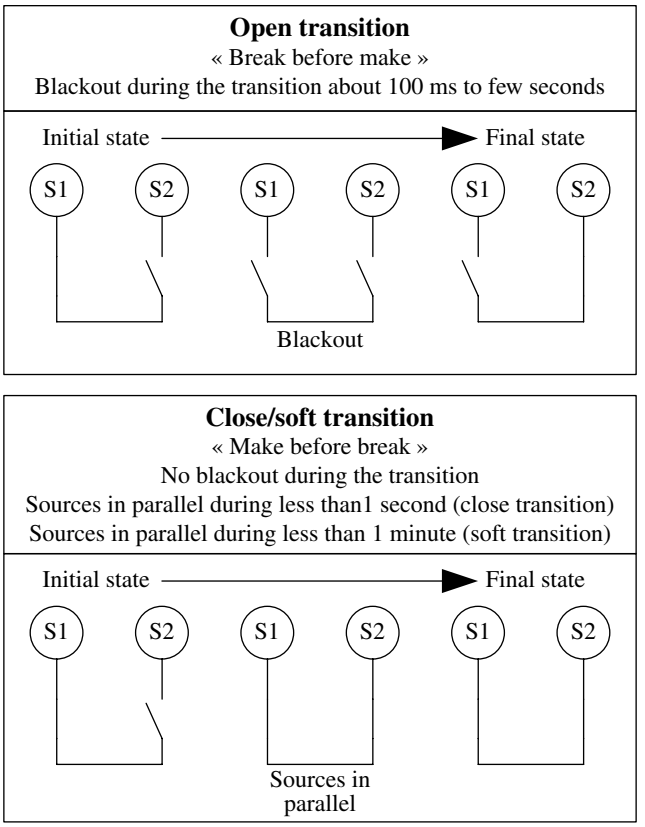


FIGURE 25.25 Different ways to make a changeover. Source: © Schneider Electric.

25.4.3.4 LV Generator Starting Sequence

For a single generator, the starting time to reach the rated speed is about 5 seconds. When the generator is ready to take the load, the ATS transfer the switchboard load on the generator with a blackout in the case of an open transition. The main loads that will be reenergized are the UPS and the nonsecure cooling loads such as the chillers. The load impact depends on the load characteristics:

- The UPS will transfer the IT load from battery to the input main 1 after a delay; classically the UPS will transfer the load with a ramp up/down, which will minimize the load impact on the generator.
- The chillers will start after few minutes; classically the chiller includes some drives that will provide a smooth starting sequence.

The load impact has to be defined for each project and checked according to the generator capability.

25.4.3.5 MV Generator Power Plant Starting Sequence

In the case of an MV power plant with several generators, all the generators start at the same time and are coupled one by

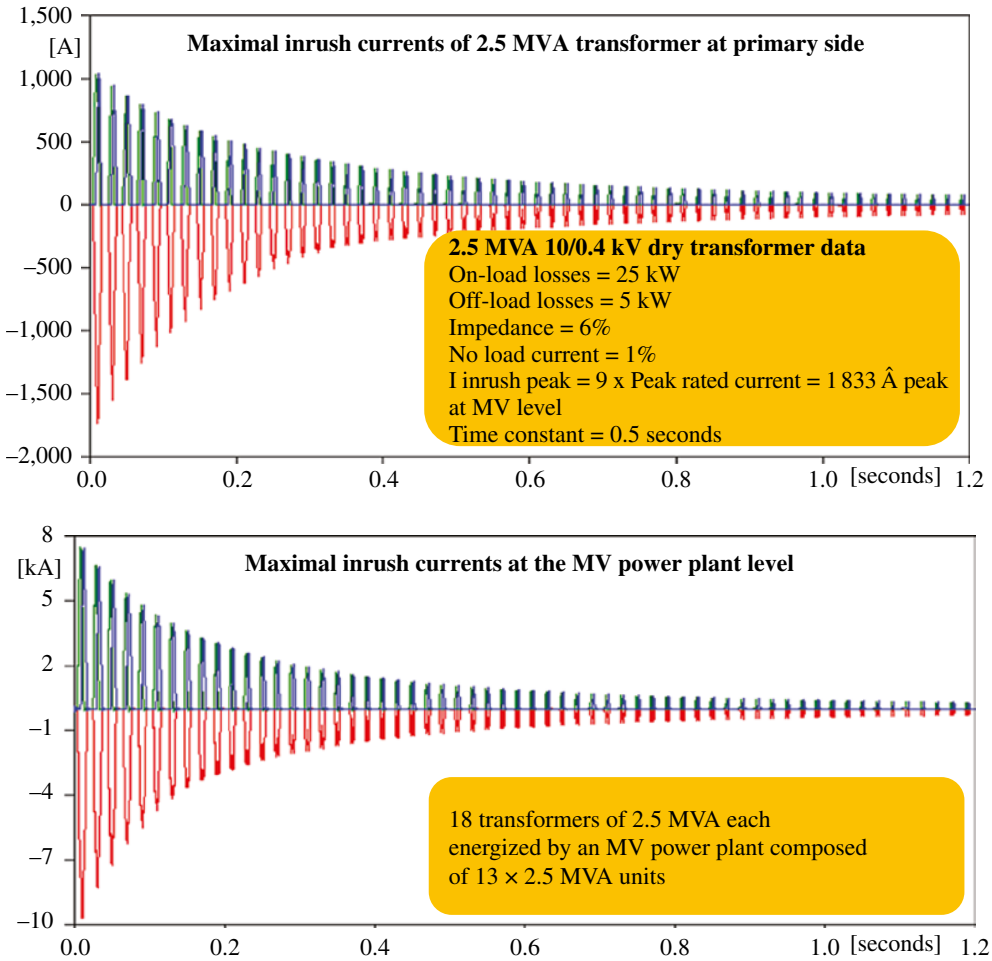


FIGURE 25.26 Simulation of MV/LV transformer inrush currents. Source: © Schneider Electric.

one to the generator bus taking less than 10 seconds per generator for synchronizing.

When the power plant is ready to take the load, the ATS transfer the switchboard load on the power plant with a blackout in the case of an open transition. As the main loads are supplied through the MV/LV transformers, the power plant will need to supply the inrush current due to magnetization of the transformers. For an MV/LV transformer, the maximal peak inrush is about 9 times the peak rated current when the transformer is energized by a perfect source. However, this value is much lower when the transformer is energized by a source with a lower short-circuit power as shown in Figure 25.26. Moreover, when considering several transformers energized at the same time, the total current will be above the sum of all individual maximum inrush currents.

These inrush currents will have two main impacts:

- Overcurrents that could trip the overcurrent protections
- Voltage drop due to reactive power impact that could trip the undervoltage and overvoltage protections
- Overvoltage due to the alternator voltage regulator (AVR) overshoot

To ensure correct operations:

- The overcurrent protections and undervoltage protections must be delayed to avoid any spurious trip;
- The reactive power impact has to be checked according to the generator AVR capability to avoid any voltage overshoot, but generally the generator does not experience any stability issue when energizing a transformer with an apparent power that is twice the alternator power.

25.4.3.6 Generator Power Rating

ISO-8528-1 standard defines the generator ratings according to four operational categories. In each category, the generator rating is defined by the maximum power output considering its running time and its load profile as shown in Figure 25.27.

ESP: Emergency standby power
For supplying emergency power for the duration of a utility power failure
Limited number of hours per year: 200 hours
Average load factor of 70% of the standby rating over a period of 24 hours

LTP: Limited-time power
Limited number of hours per year: 500 hours
Non-variable load
Applications not to exceed 100% of the prime power rating

PRP: Prime rated power
Unlimited number of hours per year
Variable load, not to exceed 70% average of the prime rating during any operating period of 24 hours

COP: Continuous operating power
Unlimited number of hours per year
Applicable for supplying continuously at a constant 100% load for an unlimited number of hours per year

FIGURE 25.27 ISO 8528-1 generator set ratings.

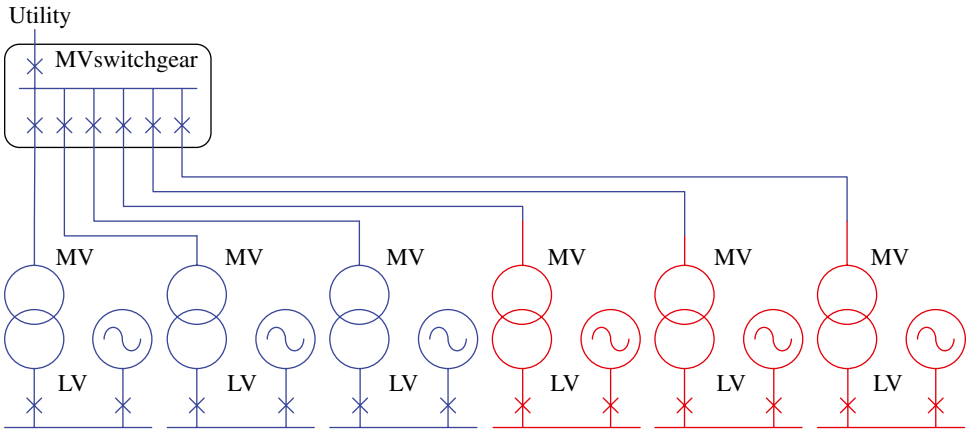


FIGURE 25.28 2*N* architecture with LV generators and nonredundant MV distribution. Source: © Schneider Electric

The same generator has different ratings according its operational categories. For data center applications, the mission profile of a backup generator is the following:

- Emergency operation in case of utility failure at 100% load (about less than 1% of the time)
- Nonemergency running time (for tests and maintenance purpose) less than 200 hours

According to the data center generator operational characteristics, the best solution is in between PRP and COP ratings, that is why, when specifying the generator power rating, the data center designers should not refer to the ISO-8528-1 but should mention the real mission profile and let the generator supplier providing its best solution.

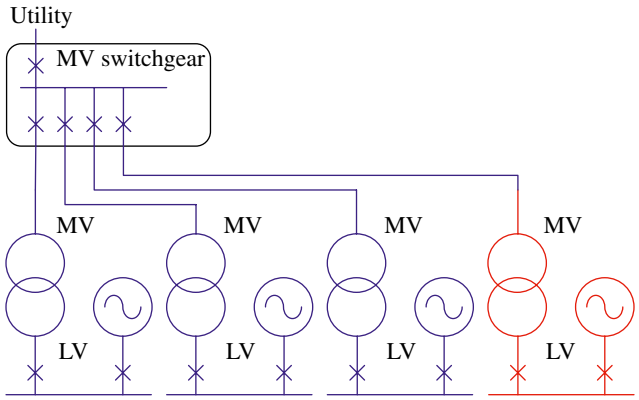


FIGURE 25.29 $N+1$ architecture with LV generators and nonredundant MV distribution. Source: © Schneider Electric.

25.4.4 MV Power System Design

The main topics for the MV power system are first the MV topology that should be in adequacy with the resilience principle depending if the generators are located at LV level or MV level.

25.4.4.1 MV Distribution Topologies

When the UPS and generators are located at LV level, the redundancy is full ensured at LV level. In this case, two alternatives are possible. Suppling all the MV/LV transformers

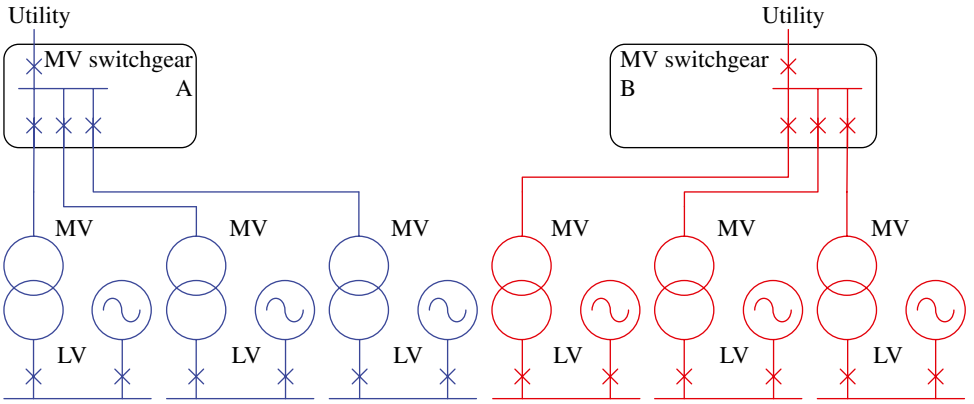


FIGURE 25.30 2*N* architecture with LV generators and redundant MV distribution. Source: © Schneider Electric.

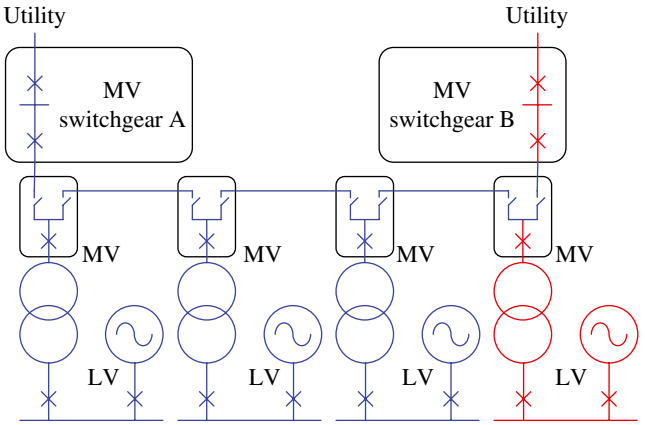


FIGURE 25.31 $N+1$ architecture with LV generators and redundant MV distribution. Source: © Schneider Electric.

without any redundancy as shown in Figures 25.28 and 25.29 will expose the data center site to run for a long period (1 month or more) in case of a major failure on the main MV substation.

However, running the data center on its generators for a long period can be a nuisance to suppliers and others. For example, if running a large data center continuously, it will have to refill the diesel tank frequently in a constraint area as well as air pollution emission that is not allowed by local regulatory. To avoid such situation, a solution is to provide redundancy even at MV level with two redundant MV switchboards as shown in Figure 25.30. A more cost-effective alternative is to use an open loop distribution as the utility does as shown in Figure 25.31.

In the case of an MV power plant, a redundant MV distribution is needed to ensure the overall redundancy of the data center. For 2*N* LV redundancy principle, the simplest MV distribution consists of two redundant MV switchboards supplied by a utility incomer or a generator incomer using an automatic transfer logic and a single MV generator board as shown in Figure 25.32.

If MV/LV power trains are designed with $N+1$ redundancy, then the MV/LV transformers needed an MV automatic transfer switch to be able to be supplied by MV switchgear A or B as shown in Figure 25.33.

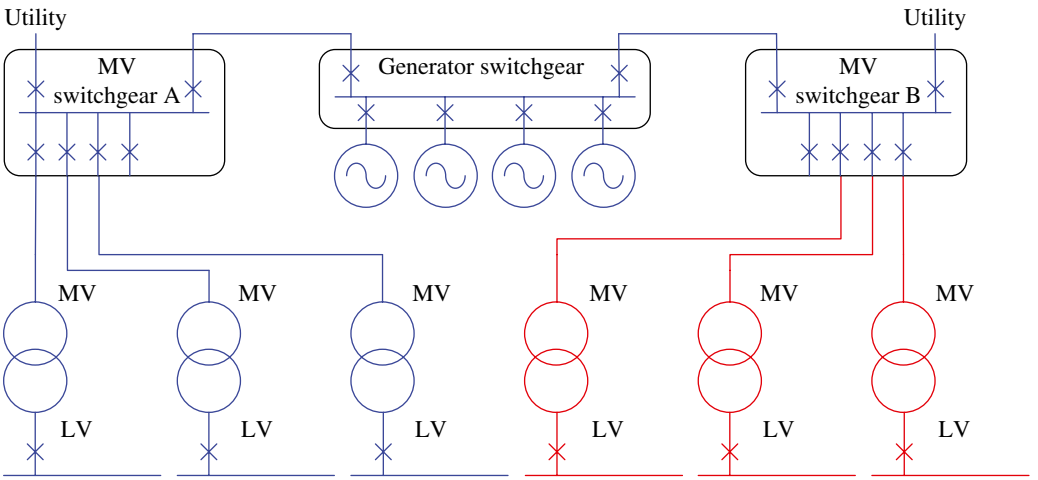


FIGURE 25.32 2*N* architecture with MV generators. Source: © Schneider Electric.

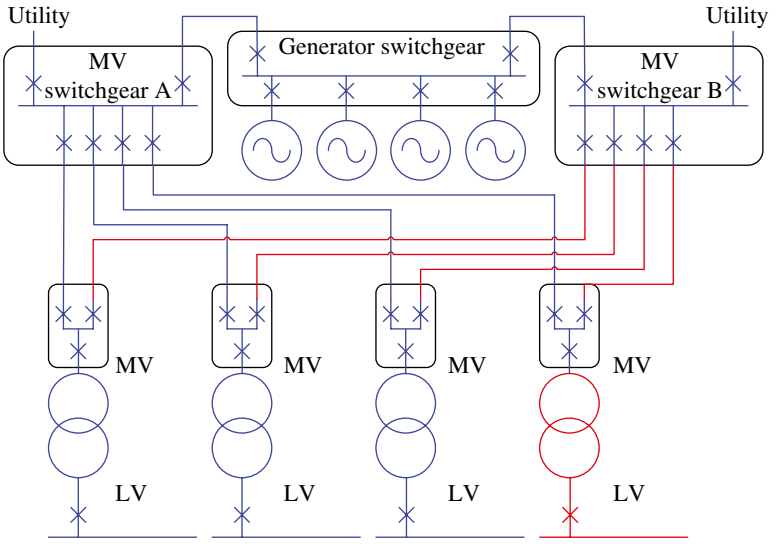


FIGURE 25.33 $N+1$ architecture with MV generators. Source: © Schneider Electric.

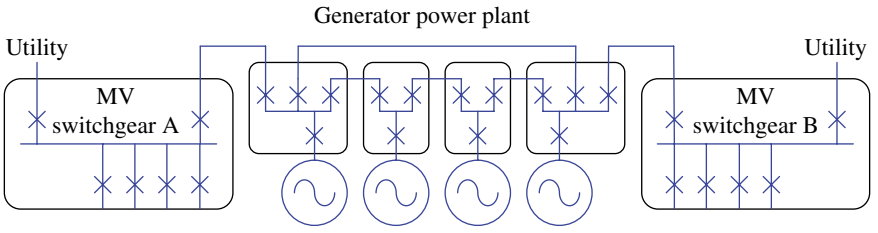


FIGURE 25.34 Full redundant MV generator power plant using closed ring architecture. Source: © Schneider Electric.

If the data center needs to meet Tier IV requirements according to the Tier classification, the MV generator power plant needs to be fully redundant by itself without considering the utility. In this case, the generator MV distribution cannot be achieved with a single bus. As generators are classically designed with $N+1$ redundancy, two alternatives are possible:

- Closed ring with a single switchboard per generator as shown in Figure 25.34
- Double feeder from each generator to generator A and generator B switchboards as shown in Figure 25.35

25.4.4.2 MV Switchboards

Several kinds of MV switchboard technologies are available and can be classified in several ways as following:

- Types of installation:
  - Indoor classically or outdoor
  - Pole mounted, ground mounted, or pad mounted
- Types of insulation technology:
  - Air-insulated switchgear (AIS)
  - Oil-insulated switchgear (OIS)
  - Gas-insulated switchgear (GIS)
  - Solid-insulated switchgear (SIS)
  - Shielded solid-insulated switchgear (2SIS)
- Types of application:
  - Primary (rated current up to 4,000 A and maximum short-circuit current withstand up to 50 kA)
  - Secondary (rated current up to 1,250 A and maximum short-circuit current withstand up to 25 kA)
- Types of loss of service continuity (LSC) when opening accessible compartment (defined by the standard IEC 62271-200):
  - LSC1 (other functional units must be disconnected)
  - LSC2A (other functional units can remain energized)
  - LSC2B (other functional units and all cable compartment can remain energized)

To select the adequate MV CB short-circuit breaking current, a particular attention shall be paid when considering the short circuit of both utility and MV generator power

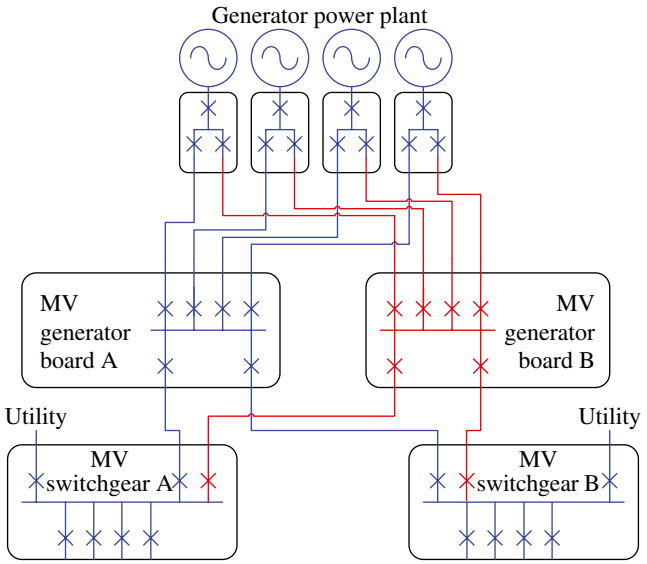


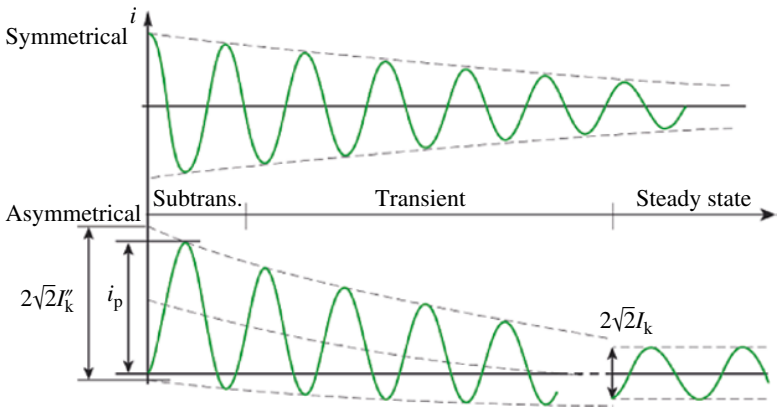
FIGURE 25.35 Full redundant MV generator power plant using double feeder architecture. Source: © Schneider Electric.

plant. A high aperiodic component of the short-circuit current as shown in Figure 25.36 may lead to a derating of the CB short-circuit breaking capacity according to IEC 62271-100.

The best switchgear range will be selected according to each switchboard characteristic such as the operating voltage, the rated current, the maximum short-circuit current, the expected number of operations, the required number of current and voltage sensors for each cubicle, and the service continuity during maintenance. Other considerations such as the installation constraints (cable entry options and requirements for gas exhaust in case of internal arc) need also to be investigated when selecting the appropriate switchboard range. To meet the best cost-effective solution, a general way is to use as possible secondary MV switchboard ranges of the country where the data center is located.

25.4.4.3 MV/LV Transformers

When installed in the data center building, the MV/LV transformers are preferred to be dry-type insulation instead of liquid immersed because of the installation constraints. However, the option to install the MV/LV transformers outdoor has several advantages such as reducing the footprint of the electrical rooms in the data center building and avoiding any transformer room cooling system. The apparent power range goes generally from 1 to 4 MVA with natural cooling rating. In terms of efficiency, the best option to minimize the total cost of ownership (TCO) is to select a transformer with low losses. Power transformers that meet the European eco-design directive are efficient products to use.



I''_k rms value of the initial symmetrical short-circuit current
i_p Short-circuit current peak value
$i_{d.c.}$ Aperiodic component of short-circuit current
I_b rms value of the symmetrical short-circuit breaking current
I_k rms value of the steady-state short-circuit current

FIGURE 25.36 Definition of short-circuit current values. Source: © Schneider Electric.

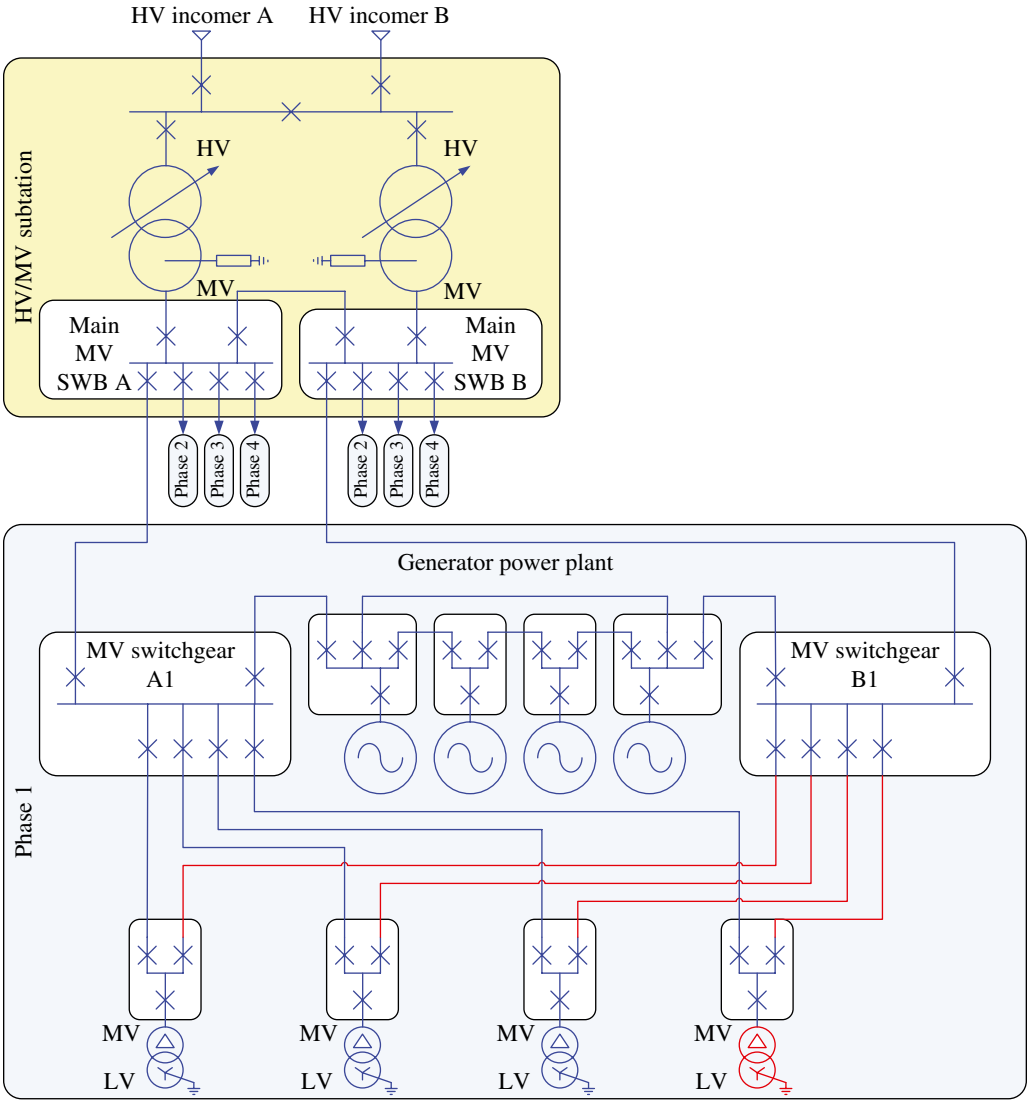


FIGURE 25.37 Example of a large data center MV power system. Source: © Schneider Electric.

25.4.4.4 MV Protection System

Considering a typical large data center site connected at HV grid with several phases as shown in Figure 25.37, the MV protection system consists of the protection of the cables from the HV/MV substation down to the MV/LV transformers, the protection of the generators and the generator switchboards, and the protection of the MV/LV transformers and its MV switchboard.

The neutral is classically earthed with a resistor. The ground fault level is set as low as possible to minimize the equipment damage in case of an earth fault. For the MV backup generators, two alternatives are possible depending on the habits:

- The generator power plant is equipped with an earthing transformer and a resistor.
- Or each generator is grounded with a resistor.

The second alternative has the benefit to be bit more reliable but has the drawbacks to increase the ground fault level.

The simplest protection system to protect the switchboards and the cables against short circuits is to use basic overcurrent protection (ANSI 50/51) with time-graded discrimination between protections. The drawback is that a 300 ms delay is needed for each stage so that the main protection could be set with a delay of more than 1 second.

To reduce the fault clearing time and achieve a better protection level, several options are possible:

- The first option is to use logic selectivity between the incomer CB and the feeder CB of the same MV switchboard avoiding any wiring outside the switchboards.
- If it is not enough, the logic discrimination can be used between different switchboards, and some additional pilot wires are needed between switchboards.
- If it matches the local habits, another alternative is to use line and bus differential protections, but this induce an over cost of the protection system.

The alternator of the generator is classically protected by the generator controller. The MV CB relay of a generator includes generally only the overcurrent protection and a directional overcurrent protection to be able to clear an alternator fault by disconnecting only the failed generator. Other functions like generator differential protection, undervoltage/overvoltage protections, and synchro check protection can also be selected.

The MV/LV transformers are protected classically with phase overcurrent protections (ANSI 50/51), earth fault overcurrent protections (ANSI 50/51G) at MV level, and other protections depending on the transformer type (dry or liquid immersed). Additional protections can be selected to trip faster and limit the damage on the transformer such as transformer differential protection (ANSI 87T).

25.4.4.5 MV Automatic Transfer Switch (ATS)

When a MV automatic transfer switch is needed to transfer from a source to another, a particular attention shall be paid to the specification, the design, and the commissioning because the ATS are one of the main equipment that ensure the reliability of the data center infrastructure. The ATS must transfer when the voltage is not suitable for the loads. The key points when designing an ATS are:

- The power configuration according to the available MV cubicle functions and the requirements in terms of protection, durability, and maintenance
- The control design according to the type of voltage sensors available (voltage presence indicators or voltage transformer), the way to achieve the ATS control logic (a dedicated ATS controller, a PLC, a single protection relay or two protection relays)

Moreover a detailed failure analysis and a test plan can minimize the risk of common mode failure by checking the MV ATS behaviors in any case including different scenarios of failures or disturbances upstream the ATS (voltage drop on one or several phases, voltage loss on one or several phases, etc.) and different failures in the MV ATS itself (unintended opening of the CB, loss of control power, wrong settings, etc.).

25.4.4.6 MV Power Monitoring

The monitoring at MV level includes the monitoring of the MV switchboards (states of the switches and CB, state of the protection relays and PLCs, and voltage presence indicator). Some metering equipment are also classically located at the incomer of the primary and secondary distribution switchboards to provide both energy metering and power quality monitoring (voltage and current harmonics, voltage sags).

25.4.5 LV Power System Design

The load specificities impact the LV power system architecture.

25.4.5.1 IT Racks

Nowadays, IT equipment such as web or application servers, storage servers, switches, and routers are mounted in a server rack. A server rack using rack-mounted technology as shown in Figure 25.38 consists of:

- The servers that are single phase, whereas the rack can be single phase or three phase depending on the rack rated power;
- Dual rack-mounted power strips provide the different sockets for each server;
- Rack-mounted server with embedded dual PSU; the server PSU supply both cooling fans and different IT subsystems.

A server rack using open compute project (OCP) technology as shown in Figure 25.39 consists of the following:

- The rack can be single phase or three phase depending on the rack rated power.
- The PSU are mutualized for several servers; several PSU in parallel are providing a single DC bus at 12 or 48 VDC; optional redundant PSU can improve the reliability; optional battery bloc can provide the UPS function at the rack level.
- The server cooling fans are also mutualized for several servers with optional redundant fans.
- The severs are supplied by a single connection to the rack DC bus.

25.4.5.2 Cooling Loads

The cooling system of a data center is designed to put outside the heat produced by different rooms (technical rooms and IT rooms). Two major cooling technologies are currently

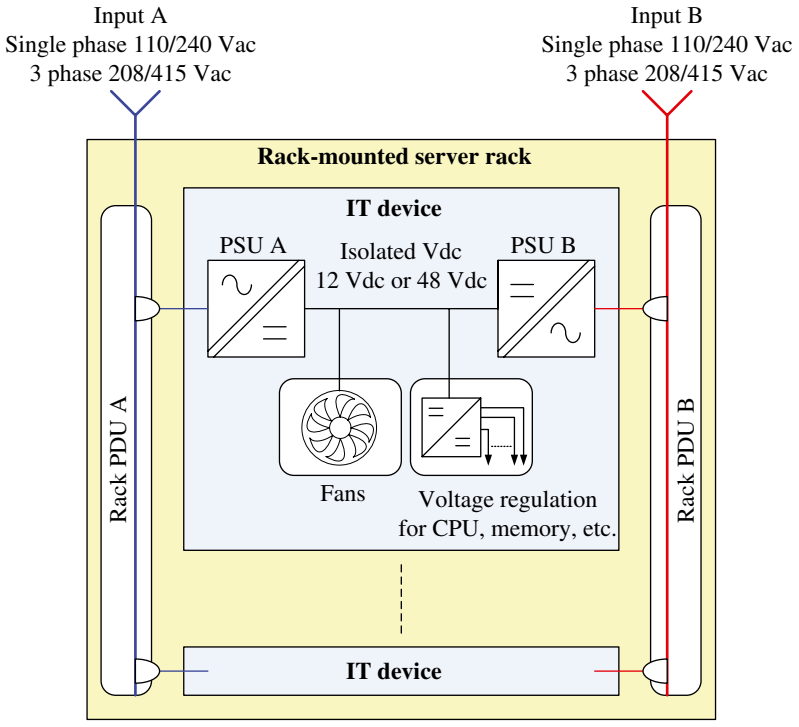


FIGURE 25.38 Server rack—rack-mounted type. Source: © Schneider Electric.

used for data center cooling: chilled water cooling and direct or indirect cooling.

Chilled water cooling is using chillers, water pumps and piping for distribution, and air handling units (AHUs) in the different rooms of the data center. Classically, the AHUs and the water pumps need a secured power supply to be able to maintain the IT room temperature during the backup power plant starting sequence in case of the primary source outage. The chillers are not secured.

Direct or indirect cooling is using fans for supply air and exhaust air, filters, and water misting when needed. In this kind of system, all electrical loads are secured with UPS.

The cooling load maximum electrical consumption can vary depending on the cooling technology, the cooling equipment, and the outside air extreme conditions. Classically the maximum power usage effectiveness (max PUE) can vary from 1.5 to 2.

Cooling loads are also classically equipped with variable frequency drives (VFD) embedded in the equipment or not. As VFDs can lead to electrical perturbances due to leakage currents in the ground that may disturb sensitive communication equipment, a particular attention shall be paid when supplying perturbating loads such VFDs and IT loads with the same MV/LV transformer. As shown in Figure 25.40, the best option in terms of electromagnetic compatibility (EMC) is to have a galvanic separation between sensitive loads and disturbing loads.

However, supplying both the IT loads and the cooling loads by the same MV/LV transformer can be interesting when considering the scalability of the data center infrastructure.

25.4.5.3 LV Distribution Alternatives with Classical Static UPS Technology

When selecting the LV topology with static UPS, the LV distribution for data center consists of the system composed by the MV/LV transformer, the main LV switchboard, the UPS units, and the final distribution using busway distribution or cable distribution as shown in Figure 25.41.

Depending on the habits and the rated voltage of the loads, several ways to design the LV distribution are possible. The use of LV/LV transformer in North America was initially used to convert the voltage from 480/277 V to 208/110 V in North America applications as shown in Figure 25.42. In other areas, as the voltage of the main LV switchboard and UPS are the same, it is possible to delete the LV/LV transformer as shown in Figure 25.43.

When the LV distribution is:

- Only TNS (Terra Neutral – Seperate) earthing system from the source down to loads;
- Without any insulation transformer;
- Equipped with an LV ATS to switch on an LV backup generator;

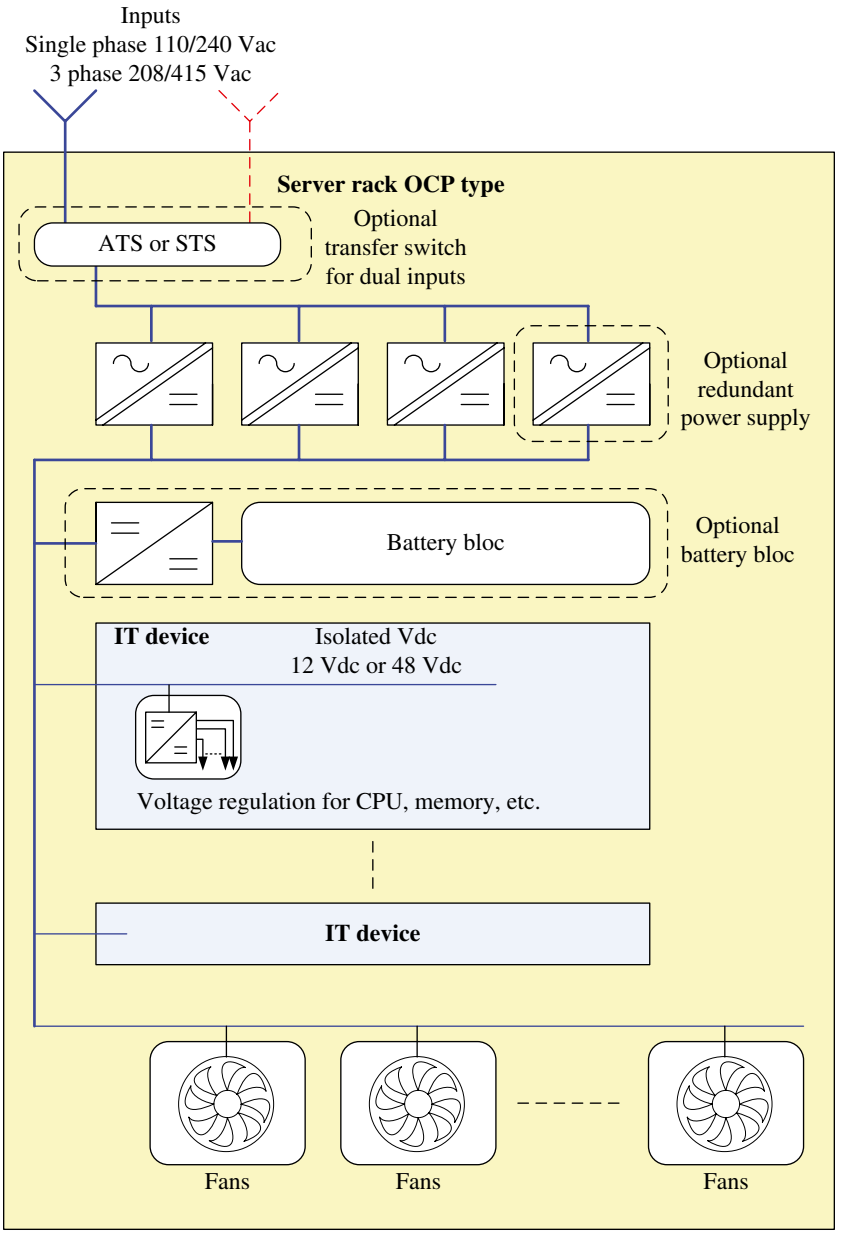


FIGURE 25.39 Server rack—open compute project-(OCP) type. Source: © Schneider Electric.

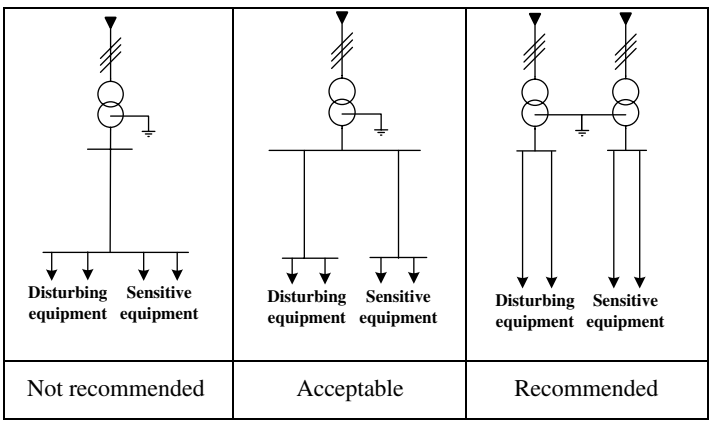


FIGURE 25.40 Load separation principle. Source: © Schneider Electric.

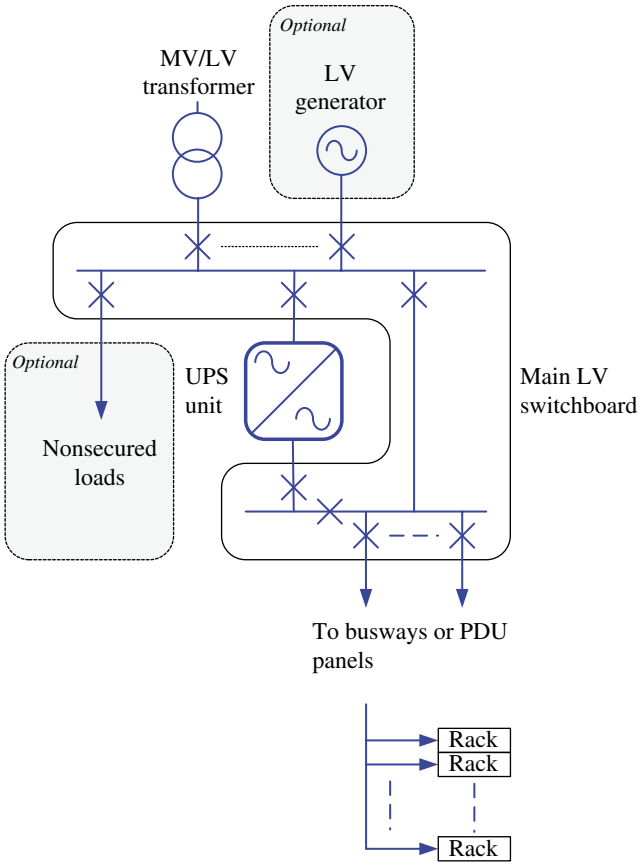


FIGURE 25.41 Overview of LV distribution for data center. Source: © Schneider Electric.

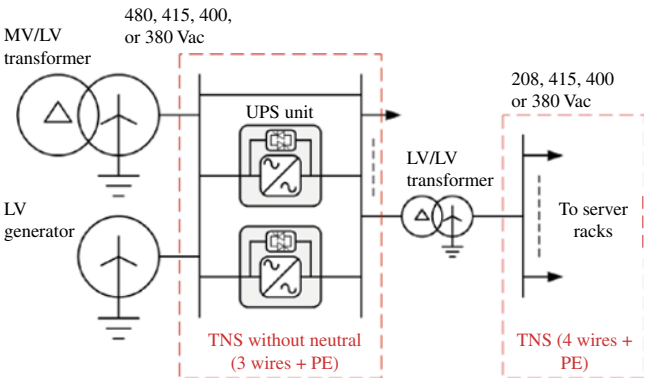


FIGURE 25.42 LV distribution using LV/LV transformer. Source: © Schneider Electric.

there is a sequence of operation where the UPS can operate ungrounded with possible overvoltages that can bother the server racks. Indeed, during the four-pole ATS switching from the MV/LV transformer to the generator supply, when the ATS has opened the transformer LV CB, the UPS loses the neutral reference to the ground. The UPS will suddenly operate in battery mode ungrounded and recovering the TNS

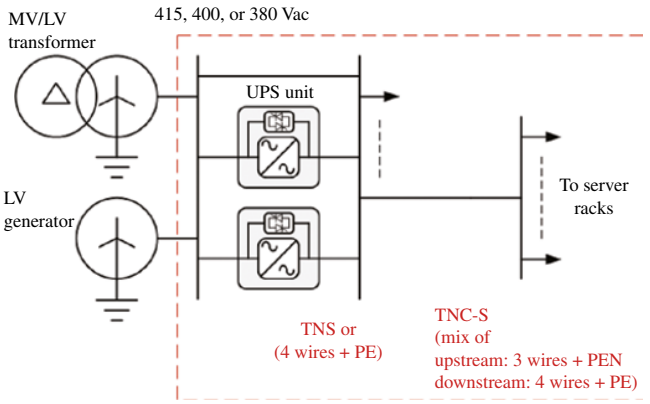


FIGURE 25.43 LV distribution with no insulation transformer. Source: © Schneider Electric.

earthing system when the ATS close the generator incomer CB. During the switching operations, the server racks downstream experience some phase-to-ground overvoltages that may stress the overvoltage protection of the server PSUs and that could lead to increase the PSU failure rate. To avoid such situations, the best option is to avoid the loss of neutral reference to the ground during the source transfer.

The distribution from the main LV switchboard to the server rack can be done using divisionary panels and cable distribution to the racks or using busways with tap-off boxes in each row of racks as shown in Figure 25.44.

The choice between busways and cable distribution has to be done according to the flexibility during installation and operating phases, the equipment cost, and the labor cost.

25.4.5.4 Main LV Switchboard Architecture

A standard LV switchboard architecture for a data center application, as shown in Figure 25.41, includes:

- The source incomers (transformer or generator)
- The UPS input CB
- The UPS output switches
- The UPS main bypass switch
- The outgoing feeder to supply the busways or the PDU panels
- The automatic transfer controller when needed
- The metering equipment
- The communication equipment

The main LV switchboard is an important piece of equipment in the data center distribution as:

- It is a key part of the UPS system;
- It deals with high constraints in terms of permanent current (up to 6,300 A) and high short-circuit current

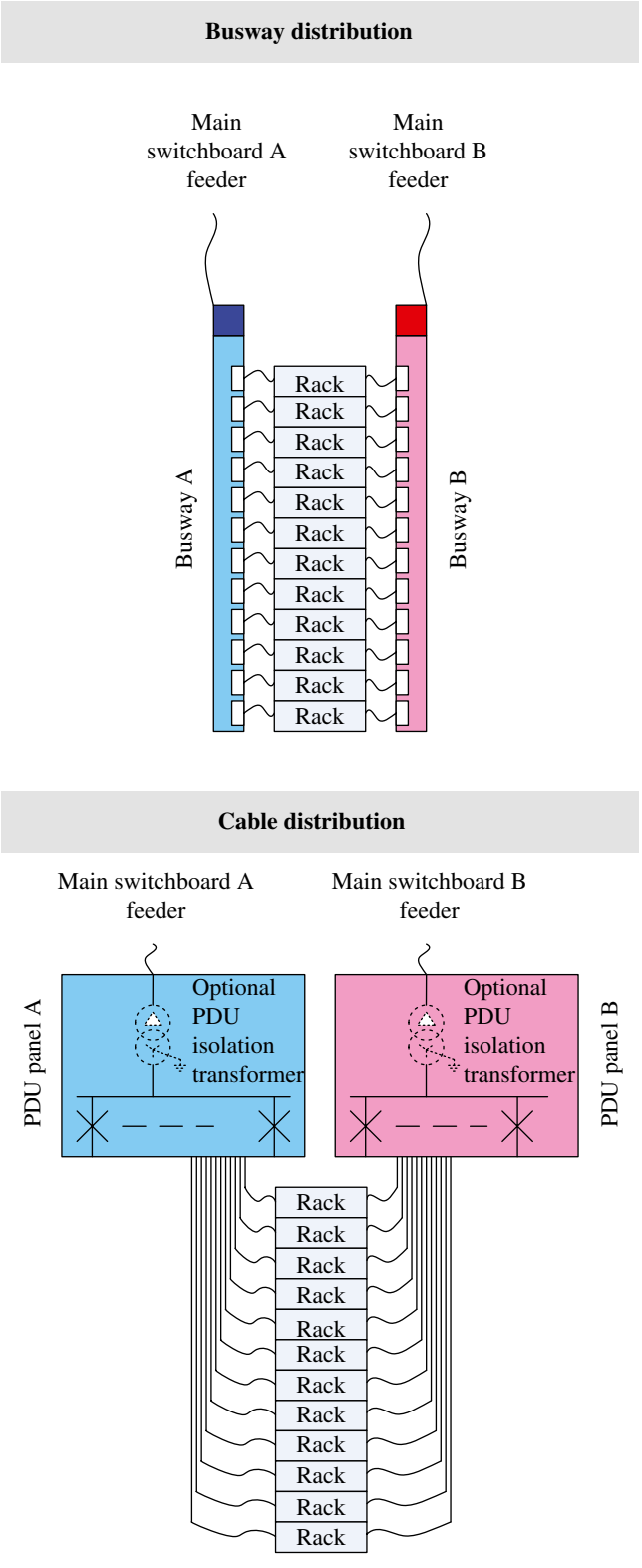


FIGURE 25.44 Alternatives for final LV distribution. Source: © Schneider Electric.

(up to 100 kA rms) and in terms of durability as the CB are operated regularly when doing the generator or UPS test and the maintenance of the UPS;

- It needs high performances in terms of safety (ensure people and equipment safety), reliability (ensure high reliable protection level, ensure very low risk of internal failure), and availability (ensure low unavailability duration in case of maintenance);
- It represents a significant part of the LV equipment cost.

LV switchboards are composed of LV devices such as CB and switches defined by the standard IEC60947 and the enclosure with its bus bar arrangement and its connections defined by the standard IEC61439.

The different technologies of LV CB are:

- Air circuit breakers (ACB) for current ratings from 1,000 to 6,300 A with parts that can be maintained (breaking poles, drive mechanism, and the trip unit); the tripping curve can be adjusted, and the trip order can be delayed for selectivity purpose;
- Molded case circuit breaker (MCCB) for current ratings from 100 to 1,600 A (only electronic trip unit is able to be tested); the tripping curve can be adjusted;
- Miniature circuit breakers (MCB) for current ratings from 2 to 125 A.

An important element is that LV CB have current deratings according to the temperature in the switchboard. Depending on the switchboard technology and the room ambient temperature, the CB current derating has to be taken into account when selecting the right current capacity.

Several specifications can have a significant impact on the design of the switchboard such as:

- Electrical main characteristics of the switchboard and its CB such as the voltage level that varies from 380 to 480 Vac and the permanent current rating goes up to 6,300 A;
- The short-circuit current values for the switchboard such as the rated peak withstand current (I_p that represents the electrodynamic constraint) and the rated short time withstand current (I_{cw} that represents the thermal constraint during a short circuit);
- The short-circuit current characteristics for the CB such as the rated peak withstand current (I_{cm} rated making capacity), the rated short-circuit breaking current (I_{cu} ultimate breaking capacity), and the rated short time withstand current (I_{cw});

- The internal arc protection options such as the internal arc-type test or arc free using solid insulation;
- The switchboard functional unit type (withdrawable or fixed type);
- The form factor of the switchboard (defined by IEC 61439 that specifies several levels named "form" of internal separation);
- The IP index (protection index code according to IEC 60529 that defines the level protection of electrical enclosure against intrusion, dust, accidental contact, and water) and the room temperature that can lead to increase the derating of the CB and the switchboard column;
- The CB duty factor that can provide useful information to optimize the number of CB in a single column.

Classically, main LV switchboards for data center application have the following features:

- Switchboard technology with form factor 4b (separation of the bus bars from the functional units and each functional unit from the other units, separation of the terminals for a functional unit from the bus bars, and those of any other unit).
- Withdrawable or equivalent (such as plug-in mounting plate) to allow to add new feeder CB or maintain any CB without shutting down the switchboard.
- Internal arc-type test.
- IP31 and 40°C maximum ambient temperature are enough as the switchboard is generally installed in a room-equipped air conditioning.

When designing the LV distribution and the main LV switchboard, it is important to remind few key elements to optimize the cost and footprint:

- Use as possible the full capacity of the CB.
- Simplify the LV switchboards by grouping the input and output UPS switchboards.
- Check that the specified short-circuit values do not lead to oversize the CB such as going from single pole CB to double pole CB.
- Optimize the switchboard by specifying the right rated short time withstand current Icw according the effective fault clearing time.
- Check that all options are useful for the customer such as:
  - The close transition between the MV/LV transformer and the LV generator
  - The UPS load bank
  - The generator load bank CB
  - MV/LV transformer advanced protection that can require additional LV current sensors
  - The power monitoring equipment that requires additional current sensors
- Decrease the number of columns by:
  - Taking into account the CB duty factors to optimize the number of CB per column (example of the transformer incomer CB and generator incomer that never runs in parallel),
  - Designing special bus bar arrangement.

An example of classical LV switchboard architecture is shown in Figure 25.45.

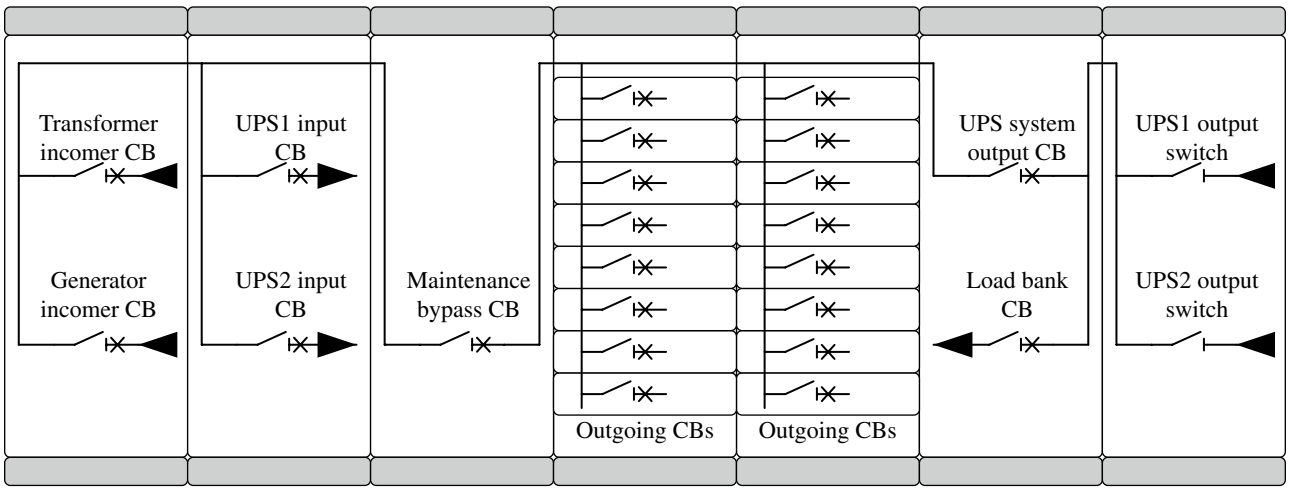


FIGURE 25.45 Example of main LV switchboard with a power train composed of two UPS units and a UPS load bank. Source: © Schneider Electric.

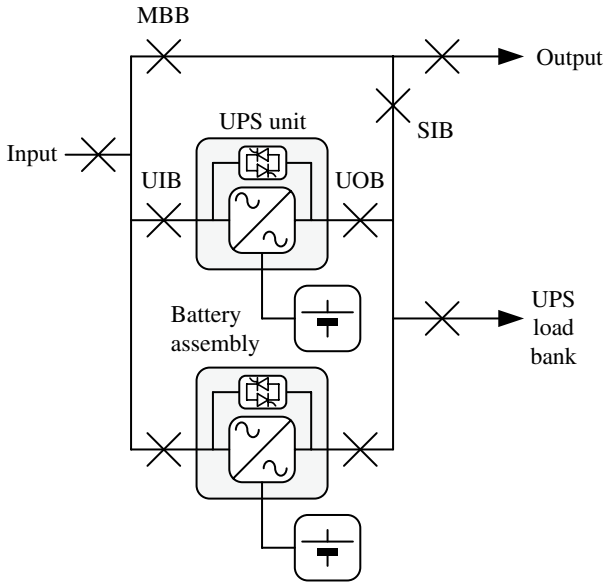


FIGURE 25.46 Example of UPS system for data center. Source: © Schneider Electric.

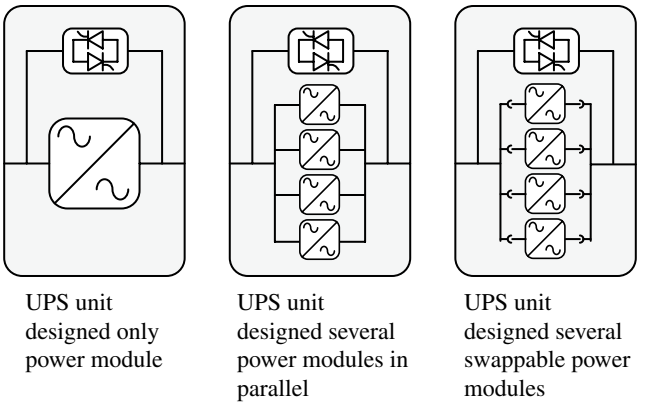


FIGURE 25.47 UPS unit internal modularity. Source: © Schneider Electric.

25.4.5.5 Static UPS System

An example of a classical architecture of 3-phase UPS system for data center is shown in Figure 25.46. Two units are put in parallel to reach the required capacity to supply the load. Each unit is equipped with its battery system and its static bypass switch. Depending on the UPS products, it is also possible to have one single bypass static switch for all UPS units. A main manual bypass switch (MBB) is able to take the whole load and allows to perform the maintenance on both UPS units while supplying the load. Thanks to the system isolation breaker (SIB), it is also possible to supply the load through the manual bypass switch and make a UPS test on a load bank.

Depending on the UPS products, several UPS system configurations are possible depending on the UPS unit size, the number of UPS in parallel, the redundancy level of UPS units or modules, the UPS unit modularity, and the option to have a centralized static bypass (Figs. 25.47 and 25.48).

To optimize the best UPS system design, several parameters have to be considered as follows:

- The UPS size and modularity that should fit the data center load growth
- The reliability and availability performance of the UPS system according to the mean time to failure of a single unit, its ability to be fault tolerant, and the ability to be maintained quickly and without shutting down the UPS system
- The overall UPS system efficiency according to the UPS conversion mode and the efficiency performances from low load to full load conditions
- The compatibility of the UPS product with the specified battery technology

To select the best UPS product, the size of the UPS needs to fit the IT room specification. The scalability of the UPS system can be defined at different stages:

- The granularity can be set at the UPS system level, so that at each step of IT load, the data center owner will add a one or several UPS system.
- The granularity can be set at the UPS unit level, so that at each step of IT load, the data center owner will add a new UPS unit in the UPS system.
- The granularity can be set at the UPS power module unit level, so that at each step of IT load, the data center owner will add a new module in the UPS unit.

The reliability and the availability performance are functions of the UPS mean time to failure, the UPS system redundancy level, the UPS behavior in case of fault, and the way to repair or replace the failed element.

UPS units are exposed to internal failures such as:

- Power electronic component failures
- Control of power electronic failures
- Cooling fan failures
- Main control of the unit failures
- Other failures (insulation failure etc.)

A great part of the UPS failures affects only a single power module. Depending on the load factor, on the UPS system modularity, and on the UPS unit internal modularity, a failure can affect only one module, or only one UPS unit, or the

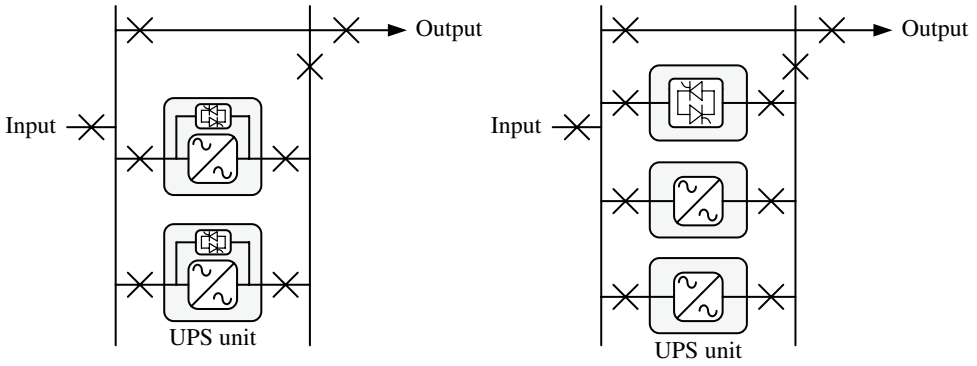


FIGURE 25.48 UPS unit static bypass switch alternatives. Source: © Schneider Electric.

whole UPS system. The general behavior can be summarized as follows:

- The failure affects only one UPS module or one UPS unit. The failed module or unit is shut down and automatically isolated from the rest of the UPS system. If the UPS system still has the capacity to supply the load, the UPS system will continue to work. Otherwise the UPS units will be switched to bypass static switches. To replace or repair the failed element, the operator would need to lock out the UPS unit using the input and output CB or only switch on static bypass to perform the maintenance operation.
- The failure affects the entire UPS system and will lead to a complete blackout.

In terms of power system energy efficiency, the UPS is a key element as it can lead to 10% losses for the worst solution. The energy efficiency performance of a UPS relies on the UPS unit design performance and the UPS ability to optimize the number of running units.

The three main UPS conversion modes give different performances in terms of output voltage quality and energy efficiency.

In double conversion mode as shown in Figure 25.49, both units are running in parallel in conversion mode: the power is going through the UPS rectifiers and inverters; the static bypass switches and the manual bypass switch are normally opened. The protection against input disturbances is maximal. In this mode, the voltage quality is guaranteed thanks to the inverter voltage control. The UPS input rectifier stage is classically designed to take an input current with very low harmonic distortion and a power factor close to 1 at full load.

The operating principle of the ECO (economical) mode, as shown in Figure 25.50, is that the power is going through the static switches. The UPS rectifiers and inverters are in standby, and the manual bypass switches are opened. If an input disturbance occurs, the UPS will open the static

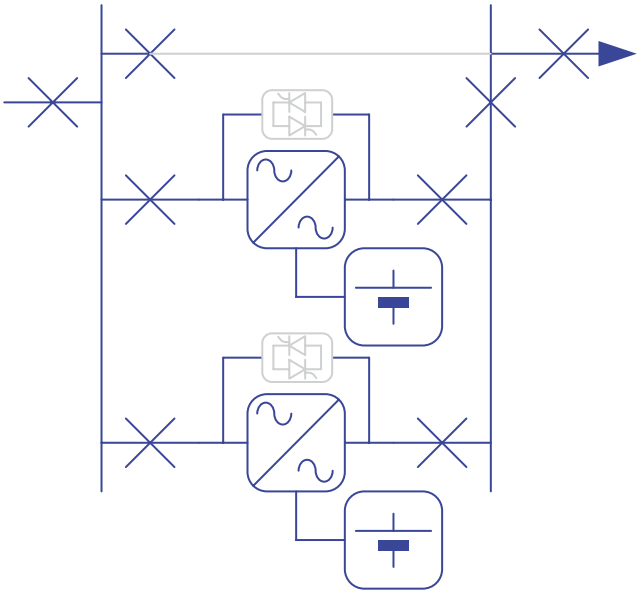


FIGURE 25.49 UPS system in double conversion mode. Source: © Schneider Electric.

switches and switch on the UPS in double conversion mode or battery mode if the input voltage is out of tolerance.

The main advantage of the ECO mode is the efficiency that can reach 99%. However, two main risks are existing with a UPS in ECO mode and need a detailed analysis to avoid any risk of blackout:

- The logic used to determine the input voltage status (to switch from bypass to conversion mode) needs to take into account all possible scenarios to avoid any situation where the voltage is not suitable for the loads and the UPS remains in standby mode.
- The time to switch to conversion mode (including the time to detect abnormal conditions, the time to switch off the static switch, and the time to switch on the inverter) has to be compatible with the voltage quality required by the loads.

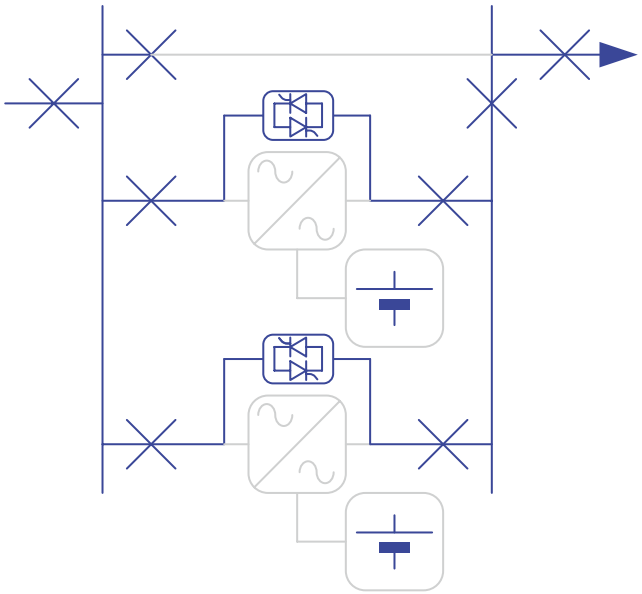


FIGURE 25.50 UPS system in ECO mode. Source: © Schneider Electric.

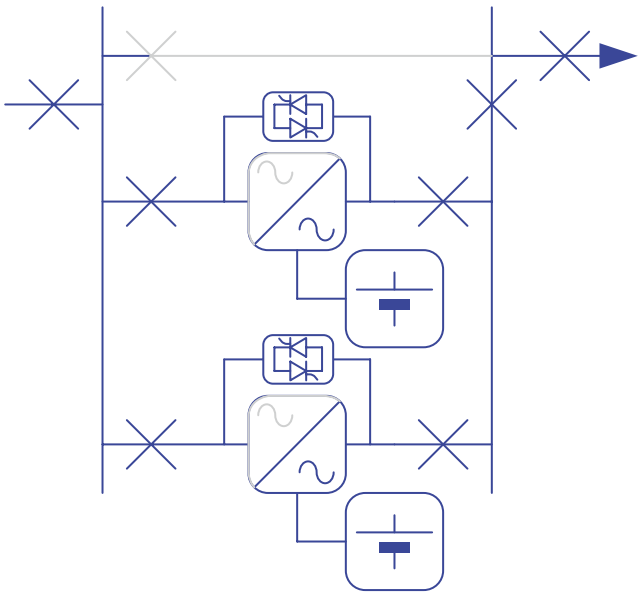


FIGURE 25.51 UPS system in active filter mode. Source: © Schneider Electric.

The active filter mode principle is to use the static bypass to supply the load as in the ECO mode except the UPS output inverter is active and used as an active filter to compensate the downstream load current harmonics and reactive power. The active filter mode gives also the advantage to have a better response when the UPS needs to switch to battery mode in case of input disturbance. The active filter mode efficiency is closed to the efficiency of the ECO mode (Fig. 25.51).

When designing a UPS system, the maximum input current will be needed when selecting the input LV CB of the UPS. A particular attention should be paid to the calculation of the maximum input current. The designer can choose to take the UPS maximum input current coming from the product data sheet that is given for the worst conditions (lowest input voltage, full load, highest charging rate), but sometimes it can lead to oversize the LV input switchboard. To avoid this oversizing, it is possible to calculate the maximum current according to the real worst conditions in the data center.

The battery system for a UPS consists of several battery strings connected in parallel, each string being composed of several battery module in series (Fig. 25.52).

The main steps for the battery system design are the following:

- Step 1 "Select the battery type"
  - According to the CAPEX, the lifetime, the mission profile, and the installation constraints, the end user will select the best option to optimize the overall TCO.
  - Nowadays, the two major battery technologies used for data center application are valve-regulated lead–acid (VRLA) battery and lithium-ion (Li-ion) battery. For VRLA, it is the most popular reserve power design because the electrolyte is captive. There are two types of VRLA batteries: AGM (absorbent glass mat) in sulfuric acid and paste-like gel cell. Lithium-ion or li-ion batteries are used for electronics, electric vehicles, or aerospace applications. Li-ion batteries deliver more power but use less space than VRLA. Although Li-ion batteries have a higher cost, but there are several benefits over VRLA batteries. A Li-ion battery has longer service life, it requires less cooling due to it can withstand at higher temperature range.
- Step 2 "Size the battery strings"
  - The number of battery in series will be set to match the UPS battery bus voltage range.
  - The number of battery string in parallel is set to match the required load power and autonomy according to the battery data sheet; a derating can be applied to the battery maximum power if the autonomy is specified as the battery end of life.
- Step 3 "Select the conductors and the protection devices"
  - Calculate the maximum discharging current, the maximum and minimum short-circuit current for both states "full charged" an "discharged."
  - Select the conductors and the protection devices; for VRLA batteries, the protection of the battery system is better when using a CB per string.
- Step 4 "Implement the battery system"
  - The battery modules are installed and wired in a battery cabinet or a battery shelf.
  - Lithium-ion batteries are typically designed with a monitoring system that includes protection functions

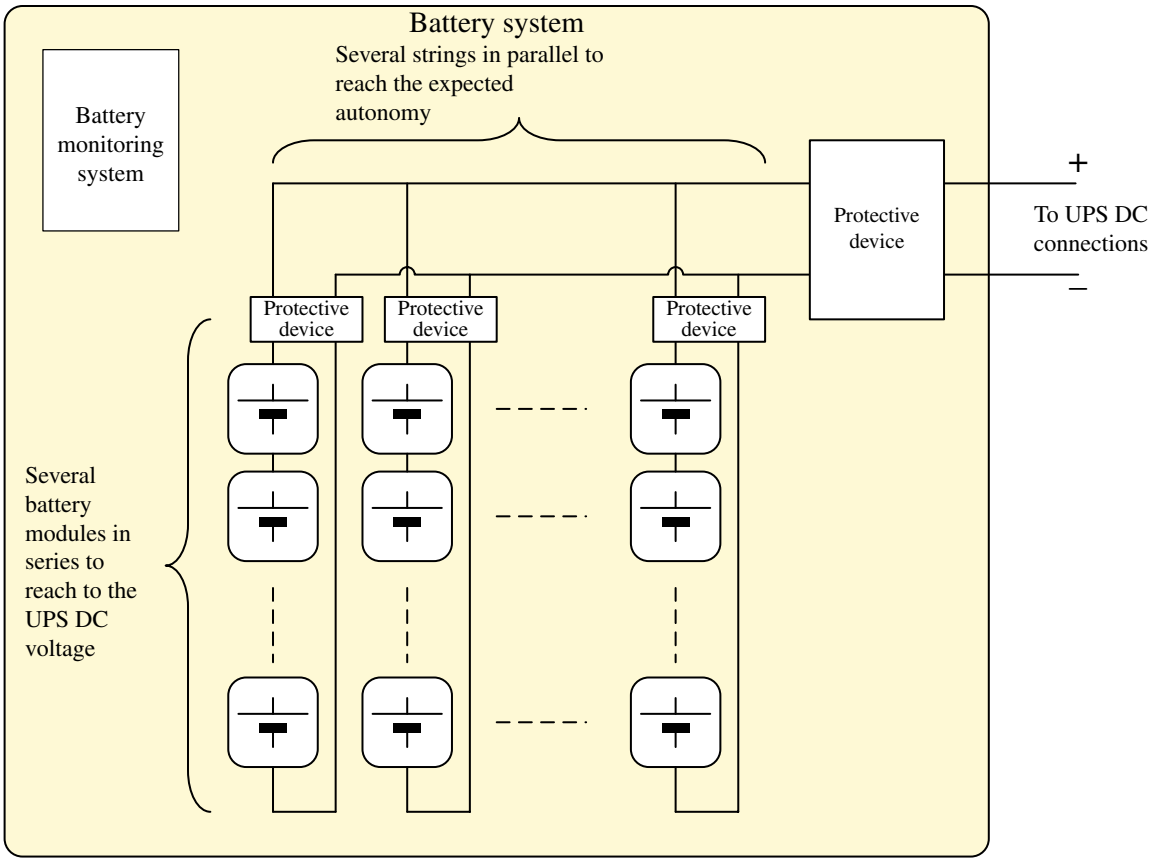


FIGURE 25.52 Battery assembly. Source: © Schneider Electric.

and battery healthy state monitoring; for VRLA battery an optional monitoring system can be added; the UPS unit can also provide a basic battery health monitoring using the voltage and current measurements and a discharge test.

- The battery system is typically installed in a separate room with room cooling and other functions to mitigate safety risk such as the room ventilation and/or a fire detection system depending on the selected battery technology.

The UPS load bank can be an option to test the UPS at full load during the commissioning phase. However, the use of the UPS load bank during the operation phase can be discussed because some UPS units can be equipped with some test function able to test both the battery system and the UPS unit at full load. Such simplification can provide a significant cost savings on the LV power architecture as it can avoid two main CB on each main LV switchboard and the load bank busway.

25.4.5.6 Static Transfer Switch

Typically based on the thyristor technology, the STS control unit uses a fast undervoltage logic to be able to make a changeover in less than few milliseconds so that the downstream loads experience no interruption.

The STS base current range goes from 30 to 1,600 A so that it can be used at rack level or just downstream the main LV switchboard. When using the STS in four-wire system, the STS must be able to switch the phases and the neutral.

The key features of an STS are:

- The maximum rated voltage (typically 440 V maximum) and rated current (range from 30 to 1,600 A);
- The maximum short circuits at the inputs that can vary from few kA rms to 50 kA rms;
- The STS thyristor short-circuit withstand;
- The number switching poles (three or four poles);
- The option of neutral overlapping to minimize the overvoltage risks and nuisance tripping in TNS earthing system;
- The short time transfer that can vary from 3 to 15 ms;
- The optional bypass switch to perform the regular maintenance operation while keeping the output alive;
- The option to make a changeover while minimizing the inrush current of the downstream LV/LV isolation transformer;

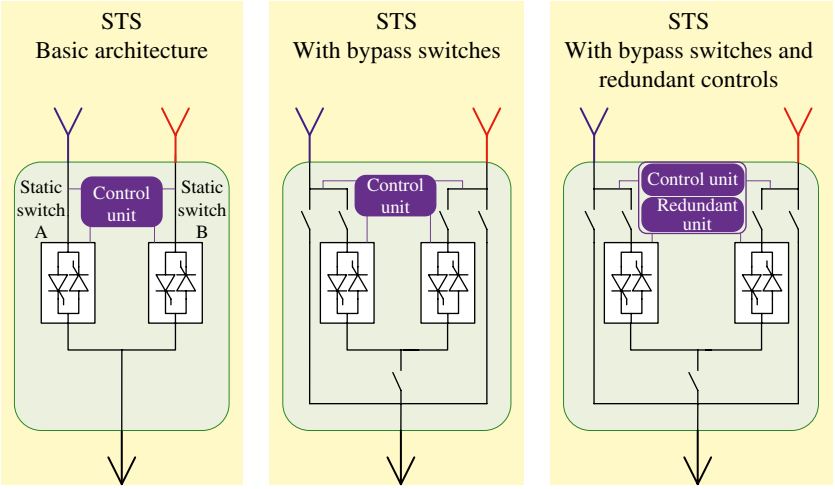


FIGURE 25.53 STS architectures. Source: © Schneider Electric.

- The internal redundancies in the unit to provide the best product reliability level such as the redundancy on the power supply, on the cooling fans, and the SCR drivers and/or on the control unit;
- Its behavior in case of downstream short circuit (option to not make a changeover when a downstream short circuit is detected);
- Its compartmentalization between the A and B incomers to avoid a simultaneous fault on both A and B incomers.

When the STS is used in a block redundant architecture, a detailed analysis should investigate all the possible fault scenarios including the STS short-circuit withstand capabilities compared with the energy limited by the upstream of downstream CB of the STS, the common mode failure between A and B incomers in the STS cabinet, and the STS automation behaviors (Fig. 25.53).

25.4.5.7 LV Protections

LV Protection Discrimination with Static UPS

In case of a short circuit downstream the UPS, the UPS behavior depends on the UPS product options and settings:

- The UPS can supply the short circuit through the inverter with a current limitation, and after a time delay if the fault is not cleared, the UPS will shut down.
- The UPS can also be set to switch instantaneously on the static bypass and supply the short circuit through the static bypass switch without any current limitation and get back to normal operation when the fault is cleared (Fig. 25.54).

When supplying a fault through its inverter, the UPS limits its output current due to the thermal constraints on the inverter IGBTs. This current limitation is done by the inverter current control loop and can be a fixed value during a period of time or a current/time curve as shown in Figure 25.55. Depending on each UPS product, the maximum current limitation can go typically from 1.5 to 3 times the rated current.

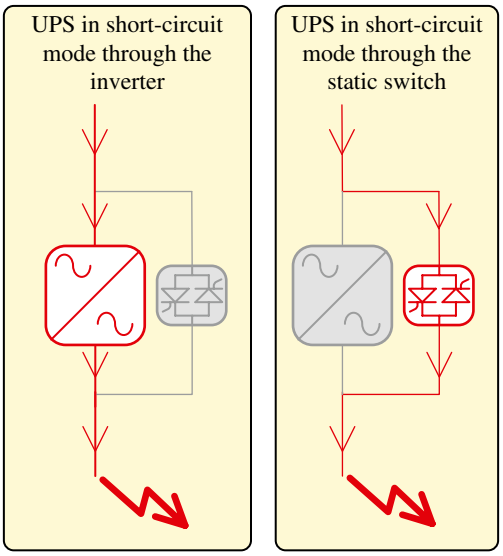


FIGURE 25.54 UPS behavior in case of downstream short circuit. Source: © Schneider Electric.

When taking the example of a typical UPS architecture as shown in the Figure 25.56 with several units in parallel, the different fault scenarios that can be studied could be a fault downstream the UPS, a fault in the UPS input (in the UPS rectifier), a fault in the UPS output (in the UPS inverter), or a fault on the UPS output bus.

The fault analysis depends on the fault current flows, on the static switch current withstand, on the UPS behavior, and

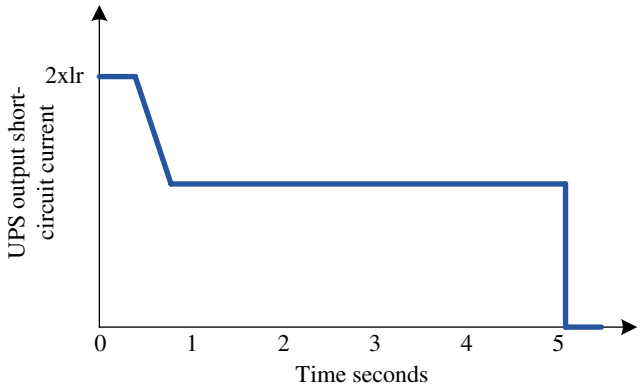


FIGURE 25.55 Example of an output short-circuit current curve of a UPS. Source: © Schneider Electric.

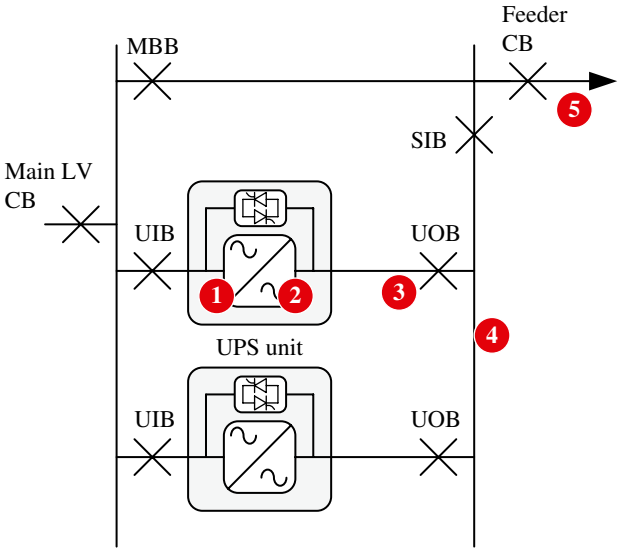


FIGURE 25.56 UPS fault scenarios. Source: © Schneider Electric.

on the protection settings. Whatever, the full discrimination between the CB downstream the UPS and the CB upstream cannot be totally because of the multiple branches in parallel. However, the most important is to achieve a full selectivity for the most frequent fault scenarios such as:

- Scenario 1: The fault should be cleared by the rectifier input fuses or at least the UPS input CB, so that the main LV CB remains closed and the other UPS units remain alive.
- Scenario 2: The fault should be cleared by the UPS inverter output fuses so that the other UPS units remain alive without any damage of the UPS static switches.
- Scenario 3: The fault should be cleared by the feeder CB without any damage of the UPS static switches.

LV Protection Discrimination Versus Switchboard Safety

When achieving the discrimination between several air CB (such as the main LV CB, UIB, UOB, SIB), the basic way is to disable the instantaneous trip function and to make time-graded discrimination. However, it will increase the delay and the current threshold of the overcurrent protection of the main LV CB and could lead to lower protection level of the main LV switchboard (Fig. 25.57).

In this case, using logic discrimination or arc fault detection device permits to keep a fast tripping in case of an internal arc in the main LV switchboard and keep a high level of safety.

Optimize the LV Architecture Using the Circuit Breaker Limiting Effect

MCB and MCCB have short-circuit limiting effect thanks to their fast opening of their contacts in less than half a cycle. The arc voltage between the CB pole limits the short-circuit

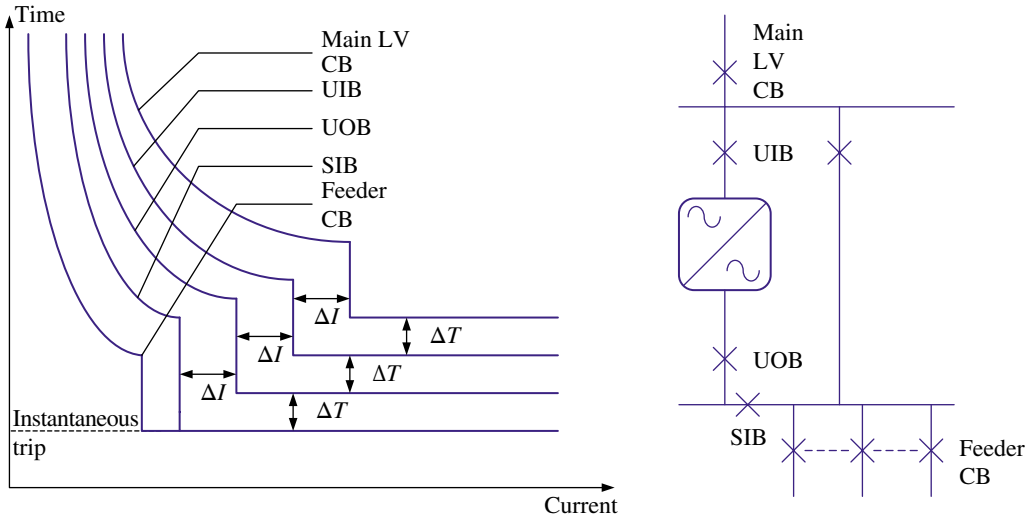


FIGURE 25.57 Time-graded selectivity between LV ACB. Source: © Schneider Electric.

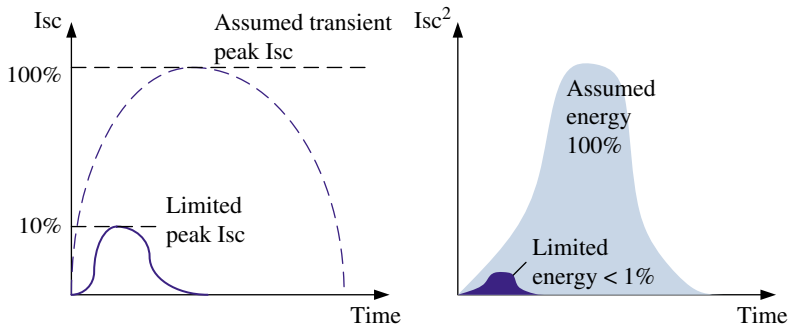


FIGURE 25.58 Limiting capacity of LV circuit breaker. Source: © Schneider Electric.

current peak and the let-through energy as shown in the Figure 25.58.

The CB limitation capacity limits the mechanical and the thermal stress induced by a short circuit. Using a coordination between the CB and the downstream equipment, the short circuit withstand of the equipment can be reduced. This coordination can be applied for products such as UPS, STS, LV panels, and busways.

Another advantage of the limitation capacity is the LV cascading where the upstream CB helps the downstream CB to open and clear the fault. Cascading provides circuit breakers placed downstream of a limiting circuit breaker with an enhanced breaking capacity. Cascading makes it possible to use a circuit breaker with a breaking capacity lower than the prospective short-circuit current calculated at its installation point. Using the cascading tables of LV CB manufacturers, the designer can optimize the cost of the LV CBs. It is also important to keep in mind that the full selectivity can be combined with cascading, also called as "selectivity enhanced by cascading."

25.4.5.8 LV Power Monitoring System

The monitoring at LV level includes the monitoring of the status of the switches and CB, the status of the trip units and the PLCs, and the voltage presence indicators. The equipment monitored are the main LV switchboards, the UPS units, and the final distribution to the racks such as the busways and/or the PDU or the remote power panels (RPP).

Some metering equipment are also classically located:

- At the incomer of the main LV switchboards to provide energy metering and power quality monitoring (voltage and current harmonics, voltage sags);
- At the UPS output bus to power quality monitoring;
- On the row busway feeder CBs to provide energy metering;
- On each rack feeder CBs to provide energy metering (for colocation application mainly).

It is important to define the right precision class needed of all the measurements to avoid any useless over costs. As an example, the metering function can be achieved by the CB trip unit, but if a high-class metering is specified on several locations of a main LV switchboard, the LV switchboard would be bigger to accommodate the current sensors. It becomes more expensive because more space will be required as well as material and labor to install more cabling.

25.4.6 Overall Design Optimization

When designing the data center infrastructure, the first key point for optimization is the scope that needed to take into account both electrical and cooling system and the right estimation of maximum power for the IT loads and for the cooling equipment consumptions.

25.4.6.1 Optimization Using Reliability Studies

The power system reliability and availability performance highly depend on the overall architecture that is composed not only of the power distribution equipment but also of the protection, control and monitoring systems, and of the maintenance for planned activities. The designers need to take decision such as:

- Choose the right level of redundancy on the grid connection substation, the generator power plant, and the MV/LV electrical distribution.
- Choose the right level of redundancy for the critical auxiliary supplies of the HV/MV substation, the generators, and the MV equipment and the LV equipment and check it does not lead any single point of failure.
- Use manual transfer equipment to keep running the maximum equipment during the planned shutdown of a device or a group of devices.
- Set the adequate tests on generators, UPS, STS, and switchboard protections and automation to be able to maintain the system in a perfect health.

A system reliability assessment can evaluate the impact of different design options on the system reliability performance. A system reliability analysis basics are to study the consequences of component failures on the system based on the knowledge of the equipment failures and the system behaviors in case of failures.

For a detailed analysis, the data needed to perform a dysfunctional analysis include the equipment specifications, the system architecture, a description of the operating modes and degraded modes, a description of the automation and protection behaviors, the layouts, the equipment reliability data (failure modes and failure frequency), the equipment planned and unplanned maintenance data, and the on-site maintenance.

By analyzing all possible failure sequences (single contingency and multiple contingencies if needed), the reliability analysis estimates the mean occurrence frequency and the probability of undesirable events such as "the loss of one IT rack," the loss of one row," "the loss of an IT room," or "the loss of the whole data center." From these results, the system weak points can be identified and give some starting points to the designers to upgrade the system.

25.4.6.2 Optimizing the Reliability and Availability During the Data Center Operation

During data center operation, the main tasks to ensure the reliability performances of the system is to manage the planned maintenance activities and to ensure the required time to react in case of failure.

The planned maintenance activities need to be defined according to the manufacturer recommendations and also fine-tuned according to the site conditions. An example of planned operations mentioned is given in Table 25.6.

When a failure occurs, the time to fix the issue will depend on several steps:

- The detection of the failure by the monitoring system that monitors basically the status of the power system equipment (status, voltage and current measurements, room temperature, etc.); a particular attention should be paid to the critical equipment to be able to detect any "hidden" failure and maintain the critical equipment functions ready to operate when required; the monitoring system can also include premium functions such as energy, harmonics, wave captures, equipment temperature, and specific services such as predictive maintenance based on the diagnostic-based detailed monitoring data.
- The site maintenance team will perform manual operations such as manual power system reconfiguration, manual restart, and lock out failed equipment; to ensure a high skill of the site maintenance teams, the site maintenance operators need to be trained on these emergency operations using if needed specific procedures.
- The failure diagnostic, the equipment spare delivery time, and the manufacturer maintenance contracts for critical equipment to guarantee the intervention time and the equipment spare availability.

25.4.6.3 TCO Optimization

The best approach is to optimize the overall architecture from the grid substation down to the loads and to consider the TCO including the CAPEX and OPEX (losses and maintenance cost).

The following good practice can help to optimize the power system architecture:

- Specify the IT room architecture to match the best electrical ratings.
- Use low loss equipment such as Ecodesign transformers, low loss UPS products, and low loss server PSU.
- Minimize the LV cable length.
- Choose the best MV voltage level and try to avoid any MV/MV transformers.
- Try to avoid any LV/LV transformers.
- Avoid very high short-circuit current rating at MV level and at LV level to keep cost-efficient equipment ranges.
- Opportunity to provide grid services with UPS and/or generators.

25.5 FACEBOOK, INC. ELECTRICAL DESIGN

These electrical topologies are not mutually exclusive; the key is to design a data center that satisfies business needs. Facebook designed a data center that merges these topologies (Fig. 25.59), resulting in a solution satisfying their requirements. The data center comprises a mix of 208 and 277 V equipment as well as single- and dual-corded servers.

The Facebook data center design team developed a revolutionary design that does not require a centralized UPS, significantly reducing losses. In this design, power flows from the utility, connecting directly to the 277 V server; battery backup cabinets are connected to the servers delivering DC power in case of an outage.

Overall, the Facebook data center follows the block redundant configuration with a reserve bus that provides power to one of the six independent systems if a failure occurs.

Figures 25.60 and 25.61 illustrate a typical Facebook-designed suite. 277 V power is distributed to the Facebook OCP servers.

TABLE 25.6 Example of planned operations for a data center site

Equipment	Preventive maintenance operation	Period
HV line incomers	Incomer locked out for HV bushing cleaning	6 months
HV GIS section	Inspection	6 months
	Controls and verification	6–12 years
HV GIS circuit breaker	Inspection	1–6 months
	Opening/closing	6 months
	Controls	3 years
	Verification and revision	12 years
HV GIS disconnector	Inspection	6 months
	Opening/closing	1 year
	Controls	3 years
	Verification	12 years
HV/MV transformers	Cleaning and control of transformer neutral point impedance	6 months
	Controls	3 years
	Verification and oil analysis	6 years
	Revision	12 years
HV protections	Protection test	1 year
HV auxiliaries	Inspection	6 months
	Battery charger capacity test Verification of LV protection	1 year
MV switchboard	Opening/closing	1 year
	Switchgear visual inspection, cleaning, MV protection relay test	3 years
	Switchboard visual inspection, cleaning	3 years
MV/LV dry transformer	Connection verification and cleaning	1 or 3 years
Generator unit	Inspection and checks (coolant heater, coolant level, oil level, fuel level, charge-air piping)	1 day
	Check/clean air cleaner, check battery charger, drain fuel filter, drain water from fuel tank	1 week
	Check coolant concentration, drive belt tension, starting batteries Drain exhaust condensate	1 month
	Change oil and filter, change coolant filter, clean crankcase breather, change air cleaner element, check radiator hoses, change fuel filters	6 months
	Clean cooling system	1 year
	Off-load test	1 week to 1 month
	On-load test with the load bank	1 onth

TABLE 25.6 (Continued)

Equipment	Preventive maintenance operation	Period
Generator power plant	On-load test with MV failure simulation	1 month to 1 year
LV switchboard	Opening/closing	1 year
	CB checks, protection relays tests	3 years
	Switchboard visual inspection, cleaning, connection checks	6 years
UPS	Visual inspection and cleaning	1 year
	DC capacitors and fans replacement	5 years
	Power supply board replacement	7 years
	Filter replacement	10 years
UPS	Visual inspection and cleaning	1 year
	DC capacitor and fan replacement	5 years
	Power supply board replacement	7 years
	Filter replacement	10 years
	On-load test	1 year
Batteries	Visual checks	1 week
	Connection checks	1 year
	Replacement	5 or 10 years
Busways	Visual inspection and thermography	1 year

Source: © Schneider Electric.

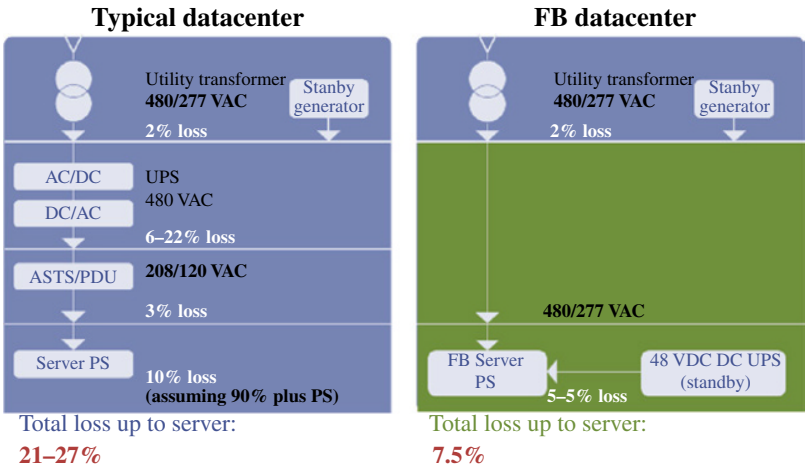


FIGURE 25.59 Typical vs. Facebook electrical topologies. Source: © Facebook.

FIGURE 25.60 Facebook data center suite. Source: © Facebook.

FIGURE 25.61 Electrical room at Facebook data center. Source: © Facebook.

FIGURE 25.62 Facebook DC UPS battery cabinet. Source: © Facebook.

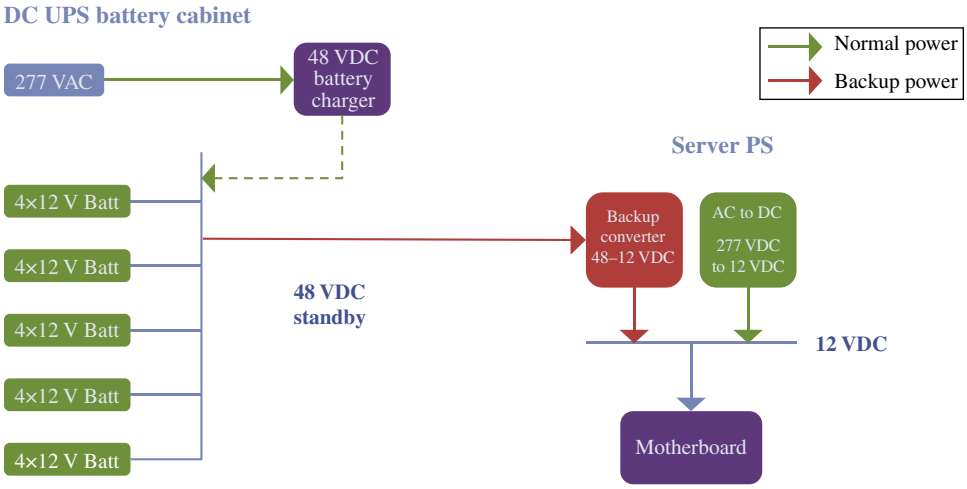


FIGURE 25.63 DC UPS backup scheme. Source: © Facebook.

Since there isn't a centralized UPS, the DC UPS battery cabinet, in Figure 25.62, distributes power to the servers when failures occur.

Figure 25.63 is a diagram that goes into depth about the power configuration of a typical DC UPS battery cabinets and 277 V server.

FURTHER READING

Cabau E. *Introduction à la Sûreté de Fonctionnement*. Cahier technique Schneider Electric, nr 144. Grenoble: Schneider Electric; June 1999

Gatine G. *High Availability Electrical Power Distribution*. Schneider Electric, nr 148. Grenoble: Schneider Electric; 1990. Available at https://download.schneider-electric.com/files?p_enDocType=Cahier+Technique&p_File_Name=ECT148.pdf&p_Doc_Ref=ECT148. Accessed on August 24, 2020.

Logiaco S. *Electrical Installation Dependability Studies*. Cahier technique Schneider Electric, nr 184. Grenoble: Schneider Electric; December 1996

Merlin Gerin. *Protection Guide (Electrical Network Protection)*. Schneider Electric; April 2006. Available at https://download.schneider-electric.com/files?p_enDocType=Catalog+Page&p_File_Name=CG0021EN+%28web%29.pdf&p_Doc_Ref=CG0021EN. Accessed on August 24, 2020.

Mitchell B. How to choose IT rack power distribution. White Paper 202, Schneider Electric; 2015. Available at https://download.schneider-electric.com/files?p_File_Name=VAVR-9G4N7C_R0_EN.pdf&p_Doc_Ref=SPD_VAVR-9G4N7C_EN&p_enDocType=White%20Paper. Accessed on August 24, 2020.

Rasmussen N. Avoiding costs from oversizing data center and network room infrastructure. White Paper 37, Rev. 7, Schneider Electric; 2012. Available at https://download.schneider-electric.com/files?p_File_Name=SADE-5TNNEP_R7_EN.pdf&p_Doc_Ref=SPD_SADE-5TNNEP_EN&p_enDocType=White%20Paper. Accessed on August 24, 2020.

Rasmussen N., Torell W. Data center projects: establishing a floor plan. White Paper 144, Rev. 2, Schneider Electric; 2015. Available at https://download.schneider-electric.com/files?p_File_Name=VAVR-6KYMZ7_R2_EN.pdf&p_Doc_Ref=SPD_VAVR-6KYMZ7_EN&p_enDocType=White%20Paper. Accessed on August 24, 2020.

Schneider Electric. Low voltage expert guides no. 5: Coordination of LV protection devices; June 2009. Available at https://download.schneider-electric.com/files?p_enDocType=Technical+leaflet&p_File_Id=9601264025&p_File_Name=DBTP107GUI_EN+%28web%29.pdf&p_Reference=DBTP107GUI_EN. Accessed on August 24, 2020.

26

ELECTRICAL: UNINTERRUPTIBLE POWER SUPPLY SYSTEM

Chris Loeffler and Ed Spears
Eaton, Raleigh, North Carolina, United States of America

26.1 INTRODUCTION

Uninterruptible power supplies (UPSs) are an extremely important part of the electrical infrastructure where high levels of power quality and reliability are required. In this chapter we will discuss basics of UPS designs, typical applications where UPS is most commonly used, considerations for UPS selection, and other components or options that are an important part of purchasing and deploying a UPS system.

26.1.1 What Is a UPS

Put simply, a UPS is a device that provides backup power when mains power fails or becomes unusable by devices requiring regulated electricity to operate. Either the UPS can provide electricity long enough for critical equipment to shut down gracefully so that no data is lost, or process is interrupted or long enough to keep required loads operational until another electrical generating source (typically a generator) comes online. Some of the different UPS topologies also provide conditioning to incoming power so that all-too-common sags and surges don't damage sensitive electrical and electronic gear. UPS systems are designed to integrate easily into the standard electrical infrastructure, so that means smaller power requirements are typically single-phase designs, with larger power requirements being handled by three-phase systems. In North America, the typical single-phase UPS design is smaller than 25 kVA, while three-phase systems start around 8 kVA and go up into the MVAs. In some counties in Europe, all systems larger than 8 kVA must have a three-phase input to make sure the mains electrical system stays balanced. Single UPS systems come in sizes ranging from 300 VA (enough power for a typical PC and monitor) to over 2 MVA (enough power for 175 homes), with larger systems being able to be installed in parallel for power levels as high as +20 MVA (enough power for a small town).

26.1.2 Why a UPS

In this age of critical computing systems and the Internet, business continuity requires that you protect your IT infrastructure from all the hidden threats of the typical facility environment. Even in today's manufacturing environments, power disruptions can cost businesses thousands of dollars in lost revenue not including the lost productivity of their workforce. Every business, no matter how small or large is at risk from internally or externally generated power abnormalities.

You may only notice power disturbances when the lights flicker or go out, but your compute, storage, network, and process equipment can be damaged by many other power anomalies that are invisible to the human eye, which can lead to degraded equipment performance or premature failure over time.

So you can worry now or worry later. One choice is proactive, the other potentially painful. Information technology (IT) systems are at risk even in the largest data centers. Of 450 Fortune 1,000 companies surveyed by Find FVP, each site suffered an average of nine IT failures each year. About 28% of these incidents were caused by power problems.

According to Price Waterhouse research, after a power outage disrupts IT systems:

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

- 33+% of companies take more than a day to recover.
- 10% of companies take more than a week to fully recover.
- It can take up to 48 hours to reconfigure a network.
- It can take days or weeks to reenter lost data.
- 90% of companies that experience a computer disaster and don't have a survival plan go out of business within 18 months.

Downtime is costly. Your IT hardware may be insured, but what about the potential loss of goodwill, reputation, and sales from downtime? Consider the number of transactions or processes handled per hour and multiply that by the value of each one and the duration of an anticipated power incident. Add the delays that inevitably occur when rebooting locked-up equipment, restoring damaged files, and rerunning processes that were interrupted. Then add the cost of lost revenue from being disconnected from your suppliers, business partners, and customers.

Could your business absorb the cost of an extended power outage or IT failure? According to the US Department of Energy, when a power failure disrupts IT systems:

- 33% of companies lose $20,000–$500,000.
- 20% lose $500,000–$2 million.
- 15% lose more than $2 million.

26.2 PRINCIPAL OF UPS AND APPLICATION

26.2.1 UPS Basics Principle

UPS designs get a base classification by the actual energy storage/delivery method used. There are two general classifications: static and rotary. The most popular design in the IT industry, the static UPS, uses some type of electronic switching components to either convert the incoming mains ac (alternating current) to dc (direct current) or take the stored energy (battery typical) during a power outage and converts that dc to ac at the correct voltage to be used by the downstream critical equipment. The rotary UPS uses a rotating device (generator) that is typically powered by the mains ac thru some type of motor system. The rotary generator, sometimes labeled DRUPS (Diesel Rotary UPS), typically uses a heavy weight flywheel assembly to store energy, which allows the generator section time to start and then provide power when ac mains power is lost. Static UPSs lend themselves to be suitable from very small to very large power requirements, where rotary systems are typically only found in very large power (kVA – kilo volt ampere) requirements

UPS systems come in different input and output voltages based on application and on the countries where they are deployed. In North America (the United States and Canada), single-phase UPS systems designed to plug into standard wall receptacles come in 120 V input. Systems deployed where the electrical contractor pulls a specialty receptacle can come in 120, 208, or 240 V. In some of the Caribbean islands and other countries where Europe or Asia influenced the electrical system infrastructure, the standard single-phase voltage is 220, 230, or 240 V. Mexico and other parts of Central America use 127 V for their standard single-phase distributed voltage. UPS for three-phase applications is typically manufactured for 208Y/120, 220Y/127, 480Y/277, and 600Y/347 V for North America and 380Y/220 V, 400Y/230Y, and 415Y/240 V (Fig. 26.1) for the rest of the world. However, some North American data centers are now deploying a 415 or 400 V electrical infrastructure to operate the IT loads close to their maximum value (typical <1,500 W IT power supply voltage range is 90–264 V), driving up the efficiency of the power supplies. In addition, the entire data center benefits as this eliminates the needs for 480 or 600 V to 208 V transformers (Fig. 26.2), typically gaining another 1–3% of efficiency. Since almost all IT applications require an input voltage less than 250 V (maximum rating of power cord), systems with higher output voltages use transformers to reduce the voltage to an acceptable level. There are a few latest-generation IT power supplies that can handle 277 V, which may become more standard as data center power use has become a large expense, and there is tremendous focus on more efficient systems.

26.2.1.1 UPS Components and Subsystems

UPSs include a number of different individual subsystems based on the system type. This section will cover most of them with some basic information on their function in the system.

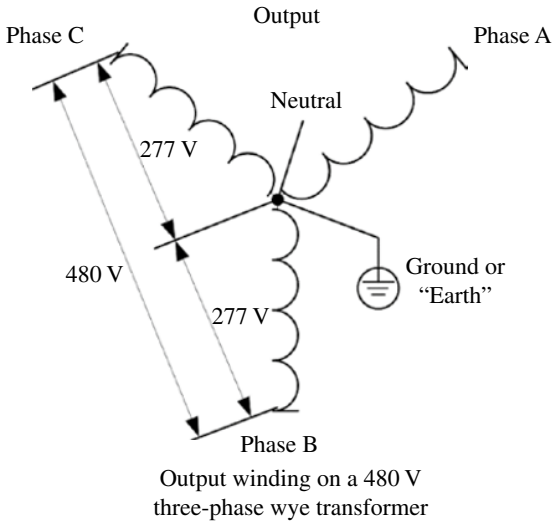


FIGURE 26.1 Electrical representation for the output of a typical three-phase wye transformer. Source: © 2021, Eaton

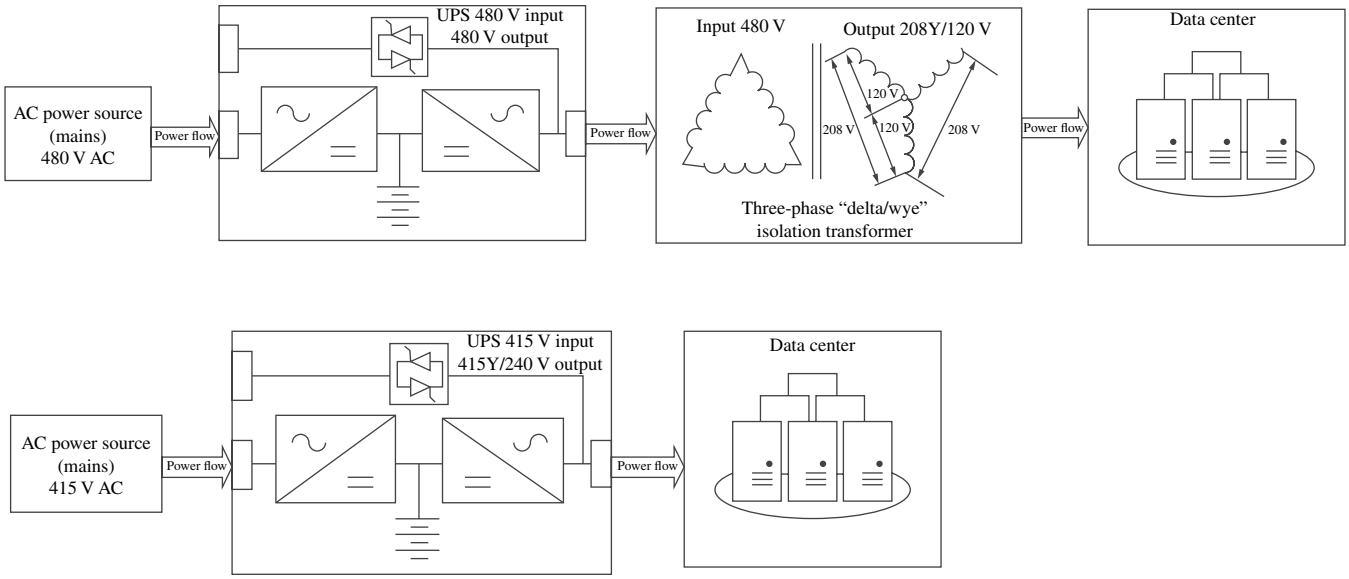


FIGURE 26.2 480V UPS system feeds transformer then to IT equipment" compared to "415V UPS feeds IT equipment directly. Source: © 2021, Eaton

Inverter

All static UPS systems include an inverter that uses the dc or backup energy source and creates an ac waveform for the connected load equipment. In the last few years, the industry has typically referred to the inverter as the output power converter. Inverter designs vary greatly based on the type of system typically based on criticality of the system and its cost. Small low-cost systems may use power transistors or MOSFETs, which typically output a basic square wave or modified sine wave. Care must be taken when applying these lower-cost inverter designs, as the non-sinusoidal output may cause a negative interaction between the UPS and protected equipment's power supply, which could result in an inoperable system. Higher-cost systems typically use devices called IGBTs (insulated gate bipolar transistors) that are used with an inverter output filter to create an almost perfect sine wave output. These inverters typically switch the IGBTs on and off thousands of times per second, in a sequence called pulse width modulation (PWM) (Fig. 26.3). As seen in the figure, the pulses at the beginning and end of each half cycle have very short "on" times and longer "off" times, with the on time increasing in duration to the peak of the sine wave, then again decreasing as the waveform decreases. The longer the device stays on, the more energy delivered to the filter network that is used to create the sine wave output. Advances in high-power IGBTs allow switching frequencies of inverters typically <50 kVA in size to be above the human audible range (18 kHz), therefore reducing UPS operational noise.

Most PWM inverter designs released prior to 2015 were known as two-level designs, (Fig. 26.4), as energy efficiency was not as important as system design reliability. These systems typically operated with efficiency levels of 91–94%, which was related to output voltage and the voltage ratings of the IGBTs used in the design.

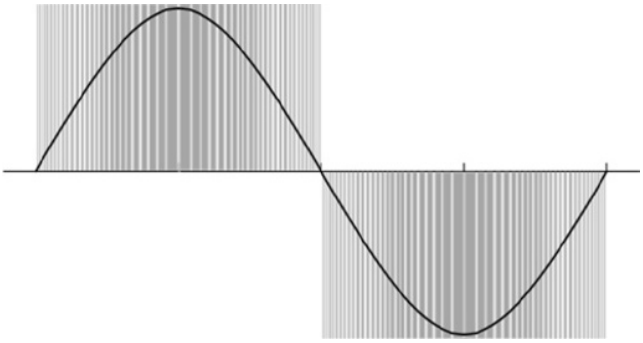


FIGURE 26.3 PWM switching pattern is filtered into clean sine wave ac output waveform. Source: © 2021, Eaton

Latest-generation double conversion UPS systems may use a three-level inverter design that has a noticeably different switching pattern (Fig. 26.5). Three-level designs typically double the number of IGBTs per phase to allow for lower-voltage-rated devices to be used in a series relationship.

There are two general types of three-level converters, the type "I" and the type "T" (see Figs. 26.6 and 26.7). The selection of I or T type is application driven, depending on the switching frequency and filter inductor size desired. Three-level inverter designs use two transistors in series, replacing a single transistor in the older traditional two-level design.

So, the conduction losses (Fig. 26.8) will be greater in the three-level converter, but the switching losses are dramatically lower, resulting in overall less losses for the three-level design. This helps to raise the efficiency of the inverter by typically one to 3% over typical two-level inverter designs, therefore reducing the overall cost of operating the UPS.

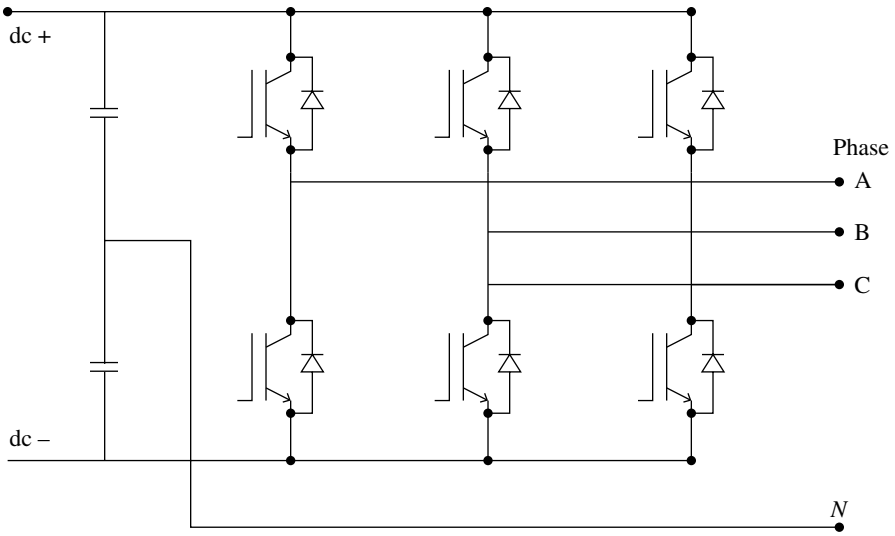


FIGURE 26.4 Three-phase IGBT-based two-level PWM inverter design. Source: © 2021, Eaton

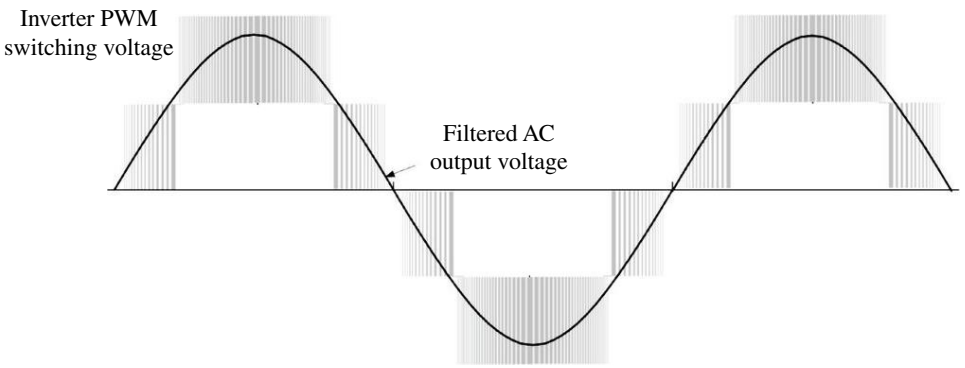


FIGURE 26.5 Three-level inverter PWM switching pattern. Source: © 2021, Eaton

The series IGBT arrangement results in a larger component count, but the electrical stress on each IGBT is much lower, allowing for reliability that is the same or better than a two-level design. Some manufacturers are even experimenting with four- or five-level designs; however, it is yet to be seen if the actual energy savings outweigh the cost of the extra silicon (Fig. 26.9).

Silicon carbide (SiC) technology and gallium nitrate (GaN) (Fig. 26.10) are the latest semiconductor technology being used in UPS rectifier and inverter power trains. SiC material features a wide bandgap that allows the device to switch on and off with less heat losses, directly improving the efficiency of the power converter. This results in a 1–1.7% improvement in UPS system efficiency while potentially allowing for smaller, less expensive semiconductors. The higher switching frequency also reduces the physical size of the filter inductors in the converter design, helping to offset the typically higher cost of the SiC when compared with traditional IGBT designs. Since this technology is a significant departure from the tried-and-true IGBT, it requires some differences in hardware and control designs. For example, the higher switching frequency presents challenges in filter design and EMI (electromagnetic induction) control. As these challenges are overcome, SiC is slowly beginning to appear in new UPS products.

GaN transistors are also a wide-bandgap device and the latest to move into the picture as suitable devices for power conversion applications. They have very high efficiency while also handling high frequencies, voltages, and temperatures. Their ability to do so, in a typically smaller die area, less than other power devices, lends well to the possibility of lower-cost power electronic devices. However, cost is always a function of mainstream manufacturing capability, and it will take time for it to be fully implemented in the automobile and UPS industries. Use of wide-bandgap devices could allow manufacturers to return to two-level inverter designs while still maintaining very high efficiencies.

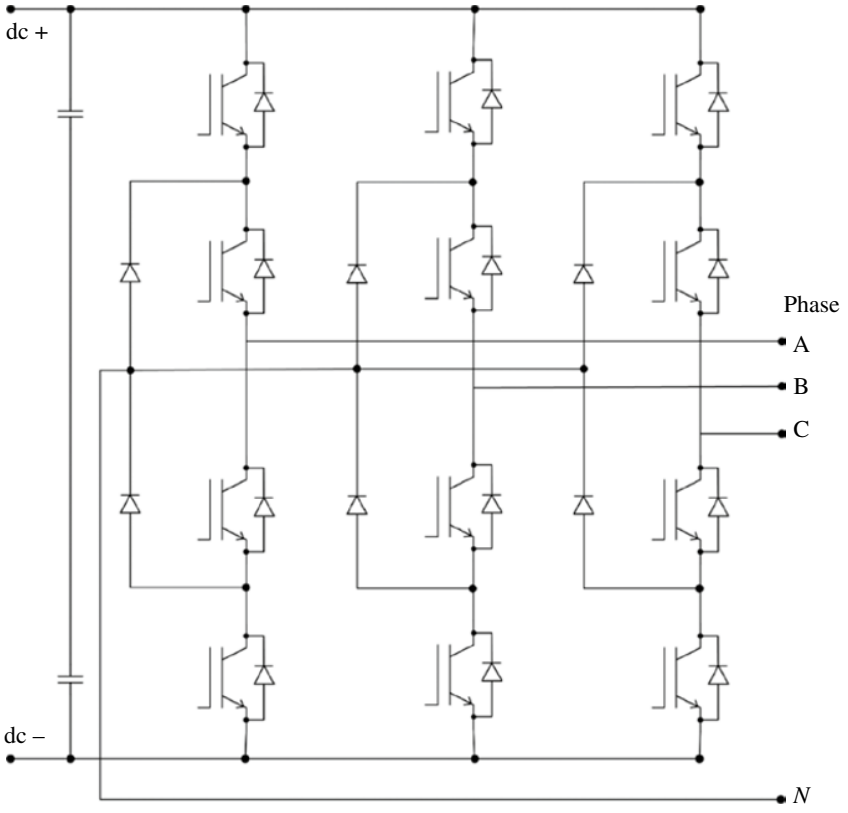


FIGURE 26.6 Three-level I-type NPC PWM inverter design. Source: © 2021, Eaton

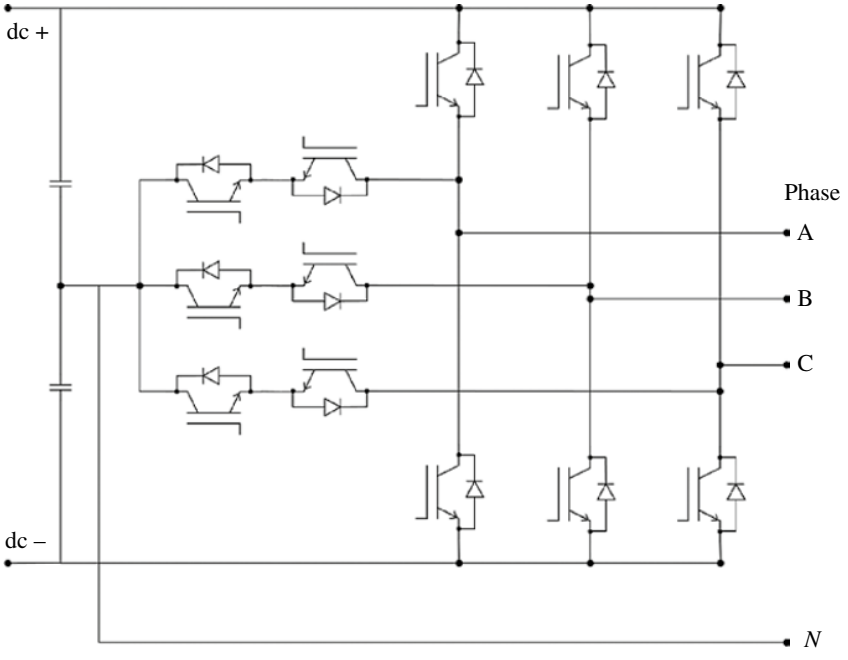


FIGURE 26.7 Three-level T-type NPC PWM inverter design. Source: © 2021, Eaton

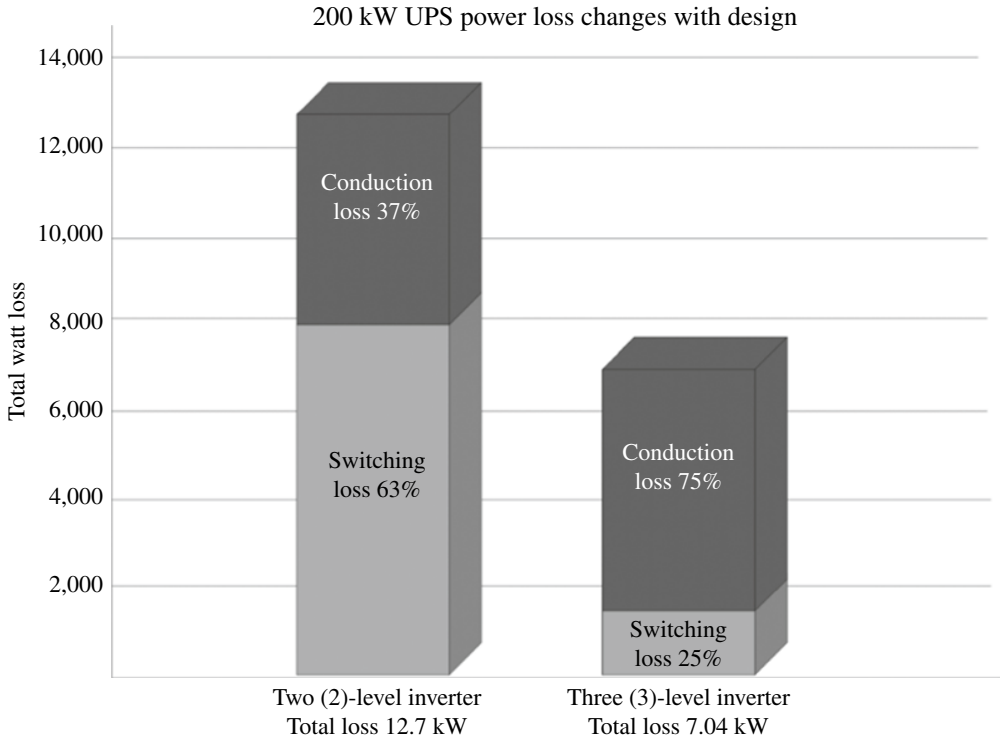


FIGURE 26.8 Comparison of conduction and switching losses between two-level and three-level designs. Source: © 2021, Eaton

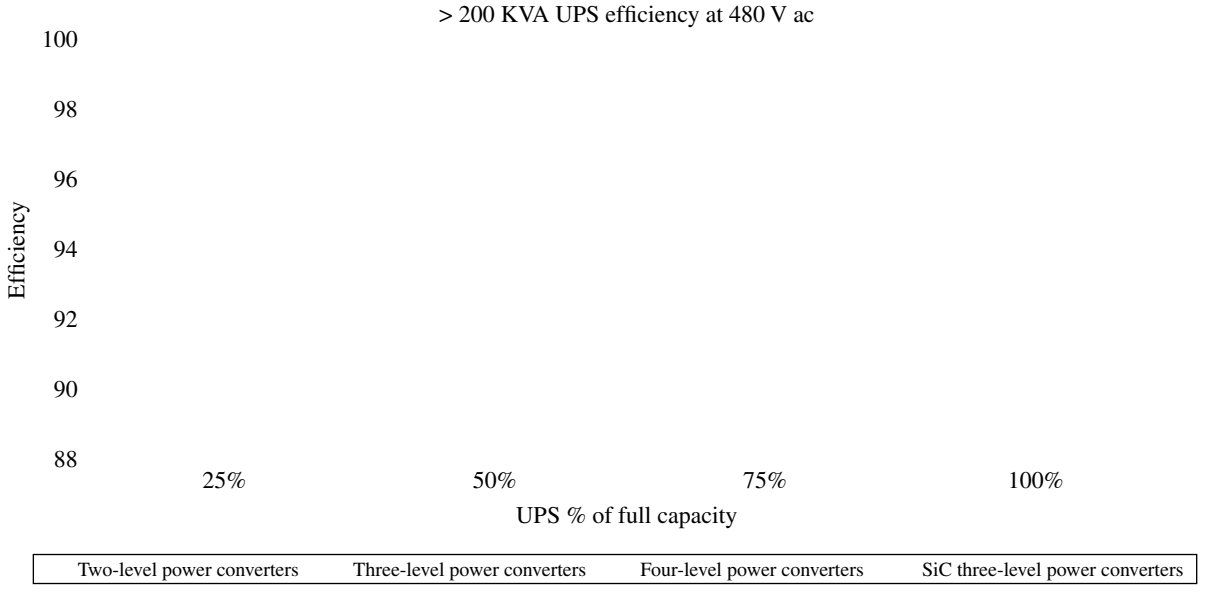


FIGURE 26.9 Efficiency comparison of two-, three-, and four-level designs, including SiC three-level design. Source: © 2021, Eaton

Rectifier

Double conversion and multimode UPS systems include a subsystem known as a rectifier (input power converter). The rectifier is the first device that sees the input voltage from the mains, and it converts the available ac power to a dc voltage for use to power the inverter and charge the batteries. Most UPSs use IGBT transistors to perform high-speed switching, similar to the PWM techniques used in the inverter described above; however it does it backwards, converting the ac to dc. This allows the UPS rectifier to be controlled such that its input current is sinusoidal, providing low input current distortion (iTHD), and its input power factor is near 0.99. Both traits allow for easy interface with an on-site generator, and they allow the optimization of the input filter to require

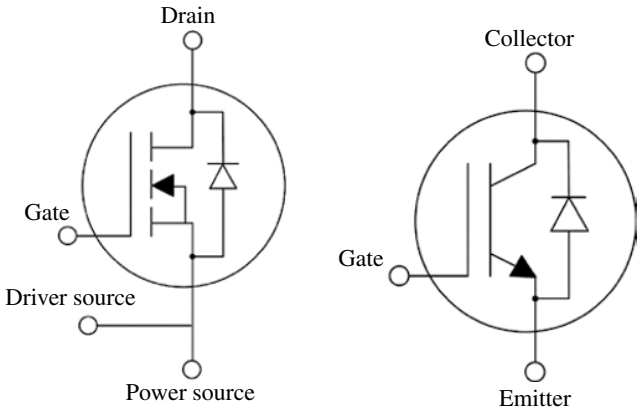


FIGURE 26.10 SiC MOSFET compared with silicon-based field-stop IGBT with a SiC Schottky freewheel diode. Source: © 2021, Eaton

the fewest capacitors possible, positively impacting reliability. The high-speed switching of the rectifier also reduces the size of the filter inductors, helping to reduce the UPS footprint and weight.

UPS Logic Control

Every UPS system has a main controller that takes in multiple inputs and adjusts or changes operational modes when required. All large UPS systems have eliminated most analog control functions and use digital control algorithms coded into the system firmware. The logic control may also contain external communication capability to talk to externally connected devices or may be responsible to send information to other communication-only controllers that are externally connected for purposes of monitoring and control.

dc-to-dc Converter

The purpose of the dc-to-dc converter is to provide proper battery charging voltage for systems with either a higher or lower dc link ("dc link" is the physical link between the rectifier output and the inverter input) voltage than that of the battery or backup energy source. Many of today's transformerless UPS designs rectify the incoming ac voltage to a level that is adventitious to ensuring the highest total system efficiency. This voltage level may not be a level that works well with the dc energy storage, so the dc-to-dc converter converts the dc link voltage to the levels needed to charge the battery, and when ac power is lost and the battery is called on, it converts the battery voltage to the levels needed for the inverter to provide the correct output.

Static Bypass Switch

The static bypass is typically found on most double conversion UPS systems, some higher-powered single conversion systems, and rotary systems. The switch's job is to supply a direct path for the mains to power the connected load equipment; however the mains voltage must be within a certain tolerance range, typically ±10%. There are usually several reasons the switch is used, first may be for maintenance of the system, activating the switch to ensure that the mains and UPS output voltages are synchronized so a maintenance bypass breaker can be closed. The second is to quickly transfer power to the mains source if for some reason there is a large output overload or internal fault in the UPS itself. Larger UPS systems include a separate machine control, power supply, and fans for the static bypass to raise the fault tolerance of the system. Another way the static switch is used is to be the main power path for high efficiency modes in multimode UPS systems. Static switches are built using silicon-controlled rectifiers (SCR), as they are much faster (1–10 ms) than large mechanical type contactors (50–100 ms) to close. However, some smaller single-phase UPS systems may use a mechanical contactor type bypass as some small contactors can transfer in 10 ms or less. The bypass must typically close within 10 ms or less to make sure the load equipment does not see a disruption in ac power. ac power loss longer than that could make the IT power supply fail, causing the load equipment to reset. Some UPS designs use both SCRs and a mechanical contactor in parallel, as the SCRs make the quick transition and the contactor closes providing a sustained bypass path. This type of static bypass is known as a "momentary static switch," and most times it is not rated to operate continuously, and if it had to operate under full load, it could fail. Latest-generation multimode systems must use fully rated SCRs as they need to be able to react by either turning on or off very quickly to make sure the load does not see any interruption in power, as mains voltages fluctuate or fail.

Maintenance Bypass

The maintenance bypass is installed to make sure the UPS can be taken off-line without interruption to the load and to safely preform service work including repairs, preventative maintenance, and upgrades. Maintenance bypasses can also be used to power the loads incase the UPS system is down due to a failure. There are several ways that electrical designs can be deployed to ensure proper service can be done on the UPS. The most basic and typical for a single UPS installation is with a stand-alone bypass cabinet, typically called a "wrap around" maintenance bypass switch. Some UPS systems do come with an "internal" maintenance bypass switch, but its functionality is limited as only part of the UPS is serviceable with only this switch installed. Smaller UPS designs (<50 kW) may use an internal IT rack-mounted or wall-mounted rotary switch. Larger systems typically use molded case switches or breakers, and very large UPSs use high-power rack out breakers in electrical switchgear. Large multi-module parallel systems will typically include many of the same devices as mentioned below, but in a highly custom-configured system. At a minimum the bypass for a

single UPS should include a maintenance wrap around switch and a UPS output disconnect. This would be typically known as a two-breaker or switch bypass. Some bypass switches include a third breaker as a UPS input breaker (three breaker), and when deploying a UPS with what is known as "dual feed" (static bypass and rectifier have separate input breakers), the switch becomes a four-breaker (Fig. 26.11) bypass. Adding UPS input breakers to the bypass is not a necessity, but it does help ensure a safer installation, as the source breaker is within sight of the UPS, and the UPS technician can more safely control the service environment.

Another option for maintenance bypass switches is a load bank breaker, or connection point, so the UPS can be tested under load before bringing it online with the critical loads. This can help eliminate a problem that may not manifest itself until load is applied to the UPS. However, caution must be exercised when designing the electrical infrastructure when using a load bank breaker to make sure its use does not overload an upstream breaker causing it to trip off-line and drop the still operating loads. Sometimes the maintenance bypass switch may be included in a system cabinet that could include a transformer and/or power distribution panels. This type of design does typically save space and cost, but you must take into consideration that an issue with one of the packaged components may require you to power down the entire cabinet. Note that the order of operation of the breakers in a maintenance bypass circuit is important, as breakers operated out of sequence can result in the loss of the critical load. Maintenance bypass panels are almost always equipped with, at a minimum, a printed operation procedure, located on the front of the panel. The operator should always refer to this procedure when preparing to transfer to or from the maintenance bypass. Frequently a "kirk key" interlock with a solenoid key release unit (skru) is installed on the maintenance bypass breakers (Fig. 26.12). This interlock is both electrical and mechanical and forces the proper order of operation in either direction, UPS bypass to maintenance bypass or maintenance bypass back to UPS bypass. The actual bypass breaker in the circuit should also have an interlock to ensure that the operator does not manually transfer the UPS inverter output directly to the maintenance bypass, as this paralleling of the UPS inverter and the mains source could cause an uncontrolled, damaging back feed condition.

FIGURE 26.12 Bypass breaker with KIRK® Key interlock installed. Source: © 2021, Eaton

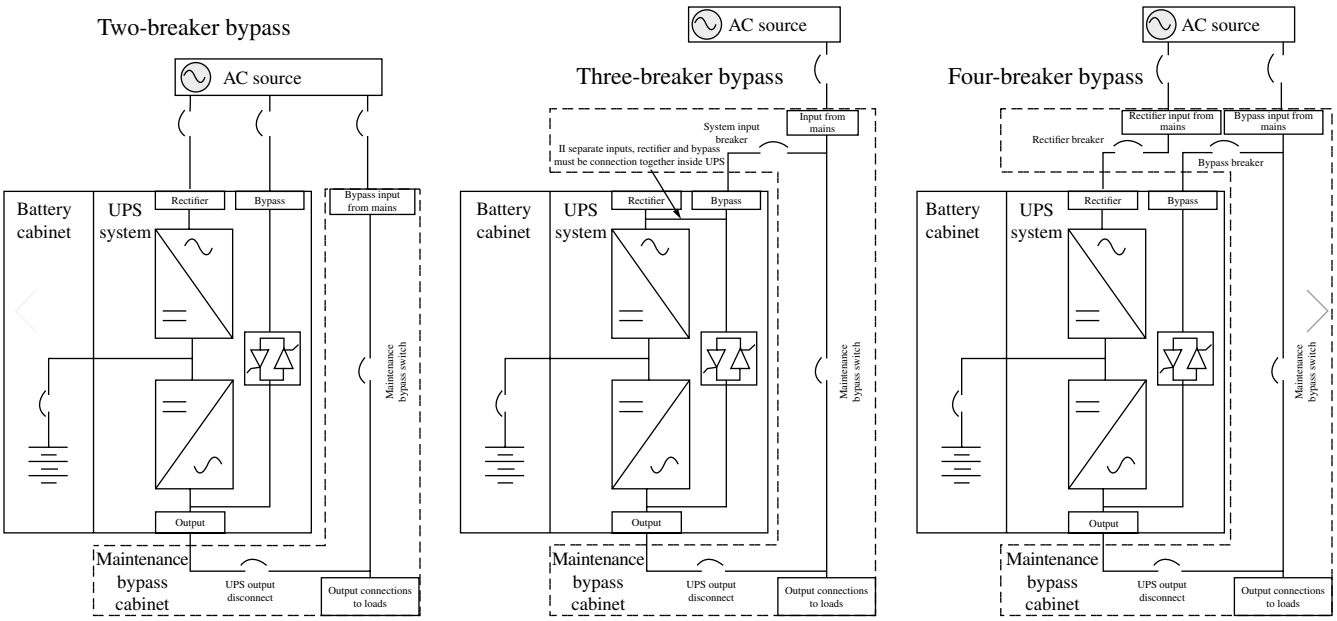


FIGURE 26.11 Two-, three-, and four-breaker schematic diagrams. Source: © 2021, Eaton

Backup Energy Source

The primary stored energy source for most static UPS systems deployed today is still the lead–acid battery. These batteries have been around for many decades providing a low-cost energy storage medium that is highly predictable. In addition, lead–acid batteries are one of most highly recycled components in production today, with the recycled lead and plastic used to build new batteries. There have been some advancements in lead–acid battery technology since the first UPS systems were deployed, and these include making the battery easier and safer to transport as well as maintain. In the past all the lead–acid batteries were a type of "wet or flooded cell" that required that the battery be shipped dry and filled with electrolyte (sulfuric acid) upon installation and then periodically monitored for electrolyte level when in use. Large UPS installations may still use this energy storage medium as they come in high capacity sizes in the thousands of amp hours. Since the lead and sulfuric acid are toxic materials, and as the battery is recharged hydrogen is released, the typical large battery installation includes a separate room, with separate backup powered ventilation fans, acid spill containment, and acid neutralizer. This special infrastructure and maintenance can be quite costly, so about 35 years ago a type of battery known as the VRLA (valve-regulated lead acid) was introduced. This battery was originally released in smaller amp hour ratings, but today many larger sizes are also available. This battery does not include the typical battery fill caps, as the electrolyte is in a gelled form and/or may be impregnated in a fiberglass mat between the batteries positive and negative plates (absorbed glass mat AGM). This change in electrolyte and way that it is contained in the battery earned the battery the notoriety of being nonspillable. This allowed the battery to be shipped full of electrolyte. Also, the battery design uses a pressure valve to utilize the properties of recombination of the emitted gas back into the electrolyte before it leaves the battery case, so a fully sealed battery with low maintenance was the outcome. These advancements in battery design did came with a couple shortfalls; first the design life is typically shorter than that of the wet cell designs, 5–10 years for VRLA with 20 years for flooded cells (some large VRLA batteries are available in 20-year design). VRLA batteries are also subject to higher cell failure rates if overcharged or operated in higher ambient temperatures, and they typically cannot be recharged as quickly as flooded cells. In recent years, "pure lead" VRLA batteries have come on the scene, and they offer some distinct improvements and capabilities over the traditional lead-calcium products. Firstly, pure lead batteries offer a wider operating and storage temperature range; they can be stored over a range from just above 0–50°C without significant degradation. The allowable storage time limits approach 2 years compared with 9–12 months for traditional VRLA. Pure lead batteries also offer a true 10-year operating life in UPS applications. All these improvements come at a cost point that is approaching parity with conventional VRLA, on a cost-per-kWh basis. The only disadvantage compared with traditional VRLA is a slightly larger physical size for a comparable capacity pure lead battery.

Lithium-ion batteries are quickly becoming a viable choice for users, in small to very large UPS sizes. Once considered too expensive for this application compared with VRLA, the lithium-ion battery costs have plummeted in recent years. They have a significant advantage in size and weight over traditional battery choices, and reduced installation space remains an important factor in purchasing decisions. More importantly, the vastly improved cycle count performance, i.e. the capability for more charge–discharge cycles, is what allows the lithium-ion battery to exhibit a 10- to 15-year service life in a UPS application. That's double the service life of traditional VRLA batteries and may allow the battery to last as long as the UPS itself. This longevity could eliminate costly and disruptive battery trade-outs, which normally are required every 5 years during the time a given UPS is in service. If this performance can be obtained at near parity with VRLA battery cost, there exists a clear cost-of-ownership advantage and reasonable ROI. If the user is considering a dual-purpose UPS application, where the UPS is used to interact with the grid, lithium ion is a much more viable energy source with the ability to sustain thousands of charge–discharge cycles during its life. Again, disadvantages may include a somewhat higher cost, the need for strict thermal control via a built-in battery management system (BMS), and special building and fire code restrictions around the installation of large-format lithium systems.

Other energy storage types are also briefly mentioned in later Section 26.5 however, most are only now starting to show some promise to even be thought about for use with UPS. Rising energy costs and internal company environmental directives are slowly cutting away on the lead–acid battery dominance in the market; however it will be a number of years before some of these other energy sources overtake lead–acid batteries for mainstream deployment.

26.2.2 General Classifications of UPS Systems

26.2.2.1 Static UPS Systems

The most basic static type UPS designs include a battery, some type of electronic switching devices to convert dc voltage at the battery to an ac voltage (inverter) that can be used by the connected IT equipment (ITE), and another electronic or electromechanical device (switch) that switches between the mains (power company) to the battery backup if the incoming power is lost or out of tolerance (Fig. 26.13). The basic UPS must also have a way to keep its battery charged and recharge the battery if used to supply power

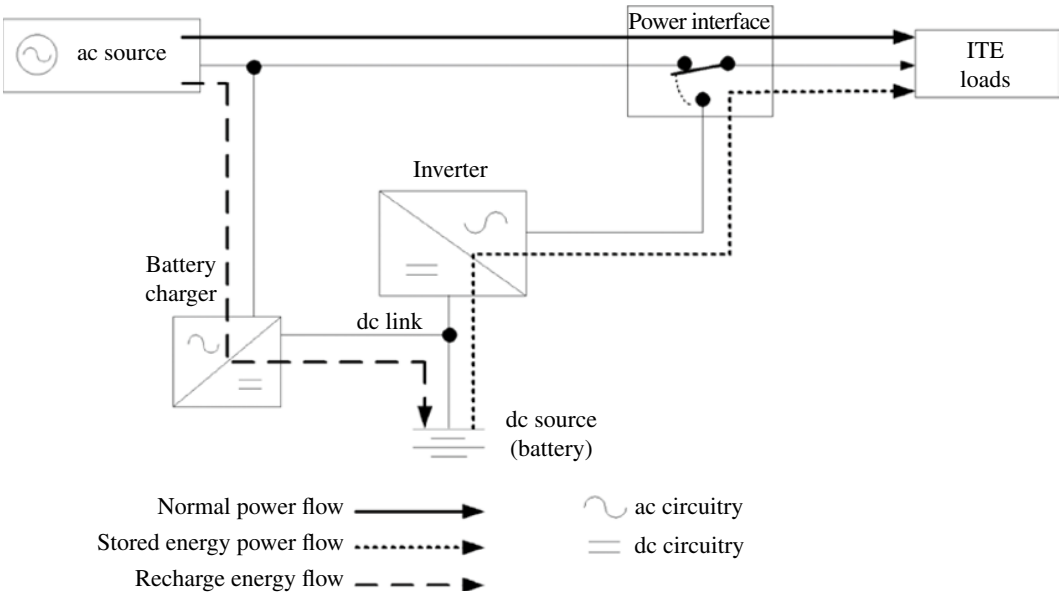


FIGURE 26.13 Simplified power flow diagram for typical small single-phase UPS. Source: © 2021, Eaton

during an outage, so it contains some type of battery charger. More complex static systems contain additional components such as automated bypassing devices (static switch) and devices that convert incoming ac to dc (rectifier or power converter).

26.2.2.2 Rotary UPS Systems

Since rotary systems use rotating mass for energy storage, they have more mechanical moving pieces than a static UPS. The spinning flywheel of a rotary system can typically only provide ac power for 5–15 seconds at full load, so a backup source such as a diesel engine may be coupled to the generator to provide longer runtimes (Fig. 26.14). Rotary UPSs are typically very large power designs that are not suited well for smaller data center power requirements, however quite well suited for large manufacturing or process control applications. Data centers wanting to move away from lead–acid batteries to an alternate stored energy medium such as the flywheel are however deploying variations of the rotary design. These deployments use a rotating flywheel for dc storage, supplying the UPS dc energy similar to what the battery does. When ac mains are lost, the flywheel generates the needed dc energy for the UPS, which is then converted to ac for the protected equipment. Many of the static UPS manufacturers have adapted their UPS designs to allow their systems to use a flywheel for energy; however, the short standard runtimes (15–30 seconds) are still a consideration when taking this path. In addition, the UPS rectifier, or bidirectional converter in the rotary systems, usually supplies the electrical power to again spin up the flywheel to the correct operating speed (up to 40,000 RPM on some designs). So, if you are looking to deploy an energy storage system using a rotating flywheel, you should first check with the manufacturers of the separate components to make sure they are compatible.

26.2.3 UPS Topologies

As mentioned earlier the static UPS comes in a number of different topologies; however, they fit into two different classifications per the IEEE, either single conversion or double conversion (Fig. 26.15).

26.2.3.1 Single Conversion

In a single conversion design, the incoming ac power is used to directly power the connected critical equipment; however there may be some voltage regulation done inside the UPS to either "buck" (lower) or "boost" (raise) the voltage to a level that is acceptable to the loads. Some single conversion UPS designs supply power to critical loads through a series inductor or a linear or ferroresonant transformer, therefore giving "line to load" isolation and transient protection. Single conversion UPS may also have a separate battery charging circuit or use the inverter to make sure the battery stays charged properly to support the critical load when called upon. The two most popular single conversion static UPS designs are the standby UPS and line-interactive (LI) UPS (Fig. 26.13).

Standby UPS

Standby UPS is most often found in small power applications such as desktop computer or home office or home

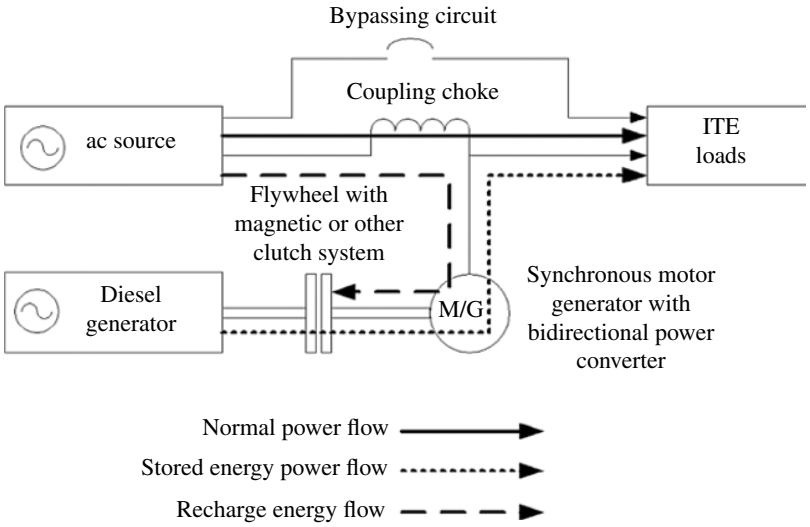


FIGURE 26.14 Rotary UPS system combined with diesel generator. Source: © 2021, Eaton

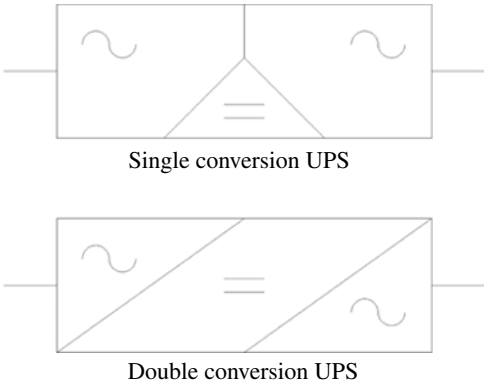


FIGURE 26.15 IEEE accepted drawings for different static UPS designs. Source: © 2021, Eaton

theater protection. For these applications, some transient protection and short battery backup are the main drivers, as well as low cost. Standby systems operate by passing mains power from input direct to the output, thru only a switching device, such as a static switch or small relay. When ac input voltage is within the tolerance level of the UPS output voltage specification (typically −15 to +10%), the UPS stays in this mode, as a very efficient pass thru device. If there is a power outage or large voltage swing, the UPS inverter turns on, and the internal switch switches to the backup source (battery) for power. Standby designs all wait until ac input voltage is lost; then they switch to battery operation, so there is always a "transfer" or switching time, where output voltage goes to 0 V. The amount of time is typically based on the system design and the fault condition; however, most manufacturers publish times ranging from 6 to 12 milliseconds (ms), which is usually fine for PC operation, but some slower transfer times could affect operation of larger servers and network equipment. Many low-cost standby UPSs have an output voltage on battery that may look almost square in nature rather than sinusoidal like the mains ac provides. This is due to the lack of output filtering circuit to reshape the output voltage to a sinusoidal shape. While this is okay for many simple PC power supplies, it can create issues with higher end server or network equipment power supplies, which react differently to the abnormal voltage waveforms. In addition, you will need a true RMS voltmeter if you are checking the output voltage reading of one of these "modified waveform" types of systems, as they will appear to have a large voltage drop when they switch to battery if using an averaging type meter.

Line-Interactive (LI) UPS

The LI UPS is found in many applications from home office to network closets to the factory floor backing up IT and even some processing equipment. LI systems come in a number of different basic power flow architectures; however, they all follow that same principle that they will regulate the incoming ac voltage to a certain output voltage specification, allowing a wider input voltage range than the standby system design. In doing this they are slightly less efficient than standby systems, but they offer the benefit of using the backup battery less, therefore extending operation during possible brownout or over-voltage conditions. The different ways that LI systems regulate voltage range from continuously operating the inverter in parallel with the mains to having a simple tap switching transformer that switches to a buck or boost mode, based on the incoming voltage. Systems that operate their inverter to regulate the voltage will be less efficient than the tap switching models; however they typically provide an output voltage that is more tightly regulated. Many LI designs also have a break in output voltage, while the inverter is switched on, ranging

from 4 to 10 ms, however systems that keep the inverter operating continuously should have no or a very short break in output power in most power failure conditions. Most LI and standby UPS designs build in some type of transient (surge) protection circuits by using MOVs (metal oxide varistors) or similar devices to clamp or shunt high voltage and frequency surges or spikes to ground, thereby protecting the connected loads.

Most ferroresonant UPS designs fall under the LI category, utilizing a ferroresonant transformer as the regulating device in normal conditions, with a separate winding on the same transformer, connected to the inverter, for operating during a power outage (on battery). A ferroresonant transformer differs from standard linear transformers as it has the capability to regulate voltage by using the properties of ferroresonance in combination with an output winding circuit known as the "tank" circuit. The tank circuit stores energy and provides enough ride thru during an outage to provide a no-break ac waveform. The input winding of a ferroresonant transformer is operated in saturation, with the tank circuit creating a resonant winding that creates the very constant sinusoidal output voltage. Ferroresonant UPS does suffer from typically large physical size and weight, with lower than typical efficiency, particularly at light loads. However, they are a very rugged design due to the use of very few active devices. These types of UPS are frequently used in poor power and environmental locations inherent in industrial, shipboard, and military applications.

Some rotary UPS designs may also use single conversion design traits to help raise efficiency, by not loading the rotating mass (generator) with the load continuously but bypassing this to a regulation circuit using in-line inductors. If ac power is lost, the generator takes over supplying power to the loads until they can be shut down or another source like an engine comes online to supply power to the system.

26.2.3.2 Double Conversion

A double conversion UPS (Fig. 26.16) differs as it has the capability to take the input ac voltage and rectify it to create a dc voltage; then that dc voltage is used to create a new ac voltage waveform; therefore it converts the energy twice. This newly regenerated output waveform is entirely and constantly controlled by the UPS inverter. However, it must be noted that even a double conversion UPS will typically track the utility's frequency to a certain point in order to keep itself in synchronization with the source frequency. This is important as if the UPS does need to make an emergency transfer to bypass, it can do so with a minimal break in power, ensuring the loads don't see the interruption. Many double conversion UPSs are programmed to operate up to 3 Hz (hertz) above or below the standard mains frequency to ensure this "lock" to the bypass source.

The double conversion UPS control logic regulates the voltage, frequency, and the waveshape of the UPS output at all times (except if on bypass due to overload or failure). These systems are sometimes referenced as rectifier–inverter systems (3). Double conversion is one of the oldest topologies, having been available for more than 50 years, typically used in highly critical or poor ac power quality environments. Since the system isolates the incoming mains ac, from the newly generated ac from the system inverter, they were always considered the ultimate in protection. However, that level of protection came with a price, higher cost, and lower operational efficiency.

Transformer-Based Double Conversion Designs

Some of the older UPS designs still on the market today use large transformers and inductors in their design to operate properly. These systems usually contain more components than the transformerless double conversion systems, as most systems include input and output transformers, a rectifier (ac to dc), battery charger (either the rectifier or a dc-to-dc converter), inverter

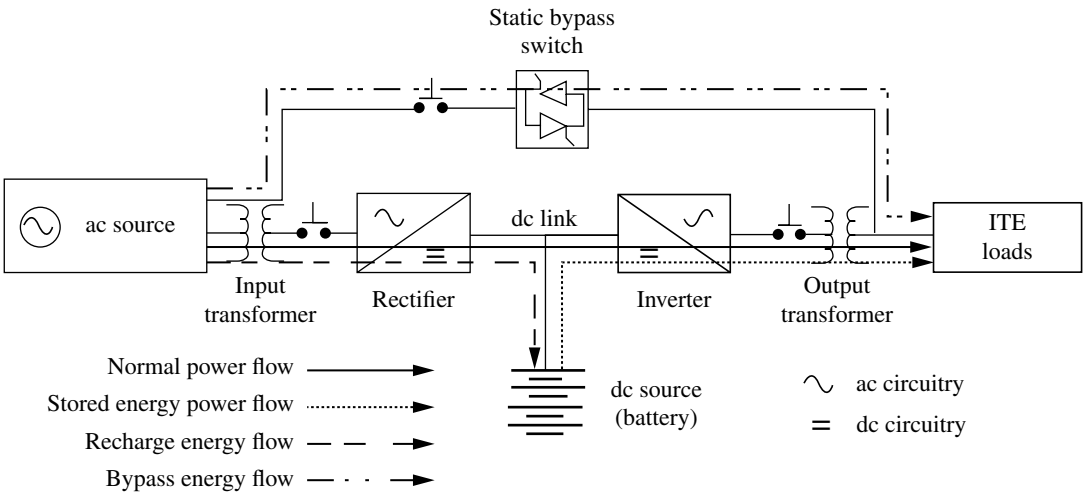


FIGURE 26.16 Typical transformer-based double conversion UPS system. Source: © 2021, Eaton

(dc to ac), and an emergency bypass mechanism, typically referred to as a static switch. The static switch is used to directly supply the loads with mains ac in case of a severe overload on the UPS output or a failure of the UPS internal systems.

As newer generations of transformer-based UPS systems were released, inverter IGBT's switching changed to start using PWM switching technology (Fig. 26.17). This was possible due to advances in IGBT designs, as higher power devices could operate at higher frequencies. PWM switching changed the requirements in the design of the inverter output filter, as filters became smaller, and it also allowed the UPS to change the inverter output voltage regulation faster in case of load changes on the UPS output.

Transformer-based systems have been phased out in most UPS designs, due to their detrimental size, weight, efficiency, and cost. The exception is industrial UPS applications, where galvanic isolation is typically requested and the available dc voltage of the energy source fares well for a transformer-type design.

Transformerless Systems

Starting in the 1990s manufacturers started offering higher-powered UPS systems using a transformer-free design (Fig. 26.18). Some of the benefits of these designs were better dynamic response, smaller physical size and weight, slightly better cost, and higher efficiency. Transformer-free designs were only possible due to the availability of high-speed digital controls, PWM inverters, and transistorized or "active" rectifiers that replaced the SCRs with IGBT transistors. These designs feature low iTHD (<4%), achieved without the need for the input transformer and the low frequency harmonic input filter, saving significantly on cost and complexity. Additionally, the faster switching rates and higher current ratings available with modern power IGBTs allow the UPS inverter and rectifier to respond instantly to transients and faults. This means that the "buffer impedance" provided by input and output transformers in older UPS designs is no longer required. The user benefits from lower weight, smaller footprint, and better operating efficiency.

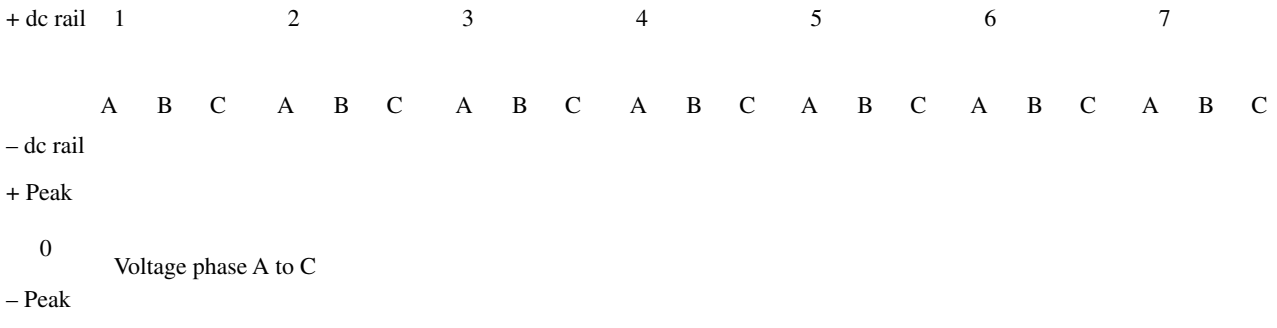


FIGURE 26.17 The inverter PWM switching sequence for a three-phase UPS. The sequence is for phase A to C. Source: © 2021, Eaton

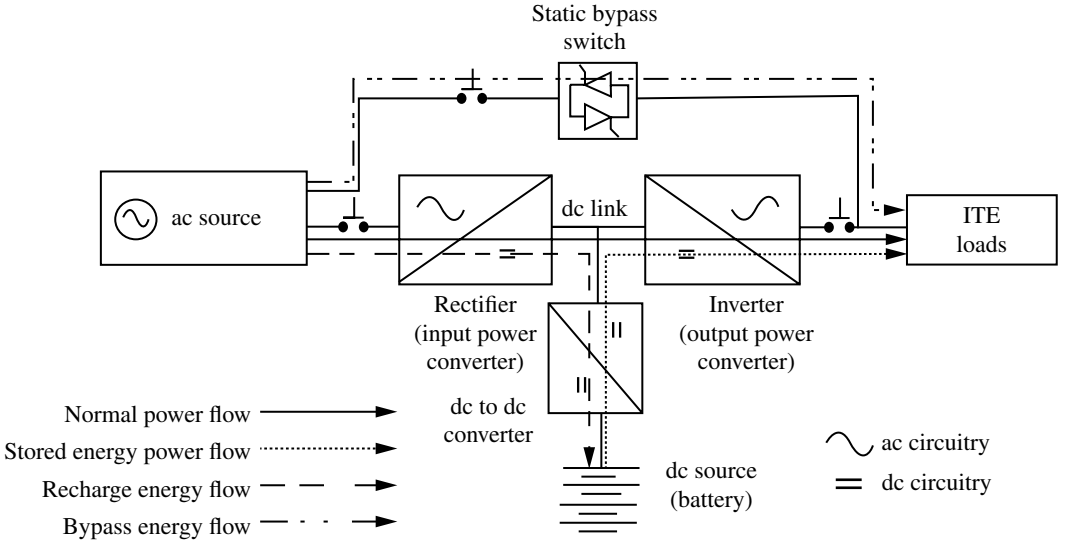


FIGURE 26.18 Basic transformerless double conversion UPS power flow drawing. Source: © 2021, Eaton

Multimode UPS

The latest generation of UPS, the multimode UPS (Fig. 26.19), came about due to the concerns with the high costs of electrical power and the need to raise the operating efficiency of the system while still providing highly reliable backup power. The multimode UPS uses operating modes found in the LI and standby UPS, as well as double conversion UPS. The reason the multimode UPS does this is to greatly improve operational efficiency while reducing wear and tear on components. In the normal mode of operation, when the mains is within the output voltage tolerance of the loads, the static switch remains closed, feeding the loads from the mains. When a voltage aberration outside of the acceptable voltage limit occurs, the UPS immediately opens the static switch and turns on the power devices to create a new ac waveform from the UPS dc source, either the rectifier if mains voltage is still present or the UPS battery. The most advanced of these systems returns more than 99% efficiency while giving transfer times of less than two milliseconds (2 ms) when mains power is lost. This arrangement continues to provide the highest possible efficiency of any UPS design, exceeding even state-of-the-art SiC systems, while ensuring excellent reliability due to its lack of dependence on complex semiconductor switching routines and moving devices like fans, all reducing the thermal stresses on power train components.

High efficiency modes on double conversion systems are not something new. In the past some double conversion systems included a high efficiency mode; however they were typically subject to several performance issues, so the feature was not used very often. The issues with high efficiency operation included inconsistent transfer times to battery if mains was lost, no filtering of high frequency transients, and requirement to operate the inverter to keep the output transformer energized so the transfer to battery could be made in time to support the load. Another issue occurred when the data center design called for large downstream static switches (see Section 26.4.5 for system designs) on the output of the UPS. If the static switch saw even short breaks in power, it would transfer the loads to the secondary power source every time mains power was lost to the upstream UPS.

Let's look at each of these issues and see how newer designs are handling this. Remember, the static bypass switch in the UPS, which is based on SCRs, requires that the current goes to zero before it can be switched off. If this device is on and the mains fails upstream of the UPS, the system would wait until the static switch shut off before turning on the inverter. If you didn't wait, all the loads upstream of the UPS (on the same mains feed) may appear as loads on the UPS inverter, causing a severe overload of the inverter until the static switch can be shut off. If that happens, the UPS output voltage momentarily dips so low the connected loads shut down. Some manufacturers have found ways to force the static switch off in microseconds, rather than the milliseconds that it took in the past. Leading UPS manufacturers are using predictive algorithms to predict what the incoming waveform should look like, and if it deviates from normal, it immediately forces off the static switch and changes the UPS to its double conversion or battery modes of operation to provide the highest protection. In addition, these leading UPS designs can determine if a fault condition is on the input or output of the UPS, which is extremely important when determining if the static switch should remain on or be forced off.

In the second case, transient surge protection of the ac is now possible due to the transformerless designs of the higher-powered UPS systems. Surge suppression is accomplished by

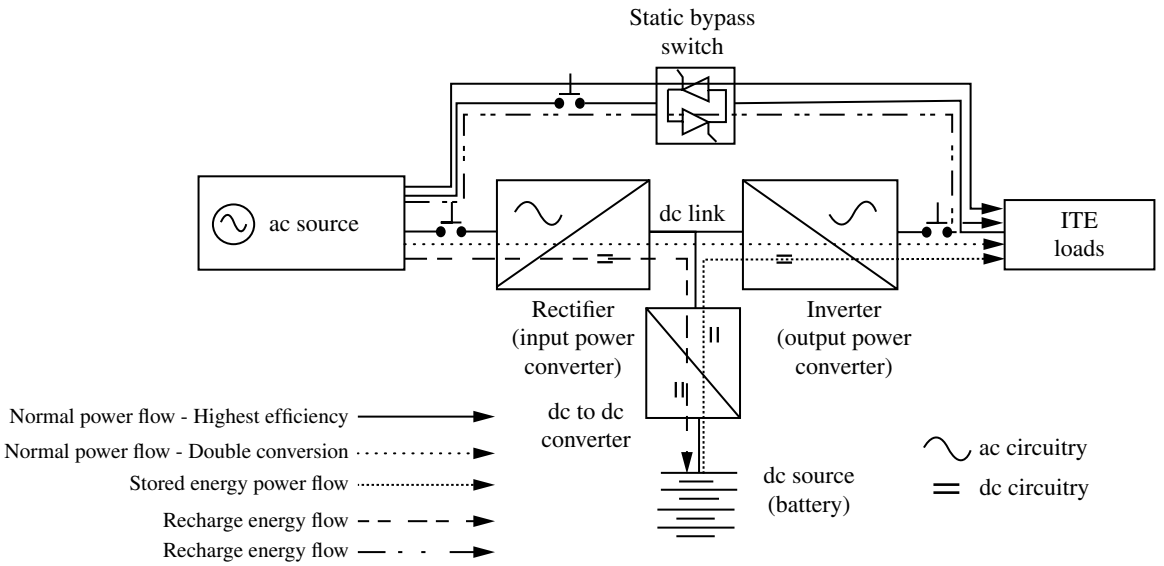


FIGURE 26.19 Multimode UPS uses multiple power paths for highest efficiency and protection. Source: © 2021, Eaton

keeping the rectifier and inverter high frequency filters attached to the mains source during high efficiency operation. The capacitors in the filters do the job of greatly reducing very fast risetime transients, such as lightning, from thousands for volts to just a few volts. Some UPS designs may not or cannot use the filters for this purpose, so you need to check the manufacturer's specification for transient suppression. If little or none is provided, install other surge protection devices in front of the UPS somewhere in your electrical system design.

The last issue with downstream static switches is still sometimes a concern but easily overcome with some timing reprograming of the static switch. However not all high efficiency UPSs are equal, and this is typically only accomplished if the UPS internal transfer times are below 2 ms. The reason is that most sites would like a maximum break in ac power of 4 ms from switching between two different ac sources. If one of the sources, the UPS, takes more than 4 ms to make a transfer itself, the static switch will have already moved the load to another source. Many static switch designs are programmed from the factory to switch if they see as little as a 1 ms break; however they can be reprogrammed to allow a 2 ms or longer break before switching sources. If you are using downstream static switches, make sure the UPS can make consistent transfers between high efficiency and other modes in 2 ms or less. This will typically ensure a trouble-free operation between the different systems.

The highest efficiency multimode UPS designs were not possible without the advancements in the industry of transformer-free large UPS system designs, high frequency sensing and control systems, and advanced predictive control algorithms used in latest-generation designs.

Yet another form of multimode UPS is the "dual purpose" also known as "Energy-Aware" UPS. A multimode dual-purpose UPS is controlled in such a way that its power conversion circuitry and battery storage system can be used at any time to perform tasks such as peak shaving (Fig. 26.20) and power demand management/optimization or even be used to act as a distributed energy resource (DER) with the mains power network. These UPSs, typically deployed with high cycle capability lithium-ion batteries, introduce a significant shift in the purpose and potential benefit of the user's power protection investment.

It is a well-accepted fact that UPS systems provide a necessary and irreplaceable benefit to the mission-critical user. Without it they are vulnerable to power anomalies, and the cost of that resulting downtime far exceeds the CapEx and OpEx costs of a UPS and battery. But those costs are still high; the UPS, battery, and generator systems in a data center spend more than 99.99% of the time simply sitting idle and "waiting" for a power outage or aberration to occur. Of course, when that outage does happen, that UPS is absolutely indispensable. But the rest of the time, these systems use some energy, generate unwanted heat, incur maintenance and repair costs, require battery replacement periodically, and sometimes most importantly, take up valuable floor space that could be used to generate revenue in a hyperscale or multi-tenant data center.

Financial and operations experts are asking if that large investment in power protection equipment could be used (directed) to provide a constant benefit, when not being used to back up the data center during those infrequent power outages. That is, "Can this UPS become a constantly working asset in my facility, rather than simply an expensive necessity that is rarely used?" A UPS deployed as a DER functions with other local DERs to assist the grid operator in balancing loads and managing peak power requirements. Many utilities have existing programs that pay the user to act as a DER on the grid. In normal operation, the UPS can be set to draw slowly from its battery during times of peak demand and then recharge that battery during low demand periods, typically late at night. This allows the user to take advantage of "time-of-use" electricity rates or avoid peak demand charges, which save money on power costs in day-to-day operation (Figs. 26.21 and 26.22)

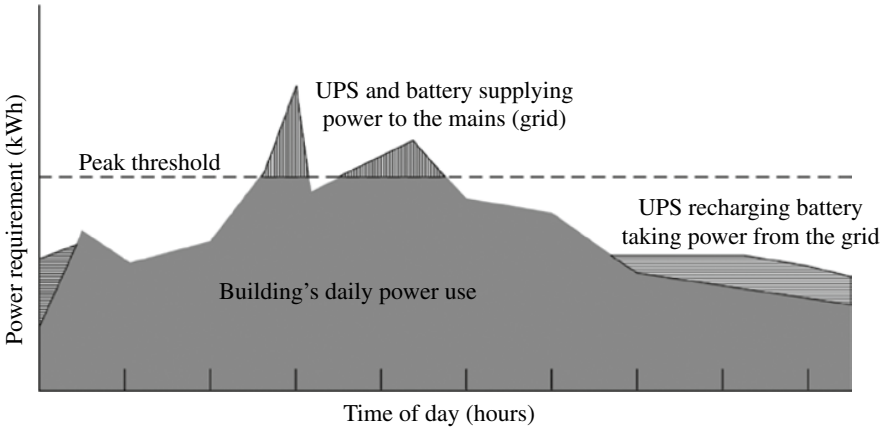


FIGURE 26.20 Peak shaving allows a facility to use the UPS system to avoid costly peak billing charges. Source: © 2021, Eaton

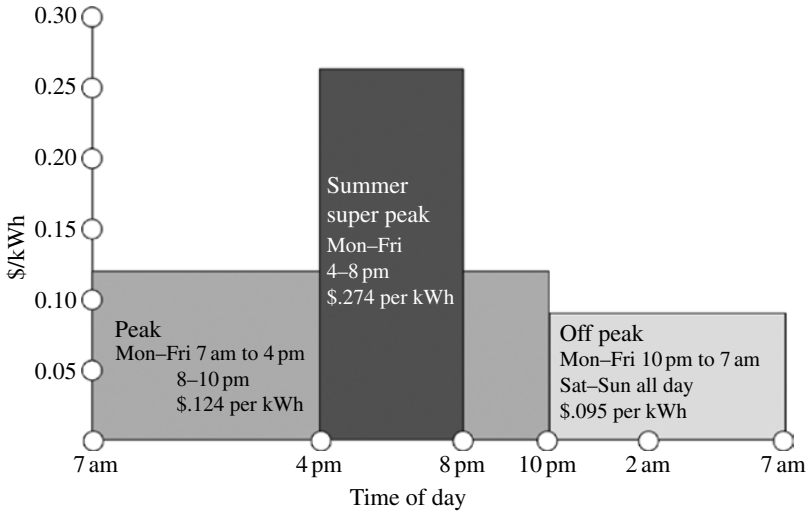


FIGURE 26.21 Typical time-of-day electrical costs from the electric utility. Source: © 2021, Eaton

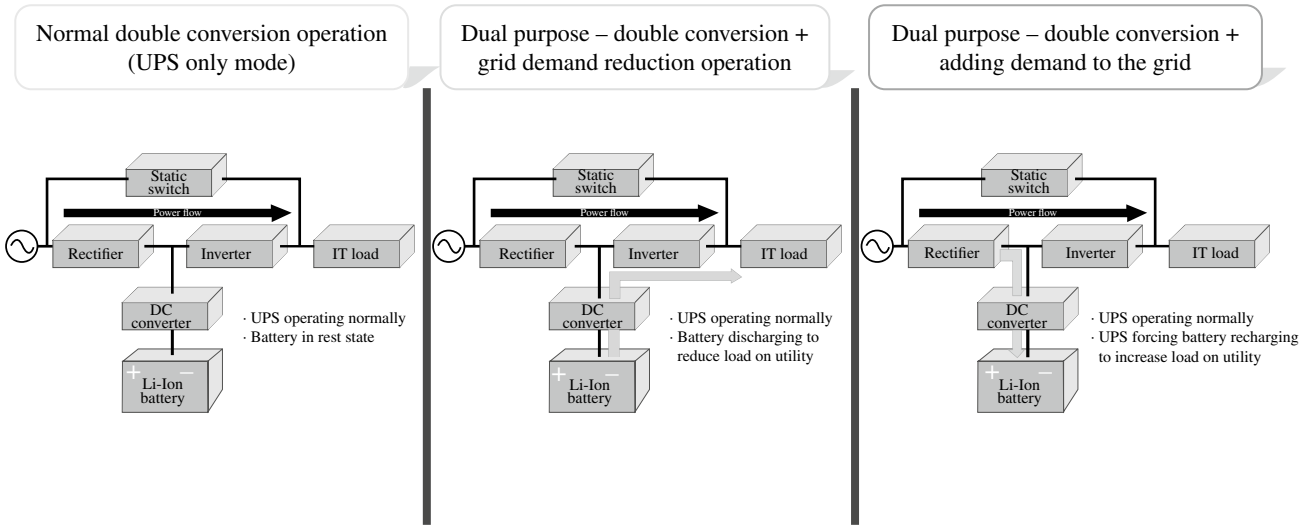


FIGURE 26.22 Three modes of operation for an energy-aware UPS system, utilized to manage demand or perform peak shaving for the facility. Multiple UPSs can operate together (aggregated) in a multi-building or campus environment. Source: © 2021, Eaton

Over time, these savings can be significant and thus reduce the ROI for their UPS investment. These cost-saving modes of operation would be used only when that UPS has spare battery capacity, beyond what is required for support of the data center. In this kind of application, the battery must be able to handle a large number of charge–discharge cycles, and this had not been viable for traditional UPS battery technologies. The advent of high-cycle lithium-ion batteries for UPS has made this application possible.

26.3 CONSIDERATIONS IN SELECTING UPS

As with any significant purchase, the designer and the user must evaluate the benefits provided by a product against the costs. UPS systems are similar to other critical data center infrastructure items, like power distribution and HVAC systems, in that a balance must be struck between required performance and reasonable cost. The following are key components of a typical evaluation of UPS characteristics.

26.3.1 UPS Response Time

The ability of a UPS to correct power anomalies by responding quickly to regulate its output is the subject of much variation and debate. The lesser cost topologies like standby systems described above may have a "switching time" from mains fed to battery fed of as little as 8 ms to as much as 20 ms. Larger systems may have a response time of 2 ms, to as much as 16 ms (one power cycle at 60 Hz), and this

response time must be evaluated against the tolerance of the IT load devices and possibly against the reaction time of downstream static switches in an A/B bus or a distributed redundant architecture. It is important to note that double conversion and some LI systems are designed with an inverter that operates constantly and can transition to and from battery operation with no interruption in output. These UPS systems also typically utilize a "make-before-break" overlapping transition when transferring the critical load to/from inverter to mains bypass, again allowing no break in output voltage continuity.

How fast is fast enough? In general, as prescribed by the ITIC/CBEMA and IEC guidelines (Figs. 26.23 and 26.24), IT equipment will not operate reliably if their input power is interrupted for more than 20 ms, and some devices can fail with only a 10 ms outage. Thus, it is a goal of every UPS design to provide response times that are as brief as technically possible and ideally make transitions with no loss of output or zero response time. For a large fraction of the mega data center market, 0- to a 3-ms response time is preferred.

26.3.2 Efficiency

UPS efficiency is one of the most aggressively advertised and competitively debated features of any UPS. This is true in part because the user will readily appreciate that higher efficiency saves money. These savings include both a reduction in the cost to power the UPS and a reduction in the cost to cool the UPS environment. In larger UPS systems these savings can be quite significant, and a UPS that is a few percent more efficient can often justify a higher initial cost versus a cheaper alternative or even pay for itself when evaluated against an existing legacy UPS system. Additionally, the more efficient product will provide tangible benefits to the user as they make their case for a more environmentally friendly data center, which is less of a drain on local community resources.

For modern UPS products a typical efficiency at full load is 95–97%. The best double conversion products may provide up to 98% with multimode UPS systems reaching 99+% efficiency. These high efficiencies are seen over a much wider load range compared with older products, where efficiency dropped significantly at loads under 50%. Note

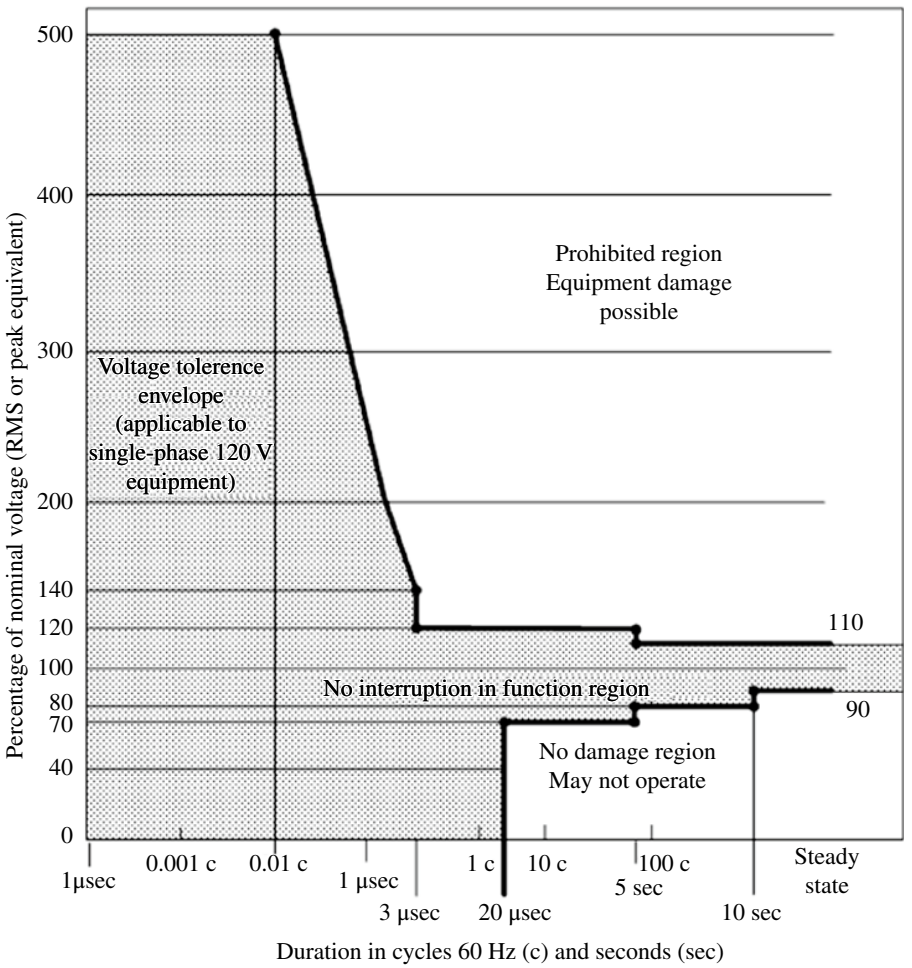


FIGURE 26.23 ITIC/CBEMA voltage envelope for guidance in IT power supply design. Source: © 2021, Eaton

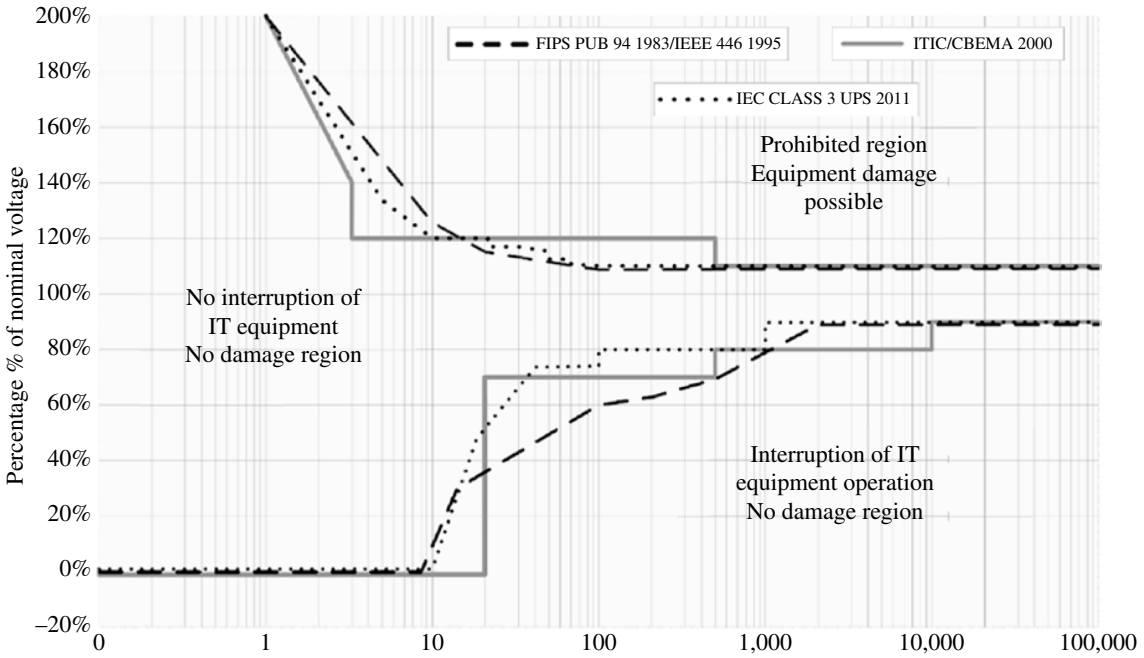


FIGURE 26.24 A compilation of multiple voltage envelopes, including the more universal IEC 62040-3 curve, that are generally agreed upon by IT equipment manufacturers. Source: © 2021, Eaton

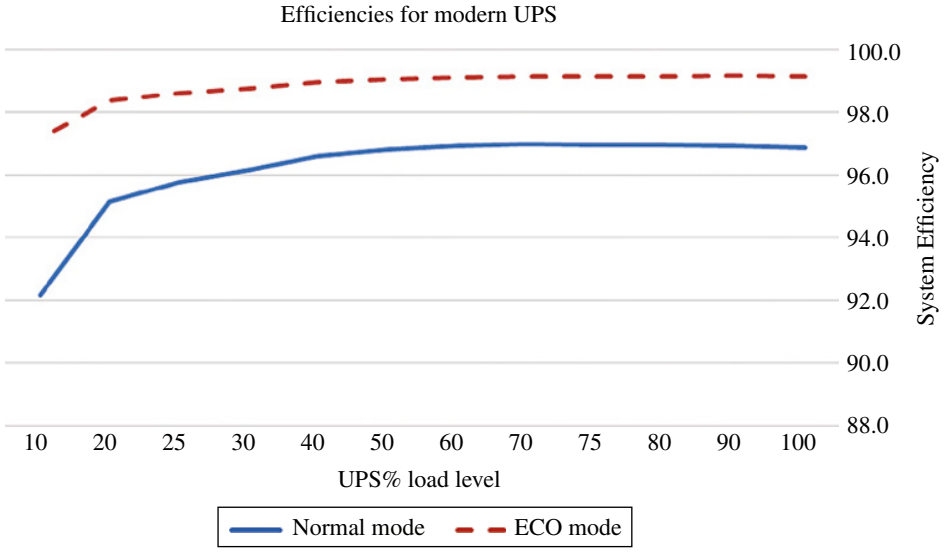


FIGURE 26.25 Different UPS topologies have efficiencies that vary with load levels. Source: © 2021, Eaton

that many if not most UPSs are loaded at less than 50%, especially in highly redundant facilities, where two separate power systems are provided in complete redundancy. The multimode UPSs excel at maintaining high efficiency at loads as low as 15–20% (Fig. 26.25), with modern three-level double conversion UPS maintaining efficiencies close to 96% at loads as low as 25%. Some modular UPS vendors offer "load optimization" systems that improve light load efficiency by automatically and instantly flexing their available capacity in response to output load changes. This is done by monitoring load levels and controlling the number of UPS modules in play at any moment while suspending modules that are not required to support the sensed load. This optimizes the load levels in the UPS so that every module continually operates at its highest efficiency, regardless of total system loading.

Keep in mind that an older UPS that has been in service for the past 10–15 years may be 5–10% less efficient than a

new product. ROI (return on investment) calculations are easily done comparing legacy performance against multi-mode performance, with some evaluations showing the entire cost of a new UPS system may be recouped in 2–3 years, making the choice to upgrade very attractive.

26.3.3 Environmental and Safety

While it is tempting to select the UPS based on technical performance alone, users have a responsibility to evaluate its environmental performance and its impact on the safety of their employees and service personnel.

26.3.3.1 Sustainability

Users that emphasize "sustainability" as a key component of their enterprise will want to evaluate the UPS based on its use of sustainable materials and sustainable manufacturing practices. A "cradle-to-grave" life cycle analysis may be requested as a part of a LEED compliance process, for example. One key component is the environmental cost of production, which includes the cost to procure and ship raw materials and the power cost required to construct and test these large electrical systems. There are guidelines and even legal requirements enforcing the mandate to avoid certain hazardous materials in the UPS, and its internal or external battery, and still more laws regarding disposal of the UPS at the end of its service life.

In certain countries in Europe and Asia, there are updated sustainability requirements for UPS manufacturers to meet to allow them to sell products into those countries. One of the biggest requirements is to meet the IEC 62040 directives that include being RoHS (restriction of hazardous materials in electrical and electronic equipment) compliant. Products compliant with this directive do not exceed the allowable amounts of the following restricted materials: lead, mercury, cadmium, hexavalent chromium, polybrominated biphenyls (PBB), and polybrominated diphenyl ethers (PBDE), Bis(2-ethylhexyl) phthalate (DEHP) Butyl benzyl phthalate (BBP) Dibutyl phthalate (DBP) Diisobutyl phthalate (DIBP) with a few limited exemptions allowed.

The battery associated with the UPS is often scrutinized as a hazardous material and subject to special rules for servicing and disposal. In addition, the sheer weights involved will require careful handling during installation, service, and maintenance activities. VRLA or "sealed" batteries still contain lead and sulfuric acid but, because they are "sealed," may be considered less of a hazard than the larger, heavier, flooded electrolyte batteries. Flooded batteries usually require a specialized room with provisions for seismic bracing, containment of spilled electrolyte, and ventilation of hydrogen gas.

Lithium-ion batteries are now being used with UPSs of every size, from rack mount UPS to multi-megawatt parallel systems. These lithium-ion batteries of course contain different internal materials and are less likely to be recycled than lead-based batteries. This is countered by the fact that they have a longer service life of 10 years vs. 5 years for VRLA. Additionally, at the end of their useful life in UPS service, lithium-ion batteries can often be "repurposed" in other applications like energy storage or emergency lighting backup and continue to provide useful capacity for an additional 5 years or more. Since these batteries contain no toxic materials, disposal is simplified due to fewer legal requirements and limits. In the end, however, it is expected that the vast majority of lithium-ion batteries will eventually be recycled, and their internal and external component materials reused. Whichever battery chemistry is used, operational safety procedures and disposal requirements are often legally mandated and strictly enforced.

26.3.3.2 Serviceability

While smaller UPS systems may feature safe user serviceability via "field replaceable units" or FRUs, the bulk of larger systems are not user serviceable beyond the external air filters covering the air intakes. Even so, the user should be vigilant and careful whenever one is in close proximity to the UPS or the power distribution system, keeping in mind that a significant amount of stored energy is contained within the UPS, even when mains power is absent. Due to concerns about arc flash, the aptly named "dead front" covers internal to the UPS should never be removed by the untrained user. When considering the UPS from a safety perspective, the selection of the UPS should include verification of UL listing or local international safety certifications, along with the use of a certified, experienced installation contractor that will observe proper wiring and grounding requirements as per local codes and the national (NEC) or international electrical codes.

26.3.4 Cost

As with most purchases, better performance and quality internal components associate directly with higher cost. But that's not the whole story; there is also the need to consider total cost of ownership, or TCO, for these systems. TCO includes the up-front cost to procure, manufacture, factory test, and ship the system, along with the following:

- Cost to perform the electrical installation and testing.
- Floor space cost per year of operation.
- Power cost to operate over the UPS service life (efficiency).
- Cooling power cost (also affected by UPS efficiency).
- Cost to maintain and repair the UPS over the service life.
- Cost to maintain the system battery and planned cost to replace the batteries.

  - VRLA batteries should be replaced every 5–6 years during the life of the UPS.
  - Lithium-ion batteries are expected to last 10–15 years.
- Cost to replace consumable parts like capacitors and fans that are not intended to last the entire life of the UPS.
- Disposal cost for the UPS and batteries at the end of its service life.

Knowledgeable users and their design/consulting partners are factoring TCO into their purchasing decisions. Some will collect cost and maintenance data from the vendor and calculate the TCO on their own. Most vendors offer online tools that provide a way for a potential owner to quickly calculate and compare the TCO of various UPS topologies, eco-mode capabilities, and power and cooling costs. They can even compare TCO of multiple competitors' products or compare an existing legacy UPS with the cost savings provided by a new, more efficient UPS.

Then there is the significant cost if the user chooses a redundant UPS system or a 2*N* or "dual-bus" architecture. These systems add the "extra" redundant UPS and importantly the extra battery system for that UPS. This may double many of the costs listed above, affecting footprint, testing, maintenance, power cost, and cooling capacity cost. Some modern UPS systems feature internal or inherent redundancy, due to their modular and scalable construction. In many cases, this can allow the user to have the reliability benefits of an $N + X$ redundant system, without the traditional penalty in capital cost expenses and increased footprint.

26.4 RELIABILITY AND REDUNDANCY

Historically, mean time between failures (MTBF) has been a key metric that UPS manufacturers use to measure and express reliability. In truth, however, MTBF is generally a poor means of predicting UPS availability.

To understand why, consider a UPS with an MTBF of 200,000 hours. A layperson might expect such a device to experience one failure in 200,000 hours – or 23 years – of operation. In reality, UPS manufacturers can't and don't test their products for 23 years. Instead, they calculate an initial MTBF based on the projected lifespan of the UPS's components. Then, after they've shipped a statistically meaningful number of units, they may replace those preliminary estimates with new ones based on actual performance in the field. Those revised numbers can be misleading though. For example, if 2,500 UPSs perform flawlessly over a 5-year study period, the result will be an impressively high MTBF rating. But if those systems contain a component with a 6-year lifespan, 90% of them could fail in the next year following the study period.

In addition, there is no universal standard for measuring MTBF. For years, most government agencies have required manufacturers to provide calculations based on the latest revision of the MIL-HDBK-217F handbook, while many commercial customers have adopted the Telcordia (Bellcore) SR-332 process. However, like many differing standard bodies, these two standards will give markedly different results. More recently, the technology industry has concluded that these measurements, while helpful, should not be the only way manufacturers grade a product's reliability. As a result, manufacturers today increasingly focus on design for reliability (DFR) as well. Unlike past standards that concentrate on individual electronic components and their relationship to the circuits used in the product's design, DFR methodologies pay greater attention to a product's intended and expected use under varying conditions.

Still, at the end of the day, there remains no one standard for measuring how a UPS performs its mission, which is keeping connected loads powered. As a result, it's nearly impossible to compare one UPS manufacturer's MTBF numbers to another's. Availability offers a somewhat more realistic measure of critical power backup systems. Given the vital role that UPSs play in the data center, the ability to replace aging or failed parts rapidly is crucial. Availability combines MTBF with a second metric called mean time to repair (MTTR) that measures the time required to acknowledge a problem, respond to it, and complete a repair.

$$\text{Availability} = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}}$$

Availability is typically expressed as a number of "nines" representing the percentage of time over a year's worth of use that a given system is operational. For example, a UPS with an MTBF of 500,000 hours and an MTTR of 4 hours would have an availability of 0.999,992, or 99.999,2% (500,000/500,004) that translates to an expected downtime of 4.2 min/year.

Still, though it's a better gauge of reliability than MTBF numbers alone, availability is flawed in important respects. In particular, it fails to account for time spent on routine service functions. If a system has to be taken down once per year for inspection, recalibration, or general maintenance, its actual operational availability will be lower than the formula above suggests. Therefore, the MTTR number should also include the time off-line to service the equipment throughout the year to get a better gage on total system availability.

26.4.1 Strategies for Increasing the Availability of UPS Power Paths

One of the most common ways to increase availability is to increase the number of power paths for ac power to be

delivered to the connected load. This can be internal to the UPS or could consist of paralleling multiple UPS systems. When internal to the UPS, there may be multiple independent subsystems that can support themselves, or it may be separate replaceable or repairable modules that can be quickly serviced. Most data center grade UPS systems include an automated bypass, the static switch, that gives the UPS two internal paths (Fig. 26.26) for getting power to the load. As the diagram shows a path through, a maintenance bypass is also typically specified for most data center applications.

Some latest-generation UPSs have been designed with multiple modular power components that can be fully isolated from the live power bus and repaired or tested (Fig. 26.27). These types of designs will typically have the main power components, the rectifier and inverter, replicated so they can give some internal redundancy in case of a failure of either of these components. However, these systems still have some common components that may require the entire load be bypassed or shut down in case of a failure of one of these components.

When evaluating the UPS reliability, a simple diagram is typically used to help understand the relationship of the components to the entire system reliability. For this we will consider that any one of the three redundant components in Fig. 26.28 can be shut down and the two remaining devices can support the output's requirement. In the upper portion of the figure, the subsystems in series (A, C, D) from the input to the output are considered a failure point that will jeopardize total system reliability. Subsystem B is redundant, and one module could be replaced, while the other systems support the load. The lower diagram in Figure 26.28 shows a typical parallel redundant configuration using three separate modules, each containing all subsystems needed to operate on its own. So a failure in any one of the single systems (1, 2, or 3) will not affect the entire system reliability.

One thing that must be considered when paralleling multiple systems or subsystems in redundancy is how many is too many. As you can imagine, the more components that you add to the system, the more components there are to fail, so you may end up in a situation where you have diminishing returns. This can be avoided by selecting the proper building block up front and limiting the number of parallel systems. It is recommended that you try to start with power blocks where you can handle the full load capacity with four to six parallel systems, therefore reducing the number of components being deployed. However, there could be a situation that your application is so large that you are using the largest block available, therefore increasing the number of modules needed and increasing the possibility of replacing failed parts in the future.

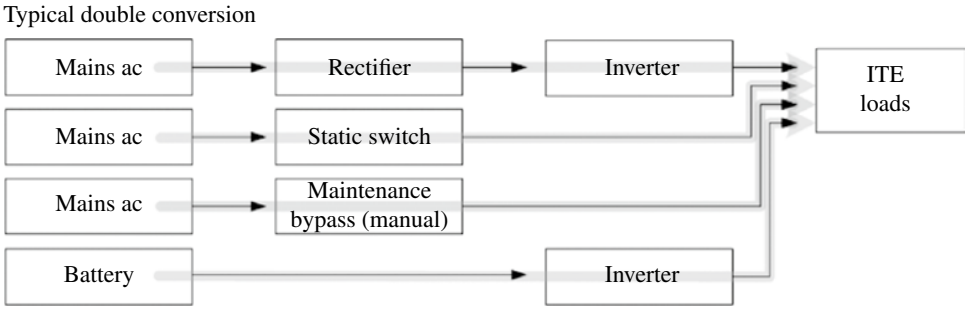


FIGURE 26.26 Multiple power delivery paths to get power to the connected loads. Source: © 2021, Eaton

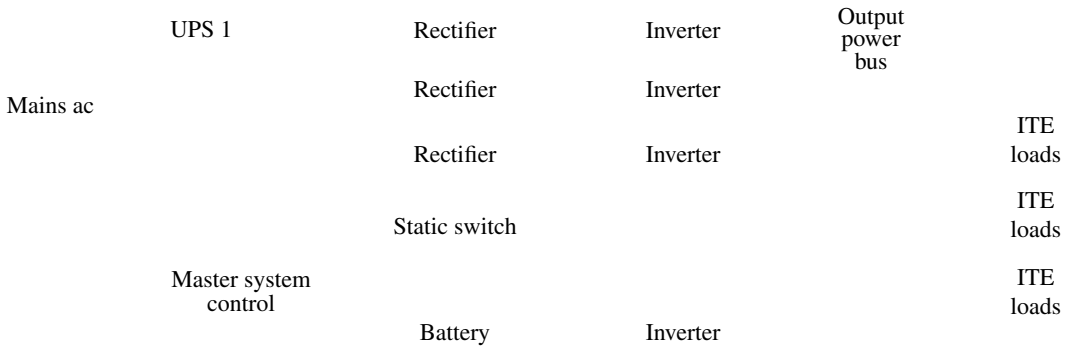


FIGURE 26.27 Single UPS system with some internal redundancy provided for rectifier and inverter. Source: © 2021, Eaton

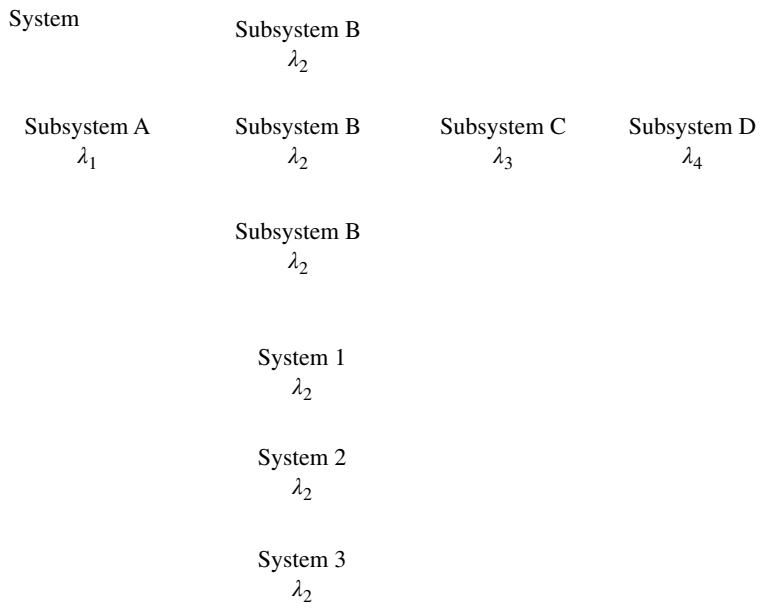


FIGURE 26.28 Upper drawing shows subsystem redundancy, while lower drawing shows parallel redundancy. Source: © 2021, Eaton

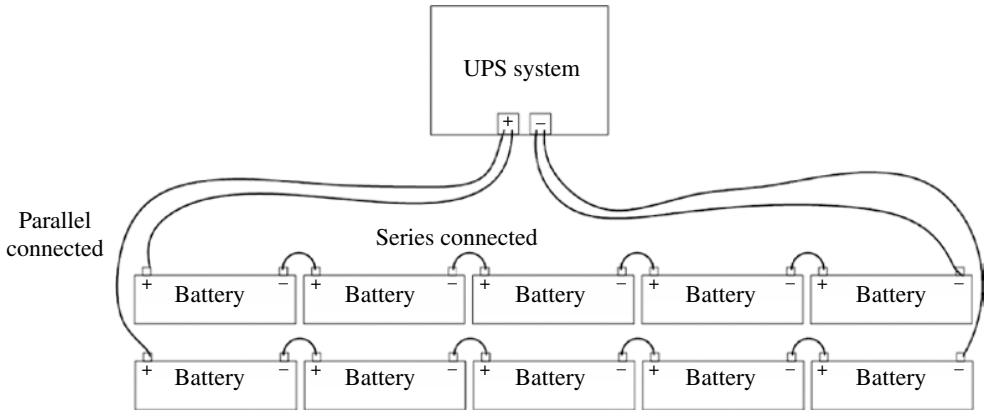


FIGURE 26.29 A way to increase UPS backup energy (dc) reliability is to parallel multiple battery strings. Source: © 2021, Eaton

Statistics have shown that the most common failure point in any static UPS system is the battery. Even with most new UPS designs completing multiple battery tests on a monthly or more frequent schedule, a single internal or battery connection problem could cause a failure. A simple way to increase the battery reliability is to add a parallel battery string. Equipping a UPS with a single string of series-connected batteries can dramatically increase risk of load loss. Say, for example, that a large UPS has 40 batteries connected in a series (+ of the first battery to − of the next and so on). If a problem occurs in any of those batteries, the entire string will probably fail, causing the UPS itself to fail. Adding another 40 batteries and then tying the most positive and most negative points together give you two parallel strings of batteries (Fig. 26.29). If either string fails, the UPS can typically run for a limited time on the other string until either a backup generator comes online or load equipment is shut down gracefully.

Increasingly, organizations are finding that the risk of running off straight mains power—even briefly—is too great to ignore. So they deploy redundant UPS modules to ensure conditioned power even if one UPS module fails. In *paralleling*, two or more UPSs are electrically and possibly mechanically connected to form a unified system with one output—either for extra capacity or redundancy. In an $N+1$ redundant configuration, you would have at least one more UPS module than needed to support the load. As a conjoined system, each UPS stands ready to take over the load from another UPS whenever necessary, without disrupting protected loads. Let's take a closer look at parallel UPS architectures—how they work, what challenges must be overcome in establishing parallel configurations, how modern paralleling technology enhances availability, and what difference it makes in your power protection scheme. Redundant UPS configurations were once relatively rare. Organizations

balked at the expense of buying two UPSs to do the work of one. Only the most substantial organizations—or those with the most critical power requirements—made the investment. That has changed. Data center managers and facilities managers have concluded that running off raw mains power, even briefly, represents unacceptable risk. The cost of downtime is now so high that even small data centers can justify the cost of redundant UPSs. In fact, redundancy is a requirement of data centers that attempts to meet reliability levels defined by some industry experts, such as the Uptime Institute.

The Uptime Institute requires a minimum of $N+1$ redundancy in the power systems as low as a Tier II redundancy level, with greater electrical system redundancies at higher levels. As a result, parallel UPS configurations are becoming commonplace. At least 50–60% of large UPS systems (300 kVA and up) are configured as parallel systems. Ten years ago, it was uncommon to parallel smaller systems (10–20 kVA in size), but now up to 40% of these smaller systems are paralleled—particularly in Europe and Asia.

26.4.2 How Do Parallel UPS Configurations Work?

On the surface of it, the concept of paralleling UPSs for redundancy is simple enough. Multiple UPS modules are linked to perform in unison (like one big UPS), sharing the critical load among them via a common output, with each module ready to take over for any other module if necessary. In an $N+1$ configuration (a typical redundancy arrangement), there would be enough spare capacity to support the load if any one module became unavailable. For example, you could protect a 500 kVA load by deploying two 500 kVA UPS systems or an 800 kVA load by deploying three 400 kVA UPS modules. During normal operation, the three 400 kVA modules would each carry one-third of the total 800 kVA load. If one module went off-line, the remaining two modules would have sufficient capacity to support the load. Figure 26.30 shows a typical parallel configuration with two UPS modules. In normal operation, ac power flows from the mains source to each UPS. Each UPS has two inputs, or what is known as dual feed, where one input goes into the rectifier and one into the internal bypass (static switch). The UPS converts incoming ac power to dc and then back to ac and then sends this clean power to a tie cabinet, where outputs from both UPSs are merged into a single output to protected loads.

Should a failure of any kind occur with either UPS module (Fig. 26.31), the critical load is still UPS protected. Internal diagnostics immediately isolate the faulty UPS module from the critical bus, while the other UPS assumes the full load, remaining in normal operation, not needing to activate the internal static switch to go into a bypass mode. When the UPSs installed in a parallel configuration retain their own internal static switches, the installation is said to have a "distributed bypass."

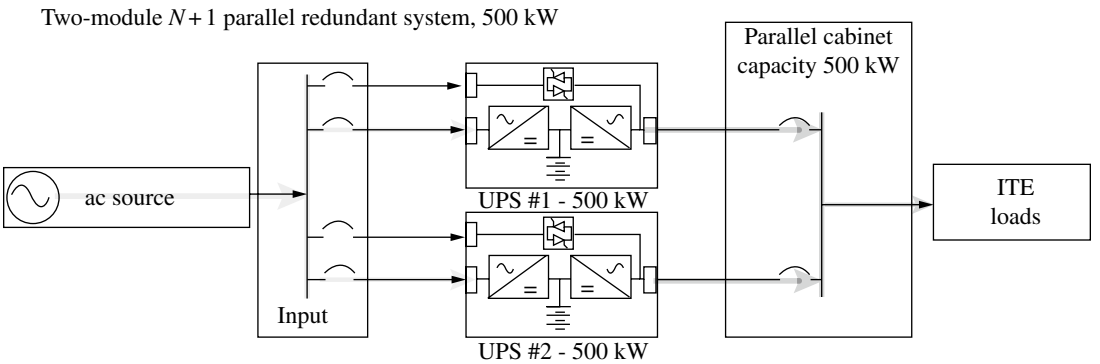


FIGURE 26.30 In normal parallel operation, both UPS modules contribute equally to shared output. Source: © 2021, Eaton

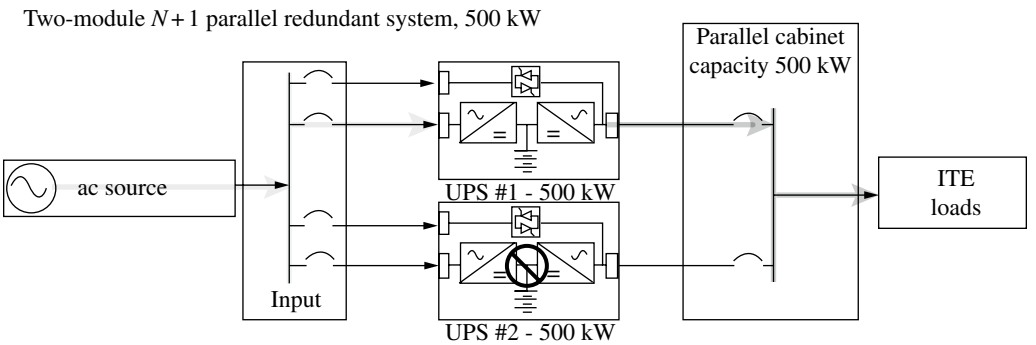


FIGURE 26.31 If either UPS module becomes unavailable, the remaining module assumes the load. Source: © 2021, Eaton

During a mains failure, each UPS module is supported by its battery system and can continue operating for minutes or hours, depending on how much battery runtime has been provisioned.

You can (and should) provision separate battery backup for each UPS, for even greater backup protection, and a higher level of redundancy; however sometimes that may not be economically feasible. A parallel UPS configuration is not limited to two UPS modules, as it frequently includes up to four modules, and some installations may contain eight modules or more. With the newest rack-mounted UPSs designed for high-density server environments, no freestanding tie cabinet is required. Paralleling is accomplished using a plug-and-play bus structure that mounts easily in the back of a standard IT rack, where the UPS modules also are installed.

The configuration shown in Figure 26.32 has a single bypass cabinet rather than the standard tie cabinet, which is known as a centralized bypass configuration (the static switch is "centralized" in an external cabinet). In the tie cabinet with a separate bypass is its own full system-rated static switch. This provides an alternate route for power during a failure, which is automatic and an instant wraparound bypass. Such an event would be rare, however may be activated during service or repair instances. The wraparound bypass would be activated only if the connected UPSs were unable to support the load in normal operation. Perhaps a short circuit caused an extraordinary overload that exceeded the capacity of all three modules together. The system would identify a failure on the critical bus and transfer to bypass mode with virtually no interruption.

Another alternative is known as a distributed bypass parallel system (Fig. 26.33). In this system, each UPS retains its own internal static switch, and they all operate in unison when a transition to or from the bypass source is required. In this type of system, when many UPSs are linked in parallel, the load they collectively support will exceed the capacity of the internal static switch and bypass circuit in any one UPS. So there is a need to ensure that all UPS modules equally share the load, even when powered by the internal static switches in each UPS. The more UPS modules that are tied together, the more important that the electrical cabling that connects the modules from the mains power to the output tie cabinet becomes. The individual

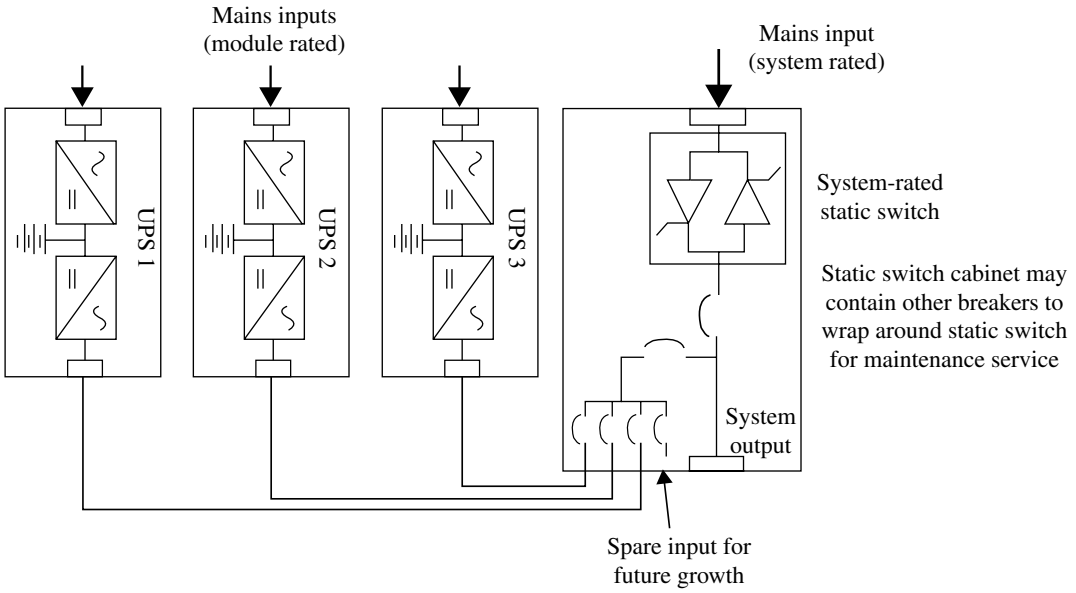


FIGURE 26.32 In a centralized bypass system, power flows to critical loads, even if all three UPS modules were off-line. Source: © 2021, Eaton

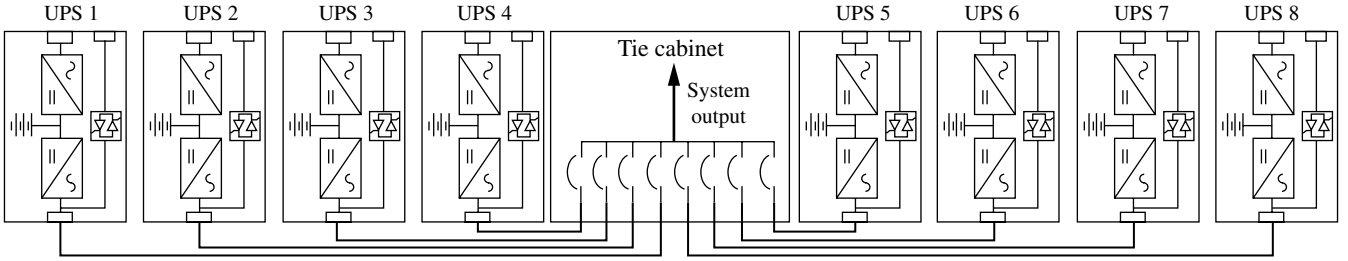


FIGURE 26.33 This distributed bypass configuration has eight UPS modules paralleled into a single system. Source: © 2021, Eaton

static switches in each UPS have no way of controlling load sharing, so it is only the electrical paths impedance that controls how much load each UPS carries when in bypass. If the total cabling route is very long to one system and very short to another, the system with the shortest route will have the lowest impedance, so it will assume more load than any of the other systems when in bypass. If it is too unbalanced that UPS may become overloaded and either take itself off-line due to overheating or a breaker may trip off-line, creating a cascading effect that could cause the rest of the systems to fail. Most manufacturers recommend only paralleling four or five systems in a distributed parallel architecture. Impedance matching of more systems than this can be done using in-line inductors; however any future wiring changes will require a rebalance of the impedances at the same time.

26.4.3 Four Key Challenges in Parallel UPS Systems

As soon as you connect multiple ac power sources into a unified parallel system, there are four key challenges to address:

- Controlling how the separate UPSs should cooperate as a unified system.
- Synchronizing the output of each UPS so it can flow into a shared output.
- Balancing the load equally among all UPSs in the parallel configuration.
- If trouble occurs, identifying and temporarily decommissioning the UPS with the problem.

These issues can be complex, and they must be managed in a way that does not compromise the high reliability for which UPSs are paralleled in the first place.

26.4.4 Parallel Systems for Added Capacity

Most organizations plan to grow, but when and how much? How much power will you consume next year, or in 5 years? You don't want to overbuild the power system today for future demands that may or may not materialize. Even if you could justify the cost, the power infrastructure would operate far below capacity and may be very inefficient as a result. And you certainly don't want to rip out and replace today's UPS just because next year's moves, adds, and changes suddenly double the need for power.

Paralleling provides an excellent solution for matching growth while extending the value of existing UPSs. The architecture to parallel for capacity looks very similar to paralleling for redundancy. Hardware components are the same; there are just small differences in operation. A system paralleled for capacity allows you to add load until it reaches capacity and then notifies you to add another module. In contrast, a redundant parallel system constantly ensures that there are enough modules to take over the total load if one drops off ($N+1$). For example, if the parallel system has five 100 kVA modules, the system would issue an alarm if the load exceeded 400 kVA—the load that four of those five modules could support.

26.4.5 Dual-Bus, or 2N Architecture for Dual Corded Data Center Equipment

In this arrangement, the UPS modules feed separate distribution panels that support separate power supplies (PSU) within every piece of IT equipment. UPS A supports one power path and one of the power supplies in the IT equipment, and UPS B supports the other power supply (Fig. 26.34).

This configuration offers a lot of flexibility, because the UPS modules do not have to be equivalent. They can be different sizes, carry vastly different loads, and even come from different manufacturers. But it only makes sense for a data center that *exclusively uses dual-corded IT equipment* (Fig. 26.35). Most of today's data centers still have legacy equipment that uses a single power supply, such as smaller networking and other communication gear.

How do you provide redundancy for those single-corded loads? Some sort of static switch arrangement would be required to switch single-corded loads from one UPS to the other, in the event of a failure of the primary UPS. In Figure 26.36, a static switch serves the single-corded equipment in the data center (Fig. 26.37). Alternately, you could use a small relay-based dual-source transfer switch mounted in the rack to feed any single-corded equipment in that rack. Whatever type of dual-source switch is deployed, the switch would transfer the load from a failed power source (UPS in this case) to the available UPS in milliseconds, without disrupting the protected load.

This arrangement adds complexity to the power distribution architecture. The more components in the power delivery chain, the more points to monitor, maintain, and troubleshoot, and the more possible points of failure. However, an even more troubling issue is synchronization. If the UPSs are not in sync with each other, the rapid switch of power from one to another via the static switch could introduce a voltage transient that could shut down or cause operational issues with those single-corded loads.

So now we have a situation where, even though the UPSs are not providing all the benefits of paralleling, their outputs still must be synchronized. This can be accomplished with an external power synchronization control (PSC) unit, which sets up a master synchronization arrangement (Fig. 26.38). Now the availability of those single-corded systems rests on the reliability of the static switch and synchronization controller. For that reason, this arrangement is best used as a stopgap measure as single-corded loads are phased out in favor of dual-corded devices.

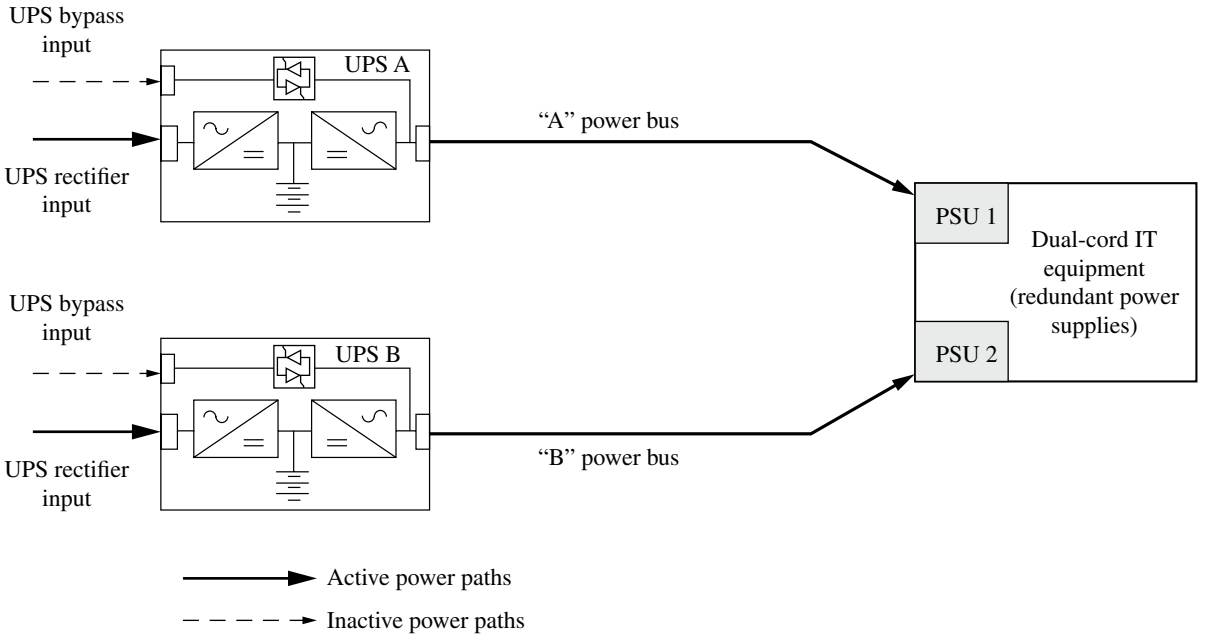


FIGURE 26.34 Typical dual-bus power system with separate UPS systems feeding the different power distribution buses. This type of deployment puts the power failure point at the individual IT device, ensuring the highest levels of availability. Source: © 2021, Eaton

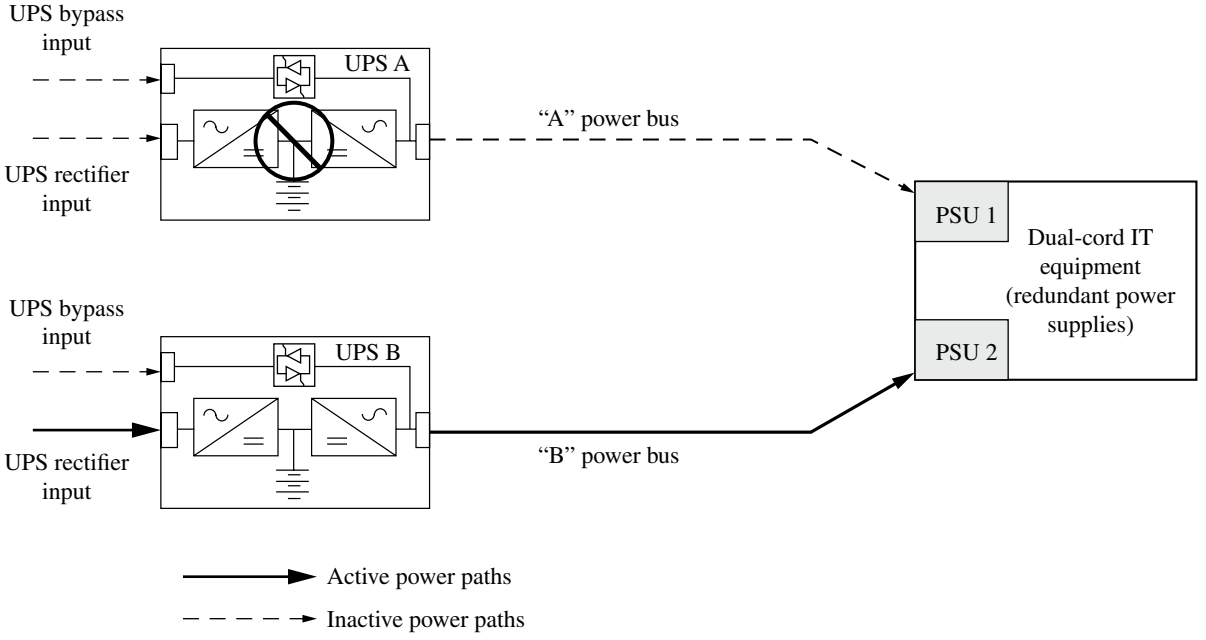


FIGURE 26.35 Failure of one of the power paths or UPS forces the other UPS to assume the entire load; therefore, designs like this should be sized properly and continually monitored to reduce the chance of a UPS or power distribution system overload. Source: © 2021, Eaton

26.4.6 Separate UPSs with Multi-Level Redundancy

Higher levels of redundancy can be achieved with a dual-bus system, especially if each bus gets its power from a different mains substation. In Figure 26.39, each side has two UPS modules (a primary and standby UPS, for $N+1$ protection), a system bypass module (SBM) to transmit power from the UPS modules or mains source, and its own backup batteries and diesel generator. Under normal operation, one bus feeds power to distribution panels serving one power supply in the dual-corded IT equipment; the other bus feeds distribution panels serving the other power supplies.

If any UPS drops off-line, the standby UPS on that side goes into action. Even if both UPSs on a side became unavailable, the IT equipment would still be powered from the other side. If a

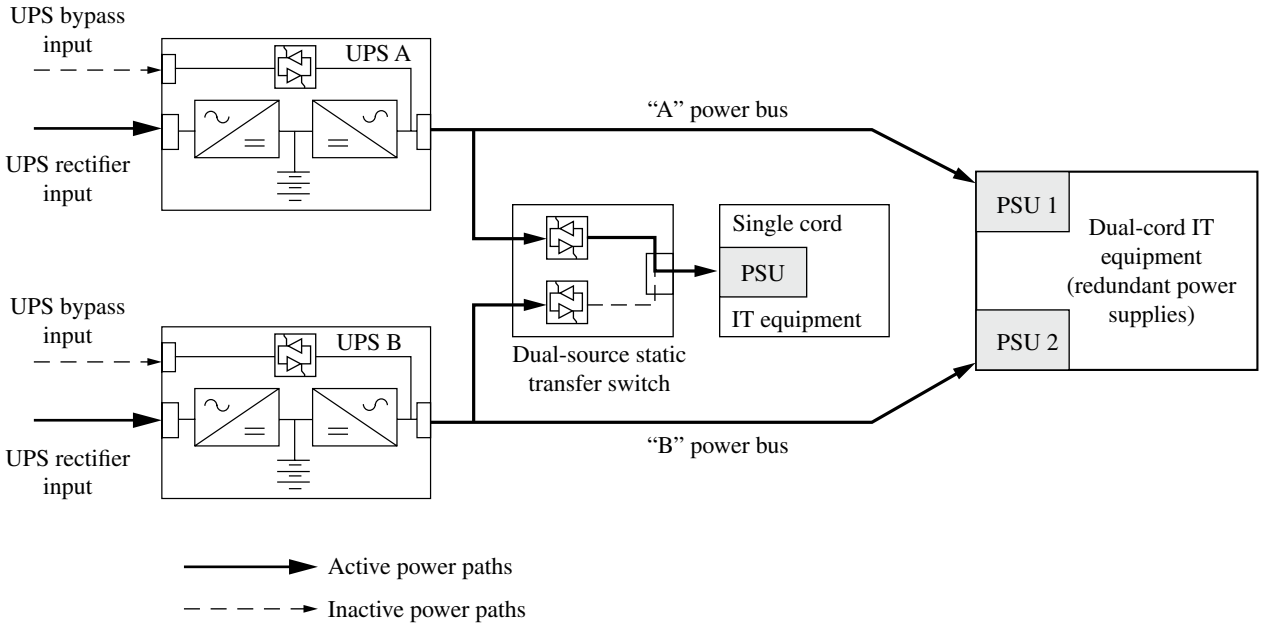


FIGURE 26.36 The dual-source static switch above is connecting the "A bus" to the single power cord loads. Source: © 2021, Eaton

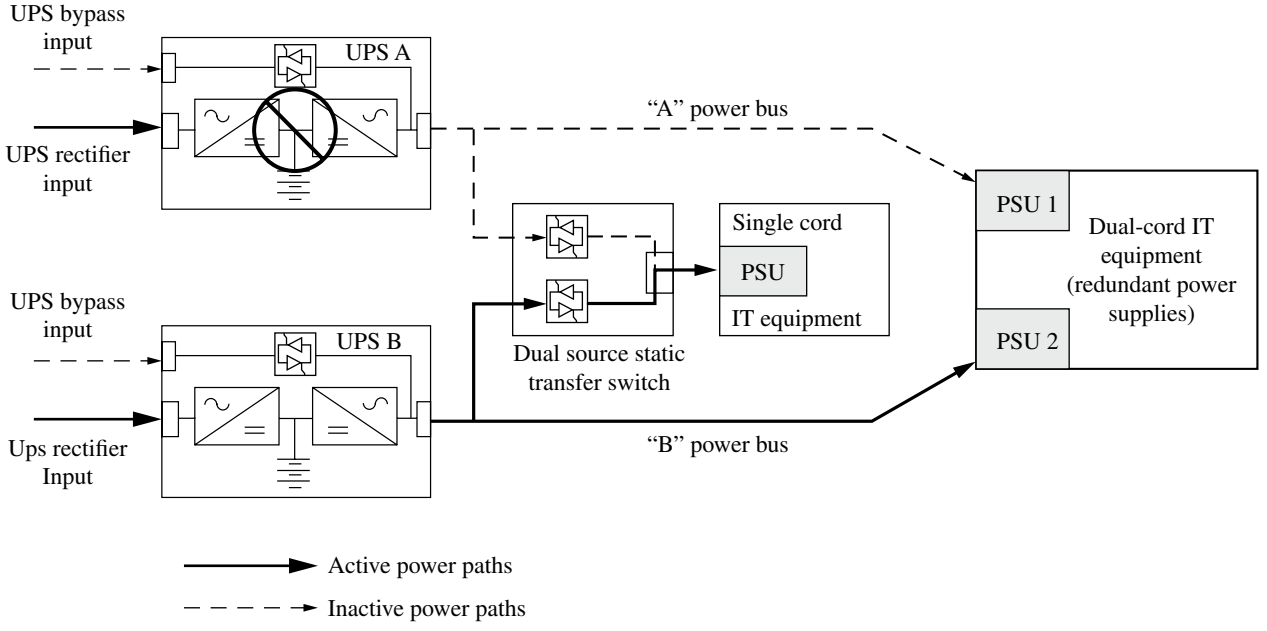


FIGURE 26.37 Failure mode: the static switch must instantly switch the single power cord loads to the "B bus". Source: © 2021, Eaton

mains substation went out, the power would remain up, because the other side is served from a different mains source.

For its high availability, this is a widely used arrangement. But "redundant redundancy" is expensive, and there's still the issue of what to do with single-corded loads. You can add a power synchronization controller that resolves the synchronization issue described earlier and simply accept a small point of vulnerability. In the arrangement shown in Figure 26.40, those redundant UPS systems are linked via a hot tie cabinet (HTC). The HTC has breakers that can isolate either side from the power chain entirely or link them together in parallel.

In normal operation, the breaker in the middle would be open, isolating the two redundant UPS systems from each other. The UPS system on the left feeds its output to the left-side bus. During a failure condition or routine maintenance of, say, the left side, the breaker in the middle would be closed and the left breaker open. Then the right-side UPS is powering both the A bus and B bus. The loads see no change in the voltage, frequency, or quality of the power they receive.

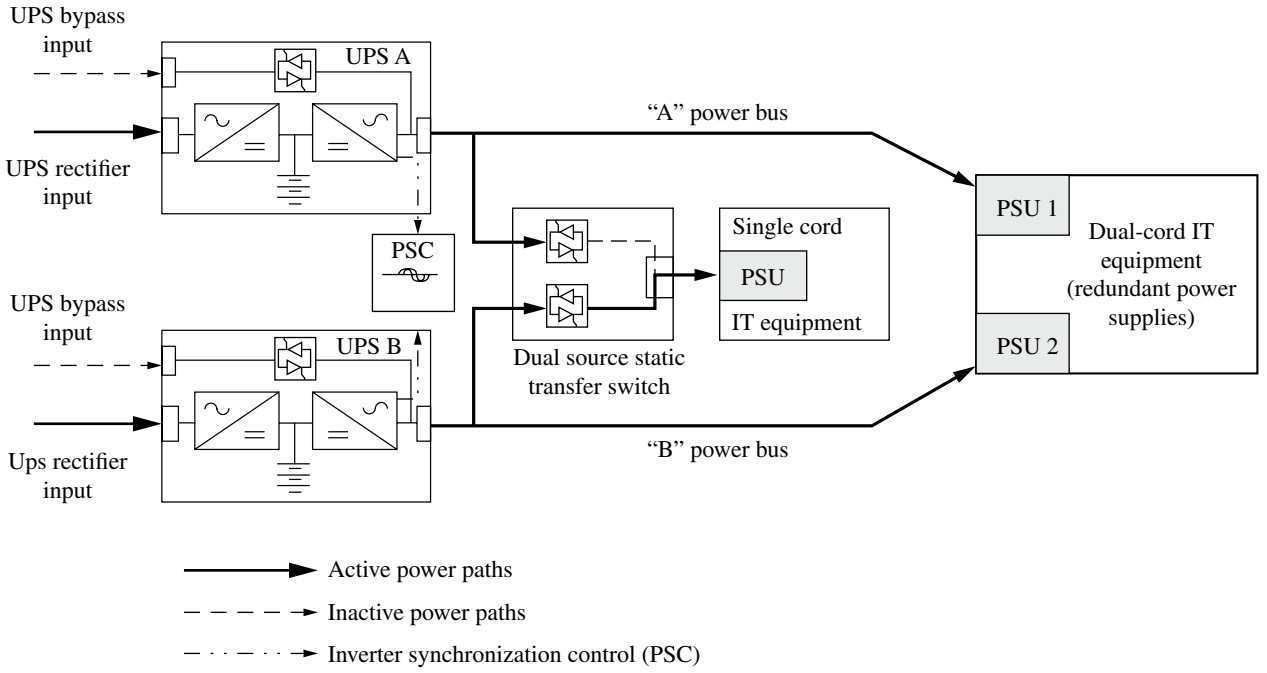


FIGURE 26.38 Power sync control (PSC) is necessary to ensure the static switch can change sources without disturbing the single power cord loads. Source: © 2021, Eaton

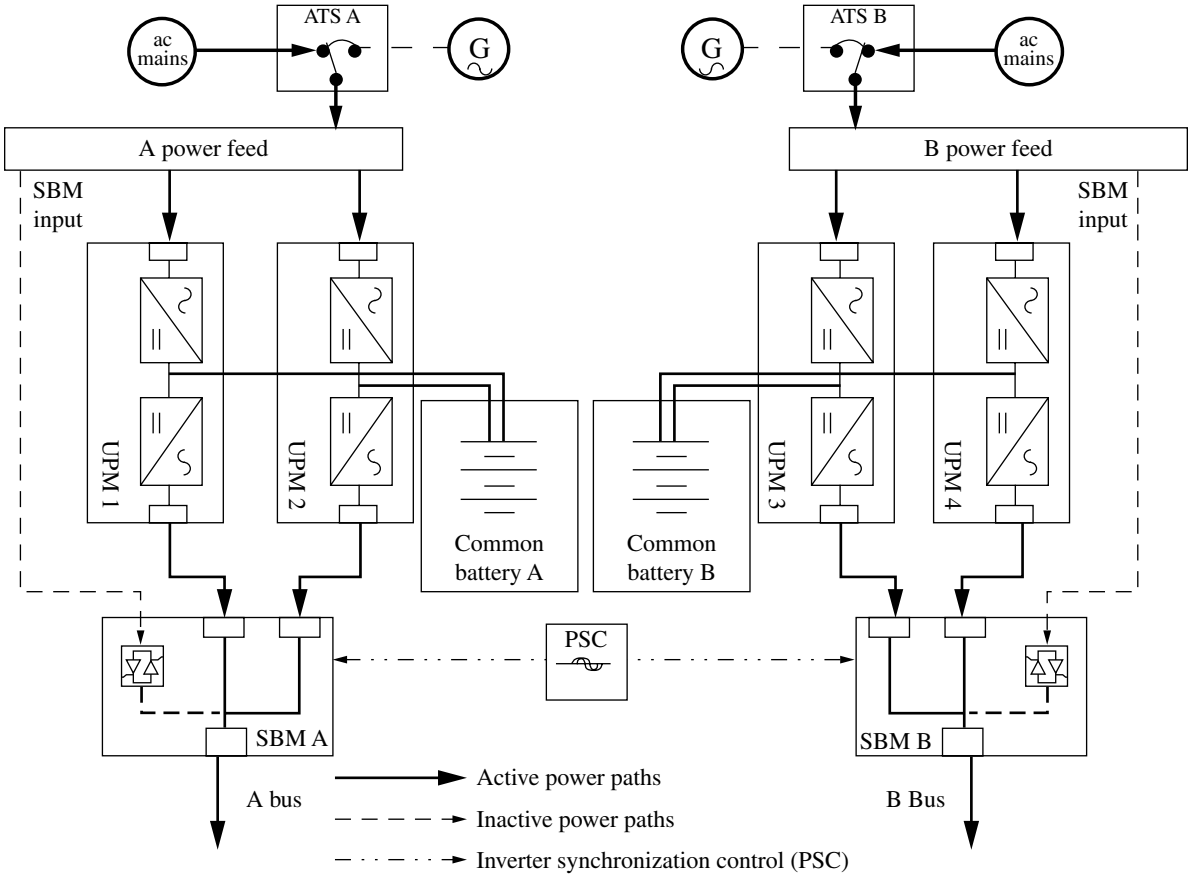


FIGURE 26.39 Multiple $N+1$ UPS systems can be used in a dual-bus configuration. Each bus has its own system bypass module (SBM) and upstream ATS with generator. A common battery is used for each group of uninterruptible power modules (UPM). Source: © 2021, Eaton

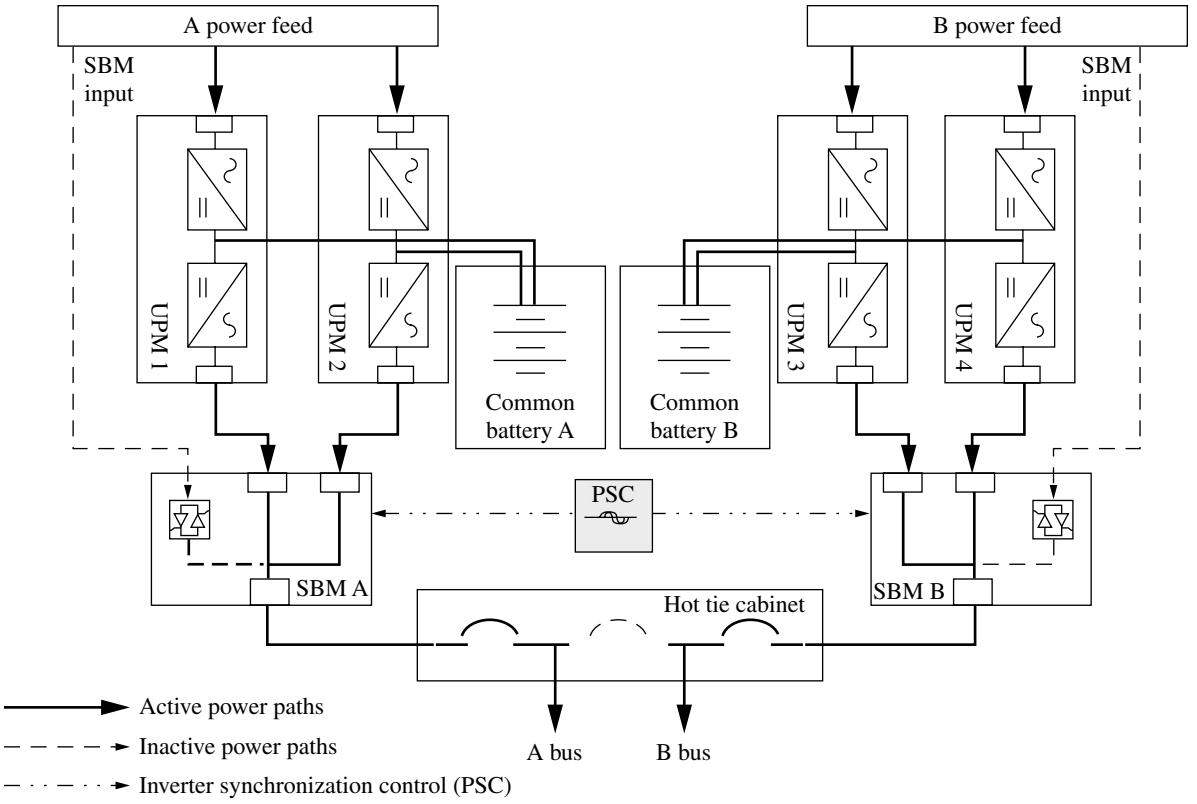


FIGURE 26.40 The hot tie cabinet (HTC) allows concurrent maintenance of the upstream A or B bus, while the downstream loads stay powered by both buses. Source: © 2021, Eaton

26.4.7 Customization Options for Large Parallel Systems

In practice, large customers need one-of-a-kind specialized configurations that match their unique needs for availability and manageability. Many options are available for parallel UPSs, such as:

- Wraparound maintenance bypass, to allow loads to keep running (off straight mains power) even if the parallel system is unavailable, such as during a catastrophic disaster.
- Redundant breakers in the tie cabinet, to permit maintenance of the primary breakers without turning the system off.
- Separate load bank breakers in the switchgear, to enable use of a load bank to test the UPS system under load while it is isolated from protected loads.
- Communication cards and a monitoring system for remote monitoring.

26.4.8 Other Options for Establishing Redundant UPS Protection

Redundancy doesn't always require paralleling. There are other options for deploying multiple UPS modules—separate rather than paralleled—to provide an added layer of assurance in the power protection architecture. For example, separate UPSs can be set up to provide serial redundancy, where even if the primary UPS is off-line, its bypass path is protected by another UPS system upstream of the failed UPS module. This is sometimes called a "catcher system". Or a data center could be divided into separate zones served by separate UPSs, thereby minimizing the impact of any single UPS failure. Or separate UPSs could serve either side of dual-corded loads or source power from different mains substations. Furthermore, any of these options can be set up for duplicate redundancy. However, each option presents some compromises, compared with peer-to-peer configurations described earlier.

26.4.8.1 Catcher Systems

Also known as an isolated redundant system, a catcher system is a method of providing $N+1$ redundancy while utilizing fewer individual UPSs (saves capital investments) and avoiding the challenges in paralleling multiple UPS on a common bus. As shown in Figure 26.41, there are three primary UPSs not paralleled, each dedicated to its own discrete load. But instead of having the bypass sources for these three UPSs connected to mains power, the bypass

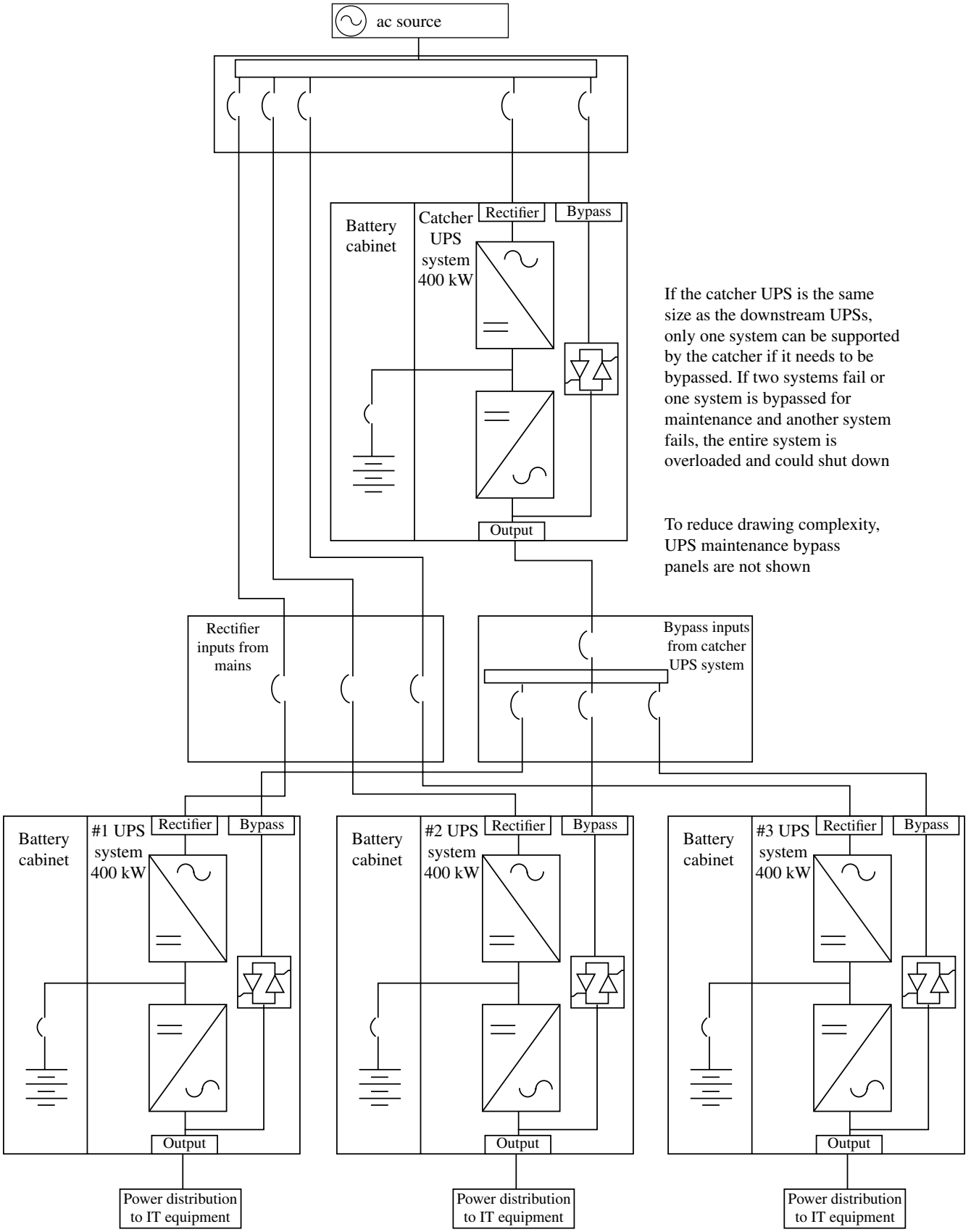


FIGURE 26.41 UPS deployed in the typical catcher UPS arrangement. Source: © 2021, Eaton

input to all three UPSs is sourced by the output of a single "upstream" UPS, called the catcher. So, the failure of any primary UPS results in a transfer to bypass, but that bypass is in fact the catcher UPS output, which provides conditioned power and battery backup, while the primary UPS is maintained or repaired. This arrangement creates $N+1$ redundancy for all the primary UPSs and requires a total of four UPSs to do the job. Contrast this with three conventional $N+1$ parallel systems, where each discrete load would require two identical UPSs to provide redundancy. This takes 3×2, or 6, UPSs to do the same job. Thus, a catcher system saves two UPSs and their respective switchgear and batteries while providing a similar level of redundancy. A catcher system does come with some obvious disadvantages (or at least compromises). Clearly, if more than one primary UPS fails simultaneously, the catcher UPS is at great risk of overload, and its bypass source may not be rated to handle multiple downstream UPS loads. While the likelihood of this happening is extremely remote, it must be considered versus the savings in initial UPS cost and TCO for the system. Another challenge is that conventional catcher systems use a single catcher with no more than three primary UPSs, due to control complexity, and the fact that once the catcher is "taken" by a single primary UPS, the remaining UPSs lose access to the catcher, and they are technically no longer $N+1$ redundant. Again, this is balanced against the cost of a conventional parallel system for each discrete load.

Fortunately, the advent of "smart catcher" systems can easily address both disadvantages described above. Consider the catcher system in Figure 26.42. In this type of smart catcher system, each primary UPS has two sources to feed its bypass input: one is from the catcher UPS output (CIB), and the other is from the mains bypass source (BIB). Each possible bypass source has a motor-operated circuit breaker in series, and only one breaker will be closed at any time. In normal operation the CIB breakers to each primary UPS are closed, and the BIB breakers are open, providing catcher-sourced bypass to each UPS. The real-time load level on all primary UPSs is sensed, and via the use of a PLC (programmable logic controller), the capacity of the catcher UPS is allocated among the primary UPSs. Thus, in the event of the failure of a lightly loaded primary UPS, it will transfer to the catcher and will use only a small portion of the catcher's capacity. This means the catcher's full capacity is not commandeered by a single failing UPS. The remaining capacity is available to support the other primary UPSs if needed, up to the full rating of the catcher UPS. This eliminates one of the disadvantages of a conventional catcher system. Depending on the actual load of each of the primary UPSs, their bypass source is determined by the available catcher capacity. That is, if a given primary UPS load level would overload an occupied catcher system, then the motor-operated CIB breaker will automatically open, and the BIB breaker will close, allowing that UPS to be transferred to the mains bypass source in response to any failure. So, in the unlikely event of multiple successive UPS failures, their ability to access the catcher is automatically controlled to ensure that they receive conditioned power with battery backup, even upon a failure or overload. This provides similar redundancy to that of an $N+1$ parallel system while eliminating the chance of catcher overload. Furthermore, the PLC controls used in a smart catcher system provide the ability for a single catcher to support more than the "3 primary UPS limit" that was mentioned above. Smart catcher systems with four, five, even up to eight primary UPSs have been deployed on a single catcher UPS. If required, the PLC controls may easily be made redundant.

26.4.9 Total System Installation

There are other considerations as the facility UPS system and its ancillary equipment are being defined and as the entire critical infrastructure is designed. These include items like alternative energy sources, maintenance and commissioning requirements, service level agreements, and monitoring and management requirements to ensure your critical infrastructure stays in top shape.

26.5 ALTERNATE ENERGY SOURCES: AC AND DC

26.5.1 Alternate ac Energy Sources

There are choices for the ac input to the UPS system. While the mains power grid is far away the most common source for any UPS, there are a few other possibilities. When designing for mobile installations or other harsh environments, where a stable mains source may not be readily available, one will need to evaluate other possibilities including:

- Wind or solar power (or wave power for marine installations).
- Diesel, natural gas (clean fuel), or turbine generators.
- Fuel cells.

26.5.1.1 Alternate DC Energy Storage Sources

Given the lead–acid battery's many flaws, it's no surprise that data center managers have long been clamoring for alternatives. At present, five such technologies show particular promise. Though only some are in widespread use today and a limited number of existing UPS models are equipped to support them, all are likely to gain increased traction over the years ahead.

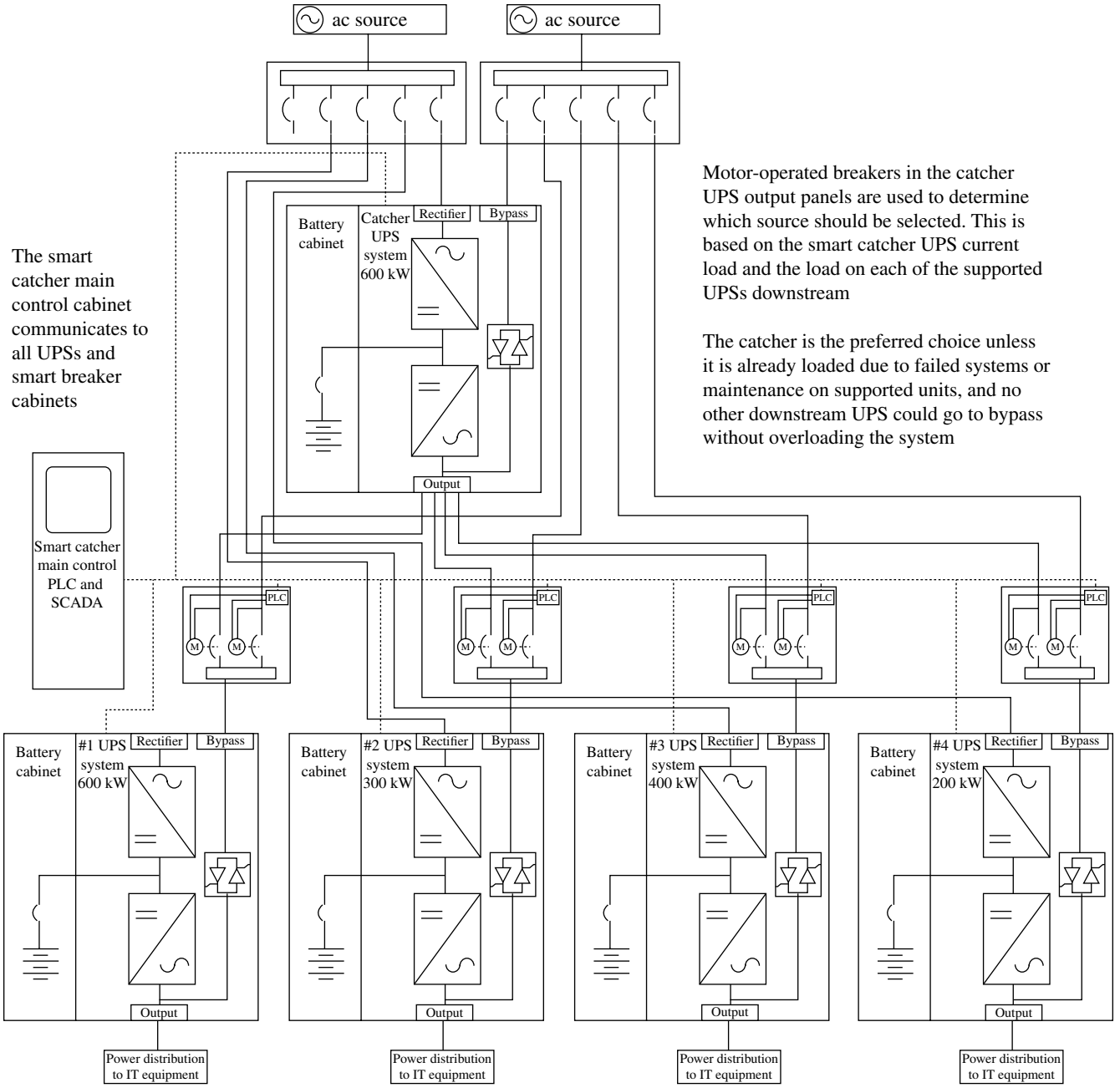


FIGURE 26.42 A smart catcher UPS arrangement with smart catcher control system. Source: © 2021, Eaton

Flywheels

A flywheel is a mechanical device typically built around a large metal disk. During normal operation, electrical power, via a motor, spins the disk rapidly. When a power outage occurs, the disk continues to spin on its own, generating dc power that a UPS can use as an emergency energy source. As the UPS consumes that power, the disk gradually loses momentum, producing less and less energy until eventually it stops moving altogether. Due to the cost and requirement to parallel multiple flywheels for longer runtimes, most are deployed with a backup time of typically 15 seconds to about a minute maximum, with the average time being 30 seconds. This is typically plenty of time to ensure that the backup generator (s) come online to support the critical power requirements. Typical flywheel power ratings are 200–500 kVA each. Modern flywheel systems have a service life of greater than 10 years, if mechanical maintenance is performed as recommended, usually yearly.

Ultracapacitors

Also known as supercapacitors, ultracapacitors are specialized extremely high-density batteries. They typically contain

nontoxic, carbon-based materials such as activated carbon and graphene. Available runtimes are very short, typically less than 30 seconds, so operation in instances where there is a backup generator is typically a must. The cost of modern ultracapacitors is approaching that of flywheels while avoiding the mechanical maintenance inherent in a flywheel system. Ultracapacitors can survive tens of thousands of charge–discharge cycles over their useful life without significant degradation in performance.

Fuel Cells

Unlike batteries, fuel cells generate power rather than store it. A fuel cell is basically an electrochemical device that converts fuel (typically hydrogen or natural gas) into energy. However, unlike an internal combustion engine, which also converts fuel into energy, a hydrogen-powered fuel cell's only exhaust product is water. As a result, everyone from auto makers to electrical utilities to UPS manufacturers are presently either introducing fuel cells to their product lines or investigating their use. Note that an idle fuel cell requires about 30 seconds to a minute to become functional, so some users parallel a small system battery string in parallel with the fuel cell system to bridge the gap while the fuel cell energizes. Note that most fuel cells are only about 85% efficient, but this is somewhat offset by the "clean exhaust" capability described above.

Lithium-Ion Batteries

Most cell phones and laptops use lithium-ion batteries, which have grown steadily smaller, lighter, and more energy dense over the last 15 years. Until recently, large-format lithium ion was rarely used in industrial settings or data centers. The cost was prohibitive compared with traditional VRLA or flooded electrolyte batteries. But the cost point for lithium has dropped dramatically, in just the last few years, making them a highly desired, affordable, and completely viable alternative to lead–acid batteries. Lithium-ion batteries can perform most of the same functions as lead–acid batteries, and their lighter weight and much smaller footprint, per kWh (kilowatt-hour) of capacity, are driving users to specify lithium for UPSs of all sizes. Additional benefits of lithium ion are a slightly wider acceptable operating temperature range and a very high charge–discharge cycle capability that allows a 10–15-year service life in a UPS application. It is this longer service life (vs. 5–6 years for VRLA) and resulting longer performance warranty that make them so attractive for both new installations and retrofit of old battery systems. A lithium-ion battery may well last the entire life of the UPS to which it is connected, dramatically improving the TCO of the entire system. Lithium-ion battery systems always include a BMS (battery management system) that does more than just monitor the voltage and current. These systems issue alarms and monitor internal temperature of each individual battery cell, along with balancing recharge current to individual cells to ensure safe operating temperatures and optimum performance. Unlike a monitoring-only scheme, the BMS can take direct and independent action in response to any condition that could result in a thermal runaway or severe voltage unbalance that could damage the battery. For example, a high temperature condition that does not respond to automatic mitigation attempts will trigger the UPS to stop charging the battery and, if that doesn't resolve the issue, will simply disconnect the entire battery string by tripping the cabinet circuit breaker, well before any risk of fire is encountered. So, the BMS, along with safer lithium chemistries and better packaging, have alleviated some earlier concerns around the use of lithium-ion batteries in UPS applications.

Challenges with lithium-ion applications are mostly centered around the "newness" of this technology. Recognizing that there would be some confusion and many questions from designers, installers, inspectors, and users, the NFPA and other organizations have updated fire codes (IFC 1206, NFPA 855) and building codes to provide requirements governing UL listings, cabinet placement, minimum thresholds in kWh, and maximum quantities in lbs. of electrolyte that can be placed in a given environment. Most of these new requirements are not much different than those required for VRLA battery installations, but engineers, installers, and building owners should be made aware of the changes reflected in these new codes prior to making final decisions on which battery technology to deploy, for both new and retrofit systems.

Nickel–Sodium Batteries

These batteries offer high capacity, rugged construction, good power density, extreme temperature operation, and excellent environmental characteristics, with no toxic material or byproducts of disposal. Nickel–sodium battery systems are already being used for outdoor applications like mains power and telecommunication facilities, where long runtimes are required. Smaller systems are already being offered for UPS applications as well.

26.6 UPS PREVENTIVE MAINTENANCE REQUIREMENTS

With proper servicing and a stable environment, a well-made UPS can operate safely and reliably for as long as 20 years. Without proper servicing, even the best UPS is significantly more likely to fail when you can least afford it. Companies in the market for UPS hardware, therefore, should also choose an appropriate UPS service plan from a service provider with the experience, know-how, and resources to provide comprehensive, high-quality support and do it safely and quickly!

Research indicates that regular preventive maintenance—which affords the opportunity to detect and repair potential problems before they become significant and costly issues—is crucial in order to achieve maximum performance from

your equipment. In fact, studies show that routine preventive maintenance appreciably reduces the likelihood that a UPS will succumb to downtime.

Selecting a service provider for your UPS can be a complex decision. Some customers simply purchase a service contract or extended warranty from the UPS manufacturer, while others prefer to contract with an independent service provider. A handful of companies employ internal engineers who can maintain all or certain parts of the power equipment. Still others choose to engage in UPS service only when something goes wrong. All these options have advantages and disadvantages, with no one choice being the best solution for every organization.

26.6.1 Common Questions for Choosing a Service Provider and Plan

1. If my UPS fails to provide reliable backup power, what is the cost of downtime to my organization?
2. How critical is continuous power to my application? Is it simply an inconvenience or do I lose customer sales, destroy products, or shut down a network of critical servers?
3. How long can I wait to obtain an emergency repair on my UPS? A week, a day, or an hour?
4. What's my position in the "priority list" during an environmental disaster?
5. How many trained field technicians for my specific UPS model are within 100 miles of my site, and do they carry the correct parts?
6. Do I have budget or cost constraints for UPS service?
7. How much scheduled preventive service do I need and what can I afford?
8. What level of service is recommended by the manufacturer?
9. Have I budgeted for battery, capacitor, or other unplanned part replacement costs?
10. Do I have competent electrical staff resources to do some or all necessary maintenance?
11. What is my risk tolerance for a UPS failure, and what happens to me personally if this UPS fails?

Regardless of the exact course of action you implement, an effective preventive maintenance plan saves time and money by minimizing business interruption and the costs of downtime, as well as enhancing your overall ROI by extending the lifespan of your critical power equipment.

26.6.2 Option 1: UPS Manufacturer's Internal Service Organization

Engaging in a service contract with the manufacturer of your UPS affords a number of benefits. To begin with, customers receive the extensive knowledge, capabilities, and expertise of factory-trained field technicians who receive ongoing and in-depth training on the manufacturer's specific UPS products. As a result, technicians are armed with the most up-to-date and comprehensive information pertaining to the functionality of the UPS, as well as access to the latest firmware and upgrade kits to maintain the highest level of performance from the UPS. Furthermore, the advanced troubleshooting capabilities of technicians translate to a reduced MTTR. When performing service on a UPS, the day-to-day familiarity and knowledge that comes from being brand specific cannot be underscored enough.

In addition to offering a deep support infrastructure of design engineers, technical support personnel, and other experts to back up its field technicians, UPS manufacturers generally possess the greatest number of field personnel and back-office resources. Furthermore, manufacturers most often have in place risk mitigation programs that are frequently overlooked by customers, such as appropriate safety programs and proper levels of insurance.

Another significant advantage to manufacturer-provided service is that technicians have spare parts readily available either from a stocked van or from a central location, ensuring that UPS problems are quickly resolved, most often on the initial service call. Furthermore, many service plans include discounts on part kits and product upgrades, which can significantly reduce the overall cost of maintenance.

To meet the varied needs of customers, UPS manufacturers offer a wide variety of service plans, including standard warranty, extended warranty, preventive maintenance, numerous service contract levels, and time and material (T&M) billing. Many also feature value-added support such as remote monitoring. Even more, most manufacturers offer service contracts that include options such as 7×24 coverage, with response times ranging from 2 to 8 hours or next-day response—an especially appealing benefit for customers in mission-critical environments. While the price of service may be slightly higher from a manufacturer compared with an independent service provider, the advantages that only a UPS manufacturer can offer may outweigh any additional costs.

26.6.3 Option 2: Independent Service Provider

An independent service provider is a third-party organization that often offers a range of services for UPSs or power quality equipment, such as professional maintenance, consulting, start-up, installation, and emergency service. Although independent service providers are frequently priced lower than a UPS manufacturer, they also generally have fewer resources available and may not be comprehensively trained on your particular UPS model.

While an independent service provider's field technicians have usually received training on either a specific UPS product or brand and may or may not be certified by a UPS manufacturer, it is virtually impossible to fully train a technician

on every UPS model from every manufacturer. Furthermore, because UPS products are continually being updated and changed, if a technician has not been recently trained by the manufacturer, he or she may not have the knowledge to adequately service the UPS.

When it comes to having access to repair parts, some technicians may carry the appropriate parts with them or have them available from a central location. However, it is difficult to carry a local supply of adequate parts for all brands. Generally, independent service providers will access a UPS manufacturer's deep support infrastructure of design engineers, technical support, and experts to back up their own field team, as the depth of their own resources can be limited. Insurance and safety records may or may not be maintained at an acceptable level. While independent service providers generally do not deliver a factory warranty unless contracted by a manufacturer, they do offer preventive service, a variety of service contract levels, and T&M billing. Some may offer value-added support such as remote monitoring.

26.6.4 Option 3: Self-Maintenance

If an organization has an internal resource that possesses sufficient electrical and safety skills, it may make economic sense to perform self-maintenance on a UPS. The most important aspect of self-maintenance is to have an efficient plan in place, in which routine scheduled maintenance is performed and common wear items such as batteries and capacitors are proactively addressed.

First responder training enables a skilled person to understand the operation, safety, environmental concerns, and basics of preventive maintenance on a specific UPS. This person must also understand the various alarm conditions and responses required for specific events, as well as the steps to start and stop a UPS correctly in various applications.

A spare parts' kit obtained from the UPS manufacturer can supplement those who choose to self-maintain their UPS equipment. However, it is important that an organization also has access to a professional service provider for more critical repairs, upgrades, or routine maintenance that may be required to supplement a self-maintenance resource.

26.6.4.1 Questions to Ask When Considering Self-Maintenance

Before opting to perform self-maintenance on a UPS, consider the following questions:

1. Is there an internal resource within your company that possesses basic UPS knowledge and electrical skills? If so, does this individual have time that can be designated to UPS maintenance?
2. Has your organization developed a specific plan for self-maintenance, including a schedule for replacing common wear and tear items?
3. Has a spare parts' kit been purchased from the UPS manufacturer?
4. Has an external service resource been identified for more critical repairs?

26.6.5 Option 4: Time and Material

Paying as you go is a common UPS maintenance approach that can be appropriate in certain situations, primarily for very old UPS models where no service contract is available. However, this tactic does not make good economic sense for complex, multi-module, or redundant UPS configurations.

Available at any time to all customers, T&M is typically charged per hour of labor, often with a minimum number of hours required. Charges are also generally more for after-hours and weekends compared with normal business hours. Response time for T&M is typically "best effort" with no guarantee of arrival, as customers with existing service agreements are always given priority over T&M customers.

Another downside to T&M is that replacement parts are usually very expensive. For example, the average board for a common three-phase 80 kVA UPS costs more than $5,200, while power modules that integrate several components exceed $10,000 each, with larger models containing several pairs of modules.

The uncertainty of response time during an emergency and financial exposure to unplanned repairs may make T&M less attractive to more mission-critical organizations. On the other hand, T&M may be appropriate for a self-maintainer, in situations where a UPS is not fully utilized or where preventive maintenance is being performed by a manufacturer or independent provider and the insurance portion of a service contract (parts and labor coverage and emergency response) is deemed unnecessary by either self-insuring or other reasons.

26.6.5.1 Questions to Ask When Considering T&M

If you are considering the pay-as-you-go approach, it is important to first consider the following questions:

1. Is there a service plan available for your particular UPS?
2. How complex is your organization's UPS?
3. Is your UPS utilized regularly or occasionally?
4. Is your UPS supporting mission-critical applications?
5. In the event of a UPS failure, can your organization afford an uncertain amount of downtime until a technician is able to schedule a service call?
6. Does your company have sufficient funds allocated for T&M service, parts, and repairs?

26.7 UPS MANAGEMENT AND CONTROL

Even with a UPS, your IT system could still go down in case of an extended power failure or if the UPS gets overloaded

for too long. Communication software can not only provide real-time notification of UPS status but also lets you assign automatic actions to perform in case of a power event. This is extremely useful if your system operates continuously without users being present to manually shut down affected equipment.

For the past 20 years, most UPS systems have come with software that would signal one or more servers that ac power was lost and that the UPS was on battery. If ac did not return and the battery energy was near depletion, the software would close all open applications to prevent any data loss. When ac power was restored, the system would automatically reboot, bringing the system back to its previous state. This solution was initially implemented on small PC servers protected by a single UPS and then moved to larger systems with an array of operating systems, many of which were proprietary to the IT equipment manufacturer. Communication was established through an RS232 serial port or via relays to a simplistic control port.

As IT systems grew bigger in size and numbers, serial communication (be it RS232 or through a USB port) was replaced by network-based communication to enable communication between the UPSs and multiple servers. In this type of installation, the UPS is assigned its own IP address on the network and could be accessed remotely by all servers being powered by that UPS, so each server could be programmed to shut down or monitor the UPS for power issues.

As networks and UPS communication hardware and software became more complex, other automatic features were developed through power management software, including remote notification via email, pager or SMS, data accumulation allowing report generation and trending analysis, complex script programming to shut down a database or a program before stopping the server operating system, and much more. Even with all these advancements, the typical installation involved servers with single operating systems and with a single application running on each server.

Virtualization is now bringing a new set of complexities, as the bond between operating system and physical hardware is no longer the standard (Fig. 26.43). Some suppliers of UPS software must ensure that shutdown software agents are installed on each virtual machine as well as on the host machine. This can be quite tedious if the number of virtual machines is large, which is becoming the standard in many virtualized environments. Leading-edge UPS manufacturers have developed new software platforms that reduce this management complexity by integrating their software into virtualization management platforms like VMware's vCenter® or Citrix XenCenter®. In these environments, one single software installation can control and shut down any cluster of servers. Another benefit is the enablement of automatic live migration of virtual machines in case of a power outage, as you are no longer limited to the option of shutting down the servers and stopping operations. Business continuity is now possible through this integration, which is available not only on vCenter but also on Microsoft SCVMM or Citrix XenCenter.

To summarize, logical and complete power management applications can help companies:

- Monitor and administer their UPSs from any location with Internet access.
- Automatically notify key personnel of alarms or alerts.
- Perform orderly unattended shutdowns of connected equipment or better work with virtualization software to move virtual machines so as to maximize availability of key applications and hardware.
- Selectively shut down noncritical systems to conserve runtime.
- Analyze and graph trends to predict and prevent problems before they happen.
- Integrate with existing network and management systems via open standards and platforms.

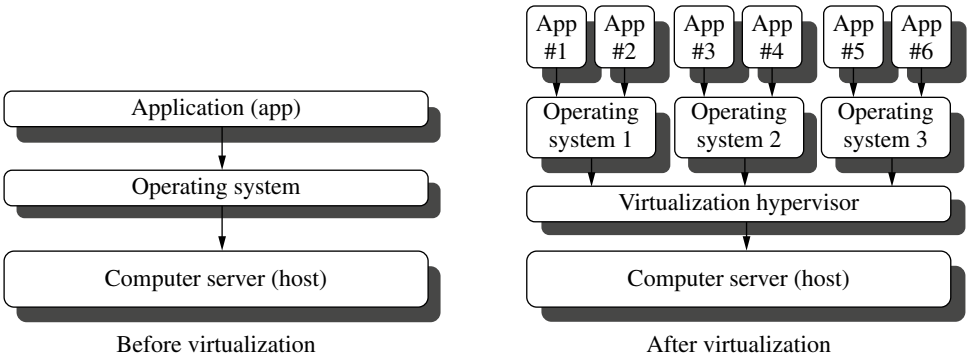


FIGURE 26.43 Advancements in computer hardware and software have led to the need to more efficiently use the computing power of each server, so the technology of virtualization was brought down from the mainframe architecture to the individual server level. Source: © 2021, Eaton

26.7.1 What to Look for in a Data Center Monitoring/ Management Product

A lot can be said for power insight and intelligence that resolves problems as they happen. But a solution should be flexible enough to proactively address issues with the right data at the right time. A comprehensive platform will work to automate, monitor, visualize, and predict these problems before they strike.

A well-designed set of power management applications will allow the client to take on the unique challenges you face within their environment by empowering them to:

- Make tasks simpler via advanced alerts and automated resolution.
- Make data actionable through fast, scalable interpretation and analysis.
- See beyond power consumption through 3-D infrastructure visualizations.
- Predict power component failure with cloud-based analytics.

And as your environment scales, so should your power management solution. Whether yours is a small deployment of UPS units and rack PDUs, a sophisticated data center housing thousands of servers and millions of datasets, or anything in between, the platform must be versatile enough to adapt with you.

26.7.1.1 Gain Control of Your Power Infrastructure Locally or Remotely

When managing any type of IT infrastructure, the ability to control the physical, electrical, and environmental resources in conjunction with the actual IT infrastructure and applications is paramount. Using policy-based automatic remediation to trigger advanced actions like migrating virtual machines during specific power and environmental events ensures that the critical applications stay operational through these aberrations. It is also important to look to integrate with industry leaders like Cisco, Dell EMC, HPE, Microsoft, NetApp, Nutanix, and VMware to keep critical applications running and automate resolutions for your entire network rather than a simple server-by-server shutdown that risks potential downtime.

Other benefits of a competent system are the ability to initiate the migration of workloads to increase system uptime and minimize generator load by suspending noncritical virtual machines. This will have the effect of extending UPS battery runtime without having to buy additional battery strings and cabinets, therefore saving on installation costs. In addition, using server hardware with power capping capability can benefit the battery runtime as well as tailor the power requirements in cases where there could be a potential rack or electrical system overload, or even adverting a high temperature shutdown in case of air-conditioning failure.

26.7.1.2 A Visual User Interface Is Useful and Should Include These Features

Being able to visualize the entire data center operation on a single screen may be one of the most beneficial aspects of a robust management system. This includes the ability to be "vendor agnostic," as most data centers have a mixed environment of systems from different manufacturers, and using standard protocol-based data collection makes deploying and updating the system much easier throughout the data center lifespan.

Capacity management with future growth forecasting or "knowing what's coming" is another benefit that should be high on the list for a management platform. The ability to analyze power devices visually and get notified ahead of replacement through reports that analyze the data and identify anything trending in the wrong direction can lead to much quicker moves, adds, and changes that are typical in the data center environment. Real-time rack-level statistics can quickly be remediated by using up-to-the-minute analytics, notifications, and alarms for events like a rack losing redundancy or power utilization that could exceed what the electrical branch can support. The management program should also include floor layout and graphic rack visualizations to plan the physical space for any required changes. This ensures that each rack is optimized for available space, power, and cooling.

26.7.1.3 Consider Using Predictive Data to Forecast Data Center Power Component Failure and Proactively Replace Components Before Failure

So, using the best management system in the world, you can still only try to manage around failures or issues, but not avert them. Wouldn't it be nice to proactively manage parts of the data center, rather than react to failures? Today's backup power systems have hundreds of sensors designed into them to monitor the performance of the equipment if used to do so. If this huge amount of individual device data is gathered and analyzed, could it be used? That answer has finally become a yes, through some the benefits provided by "big data" analytics.

New backup systems can proactively send this data to a manufacturer's cloud infrastructure, so big data algorithms can digest the data and compare it to what it should look like in normal operation and review it for characteristics of possible upcoming failures.

By providing this capability, the customer saves time and money. Site personnel can focus on other critical IT tasks, not on continually monitoring power system performance. Unplanned emergency maintenance expenses can also be

eliminated as the predictive analytics will warn you of the need to perform proactive maintenance before a hard failure causes possible downtime or time without backup protection. This type of proactive monitoring reduces risk as you are now capable of proactively replacing consumable parts before issues arise. As you can see, the fast-changing management platforms and proactive analytics are finally available to make today's IT infrastructure more predictable and even more reliable than in the recent past.

26.8 CONCLUSION AND TRENDS

Businesses today invest large sums of money in their IT infrastructure, as well as the power required to keep it functioning. They count on this investment to keep them productive and competitive. Leaving that infrastructure defenseless against electrical dips, spikes, and interruptions, therefore, is a bad idea.

A well-built power protection solution, featuring high-quality, highly efficient UPS hardware, can help keep your business applications available, your power costs manageable, and your data safe. By familiarizing themselves with the basics of what a UPS does and how to choose the right one for their needs, data center operators can ensure that mission-critical systems always have the clean, reliable electricity they need to drive long-term success.

As IT solutions progress, there are always industry leaders that are looking to challenge the ways of the past and deploy equipment in new and somewhat unproven configurations, pushing the economic envelope to their favor. With increasing reliability placed onto the IT software redundancy platforms, they are starting to look at ways to reduce the amount of equipment needed on the power redundancy side. While the economic impact of this can be shown on paper to be very attractive in the short term, the total business impact may not be known for years.

FURTHER READING

A current-dependent switching strategy for Si/SiC hybrid switch-based power converters. IEEE Trans Ind Electron 2017;64(10).

Crow LH. Achieving high reliability. J Reliabil Anal Center 2000, Fourth Quarter:1–3.

DoD Guide for Achieving Reliability, Availability, and Maintainability; August 3, 2005.

IEC 62040-1 Ed. 1.0 b Cor.1:2008. Corrigendum 1—Uninterruptible power systems (UPS)—Part 1: General and safety requirements for UPS.

IEC 62040-3 Ed. 2.0 b:2011. Uninterruptible power systems (UPS)—Part 3: Method of specifying the performance and test requirements.

IEEE Std 1184-2006 (Revision of IEEE 1184-1994): IEEE guide for batteries for uninterruptible power supply systems.

Miesner C, Rupp R, Kapels H, Krach M, Zverev I. thinQ!™ silicon carbide Schottky diodes. No. B112-H7804-X-X-7600, Infineon Technologies AG.NEMA PE 1-2003. Uninterruptible power systems (UPS) specification and performance verification.

Reliability Prediction of Electronic Equipment: MIL-HDBK-217F; 28 February 1995.

Understanding the Cost of Power Interruptions to U.S. Electricity Consumers. Ernest Orlando Lawrence Berkeley National Laboratory; September 2004.

Wolfspeed. Silicon carbide power MOSFET product overview.

27

STRUCTURAL DESIGN IN DATA CENTERS: NATURAL DISASTER RESILIENCE

David Bonneville and Robert Pekelnicky
Degenkolb Engineers, San Francisco, California, United States of America

27.1 INTRODUCTION

Natural hazards pose special design challenges for buildings that house data centers because of the value and fragility of the contents. Since building codes have as their primary purpose the protection of life safety (occupant safety or public safety), as opposed to business continuity, special structural design consideration is desirable for many data center buildings. This chapter provides an overview of the risks to buildings and contents due to natural hazards and addresses design considerations for critical buildings that may exceed those needed in the typical commercial building.

In the United States, all states currently adopt some edition of the International Building Code (IBC) for the regulation of structural design requirements. The IBC is based on a collection of adopted standards, including a load standard adopted from ASCE/SEI 7,[1] *Minimum Design Loads for Buildings and Other Structures* [1], and various construction material standards, addressing, for example, design requirements for steel, concrete, masonry, and wood. ASCE 7[2] is primarily intended to provide protection against structural failure, which it does through four sets of performance objectives, known as *Risk Categories*, discussed in more detail later. Inherent within each of these categories are an assumed probability of failure and a resulting level of reliability that is considered appropriate for the intended occupancy of buildings assigned to that category. There is inherent uncertainty associated with each category due to uncertainties in the loading intensity, material strengths, and construction quality. While the protection against structural failure is the primary basis for design requirements within each category, there is also intent to provide some level of protection of property, at least at lower force levels, although this protection is not clearly defined.

27.1.1 Structural and Nonstructural Components

Referring to a building's resilience to natural disasters, the building is typically broken down into its structural system and the nonstructural components. The structural system consists of all the floor and roof-decks or slabs, beams, columns, foundations, and any load-bearing walls. The nonstructural components are everything else. Exterior cladding; mechanical, electrical, and plumbing equipment; piping; ductwork; access floors; and server racks are all considered nonstructural elements.

Although the structural system largely controls the overall building performance under natural hazards, it represents a small percentage of the total building value. In the case of an office building, the structural system may, for example, represent about 20% of the total building shell value, the remainder being controlled by architectural, mechanical, and electrical components and systems. Tenant improvements reduce the percentage further. In the case of a data

[1] ASCE/SEI 7 describes the means for determining dead, live, soil, flood, tsunami, snow, rain, atmospheric ice, earthquake, and wind loads, as well as their combinations.
[2] https://www.asce.org/asce-7/.

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

center building, the structural system may represent as little as 5% of the total cost of the shell and core and a fraction of that when fully equipped. Since relatively modest increases in costs associated with enhancements in the structural system and anchorage of nonstructural components can have major impacts on performance, the return on this investment can be significant.

27.1.2 Environmental Design Hazards

Obviously, a building must support its own self-weight and the weight of all the equipment and people inside the building. Those loads are typically referred to as gravity loads. In addition to gravity loads, buildings are designed to safely resist loads imposed by their anticipated exposure to natural hazards. Earthquake effects (ground motion, ground failure, and tsunamis) probably have the greatest impact on data centers, followed by wind loading including hurricane, typhoon, and tornado. The resulting flooding associated with both hazards also poses a significant risk. Snow and rain can also impose significant loadings on the building's roof. Both seismic and wind loads impose forces on the primary structural system and on nonstructural components, such as exterior cladding. In addition, earthquake loads affect interior equipment and utility systems.

27.1.2.1 Earthquake Effects

In an earthquake, the amount of damage to the structural system and to the various nonstructural systems and components will vary significantly from building to building based on the intensity of the ground motion, the type and quality of structural system, the quality of the anchorage and bracing of nonstructural systems, the strength and ruggedness of the components internally and externally, and the quality of construction. As discussed in the following, it is possible to influence the damage and associated financial loss, as well as the loss of operations (repair time), through design processes that consider hazard level and performance more directly.

27.1.2.2 Tsunamis

Many coastal regions are affected by tsunamis (ocean waves) created by earthquakes. There are two manners by which a tsunami affects a building. The first is the impact of the waves on the exterior walls and the forces that are imparted to the structure. The walls can be significantly damaged or even blown out. The wave can generate a huge force that could also cause permanent lateral deformations to the structure or in the most extreme cases push the entire structure over. The second is the flooding that occurs due to the waves, which can cause significant damage to many mechanical and electrical components.

27.1.2.3 Wind Effects

Wind, whether resulting from a hurricane, tornado, cyclone, or storm, affects data centers in a similar manner. As the wind blows against the exterior of the building, pressure is generated against the exterior cladding, which translates into the supporting structure. Additionally, there is typically an upward force generated on the roof as the wind blows over it. Typically, wind-induced damage affects isolated areas of the exterior or the roof, where local failures of the cladding, roof, or supporting framing occur. In more extreme cases, the entire structure can be deformed laterally or in rare cases completely blown over. An additional issue in windstorms is that strong winds can pick up objects and then blow them into buildings. The object that is made airborne is termed a missile, and its impact, or "strike," can damage exterior cladding, in particular windows.

27.1.2.4 Rain Effects

Rain from storms only affects a data center if there is insufficient drainage on the roof to allow for ponding to occur or the envelope of the building is damaged, allowing water to seep into the building. Ponding on the roof occurs when there is a pooling of water due to insufficient or blocked drainage systems. The accumulation of water can become significant enough to overstress the roof framing, resulting in local collapse.

27.1.2.5 Snow Effects

Snow, similar to rain ponding, primarily affects structures by overloading the roof framing. Snow drifting, where an uneven accumulation of snow occurs, or rain on snow, which increases the mass of the snow, can result in increased loading on the roof framing, potentially leading to local collapse.

27.1.2.6 Flooding

Among the most destructive effects of a hurricane or tropical storm is the flooding in coastal regions due to storm surge and heavy rainfall. Flooding is the most common natural disaster in the United States and also results from other causes such as dam failures or river overflow. The Thailand flood of 2011 resulted in over $45 billion in economic damage, mostly associated with manufacturing facilities, and the 2012 flooding from Superstorm Sandy resulted in over $30 billion in economic losses. The most significant effect of flooding is the water damage affecting the nonstructural components. However, in very significant floods, the floodwaters can act like waves and impact the structure, causing damage similar to tsunami waves.

27.2 BUILDING DESIGN CONSIDERATIONS

27.2.1 Code-Based Design

Because of geologic, environmental, and atmospheric conditions, the probabilities and magnitudes of natural disasters vary. In order to develop design and evaluation standards, the engineering community has selected "maximum" probabilistically defined events for each hazard, which are felt to capture a practical worst-case scenario for the specific region. For example, it is theoretically possible that a magnitude 9 earthquake could occur in some parts of the country. However, the probability of that occurring is so remote, and the associated forces on the structure so great, that it becomes impractical to consider. On the other hand, it is not impractical to consider a magnitude 8 earthquake in the San Francisco Bay Area given that there was a magnitude 7.9 earthquake that occurred in 1906 and that geologic studies of the region indicate that the probability of such an occurrence, while small, is high enough that it should be used as the "maximum considered earthquake" for parts of that region. Conversely in Phoenix, the probability of a significant large earthquake is so remote that it is unnecessary to require consideration of a large-magnitude earthquake. Buildings in both regions, and the rest of the United States, are designed considering earthquake ground shaking that has a 2% probability of being exceeded in 50 years.

For structural design purposes, U.S. building codes and standards group buildings into Risk Categories that are based on the intended occupancy and importance of the building. Minimum design requirements are given within each of four such categories, designated as Risk Category I, II, III, and IV. The general intent is that the Risk Category numbers increase based on the number of lives affected in the event of failure, although higher risk categories can also offer greater protection against property damage and downtime. Risk Category II is by far the most common category and is used for most commercial and residential construction and many industrial buildings. Risk Category III is used for buildings that house assemblies of people, such as auditoriums; for buildings housing persons with limited mobility, such as K-12 schools; and for buildings containing hazardous materials up to a specified amount. Risk Category IV is used for buildings housing essential community services, such as hospitals, police and fire stations, and buildings with greater amounts of hazardous or toxic materials. Risk Category I is used for low occupancy structures such as barns. The ASCE 7 standard attempts to provide the higher performance intended with increasing risk category by prescribing increasing design forces and stricter structural detailing requirements for the higher risk categories. Naturally, construction costs tend to increase in higher risk categories as they do with increasing hazard level. Loads associated with each natural hazard are addressed separately in each Risk Category with the intent being to provide improved performance in the higher categories.

Data centers in accordance with code requirements generally fall into Risk Category II, suggesting that special design consideration is not warranted. This is consistent with the primary code purpose of protecting life safety while leaving the consideration of business risk to the owner. This introduces a building performance decision into the design process that is sometimes overlooked in the case of high-value or high-importance facilities, such as data centers. The desire to provide protection against substantial damage to the building and contents and the desire to protect ongoing building function would need to be addressed through performance-based design that exceeds prescriptive code requirements. Much of the desired protection of data center buildings, equipment, and contents can be achieved by treating them as Risk Category IV structures and by using performance-based rather than prescriptive code requirements.

27.2.2 Performance-Based Design Considerations

Most buildings are designed based on the design requirements specified in the standard for the applicable risk category of the building, and this is the approach that is taken where performance goals are not articulated by the building owner. In such cases, the performance expectations are not actually assessed by the design team, meaning that the owner's performance goals relative to financial loss and downtime may not be addressed. Where building performance goals are more clearly understood, as is often the case for data centers, performance-based design requirements may be appropriate. For new building design, performance-based design may be used in two different ways. First, ASCE 7 permits alternative performance-based procedures to be used to demonstrate equivalent strength and displacement to that is associated with a given Risk Category without adhering to the prescriptive requirements. Such procedures may facilitate the use of more creative design approaches and allow the use of alternative materials and construction methods, resulting in a more economical design.

The second way that performance-based design is used may be more relevant to data centers since it relates more directly to the consideration of expected financial loss associated with damage to the building and contents and to the facility's loss of operations after the event. Recently, a comprehensive methodology was developed for performance-based seismic design called FEMA P-58, *Seismic Performance Assessment of Buildings, Methodology and Implementation* [2]. The FEMA P-58 methodology involves simulating the performance of a given design for various earthquake events and characterizing the performance in terms of damage consequences, including life safety, financial loss, and occupancy interruption. The design can then be adjusted to suit the objectives of the owner.

27.2.3 Existing Buildings

Data centers are often housed in existing buildings that may have been constructed under older building codes using structural standards that have been superseded. This results in a broader range of performance expectations than defined for new buildings considering their risk categories. Many existing buildings do not meet the current Risk Category II requirements so that lower performance is expected. However, existing buildings may be upgraded to provide performance similar to new buildings designed to various risk categories.

The building codes are evolving documents. Every major disaster provides engineers with new information on how buildings perform and what did or did not work. Over the years, code requirements have become significantly more robust. In some cases, methods of design and construction that engineers once thought were safe and thus were permitted by code were found to be unsafe, and later editions of the building code reflected those findings. Also, as scientists study natural disasters, a greater understanding of their impact is realized, and this has often translated into improved design requirements.

This is not to say that all modern buildings pose little risk and all older buildings pose great risk. Performance of buildings, new or old, can vary considerably and is influenced by many factors. The type of structure chosen, the quality of initial design and construction, modifications made after the initial construction, and the location of the building can all affect the performance of the building. Because of that, the risk to a data center of a natural disaster, both in terms of life safety and financial loss, requires special attention.

Even in cases involving modern buildings, the design generally will not have specifically considered enhanced performance. Therefore, during the acquisition of a building for data center usage, it is important that the due diligence process includes for budgeting purposes an understanding of the vulnerability of the building to natural hazards, just as electrical and mechanical system requirements are understood.

27.3 EARTHQUAKES

27.3.1 Overview

The ground shaking, ground failures, and ensuing fires caused by major earthquakes have rendered parts of entire cities uninhabitable, as was the case in San Francisco in 1906, Port-au-Prince, Haiti, in 2010, and Christchurch, New Zealand, in 2011. The vast majority of earthquakes, however, do not destroy entire cities, but still do considerable damage to buildings, transportation structures, and utility infrastructure. This can render a data center inoperable, either through damage to the physical structure or loss of critical utility services like power. The occurrence of major damaging earthquakes is relatively rare when compared to the lifetime of a data center facility but also unpredictable. Therefore, earthquakes exemplify the high-consequence, low-probability hazard. The occurrence may be rare, but the consequences are too great to be ignored in the design of a specific facility. Earthquake effects may represent the most challenging natural hazard for data centers that are exposed to them. This is due to the high value of equipment and contents that are susceptible to damage from shaking and the possibility that such damage may cause a loss of operation of the facility.

Building codes recognize the need to provide specific provisions to reduce the impact of a major earthquake on communities in locations prone to damaging events. Because earthquakes occur infrequently but produce extreme forces, our codes recognize that it is cost prohibitive to design for buildings to remain damage-free in larger events. Therefore, provisions have been developed that will reasonably ensure life safety in a relatively large earthquake while accepting that there might be a very rare earthquake in which the building would not be safe. There has always been some consideration given to protecting function and property in lesser earthquakes, but no explicit provisions have been provided.

27.3.2 Earthquake Hazard

While individual earthquakes are unpredictable, there has been much study to document locations where earthquakes will have a greater likelihood of occurring and their maximum potential magnitude. Earthquakes most frequently occur on regions adjacent to boundaries of tectonic plates. The most common example of this is the region known as the Pacific "Ring of Fire," which runs along the western coast of North and South America, the eastern coasts of Japan, the Philippines, and Indonesia and through New Zealand. Due to the relative movement of tectonic plates, stresses build up to the point where the Earth crust fractures, resulting in a sudden release of energy. Typically, this occurs along previously fractured regions, which have been classified as fault lines. The San Andreas Fault, which runs along the western coast of northern California and then inland through Southern California, is an example of this. The Great 1906 San Francisco Earthquake occurred along this fault.

In addition, there is the potential for earthquakes to occur in regions within tectonic plates. These earthquakes, called intraplate earthquakes, occur because internal stresses within the plate cause fractures in the plate. An example of this is the New Madrid Seismic Zone near the tip of southern Illinois, which produced massive earthquakes in 1811 and 1812.

These regions, and specifically the faults, have been studied by geologists and seismologist to the extent that maximum estimates of earthquake magnitudes and probabilities

of those earthquakes occurring in a given time frame have been established. That information is then translated into parameters that engineers can use to assess how an earthquake would affect structures, typically a representation of the maximum acceleration of the ground during the event. From that, information maps of earthquake hazard are produced, which provide engineers with information on the earthquake hazards that they should consider in the design or evaluation of buildings.

Since it is impossible to predict earthquake occurrence precisely, the hazard maps are typically based on some assumed probability of an earthquake occurring within a given time frame. In some locations, there is a possibility of not only a very large earthquake but also the possibility of more frequent, smaller, but still damaging events. In other locations, there is mainly the likelihood of having a rare but extremely damaging earthquake. The San Francisco Bay Area is an example of the former, and the middle Mississippi Valley (Memphis and St. Louis) is an example of the latter. This may be a factor to be considered in siting a new data center or evaluating an existing one.

27.3.3 Common Effects on Buildings

The release of energy in an earthquake translates through the ground and is expressed as horizontal and, to a lesser extent, vertical shaking at the surface. How the shaking propagates up through a structure affects how the structure will respond to the earthquake. Ideally, the structure when shaken would be undamaged, and the nonstructural components would remain securely fastened against shifting or toppling. Unfortunately, in the most seismically active regions, it is generally not economically feasible to design a structure to be so robust that it can withstand the largest possible earthquakes without damage. Therefore, engineers have recognized the need to design structures allowing damage to occur but in a controlled manner that does not compromise their overall stability.

As a structure is shaken, it may either deform in a ductile manner that absorbs energy or in a brittle manner allowing portions of the structure to fail. Older structures, particularly those designed before the 1970s or 1980s, often lack the features required to ensure ductile behavior. Consequently, those buildings and some marginally designed modern buildings can experience failures of structural connections or columns, which can result in partial or even large-scale collapse. Even when the failure is not significant enough to cause a collapse, it can damage a building to the point where the structure would be sufficiently weakened to make it vulnerable to aftershocks.

Even for structures that deform in a ductile manner during earthquakes, there is still the potential for damage. In some cases, it may be extensive enough to require repair prior to reoccupancy. Because of this, the engineering profession has adopted common performance states to describe a building's postearthquake state. They are illustrated in Figure 27.1.

The description of building earthquake performance is incomplete without discussion of how the nonstructural systems—the architectural cladding, finishes, furnishings, and mechanical, electrical, and plumbing equipment—are affected by the earthquake. Earthquakes can cause equipment to shift and topple if not anchored to the structure. Building codes provide requirements for anchorage design based on expected floor accelerations. Additionally, the swaying of the building can cause distribution systems, such as piping or ductwork, to break their seals, allowing contents to leak. Movement can also cause the building's envelope to break its watertight seals, allowing water intrusion in a rainstorm. Another damage consequence of building drift is the swaying of suppression systems, resulting in sprinkler head impact and resulting water damage.

Nonstructural damage can, and historically does, occur at earthquake intensities much lower than those that cause significant structural damage. This is of significant concern for a data center because most of the value and importance of the building is related to the equipment inside the structure.

27.3.4 New Building Design Considerations

For design purposes, earthquake ground motions are defined in terms of design accelerations given in hazard maps developed by the USGS and provided in ASCE 7. ASCE 7 requirements are based on seismic concepts developed in the *FEMA 750*, *NEHRP Recommended Seismic Provisions for New Buildings and Other Structures* [3], which is updated on a 5-year cycle. The maps provide *Maximum Considered Earthquake* ground motions that are adjusted (reduced) for design purposes and combined with factors related to dynamic response and system performance to provide design seismic parameters. Because buildings cannot be practically designed to resist large (design-level) earthquake ground motions while the structural system remains elastic, seismic design provisions incorporate ductility requirements that are intended to permit postelastic energy dissipation that may be accompanied by damage to the structural system, as well as nonstructural components. The general intent is to provide reasonable assurance that life safety is protected in Risk Category II buildings and that an enhanced level of safety is achieved in Risk Category III and IV structures, along with some level of protection against damage and loss of function. These inherent levels of performance are assumed within the standard, although the more clearly defined objective for typical buildings (Risk Level II) is collapse prevention in the very rare earthquake event, which is assumed to be achievable in most cases. The assumption is that the collapse probability is 1% or less in a 50-year period for Risk Category II buildings designed in accordance with the ASCE 7 standard

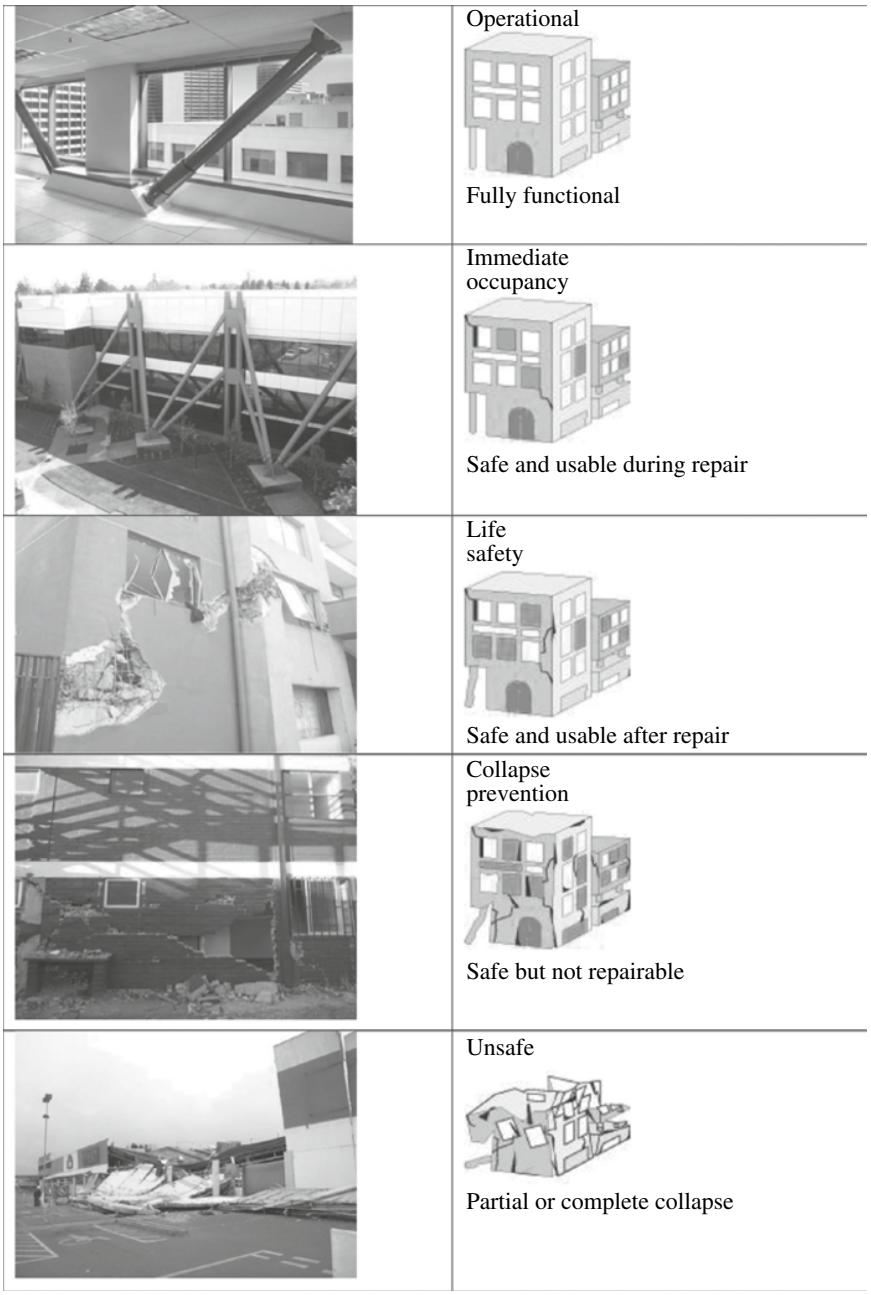


FIGURE 27.1 Building performance states. Source: Courtesy and © 2020 Degenkolb Engineers.

and that lower probabilities are associated with the higher Risk Categories.

Because a common building is only designed to prevent collapse and not to protect against damage in a large earthquake, new data centers should ideally be designed for higher performance. That higher performance should consider, at a minimum, a level that controls structural damage. Controlling damage is essential for ensuring that a data center does not get flagged as unsafe following a major disaster, something that could hinder reoccupancy. The repair time following a major earthquake may be quite long, also suggesting the need for enhanced design. In many instances, it is prudent to design a data center structure to the same level as essential facilities, such as hospitals and emergency operations centers. That would involve designing for the provisions of a Risk Category IV, as opposed to a Risk Category II, facility.

As previously stated, the equipment within a data center is more critical than the structure and therefore may require as much attention to design as the structure. In a typical building, most equipment is required to be seismically anchored if the building is located in a region of moderate or

high seismicity. However, while such anchorage prevents the equipment from toppling, it does not provide assurance that the equipment will be functional following the event. U.S. codes now require seismic certification by the manufacturer for equipment in essential facilities (e.g., hospitals) if required to be operable after an earthquake and for equipment containing hazardous substances. Equipment that has undergone this testing is seismically certified and should be considered for use in a data center.

The most critical pieces of equipment in data centers, the servers and server racks units, are not normally seismically certified. Often, the electronic components inside are sensitive to large floor accelerations. One way to protect the servers is to place the server racks on isolated platforms or to isolate the entire access floor. Using seismic isolation technology decouples the equipment from the shaking of the floor it is situated on, greatly reducing the accelerations imparted to the equipment.

27.3.5 Existing Building Mitigation Strategies

When an existing building is considered for a potential data center site or if a data center is housed in an existing building, one of the first steps should be to ascertain the seismic performance expectation of the building. A commonly used standard for this evaluation is ASCE 41, *Seismic Evaluation and Retrofit of Existing Buildings* [4]. Unlike standards for new building design, which contain prescriptive design requirements, ASCE 41 recognizes that an existing building contains a combination of structural elements of varying strength and ductility and that many of those elements would not meet the standards for new buildings. In many instances, the evaluation will require detailed analysis procedures, beyond the scope of those used in the design of a new building.

ASCE 41 has five structural performance levels: Immediate Occupancy, Damage Control, Life Safety, Limited Safety, and Collapse Prevention. Life Safety is the standard commonly associate with a typical new building, while Immediate Occupancy is typically associated with a new Risk Category IV essential facility. When assessing an existing building, Life Safety should be the minimum standard, while Immediate Occupancy may be the desired level and Damage Control an acceptable level. It will be rare to find an existing building that meets Damage Control of Immediate Occupancy; therefore, the choice will be to accept a building that only meets Life Safety and contains no protection against long downtime following an earthquake or to seismically upgrade the building.

If upgrade is chosen, it is ideal to construct the retrofit before the building is outfitted as a data center. Retrofit of a vacant structure is significantly less expensive and has less risk. If the structure is already occupied, greater care is required in the retrofit design. It may be advisable to consider structural upgrade approaches that place new structural elements on the outside of the building and above the roof, so as to limit the amount of construction occurring over the servers. If work must be performed inside the facility, temporary protective boundaries should be constructed over the servers, and the retrofit measures should be laid out in a manner that avoids or at least minimizes conflicts with existing mechanical and electrical components.

Like the structure, existing nonstructural systems are typically not constructed to new seismic codes. While this is not as significant an issue, because most equipment in a data center is relatively new (<10 years old), it is still common for it to have been installed without proper consideration for seismic bracing and likely no consideration for ruggedness and postearthquake function. Therefore, it is likely that the equipment will need to be properly anchored. Because existing equipment cannot typically be tested for ruggedness, the choice will have to be made whether to keep the existing equipment in place and accept the risk of function loss, replace the equipment with certified equipment, or isolate the existing equipment. If preventing loss of function is paramount, then isolating the equipment would generally be less expensive than replacing the equipment.

27.4 HURRICANES, TORNADOES, AND OTHER WINDSTORMS

27.4.1 Hazard Overview

There are typically three different atmospheric phenomena that produce wind gusts of sufficient magnitude to affect structures. They are storms, hurricanes/cyclones, and tornadoes. Similar to earthquakes, regional conditions dictate the magnitudes of these hazards. Also similar to earthquakes, wind hazard occurs transiently, and the magnitude cannot be precisely predicted. Because of this, scientists have developed models that can predict the probability of occurrence of a given wind gust of a certain size in a specified time frame or recurrence interval. For example, the current design wind for a typical building is based on a maximum wind gust that has a mean recurrence interval of 700 years.

Wind loads are defined in ASCE 7 by uniform recurrence interval wind speed contour maps that are provided separately for each of the four Risk Categories. The maps address all geographic areas of the United States including regions affected by hurricanes. The wind speeds provide pressures that are combined with factors related to exposure, topography, and directionality to provide design loading. The hurricane-prone region of the United States includes the Southern Gulf Coast region continuing up along the east coast. In wind design, the code-specified design pressures represent loads that the building could be expected to experience during a maximum wind event. So unlike seismic

design, in which the maximum expected earthquake forces are reduced to account for energy dissipation through inelastic response, it is intended that building structural systems, including components and cladding, experiencing design wind loading remain elastic and are not substantially damaged. An exception is made in the case of tornadoes, which can generate extreme wind loads that are not generally covered by building codes.

Tornadoes are not generally addressed by the building code because the probability that a building will be located within the path of a tornado of sufficient strength to be destructive is very low. Currently, the only structures where tornado wind speeds are considered are major high-risk structures like nuclear power plants. This is not to say that tornado effects cannot be considered in the design of a data center. There are maximum tornado wind speed maps available for the United States. For areas in "Tornado Alley," the Great Plains states, the tornado wind speeds are approximately 75% greater than those typically considered in building design.

27.4.2 Common Effects on Buildings

As the wind blows against a building, pressure is generated against the exterior cladding, which is transferred to the supporting structure. Additionally, there is typically a suction force generated against the roof and the side and back walls of the building. These pressures must be resisted by the cladding and roof elements and the structural members supporting them. Failures of cladding panels, windows, and doors are common in extreme wind events. Not as common, but still observed, are failures of the roof sheathing and roofdecks. This occurs when the suction pressure is strong enough to cause the sheathing to pull upward.

In more extreme cases, the entire structure can be deformed laterally or for lighter buildings even completely blown over. For most engineered buildings, these types of failures are rare. However, if a data center is housed in a light-framed metal building, the potential for this does exist and should be evaluated.

Another design consideration in windstorms relates to strong winds picking up objects and blowing them into buildings. The object that is made airborne is termed a missile, and its impact, or "strike," can damage exterior cladding and, in particular, windows. If a missile breaks a window or part of a wall, wind can rush into the building, increasing the internal pressure, increasing the demands on the roof and walls, possibly leading to failure.

Rooftop equipment can be susceptible to damage in windstorms. If the equipment is not sufficiently anchored, it can be blown over. The equipment can also be damaged if a large missile hits it.

Other than structural collapse, the main causes of function loss following a major windstorm are damage to the building cladding and damage to the rooftop equipment. If the building cladding is damaged, the building is no longer "watertight." Therefore, rainwater can enter the building and damage equipment. Because maintaining function of the air-conditioning system is so important in a data center, the loss of rooftop air handling equipment may result in shutdown of the facility.

27.4.3 Mitigation Strategies

For new buildings, the most logical mitigation of wind hazards is to design the building assuming it is a Risk Category IV Essential Facility. This will require that the building be designed for higher wind forces than an ordinary building. If the building is located in an area with a high potential for tornadoes, it may be prudent to consider even higher wind forces in the design of the building and its cladding. Another consideration is to provide windows and doors that are resistant to impact from missile strikes. There is an ASTM standard for these types of windows and doors, E1966. If the building must have rooftop-mounted equipment, the rooftop equipment could be surrounded by windscreens that are designed for the maximum wind forces and are of a material that can protect the equipment from missile strikes.

Existing buildings should be evaluated for the wind forces that a new building would be designed to. If the roof, framing members, cladding, or associated connections are overstressed, they should be strengthened. It is likely that the windows and doors are not resistant to missile impact and should be replaced if the tornado or hurricane hazard is high. Rooftop equipment will likely be unscreened or the existing windscreen inadequate. A new, compliant screen should be added or the existing screen upgraded.

27.5 SNOW AND RAIN

27.5.1 Hazard Overview

Similar to wind and earthquake, the snow and rain hazards vary based on the location of the building. The meteorological climate of the area will dictate the potential for major snowstorms or rain events. In both cases, the hazards are defined probabilistically. Maps can be found in building codes, which provide the design snow and rain levels, or those parameters can be obtained by site-specific studies.

The design rain hazard is commonly taken as the water accumulated on the roof from a storm with a 100-year mean recurrence interval. The rain accumulation is determined based on the slope of the roof, the types of primary and secondary roof drains, and whether those drains are blocked. There are a number of features on the roof, such as the presence of rooftop equipment, depressions in the roof, and the flexibility of the roof framing, which can lead to localized

ponding of rainwater, causing greater than anticipated demands. Presently, U.S. codes and standards do not provide for increased rain loads for essential facilities.

For snow, the 50-year mean recurrence (or 2% annual probability of exceedance) interval ground snowfall is used as the basis for the snow hazard and is augmented based on the height of the roof and the presence of roof features such as parapets that can allow for snow drifts to form. Additionally, there are factors such as rain or snow surcharge that should be taken into account because the snow traps rainwater, causing an increase in the density of the snow. U.S. codes provide a factor that increases the design snow load for higher risk category facilities.

27.5.2 Mitigation Strategies

Addressing rains and snow hazards in new buildings is simply a matter of following the building code and applicable structural design standards, such as ASCE 7. Because of the critical nature of a data center, it is recommended that, as in the case of wind and earthquake, the provisions for Risk Category IV essential facilities be used. It is also important to put in place a maintenance plan that has the roof drains regularly checked for blockage and cleaned. Also to address the issue of power loss creating a "cold roof" condition, the heating system could be placed on emergency power that can operate for at least 3 days.

Many existing buildings that are desirable for data centers, industrial warehouses, and big box-type buildings are the most likely to have roofs that have very little reserve capacity and inadequate consideration of snow drifting or are flexible enough to create ponding instabilities. Therefore, it is essential to evaluate the roof framing of a building during a due diligence study. Augmenting the roof framing can be a very significant cost and will be difficult to accomplish if the building has already been outfitted with the systems, piping, ducts, and cable trays. All of those items will obstruct access to the roof framing. If new rooftop is equipment is planned, consideration of its effect on the roof drainage and the ability of snow to drift adjacent to it should be considered. The equipment may need to be relocated, placed on elevated platforms to allow for drainage, or have the roof locally strengthened under it.

27.6 FLOOD AND TSUNAMI

27.6.1 Hazard Overview

While the mechanisms that cause a flood and a tsunami are different, they are similar hazards and their effects on buildings are similar. Floods and tsunamis are characterized by the uncontrolled flow of water into a region. The force of the water impacting the building can damage the building's façade and, if the water flow is high enough and violent enough, even damage the building structure. Once the building envelop is compromised, water may flow into the building and can damage equipment and leave the building unoccupiable due to debris and environmental hazards such as mold.

Like all environmental hazards, flood and tsunami risk differs from region to region. The magnitude of risk is based on whether the building is situated in an inundation zone. That is a region that, for a given mean recurrence interval or probability of occurrence, might be subject to flooding due to the high water level. Based on the inundation height, it can be determined if the building is located at an elevation where floodwaters will impact it.

U.S. building codes address flood by requiring consideration of a 100-year mean recurrence interval flood. The Federal Emergency Management Agency (FEMA) publishes maps that provide flood inundation zones. If a building is located in an inundation zone, the designer or a consultant to the designer needs to determine if the loads from the floodwater impacting the building are significant. Currently, U.S. building codes do not require higher mean recurrence intervals be considered for essential facilities. There is some debate about this, and professional opinions that the 500-year mean recurrence interval flood should be used for essential Risk Category IV buildings, instead of the 100-year flood.

The Tsunami Loads and Effects is discussed in the ASCE 7 Standard, Minimum Design Loads and Associated Criteria for Buildings and Other Structures. Tsunami risk and inundation maps published by the National Oceanic and Atmospheric Administration (NOAA) and some state agencies, such as the California Department of Conservation, show coastal regions subject to tsunami hazard. Many other countries with coastal regions subjected to tsunami risk also have such hazard maps. To assess the vulnerability of a site, the extent of inundation, height of run-up, and velocity of flow are needed. Where maps are not available for a specific location, site-specific studies can be performed.

27.6.2 Common Effects on Buildings

The most common effect of tsunami and flood on buildings is the inundation of the basement and first floor. The water damage can be significant and immediately renders a building or much of its equipment nonfunctional. For example, the Fukushima Daiichi nuclear disaster was initiated by tsunami waters flooding the rooms housing the emergency generators, rendering the plant without power and unable to maintain the coolant system. Additionally, after the water subsides, there can be issues with mold and other environmental hazards that would need to be mitigated before it is safe for people to reoccupy the building.

As stated previously, the floodwaters flowing into the building can damage the building's façade and even the structure if the floodwaters are high enough and have a

sufficient flow velocity. Weak points in the building envelope such as windows and doors are the most susceptible. In the most extreme cases, a flood or tsunami can produce such a violent flow of water as to literally push a building off its foundation. Debris in the flowing water can also damage buildings significantly.

27.6.3 Mitigation Strategies

The most straightforward way to address the risk of tsunami or flood is simply to not build the data center in the inundation zone or locate a data center in an existing building in an inundation zone. When this is not possible, consideration can be given to the inundation height, the maximum water flow height and velocity, and the presence of surrounding elements, which could become waterborne debris hazards.

In many cases, the flood or tsunami flows will not be very high or violent. In those cases, mitigation may simply involve locating all equipment required for continued operation of the data center at a level above the inundation height. Locating critical equipment in the basement should be avoided. It is common in flood regions to build new buildings on platforms, so the first occupiable floor is above the inundation height. If the data center is of critical importance, consideration of the 500-year as opposed to the 100-year flood inundation should be given.

27.7 COMPREHENSIVE RESILIENCY STRATEGIES

27.7.1 Predisaster Planning

The first step in any natural disaster risk mitigation plan is to understand what the natural disaster hazards are at each facility in the company's inventory. Federal and local government agencies publish information on earthquake, hurricane, flood, snow, and tornado hazards. A simple but effective approach to begin a plan would be to develop a matrix of all the facilities' sites and rank the hazard for earthquake, tornado, hurricane, and flood as high, medium, or low.

Once an understanding of the hazard at each site is known, then the process of assessing the resilience of the different facilities on each site can occur. As noted previously, not all buildings, even those designed to the same code, will perform the same in a natural disaster. The code sets forth a minimum standard, but not explicit performance objectives related to downtime and damage control. Therefore, it is helpful to have a common approach to define building performance.

For any disaster, there are three main concerns related to facility performance: loss of life, physical damage, and downtime following the disaster. These three metrics may or may not be related to each other. For example, a facility could be "Life Safe" but sustain damage to the point of being not economically feasible to repair. Conversely, a building can have very little damage overall, but localized collapsed, resulting in several people killed or injured.

For many buildings, Life Safety is the only performance level that needs to be considered. Most office buildings fall into this category. The lives of the people inside need to be protected, but the building could be significantly, even irreparably damaged from a disaster. The workers may be able to work off-site until a new facility is found or the building is repaired. On the other hand, the building may house a critical, nonredundant data center, or manufacturing operation, the loss of operation of which would cause significant business impact. In those cases, the target performance level may be significantly higher than simply Life Safety, and considerations for minimal postdisaster downtime may need to be addressed.

It is important to understand what the needs are for each facility in the organization's inventory so the right performance level can be selected for each specific building. Standards can be tailored to a company's specific needs and should be agreed upon, at least in concept before any evaluation is to begin. Once the standards are set for which buildings need only to be Life Safe, which ones have critical functions and need to be immediately occupiable, and which ones need some level of damage control, then evaluations of the buildings can be carried out and mitigation strategies developed.

Evaluating building performance for different natural disasters does not have to be a major undertaking for each building. There are methods that can be employed to provide cursory assessments of all the buildings within a portfolio. Following the preliminary evaluation, it can be determined which buildings warrant more in-depth evaluation. Typically, more detailed evaluations are done for critical buildings and for buildings where the cursory evaluation indicates there may be a problem, understanding the range of conservatism built into the cursory evaluation methods.

Following or concurrent with the initial evaluation of the facilities, it is recommended that company-specific natural disaster guidelines be developed. These guidelines should set forth the minimum performance level for each type of facility in the company's inventory. The guidelines can then be used to direct new construction projects, set forth standards for prelease and prepurchase evaluations, and determine which current facilities are not up to the required performance.

It is important to avoid adding buildings to the inventory that represent moderate or high risk. Therefore, the guidelines should be used for all new construction projects and for assessment of any building the company plans to purchase or lease. Depending on the type of facility, the cost of added natural disaster resilience may be very small.

However, without guidelines, building designers are frequently not aware that greater resilience is appropriate for a given facility.

For a typical office building, the structural cost makes up only about 20% of the total building cost, so making the structure more resilient may only add 5% or less to the total building cost. For manufacturing and data centers, the structural cost is an even smaller portion of the total building cost, and thus, added disaster resilience would cost much less as a percentage of the total cost.

When considering acquisition or lease a new building, a proper natural disaster due diligence study should be performed during the due diligence process. The risk study should focus on assessing the building's risk to life safety, damageability, and potential loss of function with respect to any significant natural disaster that may be present at the site. The company's risk guidelines should have a section that addresses acceptable risk levels for owned and leased buildings based on the occupancies and functions of the buildings.

For existing buildings currently in the company's inventory that do not meet the facility standards, there are four options: retrofit, replace, insure, and accept. Retrofitting to bring a facility up to the required performance may require significant structural modifications or may only involve addressing isolated deficiencies or bracing equipment. Structural retrofits can vary from modifying the structure in isolated areas to the addition of exterior buttress, augmenting existing member connections, or even additions of new structural elements to the interior of the building. Nonstructural elements, such as mechanical and electrical equipment, piping and ducts, and architectural elements, may need to be braced so they can stay intact during earthquake shaking or not be blown over by strong winds. Some nonstructural elements may need to be relocated so they are not located in an area that will be inundated with water if a flood occurs. In any retrofit, it is advantageous to perform the work when the building is vacant or in conjunction with a major tenant improvement. For cases in which that is not feasible, a retrofit can be designed to minimize the amount of temporary relocation and be constructed in phases or the new structural elements concentrated at the exterior of the building.

In some instance, the cost of a retrofit may be excessive and approach that of building a new facility. In certain cases, that may still be the appropriate business decision. In other cases, several options should be explored. One is to build a new, disaster-resilient facility. The other might be to build a second facility in another location, which can create sufficient redundancy so the loss of one does not significantly impact the company's business operations.

The last two options—insure and accept—are predicated on the cost of retrofit or replacement being too large to justify in conjunction to the risk exposure. Natural hazard insurance can be costly but in some cases may be sufficient to mitigate the natural disaster risk. In some cases, the time between the disaster and when the insurance claim is fully paid can be quite long. On the other hand, if the facility is redundant and does not pose a threat to the lives of the people inside, the company may choose to accept the risk and self-insure.

27.7.2 Postdisaster Planning

The moments after a major natural disaster can be chaotic. However, a well-developed postdisaster plan can serve to allow the immediate recovery to begin in spite of the chaos. There are a few important concepts that every postdisaster plan should have. The first one is to educate all employees to protect themselves during the disaster. For example, it is common for people to run out of the building during an earthquake. However, the more appropriate response, advocated by the FEMA, among others, is to drop, cover, and hold. The second is having on-site personnel trained on how to properly inspect buildings to determine if there are any glaring safety hazards. The default position should be to evacuate and wait for an engineer or building official to evaluate the building to determine it is safe.

It can take weeks for a jurisdiction to inspect a specific facility. This is because the demands on the local building department, even when bolstered by volunteer engineers, are so great that response times are unpredictable. Additionally, finding a consulting engineer to hire may be difficult because of the increased demands on their time due to the disaster. Therefore, it is important to have in place prearranged retainer agreements with an engineer who can inspect the facility or multiple facilities. It is ideal for the retained engineer to have previously evaluated the facilities so as to have an understanding of them and where the potential damaged areas may be. This will make their evaluation much more effective and can also be used to pretrain the on-site personnel for specific hazards to be aware of.

In San Francisco, following the 1989 Loma Prieta Earthquake, a program was enacted in conjunction with the Structural Engineers Association of Northern California called the Building Operation Resumption Program (BORP). In this program, the building owner contracts with the evaluating engineer, who then prepares a postearthquake inspection plan that is submitted to the jurisdiction. The city officials then approve the plan and that engineer is registered and required to post the safety rating of the building within 3 days of the disaster: Green, Safe for Reoccupancy; Yellow, Only safe for limited reoccupancy by trained personnel; and Red, Unsafe. While other cities do not have a specific program like the BORP, many have been willing to adopt building-specific BORP-like programs if the building owner brings a proposed program to the building official or planning department.

REFERENCES

[1] American Society of Civil Engineers. *Minimum Design Loads for Buildings and Other Structures*. Reston: ASCE; 2016. ASCE/SEI 7-16.

[2] Applied Technology Council (ATC). Seismic Performance Assessment of Buildings, Methodology and Implementation, FEMA P-58-1, developed for the Federal Emergency Management Agency, Washington, DC; 2019.

[3] Building Seismic Safety Council (BSSC). NEHRP Recommended Seismic Provisions for New Buildings and Other Structures, FEMA 750, developed for the Federal Emergency Management Agency, Washington, DC; 2015.

[4] American Society of Civil Engineers. *Seismic Evaluation and Retrofit of Existing Buildings*. Reston: ASCE; 2017. ASCE/SEI 41-17.

28

FIRE PROTECTION AND LIFE SAFETY DESIGN IN DATA CENTERS

Sean S. Donohue[1], Mark Suski[2] and Christopher Chen[3]

[1] *Jensen Hughes, Colorado Springs, Colorado, United States of America*
[2] *Jensen Hughes, Lincolnshire, Illinois, United States of America*
[3] *Jensen Hughes, College Park, Maryland, United States of America*

28.1 FIRE PROTECTION FUNDAMENTALS

Fire is a risk every business must deal with. For data and telecommunication centers, that risk includes not only the safety of people in the building but also the continuity of operations and the value of the equipment and data. Today, these centers are the nervous system of businesses and organizations throughout the world, and the more critical the site, the less acceptable the risk of interruption or downtime. Fire protection comes in many forms, but the goals are simple:

1. Construct buildings and systems that guide people away from and protect them from harm.
2. Give the user and responders accurate information in order to make informed decisions.
3. Limit loss (life, downtime, equipment, data, or other).

This chapter will discuss life safety and active and passive fire protection and will present the choices available to the designers typically used in data centers.

28.1.1 Fire and Data Centers

Electronic equipment and data center rooms contain a variety of combustible fuel, from printed circuit boards to wiring insulation and cabinet enclosures, which increasingly contain more and more plastic. Furnishings, backboards, batteries, and floor tiles also contribute to the overall fuel load.

In recent years, the trend has been to increase the rack power consumption density. With increased power density comes more heat and a higher risk of overheating if the ventilation systems cannot keep up. From a risk standpoint, it is critical to maintain good housekeeping within data centers and remove furnishings, paper, or other combustible load that does not contribute to the core function of the data center. Batteries and nonessential equipment should be housed in a separate room, if possible.

When electronic equipment combusts, it produces many different gases generically referred to as smoke or products of combustion. These can include corrosive gases such as HCN and HCl that can do more damage to printed circuit boards than the heat from a fire. Because of this, early detection is often desired so that staff can respond to an incipient condition before it becomes an emergency. Detection systems can continue to alert occupants to developing stages of a fire and can be programmed to provide suppression system activation.

When a fire grows beyond the ability of occupants to control, an automatic fire suppression system can extinguish or control the fire until the fire department arrives and completes extinguishment. Many buildings are required by building codes to be equipped with automatic fire sprinkler systems, based on the size and use of the building. Gaseous fire extinguishing systems are also used as alternatives to sprinklers, when permitted by the local authority having jurisdiction (AHJ).

The prime differentiator between the two systems is that sprinkler protection is considered a life safety system because it (very often) contains a fire to its room of origin, limits fire spread, and protects the remainder of the building, whereas a gaseous system is considered as equipment protection because it mitigates a specific loss other than life.

Table 28.1 illustrates how a fire in a data center may develop.

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

TABLE 28.1 Stages of fire growth for an electrical fire in a data center

Fire growth stage	Description	Possible response
Incipient	Overheating of equipment/circuits; trace amounts of combustion gases equal to lowest amount detectable by an aspirating system. No other detection	Occupant alert Occupant action Pre-alarm signal
Smoldering (visible smoke)	Increased burning, detectable by human smell. Activation of spot-type smoke detection. Highest alert level for aspirating systems	Occupant action Pre-alarm signal
Flaming	Pyrolysis and flaming combustion. Activation of multiple spot-type detectors. Increased room temperature and development of an upper gas layer	Fire alarm Shut down HVAC Initiate clean agent system countdown or release solenoid valve in pre-action system
Fire growth/spread	Copious production of smoke in quantities sufficient to quickly activate multiple spot detectors. Rapid acceleration of heat release and fusing of nearest sprinkler	Fire alarm Sprinkler system or gaseous system discharge Fire/smoke dampers close

28.2 AHJS, CODES, AND STANDARDS

The term AHJ is often misconstrued to mean a government entity enforcing statutory or regulatory fire and/or life safety requirements within the site's geographic or jurisdictional area. While this is a group that is certainly included, an AHJ can be any public or private entity to which ownership is subject to and can include the following:

- Local, state, or federal authorities.
- Insurance companies.
- Ownership (self-regulation).
- Industry groups.

These groups either adopt national standards that address construction requirements or create their own. They also replicate much of the required compliance information, so the provisions will be similar but not always the same. For example, the Telecommunications Industry Association (TIA) Tier III requirements mirror FM Global requirements for 1-hour rated rooms, whereas the building code does not. Sometimes, requirements can conflict so it is important to understand the priority. Statutory code requirements are legally required, insurance guidelines can have a financial impact, and ownership or industry guidelines are a matter of internal policy.

28.3 LOCAL AUTHORITIES, NATIONAL CODES, AND STANDARDS

Data centers are highly specialized spaces with extensive technical demands, yet they represent a small percentage of what a typical jurisdiction reviews or inspects. As with any specialized systems, it is important to communicate with authorities early in the design process because requirements may not be the same from one jurisdiction to another. This is true for site location, construction, power, ventilation, and fire protection among other requirements. Information that is typically available online or can be attained by contacting the planning, building, or fire department includes the following:

- Geographic area of jurisdiction.
- Adopted codes and standards.
- Amendments and local policies.
- Special interpretations.

The local code reviewer will typically appreciate the designer contacting them early for a special project. For jurisdictions that are not as easily approached, a local designer may need to be brought on to assist the team.

In the United States, the International Building Code (IBC) [1] and the International Fire Code (IFC) [2] apply in most jurisdictions as a base construction and maintenance code. Smaller rural jurisdictions will tend to adopt a code with little modifications, whereas large jurisdictions and cities will more heavily amend the code. An early code review is critical to ensure the design team understands all local constraints. An installation that was used in one location cannot necessarily be repeated in another.

The National Fire Protection Association (NFPA) publishes hundreds of standards addressing topics ranging from storage of flammable and combustible liquids to protective gear for firefighters. NFPA standards that apply to data centers and are referenced by the IBC and/or IFC include:

- NFPA 10: *Standard for Portable Fire Extinguishers*.
- NFPA 12: *Standard on Carbon Dioxide Extinguishing Systems*.
- NFPA 12A: *Standard on Halon 1301 Fire Extinguishing Systems*.

- NFPA 13: *Standard for the Installation of Sprinkler Systems*.
- NFPA 20: *Standard for the Installation of Stationary Fire Pumps for Fire Protection*.
- NFPA 70: *National Electrical Code*® (NEC).
- NFPA 72: *National Fire Alarm and Signaling Code*®.
- NFPA 750: *Standard on Water Mist Fire Protection Systems*.
- NFPA 2001: *Standard on Clean Agent Fire Extinguishing Systems*.
- NFPA 2010: *Standard for Fixed Aerosol Fire Extinguishing Systems*.

Additional standards that are not referenced in the IBC or IFC, but are applicable to the data center and telecommunications industry, include:

- NFPA 75: *Standard for the Fire Protection of Information Technology Equipment*.
- NFPA 76: *Standard for the Fire Protection of Telecommunications Facilities*.
- NFPA 101®: *Life Safety Code*® (LSC).

NFPA 75 [3], for example, covers active and passive protection and risk analysis. As of this publication, NFPA 75 is not referenced by IBC or NFPA 101; therefore it is not enforceable unless specifically adopted. It is referenced by the NEC in Article 645, but not as a required standard; therefore, designers must choose to use this standard unless required by some other AHJ. Among other provisions, NFPA 75 requires fire separation of IT rooms, sprinkler protection if the room is located within a building that requires sprinkler protection, and automatic detection.

An NFPA standard in development as of the publication of this handbook is NFPA 770: *Standard on Hybrid (Water and Inert Gas) Fire Extinguishing Systems*. This newer technology will be discussed later in this chapter.

28.3.1 Insurance Companies

The goals of insurance companies are clear: mitigate loss, reduce risks, and ensure business continuity. In order to meet these goals and keep premiums low, insurance companies will often place requirements on their customers. Some companies, such as FM Global, have created their own list of standards known as FM Data Sheets. Examples include FM Data Sheet 5021, Electronic Data Processing Systems, or FM Data Sheet 4-9, Clean Agent Fire Extinguishing Systems. These data sheets prescribe compliance, which may exceed that found in building and/or LSC.

The user should be aware that ownership may be held to these standards in the future and should incorporate any discrepancies into the design.

28.3.2 Ownership Standards

Ownership (e.g. the U.S. federal and state governments and large companies) may have specific requirements that exceed code or insurance requirements for their own protection including and many times based on their own experience with previous installations or loss. Some examples include the following:

- No wet piping above the data center.
- Security measures that must still allow code compliant egress.

28.3.3 Rated or Tiered System

Lastly, industry groups such as the TIA [4] and the Uptime Institute (UI) [5] have published standards based on a rated or tiered system that affect, among other requirements, passive fire protection, fire detection, and suppression. A rated or tiered system describes various levels of availability and security for the data center infrastructure; the higher the tier, the stricter the requirement. UI's Tier I and II facilities are typically only required to meet minimum code requirements, whereas Tier III and IV facilities exceed minimum code requirements. For example, Tier III and IV facilities may require both sprinkler and clean agent fire suppression, while Tier I and II facilities do not specify clean agent systems. Standards such as UI and ANSI/TIA 942-B-2017, *Telecommunications Infrastructure Standard for Data Centers*, should be consulted for additional information.

28.4 LIFE SAFETY

The first goal of building and fire codes is to safeguard the lives of people within a building. When it comes to data center layouts, rooms are typically configured to support the data processing equipment and processes over comfort of the occupants. As the need for larger and more capable data centers have increased, more and more are being designed with the ability to be maintained remotely, and only minimal on-site supervision is required. In either case, the building and LSC address the typical life safety concerns appropriate for the intended use of the building. The following are some highlights that will assist the designer in addressing occupant-specific code requirements, based on the IBC and LSC (NFPA 101).

28.4.1 Occupancy Classification

Occupancy classification describes the use or function of a space and sets in motion different requirements for different hazards. For example, a tire storage warehouse will have very different construction and fire protection requirements

than a hospital. Data centers have historically been defined as business occupancies because of their accessory function to a business that employs people and provides services. As data centers evolve into larger more autonomous buildings, it could be conceivable to consider these as storage occupancies, because of their function of providing *data storage*. However, a business occupancy is still the more common classification.

28.4.2 Occupant Load

Occupant load refers to the number of people the code considers to be in a space at the same time. This is a conservative number meant to represent a "worst-case" scenario and normally cannot be exceeded. It is common, for example, to see a posted occupant load in assembly venues such as theaters and large meeting rooms. Occupant load is used to determine egress width, number of exits, plumbing fixture count, and ventilation. However, it should be noted that for data centers, ventilation is driven by cooling requirements of the equipment.

Occupant load is derived as a function of gross floor area based on function of the space as follows:

$$\frac{\text{Floor area}\left(\text{m}^2 \text{ or ft}^2\right)}{\text{Occupant load factor}\left(\text{m}^2 \text{ or ft}^2 \text{ / occupant}\right)} = \text{Number of occupants in space}$$

The actual use of the space should be openly discussed to most closely match the highest number of people anticipated during the normal use of the space. The trend for data centers is to employ fewer and fewer on-site personnel. Spaces such as lights out data centers are designed to eliminate personnel entirely, except in emergency circumstances.

The applicable building or LSC should be consulted, but the typical occupant load factor for a data center will range from 9.3 gross m^2 (100 gross ft^2) per occupant to 46.5 gross m^2 (500 gross ft^2) per occupant depending on the occupant density. The designation of "gross" indicates that the entire floor area must be used to calculate the occupant load including the following:

- Space used by equipment.
- Interior walls and columns.
- Supporting spaces such as corridors and restrooms.

28.4.3 Egress

Building and LSC should be consulted for the full set of requirements regarding egress. Examples include the IBC and LSC [6] (NFPA 101). A few of the more common egress design concerns are presented below.

28.4.3.1 Number of Exits

It is critical to ensure that there are ample exits in a room and building. All occupied rooms require at least one means of egress. A business occupancy will require a second means of egress when the occupant load is greater than 50. Using a conservative occupant load factor of 13.9 m^2 (150 ft^2) per occupant, the designer should be concerned about a second exit when a data center exceeds 697 m^2 (7,500 ft^2). If a data center is equipped with sprinkler protection, the distance between exits are required to be at least one-third of the diagonal distance of the room itself. If there is no sprinkler protection provided, the distance between exits increases to one-half the diagonal distance of the room. When the occupant load exceeds 500, a minimum of three exits are required. When the occupant load exceeds 1,000, a minimum of four exits are required.

28.4.3.2 Egress Width

It is important to provide wide enough doors and stairways for occupants to egress safely in the case of an emergency.

Doors
A 915 mm (36 in) wide door provides at least 813 mm (32 in) of clear width. Using a width capacity factor of 5 mm (0.2 in) per occupant required by code, this equates to about 160 occupants per door. For code compliant purposes, assuming 13.9 m^2 (150 ft^2) per occupant with two exits, the occupant load would need to exceed 320 occupants, or a floor area of 4,460 m^2 (48,000 ft^2), before a width of more than two typical 915 mm (36 in) doors would need to be considered.

Stairways
If the occupant load is less than 50, the minimum staircase width will be 914 mm (36 in). Otherwise, staircases are required to have a minimum width of 1,118 mm (44 in). Stairway is required to rise between a minimum of 102 mm (4 in) and a maximum of 178 mm (7 in). Additionally, there should be at least 2,032 mm (80 in) of headroom. The applicable codes should be consulted for further requirements.

28.4.3.3 Travel Distance

Travel distance is a function of the occupancy type discussed previously and whether or not a building is protected by a sprinkler system. Travel distance is the maximum distance a person should travel before reaching an exit. It is measured to the closest exit from the most remote location in a room and should be measured orthogonally to account for equipment and furnishings. The applicable building or LSC should be consulted for these requirements, but these typically range from 61 to 91 m (200–300 ft).

28.4.3.4 Aisles

Equipment placement is a function of operational needs; however occupants need to fit in between pieces of equipment for maintenance and egress. Based on disability requirements, aisles should be maintained at 812 mm (32 in) clear minimum. In large data centers, primary aisles will need to be larger to accommodate the additional occupant load and number of exits required, but not smaller than 1,118 mm (44 in).

28.5 PASSIVE FIRE PROTECTION

Walls, floors, and ceilings of rooms and buildings are required to be fire-resistance rated for a variety of reasons, in accordance with building codes, including separation of hazards, protection of the means of egress, or to allow the construction of larger buildings. Often, the building code does not require any rating at all, especially in the case of data centers, but the sensitivity of the equipment, process, or data may drive the insurer or owner to require fire-resistance rating as previously discussed. Additional hazards, such as UPS batteries, may require fire-resistance rating per the IFC.

The goal of passive fire protection is to delay the spread of fire from an adjacent space to allow time for egress and to give firefighters time to contain a fire. The higher the hourly rating of the assembly, the higher the thermal resistance. The hourly rating assigned to fire-resistance-rated assemblies should not be construed to imply a guarantee against adjacent fire events for the duration of the stated rating, but represents the minimum time an assembly is capable of resisting a predetermined fire curve. Actual fires may burn cooler or hotter than the ASTM E-119 [7] standard fire curve because heat output is heavily dependent on the type of fuel burning. A fire in a data center, started by the overheating of electrical insulation, could actually smolder for quite some time before developing into a flaming fire, meaning that it would likely not be as severe a fire exposure as the ASTM E-119 standard fire curve.

Typically, 1 or 2 hours assemblies are typically encountered through the model building codes. Some standards, such as TIA, require bearing walls to have a fire-resistance rating as high as 4 hours for Tier IV centers. An example of a 1 hour assembly from the UL Online Certifications Directory is provided in Figure 28.1. This type of wall is one of the more common 1 hour assemblies and is composed of light gauge metal studs, insulation, and 5/8 in type "X" gypsum board. Refer to the full UL listing for complete information concerning all the materials permitted with this assembly.

The designer may consult several sources for examples of assemblies that provide the requisite fire-resistance rating. Popular sources include the IBC, the UL Online Certifications Directory, and the US gypsum manual [8].

Openings in fire-resistance-rated walls such as doors and windows require intrinsic rating, closers, or shutters to maintain the intended fire rating of the room; this is addressed by the building code. Penetrations such as ducts, pipes, and conduit through fire-resistance-rated construction must be protected when the assembly is required to have fire-resistance rating and again, codes and standards address this. Fire and smoke dampers serve to protect the duct penetration into the protected room in case of fire. While fire dampers are activated by a fusible link, smoke dampers are activated via duct-mounted or area smoke detection. It is important to evaluate user goals and ensure that HVAC flow to a room not be shut off unless there truly is a fire.

28.6 ACTIVE FIRE PROTECTION AND SUPPRESSION

Automatic fire suppression is often required in buildings housing data centers; therefore, designers need to be aware of the choices, risks, and costs involved for each type of suppressing agent. Halon 1301 used to be synonymous with data center fire protection for suppressing fires, but the use of that agent is now limited to maintenance of existing locations. A number of chemical extinguishing and inerting agents are offered as alternatives to Halon 1301, although automatic sprinklers still remain a viable option for containing fires of low-risk installations.

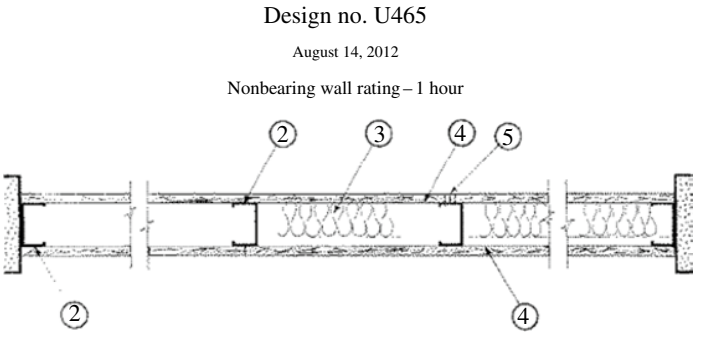


FIGURE 28.1 UL design U465. Source: Courtesy of Underwriters Laboratories. https://standardscatalog.ul.com/ProductDetail.aspx?productId=UL263 accessed 9/20/2020.

28.6.1 Automatic Sprinkler Systems

Water has long been a fire suppressant of choice. It is readily available, relatively inexpensive, and nontoxic and has excellent heat-absorption characteristics. That being said, water is electrically conductive and will damage energized equipment in data centers. However, automatic sprinkler systems are the fire suppression system of choice for the majority of built environments including occupancies that may be located in the same building as a data center.

Sprinkler activation is often misunderstood due to frequent misrepresentation by the entertainment industry. Contrary to popular belief, sprinklers are only activated by thermal response (not smoke) and activate individually one at a time. Although there are many types of thermal elements, the most popular one is the frangible glass bulb. A bulb filled with a proprietary glycerin-based fluid keeps pressurized water, air, or nitrogen from being released. When the fluid in the bulb reaches a predetermined temperature, it expands to fill the volume and breaks the glass bulb enclosure. Water or air then escapes the piping network through the opening created (Fig. 28.2).

Due to the excellent track record sprinklers have achieved in controlling the spread of fire, current building codes offer many incentives when designers specify sprinkler protection, including the following:

- Larger buildings.
- More lenient egress requirements.
- Less restrictive passive fire protection.

It is imperative that design teams discuss the use of sprinkler protection for a building. When the incentives are taken as indicated, sprinklers are required throughout a building, regardless of whether an alternative system is installed, unless specific omission is permitted by all AHJs.

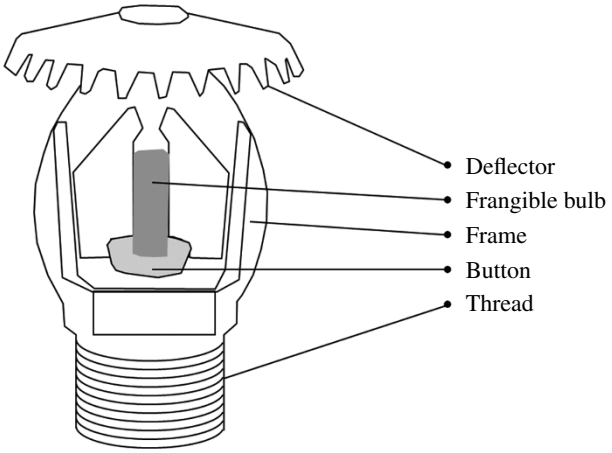


FIGURE 28.2 Standard upright sprinkler.

It is important to note that while damage to data centers from water is a concern, redundancies are built in many large data centers. There are more than one data center with the same data in the case of any shutdown, including fire. With redundancies already being built for non-fire-related concerns, the use of cheap water-based sprinkler systems is much more viable.

NFPA 13 covers the installation requirements for sprinkler systems. It should be noted that this standard, along with many of the installation standards promulgated by the NFPA, tells the user *how* to install a system and its components. Building and LSC tell the designer *when* these systems are required.

When sprinkler protection is required, it is required in all occupied spaces. It may also be required in accessible interstitial spaces depending on the fuel load and combustibility of those spaces. Thought should also be given to how water will drain after actuation. The activation of a sprinkler system can produce several hundred gallons of water before a fire is deemed controlled. Provisions to limit the spread of sprinkler water in a building should be incorporated into the building construction wherever possible.

28.6.1.1 Wet Pipe Sprinkler Systems

As the name implies, a wet pipe sprinkler system is filled with water, which is connected to a water supply system so that water discharges immediately from sprinklers activated by heat from a fire. Wet pipe sprinkler systems are the most common of sprinkler systems because of their simplicity and reliability. With low installation and maintenance costs, they make up over 80% of sprinkler systems installed [9].

Buildings in which wet systems are installed must be maintained at or above 4.4°C (40°F) and should be coordinated to avoid proximity to cooling systems operating below this threshold temperature.

Wet pipe sprinkler systems are not typically used in data centers but can be used where the risk of loss due to accidental release of water is low and/or the cost of system installation is an issue.

28.6.1.2 Dry Pipe Sprinkler Systems

A dry pipe sprinkler system employs automatic sprinklers attached to a piping network containing pressurized air or nitrogen, rather than water. The activation of sprinkler(s) releases the pressurized air or nitrogen and permits water pressure to open a dry pipe valve, allowing water to flow into the piping network and out of the activated sprinkler(s).

Dry pipe systems are typically reserved for unheated buildings or portions of buildings where the ambient temperature is not maintained at or above 4.4°C (40°F). This prevents the freezing of water found in wet pipe systems. Since data centers are normally maintained to approximately

20–23°C (68–73°F), dry pipe sprinkler systems are rarely used in this situation. Designers will encounter dry pipe systems in exterior environments such as docks, garages, or canopies in colder climates.

28.6.1.3 Pre-Action Sprinkler Systems

A pre-action sprinkler system utilizes automatic sprinklers attached to a piping network that contains air or nitrogen that may or may not be under pressure. A supplemental detection system is installed in the same area as the sprinklers. The pre-action system requires the prior activation of a detection system to open a control valve and allow water into the piping. This is most typically accomplished with smoke or heat detection but can be done with any fire alarm signal, including a manual fire alarm box or pull station.

Pre-action systems are frequently specified for data centers because they reduce the risk of accidental non-fire release of water over the electronic equipment. They may be installed to incorporate additional strategies to help prevent accidental discharge. There are three fundamental types of pre-action sprinkler systems, as discussed below.

Non-Interlocking
A non-interlocking system utilizes a deluge valve that is released by the activation of either a sprinkler or a detection system such as a smoke detector. Without detector activation, the system behaves like a dry pipe system. If a sprinkler activates, water will discharge. If a detector activates, the piping network will fill with water, but no water will discharge until a sprinkler activates.

Single Interlock
A single interlock system requires the activation of a detection system to operate a solenoid valve. The solenoid valve is an electronic valve that holds back water from entering the sprinkler piping network. Once the solenoid valve opens, water will enter the sprinkler piping network, but discharge will not occur until activation of a sprinkler. Operation of the solenoid valve alone turns the system into a wet pipe sprinkler system.

Single interlock systems provide an extra fail-safe over non-interlocking systems. Discharge requires both the activation of the sprinkler system and the detection system.

Double Interlock
A double interlock system requires two events to occur before water is released into the sprinkler piping network. Both the activation of a detection system and the activation of a sprinkler are required. Both the solenoid valve and the deluge valve must open for water to be admitted. Operation of the solenoid valve alone turns the system into a dry pipe sprinkler system.

Double interlock systems have the most complex fail-safe mechanisms. Water does not enter the piping network until activation of both the sprinkler and the detection system. The application of this type of system includes conditions in which it would be hazardous to have water in the piping for an extended amount of time such as an unoccupied or remote site where response will be delayed or for sites that cannot tolerate overhead water except in an emergency condition. This type of system is beneficial in data centers where condensation from the sprinkler piping network could drip onto and damage electrical equipment.

Both non-interlock and single interlock systems admit water; therefore, the sprinkler system could remain charged for some time before it is drained and reset. A double interlock system will not admit water into piping until a sprinkler is activated. Water never sits in the sprinkler piping network. Because of this reduced risk and relatively lower cost, double interlock pre-action systems are used most often, especially in large data centers.

28.6.1.4 Deluge Sprinkler Systems

A deluge sprinkler system utilizes sprinklers that are always open (with no fusible link or frangible bulb). A deluge valve is used to hold back water from entering the sprinkler piping network until a detector, placed in the same location as the sprinklers, is activated. Upon activation, the deluge valve releases water into the sprinkler piping network and is immediately discharged through all the sprinklers connected to the system.

Deluge systems are not ideal for data centers as they discharge large quantities of water in a short amount of time. While deluge sprinkler systems can require two detectors in different zones to be activated before releasing water to prevent accidental discharge, the risk of damage and downtime from discharge, accidental or not, is still too great.

28.6.1.5 Galvanized Piping

Although galvanized piping has historically been used in dry and pre-action sprinkler system piping, a study [10] suggests that galvanized steel corrodes more aggressively at localized points in the piping compared with unprotected steel that corrodes over a more uniform distribution. This can result in pinhole leaks in locations that are precisely designed to avoid water except in a fire condition.

When an air compressor is used, oxygen is continually fed into the system as the system maintains pressure, interacting with trapped water to corrode the system from the inside out. To combat the effects of corrosion, nitrogen or "dry air" can be used in lieu of air. When using a nitrogen inerting system, the same study suggests that corrosion is virtually halted and that performance between galvanized and black pipe is roughly identical.

28.6.2 Water Mist

Water mist systems are based on the principle that water is atomized to a droplet size of no larger than 1 mm (0.04 in). The large surface area to mass ratio results in a highly efficient heat transfer between hot gases and the water droplets, and a large amount of heat is absorbed with a relatively small amount of water. Water mist systems were initially researched in the 1950s as "fine water sprays" [11, 12] but resurfaced in the 1990s in response to the search for halon system alternatives.

Water mist systems have been tested and approved for use in computer room subfloors and for in-cabinet suppression systems. One advantage is that a properly designed water mist system can achieve fire suppression equivalent to standard sprinklers, but only using 1/3 of the water or less. Therefore, if a fire occurs, the collateral damage that could be caused by discharged water may be reduced, compared with sprinklers. However, accumulated water droplets are conductive, and accumulated moisture on circuit boards will cause problems for electronics. Where electronic equipment is likely to suffer irreversible damage due to water deposition, clean agent suppression systems are typically preferred over water mist.

Water mist systems utilize higher pressure than standard sprinkler systems to generate the small water droplets required to create the mist. Water mist systems are designed to operate at pressures of anywhere between 175 psi (12 bar) and 2,300 psi (158 bar). Pressures in this range require positive displacement pumps and high-pressure stainless steel tubing and incur much higher materials and installation costs than standard sprinkler systems. Corrosion resistant tubing, such as stainless steel, and low-micron media filters are critical to prevent the plugging of small orifices in water mist nozzles.

NFPA 750 [13] should be consulted for the application of these types of systems. The most important requirement of NFPA 750 is that the water mist system design must be based on fire testing to a test protocol that matches the actual application. Therefore, a water mist system that has been tested and approved for machinery spaces would not be approved for application in a data center.

The current status of water mist systems is accurately summarized by J.R. Mawhinney as follows: "Although FM Global has shown support for the use of water mist for telecommunication central offices, general acceptance by end users has been slow in North America. Similarly, the use of water mist as a halon replacement for computer rooms has been mixed. The fundamental issue has to do with comparing the performance of total flooding gaseous agents that can penetrate into electronic cabinets with water mist that cannot extinguish a fire inside a cabinet, at least not in a total compartment flooding mode." [14]

28.6.3 Clean Agents and Gaseous Fire Suppression

Water and electricity don't mix, and today's centers are so critical that they often need to keep running even during a fire event. Since the early 1900s several forms of gaseous fire suppression have been explored and tested to the point that their use is now very well documented and well understood. A gaseous fire suppression system acts on several branches of the fire tetrahedron (Fig. 28.3).

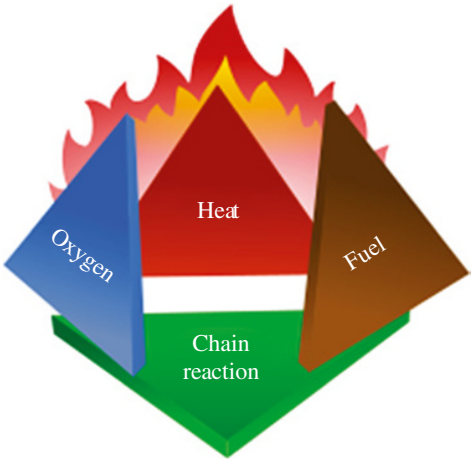


FIGURE 28.3 The fire tetrahedron. Source: "The Process of Combustion," https://www.como.gov/fire/wp-content/uploads/sites/26/2018/08/Fire-Safety-7th-Grade-presentation-with-investigation.pdf.

The process of combustion. https://www.como.gov/fire/wp-content/uploads/sites/26/2018/08/Fire-Safety-7th-Grade-presentation-with-investigation.pdf

Primarily, most agents displace oxygen, which slows down or halts the combustion process. A high specific heat allows many agents to remove heat, which would otherwise continue to accelerate combustion. Lastly, a more recent discovery has been the ability of some agents to interrupt the flame chain reaction. In reality, all three modes work together to suppress fire.

Clean agents are covered by NFPA 2001 [15], and the definition of a clean agent states that the material be electrically nonconductive and nontoxic at concentrations needed for suppressing a fire and does not leave a residue upon evaporation. These properties make clean agents very attractive from an owner standpoint because these systems can protect a data center while allowing a relatively quick resumption of operations after an event.

Clean agents are typically designed as a total flooding system meaning that, upon detection or manual activation, a system will discharge the contents of pressurized cylinders into the protected hazard (which is a defined volume) to create a predesigned concentration necessary to extinguish a fire. Clean agents can also be used in manual fire extinguishers for local streaming applications.

Here are a few design considerations to be aware of regarding a total flooding application:

1. Total flooding is a one-shot approach; once the agent has been discharged, it will either extinguish the fire or it won't. If a fire begins to grow again, another extinguishing

method will be required, such as an automatic sprinkler system or manual suppression. As previously stated, gaseous extinguishing is not a replacement for sprinkler protection.

2. NFPA 2001 requires that discharge be designed such that the agent reaches at least 95% of the design concentration within 10 seconds with a 35% safety factor for Class C (electrical) hazards and within 60 seconds for inert agents. This can require high-pressure systems depending on the agent. Systems require regular maintenance to ensure pressure requirements.
3. At least 85% of the design concentration must be maintained in the protected space for 10 min unless otherwise approved. This means either an extended discharge of agent or a very tight room that can maintain the concentration. Additional agent discharge is sometimes used to maintain mechanical mixing for the heavier agents.
4. Not meeting the room leakage requirement is one of the top modes of failure during commissioning. Further, rooms can be compromised by future construction and penetrations.
5. Most clean agents are super pressurized in liquid phase in containers and then expand to gas in the piping network under discharge. The design of these systems requires a balanced hydraulic design for two-phase flow. Many manufacturers provide their own proprietary software and design services as part of the installation costs.
6. A design involving HVAC shutdown can include the protected room as well as any above-ceiling or below-floor volumes. The volume perimeter must be established and maintained via tight construction and dampers.
7. It is critical that the positive and negative pressures associated with a clean agent system discharge be considered during the design phase to avoid structural damage to the protected hazard enclosure. It may be necessary to install pressure relief vents that automatically open during system activation.
8. In order to maintain cooling load, shutdown may not be desired; therefore, designs can include the air-handling equipment and ductwork within the design volume of the protected space; however, the following should additionally be considered:
   a. Agent volume must be increased to include the volume of and leakage through the air-handling equipment and associated ductwork. This can be a substantial increase.
   b. System sequence must be modified to keep the equipment running with outside air dampers closed to maintain concentration; the equipment should not need to be rated for elevated temperatures since the agent will prevent combustion throughout. A chief benefit to this approach is that the room does not lose cooling during an event and the air-handling system provides the necessary mechanical mixing to maintain concentration.

There are many gaseous extinguishing agents on the market, each making independent claims of superiority. The best way to become educated regarding specific applications is to consult with manufacturers, vendors, or fire protection consultants. Although it is not possible to cover each agent in use today, the four agents below are discussed due to their popularity and historical significance.

28.6.3.1 Halon

Halon refers to a family of chemical compounds using halogens (predominantly fluorine, chlorine, and bromine) to replace the hydrogen atoms in a typical hydrocarbon structure. NFPA 12A [16] covers halon system installation requirements. Halon 1301 is to this day one of the best clean agents ever discovered because it requires such a low extinguishing concentration, provides deep vapor penetration, and doesn't damage data center equipment. Its biggest drawback, however, is its classification as an ozone depleting agent. Since the issuance of the Montreal Protocol in 1987 and decree in 1993 by the United Nations, halon has not been a viable choice for new installations.

That being said, the installation and maintenance of Halon 1301 systems are not banned, only its manufacture. Therefore, designers may encounter existing systems with Halon 1301 from time to time. In fact, several vendors worldwide buy, store, and sell halon for reuse. Although many of these vendors can be readily found online, the price for reclaiming halon fluctuates wildly and can be very expensive. Furthermore, there is a limited availability of replacement parts for halon suppression systems. When a halon system discharges, the owner faces a difficult decision, whether to pay for replacement halon cylinders or replace the system with a similar clean agent.

When modifying an existing data center protected by a halon system, the right choice may be to modify the system in lieu of replacing it.

28.6.3.2 HFCs

In response to the restriction placed on halon by the Montreal Protocol, DuPont developed a family of hydrofluorocarbon (HFC)-based clean agents that have a zero ozone depletion potential; the first of which was HFC-227ea, also known as FM-200. Other common HFC agents include HFC-125, sold under the trade name Ecaro-25 and HFC-23, also known as FE-13.

FM-200 is not as efficient at extinguishing fire as halon, requiring, on average, twice as much agent for the same hazard. It is also heavier than halon requiring additional

mechanical mixing to maintain concentration. Due to the larger quantities required, existing piping systems and nozzles designed for halon cannot be reused. However, FM-200 and its family of other HFC agents have remained a popular alternative because of their similar attributes.

It is important to note that a critical downside to using a halogenated clean agent is the possible production of hydrogen fluoride (HF), which is a byproduct of HFC thermal decomposition. This means that if a fire is not knocked out quickly, the extinguishing agent could break down under heat. When reacting with water, HF turns into hydrofluoric acid, which is corrosive and highly toxic. For this reason alone, it is imperative that the agent achieve a quick knockdown through conscientious deigns and that occupied spaces be evacuated during system discharge.

Recently, HFCs like FM-200 have come under fire for their global warming potential (GWP) and are being monitored under the Kyoto Protocol. However, HFCs used in fire extinguishing systems have received different treatment than use in other applications such as refrigeration. Many environmental protection agencies, such as the US Environmental Protection Agency, agree that HFCs used in fire protection are negligible when compared with their use in other applications. In fire protection systems, the HFC is in a closed system unless it is discharged to extinguish a fire. While the United Kingdom and European Union have begun reducing the use and application of HFCs, their use in the fire suppression industry is only under monitorization. However, this monitorization still requires licensed technicians and special disposal methods when working with HFCs. In conclusion, it can be expected that regulations on fire suppression use of HFCs will only become more stringent as climate change matures into a global concern.

28.6.3.3 Inert Gases

Inergen is one of the more widely used of the inert gases, so-called because it is primarily made up of physiologically inert species including 52% nitrogen and 40% argon. In addition, around 8% carbon dioxide is added to increase breathing rate and oxygen absorption. Other inert gases include Argonite and Proinert, both of which offer a 50–50 blend of nitrogen and argon.

Inert gases suppress fire through oxygen depletion. The discharge of inert gas reduces oxygen concentration to roughly 12.5%, where flaming ignition is no longer supported. Normally, at this concentration, the physiological effects of hypoxia include confusion and loss of mental response; however a small amount of carbon dioxide has been shown [17] to allow occupants to function for a period of time necessary for egress. The hypoxia effects dissipate when the occupant is introduced to normal atmospheric conditions. Although Argonite and Proinert do not introduce carbon dioxide to the system upon discharge, the lowered oxygen concentration is high enough to allow for egress of occupants though nonpermanent effects of hypoxia can be expected.

Inert gases are relatively inexpensive when compared with other clean agents and, because they are made up of inert compounds, will not break down into hazardous species or harm the environment. However, because inert gases are so light and requires such a high extinguishing concentration, the volume of agent required is among the highest. Some hidden costs include the numerous cylinders, movement of the cylinders, and, most importantly, real estate required to store the cylinders. Inert gases also require the highest delivery pressure among clean agents, increasing the cost of piping and proper pressure ventilation. Lastly, a long-term cost that should be taken into consideration for inert gases is the regular hydro testing of the system including cylinders, hoses, and piping. Due to the high pressures required, the system must be taken out of service for extended periods of time.

28.6.3.4 Novec 1230

Novec 1230 was released in 2004 by 3 M as a new type of clean agent known as a fluoroketone. While Halon 1301 came under scrutiny in the 1990s for its ozone depletion potential, the HFCs such as FM-200 have come under fire for enhancing global warming. Novec 1230 goes by the technical designation of FK-5-1-12. It has a zero ozone depletion potential due to its lack of bromine and advertises a GWP of less than 1. Compared with FM-200's GWP of 3,350, the risk is relatively low that Novec 1230 will find itself under the reduction challenges that HFCs are currently facing.

Unlike other clean agents, Novec 1230 does not extinguish fire by removing or displacing oxygen, but rather removes energy from fire by absorbing heat. Design concentrations range from 4 to 8% depending on type of fuel. For data centers, the typical design concentration is approximately 4.2%. Pre-engineered systems are designed to meet the NFPA 2001 requirement of achieving 95% of design concentration in 10 seconds. Although Novec 1230 is a liquid at standard atmosphere and temperature, it readily vaporizes when discharged. As a heavy gas, it requires high discharge velocities and mechanical mixing during the 10 second release to achieve uniform concentration throughout the enclosure.

A primary benefit of Novec 1230 is that, stored as a liquid, it can be transported via air and can be hand pumped from one container to another without significant agent loss to atmosphere. All other clean agents must be transported over ground and delivered under pressure.

28.6.4 Fixed Aerosol Fire-Extinguishing

Fixed aerosol fire extinguishing systems, regulated by NFPA 2010 [18], may appear to be another option for fire protection.

When aerosol extinguishing is used, it is typically in a total flooding application to disrupt the chain reaction of fire. Free radicals from potassium found in the aerosol bind with the free radicals of a fire, including hydroxide, hydrogen, and oxygen. This bonding stabilizes the free radicals that would normally react with the fire. This process also removes heat from the fire.

At first glance, fixed aerosol extinguishing may appear to be a cost-effective option since it is a self-contained system. It does not require piping nor does it require pressurization of the aerosol agent. Additionally, since aerosol canisters are mounted to the walls, they do not take up any meaningful real estate.

However, there are several disadvantages to using a fixed aerosol extinguishing system. The primary concerns are listed below:

1. The small particles created during discharge can enter occupant's lungs and are generally toxic. This is a concern for occupiable, non-remote data centers.
2. Unlike clean agents, aerosol particles leave some residue in the area of discharge, especially in small, tight, and difficult-to-clean spaces. This causes process downtime for cleanup effort.
3. Fixed aerosol systems generate heat at the canister during a discharge. Activating the chemicals within the canisters is an exothermic reaction, which creates high temperatures near the discharging portion of the canisters.
4. As with clean agents, a fixed aerosol system is a one-shot principle. An electric, thermal, or manual signal actuates the aerosol canisters, releasing aerosol. Once actuated, the aerosol discharges until it is emptied. If the aerosol does not extinguish the fire, a different suppression system, such as pre-action sprinklers, will still be necessary as a precaution. Because of its one-shot principle, aerosol suppression is more commonly used in small server rooms rather than large data centers.

While fixed aerosol extinguishing may be a cost-effective option, there are many risks that should be seriously considered.

28.6.5 Hypoxic Air

An emerging technology that is being used sparingly in some countries is hypoxic air, also known as an oxygen reduction system. A compressor/membrane system is used to reduce the concentration of oxygen in the air to approximately 14%. At that concentration, studies [19] have shown that flaming ignition of most materials can be prevented. The low oxygen content, however, causes safety and health concerns over prolonged exposure.

Unlike other agents, hypoxic air is not discharged during a fire condition, but is the constant atmosphere maintained within the protected space. The chief benefit is that there is no need for integration of fire detection to initiate the system. Also, by reducing the oxidation process, products of combustion are not produced in the same quantity or at the same rate as they are in a normal 21% oxygen environment. Lastly, a hypoxic air system may have lower installation costs than the standard piping or cylinder networks needed for a clean agent system.

Conversely, hypoxic air has not been accepted in the United States, and there are no US standards that cover its use as a fire suppression system or extinguishing agent. British Standard BSI PAS 95:2011 does address its use in the United Kingdom. A recent white paper [20] discusses the current state of acceptance for this methodology.

Second, hypoxic air must be constantly generated, so it consumes energy to maintain the low oxygen concentration where other agents are stored in pressurized cylinders until needed. Oxygen sensors must be installed to control the system as the O_2 concentration seeks equilibrium with adjacent spaces. To be economical, the enclosure must be even more "airtight" than is required for total flooding clean agent systems, because the hypoxic air generation rate must exceed the leakage rate from the enclosure. Therefore, service and long-term operational costs must be considered. Furthermore, the systems require many hours to bring the oxygen concentration in a space down from 21% oxygen to 14%. During the hours required to achieve the desired low oxygen level, the space is left unprotected. Lastly, the OSHA requires all the safety precautions associated with confined space entry for personnel who might need to enter the space. It may be necessary to ventilate the enclosure to bring the oxygen level up to at least 19% before anyone can enter the space. It may then require up to 24 hours to reestablish the hypoxic level needed for fire prevention. This extended downtime can result in many negative costs, especially in critical systems.

Hypoxic air is not a viable option in the United States for fire suppression in occupied spaces, but may be an option elsewhere.

28.6.6 In-Cabinet Fire Suppression

In-cabinet fire suppression in the form of clean agents or carbon dioxide (CO_2) is an alternative. CO_2 use is limited because concentrations needed for extinguishment exceed human survivability levels. Also, thermal shock to potentially thermally sensitive equipment is a problem due to a discharge temperature of –65°C. Clean agent is the best choice for this application.

Firetrace is an example of a manufacturer of in-cabinet fire extinguishing systems. Firetrace provides two types of systems: direct and indirect. The direct system uses

heat-rupturable tubes that run through specific cabinets, connecting to an extinguishing agent cylinder. When heat is detected along the tube, the specific location will rupture into a nozzle. The rupture allows the pressure to drop and release the extinguishing agent. Similarly, the indirect system runs the heat-rupturable tubes through the cabinets; however there is a built-in nozzle at the top of the cabinet. The rupturing of the tube drops the pressure, indicating a valve to release the extinguishing agent from the built-in nozzle.

Alternatively, there are cabinets with built-in fire extinguishing, such as Vertiv's SmartRow. These cabinets prevent the necessity for extra costs to retrofit cabinets to include a system like Firetrace.

In-cabinet fire extinguishing can be a cost-effective alternative to total flooding systems. For equipment that may be susceptible to higher energy loads and serving as a possible source of ignition, these systems provide a quick knockdown and local application before other systems activate. The downside is that, similar to portable fire extinguishers, these systems cannot maintain any extended concentration. If the equipment is still energized, the source of ignition and fuel has not been removed, and the fire will continue to grow. These systems are best used when the subject equipment can be de-energized as a portion of the sequence of operation and when system activation will be quickly investigated.

28.6.7 Portable Fire Extinguishers

NFPA 10 [21] is the standard governing selection and placement of portable fire extinguishers, regardless of suppressing agent. The standard breaks fire hazards into four categories:

- Class A: Cellulosic combustible materials.
- Class B: Flammable and combustible liquids.
- Class C: Electrical fires.
- Class D: Metal fires.

Data centers represent a combination of Class A and C fire hazards. For sensitive electronic equipment, the standard requires selection from types of extinguishers listed and label for Class C hazards. Dry chemical fire extinguishers are expressly prohibited because the solid powder can irreversibly damage sensitive electronic equipment. That is why the most common portable extinguishers for data centers are those containing clean agents or carbon dioxide.

As with any manual suppression component, a portable fire extinguisher is only as effective as the person using it. Factors that limit the value of portable fire extinguishers in data centers include training, human sensitivity to an incipient stage fire, the length of time to reach an extinguisher, and access to equipment.

28.6.8 Hybrid Fire Suppression Systems

A new emerging technology is the hybrid fire suppression system. Hybrid fire suppression systems use an inert gas along with atomized water to create a "hybrid" system for extinguishing fires. NFPA 770 [22] has been proposed and is being developed to address these systems, but is not expected to be published as a standard until 2021.

Hybrid systems use nitrogen and atomized water droplets to extinguish fires. They are self-contained systems that connect directly to the fire detection system similar to clean agent systems. Upon detection, the hybrid system activates, atomizing and releasing the suppressant. The hybrid mix extinguishes fire by reducing oxygen to approximately 14.5% and cooling the fire through heat absorption.

Advantageously, the system does not require piping and functions under normal ventilation. Additionally, it does not leave a residue, and because of the water droplet size, very little wetting occurs. The amount of water released is as little as 1 gallon/min in total flooding applications, when compared with 8 gallons/min from high pressure water mist systems and over 25 gallons/min for traditional sprinkler systems. However, it is important to note that the effectiveness of the hybrid system is affected by elevation.

Hybrid fire suppression system is a new technology that shows potential and should be evaluated in the years to come.

28.6.9 Hot and Cold Aisle Ventilation

The trend to provide hot and cold aisle containment for energy efficiency can have negative effects on fire suppression and fire detection systems. For sprinkler systems, this primarily includes obstructing the sprinkler pattern at the ceiling. For total flooding systems, this could mean a delay to reach concentration at the source of ignition if the agent is released in a different aisle. For fire detection systems, excess airflow decreases the ability for detectors to observe a fire/smoke situation.

Aisle partitions or curtains are sometimes designed to drop in a fire condition either via thermal or signal response. This can blanket equipment upon release and slow extinguishment if not properly designed. It is imperative that the design team coordinate this prior to installation.

28.6.10 Fire suppression systems comparison

There are a variety of active fire suppression/extinguishing systems that can be used in data centers. Most of these systems rely on water or gaseous agent/blend to extinguish a fire. Table 28.2 highlights the agents discussed in this section, and Figure 28.4 provides a rough comparison of agent quantity.

TABLE 28.2 Comparison of common gaseous fire suppression agents [23]

Agent	Halon 1301	HFC-227ea	HFC-125	IG 541	FK-5-1-12
Trade name	Halon 1301	FM-200	FE 25	Inergen	Novec-1230
Type	Halogenated	Halogenated	Halogenated	Inert	Fluoroketone
Manufacturer	NA	DuPont	DuPont	Ansul	3M
Chemical formula	CF_3Br	C_3HF_7	C_2HF_5	$5N_24ArCO_2$	$C_6F_{12}O$
Molecular weight (g/mol)	149	170	120	34.4	316
Specific volume: m^3/kg (ft^3/lb) at 1 atm, 20°C	0.156 (2.56)	0.137 (2.20)	0.201 (3.21)	0.709 (11.36)	0.0733 (1.17)
Extinguishing concentration[a]	5%	7%	9%	38.5%	4.7%
NOAEL concentration[b]	5%	9.0%	7.5%	52%	10%
LOAEL concentration[c]	7.5%	10.5%	10%	62%	>10%
Vapor weight required kg/100 m^3 (lb/1,000 ft^3)[d]	44 (28)	74 (46)	66 (41)	66 (656)[e] m^3/100 m^3 (ft^3/1,000 ft^3)	91 (57)
Minimum piping design pressure at 20°C, bar (psi)	42 (620)	29 (416)	34 (492)	148 (2,175)	10 (150)
Ozone depletion potential (ODP)	12.0	0	0	0	0
Global warming potential (100 years) relative to CO_2	7,030	3,220	3,500	0	1
Atmospheric lifetime (years)	16	34	32	NA	0.038

Source: NFPA 2001 [15]

[a] Actual extinguishing concentration depends highly on fuel. Values represent common concentrations associated with fuels typical of a data center and are listed for comparison purposes. Engineering evaluation should be performed for each case.

[b] NOAEL: No observable adverse effect level.

[c] LOAEL: Lowest observable adverse effect level.

[d] Design calculations per NFPA 2001, 2018 edition, Chapter 5 at sea level, including a safety factor of 1.35 for Class C fuels; values are for comparison only and not to be used for design applications.

[e] Inert gases use a different formula and are represented as a volume fraction including a 1.35 safety factor.

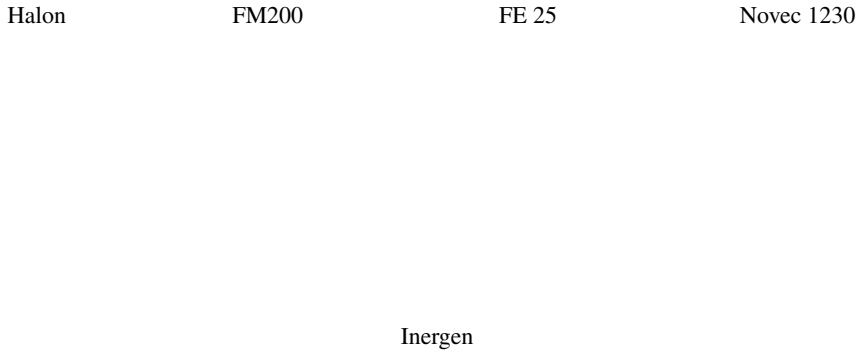


FIGURE 28.4 Comparison of clean agent quantities required for same protected volume.

28.7 DETECTION, ALARM, AND SIGNALING

A traditional fire alarm system provides early warning of a fire event before the fire becomes life threatening. In the case of data centers, the threat to life is relatively low because of sparse occupancy and moderate fuel load; therefore, the building and fire codes don't typically require fire alarm systems unless occupant loads are higher than 500 for single story buildings or 100 for multistory buildings. So, fire alarm and signaling systems have evolved to provide early warning not only for life-threatening fire events but also for property and operational loss events. In fact, this is part of the reason for the name change of NFPA 72 [24], National Fire Alarm Code® to the National Fire Alarm *and Signaling* Code as of the 2010 code cycle.

Documents such as TIA 942 and the Uptime Institute's *Tier Standard: Topology* recommend detection and fire alarm for Tier III and IV centers, as do FM Global data sheets. Therefore, for data centers, fire alarms provide emergency signals to serve the following functions:

1. To detect a fire event, system trouble, or supervisory event at some predetermined risk level. This could include any of the following:
   a. Smoke detection
   b. Heat detection
   c. Sprinkler system activation
2. To alert occupants, owners, and off-site monitoring services to varying degrees of a fire event, system trouble, or system supervisory event via signals such as:
   a. Pre-alarm alert to on-site personnel.
   b. Notification appliances in the building.
   c. Wired and wireless communication to off-site personnel and responders.
3. To initiate various sequences of suppression system activation, as programmed, such as:
   a. Releasing a solenoid valve for a pre-action sprinkler or clean agent system.
   b. Initiating a countdown for a clean agent system.

In practice, the type of system is driven by the perception of risk. Items to consider include risk to life safety, operational continuity and business interruption, and property/equipment loss. A small server room wherein the aggregate risk is low may not need any form of detection. If the owner wants some form of alert prior to sprinkler system activation, a single smoke detector can be installed to sound alarm.

For more complex and critical sites, a dedicated air aspirating smoke detection system will provide very early warning of smoldering conditions before alerting occupants of a pre-alarm condition. If combustion continues, the system will go into alarm and may initiate an optional 30-second delay prior to opening a solenoid valve to a pre-action system. If the sequence is not aborted or the conditions in the room deteriorate, then either clean agent or sprinkler water will discharge.

28.7.1 Heat Detection

Heat detection is used in spaces where smoke detection would not be practical, such as dirty or freezing locations. When a room is sprinkler protected, it is typically not necessary to additionally provide heat detection as the water flow switch from a sprinkler system will initiate alarm, unless earlier thermal warning is desired.

28.7.2 Smoke Detection

The most popular mode of smoke detection within buildings is the spot-type photoelectric or ionization smoke detector. These detectors have a low cost and high level of reliability and can be programmed for different sensitivity levels. When combined with an addressable fire alarm system, specific responses can be programmed including:

- Pre-alarm/alert
- Fire alarm
- Timer initiation
- System discharge

These detectors also have their drawbacks, most notably that the initial sensitivity level is quite high compared with other technologies; by the time a spot-type detector detects smoke, a fire originating from electrical equipment may have been already caused damage to that equipment. Second, dirt and dust accumulate in the sensing element and further affect the sensitivity of the detector over time. For this reason, spot-type detectors must be located such that they are always accessible for servicing, which can have an impact on operational costs. Another issue is spot-type smoke detectors are considered to be "passive" detection whose response is dependent upon detector placement, the smoke reaching the detector, and the airflow velocity present in the room. Lastly, ionization type detectors are sensitive to air velocity and are typically not listed for use in environments exceeding 300 ft/min (1.5 m/s). Listings exist for different ranges of velocities, but designers need to be aware of placing ionization detectors in plenums or under-floor spaces.

An extremely effective smoke detection system for data centers is the air aspirating or air sampling smoke detection system. An early manufacturer of this type of system and one that has become synonymous with the technology is Xtralis (originally IEI later vision systems); they have trade marked the term VESDA (very early smoke detection apparatus), which is now the generic term for all air sampling.

It should be noted there are multiple manufacturers and specialized applications for these systems and this document will only address the basics of how these systems work. Over the years these systems have proven to be most effective for detecting smoke within a data center. The type of fires in data centers typically start as smoldering, low energy fires. The smoke generated isn't significant and does not contain sufficient heat energy to overcome the mechanical energy created by the HVAC systems. At the early stages of combustion, it is common for the smoke-laden air to move along the ceiling and not reach the sensing chamber of the spot-type smoke detectors, due to the high airflows present.

In contrast to spot-type detectors, aspirating systems provide a much earlier warning because of two key factors; the sensitivity of the detector is so much higher and is programmable to the environment being installed, and the systems are "active" detection instead of "passive." A network of sampling pipe is arranged throughout the protected space throughout the space and continuously draws the air through a series of sampling ports. Once the air is drawing into the pipe, it is transported to the detector (this is the active component of the system). The air is analyzed within the sensing chamber. The individual sampling point spacing is determined by the type of detection being designed. The systems are modular in that a single detector can accommodate a certain length of pipe with a certain number of sampling ports, and additional detectors can be added to accommodate larger spaces or separate zones, such as under-floor or interstitial spaces.

These systems use proprietary technology for the detection of smoke particulates as it passes through the sensing chamber to be analyzed; several detection technologies include laser-based, cloud chamber or a dual-source sensor. The detector's sensing chamber measures a percentage of obscuration per lineal foot depending on the manufacturer's algorithm. For example, if a detector is set to initiate an alert condition at an obscuration of 0.25%/ft and that level of obscuration were consistent across the entire room, an occupant 40 ft away would perceive an obscuration of 10%. In reality, an overheating circuit only produces a localized level of obscuration, which is not typically perceptible to occupants, but would be picked up by the closest sampling port. Most manufacturers offer a number of preset obscuration levels that will initiate a certain signal. These preset levels can be manipulated based on the geometry and contents of the protected space and the desired smoke detection sensitivity.

There are three generally accepted methods of detection for air sampling systems: standard fire detection (SFD), early warning fire detection (EWFD), and very early warning fire detection (VEWFD). For data center applications, either EWFD or VEWFD is utilized. For an air sampling system with a total of 20 sampling holes and designed to VEWFD requirements, the sensitivities programed at the air sampling smoke detector unit may look like:

- Alert: 0.01%/ft
- Action: 0.03%/ft
- Fire 1: 0.05%/ft
- Fire 2: 0.125%/ft

In addition to providing a higher level of sensitivity, air sampling systems are not susceptible to high velocity flows the way other type of detectors are and nuisance alarms due to dust or other particulates in the air. Depending on the manufacturer, air sampling systems have differing levels of reliability in terms of segregating dust particles from actual combustion particles. One complaint is that after brief use in a new installation, the user experiences nuisance alarms and requests the sensitivity levels changed or ignores the alert and action warnings intended to protect the equipment. Because of this, it is recommended to thoroughly clean the space prior to operating equipment. A thorough understanding of what these systems are intended for is highly recommended before purchasing. Maintenance includes cleaning sampling points and changing out filters. Frequency ranges in between 6 and 60 months depending on the environment of the space.

28.7.3 Gas Detection

Gas detection is used to detect gases in the environment. For example, some types of batteries produce hydrogen when they charge. Battery rooms are typically ventilated to avoid a buildup of hydrogen; however, the owner may install hydrogen detection as an additional precaution. Gas detection is also used in the semiconductor industry for rooms using or storing hazardous production materials (HPM).

28.7.4 Sequence of Operations

The sequence of operation is the set of logical functions that a fire alarm will follow based on inputs from devices. A portion of a sample fire alarm event matrix for a data center is shown in Table 28.3. This is not to be construed as a complete sequence of operation for a fire alarm system.

28.7.5 HVAC Shutdown

As a basis of design and in accordance with mechanical codes and standards, most HVAC systems with a return air system larger than 944 l/s (2,000 cfm) are required to shut down when smoke is detected via a duct smoke detector. The duct detector is not required when the entire space serve by that unit is protected with smoke detection. It is important to coordinate this code requirement with the operational need to

TABLE 28.3 Sample portion of a fire alarm matrix

		Notification					Suppression control				
	System outputs	Actuate audible/visual alert	Actuate audible/visual alarm	Actuate abort alert signal	Actuate supervisory signal	Actuate trouble signal	Activate 30-second delay	Release clean agent	Release pre-action solenoid	Interrupt countdown cycle	Resume countdown cycle
	System inputs	A	B	C		D	E	F	G	H	I
1	Air sampling alert	●									
2	Air sampling fire		●				●				
3	Termination of delay cycle		●					●			
4	Smoke detector activation		●				●				
5	Manual fire alarm box		●								
6	Manual clean agent switch		●					●			
7	Activation of abort switch			●						●	
8	Release of abort switch										●
9	Waterflow switch		●								
10	Tamper switch				●						
11	Lockout valve				●						
12	Pre-action alarm		●						●		
13	Pre-action supervisory				●						
14	Pre-action trouble					●					

keep the data center equipment ventilated to avoid damage. When an HVAC system is dedicated to a single room, it can be argued that automatic shutdown is not necessarily the best choice because the intent of system shutdown is primarily to keep a fire from spreading to other portions of the building. Shutdown also protects the HVAC equipment from damage, but in some cases, it may make sense to leave the equipment running, at least for a period of time. Examples include:

- An alert-type alarm that does not deteriorate to a fire condition
- While data center or telecommunications equipment is being shut down
- Where the HVAC system is used to assist in mixing a clean agent

28.8 FIRE PROTECTION DESIGN & CONCLUSION

A good fire protection strategy will include a thorough evaluation of anticipated risk and will continue to evaluate that risk through the life of the building. Starting with the user and working through all the stakeholders, determine how critical the site is and what effect a complete loss would have on the organization. The effect may be minimal due to other systems in place. In this case the designer's role is to determine the appropriate authorities having jurisdiction and meet minimum code requirements.

On the other side of the spectrum, a fire event may have an enormous consequence in lost production or service. In this case, the right system may include a combination of fire suppression, detection, and alarm approaches. It will give the user the right information at the right time and focus protection on the most critical components.

Communication and consensus on goals with all vested parties will result in a successful protection strategy. Once the goals are agreed upon, the technical portion of design can start; a summary of which follows. The design team should determine:

- Site requirements (e.g. separation from other buildings on a campus).
- Type of construction.
- Occupancy classification.
- Occupant load.
- Egress requirements.
- Hazard analysis (e.g. batteries, generator fuel, storage).
- Suppression system requirements.
- Detection, fire alarm, and emergency communication systems requirements.
- Interconnection with mechanical and electrical systems.

Each data center or telecommunications room carries a certain level of risk that must be mitigated. As described in this chapter, the level of fire protection in these rooms can often exceed code minimums. Risk is evaluated in terms of life safety, business continuity, and property/equipment value. Then, the tolerance for that risk is addressed. This is where the various standards and previous experience of the design team will drive the design.

Fire protection design should be incorporated early into the design along with other systems. A coordinated team including vendors, manufacturers, and fire protection consultants experienced with LSC and integrated fire protection systems will help the team make the best decision based on established design criteria.

REFERENCES

[1] International Code Council. *2018 International Building Code*. Country Club Hills: International Code Council, Inc.; 2017.

[2] International Code Council. *2018 International Fire Code*. Country Club Hills: International Code Council, Inc.; 2017.

[3] NFPA 75. *Standard for the Protection of Information Technology Equipment*. Quincy: National Fire Protection Association; 2017.

[4] ANSI/TIA 942-B-2017. *Telecommunications Infrastructure Standard for Data Centers*. Annex G, Tables 9-11. Arlington: Standards and Technology Department; 2017.

[5] Uptime Institute Professional Services. *Data Center Site Infrastructure Tier Standard: Topology*. New York: Uptime Institute; 2018.

[6] NFPA 101. *Life Safety Code*. Quincy: National Fire Protection Association; 2017.

[7] ASTM E-119 – 19. *Standard Test Methods for Fire Tests of Building Construction and Materials*. West Conshohocken: ASTM International; 2019.

[8] Gypsum Association. *GA-600-2018 – Fire Resistance and Sound Control Design Manual*. Hyattsville: Gypsum Association; 2018.

[9] Hall JR. *U.S. Experience with Sprinklers*. Quincy: National Fire Protection Association Fire Analysis and Research Division; 2012. p 6.

[10] Kochelek J. *White Paper, Mission Critical Facilities_Is the Use of Galvanized Pipe an Effective Corrosion Control Strategy in Double Interlock Preaction Fire Protection Systems*. St. Louis: Fire Protection Corrosion Management, Inc.; 2009.

[11] Braidech MM, Neale JA, Matson AF, Dufour RE. *The Mechanism of Extinguishment of Fire by Finely Divided Water*. New York: National Board of Fire Underwriters; 1955. p 73.

[12] Rasbash DJ, Rogowski ZW, Stark GWV. Mechanisms of extinction of liquid fuels with water sprays. Combust Flame 1960;4:223–234.

[13] NFPA 750. *Standard on Water Mist Fire Protection Systems*. Quincy: National Fire Protection Association; 2019.

[14] Mawhinney JR. *Fire Protection Handbook*. 20th ed. Quincy: National Fire Protection Association; 2008. p 6–139.

[15] NFPA 2001. *Standard on Clean Agent Fire Extinguishing Systems*. Quincy: National Fire Protection Association; 2018.

[16] NFPA 12A. *Standard on Halon 1301 Fire Extinguishing Systems*. Quincy: National Fire Protection Association; 2018.

[17] Research Basis for Improvement of Human Tolerance to Hypoxic Atmospheres in Fire Prevention and Extinguishment. Philadelphia: Environmental Biomedical Research Data Center, Institute for Environmental Medicine, University of Pennsylvania; 1992. EBRDC Report 10.30.92.

[18] NFPA 2010. *Standard for Fixed Aerosol Fire-Extinguishing Systems*. Quincy: National Fire Protection Association; 2015.

[19] Brooks J. *Aircraft Cargo Fire Suppression Using Low Pressure Dual Fluid Water Mist and Hypoxic Air*. Gaithersburg: National Institute of Standards and Technology. NIST Special Publication 984.2.

[20] Daniault F, Siedler F. International standardization of oxygen reduction systems for fire protection. Presentation at 2017 Suppression, Detection, and Signaling Research and Applications Conference (SUPDET 2017); September 12–14, 2017; College Park, Maryland.

[21] NFPA 10. *Standard for Portable Fire Extinguishers*. Quincy: National Fire Protection Association; 2018.

[22] NFPA 770. *Standard on Hybrid (Water and Inert Gas) Fire Extinguishing Systems*. Quincy: National Fire Protection Association: Proposed Standard.

[23] DiNenno PJ, Forssell EW. Clean agent total flooding extinguishing systems. In: *SFPE Handbook of Fire Protection Engineering*. 5th ed. New York: Springer; 2016. p 1483–1530.

[24] NFPA 72. *National Fire Alarm and Signaling Code*. Quincy: National Fire Protection Association; 2019.

29

RELIABILITY ENGINEERING FOR DATA CENTER INFRASTRUCTURES

Malik Megdiche
Schneider Electric, Eybens, France

29.1 INTRODUCTION

Reliability engineering is defined as the science of failure. The first issues and reliability concepts appeared at the beginning of the twentieth century while engineering systems in railways, power systems, and aircraft applications. The main theories and methods have been developed for critical applications such as military, aerospace, and nuclear power plant. Nowadays, reliability engineering is widely used in many areas.

Reliability engineering uses equipment reliability statistics, probability theories, system functional analysis, and dysfunctional analysis to set requirements, measure or predict reliability, identify system weakness points, and propose improvements of the system.

Various reliability engineering techniques are used in reliability engineering:

- Equipment reliability analysis:
  - Field experience reliability statistics
  - Reliability testing
  - Accelerated life testing
- System reliability and availability analysis:
  - Qualitative analysis:
    - Hazard risk analysis
    - Failure mode and effects analysis (FMEA)
  - Reliability prediction:
    - Electronics FMEA
    - Fault tree analysis
  - Statistical simulations:
    - Maintainability analysis
    - Integrated logic support

Reliability engineering techniques can be used for reliability, availability, maintainability, and safety purposes. As this chapter is dedicated to reliability and availability engineering of data center infrastructure, safety and security are not considered.

The dependability of a system deals with the following attributes:

- Availability: Readiness for correct service
- Reliability: Continuity of correct service
- Maintainability: To undergo modifications and repairs

Dependability will be used to name the reliability and availability performance of a data center infrastructure.

The next parts will provide:

- The understanding of basic concepts of system reliability and availability analysis including equipment dependability data and dependability methods

Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center, Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

- The main features to be assessed to perform a relevant analysis
- An application guide to use to a dependability analysis for data infrastructure at design phase and operational phase

29.2 DEPENDABILITY THEORY

29.2.1 System Dependability Analysis Definition

System dependability analysis basics are to study the consequences of component failures on the system based on:

- Equipment failures and maintenance data
- System behaviors in case of failures

29.2.2 System Dependability Indexes

To analyze and estimate the dependability of system, the main system dependability indexes are the following.

29.2.2.1 Reliability

This is the ability of a product to perform a given function, in specified conditions, for a given period of time.

Mathematical indexes to be used are as follows:

The reliability $R(t)$ = "Probability to experience no failure on [0; t]"

The unreliability $\bar{R}(t)$ = "Probability to experience at least one failure on [0; t]"

$$\bar{R}(t) = 1 - R(t)$$

The mean failure frequency F = "Estimated number of failures per year (or hour)"

29.2.2.2 Availability

It is the ability of an item to be in a state to perform a required function.

Mathematical indexes to be used are as follows:

The availability $A(t)$ = "Probability to be capable of performing a required function at t"

The unavailability $\bar{A}(t)$ = "Probability not to be capable of performing a required function at t"

Asymptotic availability $A = \lim_{t=\infty} A(t)$

Asymptotic unavailability $\bar{A} = \lim_{t=\infty} \bar{A}(t)$

$$\bar{A}(t) = 1 - A(t) \text{ and } \bar{A} = 1 - A$$

29.2.2.3 Maintainability

It is the ability of an item to be repaired.

The maintainability $M(t)$ = "Probability to be repaired at t"

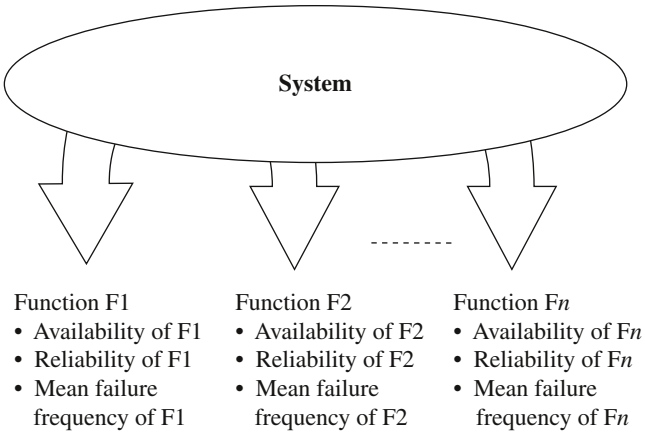


FIGURE 29.1 System external functions. Source: © Schneider Electric.

29.2.2.4 System Functions

It is important to note that system dependability indexes are associated to a function or a group of functions of the system as mentioned in Figure 29.1.

29.2.3 Equipment Dependability Data

29.2.3.1 MTTF, MTBF, MDT, MTTR, MUT

Figure 29.2 shows the equipment state as a function of time.

Time to failure and downtime are random variable. The mean values are defined as follows:

MTTF = Mean time to (first) failure
MTBF = Mean time between failure
MUT = Mean up time
MTTR = Mean time to repair
MDT = Mean down time (including fault detection time + spare delivery time + MTTR)It is important to note that:

- MTTF and MTBF are statistical values and are meant to be the mean over a long period of time and a large number of units.
- MTBF = MTTF + MDT so technically MTBF depends on the time to repair so dependent on-site service maintenance, whereas MTTF is an equipment data that is not dependent on the time to repair.
- Technically, MTBF should be used only in reference to a repairable item, while MTTF should be used for non-repairable items. However, MTBF and MTTF are commonly used for both repairable and non-repairable items.

29.2.3.2 Failure Rate λ

The failure rate $\Lambda(t)$ is defined as the probability to fail between times [t; $t + \Delta t$]:

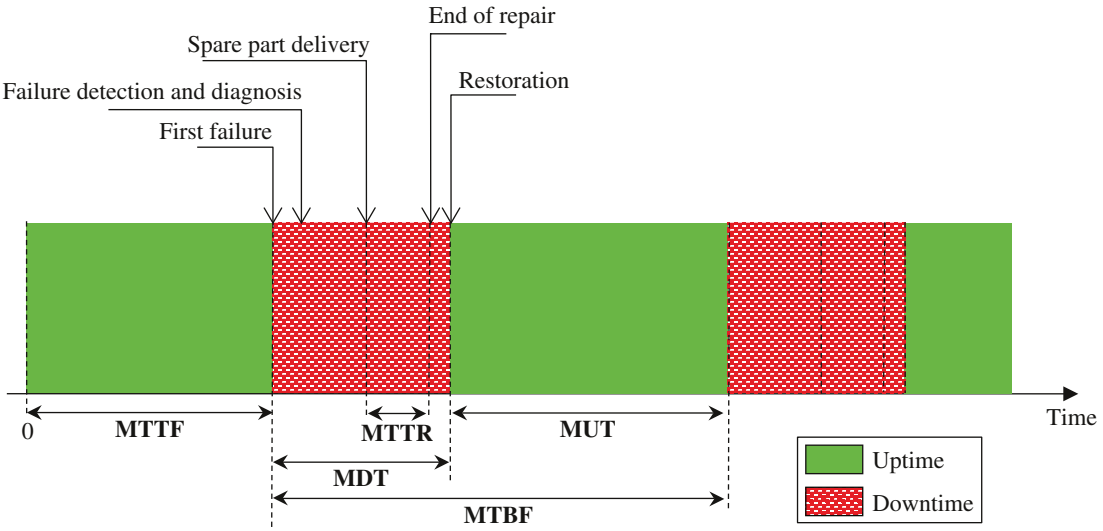


FIGURE 29.2 Equipment state during its service period. Source: © Schneider Electric.

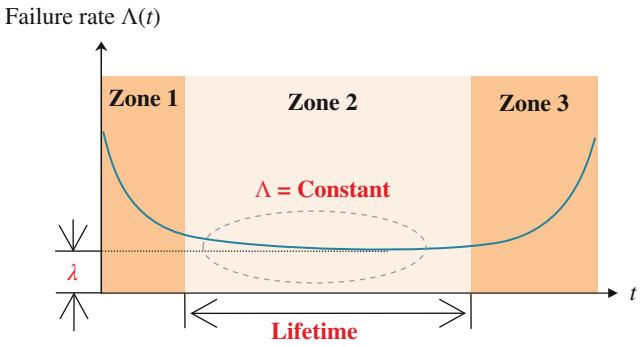


FIGURE 29.3 Failure rate (λ) curve. Source: © Schneider Electric.

$$\Lambda(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \cdot P\left[\text{failure during } (t; t+\Delta t) \text{ knowing hat no failure happened before}\right]$$

$$\Lambda(t) = -\frac{1}{R(t)} \cdot \frac{dR(t)}{dt}$$

The failure rate is not always constant, and its evolution over time can be described by the well-known "bathtub" curve as represented in Figure 29.3.

There are three distinct zones on this curve (Fig. 29.3):

Zone 1: Early-life or infant mortality phase

During this period, the lambda value decreases. Failures are due to latent faults. Burn-in is intended to eliminate latent failures in the factory and prevent defective parts from being shipped to customers. Commissioning tests are also intended to detect equipment failure before its operational phase. It should minimize the size of this zone.

Zone 2: Useful life period

Failures occur randomly, and the failure rate is constant. In every dependability studies, failure rate values used correspond to this phase and therefore are constant.

Zone 3: Wear-out phase

At the end of the useful life period, aging affects components and generate drastic and uncontrolled increase of their failure rate. It is the beginning of the wear-out phase. Commonly for electrical and electronic equipment, the failure rate (λ) is assumed to be constant during the equipment lifetime, and the random variable "time to failure" follows an exponential distribution law $e^{-\lambda t}$ (Table 29.1):

$$\Lambda(t) = \lambda$$

$$R(t) = 1 - e^{-\lambda t}$$

where e is the base of the natural logarithm—a mathematical constant approximately equal to 2.71828.

The mean of the random variable "time to failure" is MTTF = $(1/\lambda)$.

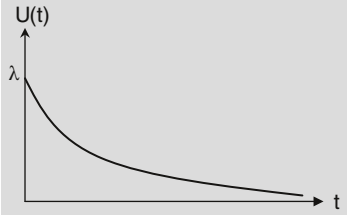
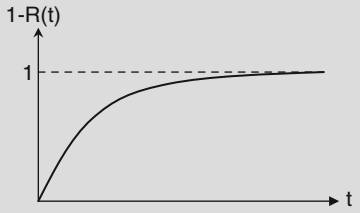
Assuming a constant failure rate, it allows:

- Simple field experience failure estimations

  An estimation of the failure rate is defined as the ratio between the number of observed failures and the cumulative operating time of the products. Therefore, its unit is consistent with the inverse of a time.
- Simple system predictive reliability and availability calculations

A common misunderstanding is to mix up the lifetime and the MTTF (or MTBF) of a component. These two parameters are different as shown in Figure 29.3. The lifetime is set

TABLE 29.1 Time to failure follows exponential distribution law

Probability density function (failure density)	Cumulative distribution function (unreliability)	Mean value (MTTF)
$U(t) = \lambda \cdot e^{-\lambda \cdot t}$	$1 - R(t) = 1 - e^{-\lambda \cdot t}$	$\text{MTTF} = \frac{1}{\lambda}$
		

Source: © Schneider Electric.

TABLE 29.2 Definition of failure mode of a circuit breaker

Failure modes of a circuit breaker	Contributions (% of total failures)	Failure rate
Fail to trip on fault	xx%	$\lambda_1 = \text{xx\%} \cdot \lambda$
Spurious opening	xx%	$\lambda_2 = \text{xx\%} \cdot \lambda$
Unintended closing	xx%	$\lambda_3 = \text{xx\%} \cdot \lambda$
Fail to open on demand	xx%	$\lambda_4 = \text{xx\%} \cdot \lambda$
Fail to close on demand	xx%	$\lambda_5 = \text{xx\%} \cdot \lambda$
Insulation breakdown	xx%	$\lambda_6 = \text{xx\%} \cdot \lambda$

Source: © Schneider Electric.

according the wear-out phase to replace the component before its failure rate increase due to aging effects, whereas the MTTF is related to the random failure frequency during the lifetime.

For some component, the failure rate γ can be calculated per number of operations (opening/closing cycle or start sequence). In this case, the following equation is used:

$$\lambda = \gamma \cdot N_{\text{operations}}$$

where

λ = failure per hour
Γ = failure rate
N = Operations/hour

29.2.3.3 Failure Modes

To perform the analysis of the consequences of each possible failure, it is important to define clearly the failures for each component of the system. As a component can fail in multiple ways and due to multiple causes, defining all the possible failures will be too long and will lead a lot of complexity when studying hundreds of equipment.

To simplify into dependability data of component, an interesting way is to group the failures leading to the same consequences: the degradation of a function. This group of failures is named "failure mode."

The failure mode is defined as the degradation of a function of the component.

Table 29.2 shows an example of the failure modes of a circuit breaker.

For each failure mode, the associated failure rate has to be determined.

29.2.3.4 Curative Maintenance Data

As mentioned previously, the global downtime after a failure can be detailed in a series of events that are represented in Figure 29.4.

The global downtime of a component is the sum of:

- The time to detect the failure

  A failure can be detected by:
  - Its direct consequence on the system functions (for example, a short circuit will trip the upstream circuit breaker, and the SCADA or the users will detect immediately the failure)
  - A protection system
  - A watchdog trip

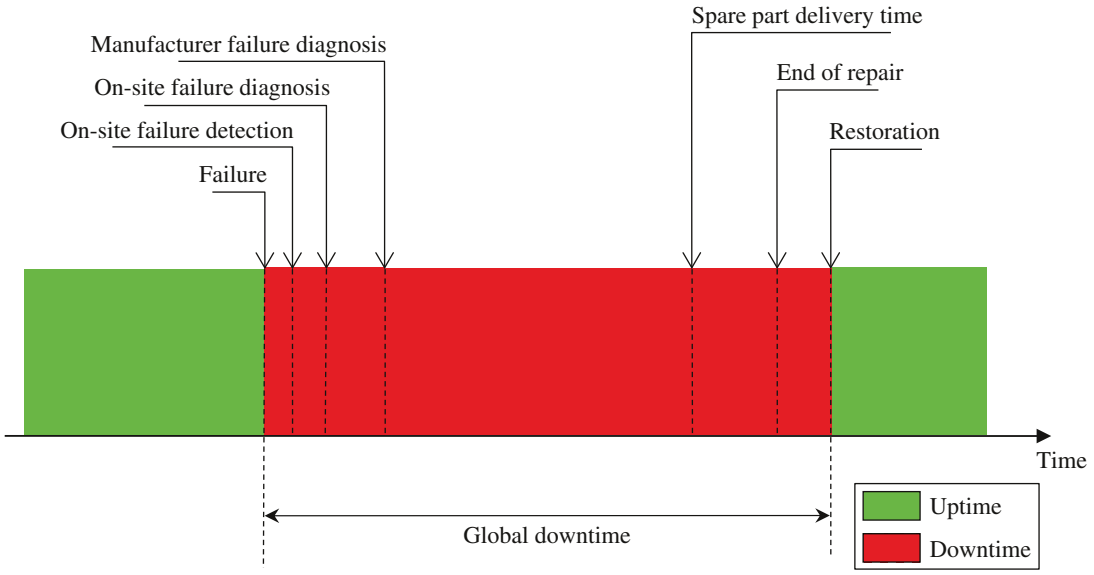


FIGURE 29.4 Curative maintenance times. Source: © Schneider Electric.

- A periodical test
- A preventive maintenance operation

Note that the monitoring system can minimize the failure detection.

- The time to diagnostic the failure

  The time to diagnostic includes:
  - The customer monitoring system functions
  - The customer on-site maintenance shift
  - The customer on-site maintenance competencies
  - The manufacturer service maintenance contract

  Note that the monitoring system and disturbance analysis functions can minimize the failure diagnostic.
- The spare part delivery time
- The time to lock out the equipment
- The mean time to repair or to replace the equipment
- The time to unlock and restore the equipment

All the maintenance times are also random variables. It is common to approximate these random times by assuming that the maintenance times are constant as shown in Figure 29.5.

However, this approximation has to be carefully assumed according to:

- The relation between some equipment MDT and time-limited redundancies
- The relation between some equipment MDT and time criticality of process interruptions

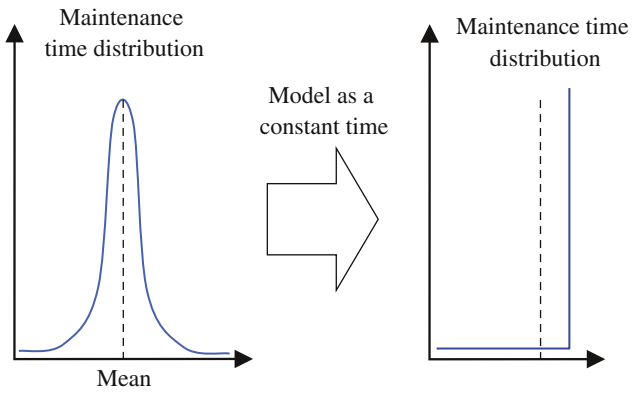


FIGURE 29.5 Maintenance time distribution model. Source: © Schneider Electric.

29.2.3.5 Preventive Maintenance Data

Preventive maintenance is planned maintenance that is designed to:

- Improve equipment life
- Guarantee no aging phenomena
- Detect latent failures

Preventive maintenance ensures constant failure rates as shown in Figure 29.6.

Preventive maintenance operation can include:

- Inspections
- Cleaning

- Tests and measurements
- Adjustments
- Part replacements

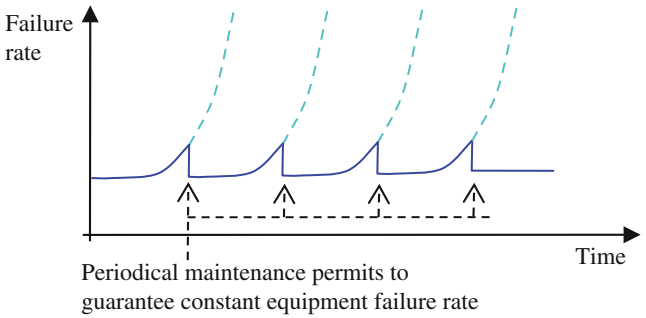


FIGURE 29.6 Failure rate evolution according to periodical maintenance. Source: © Schneider Electric

However, maintenance operation may require to lock out the equipment or to lead to some equipment function unavailability during the operation.

For each equipment preventive maintenance operation, the required data to perform a dependability analysis can be summed up as follows:

- Description of the maintenance operation including which devices need to be shut down and locked out
- Frequency and duration of the operation

29.2.3.6 Particular Case of the Utility Dependability Data

IEEE 1366 standard defines indexes to measure electrical utility distribution system dependability. Here are the definitions of these indexes:

System Average Interruption Frequency Index

$$\text{SAIFI} = \frac{\text{Number of customer interruptions}}{\text{Number of customers served}}$$

System Average Interruption Duration Index

$$\text{SAIDI} = \frac{\sum \text{Customer minutes interrupted}}{\text{Number of customers served}}$$

Average Service Availability Index

$$\text{ASAI} = \left(1 - \frac{\text{SAIDI}}{\text{Minutes per year}}\right) \times 100$$

Momentary Average Interruption Frequency Index

$$\text{MAIFI} = \frac{\text{Number customer momentary interruptions}}{\text{Number of customers served}}$$

Similar dependability data can be calculated for other utilities like water or gas utilities.

29.2.4 Basic System Dependability Modeling

29.2.4.1 Single Component

The availability and reliability of a single component is defined by the following formula (Fig. 29.7):

$$\lambda_{\text{equivalent}} = \lambda_1$$

$$\text{Unavailability} = \frac{\lambda_1}{\lambda_1 + \dfrac{1}{\text{MDT}_1}} \approx \lambda_1 \cdot \text{MDT}_1$$

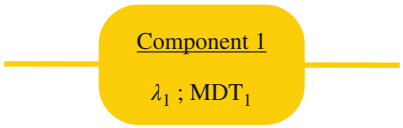


FIGURE 29.7 Single component. Source: © Schneider Electric.

29.2.4.2 Nonredundant Components

The availability and reliability of two nonredundant components (Fig. 29.8) are defined by the following formula:

$$\lambda_{\text{equivalent}} = \lambda_1 + \lambda_2$$

$$\text{Unavailability} = \frac{\lambda_1}{\lambda_1 + \dfrac{1}{\text{MDT}_1}} + \frac{\lambda_2}{\lambda_2 + \dfrac{1}{\text{MDT}_2}} \approx \lambda_1 \cdot \text{MDT}_1 + \lambda_2 \cdot \text{MDT}_2$$

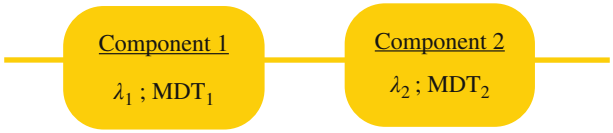


FIGURE 29.8 Two non-redundant components. Source: © Schneider Electric.

29.2.4.3 Two Components with Active Redundancy

"2 components with active redundancy" means that the two components are running together (Fig. 29.9). If one fails, the other one is able to keep the system running.

The availability of two redundant components is defined by the following formula:

$$\lambda_{\text{equivalent}} \approx \lambda_1 \cdot \lambda_2 \cdot \left(\text{MDT}_1 + \text{MDT}_2\right)$$

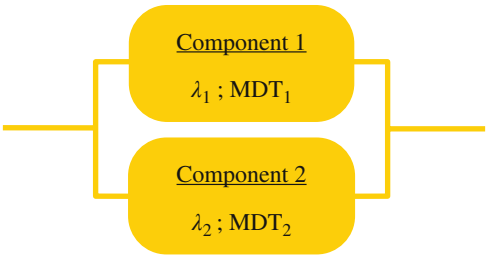


FIGURE 29.9 Two redundant components. Source: © Schneider Electric.

$$\text{Unavailability} = \frac{\lambda_1}{\lambda_1 + \dfrac{1}{MDT_1}} \cdot \frac{\lambda_2}{\lambda_2 + \dfrac{1}{MDT_2}}$$
$$\approx \lambda_1 \cdot MDT_1 \cdot \lambda_2 \cdot MDT_2$$

29.2.4.4 Two Components with Passive Redundancy

Two components with passive redundancy means that one component is in standby mode (partially activated or switched off). If the first component fails, the second one is activated to keep the system running (Fig. 29.10).

The availability of two redundant components is defined by the following formula:

$$\lambda_{\text{equivalent}} = \lambda_1 \cdot \lambda_2 \cdot MDT_2$$

$$\text{Unavailability} \approx \lambda_1 \cdot \lambda_2 \cdot MDT_2 \cdot \min\left(MDT_1 ; MDT_2\right)$$

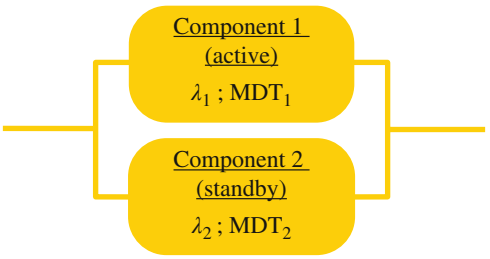


FIGURE 29.10 Two components with passive redundancy. Source: © Schneider Electric.

29.2.4.5 Partial Redundancy

Redundancy can be achieved on two components only partially due to capacity limitation (Fig. 29.11).

This capacity limitation can be:

- Time limited like components using batteries, fuel storage, or water storage

  In this case, the component is redundant for any failure that does not exceed the time-limited capacity of the component.

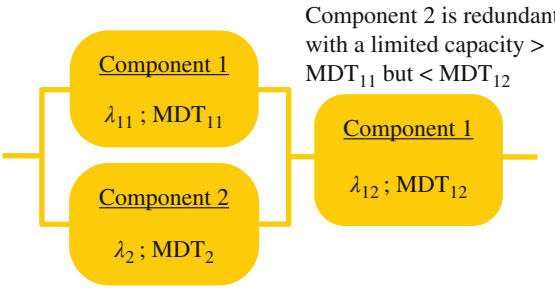


FIGURE 29.11 Multi-components with partial redundancy. Source: © Schneider Electric.

- Dependent on a variable condition
  - Environment condition (temperature, humidity, etc.)
  - Load probability that can exceed the component capacity

  To model that type of redundancy, the ratio PR defined as "the percentage of time for which the redundancy is achieved" has to be determined. Then the dependability diagram canted as shown in Figure 29.12:

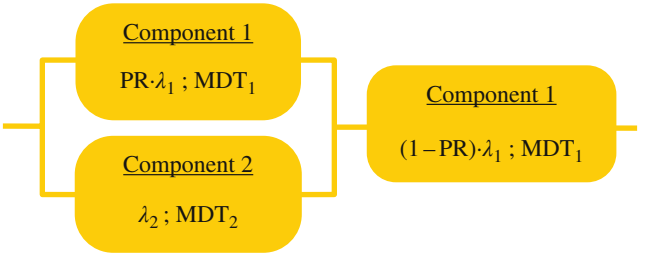


FIGURE 29.12 Component 2 is redundant partially with a percentage PR. Source: © Schneider Electric.

29.2.4.6 Common Cause Failures

A common cause failure is one in which a single failure or condition affects the operation of multiple devices that would otherwise be considered independent. Common cause failures are one of the most important points of a reliability study because a common cause failure can lead to the failure of two redundant devices.

Common cause failures can be classified as follows:

- Human error:
  - Error during design, manufacturing, and installation phases
  - Unintended action
  - Inadequate or incorrect procedure
  - Inadequate training
  - Inadequate maintenance

- Environment:
  - Fire, smoke
  - Temperature, humidity, moisture
  - Electromagnetic field
  - Animals, bio-organisms
  - Contamination, dust, dirt
  - Wind, flood, lightning, snow, ice, earthquake

The common cause failure can be modeled as shown in Figure 29.13.

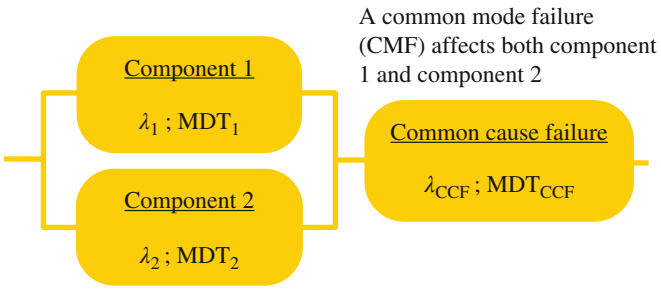


FIGURE 29.13 Common cause failure. Source: © Schneider Electric.

29.2.4.7 Hidden Failure

A failure of a function that normally requires to keep the system running will have no direct consequence on the system. This type of failure is named hidden failure or latent failure.

A hidden failure is detected:

- When the failed function is required
- During periodical test of the function

A simple way to represent a hidden failure is shown in Figure 29.14.

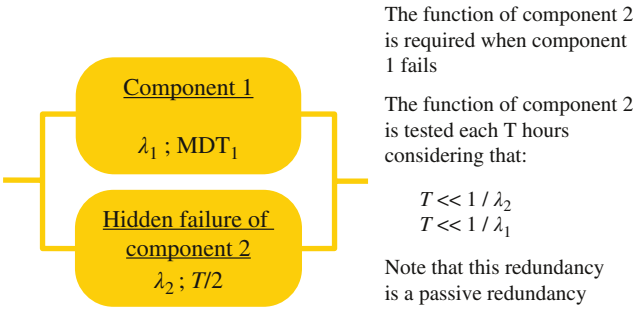


FIGURE 29.14 One component with hidden failure. Source: © Schneider Electric.

29.2.4.8 Preventive Maintenance

If the process needs to run during a preventive maintenance operation that requires to lock out some components, the maintenance operation needs to be taken into account for the dependability analysis as shown in Figure 29.15.

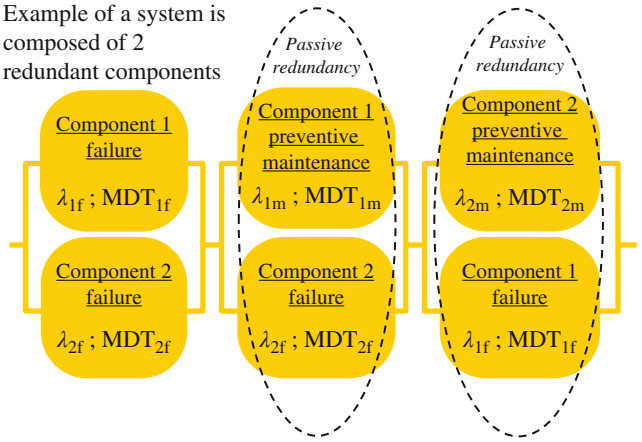


FIGURE 29.15 A system composed of two redundant components. Source: © Schneider Electric.

29.2.4.9 Lifetime and Preventive Replacement

As mentioned previously, the preventive replacement of aging part of a component permits to guarantee constant failure rate.

In some case, the customer prefers to replace the aging parts after a failure; a simple and pessimistic way to model the increased failure rate is shown in Figure 29.16.

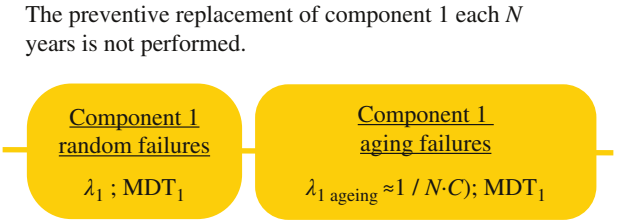


FIGURE 29.16 Simple and pessimistic approach to model the increase failure rate. Source: © Schneider Electric.

29.3 SYSTEM DYSFUNCTIONAL ANALYSIS

29.3.1 Dependability Analysis Methodology

The general dependability methodology as used in dependability engineering field is simply represented in Figure 29.17.

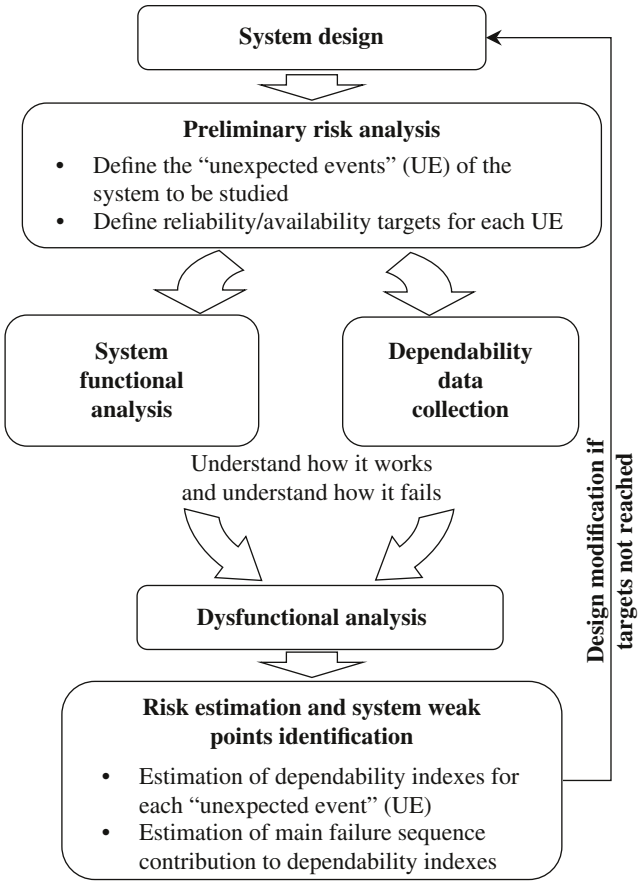


FIGURE 29.17 Dependability methodology. Source: © Schneider Electric.

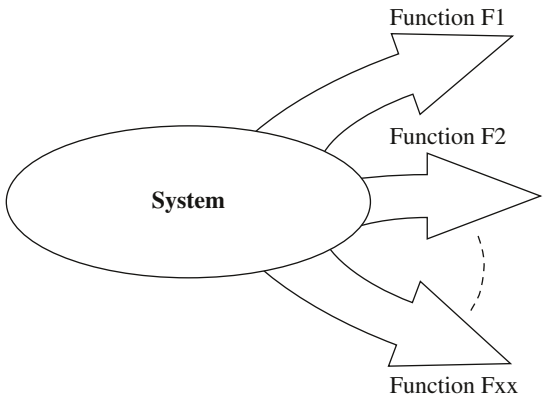


FIGURE 29.18 System external functional analysis. Source: © Schneider Electric.

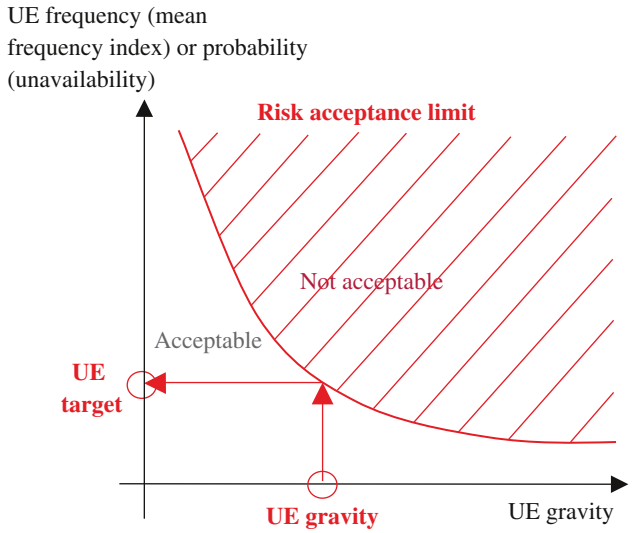


FIGURE 29.19 Risk acceptance graph. Source: © Schneider Electric.

29.3.1.1 Preliminary Risk Analysis

This preliminary step is essential to define precisely what are the critical events to be studied and to perform the right study. The first step is to identify the external functions of the system and define each undesired event (UE) of the system as the degradation or the unavailability of one or several functions of the system (Fig. 29.18).

For example, UE definitions can be:

- UE1: "Loss of function F1 during more than …"
- UE2: "Loss of function F1 during more than …"
- UE3: "Loss of functions F3 and F4 during more than …"

Note that sometimes the consequences of an expected event can be very different depending on the duration of the UE. In this case, it is important to define different UE with different durations. The second step is to define reliability and/or availability targets for each UE as a function of risk acceptance limit and the UE gravity as shown in Figure 29.19.

29.3.1.2 System Architecture Description

The technical data on the system architecture to be collected are the following:

- The topology of the main systems and the automation and protection systems
- The topology of the critical auxiliaries (such as the control power for switchgears, the fuel supply system of the generator power plant, the cooling system of the technical rooms)
- The layout for all the equipment

29.3.1.3 System Functional Analysis

The aim of this task is to understand how the system works (Fig. 29.20). It has to characterize:

- The operational modes of the system including eventual upgrade and evolution phases

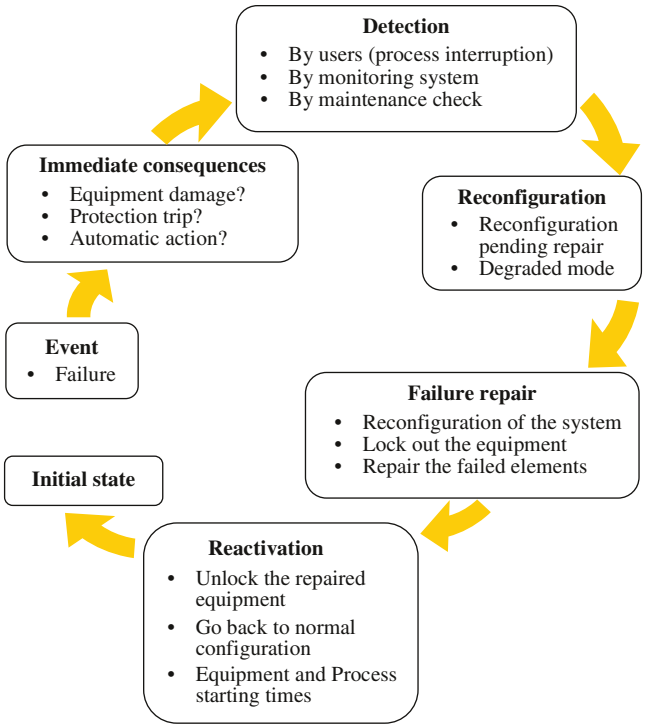


FIGURE 29.20 System behavior after a failure. Source: © Schneider Electric.

- The process automation system
- The protection and automation systems
- The monitoring system
- The emergency maintenance actions to reconfigure the system

Note that some assumptions are needed to determine the consequences of failure sequence:

- Equipment tolerance to supply interruptions
- Protection and automation system behaviors
- Emergency maintenance behaviors
- Starting time of equipment or functions after a blackout

These assumptions have to be done according to equipment data sheet, design studies, and on-site maintenance.

29.3.1.4 Dependability Data Collection

As mentioned previously, the dependability data collection needs to include both components failure data and maintenance data as shown in Table 29.3.

Some specific additional data can be added:

- The time for perform manual operations like switching operation to change the system configuration or a manual restart
- The level of on-site maintenance team

29.3.1.5 Dysfunctional Analysis and System Weak Point Identification

The dysfunctional analysis is the study of the consequences of each possible failure on the system.

As shown in Figure 29.21, based on the possible failures of each component and the behavior of the system, the dysfunctional analysis:

- Generate the failure sequence
- Determine all the actions of system until equipment restoration
- Determine if the failure sequence affect an UE
- Compute dependability indexes for each UE
- Compute failure sequence contribution to each UE

Note that a dependability assessment has to make the assessment for single failure sequences and also multiple failure sequences. However some assumptions can be done to simplify the analysis. This will be detailed in the further chapter.

The dysfunctional analysis results can be resumed for each UE as represented in Table 29.4. If the target is not reached, the system can be improved by identifying the main failure sequence contributions to UE and propose improvements that clear or minimize the critical failure consequences. This can be done into several ways by:

- Designing adequate redundancy
- Setting adequate maintenance

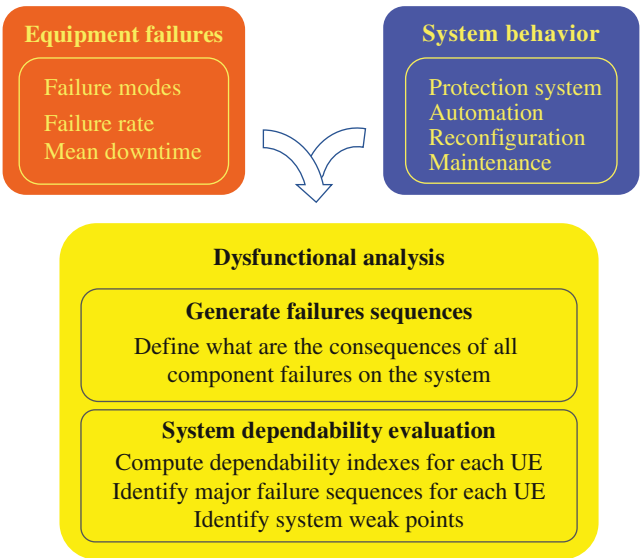


FIGURE 29.21 Dysfunctional analysis principles. Source: © Schneider Electric.

TABLE 29.3 Dependability data table

	Reliability data				Curative maintenance data				Preventive maintenance data			
Equipment type	Reliability source	Failure rate (1/h)	Failure modes	Contributions to failure rate (%)	Detection time (h)	Diagnostic time (h)	Spare part delivery time (h)	Time to repair and restore (h)	Maintenance operation features	What are the equipment or functions unavailable during the operation	Frequency (/year)	Duration (h)
Component 1												
Component 2												
…												

Source: © Schneider Electric.

TABLE 29.4 Dependability assessment results

UE *n*°xx	
Mean frequency index (xxx/year)	
Main failure sequences	Contributions to mean frequency index (%)
	xx
	xx
	xx
Mean unavailability index (xxx h/year)	
Main failure sequences	Contributions to mean unavailability index (%)
	xx
	xx
	xx

Source: © Schneider Electric.

However, the proposed solutions have to take into consideration the following points:

- Be sure that the proposed solution is technically possible and cost effective.
- Be sure that the proposed solution is more dependable than before.
- Keep as possible the system simple to operate.
- Respect as possible the customer habits in terms of system architecture design and operation.

29.3.1.6 Example on a Simple System

The system consists of MV/LV power system supplying LV critical loads as represented in Figure 29.22.

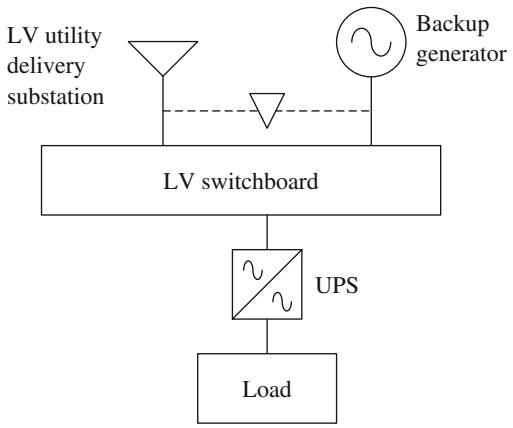


FIGURE 29.22 Single-line diagram of the power system. Source: © Schneider Electric.

Primary Risk Analysis

To determine the main functions, an external functional analysis is performed in Figure 29.23:

Function 1: Supply critical loads

Function 2: Permit maintenance operation, performs installation monitoring

Function 3: Ensure safety

Function 4: Prevent environment pollution (EMC, chemical, etc.)

The goal of the study is to improve the system architecture to optimize the critical load power supply reliability. Therefore, the UE to be studied is defined as UE1 "Loss of critical load power supply." As an interruption of the process (UE1) is critical whenever the UE duration, the target is to minimize UE1 frequency.

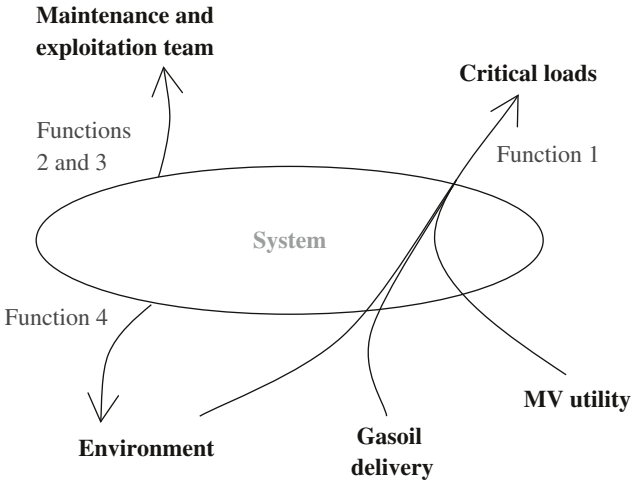


FIGURE 29.23 External functional analysis. Source: © Schneider Electric.

Functional Analysis

Dependability data collection (Table 29.5) considers the following:

- The MV utility delivery is backed up by an LV genset.
- The LV genset starts when the switchboard is de-energized and is designed for continuous operation.
- The LV genset fuel storage permits a 72-hour autonomy at full load. A specific emergency fuel delivery service permits to refill the fuel storage during genset long-period operation.
- The LV genset is tested each month.
- An automatic transfer switch (ATS) permits automatic changeover from one source to another.
- A UPS system with 5-min battery permits to supply the loads during the genset starting sequence.

TABLE 29.5 Dependability data collection

Equipment—failure mode	Failure rate (1/hour)	Mean down time (hours)
Major blackout on HV grid	1.00E-06	4
MV electrical utility failure	1.00E-04	1
MV electrical utility short interruption (<3 min)	1.00E-03	0.033
Genset—fail while in standby mode	1.00E-04	365
Changeover—fail to switch	1.00E-06	365
Changeover—unexpected opening of both switches	1.00E-08	2
LV switchboard failure	1.00E-07	2,190
UPS—short circuit on output	3.00E-07	168
UPS—loss of UPS path (switch on static bypass)	1.00E-05	50

Source: © Schneider Electric.

- Critical loads are tolerant to power supply interruptions with voltage drop of more than 40% of rated voltage during more than 100 ms.
- Maintenance team 24 hour/24 hour permits a mean time to intervention of 15 min and is able to perform manual reconfiguration in 2 hours.

TABLE 29.6 Dysfunctional analysis results of the example

UE1 "Loss of critical loads"—estimated mean frequency: 0.041/year	
Main failure sequences	Contribution to UE1 frequency (%)
"Genset failure during standby" and "utility failure"	80
"Changeover—Fail to switch" and "utility failure"	1
"Changeover—Unexpected opening of both switches"	0
"LV switchboard failure"	2
"UPS—Short circuit on output"	6
"UPS—Loss of UPS path" and "utility short interruption"	11

Source: © Schneider Electric.

Dysfunctional Analysis Results
A fault tree is performed to estimate UE1 frequency and identify main failure sequences. The results are mentioned in Table 29.6.

Weak Point Identifications and Improvement Proposal
The dysfunctional assessment shows that the main contributions to critical load interruptions are the failure sequence "genset failure during standby" and "utility failure." This means that the redundancy of MV utility with a single genset is not sufficient. Several ways can be investigated:

- Improve genset availability by more frequent tests and periodical maintenance and also reduce the genset MTTR.
- Improve genset availability by more redundant "starting system."
- Provide "$N+1$" redundant genset power plant.

29.3.2 Main System Dysfunctional Analysis Methods

29.3.2.1 Failure Mode Effects and Criticality Analysis (FMECA)

Definition According to IEC 60812
Failure mode effects and criticality analysis (FMECA) is method defined by the IEC 60812 standard:

> Failure Modes and Effect Analysis (FMEA) is a systematic procedure for the analysis of a system to identify the potential failure modes, their causes and effects on system performance (performance of the immediate assembly and the entire system or a process).
>
> FMECA (Failure Modes, Effects and Criticality Analysis) is an extension to the FMEA to include a means of ranking the severity of the failure modes to allow prioritization of countermeasures. This is done by combining the severity measure and frequency of occurrence to produce a metric called criticality.

It is important to note that the FMECA makes only the assessment of single failures and does not consider multiple failures.

The classical system FMECA features are the following:

- The system is divided into components (choice of detail level).
- The FMECA is presented in Table 29.7.
- Local and final effects are identified for each component failure mode.
- Indicators frequency (F), detection (D), and gravity (G) levels are determined using reference tables. Reference tables are defined according to the system. Examples of reference tables are given in Tables 29.8–29.10.

Note that the "failure detection" indicator means that the failure is detected and hazard risk is avoided.

TABLE 29.7 Example of FMECA table according to IEC60812

Function	Failure mode	Local effects	Final effects	*F*	*D*	*G*	Criticality ($C = F \times D \times G$)	Action

Source: © Schneider Electric.

TABLE 29.8 Frequency index table

Ranking	Frequency	Criteria failure mode occurrence (failure rate)
1	Improbable	<1e-9/h
2	Remote	<1e-8/h
3	Occasional	<1e-7/h
4	Probable	<1e-6/h
5	Frequent	<1e-5/h

Source: © Schneider Electric.

TABLE 29.9 Detection index table

Ranking	Detection	Criteria: Likelihood of detection by design control ranking
1	Almost certain	Design control will almost certainly detect a potential cause/mechanism and subsequent failure mode
2	Moderately high	Moderately high chance the design control will detect a potential cause/mechanism and subsequent failure mode
3	Very low	Very low chance the design control will detect a potential cause/mechanism and subsequent failure mode
4	Absolutely uncertain	Design control will not and/or cannot detect a potential cause/mechanism and subsequent failure mode, or there is no design control
5	Detection	Criteria: Likelihood of detection by design control ranking

Source: © Schneider Electric.

Various reference tables can be defined using several ranks, using failure rate instead of frequency levels, or using or not using detection indicator:

- Criticality level (frequency $F \times$ detection $D \times$ gravity G) is estimated for each failure mode. Criticality level is used as global risk assessment indicator. The risk acceptability is defined with the customer and can be presented in Table 29.11.

TABLE 29.10 Gravity index table

Ranking	Gravity	Criteria
1	None	No discernible effect
4	Very minor	Fit and finish/squeak, and rattle item does not conform. Defect noticed by most customers (greater than 75%)
7	Very low	Vehicle/item operable but at a reduced level of performance. Customer very dissatisfied
8	Very high	Vehicle/item inoperable (loss of primary function)
9	Hazardous with warning	Very high severity ranking when a potential failure mode affects safe vehicle operation and/or involves noncompliance with government regulation with warning

Source: © Schneider Electric.

- The criticality level of each failure mode gives the identification of main system weak points and the main critical failures for which an action has to be taken to decrease the risk.

The FMECA is widely for risk analysis of system for design or operational purpose. Its main advantage is its simplicity that:

- Makes the FMECA affordable for a lot of people to understand the method and its results
- Requires no specific tool
- Makes easy to perform an exhaustive analysis
- Brings a synthesis of all risks at the same time

However, its main drawback is the fact that the assessment depends highly on:

- The reference table definition
- The people that make the assessment of indicators

Failure Mode Effects and Criticality Analysis Customized for System Dependability Analysis

FMECA can be customized to avoid the use of qualitative indicators. The principle is to make the assessment of UEs of the system.

Basically, the FMECA consists in the study of each single failure consequence on the system as described in Figure 29.24.

TABLE 29.11 Risk acceptance matrix

Risk acceptability matrix				
	Severity = detection × gravity			
Frequency of occurrence of failure effect	Insignificant	Marginal	Critical	Catastrophic
Frequent	Undesirable	Intolerable	Intolerable	Intolerable
Probable	Tolerable	Undesirable	Intolerable	Intolerable
Occasional	Tolerable	Undesirable	Undesirable	Intolerable
Remote	Negligible	Tolerable	Undesirable	Undesirable
Improbable	Negligible	Negligible	Tolerable	Tolerable

Source: © Schneider Electric.

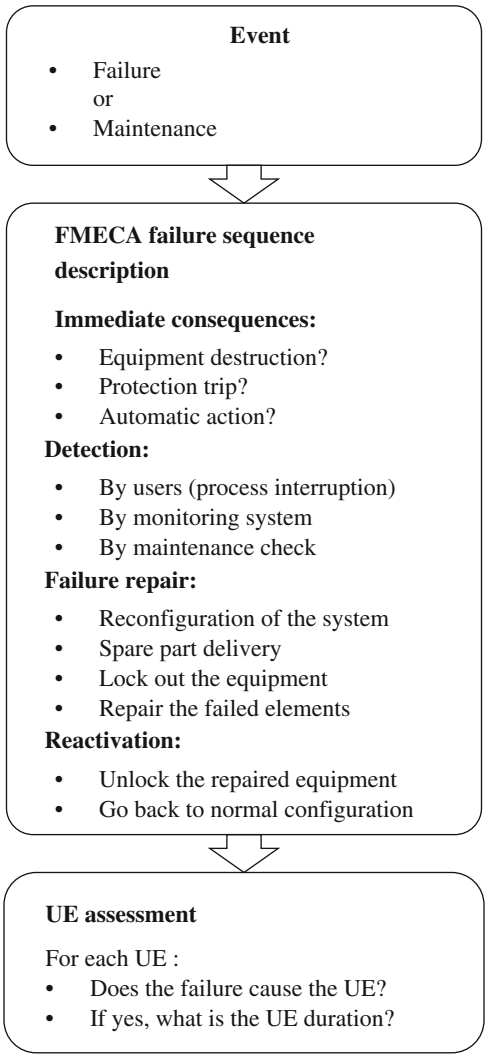


FIGURE 29.24 Assessment principle of an equipment failure mode during a FMECA approach. Source: © Schneider Electric.

The customized FMECA table is presented in Table 29.12. In case of an undetected failure (hidden failure), a pessimistic way is to consider that this failure is detected on a contingency.

In case of an undetected failure (hidden failure), a pessimistic way is to consider that this failure is detected on a contingency.

The UE estimation is done as follows:

$$\text{Mean frequency of UE} = \sum \text{Failure rate} \times \text{UE happens}$$

$$\text{Mean unavailability of UE} = \sum \text{Unavailability} \times \text{UE happens}$$

The contribution of main failures to each index is obtained using a sorting of FMECA lines with the column "Failure rate × UE happens?":

FMECA does not make the assessment of multiple failures; it only makes the assessment of single failures for each UE.

The statistical estimations of UE indexes are assuming that failure combinations are negligible compared with single failures. This assumption has to be demonstrated by verifying that the main failure combinations of the considered UE are negligible.

The FMECA results can be presented for each UE as mentioned in Table 29.4.

Simplified FMECA for System Dependability Assessment

Sometimes, it is required to perform a failure analysis within a short time. In this case, it is interesting to use a simplified FMECA to identify single points of failure without any statistical estimation with a FMECA table as mentioned in Table 29.13.

The equipment level definition is not detailed, and the failure modes are very simplified; only the worst failure modes are selected. For example, an LV switchboard is considered as a component with a unique failure mode, which results in a loss of insulation and the unavailability of the whole switchboard until repair.

TABLE 29.12 Customized FMECA table

Line *n*	Localization	Equipment reference	Equipment functions	Failure mode	Direct effects	Detection & consequences until normal state	Failure rate	UE1: "Loss of …"			UE2: "Loss of …"		
								UE happens?	Duration (h)	Unavailability (h/year)	UE happens?	Duration (h)	Unavailability (h/year)

Source: © Schneider Electric.

TABLE 29.13 Simplified FMECA

Localization	Equipment reference	Equipment functions	Failure mode	Direct effects	Consequence until repair and back to normal state	Frequency index estimation	UE1: "Loss of …"	UE2: "Loss of …"	UE3: "Loss of …"
							UE happen?	UE happen?	UE happen?

Source: © Schneider Electric.

29.3.2.2 Failure Combination Analysis

As mentioned before, the FMECA approach permits the assessment of a UE by performing a single contingency analysis. A single contingency analysis is acceptable as long as multiple failure sequences are negligible compared with single failure sequences. In some cases, when the system is highly reliable or available, this assumption is no more valid, and the dependability assessment needs an analysis of single and multiple contingencies to provide accurate results.

Principle
Failure sequences can be represented as shown in Figure 29.25.

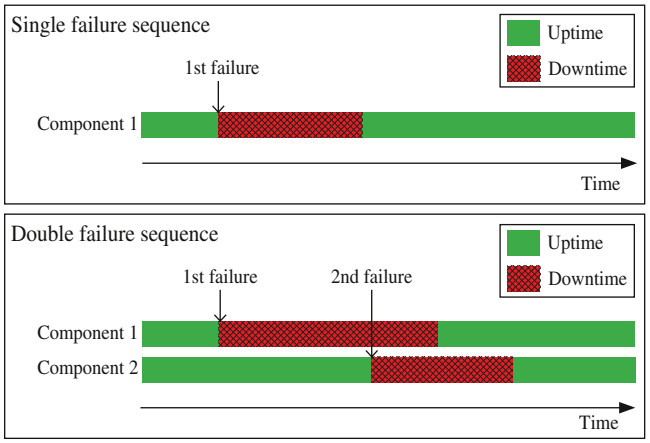


FIGURE 29.25 Single and double failure time sequence graphs. Source: © Schneider Electric.

The number of failure sequences becomes very important if multiple failure sequences are taken into account. Indeed, if a system is composed of n component with p failure mode for each element, possible failure sequences are:

- $(n \times p)$ sequences with a single failure
- 2^{n+p} sequences with two failures
- 3^{n+p} triple failure with three failures

That is why, in general, the use of a specific tool dedicated to system dependability modeling is essential. Several methods have been developed such as reliability bloc diagram, fault tree, event tree, Markov graph, and stochastic simulation.

Reliability Block Diagram
The principle of a reliability block diagram (Fig. 29.26) is described below:

- Each component failure mode is modeled by a block with the associated dependability parameters MTTF and MTTR.

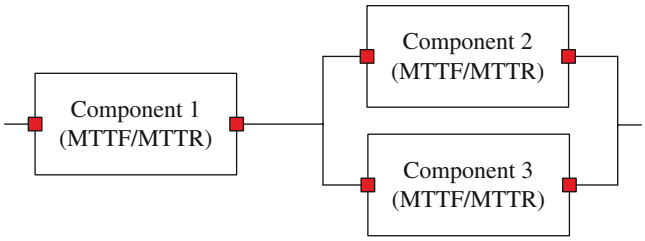


FIGURE 29.26 Reliability block diagram principle. Source: © Schneider Electric.

- Each UE is modeled by a reliability block diagram using blocks connected in series or parallel if redundant as mentioned in Figure 29.26.
- According to the model, the tool automatically computes the statistical estimation of UE probability and frequency and also a list of the minimal cut sets that are the unique combinations of component failures that lead to the UE.

Minimal cut sets can be used to determine main failure sequence contribution to UE probability or frequency.

Fault Tree
The fault concept is a top-down, deductive failure analysis in which a UE of a system is analyzed using Boolean logic to combine a series of events as mentioned in Figure 29.27. In opposite with FMEA concept that consists in an analysis of failure consequences and the identification of failures that leads to the UE, the fault tree process consists of representing the causes of the UE occurrence.

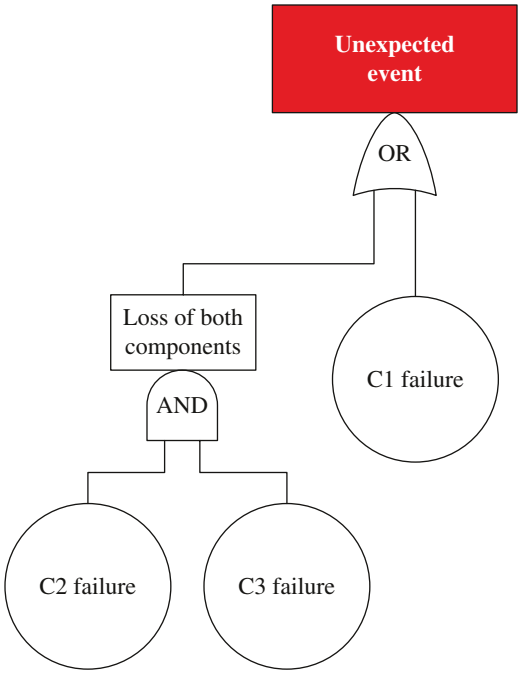


FIGURE 29.27 Fault tree example. Source: © Schneider Electric.

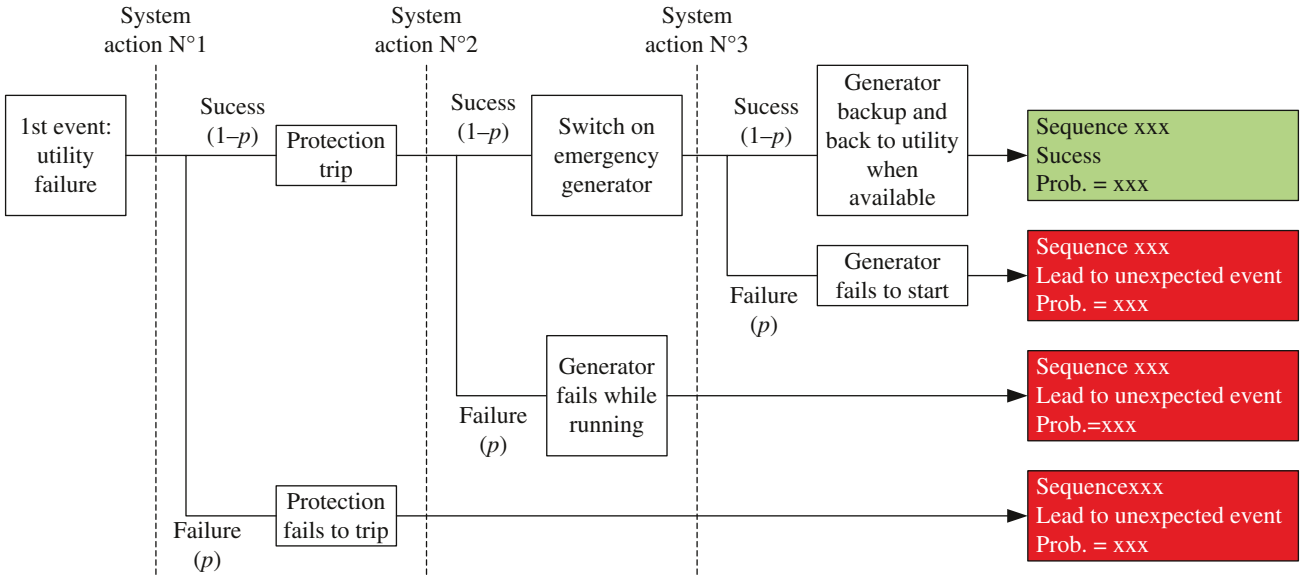


FIGURE 29.28 Event tree example. Source: © Schneider Electric.

The principle of a fault tree is described below:

- Each possible failure mode is modeled by a base event with the associated dependability parameters MTTF and MTTR.
- Each UE is modeled by a fault tree using logic gates "AND," "OR," "Voting OR (k/n)," "INHIBIT," etc. to represent the possible causes of the UE occurrence.
- According to the fault tree, the tool automatically generates the associate binary decision diagram of the UE and computes the minimal cut sets that are the minimal failure combinations that lead to the UE.
- UE probability, UE equivalent failure frequency, minimal cut set probabilities, and equivalent failure rates are computed by the tools.

Minimal cut sets can be used to determine main failure sequence contribution to UE probability or frequency.

Event Tree

As FMECA approach, the event tree is an inductive analytical diagram in which a failure (or an event) is analyzed by describing the chronological following events. The difference with a FMECA approach is that multiple failures are taken into account by using Boolean logic to determine the possible consequences as shown in Figure 29.28.

An event tree displays sequence progression, sequence end states, and sequence-specific dependencies across time. For each initiating event (1st event), a list of possible sequences are identified in terms of success or failure (UE is reached) and in terms of probability (unavailability and mean frequency).

After the analysis of all initiating events and all possible failures, UE probability and equivalent failure frequency are computed by the tool.

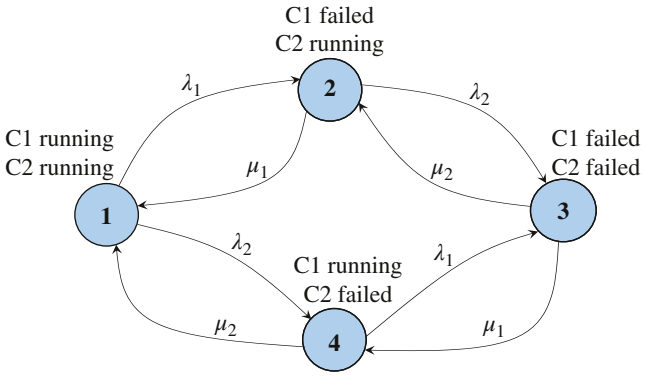


FIGURE 29.29 Markov graph example. Source: © Schneider Electric.

Discrete-Time Markov Chain

Discrete-time Markov chain is a mathematical model in which the system is modeled by its different states and the transitions from a state to another. Figure 29.29 shows a Markov chain of a system composed of two active and redundant components.

By assuming that transition is a random exponential law, the system can be mathematically modeled as follows:

$$\left[\frac{dP_1(t+dt)}{dt}\ \frac{dP_2(t+dt)}{dt}\cdots\frac{dP_p(t+dt)}{dt}\right] = \left[P_1(t)P_2(t)\ldots P_p(t)\right]A$$

where $A = \begin{bmatrix} a_{11} & \dots & a_{1p} \\ \dots & \dots & \dots \\ a_{p1} & \dots & a_{pp} \end{bmatrix}$ is the matrix of the transition rates.

The system can be computed to determine the state probabilities and then the transition mean frequency.

Time-Sequential Stochastic Simulation

As a Markov chain is a stochastic model, another possibility is the stochastic simulation of a system. The principle is to simulate the possible events according to their probability distribution and the system reactions to these events. After a sufficient number of simulations, statistical estimations of UE can be computed (Fig. 29.30).

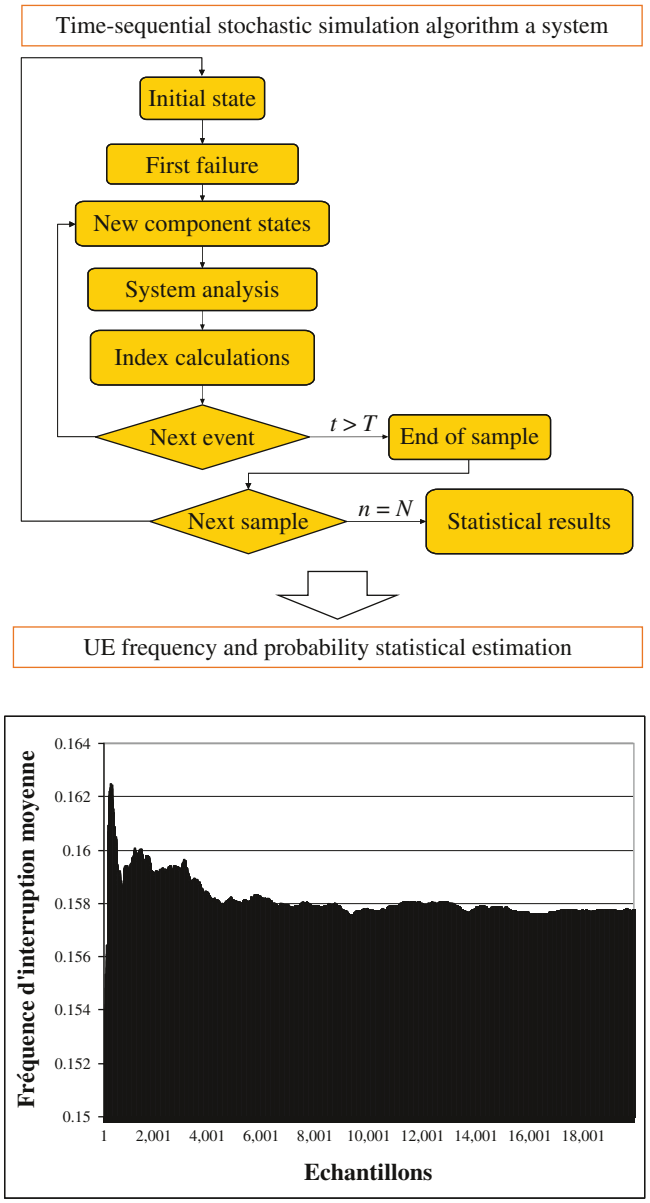


FIGURE 29.30 Time-sequential stochastic simulation algorithm of a system. Source: © Schneider Electric.

The simulation needs a model of the behavior of the system for each possible failure. This model can be based on a dedicated computer program, models based on Petri nets or others.

29.3.3 Advantages and Drawback of Dysfunctional Tools

The advantages and drawback of different dysfunctional tools are shown in Table 29.14.

29.4 APPLICATION TO DATA CENTER DEPENDABILITY

This part is dedicated to the data center dependability assessment. Some key points are presented to perform efficient and accurate dependability assessments as well as the benefits of dependability assessment during design phase and finally some key points to manage dependability assessment and tier standard architectures [1].

29.4.1 Benefits of System Dependability Assessment

The main difficulties to achieve a reliable and cost-effective design of a data center infrastructure that includes several critical systems with interdependencies are the following:

- Design too much redundancies on some equipment.
- Whereas some single failure points are not identified.
- Reliability levels could be not harmonized for the different systems (electrical power system, fuel storage, cooling system, water storage, auxiliary systems, monitoring systems, etc.).

Dependability assessment can be used at design phase during the data operation or during data center infrastructure refurbishment. It permits to:

- Estimate system dependability performances to prove that dependability targets are reached
- Identify and prioritize system weak points to design the right dependability improvements

During design phase, it is common to make iterations with the design team as mentioned in Figure 29.31.

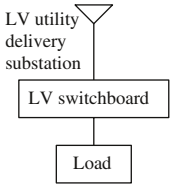
Here is a simple example of the use of a reliability study during a basic design phase. The goal is to provide sufficient redundancy to reach a failure frequency below 0.001/year.

TABLE 29.14 Advantages and drawback of different dysfunctional tools

	Simplified FMECA	FMECA	Fault tree	Event tree	Markov graph	Stochastic simulation
Time required to build the model	++	+	–	––	––	––
Large system	++	++	++	+	––	––
Complex behavior	–	+	+	++	++	++
Result accuracy	–	+	++	++	++	++
System weak point identification	++	++	++	++	–	––
Model verification	++	+	–	––	––	––
Easy to understand	++	++	+	++	––	––
Easy to use	++	++	+	++	––	––
Existing software tools	Implementable on a spreadsheet	Implementable on a spreadsheet	Many available tools for fault tree analysis	Many tools available for event tree analysis	Many available tools for Markov graph analysis	Existing tools for stochastic simulation of Petri net model
Summary	Suitable for quick analysis during architecture predesign/basic design phase Permit to identify main single failure points	Suitable for a quick dependability analysis during detailed design phase Permit to identify ALL single failure points and point out main contribution to failure frequency and unavailability	Best tool for a complete dependability analysis during detailed design phase Permit to identify main single failure points Permit to identify inadequate redundancies	Not suitable for exhaustive analysis because of too large number of initiating events to be studied	Not suitable for large systems Not adapted for quick identification of system weak points	Not suitable for large systems Not adapted for quick identification of system weak points
Remarks	—	Component failure mode level of details has to be fixed in adequacy with the expected precision of the dependability analysis and also the time and cost constraints	1. Some commercial tools provide contribution to unavailability but may not provide contribution to failure frequency. This can be an issue if the UE target is a failure frequency and not an unavailability 2. Some commercial tools propose an automatic dysfunctional model generation based on a system functional model provided by the user A particular attention has to be paid to the automatic generation: in many tools, the dysfunctional model is automatically generated using many assumptions that cannot be acceptable for the study			

Source: © Schneider Electric.

Design #1

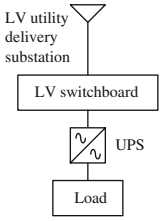


Reliability analysis results of Design #1

Case 1 - Base case Unexpected event : "Loss of the load more than 20 ms"	Failure frequency	Unavailability
	Failures/year	Hours/year
	9.65	**3.12**
	Contributions	Contributions
HV utility blackout	0%	1%
MV utility long interruption	9%	28%
MV utility short interruption	90%	9%
LV switchboard	0%	61%

The main weak point is "MV utility short interruption." Recommendation: Add a UPS with more than 3 min autonomy.

Design #2

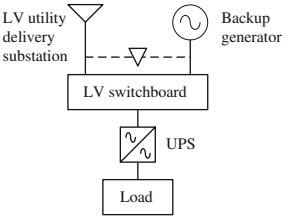


Reliability analysis results of Design #2

Case 2 - UPS with 5 mn autonomy Unexpected event : "Loss of the load more than 20 ms"	Failure frequency	Unavailability
	Failures/year	Hours/year
	0.89	**3.27**
	Contributions	Contributions
MV utility long interruption	98%	27%
LV switchboard	0%	59%
"UPS – Loss of UPS path" and "utility short interruption"	0%	0%

The main weak point is "MV utility long interruption." Recommendation: Add a backup generator.

Design #3

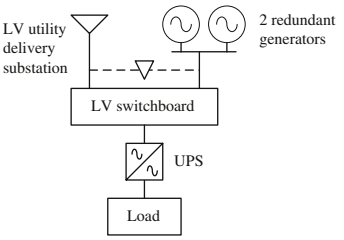


Reliability analysis results of Design #3

Case 3 - With UPS and genset Unexpected event : "Loss of the load more than 20 ms"	Failure frequency	Unavailability
	Failures/year	Hours/year
	0.04	**2.39**
	Contributions	Contributions
"Genset failure during standby" and "utility failure"	80%	1%
"Changeover – Fail to switch" and "utility failure"	1%	0%
"Changeover – Unexpected opening of both switches"	0%	0%
"LV switchboard failure"	2%	80%
"UPS - Short circuit on output"	6%	18%
"UPS – Loss of UPS autonomy" and "utility short interruption"	11%	0%

The main weak point is "genset failure and utility failure." Recommendation: Add a redundant generator.

Design #4



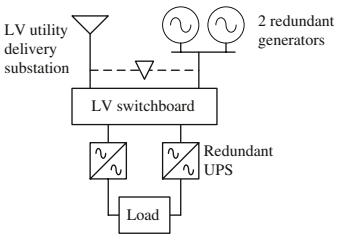
Reliability analysis results of Design #4

Case 4 - UPS and 2 redundant genset Unexpected event : "Loss of the load more than 20 ms"	Failure frequency	Unavailability
	Failures/year	Hours/year
	0.009	**2.36**
	Contributions	Contributions
"2 genset failures during standby" and "utility failure"	12%	0%
"Changeover – Fail to switch" and "utility failure"	3%	0%
"Changeover – Unexpected opening of both	1%	0%
"LV switchboard failure"	9%	81%
"UPS - Short circuit on output"	28%	19%
"UPS – Loss of UPS autonomy" and "utility short interruption"	46%	0%

The main weak point is "UPS failure and utility short interruption." Recommendation: Add a redundant UPS.

FIGURE 29.31 During design phase, it is common to make iterations (Design 1–6) with the design team. Source: © Schneider Electric.

Design #5

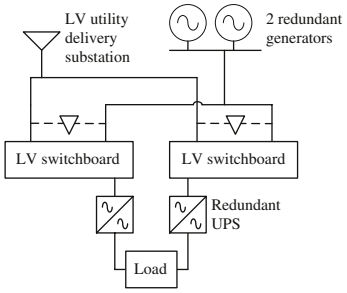


Reliability analysis results of Design #5

Case 5 - redunant UPS distributions and gensets	Failure frequency	Unavailability
	Failures/year	Hours/year
Unexpected event : "Loss of the load more than 20 ms"	**0.002**	**1.92**
	Contributions	Contributions
"2 genset failures during standby" and "utility failure"	48%	0%
"Changeover – Fail to switch" and "utility failure"	13%	0%
"Changeover – Unexpected opening of both switches"	4%	0%
"LV switchboard failure"	36%	100%
"Loss UPS1 and UPS2"	0%	0%

The main weak points are "2 genset failures and utility failure" and "LV switchboard failure." Recommendation: Add 2 redundant LV switchboards or redundant utility incomers.

Design #6



Reliability analysis results of Design #6

Case 5 - Redunant UPS distributions and gensets	Failure frequency	Unavailability
	Failures/year	Hours/year
Unexpected event : "Loss of the load more than 20 ms"	**0.001**	**0.002**
	Contributions	Contributions
"2 genset failures during standby" and "utility failure"	48%	0%
"Changeover – Fail to switch" and "utility failure"	13%	0%

The target is reached

FIGURE 29.31 (*Continued*)

29.4.2 Key Points for Data Center Dependability Assessment

29.4.2.1 Data Center Infrastructure Undesired Event (UE) Definition

Undesired Event Definition of a General Data Center

To ensure its main function, that is, the "IT process operation," the data center infrastructure consists of several systems that are briefly described in Figure 29.32.

A simple functional analysis is performed below (Fig. 29.33):

F1: Provide IT process to customer (servers and communication with access provider):

F1.1: Data center operations (IT, power system, and mechanical systems)

F1.2: Provide electrical power for data center loads

F1.3: Provide water supply for cooling systems

F1.4: Provide reserve energy to emergency power plant

F1.5: Ensure data center process during harsh environment

F1.6: Ensure data center security

F1.7: Ensure maintenance operation

F2: Ensure people safety

F3: Prevent environment pollution (noise, chemical, etc.)

This analysis is not exhaustive but permits to keep in mind that:

- The main functions are the following: "IT process operation," "ensure people safety," and "prevent from environment pollution"
- To ensure IT process, many systems are to be taken into account

The UEs of a "classical" data center infrastructure can be deduced from the main function degradation as mentioned in Section 29.3.1.1:

UE1: Loss of IT process

UE2: Safety risk

UE3: Environment pollution

Dependability and Safety Targets

Concerning the UEs related to safety and/or pollution (UE2 and UE3), classical targets can be determined according to standards or with qualitative target as follows:

- No single failure that affects the UE
- No failure combination (including an undetected failure and a 2nd failure) that affects the UE

IT process dependability targets are set according to the gravity. As mentioned in Section 29.3.1.1, UE1 can be decomposed in several "sub-UEs" if the consequences depend a lot on the IT process failure (loss of IT application during more than a minute, loss of data, etc.).

Moreover, depending on the customer specificities, the UE target can be defined in terms of unavailability (%) or failure frequency (failure per year) or both. Classically, the goal of a data center infrastructure is to provide a service without any disruption so the target should be defined in terms of failure frequency. However, some large customers can define their target in terms of unavailability because they can handle a data center blackout as their IT process is backed up by other sites.

An example of the decomposition of UE1 is given in Figure 29.34.

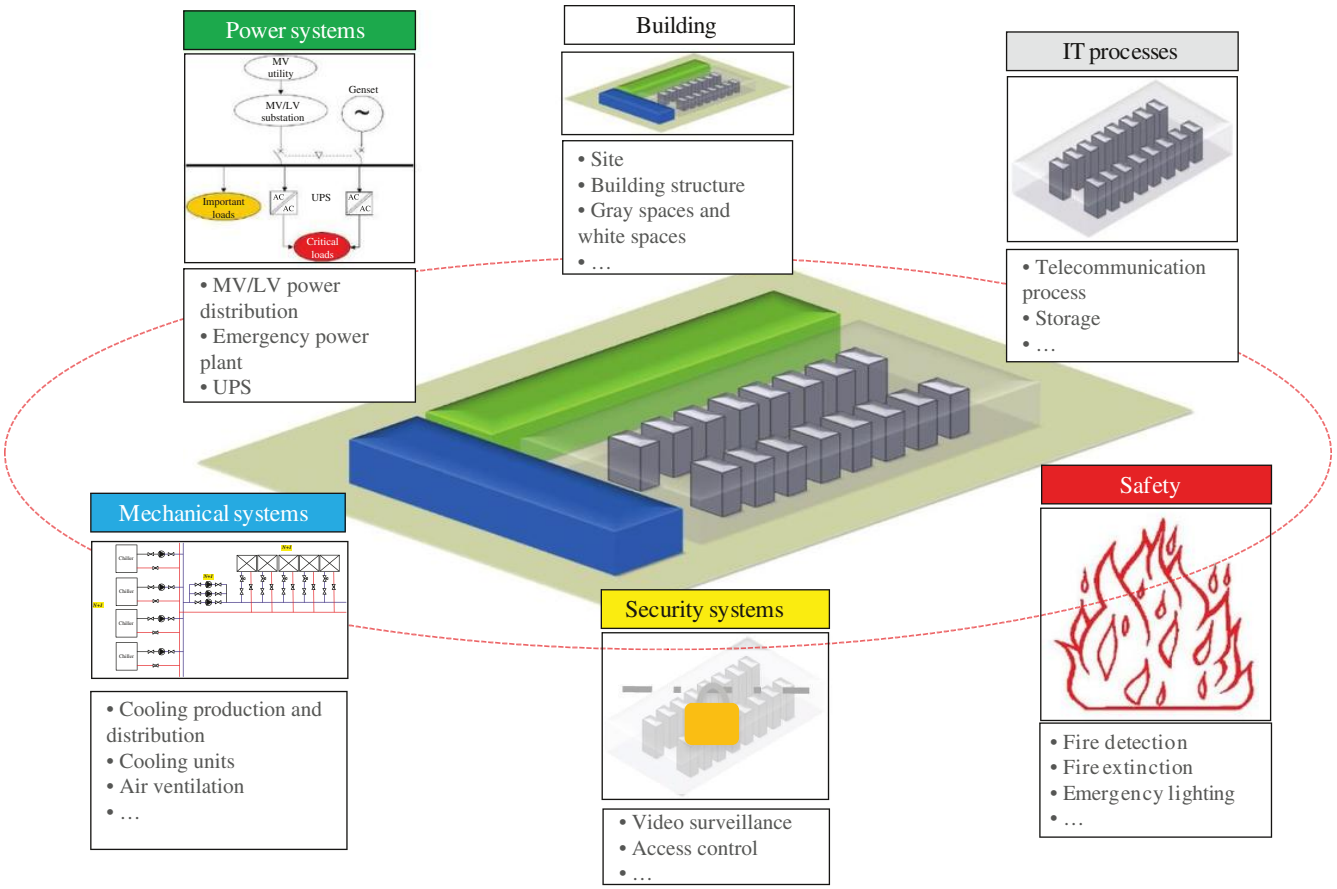


FIGURE 29.32 Data center infrastructure overview. Source: © Schneider Electric.

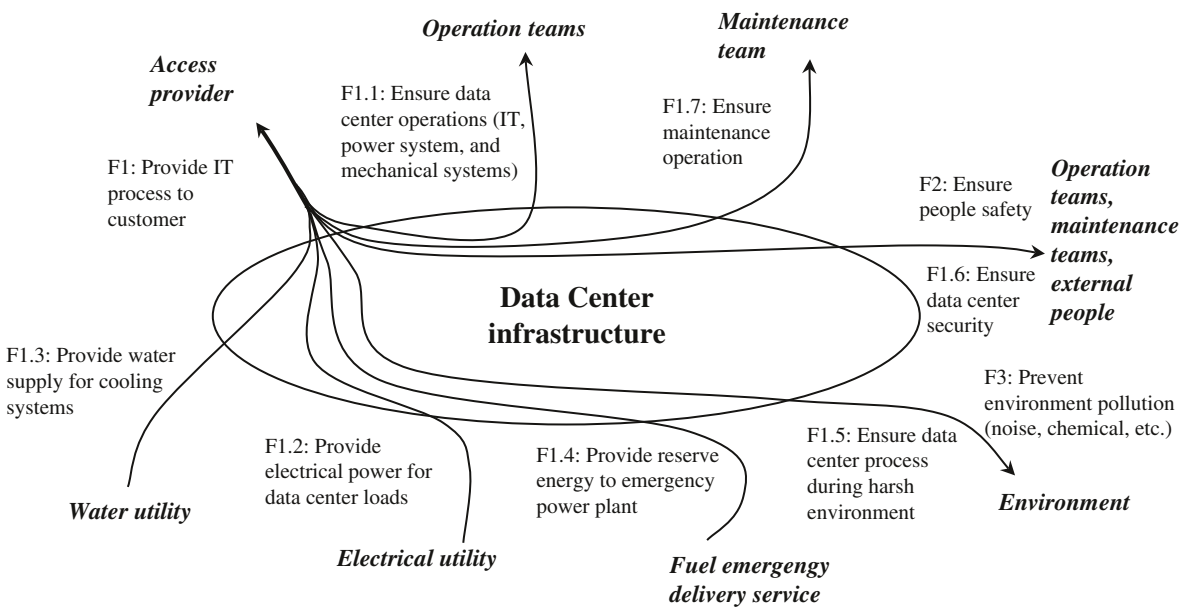
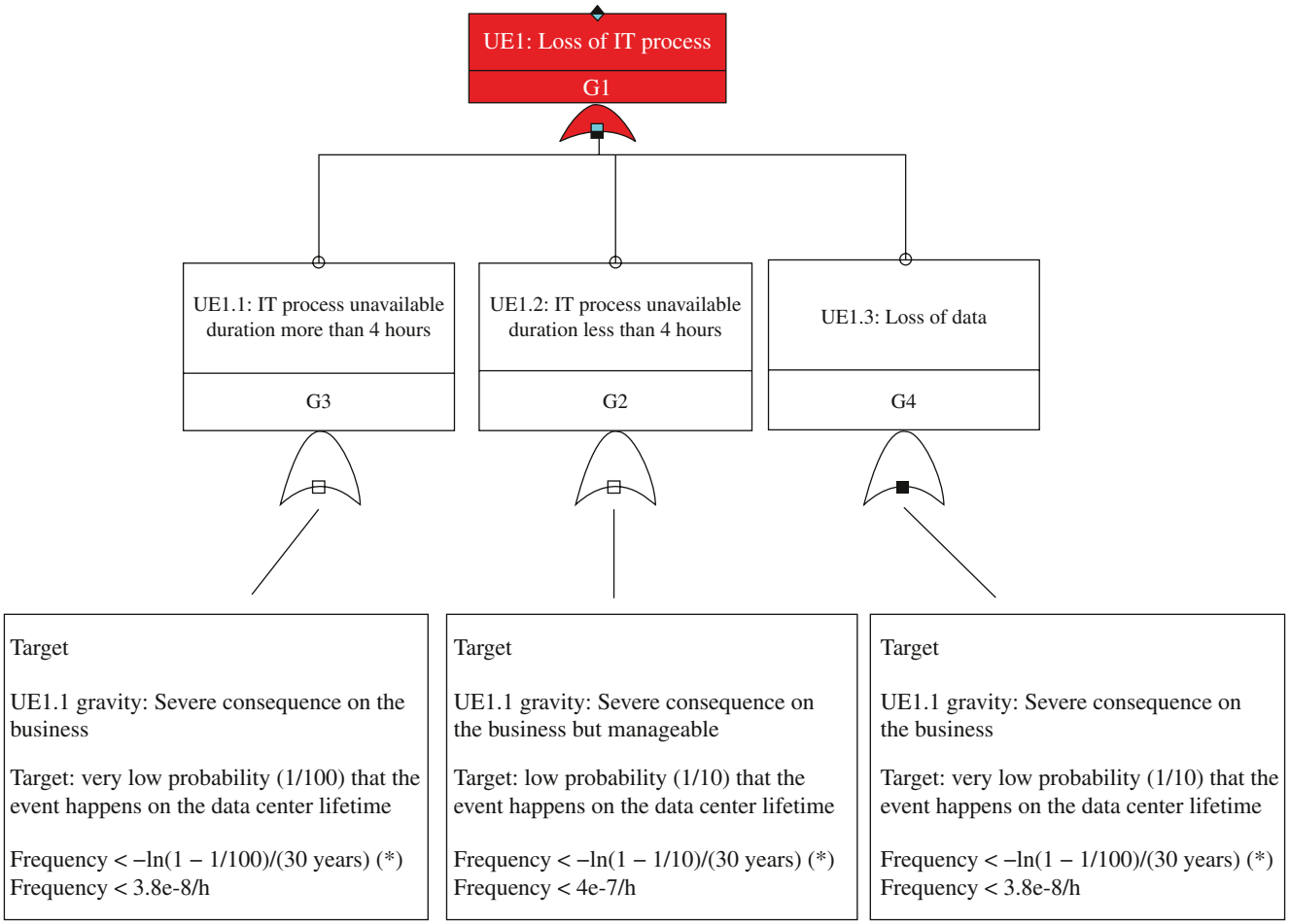


FIGURE 29.33 A simple functional analysis. Source: © Schneider Electric.



(*) The frequency target is calculated using the reliability indicator
R(t) = e–Frequency.lifetime > probability to "Experience no failure during the installation lifetime"

FIGURE 29.34 UE1 "Loss of IT process" decomposition. Source: © Schneider Electric.

29.4.2.2 System Data Collection

Technical Scope
When performing a dependability assessment, it is essential to taken into account the whole system to ensure no useless redundancies. Figure 29.35 shows the global scope of a data center infrastructure to be studied in a dependability analysis.

Technical Data Collection
Data to be collected to perform a dependability assessment are summed hereafter Figure 29.36.

Particular attention has to be paid to the following aspects.

Automation
Automation behaviors have to be characterized by:

- The equipment involved in the functions (sensors, logic, and actuators)
- Automation function global overview (to be able to determine the consequences of different possible failures)

Equipment Operating Modes and Degraded Modes
The behaviors of the system depend also highly on:

- Equipment tolerance to supply interruptions (electrical supply, water supply, air conditioning/ventilation, etc.)
- Starting time of equipment or functions after a blackout

Redundancies
Equipment redundancies have to be determined according to the system architectures (electrical, cooling, etc.) and the equipment specifications checking:

- Equipment limitation that could lead to partial redundancy
- Possible common cause failures linked to:
  - Interdependencies of events (design/production/installation errors, environment, human factor, etc.)
  - Auxiliary systems (power supply, water supply, SCADA systems, etc.)

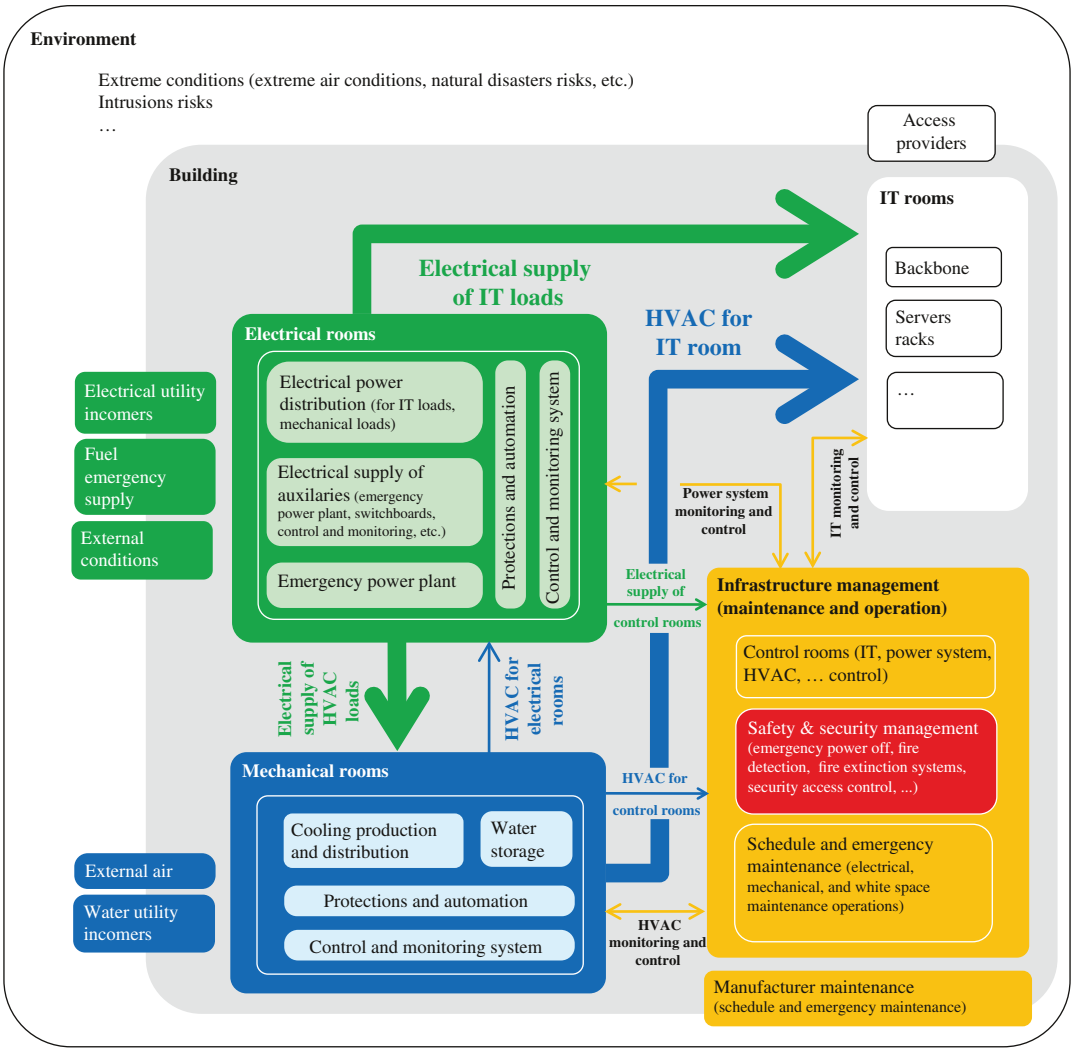


FIGURE 29.35 Technical scope of the data center infrastructure to be studied for an overall dependability assessment. Source: © Schneider Electric.

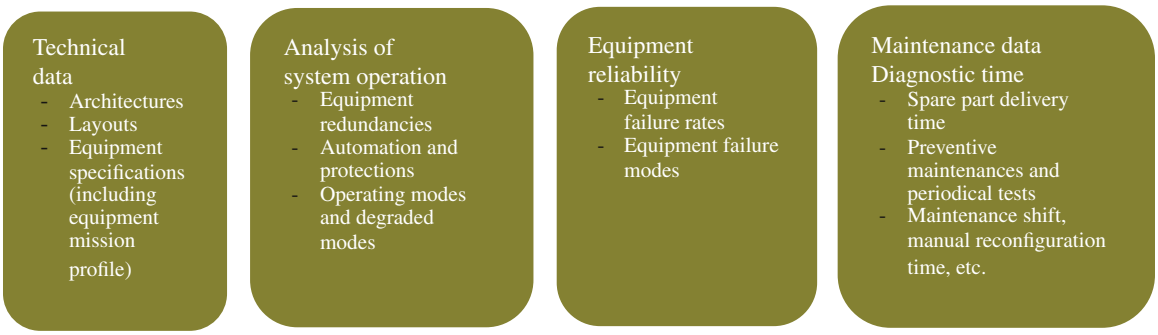


FIGURE 29.36 Dependability assessment data collection. Source: © Schneider Electric.

Failure Detection
To determine the detection time of a failure, failure detection means have to be described including:

- Monitoring system
- Periodical tests (frequency, diagnostic coverage, etc.)
- Maintenance operator intervention time

Reliability Data
Equipment reliability data can be determined by several sources:

- Field failure rate sources
- Manufacturer databases
- Reliability handbooks
  - IEEE Gold Book Std 493
  - EIReDA 1998 (French field experience of nuclear power plants on electrical and mechanical equipment)
  - NPRD from Reliability Information Analysis Center (mechanical and electrical equipment)
  - EXIDA—*Safety Equipment Reliability Handbook* (Ed. 2)
- Experts with great experience of field failures could also provide valuable information on equipment reliability.

Several warnings on failure rate assumptions are highlighted below:

Field failure rate/predictive electronic studies: Theoretical electronic reliability studies performed by manufacturer are sometimes available. These failure rates are determined according standards like IEC62380 or MIL-HDBK 217 F. These electronic reliability studies are performed to optimize the electronic design, but the failure rate value could be pessimistic. When no field experience failure rates are available, it is commonly accepted in system dependability calculation to use theoretical values.

Failure rate validity: Failure rates are valid under certain conditions (mission profile, lifetime) that need to be highlighted and checked if adequate with real conditions.

Failure modes: Failure mode contributions to global failure rate are sometimes mentioned in field failure sources. When not available, a short simplified FMEA analysis of the equipment based on equipment internal architecture data and field experience of the equipment can bring sufficient information to determine failure mode contributions with great accuracy.

Common Mode Failures
As data center infrastructures are high reliable/available systems, common mode failure identification and quantification are major key points for an accurate dependability analysis.

Maintenance Data
To determine equipment MDT, it is essential to take into account:

- The time for failure diagnostic
- The time for spare part delivery

Also programmed unavailability of equipment for preventive maintenance (verification, cleaning, preventive replacement, etc.) or for installation work during installation evolution phase needs to be taken into account.

Lack Out of Data
Sometimes, some data are not available. In this case, assumptions have to be made, highlighted, and verified further in the project.

29.4.2.3 Dependability Management During Project Phases

During the project cycle, dependability is a main customer requirement that is applied on each project stage as mentioned in Figure 29.37.

Preliminary Outlines
During preliminary studies, the customer defines its needs and the global architecture of the data center with the help of a design office:

- Site selection
- Building main characteristics
- IT process definition
- IT rack-rated power

At this stage, a preliminary risk analysis is performed to identify UEs and determine dependability targets. This step is performed by the customer, and a competency in dependability engineering provided by the customer or an external design office.

Basic Design
Within its proposal, the contractor provides a simplified dependability analysis to confirm that its basic design reaches the dependability requirements. At this step, a simplified analysis is sufficient as the design may change during detailed studies due to technical issues or customer requirement modifications. Moreover, a simplified analysis is useful when making iteration with the design team.

The simplified analysis can be limited to a simplified FMEA analysis but including all the data center infrastructures (Figure 29.35). In addition, few calculations of multiple contingency analysis on non-reliable equipment (utility/power plant, chillers, pumps, etc.) will provide that redundancies are correctly designed.

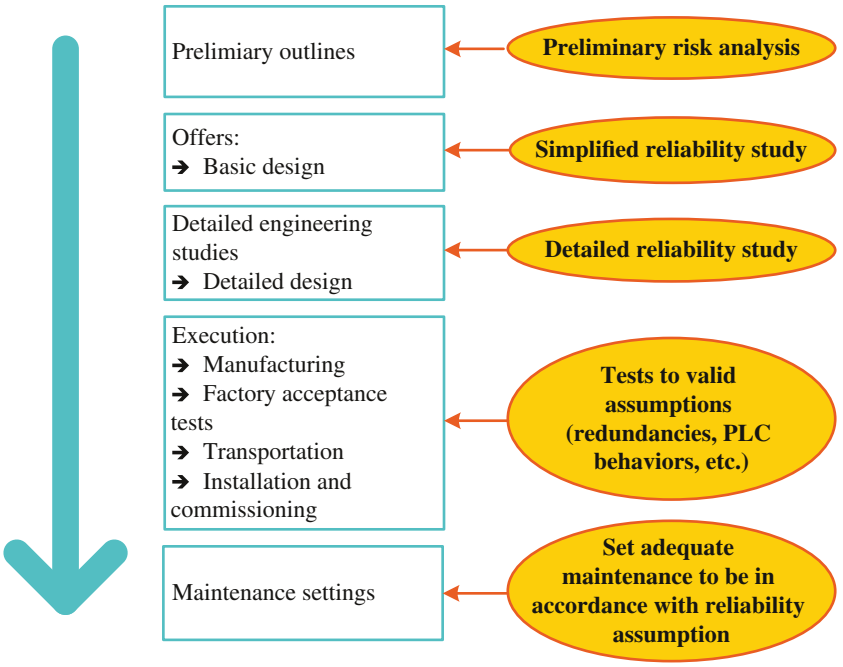


FIGURE 29.37 Dependability management during project cycle. Source: © Schneider Electric.

Detailed Studies

During detailed design stage, an efficient way to proceed is to make dependability checks during each system design phase before providing a complete dependability analysis.

When the detailed design is sufficiently defined, the dependability detailed analysis can be performed using assumptions for maintenance data.

Project Execution

Even if detailed engineering stage has been completed, some modifications may happen and lead to an update of dependability analysis (It is a common issue that design modifications are done at this stage without checking dependability consequences).

During installation and commissioning phases, assumptions of the dependability analysis need to be confirmed by inspection and tests to ensure that the system satisfies the dependability requirements.

On-site Mintenance Setting

The on-site maintenance has to be set to match dependability analysis assumptions. Some iteration may happen: maintenance assumption modifications and dependability analysis update to ensure the dependability level remains unchanged.

Some difficulties of managing overall dependability analysis during project phases are summed up below:

- During basic design and detailed design phases, the contractor is responsible for the overall dependability assessment. Some difficulties may happen so the contractor needs to synthesize the overall dependability based on several dependability analyses of each system (electrical, HVAC, security, etc.) provided each design offices. To minimize these difficulties, the dependability targets have to be clearly defined for each system at the beginning of the project (at design phase). Moreover, during the detailed design phase, as many different systems are interdependent, it is better if only one entity performs for the overall detailed dependability analysis.
- The accuracy of dependability assessment may be degraded if:
  - Some parts of the systems are not included in the system, particularly for auxiliary systems. Typically, the IT process is frequently separated from the rest of the infrastructure; this can lead to misunderstandings and architecture design problems (oversize redundancy on terminal electrical power distribution or minimize the risk to "lose at the same time the entire data center IT rooms").
  - The reliability expert who performs the analysis is not experienced with engineering and exploitation of each system.
- A common problem is that dependability is considered during design phase but not after project execution and operation phase. The customer shall keep its dependability analysis updated during all the phases of its installation.
- As data center infrastructures are intended to be upgraded severally during its lifetime, overall dependability analysis needs to be updated for each phase to ensure dependability level during all phases.

29.4.3 Tier Classification and Dependability Assessment

A basic description of Tier classification according to TIA 942 standard is given below:

Tier I: "Basic capacity"—no redundancy required
Tier II: "Redundant capacity components"—redundancy on non-reliable equipment
Tier III: "Concurrently maintainable"—Tier II requirements + each equipment can be removed and repaired without a data center blackout
Tier IV: "Fault tolerant"—Tier III + fault tolerant architecture (no single failure point)

According to the data center IT business criticality, the tier classification is a powerful tool to set adequate Tier level according to the customer business and determine which equipment needs redundancy in the infrastructure design (electrical, mechanical, building, etc.). The benefits and drawbacks of the classification are listed below (Table 29.15).

During project phase, an efficient procedure is described below:

- During preliminary study, the data center owner expresses the criticality of its IT business and then is able to set the adequate Tier level according to the Tier standard.
- During basic design phase, the contactor provides a simplified FMECA to prove that the Tier requirements are reached.
- During detailed engineering studies, the contactor provides a detailed FMECA study to prove that Tier requirements are satisfied.

TABLE 29.15 Benefits and drawbacks of tier classification

Benefits of Tier classification

Simple classification
- ⇨ Understandable and accessible to everyone
- ⇨ Powerful to perform a quick assessment at design phase

Take into account all critical systems (electrical power system, HVAC systems, critical auxiliaries, …)

Good levels of dependability classification that provides a good frame of references

Disadvantages of Tier classification

Pessimistic assumption on utility dependability that leads to oversize some redundancies

The huge gap between Tier III and Tier IV leads sometimes to design Tier III infrastructure with additional redundancies without matching Tier IV requirements

In some cases, "*N*+1" design of some equipment is not enough but the classification does not take into account

Emergency and preventive maintenance are not taken into account

All equipment failure modes are not systematically taken into account as well as the failure detection

Source: © Schneider Electric.

FURTHER READING

ANSI TIA-942: Telecommunications Infrastructure Standard for Data Centers. Arlington, VA: Telecommunications Industry Association. Available at www.tiaonline.org. Accessed on November 22, 2020. 2014.

Billinton R and Alan RN. *Reliability Evaluation of Power Systems*. 2nd ed.. New York: Plenum Press; 1994.

Cabau E. *Introduction à la Sûreté de Fonctionnement*. Cahier Technique Schneider Electric Nr 144. Grenoble: Schneider Electric; June 1999.

IEC. *IEC 60300-3-1: Dependability Management: Part 3-1 Application Guide--Analysis Techniques for Dependability Guide on Methodology*. 2nd ed.; 2003. https://webstore.iec.ch/preview/info_iec60300-3-1%7Bed2.0%7Den.pdf. Accessed on August 4 2020.

IEC. *IEC 60812: Analysis Techniques for System Reliability—Procedure for Failure Mode and Effects Analysis (FMEA)*. 2nd ed. Geneva: IEC; 2006.

IEC. *CEI-IEC-61165: Application des techniques de Markov*. 2nd ed. Geneva: IEC; 2006.

Logiaco S. *Electrical Installation Dependability Studies*. Cahier technique Schneider Electric Nr 184. Grenoble: Schneider Electric; December 1996.

Lonchampt A, Gatine G. *High Availability Electrical Power Distribution*. Cahier Technique Schneider Electric nr 148. Grenoble: Schneider Electric; 1990.

Villemeur A. *Sûreté de fonctionnement des systèmes industriels*. Paris: Eyrolles; 1988.

30

COMPUTATIONAL FLUID DYNAMICS FOR DATA CENTERS

MARK SEYMOUR
Future Facilities, London, United Kingdom

30.1 INTRODUCTION

One of the principal issues in maintaining the very high availability required for mission-critical facility in data centers is how to deliver cooling effectively and efficiently to all the equipment, wherever it is in the room.

The significant power densities in modern data centers make this a substantial task, which justifies significant focus if it is to be implemented in a way that achieves the cooling objective at low cost and without significantly interfering with operational objectives. People often forget that the primary purpose of the cooling system is to cool the electronics; historically, data centers were simply treated in the same way as other occupied spaces.

Figure 30.1 shows that to cool the electronic component it is important to take responsibility for configuration/design of cooling for the IT equipment (ITE) itself, how it is cooled in any rack or cabinet, and how those cabinets are cooled in the data hall. The broken line represents the break in ownership of the problem: at the electronics scale, the manufacturer is responsible; at the room scale, the facility or IT manager is responsible. There is an obvious danger that the rack/cabinet configuration and cooling fall between the two.

Modern environmentally considerate designs are often very data center specific. They use fluids—most commonly air, but increasingly liquid—to deliver the cooling to the IT and carry the heat away from it. This is a challenge for both the initial design and the ongoing management. Unlike a box of electronics where the internal components remain mostly fixed for the life of the system, what is installed in many data centers changes frequently (in some cases even on a daily basis).

Where air is the medium employed for cooling in the rooms of the data center facility holding the ITE (commonly known as data halls), it is often beneficial to analyze and thus optimize the cooling design and ITE configuration to make best use of the cooling available. This helps avoid hot spots that otherwise tend to build up as the data center evolves over time. The increase in popularity of economization or free cooling has also resulted in computational fluid dynamics (CFD) being applied to airflow around data centers. Analysis will also be useful in the design and configuration of liquid-cooled systems, but this will commonly be a task allocated to a cooling professional, and so this is not the primary focus of this topic.

One of the key issues about a data hall is that almost every hall is unique. For one reason or another and unlike typical electronics cooling problems (where the equipment configuration is defined and fixed during design), the design of a data hall varies from one installation to another. What is more, data hall configuration will continue to vary over time as new equipment items are deployed; older equipment items are removed, moved, or upgraded; and applications are deployed based on current demand. This creates a scenario where a "one-size" solution does not fit all. CFD is therefore deployed as a tool for capacity planning during operation as well as design assessment and optimization. In operation it forms a key part of what is being referred to as "The Digital

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

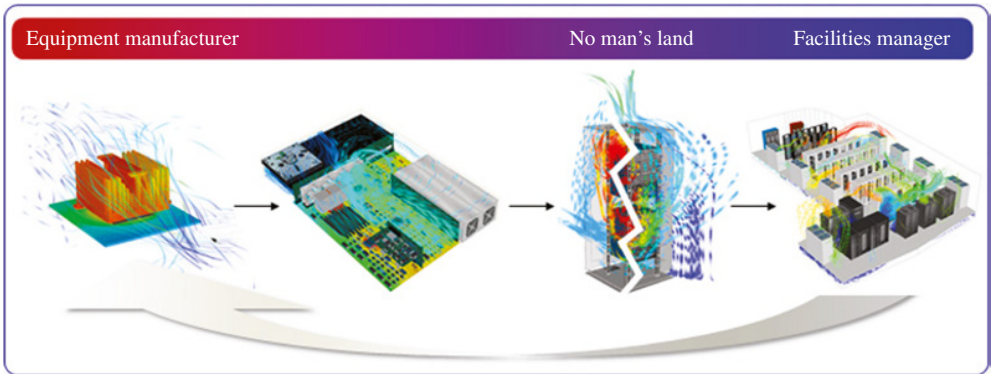


FIGURE 30.1 Data center cooling is required at all scales to effectively cool equipment.

Twin"—a computer-based replica of the data center that can be used to understand what is happening now and what might happen in the future due to changes while also enabling capacity planning considering parameters such as space, power, weight, and networking alongside cooling.

Because the cooling is achieved via a fluid cooling medium, the only theoretical technique that can predict the complex performance of the cooling system—the cool air delivery and hot air scavenging—is CFD. Since the CFD model replicates the features of the data center and must evolve alongside the real facility, it is often referred to as the "Digital Twin." In the drive for increasing energy efficiency, the heat rejection performance to the outside world is also becoming increasingly important, so airflow around the building and its power and cooling infrastructure is also a frequent subject for CFD modeling. In fact, many components in the supply chain, including the ITE and the support infrastructure such as cooling systems, are themselves designed using CFD.

30.2 FUNDAMENTALS OF CFD

CFD is the use of computers (using numerical methods) to analyze the likely behavior of a fluid (liquid or gas). It accounts for the stimuli that promote or restrict movement of the fluid within and around geometrical objects in the region for which the analysis is being made.

The term "fluid" is used because the techniques can equally be applied to liquids or gases. For example, an analysis could be made both for the chilled water flow inside the chilled water circuits of a finned heat exchanger and for the airflow between the fins around the chilled water circuits of the heat exchanger. In fact, the analysis is not limited to the flow of only a single fluid, liquid, or gas, but can in principle analyze both the chilled water flow (liquid) and airflow (gas) simultaneously, accounting for the heat transfer between the two fluids (in this example, water and air). It can also account for the impact of temperature variations in the two fluids and indeed the solids comprising the heat exchanger itself.

Already from this simple example you will begin to realize that the CFD methodology is very general and can be applied to many applications, not just cooling ITE in a data hall (largely using air as the fluid). In fact, the techniques are even more powerful than indicated thus far: they can be applied not only to segregated fluids but also to fluid mixtures as well. They can allow for changes in state (e.g., evaporation and condensation of water), known as two-phase problems. Additionally, they can be applied to time-dependent variations in the fluid flow via "transient analysis", rather than just the conditions at a snapshot in time as though nothing ever changes. The latter is commonly known as a "steady-state analysis."

It is, therefore, appropriate to outline some of the fundamental principles of CFD before considering, in somewhat more detail, the uses and application of CFD to data centers.

30.2.1 Basic Principles

When performing a CFD analysis, one uses a computer program to solve a set of equations commonly known as the Navier–Stokes equations. The equations mathematically define the conservation laws for mass, momentum, and energy and were first derived by Claude Louis Marie Navier in 1827. The equations were in fact independently derived by several people over the next two decades, and, in 1845, George Gabriel Stokes published his derivation (from a different perspective) of the equations. Hence, they subsequently became known as the Navier–Stokes equations. However, for most practical applications and in fact all but the simplest of scenarios, there is no analytic solution to these simultaneous equations.

In order to solve the equations, it is necessary to divide the space up into a grid of cells; for each cell, the Navier–Stokes conservation equations can be written down as nonlinear partial differential equations. The detailed form of this mathematics is not important in this context, but it is important to understand that they can now be used to understand the fluid flow and heat transfer.

There are several equations. The most fundamental are the basic equations describing the velocity of the fluid in the cell.

There are generally three equations for three orthogonal velocity components (in a rectangular grid, one for each axis direction: *X*, *Y*, and *Z*). Additional equations can then be added to represent the transport of other things such as thermal energy (for temperature) or contaminants. Finally, the equation for continuity (mass conservation) is added, from which pressure can be derived.

The property being calculated in each equation set is often termed a "variable," and the collection of values for each variable (typically one value per cell) covering the entire calculation volume is often called a "field of data." From a conservation perspective, the fluid (or other transported property) entering and leaving a cell through its faces must be consistent with each other and with its neighbors. This transfer of fluid through a surface is known as a "flux."

The actual formulation of the equations is dependent upon the numerical method employed (Section 30.2.2). As there is no analytic solution for most practical problems (i.e., the equations cannot be rearranged so that the answer can be calculated directly), the equations must be solved numerically. Essentially, numerical solution is achieved by making a "guess" at the answer and inserting the values into the equations for every cell and every variable. Once inserted, the inconsistency in fluid flow or other variables (e.g., heat flow) into and out of any given cell can be calculated, allowing a correction to be made to the guess.

The process can then be repeated iteratively, and, if successful, the error reduced to an acceptable level. This process is known as "convergence," and, when successful, the solution is termed a "converged solution" (Section 30.2.5). It is important to note that being converged does not mean that the prediction of flow (or any other transported variable) is accurate, just that the numerical errors (the level of inconsistency between the different values) have been reduced to less than a predefined "acceptable level."

There are other numerical reasons for the solution to be considered only an approximation of reality. The most common are as follows:

- Breaking the model (the representation of the data center, what is in it and affecting it) into pieces to form a discrete grid and the potential for numerical diffusion (Section 30.2.4).
- Not all the equations are a pure description of the physics. In particular, the turbulence model is an empirically derived relationship intended to capture the gross effects, such as mixing, that occur as a result of smaller-scale fluctuations. For some applications (aerodynamics, for example), the equations may be tuned specifically for that niche (it may even be essential to do so for the technique to be accurate enough to be useful).
- The way in which we define the model: The physical geometry, airflow, and thermal boundary conditions that define features inside the calculation space (often known as the "solution domain"). Also, the representation of the items that define interaction with the surrounding environment and other elements of the model. These representations are often referred to as "boundary conditions" and will be discussed further in Section 30.2.3.

30.2.2 Numerical Methods

There are many approaches used in numerical mathematics to solve sets of differential equations like the Navier–Stokes equations. The method adopted affects the formulation of the equations for the solution, the way the space is broken down into the cells or grid, and how much computational power (in terms of processor and memory resources) is required. It even affects whether the solver (the computer program solving the equations) is likely to always produce a solution for any scenario or whether it will require special attention/control to achieve a solution at all.

The following are methods that the reader is most likely to encounter when considering CFD for application to data centers.

30.2.2.1 The Finite Volume Method

The finite volume method is a numerical method based on dividing the space into control volumes (or cells), thereby discretizing the space into a mesh of cells each surrounding a data point. By casting the Navier–Stokes conservation equations onto the mesh of cells, by definition of conservation, the flux (airflow, heat flow, etc.) at any face of a cell leaving or entering the cell is equal to the flux entering or leaving the neighboring adjacent cell. Figure 30.2 shows for a 2D slice how, in what is known as a staggered grid approach, the scalar quantities (e.g., pressure (*P*) and temperature (*T*)) are calculated and stored at the cell center,

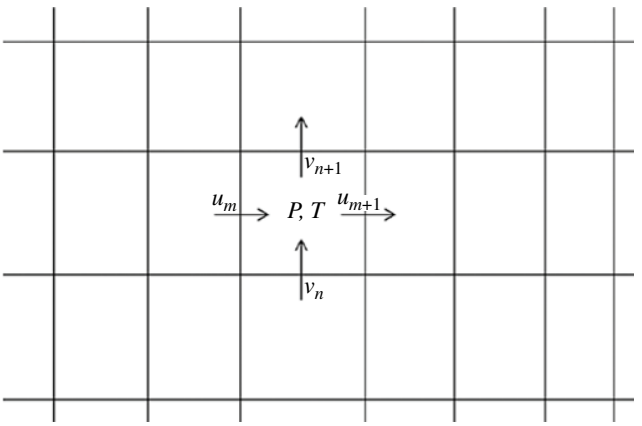


FIGURE 30.2 Slice through mesh showing connectivity, values, and fluxes.

while the velocities (u,v) and hence the fluxes (mass flow rate of the air) are calculated and stored on the cell face.

One of the advantages of this method is that it is easier to write down the equations: the fluxes through the cell faces are a direct result of the values in the adjacent cells. For example, the flow through two adjoining faces is directly dependent upon the difference in pressures in the two cells. The disadvantage is that it has just one point per cell—there is no information about gradient—and this results in numerical diffusion that can be addressed to some extent by the use of more sophisticated numerical algorithms, such as higher order differencing schemes.

At the time this chapter was written (April 2019), the finite volume method is by far the most common approach used for data center modeling; it has been demonstrated to be capable of producing usefully accurate predictions in an acceptable period of time. The remainder of this section will therefore assume this approach unless otherwise stated.

30.2.2.2 The Finite Element Method

The finite element method is more commonly used for structural analysis problems, although it has also been used for commercially available CFD programs. The advantage of the finite element method is that it has multiple points in each element (required in structural analysis to determine the stress): it carries more information per cell and, implicitly, the gradient or a variable.

Ironically, the use of multiple points is also the disadvantage of the finite element method. This additional information makes the approach much more computationally expensive, not only in terms of memory but also in terms of computational speed. For fluid dynamics applications, it is often therefore too expensive, although it is used particularly where the CFD software may be intended to work alongside structural analysis software, making it easier to integrate the two technologies.

30.2.2.3 Methods for Faster Solution

The potential flow method *was adopted by some* because the assumptions allow results to be produced in simpler and quicker solutions. The key assumption is that the flow is a perfect fluid; that is, that it is inviscid (has no viscosity) and it is incompressible. The result is that the flow will be irrotational, such that there is no flow normal to the streamlines (streamlines show the convective flow path of the fluid/air or the conductive path for heat). To use an analogy, it would be like a river flowing continuously along the river path with all the water flowing smoothly along it, around bends, and over and around rocks, all without vortices, circulations or cavities along the way. In practical terms, this will result in flow that cannot separate from a surface, and so, for example, flow may appear to unrealistically flow around corners.

While potential flow-based tools still exist, researchers are looking to less simplified approaches that might still deliver significant increases in speed. Two methods currently receiving significant attention are fast fluid dynamics (FFD) and the lattice Boltzmann method (LBM). FFD is a projection method for solving the Navier–Stokes equations. Although less accurate than traditional solution techniques, it is naturally faster than traditional solution methods. It is also easily parallelized and therefore highly suitable for implementation on graphics processing units (GPUs). Unlike FFD, LBM does not solve the Navier–Stokes equations for conservation; rather, it solves the lattice Boltzmann equations. The solution can be interpreted as fictitious particles that perform consecutive propagation and collision processes over a discrete lattice mesh. For multi-scale problems, it will struggle to match the accuracy of traditional Navier–Stokes CFD methods, since incorporating local refinement is well established for the latter. However, LBM has the advantage that it is potentially much faster and easily parallelized on GPUs, potentially giving real-time simulation and a solution that is intrinsically transient. Current challenges being addressed include developing a consistent approach to coupling thermal and hydrodynamic processes and solution dependencies on lattice size.

30.2.2.4 Other Methods

These are by no means all of the numerical methods available to solve partial differential equations, and, because of the size of a data hall and the number of objects and features inside it, there is continuing effort to identify faster solution techniques with acceptable accuracy. However, in the opinion of the author, these represent the techniques currently most commonly in use or under consideration for data centers.

30.2.3 What Defines a CFD Model

A fluid dynamics simulation is undertaken to understand the flow (and often the heat transfer) within the fluid volume. The primary variables are the mass, momentum, and energy. To define a model for simulation, a physical description must be made in a way that can be represented through one of the methods described in Section 30.2.2. Plainly speaking, one builds a three-dimensional (3D) computer-generated model of the data hall.

The first thing to define is the size of the computational space in which the calculations are going to be made. This is commonly the shape of a shoe box—a rectangular block with dimensions appropriate for the applications concerned. This space—the solution domain—can be much more arbitrarily shaped in some cases, but for simplicity's sake we will assume a rectangular box here.

Once defined, you can imagine that if this box were filled with air or any other fluid, absolutely nothing will happen inside it unless there is some input or force to disturb the equilibrium. Of course, the six faces of the solution domain represent an interface to the "world," which has an impact on the inside. Even as a sealed box, the faces can transfer heat, causing air close to the wall to heat up and rise. Once the air is moving, the walls of the box have another effect: friction from the surface will slow down the nearby air.

Where something is added (heat in the example mentioned earlier), this is a "source," and where something is taken away, it is a "negative source" or "sink." Of course, sources or sinks are not restricted to heat and friction but can be any source for mass, momentum (mass × velocity), and energy or indeed any other calculated/solved-for variable. Further, they can occur anywhere inside the solution domain as well as on the surfaces of the solution domain. The most common additional parameters are moisture content/humidity (increasingly important due to the adoption of adiabatic/evaporative cooling) and contaminant concentrations (e.g., to track the exhaust from a generator in an external flow analysis).

So, a fan *inside* the solution domain speeding up the air would be a momentum source because of the increase in velocity it creates. However, a fan bringing air *into* the solution domain would be both a mass source *and* a momentum source: it is adding air to the solution domain volume *and* bringing it in with velocity, thereby adding momentum. Furthermore, if the simulation accounts for temperature, then a source or sink of temperature (or "enthalpy") will also be present. In CFD terms, these source and sink terms are generally described as "boundary conditions."

Boundary conditions can be defined on a surface or defined over a volume, depending on what they represent. Most CFD programs include an extension to allow what is known as "conjugate heat transfer." This means that solid objects inside the solution domain can exchange heat with the fluid (or other solid objects), while the conduction of heat within the solid objects and the heat distribution in the fluid is calculated.

A key part of the boundary condition is representation of the effect of surface friction on local velocity and any resulting heat transfer. A common method is to assume a logarithmic variation of velocity with distance from the wall. This is known as the "log law of the wall," and it assumes that the flow is turbulent. To represent the surface boundary condition more completely, the logarithmic profile is replaced by a linear relationship in the region immediately adjacent to the wall known as the "laminar sub-layer" where the flow is no longer turbulent.

As well as direct mass and momentum sources representing a specified flow at a boundary condition, a flow boundary condition can be dependent on the conditions outside the solution domain. The conditions outside could represent the following:

- Pressure due to pressurization in an adjacent space
- Pressure and momentum resulting from an external flow such as the wind
- Relative buoyancy due to thermal expansion (or contraction) of the air compared with the external (reference) condition

In some CFD simulations it is quite common to treat the fluid as incompressible, where the expansion or contraction of the fluid does not significantly affect the mass of fluid in a cell. This is common in airflow and heat transfer simulations for the built environment, where the range of temperatures and pressures is usually small enough not to warrant treating the flow as compressible. Even so, the effect of temperature cannot be completely disregarded: the density variations that result do cause warmer air to rise and cooler air to fall. This can be accounted for by adding a force as a source term into each grid cell to reflect the local relative buoyancy. This approach is commonly referred to as the "Boussinesq approximation."

30.2.4 Choosing a Solution Grid

Historically, one of the most time-consuming issues for a CFD modeler has been defining the solution grid or mesh. The choice and level of refinement can substantially impact the predicted airflow and heat transfer, or even the ability for a converged solution to be achieved. Moreover, it also affects the level of resolution that can be studied.

The flow through an array of small obstructions could be predicted using a detailed model describing the geometry of each of the obstructions and the boundary layer around each. This would require a fine grid—a large number of cells in the area of the obstructions—and could therefore be very computationally expensive. However, if all that is of interest in relation to that group of obstructions is the degree to which they obstruct the airflow and the consequent pressure drop (with no interest in the details of the local flow), a simplified model is often employed. This uses empirically derived equations to calculate the pressure drop. It does so based on key characteristics of the obstructions: cylinder size and spacing for a group of pipes or cables, for example. Such a simplified approach no longer requires a grid small enough to resolve the details of the geometry and represents a significant grid saving. "What level of detail do I need to represent to capture the fluid dynamics of interest?" is a key question that a modeler should ask themselves every time they create a model. Adding more detail than necessary will increase grid count potentially making it slow or, in extreme cases, even too big to solve. Excluding important detail will produce a smaller grid but may result in misleading or erroneous predictions.

It is important to note that in most circumstances the only way to be confident that the model has sufficient grid to capture the features of interest that have been included is to undertake a grid sensitivity study. This establishes whether or not the results of interest change as the grid is refined. When refining the grid to use smaller cells does not result in a significant change in the predicted flow and temperatures, the solution is considered to be grid independent.

The simplest form of grid is one where the space is divided up into an array of rectangular/brick-shaped elements. This can easily be visualized in two dimensions as a set of parallel (but not necessarily equally spaced) lines across the page and a similar set running vertically up the page. The result is an array of rectangles within an overall bounding rectangle. The lines are placed closer together where there is need for increased resolution, but the user should be aware that because the lines run all the way through the solution domain, a small gap between lines in one direction with large gaps in another may cause very long thin cells, and these may make it difficult to solve the equations and predict the flow solution. A long and thin cell is referred to as having a "high aspect ratio." A similar problem in solution can also occur if the lines in one direction are very close together and then suddenly they are very far apart. This is known as a "high expansion ratio."

In three dimensions, the process is identical. However, now the lines should be considered as planes and a third set must be drawn in the third orthogonal direction of the 3D axis system. This type of grid is often referred to as a "structured Cartesian grid." It is structured because each grid cell has six faces (one on each side of the box) and next to each cell face is another cell with a coincident face (Fig. 30.3).

There are two key challenges with a structured Cartesian grid:

1. The difficulty of capturing varying degrees of detail throughout the model without creating high aspect ratios (long, thin cells) or high expansion ratios (adjacent cell size changing very quickly).
2. Where objects or surfaces do not align with the grid (as in Fig. 30.4), special treatment is required to avoid surface flows detaching unrealistically. Without special treatment, the surface must be approximated by a line following the cell faces it cuts, and the resulting "stairstep" will interfere with the development of a boundary layer that one might otherwise expect.

In some CFD tools, an "unstructured grid" is used in order to more easily describe complex shapes and more accurately predict surface effects.

This can be achieved in several ways. The simplest way is to retain the Cartesian approach, but to no longer insist on the lines/planes extending all the way through the solution domain, and to allow a cell face to have more than one other cell face next to it. This is often limited to dividing the cell into two in any given direction on a specific cell face. This limit is not essential, but when applied, this gridding approach is called "octri." Given that the grid is no longer structured, the cut cells can be treated by accepting that locally the cells no longer need to be rectangular, and so the shape can be explicitly represented, thus avoiding any stair-step special treatment.

An alternative approach is not to insist on the cell shape being rectangular or box shaped. This allows the grid to follow more complex outlines and permits the cell faces to more closely reflect the true surface. The cells can be distorted brick-shaped elements but are more commonly tetrahedral in shape. Examples of the two approaches are shown in Figure 30.5.

The fundamental disadvantage of unstructured grids is that they require more computer memory and are slower to calculate per cell. This is because if the cells are not a simple rectangular mesh of brick-shaped cells where the next cell is implicitly known, then additional connectivity information has to be stored and processed in order to know which cells are neighbors. Of course, this disadvantage may be outweighed by the fact that an unstructured grid may allow the model to be constructed and calculated with fewer cells by only refining where necessary, but this is not guaranteed. Refinement is arguably more effective when using tetrahedral

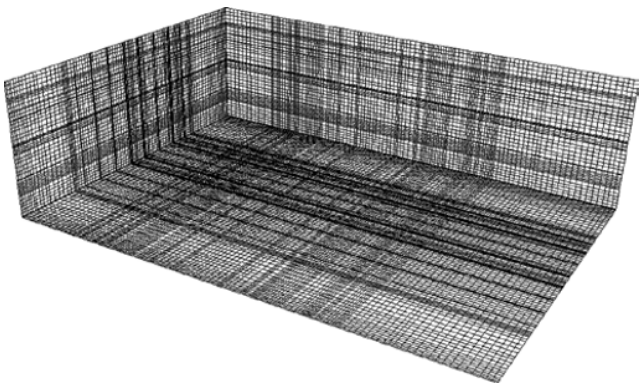
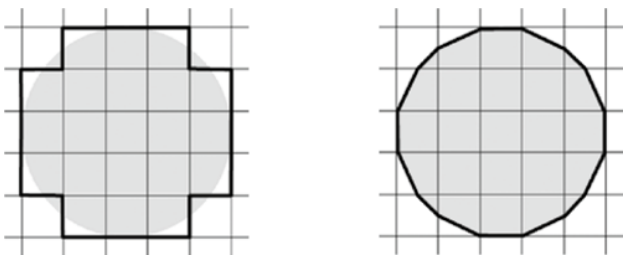


FIGURE 30.3 Structured Cartesian grid.



Simple treatment following structured Cartesian grid

Special treatment dividing grid cells into parts

FIGURE 30.4 Surface approximation in a simple Cartesian grid approach and in a modified approach.

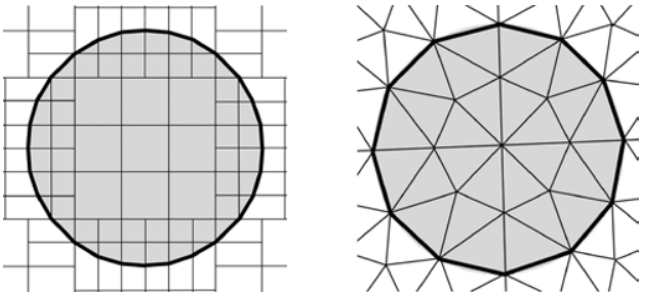


FIGURE 30.5 Shapes can be directly represented to coincide with the unstructured faces of the cells.

or body-fitted approaches, but here gridding has an additional criterion: the cells must not become too distorted with small internal angles.

An advantage of CFD designed for a particular purpose is that the grid rules can be developed to suit the application involved. These rules are based on an awareness of the types of objects that will be in the model and their grid requirement.

30.2.5 Calculating the Solution

As there is no analytic solution for anything but the most simplistic configuration, the Navier–Stokes equations must be solved numerically. The way this is done is essentially a "guess-and-correct" approach, where the solver is provided with an initial guess of the solution (normally zero velocity with a single value for temperature and pressure throughout the solution domain).

The boundary conditions are superimposed on this set of field data and provide sources and sinks of mass, momentum, and energy for the cells they are located in. Without correction, the addition of these sources and sinks locally will create errors in the conservation equations for the cells, and so the values need to be adjusted to account for the changes. These corrections are made on a variable-by-variable basis and on a cell-by-cell basis for each variable.

Of course, the values for a cell depend on the values in all its neighbors. So, because the values are recalculated cell by cell, the first variable will no longer be consistent with the other variables, and, similarly, the first cell recalculated will not be consistent with its neighbors after they have been updated. As a consequence, this recalculation of the values for all the variables in all the cells has to be undertaken many times—as many times as required until the errors in the conservation equations are reduced to an acceptable level. At that point, the solution is deemed to have "converged." Each time all the values for all the cells are recalculated for all the variables is termed an "outer iteration." In practice, some variables (typically temperature and pressure) may be recalculated several times during any given outer iteration. These repeated recalculations for a single variable carried out within a single outer iteration are called "inner iterations."

The fact that these recalculations are made does not guarantee that the values calculated for each field of data will necessarily be closer to the true answer iteration by iteration. Consider riding in a car. If the car had no suspension, the ride would be very uncomfortable because even though the intention (desired solution) for the car is to ride smoothly along, it would deviate from this smooth path bump by bump. By adding springs, the severity of the deviation of the ride position in the car can be reduced by the spring absorbing some of the energy and starting a decaying oscillation. In most circumstances, the oscillation will gradually decay, but if the car hits bumps at a rate that excites the natural frequency of the spring–mass system that the car represents, the oscillation could go out of control. For a car's suspension, the solution is to add a shock absorber or damper to limit the motion on the spring, thus bringing the car back to equilibrium ride height more quickly. Making the damper too heavy results in a hard ride with little oscillation, but once the car deviates from ride height, it takes a long while to move back to equilibrium position. With a very light damper, much of the oscillation is allowed, and the energy from the bump(s) will be only very slowly absorbed, giving long periods of oscillating ride height.

The convergence process for numerical solution in CFD can be considered similar to this analogy. If undamped, the solution may oscillate or even "diverge" (move away from the solution). So, damping is used to stabilize the solution. This is normally done in one of the following two ways:

1. *Linear relaxation*: Here the calculation of the change required for a value is ΔV, then part of the change applied, that is, $f\Delta V$, where f is the linear relaxation factor (between 0.0 and 1.0).
2. *False time step relaxation*: Where the change is calculated as though it were the change that would happen if a small amount of time were to pass.

Such false time step relaxation should not be confused with "time-varying" or "transient simulation," where the CFD is used to predict time-varying flows. For time-varying calculations, the solution is undertaken by using a similar iterative process to solve the equations for the change that will occur over a small period of time. In the same way that the grid resolution is important in space, so is the time step size in time. The time step must be small enough to capture the time-varying features of interest. Generally speaking, the smaller the time step, the fewer the number of outer iterations required to reach convergence.

30.2.6 When Is the Solution Ready for Use?

We have discussed the concept of convergence and a converged solution: the process of reducing the errors in the equations and therefore being closer to the principle of conservation of mass, momentum, and energy throughout the

solution domain. The errors left in the equations are often termed "residual errors."

One way of measuring residual errors is to add up all the errors (imbalances) from all the cells for each variable and to then compare the error sum with some reference value. Commonly, the reference value is the incoming mass, momentum, or energy (as appropriate) for the variable. The solution is deemed converged when the errors for all the variables fall below a small proportion (e.g., 0.5%) of that reference value. The incoming mass, momentum, and energy is commonly estimated before the solution takes place, and so the final performance may not be a true reflection of the true incoming mass, momentum, and energy. Also, *where* the error occurs in the solution domain can be very important. If much of the error is localized in an area of the solution domain that is of no particular interest, then a higher error can be tolerated. If, on the other hand, the error occurs at a point of key importance, then the error may be less acceptable.

Another way of determining whether the solutions are acceptable (at least for non-time-varying calculations) is to monitor the change in key variables for important interest points and see whether they have achieved a steady condition. If an acceptable residual error has been achieved and the points of interest have stable conditions, then it is likely that the solution is sufficiently converged to be considered representative for final analysis.

Once you have established that from a numerical perspective the solution is ready to use, it is important to review the model and results with the awareness that they may not be correct because the model is not a sufficient representation of reality. Put another way, one must review the model with a healthy degree of skepticism: a model is only as good as the input data that defines it. There are two possibilities for it not being correct:

1. The user has made a mistake building the model.
2. There is insufficient detail in the model to capture all the key features.

In the event this is a predictive model and there are no measured results to compare with, it is often good to have an expectation of what the solution will be. Then, when the solution is different, to question why it is different—is there a mistake in input or is there something happening that is reasonable but unexpected? Of course, when the model is of something that exists and is being used for troubleshooting or onward development, it is always best to first model the existing scenario and develop a realistic model before using it for predictive simulation.

30.2.7 What Are the Results?

The solution methodology delivers values for each of the solved-for variables. The basic set of variables includes pressure (derived from continuity), temperature, and velocity components (and thus the resulting overall fluid velocity and direction). The data is available for a grid of points throughout the room, depending on the grid that is chosen (as described earlier).

In a structured Cartesian grid, the data is stored in a set of 3D arrays, one array for each solved-for variable and (in the finite volume method) one value per grid cell. Given the rich data set, it is possible to undertake a wide range of post processing. Data center-specific post processing and metrics are described later; however, almost any CFD tool will provide the following.

30.2.7.1 Result Planes

A "result plane" is a graphical depiction of the values of a calculated variable. It is displayed in the 3D model as a plot (normally in plan or elevation orientation, but not necessarily so), where each grid cell in a selected plane is colored. The color is set according to the value in that grid cell for the selected variable in question (Fig. 30.6).

In the example shown, the plot is of temperature variation in gray scale, with white being hot and dark gray (almost black) being cold. On most computers or printouts, this would normally be in color, typically showing purple or blue as cold and red as hot.

Although it is normal to draw the variation in a smoothed way, interpolating the values between points and creating the impression of continuous variation, most tools also allow you to plot just the calculated value in each cell. This is sometimes helpful as some CFD programs are less intelligent when making the interpolations, especially near solid boundaries and a solid–fluid interface. In such a scenario, a simple interpolation may produce a misleading plot.

A result plane can also be used to plot airflow patterns (Fig. 30.7) (or to plot heat fluxes for conduction, where appropriate) by combining the three orthogonal velocity components (or three fluxes). The magnitude of velocity (or flux) in the plane is normally indicated by the size of the arrow, while the magnitude of the 3D velocity is often represented by the color scale or the gray scale. Color is also often used to represent another variable so that the relationship between flow and other variables can be seen more easily.

30.2.7.2 Streamlines

"Streamlines" are commonly used to understand the convective flow path of air (or conductive path for heat) (Fig. 30.8). They are easy to relate to visually as most people have seen streamers, smoke carried in an airstream, or dye in water. They simply follow the convective path of the fluid from a single point or a set of points.

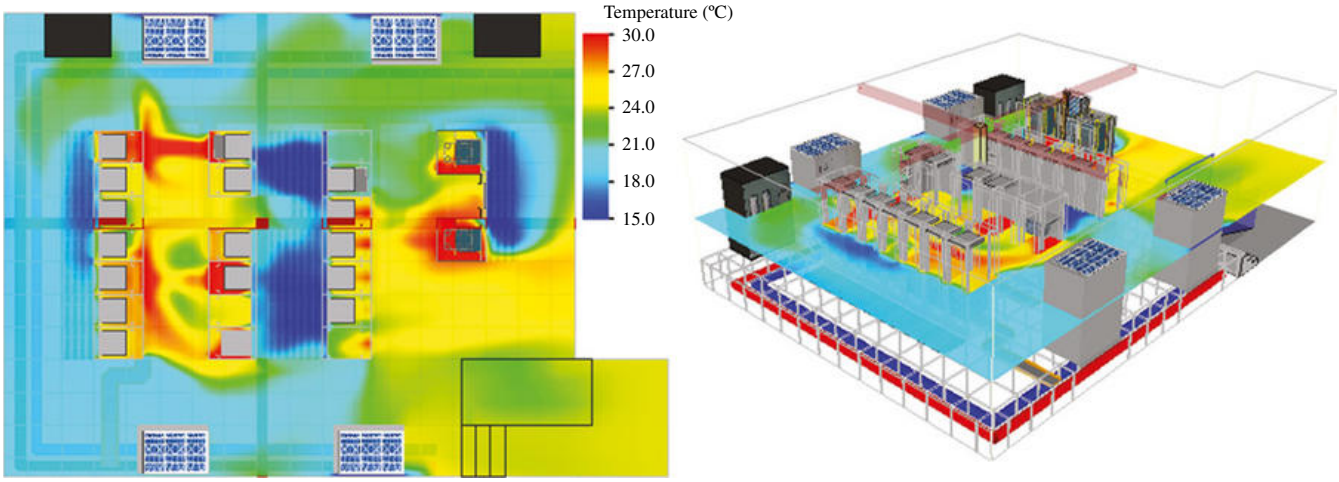


FIGURE 30.6 Result plane of temperature and halfway cabinet.

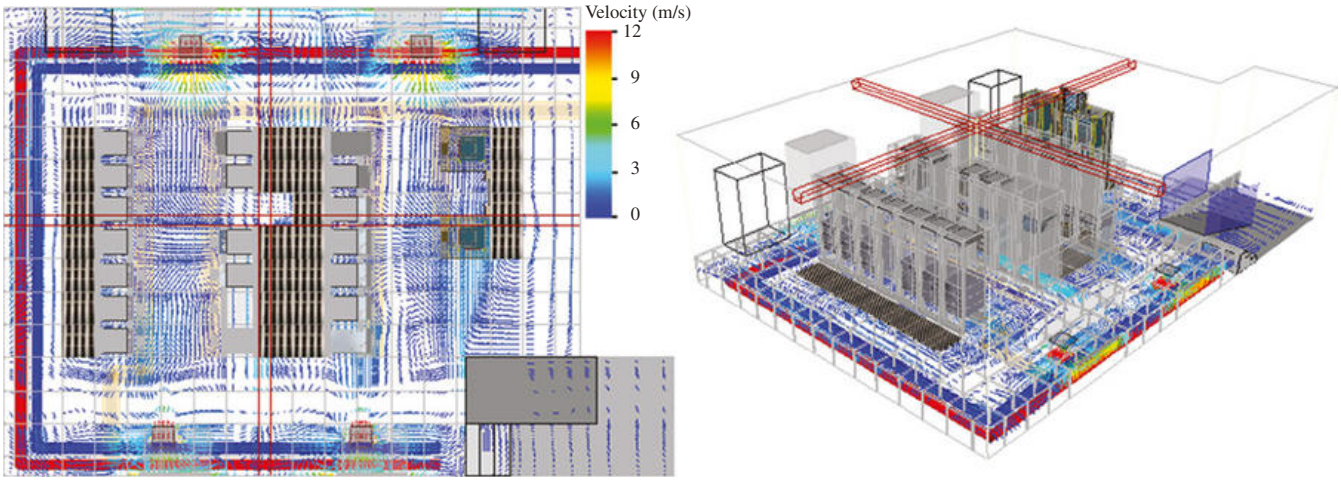


FIGURE 30.7 Result plane of flow pattern in a floor void.

30.2.7.3 Surface Plots

Another way of visualizing results is to graphically depict results on a surface. There are two basic types of surface plot:

1. *Distribution on an object surface*, such as surface temperature or surface pressure. In the first example, the surface temperature is commonly the temperature of a solid at its surface where it meets fluid, often air. This is useful, for example, when modeling electronics. Surface pressure, on the other hand, is normally the pressure in the fluid adjacent to the solid surface. An example of its use might be the pressure distribution on an aircraft or vehicle when optimizing lift (or down-force) and drag.
2. *A surface representing a constant value of a calculated variable*. This is sometimes referred to as an "iso-surface." An example of its use might be to show the volume in which a contaminant/pollutant is at a hazardous level. The surface would be drawn at the critical level with outside the surface—away from the pollutant source—being below the critical level and inside the surface—nearer the source—being above the critical level.

30.2.7.4 Post-processed Data

As described earlier, the large volume of field data lends itself to 2D and 3D visualization in graphical views. However, the data can also be post processed to provide aggregated data for quicker understanding. This is sometimes referred to as "derived data."

For example, the total flow rate through an inlet or outlet is often recorded, perhaps with its average temperature and/or concentration. The flow is summed, and the averages calculated using all the cells/cell faces that coincide with the inlet or outlet in question. The more tailored the CFD program is to an application, the more tailored this derived data can be.

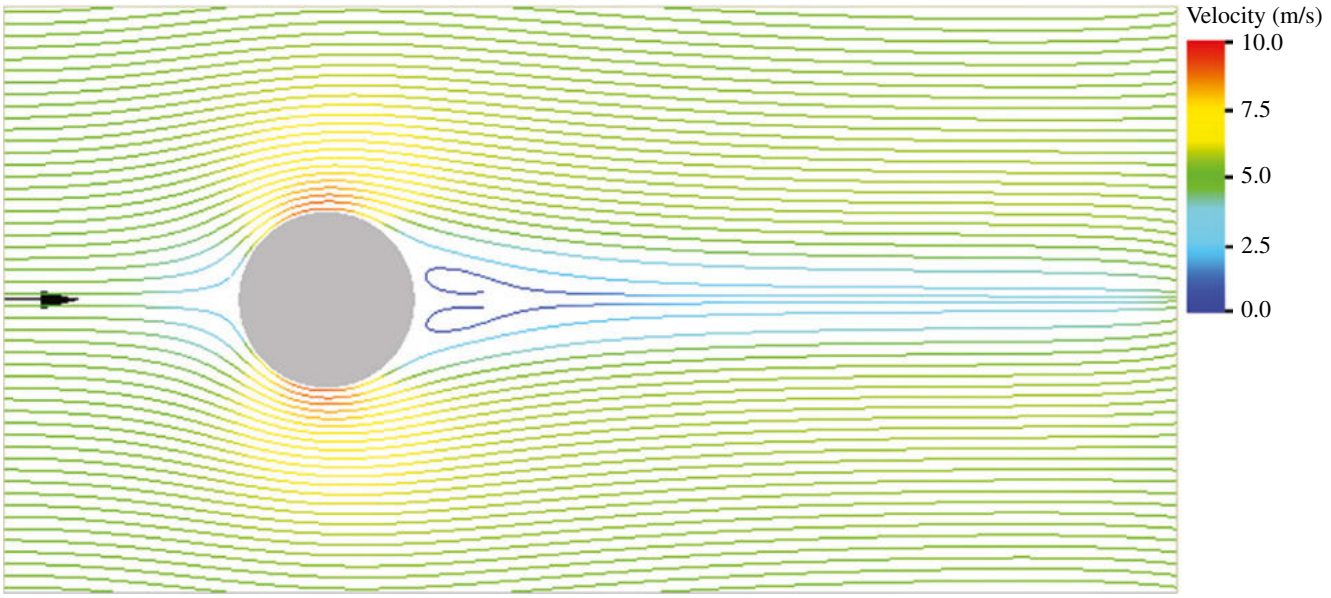


FIGURE 30.8 Streamline flow around a cylinder.

30.3 APPLICATIONS OF CFD FOR DATA CENTERS

30.3.1 Typical Uses

Strictly speaking, CFD is most commonly applied to a data hall (the room where the ITE is housed) rather than a data center (the entire facility, including the support infrastructure outside the IT room). It is sometimes applied to other aspects of the data center, such as the following:

- Airflow around outdoor chillers to determine whether the hot air is exhausted effectively without being re-entrained
- Airflow and cooling of batteries and infrastructural equipment such as uninterruptible power supply (UPS) systems
- Generator halls
- Almost any of the support spaces, whether occupied by equipment or people

Of course, the requirements of CFD vary for each type of space, but especially if the focus is human comfort rather than equipment operating conditions.

This chapter will focus primarily on the data hall but will also address some of the other issues that may occasionally be addressed for a data hall or arise in other data center applications. For the data hall, there are a number of different tasks that may need to be considered. These are as follows:

1. Conceptual design
2. Detailed design
3. Assessment
4. Detailed assessment and troubleshooting
5. Operational management

In addition, it is also common to consider the cabinet scale, configuring ITE in a single cabinet or group of cabinets independently. This means that ITE can be deployed in cabinets with the confidence that the internal configuration is effective and will not undermine room-scale management. Rack or cabinet-scale simulations are considered important because some data center professionals report that equipment cooling problems stem from internal cabinet configuration issues as frequently as they do from room configuration issues.

30.3.2 Use for Data Center Design

Most data center design scenarios do not need to consider the detail of specific equipment configurations. In general, they only consider the ability of the room cooling system to distribute the cool air throughout the facility and scavenge the warm ITE exhaust air effectively. In the author's experience, about half of overheating problems occur as a direct result of equipment configuration. Accordingly, a good design does not guarantee its successful operation, but, subject to due care at the equipment configuration stage, it does increase the likelihood of success. For this reason, it does not matter that in early design the end user cannot normally say exactly what equipment will be installed. The end user only really needs to know, conceptually at least, the different types of equipment that will be installed from a cooling methodology perspective.

The level of detail required in the model will depend on the design decisions being considered. While adding unnecessary detail will not result in poor decisions, it may take more time than necessary for the modeler to create the model and more compute time to solve the model. This will limit the number of design iterations that can be studied in a given space of time, although some tools do have the option to undertake the calculation at a lower level of granularity, even when more detail has been included in the model.

30.3.2.1 Conceptual Design

In conceptual design, CFD can be used to test and optimize the overall design philosophy. Historically, CFD simulations have been made for a full load scenario with a very uniform power and airflow distribution. As a consequence, the conceptual design scenarios often do not exercise the system for the intermediate and nonuniform loads that the system will have to address.

The challenge here is that it is necessary to create a range of scenarios that test the system for a range of realistic conditions when very little is known about what ITE will be installed or when it will be installed. Consequently, it is not appropriate to be too detailed in modeling; only sufficient detail to capture the key characteristics and test sensitivity is required.

If the appropriate decisions are made (Section 30.4), a conceptual design model can be used to:

- Test different cooling strategies, be they conventional CRAC (computer room air conditioning) or CRAH (computer room air handler) systems, in-row or overhead cooling units, cooling units with economizers, or direct fresh air economizer cooling
- Test the number, size, and layout of cooling systems to optimize the cooling distribution, accounting for the room size and shape and other architectural features (such as columns), equipment layout, and power distribution
- Optimize cooling paths, including raised floor height, false/drop ceiling height, and ventilation duct size
- Optimize floor grille, ceiling grille, and duct grille layouts
- Test segregation concepts, such as cold-aisle or hot-aisle containment systems
- Allow for the effects of notional power distribution in the data hall and optimize ITE layout
- Test part load configurations
- Test redundant cooling configurations
- Optimize energy efficiency
- Evaluate the impact of generic representations of cable

It is normal for conceptual design models to make a number of assumptions about the ITE and practices to be deployed in the data hall, including the following:

- Generic cooling system such as CRAC/CRAH units have simplified/idealized cooling distribution and nominal cooling capacity.
- Known cooling set points, such as supply air temperature and airflow rate with limited, if any, control.
- ITE is configured for front-to-back ventilation.
- Cabinets are well configured and do not allow internal recirculation.
- Cable penetrations are well managed and have little impact.

30.3.2.2 Detailed Design

Of course, a model created for a detailed design assessment can be used to make conceptual judgments, but often it is unnecessary to develop the model to this level of detail until conceptual decisions have been completed. That said, adding detail to the model allows these decisions to be confirmed with greater confidence. It also allows the modeler to test additional design assumptions that may undermine or enhance the performance. Additional considerations will include the following:

- Detail of cooling system, such as specific CRAC/CRAH units with particular fan type and consequent airflow pattern
- Detail of control system, including sensor locations and control characteristics, to include condition-sensitive capacity and variable air volumes
- The impact of more realistic equipment choices allowing for the following:
  - Non-front-to-back configurations where appropriate (e.g., in switch cabinets)
  - End-user practices, such as top-of-rack switches and not using lower slots
  - Higher ITE power density so that ITE does not fill the cabinet
- The inclusion of more realistic cabinet and equipment configurations that may affect recirculation, such as the following:
  - Whether cabinets are mounted off the floor
  - Empty slot blanking policies
  - Gaps around the mounting rails to the sides, above the top, or below the bottom

- Details of aisle containment systems, including potential leakage paths and control measures to achieve cooling system–ITE airflow balance
- The effect of realistic cable management practices, including the following:
  - Variation in cable route sizes and densities based on type of ITE deployed
  - Realistic cable penetration sealing performance and other penetrations, such as raised floor holes

Detailed design models are still expected to use typical ITE types, notional cable routes and cabling densities, and uniform damper settings. Consequently, they generally will not contain the diversity of configuration typically seen in an established data hall. In addition to these more detailed models, to better understand and optimize the data hall configuration, similarly detailed simulations may be undertaken at a cabinet scale to understand the implications on internal cabinet airflow of user operational choices (e.g., the use of top-of-rack switches with side-to-side ventilation) and optimize/influence practices, where possible.

30.3.2.3 External Airflow Modeling

CFD for external airflow is controversial but increasingly used. Why is it controversial? Traditional RANS CFD is criticized in modeling external airflows because the turbulence models used for normal indoor applications are not considered the most appropriate for external airflows. Large eddy simulation (LES) approaches are considered appropriate and are essential for such applications as a single-sided ventilation that relies on the unsteadiness for effective ventilation. However, for data center applications, the size of the solution domain and the need for highly resolved grid mean that LES is generally considered impractical. Perhaps more importantly, the uncertainty in the boundary conditions and the variety of scenarios—such as the need to consider multiple scenarios of different wind speeds and directions—means that the approach to external modeling should not be focused on precise specification and prediction, but rather on sensitivity and characteristic. For example, is the warm moist air ejected well into the free stream airflow rather than being re-entrained into the intakes of the indirect evaporative coolers? If such an approach is used, the differences in turbulent diffusion/mixing should not be important, and external flow modeling should be successful. Typical scenarios for which external airflow modeling is being used are chillers in a roof well, generator cooling performance/contaminant dispersion, and indirect evaporative cooling performance/avoidance of humidity and heat recirculation.

30.3.3 Use for Assessment, Troubleshooting, and Upgrade

CFD models created for existing data hall assessment, troubleshooting, and upgrade can be made at two levels of detail, similar to conceptual design and detailed design. The consequences of the simplifications will be similar. As such, if the conceptual design approach is adopted, only high-level issues will be predicted: any changes proposed and modeled will therefore only be appropriate at the same high level.

In order to understand issues at rack level, more detail will be required in the same vein as the detailed design option. Often calibration (Section 30.3.4) of the model will be necessary (and is always desirable) using monitored data for the systems in practice. Examples are as follows:

- Measurement of airflows and temperatures for the CRACs/CRAHs or other cooling systems to determine the actual control response.
- The power distribution as close to IT system level as possible to determine the impact of system use and resulting utilization. This is important since power consumption and consequent heat dissipation are strongly dependent upon the applications deployed on the ITE.

This level of detail and the fact that a model is "calibrated" (Section 30.3.4) should not lead the user to believe that the predictions will be perfect. Although the simulation results from a well-prepared model will provide useful qualitative results and, to a great extent, good quantitative results, a simulation cannot be guaranteed to predict individual equipment inlet temperatures precisely. The simulation may, in some instances, predict a problem in a different cabinet—perhaps an adjacent cabinet to where the problem really occurs. This happens because some airflow phenomena can be very sensitive to small details.

For example, when the jets from two opposing CRAC/CRAH units pass each other, this can cause a recirculation between the two jets. The recirculation is like a small tornado: low pressure at its center, while all around it is at a similar pressure to the typical pressure in the raised floor. If this low pressure is below a floor grille/perforated tile, then there will be little flow upward and potentially even flow downward. Such a flow feature can be disturbed by very small forces and can easily be predicted one tile away from its true placement. If this is the case, a tile that should have virtually no airflow may in reality have airflow of several hundred liters per second. Meanwhile, its neighbor that should have hundreds of liters per second is simulated as having almost none. In such an instance, the predicted flow could be significantly in error. However, the qualitative effect is likely to be predicted accurately, even if slightly in the wrong location.

With this acknowledged, CFD has been used to model many corporate data centers. Indeed, it has been able to identify and provide understanding of problems and consequently enable resolutions to be tested prior to implementation.

30.3.4 Use for Operational Management

Classic operational management using software tools has focused on what are known as DCIM (Data Center Infrastructure Management, pronounced *dee-sim*) tools that claim to provide a toolset for the facility managers to deploy new equipment accounting for space, power, cooling, and network.

The limitation with DCIM tools is that the cooling is based on the *design capacity* of any given IT rack/cabinet and can be an overprediction or an underprediction of actual capacity. The reality, however, is that *capacity varies from ITE rack to ITE rack*.

DCIM tools point to the ability to integrate data from live monitoring systems to see live power consumption and live temperatures, allowing the end user to understand the true environment. But in practice, the problem is that this information tells you what *has happened* rather than what *will happen*. What is required is for DCIM to become predictive DCIM (PDCIM, pronounced *p-dee-sim*). In practice, the only way for this to be done for a data hall is to apply CFD.

Consequently, the use of CFD in operational management is perhaps the most demanding yet rewarding application of CFD to the data hall. At the time of writing, operational management using CFD is only really practiced for some enterprise-scale data halls. This is ironic, since one of the key challenges to the use of CFD operational management is the complexity and time-consuming nature of this simulation: it could be applied much more easily and quickly to the many smaller data halls that exist in almost every city all over the world. However, the rewards for any scale data hall are potentially very significant. The difficulty in deploying a variety of ITE systems in a data hall is that the impact of a change in one area of the hall may be felt somewhere else. In fact, it is a little like a water bed—if you press down in one location, the bed is likely to rise at locations quite some distance away.

Large data halls can have many ITE changes every day. Over time, it is therefore common to arrive at a configuration where placement of a new item of ITE can cause overheating in itself or some other, quite unconnected, item of equipment. Once this happens, the typical enterprise reaction is to protect the mission-critical operations in the data hall by refusing any further deployments. This can, and often does, occur at between 60 and 70% design capacity for the data hall. In a mission-critical facility, one of the key challenges is that once deployed, it is very hard to power down ITE: going back on a problematic deployment is almost impossible.

CFD provides a method by which something unseen—air movement—can be visualized and understood. This has provided an opportunity to address and resolve some of the cooling challenges in facilities where apparently substantial design capacity had seemingly been lost. Further, these data centers are often substantially overcooled in order to address overheating in a handful of items of ITE. In practice, CFD has been used to bring problematic data centers back under control, increasing available capacity to well in excess of 90% of the original design intent and allowing cooling to be undertaken in a much more energy-efficient manner. The cost savings can be tens or even hundreds of thousands of dollars per hall in any given calendar year. Likewise, it can avoid tens of millions of dollars being spent ahead of schedule to build another data hall, not to mention the less tangible, but still very real, reduction in risk and improvement in availability.

An often unrecognized benefit of a calibrated data hall CFD model is that to define the model, it is necessary to include all key items of infrastructure and ITE present within the hall. As a consequence, it is natural (as some software tools have done) to extend the model to provide inventory tracking, space, power, cooling, and network capacity planning, as well as providing the cooling simulation. CFD, therefore, provides an almost unique opportunity for PDCIM to be applied as a matter of course for any chosen data hall. The main disadvantage of existing CFD software programs, when compared with current DCIM programs, is that the majority of current-generation CFD tools provide snapshot analysis that is primarily suited to data hall design rather than to operational management.

The adoption of the predictive capability of CFD along with its inevitable understanding of physical asset distribution and power consumption has resulted in a realization that a CFD-based model provides a more global understanding than cooling alone. To deliver a model that reflects the performance, the model needs to understand what assets are deployed, where they are deployed, and their airflow and thermal characteristics. The most direct way to do this is to understand the power network. With the power network defined, either power data from the live network can be imported to define the heat dissipation, or IT loads can be applied to individual IT assets and the power on power systems calculated as a result. As a result, once calibrated, the CFD model provides an accurate digital replica of the data center—a Digital Twin of configuration that is able to model the behavior of any known existing configuration of the data hall or indeed predict the behavior of a hypothetical future situation to determine future risk and test future plans.

30.3.4.1 Calibration

Calibration is an essential part of data center operational management using CFD. Deployment tools need to be available for everyday use by the facility and IT management teams, so the traditional snapshot use of CFD by fluid

dynamics specialists is not appropriate. In reality, however, the tools can be given user-friendly interfaces that make them attractive to data center professionals, but the complexity of a data hall means that it is easy for a data hall model to provide results that—while they look convincing—may actually be misleading and will certainly not be a true Digital Twin. For this reason, it is essential that the Digital Twin is periodically recalibrated if it is to be used to determine the most appropriate ITE deployment strategies with confidence.

Unlike calibration of an item of test equipment—where calibration of the equipment is the process of comparing the instrument's readings with readings from a more accurate tool, allowing the readings from the test equipment to be corrected to provide true readings in the field—calibration of a Digital Twin is used to provide data to see if the model is still sufficiently representative of reality to be used for operational management. If not, the measurements are used to enable the user to determine which features need updating in the model to make it sufficiently accurate.

Depending on the rate of change in the facility, such calibrations should be carried out periodically. Where changes occur as a routine part of operation, normally no less than quarterly as an absolute minimum and at any time, there is a significant update to the infrastructure configuration or IT layout. Fortunately, the increase in built-in monitoring systems for data halls is resulting in more of the data required being available automatically.

To make a calibration, the following are typically monitored:

- CRAC/CRAH or other data hall cooling system temperature and airflows
- Perforated/slotted floor grille airflow rates and temperatures
- ITE power draws as near to the ITE as possible
- ITE inlet air temperatures: As measured at the IT inlet vents or at the equipment rack inlets depending on whether the model is used for specific ITE configuration or room configuration alone.

Probably the most difficult of these to measure are the airflow parameters. In particular, airflow through modern high-open-area perforated floor tiles is difficult because of the low resistance of the tile itself combined with the large number of floor tiles on a single open floor void. Consequently, the introduction of the measuring device can significantly affect flow measurement. Flow hoods with back pressure compensation (designed for flow measurement from grilles on more typical building ventilation ducts) can help in correcting the measurements, but these can neither correct the measurement if the addition of the hood causes the flow to almost nothing, nor make satisfactory measurements where the flow is particularly turbulent/unsteady. This latter feature is commonly true at perforated tiles near CRAC/CRAH downflow units.

Similarly, making velocity measurements in the outflows from CRAC/CRAH units is difficult for the same reason, and flow measurements for these units are normally most readily achieved by measuring the return airflows.

30.4 MODELING THE DATA CENTER

Like any simulation, the value of a data center model will depend on the quality of the model in question as much as the simulation tool of choice.

At present, the user of the tool is almost entirely responsible for the model, including the representation of proprietary items such as CRACs/CRAHs, power distribution units (PDUs), ITE, etc. Although some CFD tools offer libraries of equipment (commonly referred to as "symbol libraries"), these are limited in scope and often require review and tailoring for their use in the chosen facility model. The modeling decisions made can, therefore, critically affect the outcome of modeling: "garbage in–garbage out." It is therefore important that the user should use traditional approaches to gain confidence in the model, including the following:

- Having an expectation of the result and questioning why the result is different. That is, is there a flaw in the model, or is something genuinely happening that was not anticipated?
- Undertaking sensitivity studies where there are uncertainties. One of the key advantages of simulation is that it can be run for a variety of conditions, and so, where there is uncertainty, parametric variations can be undertaken to test sensitivity.
- For a real facility, where possible, use calibration of the model to ensure it is well specified so that the effect of changes can be expected to reflect reality.
- When making models of items to be included in a Digital Twin, first test them independently (in a separate purpose-built test model) before using them in the model.

There now follows some high-level (and not exhaustive) guidance on data center modeling.

30.4.1 Architecture

In a data hall application, the architecture of the envelope is normally only important from a shape point of view. This is because the internal heat gains are normally so large that fabric heat transfer is at most a second-order effect. It is also

common for data halls to be internal spaces and, as such, to be surrounded by other controlled environments, so the temperature differences are not significant. Internal architecture such as columns and partition walls need to be considered primarily from an airflow perspective where they impact the air distribution associated with the cooling system, particularly—but not exclusively—on the supply side. This section discusses the two classes of architecture identified, focusing on when it is necessary to include additional details.

30.4.1.1 The Data Hall Envelope

Consider a typical enterprise data hall with floor area of 1,000 m^2 (~10,000 ft^2). A typical plan might be 40 m in length and 25 m in width. It is common for the data hall to be quite tall so, including a floor and ceiling plenum, let us consider the wall height to be 5 m.

Given modern building codes, it is likely that the *U*-value (heat transfer coefficient measured by how much heat is passed from the air on one side of the material to the other over an area of 1 m^2 for every 1°C difference in temperature) for the envelope will need to be significantly less than 1 W/m^{2}K. Because this represents a worst case, we will use it as the heat transfer coefficient. If the walls were all exposed to the external environment, then total heat transfer through the vertical façade would be heat transfer coefficient (U) × area (A) × temperature difference (ΔT):

$$\begin{aligned} UA\Delta T &= 1.0 \times \left(2 \times 40 \times 5 + 2 \times 25 \times 5\right) \times \Delta T \\ &= 650 \times \Delta T \text{ W} \end{aligned}$$

This represents 650 W of heat transfer per degree of difference.

If the data hall is an internal building space, then, given a temperature difference of only a few degrees between inside the data hall and its surroundings, the heat gain is unlikely to be at all significant.

If the data hall is external and placed in a hot climate with an extreme temperature difference of 20 K,[1] the heat transfer through the walls would be 13,000 W. This is likely to be around 1% of the design internal full load heat gain.

The roof can be considered similarly. With an area of 1,000 m^2, the heat transfer would be of the same order, giving a conducted heat gain of a few percent of the design internal full heat load.

The floor is likely to be relatively neutral in normal operation for a modern, low-energy data hall, since data halls are commonly on the ground floor and increasing supply air temperatures are likely to be similar to the ground temperature. This should be considered when considering cooling margins, but is probably not important to airflow and heat transfer calculations in the data hall.

[1] Temperature differences in SI (metric) units are normally quoted in Kelvin (K). In practice, this has no impact since 1°K = 1°C.

Now let's consider a situation where the façade is exposed to solar radiation. We will again consider a scenario that will be close to worst case in order to understand the magnitude of the solar radiation effect. For a rectangular data hall, only two of the walls and the roof can be exposed to direct shortwave solar radiation at any given time. Assuming that:

- The angle of incidence onto all surfaces was 45°, reducing the incident radiation per m^2 by a factor of 0.707.
- If the sun were immediately above the data hall, then the area subjected to shortwave radiation would be reduced from 1,325 to 1,000 m^2. If the solar intensity were 1,000 W/m^2, then the heat gain falling on the building could be of the order of 1 MW. If all the heat were absorbed by the surface and given the *U*-value of the surface of 1 W/m^2-K, then it would be reasonable to assume that the resistance of the wall plus internal surface resistance is the order of 10 times the resistance resulting from the external surface heat transfer coefficient. As a consequence, it would be possible for 10% of the solar radiation to occur in the data hall. The heat gain from solar radiation could be of the order of 10% of the internal heat gain. This clearly cannot be ignored from a capacity standpoint at least. If the data hall has a substantial area of glazing or is an older building with lower thermal performance, then the thermal performance of the façade must always be considered from a capacity and internal temperature perspective.

There is one more circumstance where building surfaces start to play a thermal role. This is when a complete cooling failure scenario is considered. With a surface heat transfer coefficient of the order 10 W/m^2, a surface of 1,000 m^2 could absorb 10 kW per degree of difference. As the temperature rises in the data hall, while this will be insufficient to offset the heat load, it can contribute toward a slowing in the rate of temperature rise.

30.4.1.2 Internal Architecture

Temperature differences across elements of internal architecture are not normally of key importance. However, the presence of internal architectural elements can significantly affect airflow.

For internal partitions, be they to provide independent airflow plenums or to segregate areas of the data hall for other reasons, their segregation efficiency is of great importance. As a consequence, the user should pay particular attention to any leakage paths. For conceptual design, the designer may assume that the segregation is perfect, but for detailed design or assessment—and operational management in particular—care must be taken to ensure a good understanding of any leakage paths.

Over recent years, considerable attention has been focused on the segregation efficiency of the raised floor. The piecewise construction itself does allow leakage between the floor tiles, but in a typical data center, this leakage is small—of the order of 1%. However, in legacy data centers, or indeed any data center where little or no attention is paid to the management of floor penetrations, open cable penetrations and other poor tile cuts (e.g., around cooling and power infrastructure) can result in unmanaged leakage of the order of 50% of the cooling airflow. From a modeling perspective, it is critical that the CFD tool can account for this leakage not only in magnitude but also in location: the location may control whether or not the cooling air is useful.

The introduction of segregation as a key element of energy-efficient cooling, such as aisle containment, has made this aspect even more critical. Incorrect assumptions about the leakage can completely undermine the theoretical performance of the cooling system by allowing unexpected recirculation (potentially hotter and more dangerous than without containment in place) and can also result in flow control challenges that otherwise might not be experienced.

Detailed design and assessment models and operational management models must also be able to capture the behavior of flow devices such as perforated floor tiles. It is important to recognize that one tile of 40% open area will not necessarily perform in the same way as another tile of 40% open area. This is because the air volume passing through the tile will depend not only on the resistance to airflow and the pressure difference but also on the geometry and its effect on turning the air through the tile and on the tile's outlet flow pattern/velocity distribution. The ability of a CFD program to characterize key airflow devices is also, therefore, critical not only to the prediction of airflow in the room but also to the fundamental prediction of airflow distribution across the array of perforated tiles.

Other internal obstructions should be considered on the basis of their position relative to key airflows. For example, columns are often important because they interfere with the underfloor airflow and may result in an inability to have regular perforated floor tile and equipment distributions. Beams, on the other hand, are often unimportant because they are typically at high level where there is little air movement. An exception to this is where the return air systems are at high level and the beams create segregation that prevents or aids the hot air in its return to the cooling system.

30.4.2 CRAC/CRAH Units and Cooling Infrastructure

The cooling distribution system is particularly critical in a CFD model of a data hall. For most data hall modeling, however, the CRAC or CRAH units—be they conventional downflow units or more recent alternatives such as in-row, in-cabinet, or overhead cooling systems—are represented as a black box. The assumption is made that any temperature variation on the return, and any cooling that is applied, results in a homogeneous fully mixed airflow at the CRAC/CRAH outlets/supplies. For most scenarios, this is a perfectly adequate representation, as the airflow leaving CRAC/CRAH units is usually highly turbulent, and so any nonuniformity at the outlets normally mixes out very quickly. Of greater importance is the representation of cooling capacity and the controls for air volume, cooling, and any humidification/dehumidification, if applicable.

For conceptual design or concept assessment, all that is generally required is the ability for the black box cooling unit to deliver the design airflow with the cooling to match the heat load in the space. Most data center CFD tools provide simple controls to govern the cooling applied, based on either a user-specified set point for the average supply air temperature or the average return air temperature. Given the lack of knowledge of the heat load distribution for these types of simulation, such a model is a fit-for-purpose approach that can distinguish between basic assumption changes, such as return air temperature control versus supply air temperature control.

For more detailed analysis, especially for detailed assessment or operational management, it is important to a use a more representative model of the chosen cooling unit and any associated controls. The following are critical to the effective prediction of the cooling performance:

1. A representative location for the cooling sensor(s): Temperature distribution can vary dramatically over a small distance. For example, it is not unusual for the air temperature at the return to a downflow unit to vary by 5°C across the entire return area (Fig. 30.9).
2. Careful representation of the outlet/supply airflow distribution: Particularly since the introduction of radial blowers can provide a very different air distribution (depending on configuration) compared with more conventional centrifugal blowers (Fig. 30.10).
3. The ability to vary airflow and cooling in response to controls based on feedback from temperature, pressure, or flow sensors, as well as accounting for the physical capabilities of the system (e.g., the fan curve).
4. The ability for cooling capacity to vary based on on-coil conditions: During failure scenarios as the return air temperature rises and, therefore, the air temperature reaching the coil rises. This increased air temperature and the consequent rise in coolant temperature often results in increased capacity due to higher temperature differences at the coil and/or at the external heat exchanger.
5. Flow distribution from local cooling resources such as in-row coolers: The pattern and velocity distribution of the supply air can dramatically alter how much room air mixes with the cool air entering the cold aisle.

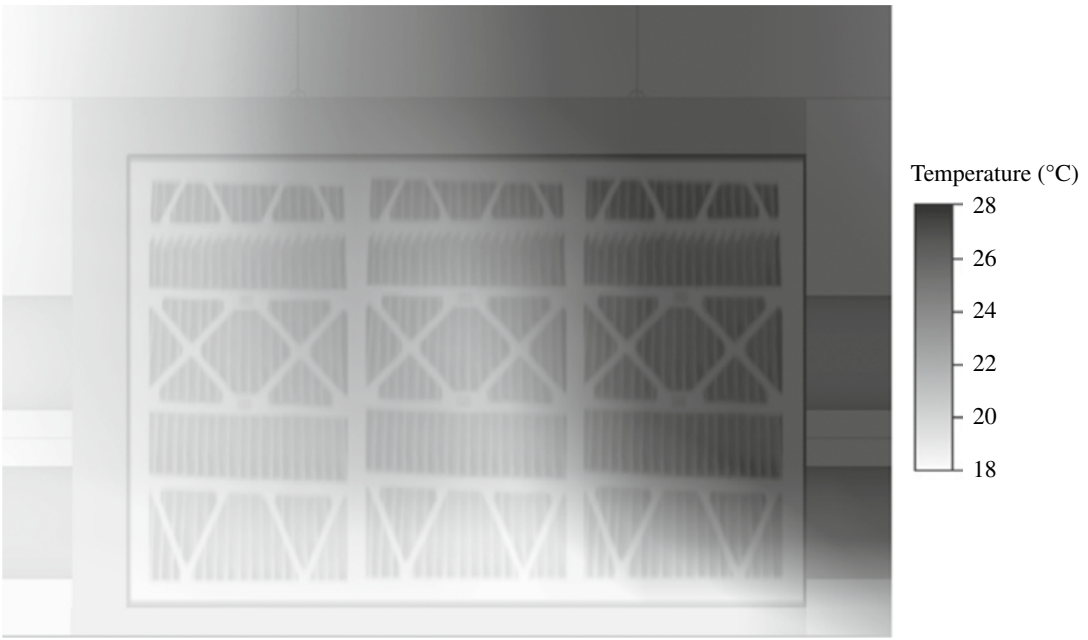


FIGURE 30.9 Typical return air temperature variation at a downflow unit return.

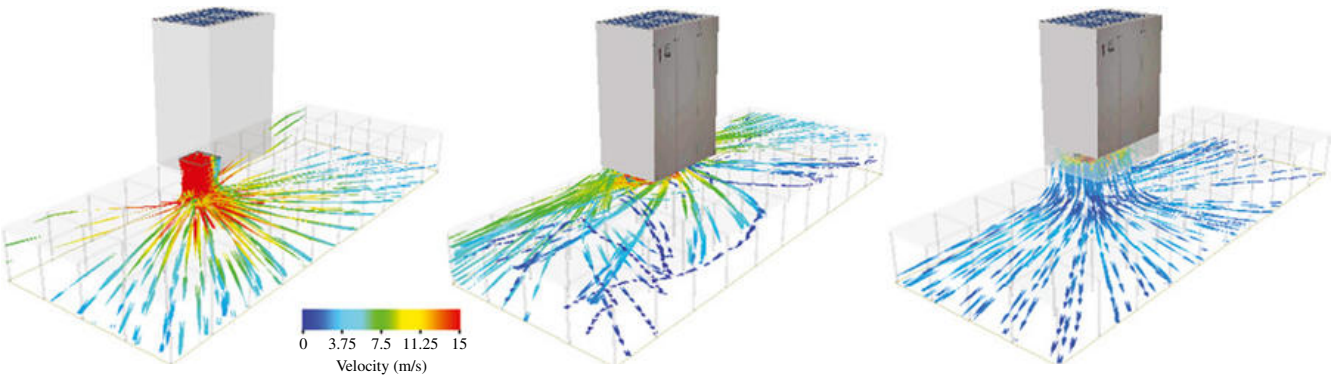


FIGURE 30.10 Typical flow patterns—centrifugal blowers (left) and radial blowers (center and right).

30.4.3 Other Infrastructure

The presence of other infrastructure in the data hall can critically affect the cooling performance, but this is easily overlooked. This can be important because the infrastructure uses cooling resources that would otherwise be intended for cooling the ITE. Alternatively, the infrastructure may be important because of its geometry and its impact on the path of the cooling airflow.

30.4.3.1 Use of Cooling Resources

Some infrastructure that requires cooling will normally take it from the air intended to cool the ITE. This is particularly so when UPS, PDU and transformer systems are placed inside the data hall. Likewise, other heat sources such as lighting should also be considered, even if only from the point of view of their impact on reducing the cooling capacity available for IT.

For example, in-room power distribution units can be cooled directly from the raised floor, with some units having pressure-driven flow through openings in the bottom to vents in the top for cooling internal components such as transformers. Other PDUs may have built-in cooling systems that draw air from the room and return it to the room.

It is also worth noting that these infrastructure items, because of their weight, often stand on frames that are mounted on the floor slab rather than being mounted on the raised floor itself. If insufficient care is taken during installation, unintended and uncontrolled leakage can occur through the hole in the raised floor under and around the infrastructure item. To make detailed assessments and undertake operational management using CFD, it is important to inspect and account for these systems.

30.4.3.2 Obstructions to Airflow

In addition to the items of infrastructural equipment that are typically placed above the raised floor, the common items to allow for are cooling pipes, power cables, and data cables. These were historically installed in the raised floor and can significantly impact the air distribution throughout the room. Problems with airflow often resulted from poor cable management and the raised floor becoming blocked.

The natural response to these issues has tended to be an overreaction—entirely removing almost all of these obstructions from the raised floor (at least the power and data cable elements) to leave the raised floor available as an air distribution plenum. CFD analysis can, however, be used to assess whether this is indeed the appropriate thing to do; experience suggests that some degree of distributed obstruction is helpful in achieving a more uniform airflow through raised floor perforated tiles. This is because they help to break up the highly directional airflows from the cooling units, thus generating a more uniform static pressure throughout the raised floor plenum.

For conceptual design, it is normally straightforward to account for infrastructural obstructions to airflow. However, for more detailed analysis and detailed assessment, troubleshooting and operational management in particular, more care is required in order to capture the true influence.

Structured pipe and cable routes are normally laid out during design, with only the main routes being critical in a conceptual design model.

Cooling Pipes

The main cooling pipes are often significant enough in size to be included individually. Individual branches to each cooling unit are often ignored for conceptual design and may also be ignored even in more detailed analysis. Whether they are included will depend on their size and, importantly, their number and location. The decision whether to include smaller pipes can usually be made based on whether they, together with any other objects close by, represent any significant obstruction to airflow.

Small-percentage obstructions of the air path by one-off circular cross-sectional pipes are unlikely to significantly affect air distribution. The placement of the main cooling pipes in the path of the cooling jet from the cooling unit is likely to significantly affect the path and penetration of the jet.

Power Cables

Power distribution can be made in several forms, from individual cables to significant sized power conduits. Power conduits can be treated in a similar manner to cooling pipes. However, care must be taken to account for conduit junction boxes. These can be significant in size and, when placed in a raised floor, may block half the height of the raised floor and significantly disturb the local airflows.

Power cables, when laid in bundles or groups on the raised floor to be distributed around the data hall, are normally satisfactorily represented by solid obstructions capturing the height, width, length, and depth of the bundles. Often, the most difficult power cables to model are groups of cables descending vertically from the PDU or UPS to the solid floor. They are difficult to model because they are often in a loose bundle where air can pass between the cables. Visual inspection of the bundle will give the appearance of the region being more heavily blocked than it really is because of the visual obscuration. Consider the example shown later.

In one direction, cables are separated from each other by 150% of the cable diameter (Fig. 30.11 *left*). From the perpendicular direction, the cables are separated by the same distance as the cable diameter. Taking any row in either direction, the open area is 50% or more. Inspecting the bundle from the side from some directions would show the gaps to be blocked by the next row of cables (Fig. 30.11 *center*), while from others the bundles may appear to have a small gap (Fig. 30.11 *left*). In actual fact, a calculation will show that the volume is actually only around 32% blocked (Fig. 30.11 *right*). So, the judgments made using visual inspection should be treated with care!

If your CFD tool of choice does not have a loose cable bundle object, then use its porous obstruction modeling

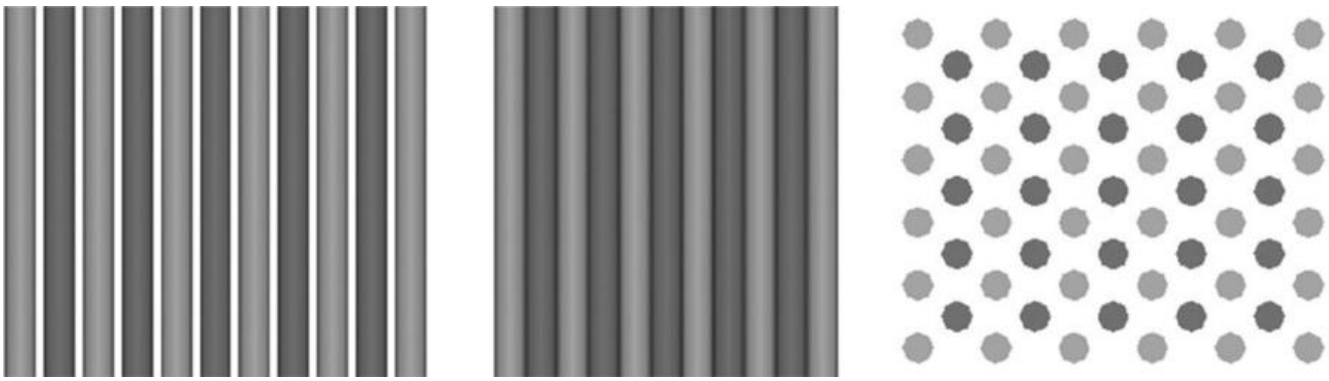


FIGURE 30.11 Staggered array of cylindrical obstructions.

object to represent the partially obstructed volume. Coefficients for general purpose volume obstruction of this nature can be found in [1].

Data Cables

For conceptual modeling, as the number and size of data cables will not be known in detail, it is normal to include a simplified obstruction represented as a solid: a bundle of cable tied together. For structured cabling, this is a reasonable representation since, unlike power cables, data cables are typically much smaller and can be bundled together and be changed in direction on a reasonable radius of curvature while still retaining the bundle properties. In some instances, it is necessary to also include the cable tray that follows the cable route and supports the cable bundles. This is particularly so when the cable tray has solid faces, either as a result of a sheet metal construction or because the wire mesh construction is supplemented by solid sheet of a material (like Correx, a corrugated polypropylene sheet) to prevent damage that may otherwise occur when cables are supported on a large open mesh wire basket tray.

For more realistic modeling in assessment, troubleshooting, or operational management, it is normal to capture a more realistic distribution of cable densities based on what is actually installed. Like conceptual design, cable route modeling is relatively straightforward when the cabling is structured. However, where the cabling is not structured (Fig. 30.12) (particularly so for legacy data centers), then a similar challenge occurs in determining the true degree of obstruction when the cables are haphazardly deployed.

For data-hall-specific modeling tools, particularly when they are being used for operational management, including the cable routes can also have the added benefit that the tool may be able to automatically route cables when new equipment is added to the models. This itself has a number of additional benefits, including better representation of the cable obstructions and a calculation of the cable lengths required.

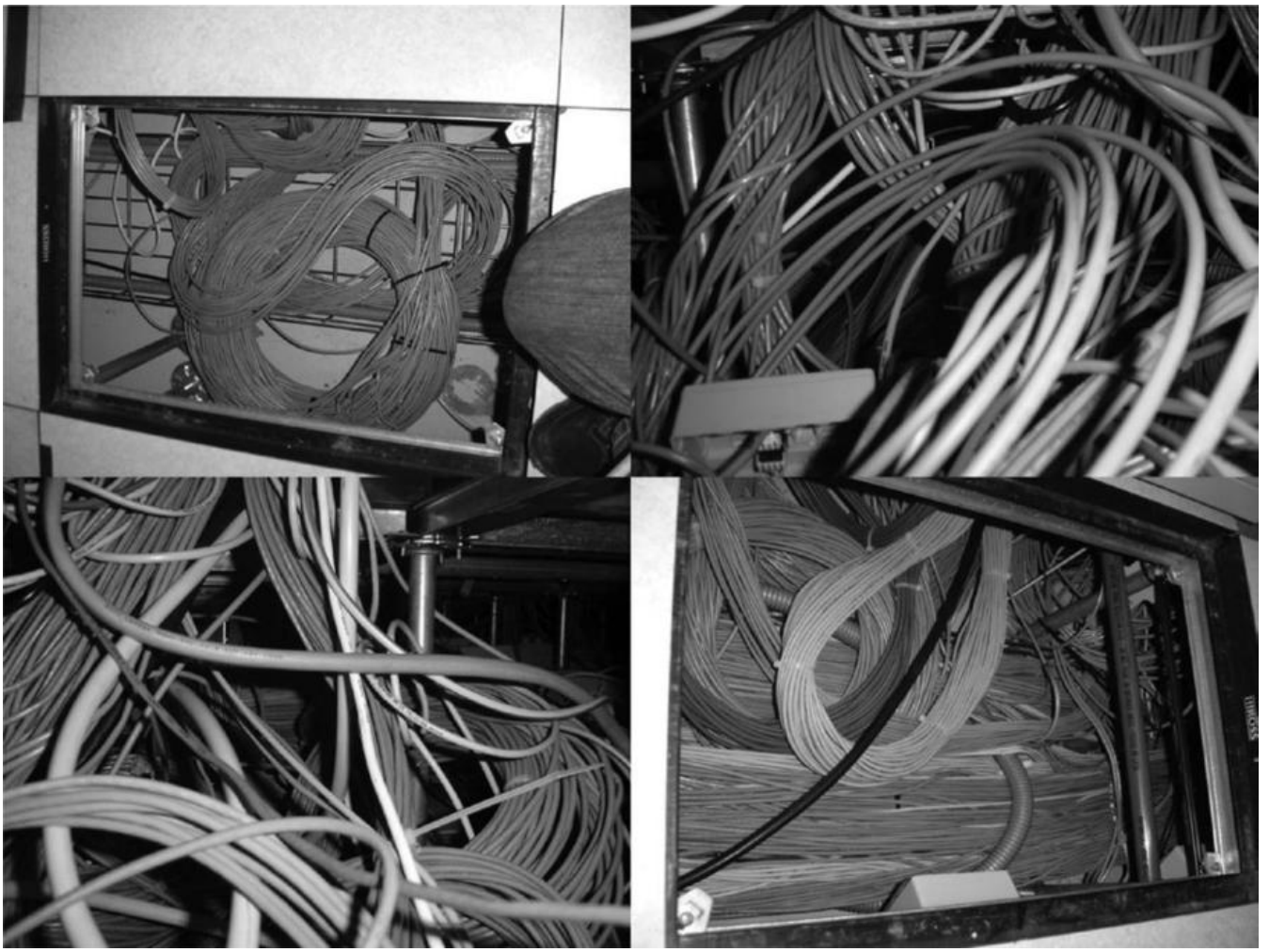


FIGURE 30.12 Classic unstructured cabling. Source: Courtesy of Future Facilities.

30.4.4 White Space ITE

Clearly of great importance, modeling the white space is critical to the purpose of CFD for data centers. This area of modeling poses the greatest dilemma to the data center CFD modeler. It is known that the characteristics of the ITE installed can change not only the local flow and heat transfer but also the overall cooling performance within the data hall. Yet at the design stage for a data center, it is nearly always the case that little, or nothing, will be known about the actual ITE that will be installed. During conceptual design, it is therefore common to assume the following:

- The equipment ventilates from front to back.
- The cabinets are filled with ITE.
- The entire data hall is populated with cabinets in which equipment is installed to the average design limit of heat load per cabinet.
- A nominal airflow per kW of IT load typically of the order of 120 cfm/kW (56.67 l/s/kW) of IT load.
- There is no opportunity for warm air recirculation in or under the cabinet.

Given the lack of certainty about the ITE to be installed, it is reasonable to make such gross assumptions provided that the user of CFD for data center conceptual design recognizes that the cooling design may be sensitive to a variety of design parameters. A more comprehensive and reliable design study can be undertaken if the sensitivity studies are made for the following:

1. Airflow rate per kW: Typical values range from about 80 cfm/kW (modern high-density equipment operating at high utilization) to 160 cfm/kW (low-power legacy equipment or modern equipment operating at low utilization). This is important because it can test the ability of the cooling system to match IT demand in terms of airflow (m^3/s) as well as power (kW).
2. How dense the equipment is in the cabinet: Is the design sensitive to whether the cabinet is loaded over its entire height or whether it is half filled at the bottom or the top of the cabinet?
3. Variation of heat load per cabinet in different zones of the data center to represent some high-density zones and some low-density areas (but with the same overall average cabinet power density). This tests the ability of the cooling to adapt to more realistic nonuniform equipment distributions.
4. Partially loaded scenarios to understand how the design will cope when it is only part occupied in the early years.
5. Cooling system failure scenarios to investigate whether the redundancy strategy employed is effective.

A classic modeling error for ITE is to assume that the airflow exhausted from the ITE is distributed across the full face of the cabinet or at least the full area of the perforated door. This often results in unrealistically low momentum from the ITE, and so the warm exhaust air tends to rise more rapidly and travel less distance horizontally than it will in reality.

30.4.4.1 ITE Detail

When modeling real installations for detailed assessment or operational management, more attention must be paid to the specific ITE. One of the key challenges is that the ITE will have variable heat output depending on its utilization. It is also likely to have variable airflow that can be dependent not only on the utilization but also on the operating environment and its consequent impact on the temperature of key electronic components. Further, knowing the manufacturer and model of many items of ITE is an insufficient specification since many platforms can be configured not only in terms of memory installed but also in terms of the number and type of processors, i/o cards, etc.

For most deployment decisions, what the IT manager or facility manager is most interested in are the maximum demands that will be placed on the facility infrastructure. As a consequence, most of the readily available data for ITE power and airflow is based upon these maximum requirements that, in normal operation, are never realized. Hence, for detailed design, assessment, and operational management, it is important that the ITE model behaves appropriately for the conditions that are likely to be present. Consequently, some enterprise organizations will bench test their standard configurations of hardware running their chosen applications to better understand their operational characteristics.

The adoption of measures to make the data center more energy efficient is making the need for such bench test data even more important. For example, the rise in ambient operating temperature makes it more likely that temperature thresholds that result in IT airflow changes will be reached. Also, the adoption of aisle containment requires greater attention to the airflow balance between the ITE and cooling system.

Ideally, any ITE will be defined by the following:

1. The physical geometry of the ITE
2. The location and size of inlets and outlets
3. The power consumption of ITE and any dependence upon configuration and utilization
4. The airflow and its dependence upon pressure and temperature
5. Distribution of airflow and heat dissipation into separated flow paths through the ITE

Without this data, what can practically be done to achieve a useful model of the ITE?

- In the worst case, where design/modeling must be done based entirely on available published data:
  - Use nominal operating power consumption, if provided. Otherwise, use the name plate power (the quoted maximum power consumption) provided for electrical safety factored by a multiplier to reflect the fact that this power is not normally achieved in normal operation. For legacy equipment, this factor can be as low as 25%, but typically modelers use a factor of 50%, which reflects increasing utilization and the current state of technology.
  - Use the published nominal flow rate and, where no data is published, assume a flow per kW. Typically, modeling often assumes a value of about 56.67 l/s/kW (120 cfm/kW), but values can vary dramatically. Modern high-density equipment can have flow rates as low as 35 l/s/kW (and the current trend is toward these lower flow rates), while older/low utilization equipment may have flow rates as high as 75 l/s/kW. Indeed, this latter figure can be considerably higher if the equipment is low power.
- Where there is access to an operating data hall:
  - Use measured data for power to improve the estimate of power consumption at the ITE. In almost any data hall, the power consumption is known at PDU level. With the advent of intelligent power strips, power consumption is now often monitored down to the cabinet power strip level and even the individual socket level. In any case, the data available should be used at its most refined level to adjust the estimated power dissipation to match the measured data. So, if a cabinet has an estimated power consumption of 3.2 kW but the measured power consumption is only 2.4 kW, then the power for each item of ITE in the cabinet should have power reduced to ¾ (2.4/3.2) of the estimated value. This has the added advantage of enabling a much more realistic power distribution in the Digital Twin.
  - Take sample measurements of inlet and outlet temperatures on selected items of ITE and, given the power consumed, use the temperature difference to check on the airflow. So, for example, a server using 400 W of power with a measured temperature difference of 12°C has a flow rate of 27.9 l/s based on the following formula:

$$Q = \dot{m}\, Cp \Delta T$$

where
Q = the power in Watts,
$\dot{m}$ = the mass flow rate in kg/s (which in turn can be converted to volume flow rate by dividing by the air density),
Cp = the specific heat capacity of air in J/kg K,
ΔT = the temperature rise in °C (or K).

To get the flow rate per kW (1,000 W), simply multiply the flow rate by 1,000/400: in this case, 69.7 l/s/kW.

30.4.4.2 Cabinet Detail

As has been said, for conceptual design, it is normal for modelers to assume that the cabinet prevents recirculation from the hot exhaust side of the equipment to the cooler inlet side (and that it does not allow cool air bypass). In this scenario, many simulation tools assume that certain vents on the cabinet are inflows and others are outflows, even though, in practice, a perforated vent can easily allow both in different locations.

Further, some argue that this is a sufficient assumption for assessment and operational management, given that one is primarily interested in providing a good operational environment for the ITE. However, in the author's opinion, this is not the case:

- First, the primary interest is not the room conditions but the local conditions for each and every piece of ITE.
- Second, if recirculation or bypass occurs in the cabinet (Fig. 30.13), then the air volume passing through

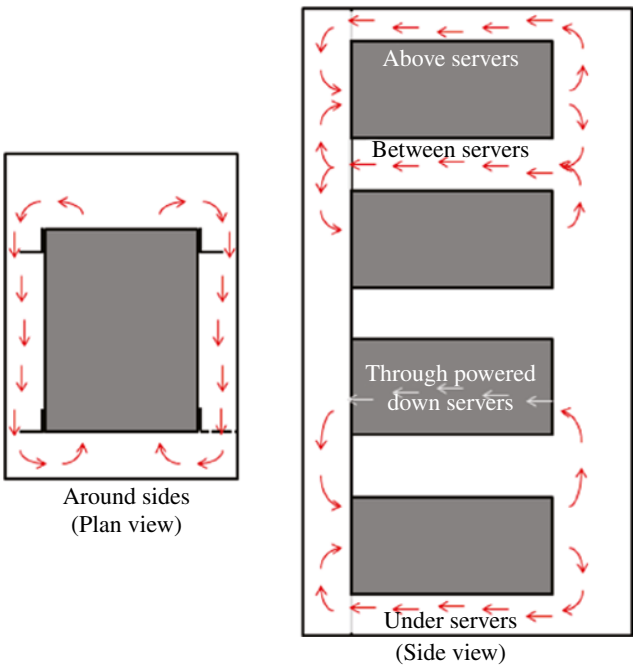


FIGURE 30.13 Schematic of internal cabinet recirculation.

the cabinet and the temperature rise across the cabinet from inlet to outlet will not be the same as for the ITE itself.

- Finally, the inlet conditions at the ITE inlets are likely to be different from the surrounding room conditions.

Let us consider an example where 20% of the air flowing through the ITE recirculates inside the cabinet. As a consequence, only 80% of the airflow demand from the ITE is taken from the room. This means that the average inlet temperature will be increased by 25% of the ITE temperature rise due to the recirculation air supplying a quarter as much air as there is cooling air. It also means that the exhausted air volume from the cabinet will be only 80% of the ITE flow rate, but at 25% greater temperature rise from the room inlet condition than the temperature rise in the ITE itself.

It is therefore critical for detailed assessment, troubleshooting, and operational management to model the potential for recirculation and bypass inside the cabinet. The most common locations for this to occur are as follows:

- Through empty unblanked slots in the mounting rails
- Around the sides of the mounting rails
- Under the cabinet and/or equipment in the bottom u-slot
- Over equipment in the top u-slot
- Between ITE that do not fully occupy complete u-slots
- Through ITE that is installed but that is switched off

One cabinet detail that is often overlooked is the presence or absence of cabinet sides separating adjacent cabinets from each other. This is particularly important if the cabinets have large open cable penetrations (Section 30.4.4.3) in the raised floor beneath them, a characteristic that normally only occurs in legacy installations. In such circumstances, high-velocity jets can generate a recirculation in the cabinets and the raised floor, drawing warm air out of the bottom of some cabinets before the mixed air reenters cabinets further down the row.

30.4.4.3 Cable Penetrations

Cable penetrations in the cabinet itself are not normally important when compared with the vents intended for ventilation in the cabinet sides and top, although there are some exceptions: for example, cabinets with little ventilation (such as those with glass doors), but these are no longer normally used to house ITE with significant power.

However, the location, size, and management of cable penetrations in the raised floor beneath the equipment can (as mentioned in Section 30.4.1.2) significantly affect the room cooling system performance. They can allow as much as half the cooling to leak unintentionally into areas of the facility that do not require cooling—the back of cabinets and the hot aisle, for example. It is now generally accepted that such cable penetrations should be managed by limiting the airflow with the introduction of foam, brushes, or gaskets. However, claims for the efficiency of such seals are often not achieved. This is because a cable turning from a horizontal distribution path in the raised floor to the vertical distribution entering the cabinet tends to disrupt the seal. It is therefore important to include this leakage when creating a CFD model for assessment, troubleshooting, or operational management.

CFD modeling can be particularly important in relation to cable penetrations when the intention is to upgrade a facility by introducing cable penetration management in a legacy data center where the cable penetrations previously went unmanaged. While the rule of thumb is that introducing management generally improves performance, unilateral introduction without analysis of the impact can be dangerous: in some cabinets, ITE may be relying on cooling arriving through the cable penetrations.

It should be noted that a CFD tool that does not offer the user any control over where cable penetrations are placed is of little value in this type of analysis. This is because a cable penetration discharging air into the inlet side of a cabinet will have a very different impact compared with a cable penetration discharging air into hot exhaust side of the cabinet, where it will probably offer no cooling value to the ITE at all. Similarly, care should be taken with tools that do not enable such discrimination to ensure that they do not overpredict cooling by assuming that air entering through cable penetrations is providing cooling when it is not.

30.4.5 Modeling Control Systems

Controls have already been discussed to some extent in relation to cooling systems (Section 30.4.2) and ITE (Section 30.4.4), and it is indeed these two areas where control is most important. The primary challenge with CFD models is the replication of the configuration and behavior of the real control system.

For conceptual design studies, most CFD tools have simple control models that allow cooling, or airflow through cooling systems, to be controlled based on the average return air temperature or the average supply air temperature. This is normally sufficient for conceptual studies, where details of the cooling systems, control systems, and ITE are unknown.

However, as the design proceeds into the detailed design phase, or where the model represents a real facility and is to be used for studies of several different configurations or conditions, the response to the control system becomes more

important. The following list (not exhaustive) represents features that a modeler might look for to enable better representation of the facility behavior in the model:

1. Point sensors, in addition to average sensors, that can be placed at a user-specified location.
2. The ability for controller response to be based on multiple sensor values and to select functions such as average, minimum, maximum, difference, or another user-specified function of the values.
3. The ability for multiple items to be controlled by a single controller.
4. For a controller to control based on sensor values for one variable (e.g., temperature), but the controller output or object response to be limited by another variable (e.g., pressure).
5. For the object response to be conditional on the conditions in which it operates. For example, when the controller output is demanding full cooling from a cooling unit, the maximum cooling should be dependent on the air temperature onto the coil, the coil coolant parameters, and the air volume.
6. The ability to control items other than cooling systems, including ITE flow, fans, dampers, etc.
7. The ability to control multiple variables including temperature, airflow, pressure, and relative humidity/moisture content.
8. The ability to sense any of the variables and use them for control.
9. Linear controller response and more general controller response.

Although this list represents features that are already present in a variety of data hall systems, it is unlikely that any CFD tool currently provides all these features in an easy-to-use data-center-focused toolset out of the box. However, this is probably not the main challenge, as tools are being enhanced rapidly and making these features more accessible.

The key issue is that equipment manufacturers tend to regard control systems as proprietary, and the control system response can be quite complex because of the number of fans and components monitored. As a consequence, equipment manufacturers tend not to openly publish the data for their equipment. This makes it difficult for software suppliers to produce good library models of equipment. Instead, they must provide libraries with limited characterizations. The user should therefore not assume that, just because a tool has the capability to represent all the characteristics of the library models supplied or built by other users, it will necessarily contain a complete characterization.

Even so, it is anticipated that over time, market forces and pressure from end users of the equipment will result in more data and better library models being available.

30.4.6 Low-Energy Designs

The focus on low-energy design is resulting in new technologies that need to be modeled using CFD. In fact, CFD in its general form has been used to model the fundamental processes used in low-energy design for many years. However, for data hall applications, it is more appropriate to use simplified models rather than to include the full detail of the entire physics. As a consequence, CFD tools for data centers are permanently evolving to capture the new approaches being used in the drive for lower-energy, more efficient designs.

At the time of writing this chapter, the technologies being employed for state-of-the-art data halls are largely associated with air-side and coolant-side economizers. The features that are now being introduced to data hall CFD tools include the following:

1. Water mist sprays and subsequent evaporative cooling. The model probably does not need to be a full model of the nozzle and droplet motion with full two-phase treatment of the liquid-to-gaseous transition. However, the model does need to account for where the moisture will be evaporated and how much will be evaporated since the evaporation process absorbs energy (heat), providing cooling that does not have to be provided by a chiller. This evaporative cooling, "adiabatic cooling," is often termed "free cooling" because it does not require a mechanical cooling system such as a chiller.
2. Wet media as a source of moisture with similar intention to water sprays: To predict the location and extent of adiabatic cooling that results from the change in phase of the water droplets into gaseous phase.
3. Mist eliminators: Where adiabatic cooling is applied to the air side of the data hall cooling system, any excess droplets that are not evaporated in the air handling section could be carried by the airstream into the data hall. Consequently, mist eliminators are deployed to remove any excess droplets.
4. Controllable dampers and vents for dynamic control of recirculation and free cooling.

These technologies, and particularly the simplified models for them, are currently in their infancy for data hall modeling. As a consequence, the options available vary widely from one CFD tool to another. The user should, therefore, consider carefully what they hope to achieve with the tools and should look for clarification of capability from the CFD tool supplier.

30.4.7 Challenges

As mentioned in Section 30.2, one of the main challenges to data hall modeling is that the size of enterprise data halls is such that to capture the full details requires detail on a scale

that results in very large calculation grids. The consequence of this is the requirement for significant computer resources and the resultant lengthy simulation times. There are currently two solutions being pursued to try to overcome this challenge:

a. Using alternative numerical techniques that try to simplify the physics in order to streamline and speed up the calculation. While the advantage of this approach is speed, the disadvantage is that it requires expertise from the user to know either when the simplified approaches are likely to deliver acceptable predictions or what to look for in the predictions to determine whether or not they are likely to be valid for the scenario(s) being considered.
b. Simplifying the data center models using current state-of-the-art CFD tools, but ignoring certain details that add complexity and, in the view of the user, do not need to be considered for the decisions under consideration. Similar to a), the key benefit is solution speed, but the key disadvantage is that it relies on the user understanding the impact of the simplifications applied. They must also understand what decisions the simulation results can be used to make or can contribute to.

Given an acceptance of the intensive nature of current data center CFD, probably the next greatest challenge to CFD is accuracy of the predictions. There are now numerous examples of CFD being used effectively to predict conditions for enterprise data halls; indeed, the tools are now being applied in day-to-day operational management.

While CFD is by nature an approximation—introducing errors into the prediction as a result of its formulation and methodology (e.g., the numerical discretization to solve the equations and the empirical formulae to represent turbulent mixing)—probably the biggest uncertainties in data center modeling occur as a result of a failure to truly describe the boundary conditions.

There are two key sources of error resulting from boundary condition description:

1. The use of simplified representations of the boundary conditions does not truly capture the behavior of the object because the true behavior is not well understood. A good example of this (that occurs for almost any data center using a raised floor to deliver cooling) is the modeling of perforated tiles.

   The traditional approach (as described briefly in Section 30.4.1.2) is to represent the perforated tile by a resistance to airflow to capture the pressure difference required to drive air through it. In practice, this is only part of the story because a real perforated tile is made up of several layers or elements.

   The top surface of the perforated tile does act, to a large extent, in the manner represented by a flow resistance to capture the pressure drop characteristic, but it also has other effects. It normally produces a localized region immediately above the perforated tile where the discrete jets from the individual openings in the tile coalesce. In this region, an additional pressure drop is produced, which can be recovered either by sucking additional air from the surroundings into the jet leaving the perforated tile or by the jet being forced to contract and slow down as it consolidates.

   Further, depending on the construction of this top surface, the direction of departure of this jet may vary from the underlying direction, resulting from the approach direction and turning that occurs as the air passes through the perforated tile substructure. As implied by this latter statement, this top layer of the tile is not the only layer of significance. There is also normally a considerable substructure present for a perforated tile, largely intended to provide the physical integrity required for the tile to withstand the loads created by heavy equipment being transported through the data hall.

   This substructure can dramatically influence how much air from the raised floor passes through the perforated tile and whether the flow direction under the raised floor plays a role in the air volume and leaving direction of airflow. In addition, a flow control damper layer may also be present, adding further geometric and flow complexity.

   At present, there are no scientifically documented methods for the creation of simplified models of such perforated floor tiles. Current research indicates that for a representative model, it is commonly necessary to add details of all the layers to a lesser or greater extent, depending on the tile construction. The user is, therefore, dependent on the CFD software provider to undertake characterization and provide simplified models on an individual perforated floor tile by perforated floor tile basis, or the user must undertake this characterization themselves.

   This is perhaps an extreme example of lack of understanding of characterization of objects, but in fact, when a user adopts a tailored CFD tool, they are accepting that each and every object characterization has been made appropriately.
2. The use of simplified representations of the boundary conditions does not adequately capture the true state of the data center. This is because the features that define it are either too detailed to survey and understand or too detailed to practically model using current CFD modeling techniques and typical hardware available to practitioners.

A good example of this category is a group of unstructured cables. While in principle the user could survey and record the precise path of every cable, the time and expense of doing so—not to mention the computational expense of modeling simulation of such a detailed representation—make such a strategy completely unviable and the improved accuracy not justified.

It is, however, this set of judgments—which details to include in a model—that will be critical as to whether or not it is sufficiently representative for the task in hand. These same choices drive the need for model calibration for a real facility model (Section 30.3.4.1).

Other challenges that occur in data hall modeling are a result of the large scale of a data hall and the time-consuming nature of modeling and simulation. Key challenges that exacerbate this performance issue are as follows:

1. The high rate of change of ITE, meaning that it may not be practical to run a simulation or simulations for each and every proposed deployment
2. The variety of equipment types available, each with unique characteristics, making it difficult to maintain a drag and drop library of all possible models and configurations
3. The dependence of ITE heat dissipation and airflow rate on utilization, compounding the lack of information available to fully characterize equipment
4. The increased resource that would be required to fully consider the numerous redundancy and failure scenarios that could occur

Despite all these challenges, CFD modeling is now a part of the data hall design and operational management, because it has been found to help in the delivery of more energy-efficient and lower-risk strategies.

30.4.8 Time-Dependent Simulation and Failure Scenarios

As mentioned, the Navier–Stokes equations and CFD are able to predict steady-state predictions of how a data hall will perform from a thermal perspective, but they can also allow for time-dependent variations. This is particularly important in the case of an entire cooling failure, but first it is appropriate to consider less extreme failure scenarios.

30.4.8.1 Redundancy Failure Scenarios

Mission-critical data halls and their support infrastructure are generally designed with redundancy in place to protect the ITE, processes, and data from failure. For cooling, the common strategy is to install additional cooling units so that should one or more fail, there is still sufficient cooling.

In a small data hall, it is normal to use $N+1$ redundancy. That is, if N cooling units are required to deliver the cooling air to the data hall, one additional unit needs to be installed "just in case." From an energy and maintenance perspective, it would normally be most efficient to operate with the additional unit in standby and to make wear and tear even for all units—to periodically change which unit is on standby. However, from a cooling perspective, this increases the risk. There will now be different $N+1$ cooling patterns for the data hall, and, in principle, all configurations should be tested. Some may suggest that the changeover is another state that should be tested by running a time-dependent simulation. However, the changeover period should be relatively short lived, and, unless the data hall is being operated right at the limit, it is unlikely that a cooling-induced failure will occur in the time for switchover between two acceptable cooling configurations.

For larger data halls, it is normal to include at least two redundant units so that while one is not in use due to maintenance, it is still possible for another unit to fail, leaving N units still operating. However, it will be apparent that as the number of redundant units increases, the number of flow configurations increases astronomically. Although in principle it would be easy to ask the computer to run simulations for every possible combination, this may not be practical. In any case, it is probably not worthwhile either, since the chance of an increasing number of units failing at the same time becomes less likely. In this instance, it is more normal to use diagnostic evaluations of cooling performance under normal operation to look for mission-critical equipment that may be heavily reliant on one or two cooling units and to select a small number of critical failure scenarios to consider. Examples of commonly analyzed configurations include (i) where a number of cooling units are bunched together because of the physical architecture and (ii) failure of the cooling units in a high-density area of the data hall.

Simulation of a $2N$ redundant cooling system should not be confused with the scenarios described earlier. In a $2N$ scenario, while the redundancy could be achieved by having twice as many cooling units as needed, the redundancy is commonly achieved by each N being cooled using independent cooling circuits. So, two independent failure scenarios should be considered: one for each of the two cooling circuits failing, leaving the units operational on the other cooling circuit.

In all these failure scenarios, it is important to recognize that, when fully loaded, the data hall may provide quite an uneven load on the different cooling units. It is therefore important to ensure that the capacity variation, as a function of on-coil temperature for the cooling units, is known. Only then can the simulation behave accordingly should the on-coil temperature rise.

30.4.8.2 Total Cooling Failure

Total cooling failure can only be addressed using a time-dependent solution. This is because the data center will change from being in a constantly cooled state to one where no cooling is actively applied. Accordingly, the model must continuously change with time. For a time-dependent simulation, the CFD program will calculate the change in conditions over small steps in time. The solution is still iterative and consequently expensive from a computational perspective. However, it should not be imagined that one time step will take as long as one steady-state calculation: a single time step normally takes a small number of iterations to converge if the time step is chosen appropriately.

Simulation of total cooling failure in a data hall by CFD programs is a special case time-dependent simulation. It is undertaken quite frequently because of the critical need to understand whether the design is likely to manage such a catastrophic failure well. However, due to the way in which data halls are currently modeled in CFD programs, there are a number of limitations that many, if not all, CFD tools are likely to suffer from and that require special attention if the predictions are to be most useful.

In practice, it is likely that CFD simulations of entire cooling system failure are likely to be conservative—they will predict faster temperature rise of the system than will occur in reality. It is also reasonable to assume that CFD tools designed for data hall modeling are more likely to have strategies in place to offset the unrealistic effects of modeling approximations and lack of data, and so they can reasonably be expected to perform better.

Classic simplifications and approximations that need to be addressed for a realistic simulation include the following:

1. The data hall model knows little about the external elements of the cooling system, such as the chilled water or refrigerant loops or the chillers themselves. While pumps are still running, these systems will add thermal inertia to the system and slow down the rate of temperature rise.
2. The cooling systems in the data hall have considerable thermal inertia, particularly in the heat exchangers. If the fans are connected to standby power systems, they will continue to provide cooling. If not, the large fans continue to run for a short period due to the inertial mass in rotation. Even when stopped, if there are no non-return dampers fitted to the cooling units, cooling will still be transferred into the air as it falls through the coils under natural convection or is drawn through the coils by the ITE fans that are running on standby (UPS power).
3. The building architecture included for a steady-state calculation of data hall conditions does not need to include the thermal inertia. So, when a failure scenario is modeled, special attention must be paid to these surfaces to allow for their thermal inertia.
4. Sheet metal/thin objects, such as cabinet walls, are normally modeled as thin. This is because their physical thickness does not need to be included explicitly to correctly represent the thermal resistance for a steady-state calculation. As a consequence and in line with the architecture, special treatment will be required to account for the heat capacity of these items.
5. The ITE itself does not immediately heat up as a result of air temperature rise at the ITE air inlets. Consequently, the temperature rise at the ITE outlet will be delayed. As these items are treated as "black boxes," generating heat and airflow for steady-state normal operation analysis, the time delays occurring in the temperature rise resulting from what may be thousands of IT items cannot easily be represented without special treatment.

If a user intends to use a CFD tool for this type of analysis, it would be prudent to investigate if and how the tool addresses some of these issues and, if not, whether there are practical methods that can be applied to the CFD models to obtain more realistic results.

When a CFD tool is used for a time-dependent analysis, it will calculate full results for each time step. Given the size of each data set and the fact that the time step is likely to be of the order of 1 s, an analysis over several minutes would result in hundreds of data sets. It is, therefore, common practice to save the full data sets only for selected times rather than saving the data for every time step. The data is normally saved after completion of time steps in a list or sometimes simply at a specified time step frequency. For every full set of simulation data saved for a selected time step, the user can use all the standard tools (Section 30.2.7) for inspecting the conditions in the data center as well as the data-hall-specific analysis tools (Section 30.4.9). In addition, the result views can normally be animated so that the user can visualize the change in conditions over time.

In addition to views and data from the complete saved sets of results, a time-dependent analysis usually automatically records the data at every time step for selected items. Data-hall-specific CFD tools have the advantage that they are normally able to automatically identify key data to be recorded as full time histories, but the user can normally identify locations/data that they would like to record. In non-data-hall-specific tools, the user is likely to have to identify all the data points that are to be stored.

These cooling-failure-specific analyses are only likely to be available for data-hall-specific CFD tools. However, most CFD tools have general time-dependent analysis capabilities. The advantage of these general time-dependent analyses is that they are not limited to cooling failure analysis and can be used for other variations over time. This includes

changes in IT utilization and the consequential changes in heat dissipation and airflow. The disadvantage is that all the special treatments that are included automatically for cooling failure would have to be manually implemented to obtain an equivalent solution.

30.4.9 Data-Hall-Specific Analysis

Once the software development team is focused on data centers rather than any other fluid flow and heat transfer application, it is natural to include some data-center-specific analysis of the wealth of data that is produced by CFD.

There are a significant number of useful metrics that can be used to distill the data to something that the mechanical engineer, facility manager, or IT manager can quickly understand. Of course, the most critical interest is whether or not the IT is efficiently cooled. One can simply plot the mean or maximum inlet temperatures for each cabinet, but this still means the user must compare the data with a reference (Fig. 30.14).

Published indices such as ASHRAE temperature compliance can be calculated, but for a live data center, it may be more appropriate to test the inlet temperature against the manufacturer's recommendations. In this case, CFD software can produce a plot showing how close the cabinet-mounted ITE is to overheating. It does this by comparing the maximum inlet temperature of the ITE with the maximum temperature recommended for the ITE by the manufacturer (Fig. 30.15).

To understand and optimize cooling efficiency, it is helpful to know where the cooling is coming from and where it is going. Simple plots can be made showing the load on each cooling system (Fig. 30.16).

Streamlines, as previously discussed, can be used more intelligently to trace air and show which air from which cooling sources goes into ITE inlets and which air is unused. This methodology can be extended to quantitatively measure the supply effectiveness of the cooling system delivering cool air to ITE. A similar methodology can be applied to the hot air exhausted by the equipment returning to a cooling system and how effective the cooling systems are at scavenging the hot air.

The same methodology can be used to track how much air recirculates either inside the cabinet or in the main body of the room.

There are several quantitative measures:

1. The Supply Heat Index (SHI) and Return Heat Index (RHI) are displayed once the model has simulation results. These indices, developed by HP [2], have been adopted by such organizations as ASHRAE and Green Grid.
   - In simple terms, SHI is a measure of the extent to which cold air from the ACUs is diluted by warm recirculated air before it reaches the inlets of the ITE. A value of zero indicates "perfect" behavior; that is, no dilution occurs, and the equipment inlet temperature is, therefore, equal to the ACU supply temperature.
   - Similarly, RHI is an indication of how much of the cold supply air mixes into the ACU return air stream without ever reaching any ITE. A value of 1 represents the "perfect" behavior, where the hot air returns undiluted to the equipment.
2. The Rack Cooling Index (RCI)® is a measure of how well the system cools the electronics within the manufacturer's specifications, and Return Temperature Index (RTI)™ is a measure of the energy performance of the air management system. RCI® is a best practice performance metric for quantifying the conformance

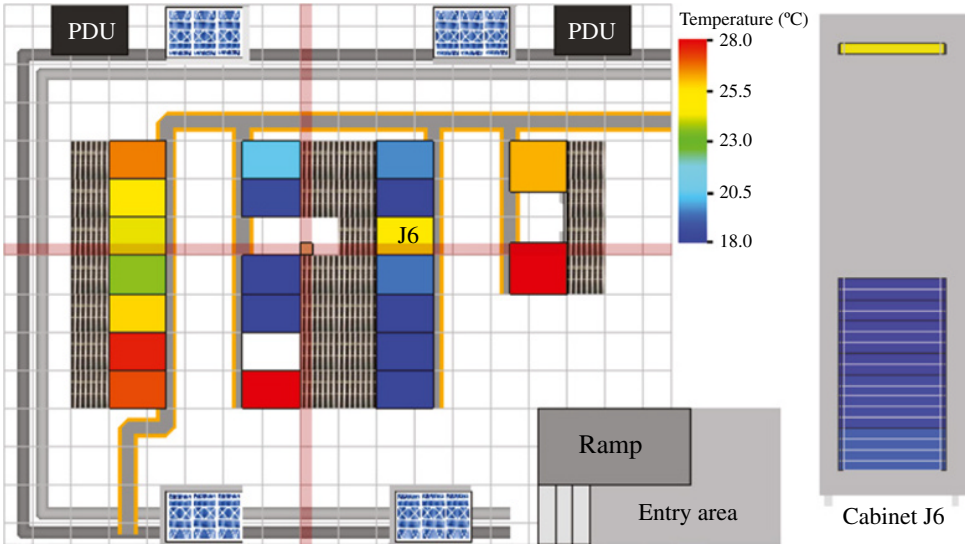


FIGURE 30.14 Maximum inlet temperature—room plan view (left) and individual cabinet view (right).

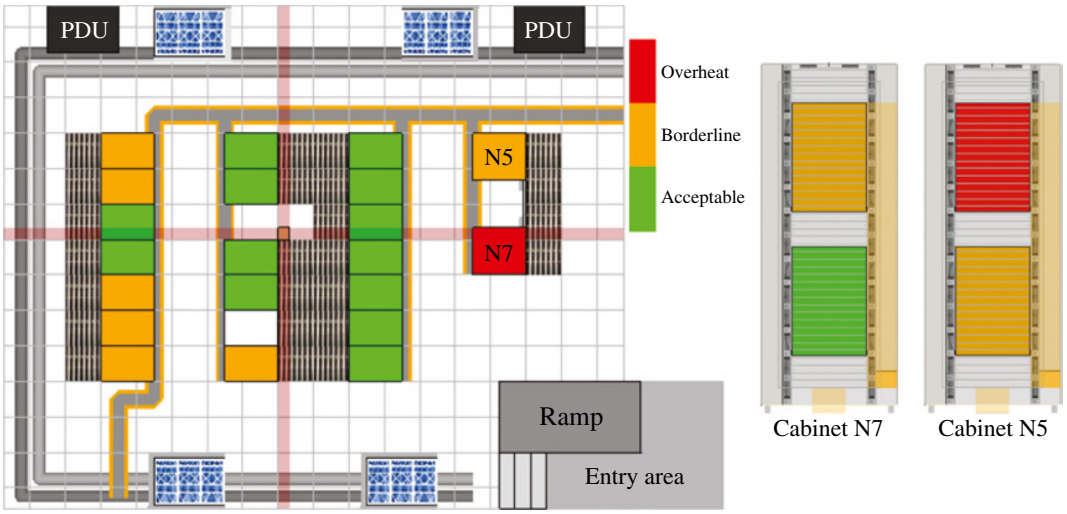


FIGURE 30.15 Plot of overheat risk.

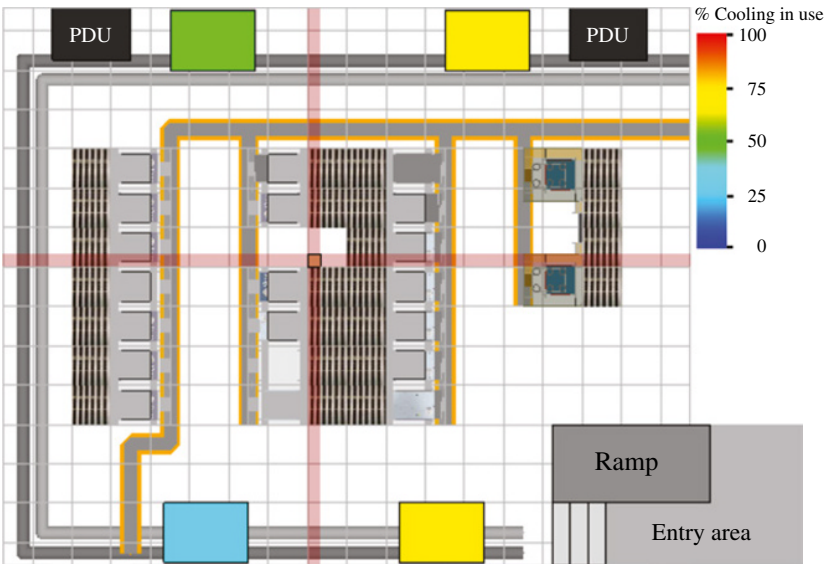


FIGURE 30.16 Cooling unit load.

with ITE intake temperature guidelines, such as those from ASHRAE and NEBS (Network Equipment Building System). RTI™ is a measure of net bypass air or net recirculation air in the ITE room. The indices [3] can be used under license from ANCIS Incorporated (www.ancis.us). RCI is calculated in two parts—RCI_{HI}™ and RCI_{lO}™:

- $RCI_{HI} = [1 - (\text{total over temp/max allowable over temp})] \times 100$, resulting in a %. A value of 100% represents ideal conditions: no over temperatures. A value below 90% is often considered poor.
- $RCI_{LO} = [1 - (\text{total under temp/max allowable under temp})] \times 100$, resulting in a %. A value of 100% represents ideal conditions: no under temperatures. A value below 90% is often considered poor.

where:

- Total over temperature represents a summation of over temperatures across all equipment intakes. An over-temperature condition exists when an intake temperature exceeds the max recommended temperature (ASHRAE Thermal Guidelines (2012) Class A1: 27°C).
- Max allowable over temperature is defined as max allowable temperature (32°C) minus the max recommended temperature (27°C) times total number of intakes.
- Total under temperature represents a summation of under temperatures across all equipment intakes. An under temperature condition exists when an intake temperature drops below the min recommended

temperature (ASHRAE Thermal Guidelines (2012) Class A1: 18°C).

- Max allowable under temperature is defined as min recommended temperature (18°C) minus the min allowable temperature (15°C) times total number of intakes.

RTI is defined as:

- RTI = $[(T_{Return} - T_{Supply})/(T_{EquipOut} - T_{EquipIn})] \times 100$, resulting in a %. A value between 80% (net bypass air) and 120% (net recirculation air) is often considered near-balanced airflow.

where:

- T_{Return} is the airflow (by volume) weighted average return air temperature to the cooling system(s)
- T_{Supply} is the airflow (by volume) weighted average supply air temperature from the cooling system(s)
- $T_{EquipOut}$ is the airflow (by volume) weighted average equipment exhaust temperature from the ITE
- $T_{EquipIn}$ is the airflow (by volume) weighted average equipment inlet temperature to the ITE

3. Simulation using CFD provides the inherent ability to trace where the air has come from and where it is going. Several indices utilize this functionality to characterize cooling system performance. Capture index is the property of APC by Schneider Electric. Capture index [4] includes the "cold aisle capture index (CACI)" and the "hot aisle capture index (HACI)," defined as follows:
   - Cold aisle capture index: The fraction of air ingested by a rack that originated from local cooling resources (e.g., perforated floor tiles, local overhead cooling supplies, or local coolers serving the same cold aisle cluster of equipment).
   - Hot aisle capture index: The fraction of air exhausted by a rack that is captured by local extracts (e.g., local coolers or return vents serving the same hot aisle cluster of equipment).
4. Cooling unit zone of influence: Since the air can be traced to determine how much of it reaches which ITE, CFD can show which ITE each cooling system affects. This is helpful in understanding the potential impact of a cooling system failure, although it should be noted that the airflow will change even if only one cooling unit fails. Accordingly, this measure is only an indicator of the actual effect.
5. Tailored analysis can also be extended to more general metrics—for example, if performance data is provided for external parts of the system, such as the chiller and associated chilled water distribution systems, power usage effectiveness (PUE) can be calculated. This is possible because the CFD model knows how much power is consumed by the IT, and, using this with the external data and the internal calculated cooling system performance, the model can calculate how much power is used by the entire system. Similarly, the data can be used to summarize energy consumption and associated costs.

It is anticipated that as CFD is further adopted in day-to-day operational management and becomes more closely integrated with other tools, these purpose-specific analyses will be further developed.

30.5 POTENTIAL ADDITIONAL BENEFITS OF A CFD-BASED DIGITAL TWIN

By definition, CFD models must describe the physical geometry of the facility, the cooling system configuration and operation, and the heat loads that are to be cooled. For a facility in operation, if the simulations are to be capable of predicting the detailed cooling performance, it is likely that the model will need to contain significant details of the infrastructure and ITE alike.

As a consequence, the models can be easily extended to address other data center equipment deployment issues relating to space, power, network, and weight issues. They could easily be used as PDCIM tools themselves or integrated with other DCIM tools, monitoring systems, or even IT application-based tools.

One way of defining an operational CFD model is to include every item of ITE explicitly in its correct cabinet and u-slot, to effectively manage how much heat it dissipates, and to know its connectivity into the power system. The result is that the CFD model will automatically provide an inventory of the equipment located in the data hall.

Another aspect of CFD modeling is the need to determine how much power is dissipated as heat by the ITE in different locations. Given that the applications installed and the loading of the ITE will significantly affect the heat dissipation, one way of identifying the power consumption of a live facility is to use live power monitoring data. When this is done, the model will probably have the power network connectivity stored to enable it to make use of available data. In such a case, the CFD data hall model can be used to analyze the power system, including load balancing and the impact of single (or multiple) points of failure on availability etc.

In a similar way, it is natural for the 3D model to be extended to hold data so it can be used as a DCIM tool. However, given the predictive character of a CFD tool, it lends itself to being used for capacity planning as a Predictive DCIM tool.

Bearing in mind the complexity of data centers in terms of construction, logistics, and operation, it is likely that separate tools will continue to be provided and used in corporate organizations for individual or small groups of tasks, including asset management and inventory, monitoring deployment, etc. It is therefore extremely valuable if the different tools can share data. The advantages in respect of the CFD model are severalfold:

1. IT inventory can be drawn from a chosen asset (or DCIM) tool.
2. Planning changes can be received from a chosen DCIM tool or transmitted to the chosen DCIM tool following analysis in the CFD tool.
3. Power network and consequent heat dissipation can be derived from a third-party power modeling and/or monitoring system.
4. Monitored data can be displayed in the 3D view provided by the CFD tool.
5. Cable routing can be made in the 3D model prior to implementation.
6. Alternative plans can be analyzed with a view to avoiding lost capacity by considering space, power, network, weight, and cooling holistically.

If this type of approach is adopted where several tools share data, the end user can treat the applications as being a whole, and the result is likely to provide a value greater than the sum of the individual parts.

30.6 THE FUTURE OF CFD-BASED DIGITAL TWINS

As computers become an increasingly significant part of our daily lives, the demand for data centers is growing astronomically. One of the primary concerns about CFD in the Digital Twin is the significant compute requirement necessary to represent a live facility and simulate changes. This has resulted in many people and organizations exploring the possibilities of using simplified methodologies (such as potential flow and other reduced order models) in order to achieve faster computing: simulation in real time, in an ideal world. Fortunately, the very driver for computing in so much of our everyday lives is also providing the potential for rapid computation. One example is the rapid development of GPUs with hundreds of processors. As they are developed and gain access to sufficient on-board memory, they offer the opportunity for massive parallelization (and consequent high-speed solution) on a relatively low-cost platform, negating the need for simplified solution methodologies.

Since it seems likely that the computational challenge will be overcome, how will the scope of these tools change? Will they be adopted for management? It is already clear that CFD tools can be and are being used for more than one-off simulation of a proposed or existing facility. They have already been applied to the management of real facilities on an ongoing basis, focusing on deployment and capacity planning and accounting for space, power, network, and weight as well as cooling.

However, as well as Digital Twins becoming the norm in all walks of life, so is artificial intelligence (AI) in the form of machine learning (ML). Simply put, ML is the use of algorithms that observe system behavior and, from the knowledge gained, predict future behavior for other circumstances. One obvious opportunity for the application of such a model in the data center would be to drive the cooling using a control system based on known behavior rather than traditional controls such as PID. The advantage would be that specific circumstances could be more directly addressed. For example, when a particular rack experiences high inlet temperature, the model could know which cooling systems actually cool that IT and the controls for that/those unit(s) could be adjusted accordingly.

How well the ML predictions match reality, and thus whether the controls are indeed effective, depends not only on the ML algorithms but also on how much data is available for training. A mission-critical facility or data center may not find this approach very confidence building since, based on monitoring and measurement, the data center would need to be operated without any understanding of behavior for a variety of conditions before any confidence could be placed in the ML-based control. This is an ideal opportunity for the Digital Twin. Many scenarios can be simulated, a priori, to gain a broad understanding of data center behavior reducing the risk of inappropriate control response. The simulated data could then be complemented by measured data in two ways:

1. As measured data becomes available, the expected behavior can be increasingly understood (i.e., the ML trained) from measured performance rather than simulated performance alone.
2. When future scenarios are outside the already understood conditions, additional simulations can be undertaken to extend the knowledge of likely behavior.

The advantage of this approach is that the instantaneous control response required by control systems can be met by the reduced order model that results from ML. At the same time, the understanding of behavior still benefits from the comprehensive analysis provided by simulation using CFD in the digital twin.

Whether CFD tools become the core platform, or whether other management tools will absorb CFD tools, is not yet clear. What is certain is that an integrated toolset including CFD will be the norm. The toolset will almost certainly be broader, being the central reference for anything you need to

know about your data centers around the world. It is also likely to go deep into the application layer as well as the facilities and physical layers.

CFD is also used to design much of the equipment being housed in a data center. Consequently, it also seems likely that the simplified models used in the Digital Twin described here will be automatically generated and be an output of the equipment design process. If members of the supply chain follow the electronics industry's lead (e.g., chip vendors making chip/component thermal models available), compact object models could substantially improve model quality and uniformity, making the modeling process much quicker, simpler, and more reliable.

One of the questions often posed about the future of CFD is, "will it continue to be necessary?" These questions have been raised because people see changing technologies as taking away the design and operation issues. In recent times, the question has arisen because of the use of hot and cold aisle containment. After all, to the uninitiated, physical segregation will prevent air mixing. However, in practice, segregation cannot be perfect and, if recirculation is allowed to take place, can be much more serious owing to the higher temperature of the undiluted return air. The emerging technology that re-raises the question is that of liquid cooling. If liquid cooling is used in place of air, why would CFD be needed? In practice, liquid-cooled systems are desirable where power densities are high. Most, if not all, liquid cooling systems still leave some of the heat in the room to be carried away by the room cooling system, and so, with high power densities, this may still be a significant cooling load. Further, the liquid cooling systems themselves will become large and complex: they are therefore likely to require some sort of simulation of their own for the purposes of optimization.

Further, the data hall is just one part of a data center. There are many other parts of the system that may warrant CFD, such as free cooling systems, generator halls, UPS rooms, etc. In summary, in the view of the author, for the foreseeable future at least, CFD is here to stay.

REFERENCES

[1] Idelchik IE. *Handbook of Hydraulic Resistance*. 3rd ed. London: Hemisphere; 2005.

[2] Sharma RK, Bash CE, Patel CD. Dimensionless parameters for evaluation of thermal design and performance of large-scale data centres. Proceedings of the 8th AIAA/ASME Joint Thermophysics and Heat Transfer Conference 24th June 2002–26 June 2002, St Louis, Missouri. American Institute of Aeronautics and Astronautics; 2002. AIAA-2002–3091.

[3] Herrlin M. Airflow and cooling performance of data centers: two performance metrics. [PDF] ASHRAE Trans 2008;114(2):182–187. Available at http://www.anci.us/images/SL-08-18_Final.pdf. Accessed on January 24, 2013.

[4] Shrivastava SK, Van Gilder JW. Capture index: an airflow-based rack cooling performance metric. ASHRAE Trans 2007;113(1):126–136.

31

DATA CENTER PROJECT MANAGEMENT

SKYLER HOLLOWAY
Facebook, Menlo Park, California, United States of America

31.1 INTRODUCTION

Projects are everywhere. Any time that resources and work are combined with an expected outcome, a project is created. These can differ wildly from quick tasks completed in succession on a project at home to multiyear, multiagency ventures that cost significant sums of money. Within the data center space, projects are no different: they are ubiquitous and can include building construction, physical network rack/stack/patching, systems provisioning, and many others. Because of how frequent projects are, successful project management can provide organizational benefits such as decreased timelines, efficient use of resources, and predictable outcomes. This chapter attempts to provide an overview of common attributes of all projects that is widely applicable across the data center space. Further, readers are encouraged to seek to implement the strategies and methods provided here to improve the outcomes of projects to which they are involved.

While many readers will understand or use the terms project management and program management interchangeably by mistake, they are not synonymous. Project management is the hands-on tactical management of a project, whereas program management is typically a higher-level oversight of multiple programs or projects. The two topics do have many common themes, such as schedules or budgets. This chapter will focus on principle of project management.

31.2 PROJECT KICKOFF PLANNING

At the initial launch of a project, a formal kickoff should occur. Led by the project manager, the intent of this step is to introduce the project to both the immediate team members and broader or extended team members as a finite project with an expected outcome, establish the team identity, outline roles and responsibilities, chain of command, and finally ensure commonly agreed objectives of the project. While this step may seem like a formality in some cases, a successful project kickoff sets the initial direction of a project and greatly minimizes the chance of confusion or misalignment between project team members. A simple, but certainly not exhaustive, checklist of topics to be covered during a project kickoff is listed below. A project kickoff can take the form of a detailed email, virtual meeting, or multiday in-person conference depending on the complexity and importance of the project. It is the project manager's role to tailor the kickoff planning to the specific project needs based on the context of the situation:

1. Create list of required and option attendees.
2. Align all project team members on roles (can be done before kickoff as well).
3. Highlight and gain consensus on project objectives.
4. Agree to standard expectations of all team members (e.g. meetings to be attended, level of responsiveness) and resources to be assigned to the project.
5. Determine project-tooling strategy and communication protocol.
6. Review all common information with stakeholders. Can take the form of a project charter if desired.
7. Assign single point of contact from each vendor.
8. Conclude project kickoff and issue formal communication of the outcomes.

31.3 PREPARE PROJECT SCOPE OF WORK

31.3.1 Prioritize Project Scope, Timeline, and Cost

The relationship between a project's scope, timeline, and cost is commonly referred to as the "triple constraint" of project management (Fig. 31.1) because the three pieces are each a

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

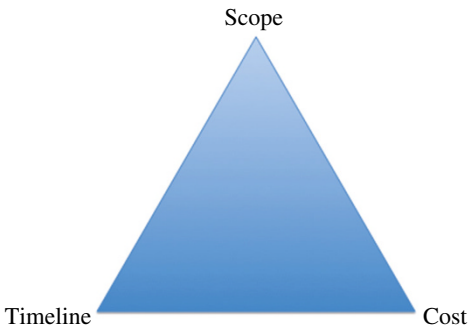


FIGURE 31.1 Triple constraint (prioritize project scope, timeline, cost).

constraint with which the project team must comply. Increasing focus on one of the constraints will result in a de-prioritization of the other(s). For example, if cost is the most important consideration for a project, then the trade-off will exist that the project will not be completed as urgently as possible (timeline). This is true because methods to shorten the timeline of a project often increase costs such as over-time pay, higher-skilled labor, fees to expedite (permits or materials), etc. In a separate example of a project to develop a software application, there will be trade-offs that need to occur between the features to be included (defined as the scope) and the cost and/or timeline to complete the project. A project with extensive requirements will certainly increase the other constraints (timeline and cost). So, it is important to prioritize which of the triple constraints should be most important for a project as this will drive team decision-making. In the case that all three constraints are said to be the first priority, no prioritization at all has taken place. Although discussions around priorities can be difficult, such as when a finance team and project delivery team have competing objectives between cost and timeline and/or scope, they should be discussed up front during initial project phases or impactful misjudgments may occur at later stages.

31.3.2 Develop Written Scope Document

In addition to determining prioritization between scope, timeline, and cost, the project team will need to create and issue written scope documentation. Referred to as "statement of work" (SoW), or simply as the scope, this documentation serves as the baseline to which a project's completion can be measured. The SoW is also the agreement between the owner and the rest of the project stakeholders regarding what the project is and, importantly, also what it is not. Development of a SoW may take many reviews and iterations as the project team creates a consensus on the triple constraint trade-offs mentioned above. In many cases, this SoW becomes a functional part of the contractually agreement between an owner and vendors.

31.4 ORGANIZE PROJECT TEAM

31.4.1 Selecting Stakeholders and Outlining Responsibilities with a RACI Chart

Stakeholders are any person, team, or company that will have an active role in the project. Creating clear expectations and responsibilities for all of the stakeholders is an important but an often overlooked step to ensuring that there is broad alignment between all team members on whom will have differing levels of involvement for each step of the project. A commonly used tool for outlining this information is known as a RACI matrix (Table 31.1)—an acronym for the four levels of responsibilities each stakeholder can be assigned: responsible, accountable, consulted, and informed. In simple terms, a RACI matrix is used to assign roles and responsibilities for all project team members. RACI should be reviewed by all parties and agreed to before project adoption. In addition, at later stages of the project, the RACI can be a source of ongoing reference if disputes or questions arise regarding responsibilities.

31.4.2 Team Building

Team building can take various forms, but the intent of time spent during team building is to develop familiarity and personal relationships between all team members. These relationships will ease communications and the solving of

TABLE 31.1 RACI matrix

	Project manager	Designer	Engineer	Owner/End-user	Financial controller
Create project budget	A	C	C	I	R
Design building layout	A	R	C	I	I
Permit plans	A	*C*	R	I	I
Complete construction	R	A	A	I	I
Public opening	A	C	C	C	C

(**R**)esponsible, (**A**)ccountable, (**C**)onsulted, and (**I**)nformed.

disagreements during the life cycle of a project. Any social, nonwork-related event can serve as a chance for team building, but the time can also include short enjoyable activities during the workday. Team building can be incorporated into the kickoff events and also periodically redone during projects to ensure new members are incorporated into the group. Often overlooked, quality team building can add to the individuals' satisfaction in their work and contribute to the creation of high-performing teams. Team building activities may require financial resources, so provisions should be made in approved project budget.

31.5 PROJECT SCHEDULE

Creating and maintaining a project schedule is a key component of project success. An accurate schedule ensures that all stakeholders have the same understanding of project timeline expectations and can effectively plan their own contributions to the overall project. Also, an understanding of the activities on the critical path can help drive project decisions (see Section 31.5.1). Without a schedule, project stakeholders are left uninformed of when key activities will be taking place, resulting in confusion and poor coordination. The project schedule should be accurate, accessible to all, updated regularly, and include input and acceptance from all stakeholders.

31.5.1 Critical Path and Float

A commonly used term in scheduling is critical path. This is defined as the set of tasks that generate the longest possible overall completion time of the project. For instance, if an activity on the critical path is delayed, it will result in an overall project delay, whereas the same will not happen with an activity that is not on the critical path. Because of this reason, understanding the critical path of any project provides the decision-makers with an essential understanding of the most critical activities to which special attention and potentially resources should be focused. Conversely, noncritical path activities can be de-prioritized to an extent without an overall negative impact to the project timeline. This amount of time (typically in days) is referred to as the "float." Activities on the critical path are considered to have zero float.

31.5.2 Basic Schedule Types

There are numerous types of schedules that present information in different ways, and each has unique characteristics. Generic examples of two of the most common approaches—Program Evaluation and Review Technique (PERT) and a Gantt chart—have been provided below. Either, or both, of these approaches can be used based on the project's scheduling needs.

31.5.2.1 Program Evaluation and Review Technique (PERT)

PERT is a useful type of schedule because it shows the relationship between individual tasks and thus helps create a priority around these tasks (Fig. 31.2). Once a list of all tasks or activities has been created, their relationships can be outlined (Table 31.2). A predecessor is defined as a task that must be completed in order for the next task to begin. For example, building the racks in a data center is required before the switches can be placed inside of these racks. Thus, building the racks (Activity A) is a predecessor to stacking the switches (Activity B). If no prerequisites are required for a project task to begin, then it simply is listed as having no predecessors (Activity A). This precedence network, or PERT chart, (Fig. 31.2) captures this relationship, which allows the project team to correctly sequence and plan the individual tasks of a project. The following Q&As are listed to better illustrate PERT technique.

Question: Before Task H can begin, what other tasks must be completed?

Answer: In order for Task H to begin, specifically Tasks F and E must be complete, in addition to all of *their* predecessors. So, F and E is the correct answer, but A, B, C, D, E, and F is also a correct answer.

Question: As the project manager, if you sense that Task D or Task E will likely suffer from a delay, which task should you consider to be more important? Why?

Answer: Based solely on the information provided including the assumption of the critical path in darkened blocks and bold or thicker lines, the delay to Task D should be treated as more urgent. The main difference between Tasks D and E is that D lies on the critical path, which means that any delay in D will delay the project completion proportionally. A delay to

TABLE 31.2 Create and outline task relationships

Activity	Predecessor(s)
A	—
B	A
C	A
D	B
E	C
F	D
G	D
H	E, F
I	G, H

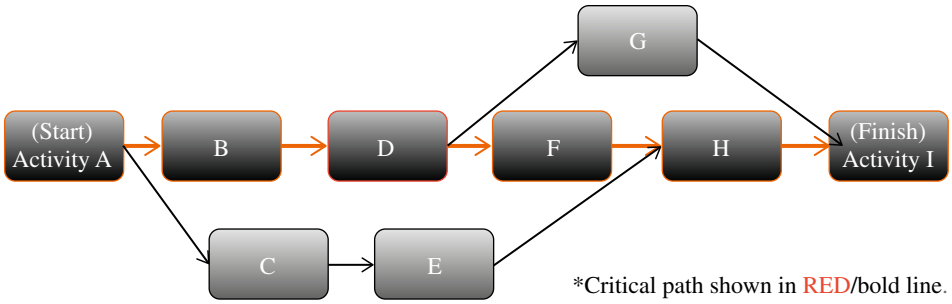


FIGURE 31.2 PERT chart captures tasks' sequence and reflects critical path activities.

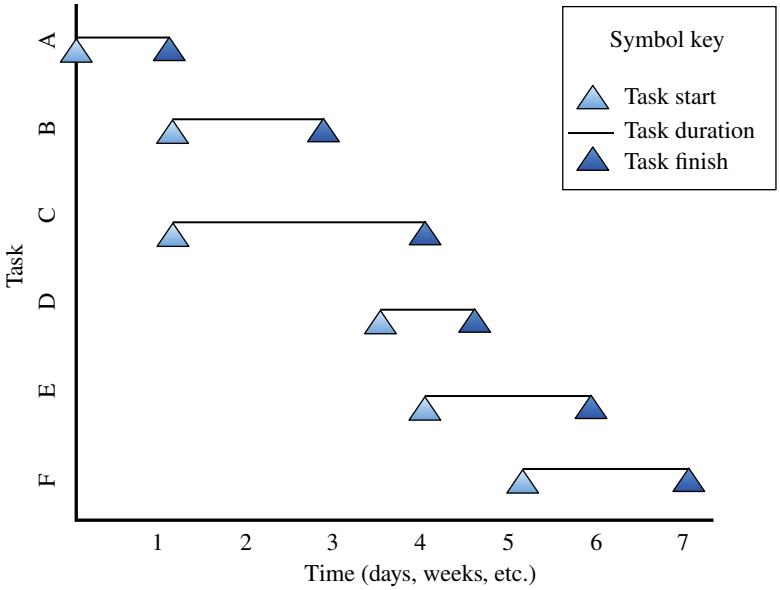


FIGURE 31.3 Gantt chart.

E will not immediately incur the same outcome until the entire float of E in the schedule has been exhausted.

31.5.2.2 Gantt Chart

A Gantt chart (Fig. 31.3) is a second popular type of schedule that provides distinct differences to a PERT chart, such as the inclusion of individual task durations and in some cases the absolute dates that a task should occur. Gantt charts are in the form of a X-Y axis or similar layout. In the example shown, the *Y*-axis lists each of the tasks and the *X*-axis shows the timeline on which each of the tasks will occur. For task A, a project team member could recognize that Task A will take approximately one unit of time and will start at the very beginning of the project. When an absolute time measure is used (e.g. measured in days and kick-off on January 1) (Fig. 31.3), the rest of the tasks can be planned for specific dates on the calendar (Task B occurring on January 2 through 3, inclusive). While Gantt charts provide useful information for project planning, they typically do not visually show the predecessors, although this information can be added in a dedicated column.

Using a Gantt chart (Fig. 31.3), comprehension checks could be performed as follows:

Question: If the project shown in Figure 31.3 is 7 days in total duration, how many days is Task B expected to take?

Answer: Task B is expected to take 2 days total.

Question: In the Gantt chart provided, what additional information could be helpful to understanding the scheduling logic?

Answer: As one example, critical path information could help the project manager understand the most crucial tasks. Alternatively, dotted lines or a chart that details each tasks' predecessors would allow for the critical path to be calculated.

31.5.3 Schedule Control

The first step to effective project management is certainly to create the best schedule possible, but in most modern-day projects this alone will not satisfy all project scheduling needs.

Projects commonly deviate from the initial schedule in many ways including delays, scope changes, resource reallocation, change in business priorities, etc. A process should exist around how the schedule is regularly updated by all team members to adjust to the realities of a dynamic project. For example, once the initial schedule has been created and issued, the project manager could seek updates and adjustments on a weekly or monthly basis to capture the changes and reissue a version-controlled schedule. This is an important step to ensure that at anytime all team members comprehend and execute tasks based on the latest schedule. In the simplest terms, it's important for all musicians to play off the same sheet of music!

31.5.4 Project Scheduling Tools

Some of the common applications of tools for project management include electronic document storage (Box, Dropbox, "X" Drive, etc.), virtual meeting hosting (GoToMeeting, Microsoft Teams, Webex, Zoom), information reporting structures (for creating progress dashboards or meeting minutes), and email and instant messaging capabilities. The number of tools that can be engaged on a project is truly limitless.

Because the best tools on the market for each application are subject to change and project needs are all unique, we will not review or suggest specific tools, but instead provide considerations for use while developing tool strategy.

These considerations include:

- How many users will require this tool
- Based on the number of users (which commonly drives pricing), what will be the cost
- If cost is an issue, would a cheaper or free option available that can be used
- Which company and individual role is responsible for onboarding (training, access, etc.) and ongoing management or troubleshooting
- How long is the intended use of the tool
- How will the information be archived (if applicable)

In summary, creating a tooling strategy and ensuring that all parties adopt the common approach to tooling is important for project success. Properly planning out this strategy proactively will result in the best outcome for the project.

31.6 PROJECT COSTS

31.6.1 Estimates and Budgets

How much a project costs is an important piece of information for all stakeholders. Estimates budgets and costs-to-date are useful tools that can help to create, predict, monitor, and record the costs over the life cycle of a project, although sometimes their meanings can be overlapping. Estimates are typically described as the expected final cost for a project through completion, but there are different types of estimates: rough order of magnitude (ROM), detailed, and budgetary estimates. Depending on the point in a project and specific needs, different estimates can be more suited for use. In accordance to financial practice of a company, ROM could be an estimate based on schematic design. Detailed estimate is established after construction drawing. And a budget estimate is a detailed estimate with additional 10% to cover contingency. The purposes for estimates during projects include setting aside business capital for use, monitoring spending versus a benchmark, and tracking spending to prevent waste or over spending. For each of these purposes, different types of estimates can be best suited for meeting each of the project's needs, even for different stakeholders. For example, company executives may desire a high-level budgetary number so they can be fully aware of and plan for the worst-case spending potential. In a separate example, the project manager may want a general sense of what an optional project will cost so that the cost/benefits can be weighed before the project is commenced: a ROM estimate (Table 31.3) can serve this purpose. Budgets are commonly held on projects to capture the expected costs that are produced during the process of creating an estimate. Although the terms can have overlapping and even combined use, estimates and budgets should be understood as different parts of projecting and controlling costs on a project.

31.6.2 Project Cost Control

Inevitably over the course of a project, the expected costs will be adjusted. This can occur in two main ways: the magnitude of the project cost can change, i.e., $10–11 million, or the certainty

TABLE 31.3 Rough order of magnitude estimate

Budget type	Expected accuracy	Explanation
ROM estimate	±15–25%	Early-stage cost estimate used for project feasibility and investment go/no-go decision. ROM stands for "rough order of magnitude".
Detailed estimate	±5%	Most precise estimate that can be considered the "best guess" because there is equal chance the project comes in over or under this budget. Contingencies are often removed or separately noted.
Budgetary estimate	–10% only	This is the maximum spendable amount, including contingency, and any other potential costs. The final cost should always come in under this amount. Commonly used for high-level capital budgeting purposes.

in the final cost can change, e.g. the team believes the project will cost $10–12 million versus the team knows the cost will be $11.51 million with little margin for error. Both types of cost adjustments are important for helping the business manage the investment that a project represents. Project cost control is a system or set of steps that ensure an accurate cost understanding is maintained over the course of project. Typically done on a monthly cadence to work within financial accrual requirements, cost control does not require specific steps, but rather a prediction process. Further description is not provided because cost control is often dictated by company policies.

31.7 PROJECT MONITORING AND REPORTING

As projects will naturally range in length from short-term to much longer, constant monitoring and periodic reporting is a necessary step in project management that must be considered by the project manager. Reporting is also how non-project team members can stay apprised of project progress and other important information. For example, one key report may be the executive update that could be provided to leaders on a monthly basis. In this example, a report would be issued specifically for the intended audience and include relevant information only such as budget update and overview of progress, but may exclude details that executives may not need to know. Each project report should be tailored to fit the unique needs of the project and issued in a predictable fashion and cadence. When aligning project reports, the project manager should seek input from team members on what information that they need and the frequency at which the information should be provided. Many example reports can be found online to gain ideas of reports that may be useful for your project.

31.8 PROJECT CLOSEOUT

Although commonly overlooked, formal closeout of a project is an important step to be completed by the project manager and stakeholders. Closeout marks the official end of the project. A simple checklist has been provided below, although the items may not necessarily need to occur sequentially based on unique project needs:

1. Review project objectives from the kickoff and ensure that each has been fully completed.
2. Ensure all outstanding payments have been made to vendors and subcontractors. No further requests for payment to be accepted.
3. Officially closure of the budget and return any unused money (both contingency and underspend).
4. With all appropriate team members, complete a Lessons Learned or similar process to review the areas where the project team could have improved the approach or outcomes. Save and distribute this information to future project teams.
5. Create and disseminate a process for additional scope requests (to be handled on a discrete future project).
6. Celebrate a job well done with the project team. Make each individual feel appreciated for their contributions to the groups' success.
7. Communicate broadly about project completion and successful outcomes. Highlight the business needs that drove the project inception.

31.9 CONCLUSION

Project management in the data center space, and more broadly, is vital to organizational success in that successful project management ensures that executive strategies are fully carried-out in an optimized fashion. Project management involves very specific steps and expectations that have been briefly reviewed in this chapter, but at times project management cannot be distilled so simply. Project management can be as much art as it science and relies on the soft skills, communication abilities, and even instincts of the project manager for ultimate project success. These skills will be developed and honed over time, but a focus on the fundamentals of project management will allow for the basic tenants to be guaranteed on all projects.

FURTHER READING

Portny SE. *Project Management for Dummies*. 5th ed. Hoboken, NJ: Wiley; 2017.

Project Management Institute. *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)*. 6th ed. Project Management Institute; 2017. Available at www.pmi.org. Accessed on June 8, 2020.

PART IV

DATA CENTER OPERATIONS MANAGEMENT

32

DATA CENTER BENCHMARK METRICS

BILL KOSIK
DNV Energy Services USA Inc., Chicago, Illinois, United States of America

32.1 INTRODUCTION

In addition to Power Usage Effectiveness (PUE), there are currently several energy and performance metrics designed specifically for data centers. Many of these metrics are used to evaluate systems beyond the facility power and cooling systems. An example of this is to judge the efficacy of the IT systems. The maturity and usefulness of these metrics ranges from fully established to ones that are still under development. The information presented here focuses on metrics and evaluation programs, such as PUE, that are currently used by data center owners to benchmark data centers and look for ways to reduce energy use.

32.2 THE GREEN GRID'S PUE: A USEFUL METRIC

In order to develop some historical perspective on the development of meaningful data center energy efficiency metrics, we need to go back to 2007, when the U.S. Environmental Protection Agency (EPA) produced a report at the request of the U.S. Congress on server and data center energy efficiency. One of the more prescient statements from the report stated: "The federal government and industry should work together to develop an objective, credible energy performance rating system for data centers, initially addressing the infrastructure portion but extending, when possible, to include a companion metric for the productivity and work output of IT equipment (ITE)." This was a clear directive to the industry to develop a uniform metric that is useful to both the private and public sector in their power, cooling and IT systems. Just a few years later, The Green Grid the PUE metric which is used to determine energy efficiency of data centers. In short order, PUE became one of the top metrics, used worldwide, for organizations that have significant data center portfolios and needing an objective method to drive energy efficiency. While the definition of PUE is generally understood and has been in use for several years, the industry is still fine-tuning how PUE is computed, both theoretically and when measuring energy consumption from a live data center.

It is important to remember that PUE should not be used to compare different data centers and especially should not be used as a marketing tool. Stating a PUE of a data center without providing design/operating details (climate, cooling system type, reliability level, percent loaded, etc.) is somewhat meaningless. PUE considers how energy is consumed for all systems within the data center, cooling, power distribution, and other ancillary systems. A PUE can be developed for the power and cooling systems (similar to mechanical load component (MLC) and electrical loss component (ELC) described in *ASHRAE 90.4, Energy Efficiency in Data Centers*), but more importantly, as a group representing a total PUE for the entire facility. The PUE is a measure of energy efficiency and is represented by the following equations:

$$\mathrm{PUE} = \frac{\sum \text{Power delivered to data center}}{\sum \text{IT equipment power use}} = \frac{P_{\text{mechanical}} + P_{\text{electrical}} + P_{\text{other}}}{P_{\text{IT}}}$$

$$\mathrm{PUE}_{\text{mechanical}} = \frac{P_{\text{mechanical}}}{P_{\text{IT}}}$$

$$\mathrm{PUE}_{\text{electrical}} = \frac{P_{\text{electrical}}}{P_{\text{IT}}}$$

Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center, Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

$$\mathrm{PUE}_{\mathrm{other}} = \frac{P_{\mathrm{other}}}{P_{\mathrm{IT}}}$$

Energy use is measured at the main utility transformer, mechanical switchgear, UPS (Uninterruptible Power Supply), and miscellaneous load panels. Measuring the input and output power indicates efficiency of different electrical components, which becomes another data point when attempting to optimize efficiency. The document *PUE: A Comprehensive Examination of the Metric*, authored by The Green Grid, provides guidance on a consistent methodology for establishing PUE.

32.3 METRICS FOR EXPRESSING PARTIAL ENERGY USE

32.3.1 The Green Grid's Partial PUE (pPUE)

The document *PUE: A Comprehensive Examination of the Metric*, authored by The Green Grid, provides guidance on a consistent methodology for establishing partial PUE (pPUE). To compare, PUE is defined as the total data center facility energy divided by the ITE system energy. There may be circumstances when only a part of the facility energy is known; in order to develop a PUE-like metric in cases like these, The Green Grid developed the pPUE metric.

As an example of when it is appropriate to use pPUE, there are situations where a data center gets cooling from another source, like district cooling or from a central plant. In this case the data center owner may want to publish values for an industry-recognized energy efficiency metric, but has no control of a major portion of the energy-consuming systems. In this case, the data center owner can still calculate the sum of the energy use and losses for equipment that supports the ITE. Instead of a system-specific metric, it is a zone-specific metric. The starting point is the ITE: any equipment and systems under the control of the data center owner will be included in the pPUE calculation. So if there are transformers and local cooling equipment that support the ITE load, they are included, even if the primary cooling is provided by others. The equation is as follows:

$$\begin{aligned}\mathrm{pPUE} = {} & \big(\text{Zone energy } (\text{power, cooling, electrical losses}) \\ & + \text{ITE energy } (\text{in zone})\big) \div \text{ITE energy } (\text{in zone})\end{aligned}$$

Of course, pPUE cannot be used in lieu of PUE and is best used for internally managing energy consumption and should not be used as a comparison tool.

Note that pPUE should not be considered a substitute for a complete PUE evaluation. It is always best to develop a comprehensive PUE model in order to fully understand a facility and manage it with complete knowledge of all components. Furthermore, as with PUE, pPUE is a metric best used as a tool for management, rather than for making comparisons with other facilities.

32.3.2 ASHRAE 90.4 Mechanical Load Component (MLC)

MLC is fundamentally different than pPUE but still demonstrates efficiency of a subset of the data center energy use (see Section 32.8.3 and also in the Chapter 11 relating to ASHRAE 90.4). While pPUE is mainly used for internal energy efficiency monitoring and can be applied in different ways for different situations, MLC is solely used for determining compliance to the minimum requirements in ASHRAE 90.4. The 90.4 standard provides details on how to calculate MLC, but an important aspect to MLC is determining energy use at 25, 50, 75, and 100% of ITE load. This reveals the energy use at part loads, which is where the data center will predominantly operate—rarely (if ever) will a data center operate at 100% of its design capacity. The equation for MLC is as follows:

$$\textbf{Annualized mechanical load component}\,(\mathbf{MLC}) = \frac{(\text{Mech_energy25\%} + \text{mech_energy50\%} + \text{mech_energy75\%} + \text{mech_energy100\%})}{(\text{Data center ITE energy25\%} + \text{data center ITE energy50\%} + \text{data center ITE energy75\%} + \text{data center ITE energy100\%})}$$

Note that all units are in kWh (energy use).

32.3.3 ASHRAE 90.4 Electrical Loss Component (ELC)

ELC uses the same fundamental concept behind its development. Using the Standard 90.4 approach to calculate the ELC defines the electrical system efficiencies and losses. ASHRAE 90.4 defines three electrical system segments that are used to determine ELC:

- Incoming electrical service segment
- UPS segment
- ITE distribution segment

The segment for electrical distribution for mechanical equipment is stipulated to have losses that do not exceed 2% but is not included in the ELC calculations. All the values for equipment efficiency must be documented using the manufacturer's data, which must be based on standardized testing

using the design ITE load. The final submittal must consist of an electrical single-line diagram and plans showing areas served by electrical systems, all conditions and modes of operation used in determining the operating states of the electrical system, and the design ELC calculations demonstrating compliance.

ASHRAE 90.4 has tables listing the maximum ELC values for ITE loads less than 100 kW and greater than or equal to 100 kW, respectively. The table shows the maximum ELC for the three segments individually as well as the total.

The efficiency of the electrical distribution system impacts the data center's overall energy efficiency in two ways: lower efficiency means greater electricity use and a greater air-conditioning load to cool the electrical energy dissipated as heat. ASHRAE 90.4 is explicit on how this should be handled: "The system's UPS and transformer cooling loads must also be included in the MLC, evaluated at their corresponding part load efficiencies."

The standard includes an approach on how to evaluate single feed UPS systems (e.g., N, $N+1$, etc.) and active dual feed UPS systems ($2N$, $2N+1$, etc.). The single feed systems must be evaluated at 100 and 50% ITE load. The dual active feed systems must be evaluated at 50 and 25% ITE load because these types of systems will not normally operate at a load greater than 50%. An added benefit to this process is a demonstration of how reliability affects energy use.

ELC is determined from tables in 90.4 that indicate minimum efficiency values for the different segments of the electrical systems. If the electrical system complies with the values in the table, it complies with the required ELC.

32.4 APPLYING PUE IN THE REAL WORLD

So how can the metric PUE be used to analyze alternative power and cooling systems in order to arrive at a decision on what provides the best energy efficiency? It is vital to identify the interdependencies between the systems to generate a holistic understanding. Climate, cooling system type, power distribution topology, and redundancy level (reliability, availability) will drive the power efficiency of these systems. The interactions between these systems will become clearer after performing an analysis to determine peak power and annual energy use of the data center facility.

There are two fundamental approaches to defining power and cooling system energy use with respect to the entire annual energy use. Figure 32.1 (which is how PUE is correctly determined) depicts the energy use of the entire data center, including the ITE energy use. In this example, the HVAC consumes 11% of the total energy use of the data center. Inspecting Figure 32.2 we see that the energy consumption of the lighting, power, and cooling systems is shown without the ITE energy use. Here the mechanical system will consume over 50% of the total. As mentioned,

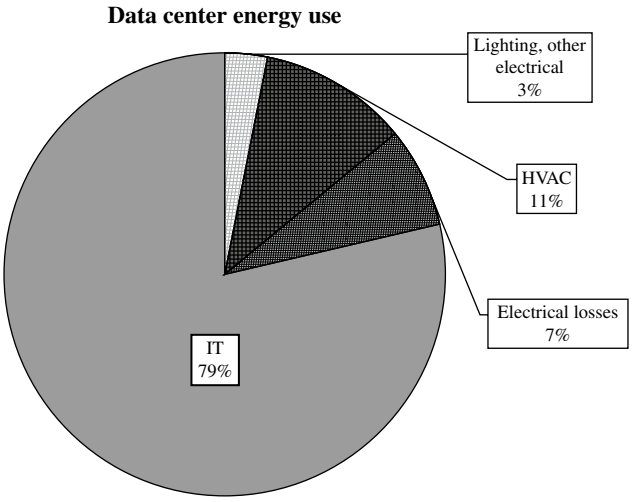


FIGURE 32.1 The energy use of the entire data center facility. Source: ©2020, Bill Kosik.

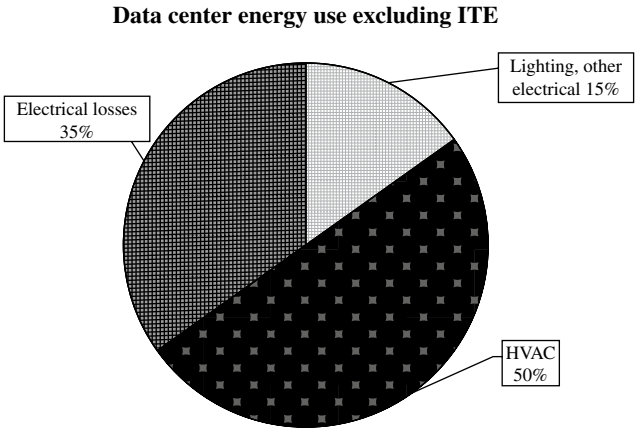


FIGURE 32.2 The energy use of the entire data center facility without including ITE. (For closer inspection of power and cooling systems, the ITE energy use can be left out of the analysis. Once the analysis is completed on the power and cooling systems, the ITE energy use can be added back to see the impact on PUE). Source: ©2020, Bill Kosik.

Figure 32.1 is the industry standard for reporting data center energy use, but organizations may have a need to illustrate the energy use in other ways, like Figure 32.2. The design, control, and operation of the power and cooling systems will have a significant influence on the annual energy use of the facility, and using reporting techniques, such as PUE, will help to identify reduction strategies.

When taking into consideration the climate, type of mechanical system, supply air temperatures, economization techniques, air distribution, etc., the mechanical system efficiency has the greatest variability. This is why PUE values should not be used to compare different data centers; there are simply too many variables that have an effect on power use. Depending on the reliability requirements of the data

center, there are a number of losses attributable to the power and cooling systems. Some examples of these elements include the following:

1. In some cases, the power that supports the mechanical systems must be transformed from a "primary customer" voltage (e.g., 4 kV) to a voltage that works with the cooling equipment (e.g., 600 V).
2. In high-reliability facilities, certain cooling equipment might be fed from UPS power such as CRAC fans or chilled water pumps to maintain operation when the mechanical system is transferring from normal power to generator power.
3. Mechanical cable distribution losses.
4. The electrical losses start at the utility and continue through the UPS, Power Distribution Units (PDUs), Remote Power Panels (RPPs), and ultimately to the IT equipment. This includes the power to the servers via static transfer switches (STS) and PDU.
5. UPS systems will typically have efficiencies of 85–96% depending on the UPS type, configuration, and the running capacity.
6. Distribution cable losses will run from 1 to 1.5%.

Since the largest power consumer in the mechanical system is the chiller (or other types of electrically driven, mechanical compression equipment), one primary strategy to decrease overall energy consumption is to elevate the supply air temperature by increasing the chilled water supply temperature and/or reducing the temperature of the air moving across the condensing coil. However, the ability to incorporate this strategy will completely depend on the type of mechanical system, the climate, and the allowable supply air temperature for the IT equipment. Consider that for fixed speed chillers, every 1°F increase in chilled water temperature can increase chiller energy efficiency by 1–2%. For variable speed drive (VSD) chillers, every 1°F increase in chilled water temperature can result in a 2–4% efficiency increase. Therefore, increasing the supply air temperature from 60 to 75°F will result in an average efficiency increase of the chiller of nearly 40%.

Care must be taken when attempting to analyze PUE. As an example, a facility may have a very efficient cooling system, but the electrical distribution system may have very high losses. The PUE might appear to be good, but without the data on the electrical system's performance, it is impossible to develop valid efficiency strategies. Similarly if a cooling system has an extremely efficient chiller but the chilled water pumps are consuming an inordinate amount of energy, granular data is required to uncover this type of anomaly. Finally PUE is not meant to compare the performance of different data centers. It is meant to assist in identifying areas of improvement in a single facility, to develop a personal best if you will.

32.5 METRICS USED IN DATA CENTER ASSESSMENTS

After data from the energy audit have been collected and analyzed, it is helpful to use the information as inputs for industry metrics that have been developed specifically to help data center owners benchmarking and judging the overall effectiveness of their data center. Many of these metrics are in a constant state of improvement so it is essential that the users of these programs provide feedback to the organizations issuing the metrics.

When performing assessments, it is vital to know if there are any specific criteria that must be adhered to when taking measurements. As an example, when using the Rack Cooling Index (RCI), the appropriate guidelines must be used to ensure accurate and consistent measurements. There are several types of metrics related to data center performance; not all these metrics must be used to determine the efficiency of the data center. When attempting to reduce energy use in a data center, the metrics that generate the most meaningful data beneficial to bettering the energy use of the facility are the most useful.

32.6 THE GREEN GRIDS XUE METRICS

Since the Green Grid released their white paper on PUE in October 2007, they have been developing a family of "xUE" metrics "designed to help the data center community better manage the energy, environmental, societal, and sustainability-compliance parameters associated with building, commissioning, operating, and decommissioning data centers." The following is a high-level overview of these metrics:

32.6.1 Energy Reuse Effectiveness (ERE)

This metric was developed to recognize that some data centers can provide energy useful for energy needs in other parts of the facility or campus. One of the more common ways this can be applied is by providing low grade heat obtained from the air exhausted from the servers (through a heat exchange process) to heat adjacent buildings or preheat domestic water before it enters an electric or natural-gas-powered water heater. The process is represented by the following:

$$\text{ERE} = \frac{(\text{Annual facility energy use} - \text{Annual energy reused})}{\text{Annaual IT energy use}}$$

32.6.2 Water Use Effectiveness (WUE)

The purpose of the metric is to determine the efficacy of water use in the data center based on the energy used by the IT equipment. The formula is as follows:

$$\text{WUE} = \frac{\text{Annual site water usage}}{\text{Annual IT energy use}}$$

Depending on the types of systems, this may end up as a zero-sum game, meaning as water use for the facility is reduced, the energy use for cooling systems may increase. This can happen when air-cooled direct expansion (DX) cooling equipment is used in lieu of water-cooled equipment. The amount of site water used will be much lower using the DX system, but due to the inherent inefficiencies in the air-cooled equipment, more site energy is used. The goal is to reduce overall energy use, so even if water-cooled HVAC equipment is used, the water consumption must be reduced. With more site energy being used, the source water—the water used at the electrical generation plant—will increase. This way the metric will yield a more complete picture of water that is consumed on a regional level, where it is most important. When the source water is taken into consideration, the formula changes:

$$\mathrm{WUE}_{\mathrm{source}} = \frac{\text{Annual site water usage} + \text{Annual source water usage}}{\text{Annual IT energy use}}$$

This involves using data on water use of the regional power plants that will most likely have to be an estimate since exact water consumption figures may not be available.

32.6.3 Carbon Use Effectiveness (CUE)

This metric illustrates the amount of carbon that is expended as compared with the annual IT energy used in the data center. Like the other metrics, the data center owner is encouraged to decrease the numerator, thereby making the CUE smaller. The CUE is defined as follows:

$$\mathrm{CUE} = \frac{\left(\text{Annual } CO_2 \text{ emissions caused by the data center}\right)}{\text{Annual IT energy use}}$$

This is a source-driven metric since the CO_2 emissions come from the power plants that feed the electric grid that supplies electricity to the data center. Like the water consumption figures used in calculating the WUE, the CO_2 emissions come from data issued by the U.S. EPA in the United States and other agencies internationally. Once the annual data center energy in kWh is determined, it is a simple calculation to determine the annual CO_2 emissions. Like WUE, these metrics are especially useful when vetting sites for building a green field data center.

32.7 RCI AND RTI

RCI (Rack Cooling Index) is a dimensionless factor developed by Dr. Magnus Herrlin.

RTI (Return Temperature Index), also developed by Dr. Magnus Herrlin, is a metric to determine the efficacy of the air management in a data center.

32.7.1 Rack Cooling Index (RCI)

RCI is a metric to determine the effectiveness of the air management in the data center. Using Computational Fluid Dynamics (CFD) modeling or measurements in an existing data center, the mean inlet temperature is used in conjunction with the maximum allowable and maximum recommended temperature to develop a percentage effectiveness where 100% means that the mean inlet temperature exactly meets the requirements. There are two metrics—$\mathrm{RCI}_{\mathrm{HIGH}}$ and $\mathrm{RCI}_{\mathrm{LOW}}$—to reflect that the temperatures of the IT cabinet at the upper and lower levels will vary. RCI is defined as follows:

$$\mathrm{RCI}_{\mathrm{LOW}} = \left(1 - \frac{\text{Total under-temperature}}{\text{Max allowable under-temperature}}\right) \times 100\%$$

$$\mathrm{RCI}_{\mathrm{HIGH}} = \left(1 - \frac{\text{Total over-temperature}}{\text{Max allowable over-temperature}}\right) \times 100\%$$

Based on the standard being used, the numerical value of these indices will vary. ASHRAE recommends that rack air intake temperatures are between 18 and 27°C (64.4 and 80.6°F). The best practice method to monitor temperature in a data center dictates three temperature sensors installed at the top, middle, and bottom of every third rack.

32.7.2 Return Temperature Index (RTI)

Similar to RCI, RTI judges how effective the air distribution system is at isolating the cold air meant for the computer equipment from the hot air that is expelled from the equipment. When the rack inlet temperatures are equal to the supply and return air temperatures, the RTI will be 100%, meaning there is no mixing and the supply air that comes from the air handling system is the same temperature as what is delivered to the IT equipment. And, when the air that is being returned to the air handling system is equal to the air temperature at the discharge of the IT equipment, there must be no mixing:

$$\mathrm{RTI} = \left(\frac{\left(\text{Return air temperature} - \text{Supply air temperature}\right)}{\left(\text{Rack outlet mean temperature} - \text{Rack inlet mean temperature}\right)}\right) \times 100$$

32.8 ADDITIONAL INDUSTRY METRICS AND STANDARDS

There are several other metrics that focus on the IT equipment performance and energy consumption. Using these metrics to drive change in the reduction of data center energy use is vital to the overall energy efficiency strategy in a data center.

32.8.1 SPEC

According to the Website, "The Standard Performance Evaluation Corporation (SPEC) is a non-profit corporation formed to establish, maintain and endorse a standardized set of relevant benchmarks that can be applied to the newest generation of high-performance computers. SPEC develops benchmark suites and reviews and publishes submitted results from member organizations and other benchmark licensees." The specific standard, SPECpower_ssj®2008, is a benchmark that evaluates the power and performance characteristics of volume server class and multi-node class computers, providing a method to measure power in conjunction with a performance metric. These data provide useful information to design engineers and facility operators in the form of actual measured power and performance of enterprise servers, helping to develop better estimates on potential future IT loads. When analyzing the performance of the computers tested using the SPECpower software (Fig. 32.3), it is immediately clear since 2007 there has been a consistent improvement in computational ability, measured in performance/watt.

32.8.2 TOP500/Green500™

Started in 1993, the TOP500 is a list of 500 most powerful computer systems worldwide. The mission of the TOP500 organization is to develop and maintain a ranked list of general purpose systems that are being used for high-end applications. The authors of the TOP500 use the LINear equations software PACKage (LINPACK) to measure the time required to solve complex mathematical problems. The results of the testing use the nomenclature floating point operations per second (FLOPS) or FLOP/s (top straight line) (Fig. 32.4 FLOP/s (Top straight line)).

During the period where the TOP500 was building momentum and improving the accuracy and quality of their reporting on computing performance, the high-performance computing (HPC) community was realizing that a coordinated effort was needed to raise the awareness on the power consumption of HPC systems. Launched in 2006 the Green500 has a mission to promote power-aware supercomputing and track the fastest systems in the world from a performance-per-watt perspective. The Green500 ranks the most energy-efficient supercomputers in the world according to the metric megaflops per watt (Mflops/W) (Mflops/watt—Lower curve line) (Fig. 32.4). Since then, in addition to computational ability, the computer's performance per unit power has become a part of the computer's overall capability. The testing can be done by the end user and must follow the Green500's testing protocol, *Power Measurement of Supercomputers*. The results are tabulated and released approximately every 6 months.

32.8.3 ASHRAE 90.4

Commercial buildings in the United States and abroad are designed using standards, guidelines, and codes that include requirements for energy efficiency. ASHRAE 90.1 is an indispensable and widely referenced standard for this, used for design, documentation, and compliance. It has proven to be very effective in reducing energy use in buildings dating back to 1976. But with the proliferation of data centers that consume unprecedented amounts of energy, it became clear that Standard 90.1 could not address certain aspects of data center design vis-à-vis protocols for system type, reliability, electrical system efficiency, etc.

For example, ASHRAE 90.1 describes a process on how to model a computer room, but this language is intended for

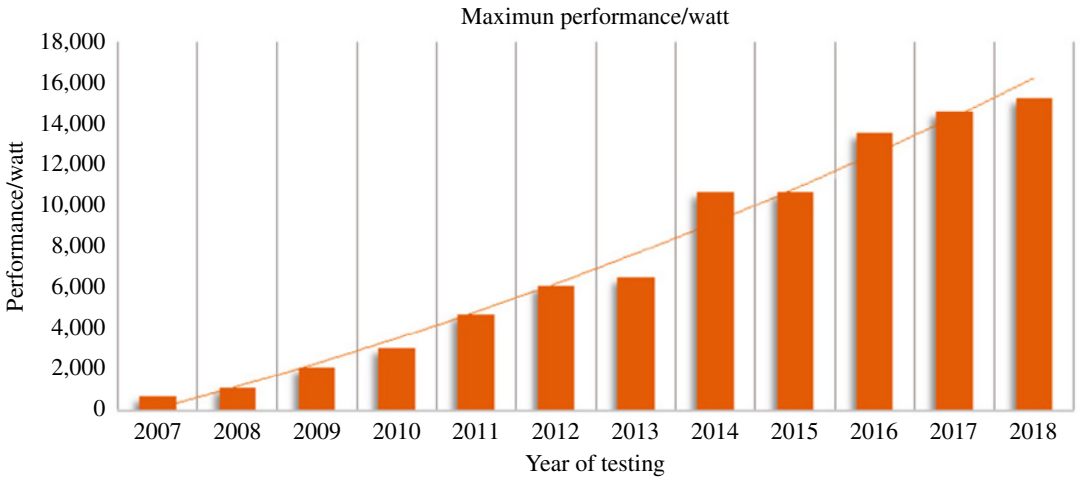


FIGURE 32.3 Based on the testing by the Standard Performance Evaluation Corporation (SPEC), in the past decade, the performance per watt for enterprise servers has increased an average of 36% year over year. Even though the server average power has remained flat over this time period, the performance has increased. Source: ©2020, Bill Kosik.

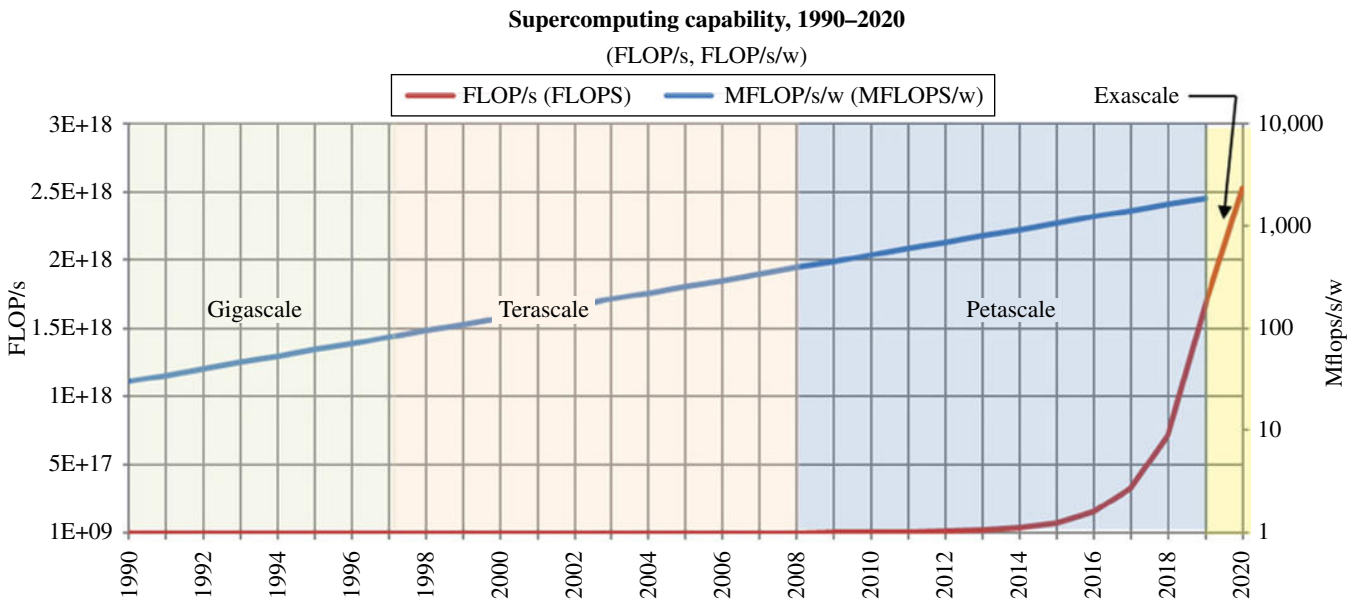


FIGURE 32.4 Supercomputing continues to produce faster machines with a modest growth in power requirements. Starting in 2014, the industry started seeing significant increases in performance (FLOP/s) and performance/power (FLOP/s/W). This is indicative of gains in energy efficiency of the machines and advances in computer technology. Source: ©2020, Bill Kosik.

a small data processing room within a larger corporate building, which is very different from a stand-alone data center with a small amount of administrative space. This is just one example of many. As data centers became larger and more sophisticated, the importance of developing a new industry-backed standard, formed using a consensus-driven process, became clear. Given its widespread use and long history, a logical conclusion was reached to use ASHRAE 90.1 as a framework for a new data-center-specific energy efficiency standard. ASHRAE 90.1 has another key advantage: municipalities can adopt and codify it for inclusion in their local building code.

ASHRAE released drafts of a new data-center-specific energy efficiency standard. It is called *ASHRAE 90.4-2016, Energy Standard for Data Centers*. After a year of public review and comment, it was formally released as a new standard.

ASHRAE 90.1-2016 is the normative reference to ASHRAE 90.4, creating a system that avoids doubling up on future revisions to the standard, minimizes any unintended redundancies, and ensures that the focus of ASHRAE 90.4 remains solely on data center facilities. Also, issuing updates to ASHRAE 90.1 will automatically update ASHRAE 90.4 for the referenced sections. In the same way, updates to ASHRAE 90.4 will not affect the language in ASHRAE 90.1.

Because many local jurisdictions operate on a 3-year cycle for updating their building codes, many are still using ASHRAE 90.1-2013 or earlier. The normative reference is ASHRAE 90.1-2016; however, the final say on an administrative matter like this will always fall to the authority having jurisdiction (AHJ).

Using ASHRAE 90.4 in combination with Standard 90.1 gives the engineer a completely new method for determining energy consumption compliance for a data center facility. In Standard 90.4, ASHRAE introduces new terminology for demonstrating compliance: design and annual MLC and ELC. ASHRAE is careful to note that these values are not to be confused with PUE or pPUE and are to be used only within the context of ASHRAE 90.4.

The standard includes compliance tables consisting of the maximum load components for each of the 19 ASHRAE climate zones. Assigning an energy efficiency target, either in the form of design or an annualized MLC to a specific climate zone, reinforces the inextricable link between climate and data center energy performance. Design strategies, such as elevated temperatures in the data center and air or water economization, are heavily dependent on the climate; when analyzing efficiency across the climate zones, the data will feed into the decision-making process to determine the location of the data center. While using the data to determine a location for the data center is not part of the process for determining compliance, it is valuable for the progress of the overall project.

ASHRAE 90.4 uses a performance-based approach that allows the engineer to develop energy-reduction strategies using computer-based modeling rather than taking a prescriptive approach. Not only does this result in more accurate comparisons, but also it accommodates for next-generation energy-efficient cooling solutions (and for the rapid change in computer technology).

Many of the provisions contained in 90.4 encourage innovative solutions:

- *On-site renewables or recovered energy*. The standard allows for a credit to the annual energy use if on-site renewable energy generation is used or waste heat is recovered for other uses.
- *Derivation of MLC values*. The MLC values in the tables in ASHRAE 90.4 are generic to allow multiple systems to qualify for the path. The MLC values are based on equipment currently available in the marketplace from multiple manufacturers that meets the minimum efficiency requirements using the AHRI (Air Conditioning, Heating, and Refrigeration Institute) Product Performance Certification Program.
- *Design conditions*. The HVAC system design is not bound to standard parameters like a delta *T* (temperature rise of the supply air) of 20°F and a return air temperature of 85°F. This gives the engineer lots of room to innovate and propose nontraditional designs, such as water cooling of the ITE equipment.
- *Trade-off method*. Sometimes mechanical and electrical systems have constraints that may disqualify them from meeting the MLC and ELC values on their own merit. The standard allows, for example, using a less efficient mechanical system offsetting it with a more efficient electrical system (and vice versa).

32.9 EUROPEAN COMMISSION CODE OF CONDUCT

In 2008, the European Code of Conduct for Data Centers was launched with the ambitious goal of reducing data center energy use, currently predicted to reach 100 TWh by year 2020. Managed by the JRC (Joint Research Centre), which has set ambitious standards, the code of conduct is a voluntary initiative identifying key issues and agreed-upon solutions, described in the Best Practices document. In addition to the participants, companies (vendors, consultants, industry associations) can also promote the code of conduct to their clients.

32.10 CONCLUSION

It is important to understand that these metrics should be used together providing a range of data points to help understand the efficiency and effectiveness of a data center; different combinations of these metrics will produce a synergistic outcome. As an example, when PUE is used in conjunction with WUE, it is possible to see how the values interrelate with each other and why it is good to look at the corresponding water consumption when different energy efficiency strategies are contemplated. Similarly, when analyzing different cities for a new data center build, using PUE and CUE will result in data that is influenced by the type and efficiency of the local power generation and how the climate affects the cooling system performance. Using these metrics alone or in strategic combinations brings great value to analyzing energy use in the data center.

FURTHER READING

ANSI/AHRI 1360 (I-P)-2017. Performance Rating of Computer and Data Processing Room Air Conditioners. Available at https://webstore.ansi.org/standards/ari/ahri13602017. Accessed on August 4, 2020.

ASHRAE Guideline 14-2014. Measurement of Energy and Demand Savings. Available at: https://www.ashrae.org/technical-resources/bookstore/supplemental-files/supplemental-files-for-ashrae-guideline-14-2014. Accessed on August 4, 2020.

ASHRAE Real-Time Energy Consumption Measurements in Data Centers. Available at https://www.ashrae.org/technical-resources/bookstore/datacom-series. Accessed on August 4, 2020.

ASHRAE Standard 90.1-2019 (I-P Edition). Energy Standard for Buildings Except Low-Rise Residential Buildings. Available at https://www.ashrae.org/technical-resources/bookstore/standard-90-1. Accessed on August 4, 2020.

ASHRAE TC9.9: Data Center Power Equipment Thermal Guidelines and Best Practices, ASHRAE Technical Committee.

Avelar V, Azevedo D., French A, (eds.). PUE™: A Comprehensive Examination of the Metric. Available at https://datacenters.lbl.gov/sites/all/files/WP49-PUE%20A%20Comprehensive%20Examination%20of%20the%20Metric_v6.pdf. Accessed on August 4, 2020.

Avelar V, Azevedo D, French A, (eds.). The Green Grid: PUE™: A Comprehensive Examination of the Metric. Available at https://www.thegreengrid.org/en/resources/library-and-tools/237-PUE%3A-A-Comprehensive-Examination-of-the-Metric. Accessed on August 4, 2020.

BREEAM International New Construction 2016. Data Center Annex – Pilot v0.0; Criteria Annex – SD233 – Annex 2; 2016. Available at https://www.breeam.com/wp-content/uploads/sites/3/2019/11/SD233-Annex-2-Data-Centre-Criteria-Annex.pdf. Accessed on August 4, 2020.

Koomey JG. Growth in Data Center Electricity Use 2005 To 2010. Available at https://www.koomey.com/research.html. Accessed on August 4, 2020.

Lawrence Berkeley Lab High-Performance Buildings for High-Tech Industries, Data Centers. Available at https://hightech.lbl.gov/. Accessed on August 4, 2020.

Shehabi A., Koomey JG, et al. United States Data Center Energy Usage Report, Lawrence Berkeley National Laboratory; 2016.

Singapore Standard SS 564: 2020 Sustainable Data Centres. Available at https://www.imda.gov.sg/regulations-and-licensing-listing/ict-standards-and-quality-of-service/

IT-Standards-and-Frameworks/Green-Data-Centre-Standard. Accessed on August 4, 2020.

(TC) 9.9 Mission Critical Facilities, Data Centers, Technology Spaces, and Electronic Equipment, ASHRAE 2016.

The Green Grid: Carbon Usage Effectiveness (CUE): A Green Grid Data Center Sustainability Metric. Available at https://www.thegreengrid.org/en/resources/library-and-tools/241-Carbon-Usage-Effectiveness-%28CUE%29%3A-A-Green-Grid-Data-Center-Sustainability-Metric. Accessed on August 4, 2020.

The Green Grid Data Center Power Efficiency Metrics: PUE and DCIE. Available at https://www.missioncriticalmagazine.com/ext/resources/MC/Home/Files/PDFs/TGG_Data_Center_Power_Efficiency_Metrics_PUE_and_DCiE.pdf. Accessed on August 4, 2020.

The Green Grid: ERE: A Metric for Measuring the Benefit of Reuse Energy from a Data Center. Available at https://www.thegreengrid.org/en/resources/library-and-tools/242-ERE%3A-A-Metric-for-Measuring-the-Benefit-of-Reuse-Energy-From-a-Data-Center-. Accessed on August 4, 2020.

The Green Grid: Recommendations for Measuring and Reporting Overall Data Center Efficiency Version 2 – Measuring PUE for Data Centers. Available at https://www.energystar.gov/ia/partners/prod_development/downloads/Data_Center_Metrics_Task_Force_Recommendations_V2.pdf?7438-21e8. Accessed on August 4, 2020.

The Green Grid: Water Usage Effectiveness (WUE™): A Green Grid Data Center Sustainability Metric.

US Green Building Council – LEED Rating System.

33

DATA CENTER INFRASTRUCTURE MANAGEMENT

Dongmei Huang
Beijing Rainspur Technologies, Beijing, China

33.1 WHAT IS DATA CENTER INFRASTRUCTURE MANAGEMENT

The data center industry is awash with change. Since the days of the dot-com era, the data center has been massaged, squeezed, stagnated, and reconstituted more than once with the purpose of cost reductions, increased capacity, compliance and control, and overall efficiency improvements. In a 2019 IT report published by Gartner "IT Key Metrics Data" majority IT budgets are spent on data center infrastructure and its operations, and surprisingly, not many companies have invested in the tools, technologies, and discipline needed to actively manage these huge capital investments.

Data center infrastructure management (DCIM) is a critical management solution for data centers that has been used in data center industry and proven successfully. The origin of the term DCIM is not clear nor is the exact definition of DCIM universally agreed at the moment. That said, the initial spirit of DCIM can be summarized much in the way Gartner has expressed it, "The integration of information technology and facility management disciplines to centralize monitoring, management and intelligent capacity planning of a data center's critical systems. Additionally, DCIM is achieved through the implementation of specialized software, hardware and sensors. DCIM will enable a common, real-time monitoring and management platform for all independent systems across IT and facilities and must manage workflows across all systems."

DCIM has transitioned well beyond simple monitoring, drawing pretty pictures, and interactive eye candy and has become the data center extension to a number of other systems, including asset and service management, financial general ledgers, etc. At the end of the day, a well-deployed DCIM solution quantifies the costs associated with moves, adds, and changes on data center floor; understands the cost and availability to operate those assets; and clearly identifies the value derived through the existence of that asset over its useful life span. And true to the original spirit of DCIM mentioned above, these business management views span the IT and facilities worlds.

Taking a closer look at Figure 33.1, you can see that DCIM is in direct support of modern approaches to data center asset and service management. Combining two well-known models from Forrester and the 451 Group, you can see how DCIM provides the view of the data center from the physical layer upward, whereas most IT management umbrellas in use currently are limited to a top-down logical view. According to a recent IDC report, 57% of data center managers consider their data centers to be inefficient, and 84% of those surveyed have issues with space, power, or cooling that is directly affecting the bottom line. Clearly, these models must converge into a single management domain with combined views.

33.1.1 DCIM Maturity: The Technology and the User

Any new technology typically emerges after a long journey. Taking any of the recent data center examples like virtualization or cloud computing, it can be seen that there are several distinct periods that must be traversed before any technology is deployed in standard production. Gartner refers to this flow as the "hype cycle" and is shown in Figure 33.2. What starts as an amazing innovative idea with all the promise in the world gets tested and retested over the subsequent years with a dose of reality thrown in for good measure. Technologists

Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center, Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

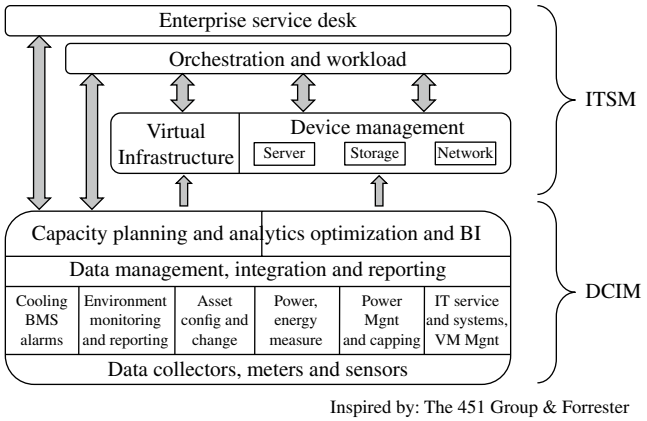


FIGURE 33.1 DCIM has been broadly defined as the layer of infrastructure that supports the enterprise IT function.

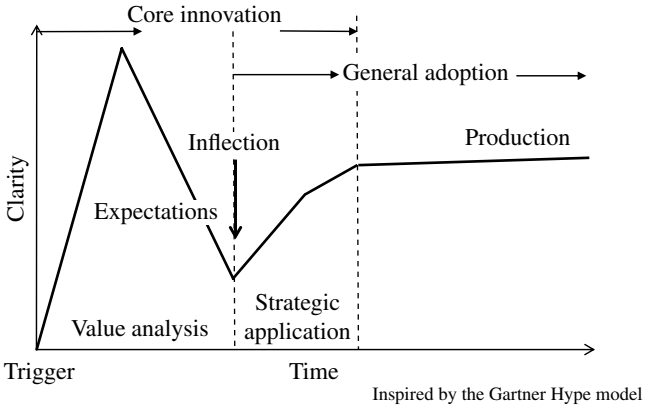


FIGURE 33.2 "Hype cycle"—The maturation of any new technology follows a very predictable adoption cycle. Wild enthusiasm is replaced by reality with the most sound ideas becoming production.

and business managers alike poke and prod at new inventions to determine how it could pertain to their own use cases. Over time some inventions vanish for various reasons, and others emerge with general adoption growth rates over time.

DCIM has been following this same curve. Referring to Figure 33.3 you can see that various organizations are currently at different stages in their ability to think about their computing needs in the future and are being challenged to self-evaluate their own IT best practices of the present. After years and years of ad hoc asset management solutions for the data center and the typical unique change processes that have abounded throughout the data center industry, we find a set of points along the maturity continuum, including integration with computational fluid dynamics (CFD) for what if analysis, that characterizes the industry's current status.

33.1.2 DCIM Is Strategically Important to the Modern Data Center

DCIM is a resource and capacity planning business management solution for the data center. It enables the data center

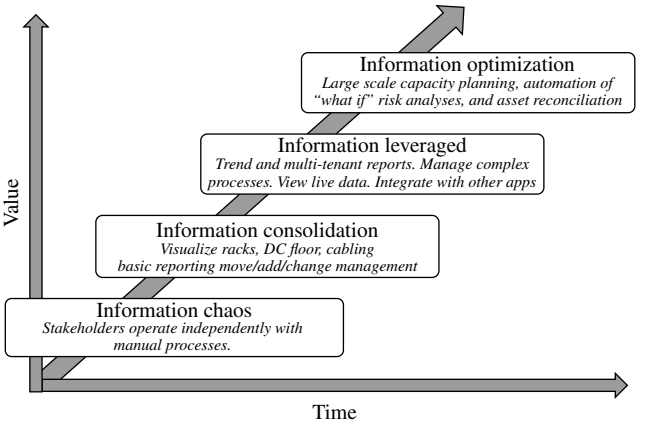


FIGURE 33.3 Existing data centers have management processes that vary wildly in their maturity. With DCIM, each can be optimized.

to leverage existing physical layer technologies including monitoring, capacity planning, configuration data bases, environmental sensing, etc., and it also enables the seamless integration into an enterprise's other business management solutions used for asset management, process management, data management, HR (human resource) planning, budgetary planning, SOX compliance, etc. The Sarbanes–Oxley (SOX) Act of 2002 is a U.S. legislation that requires company's IT in compliance of record keeping practices to support in the event of an audit.

In a twenty-first-century company, information is the most strategic asset and competitive differentiator. The data center structure itself is the factory floor that produces that strategic value. DCIM is the manager of the data center floor in which hundreds of millions of dollars of assets and billions of dollars of information flow in nearly every fortune 500 enterprise. DCIM is the business management solution for this critical infrastructure.

There are at least four stakeholders emerging as having keen interest in DCIM, and each has their own sets of needs from the adoption of DCIM: (1) the IT organization, (2) the facilities organization, (3) the operations' finance department, and (4) the corporate social responsibility individuals.

In general, IT is becoming viewed as a single entity that has a quantifiable value to the organization and associated benefits. The IT and facilities organizations are now being tasked with a common set of goals regarding the data center and in response finding themselves required to behave as a single business unit, with transparency, oversight and accountability, forecasting, and overall effectiveness all in focus. DCIM enables that focus and has become essential for the data center community.

Andy Lawrence at the 451 Group put it most succinctly, "We believe it is difficult to achieve the more advanced levels of datacenter maturity, or of datacenter effectiveness generally, without extensive use of DCIM software. The three main drivers of investment in DCIM software are economics

(mainly through energy-related savings), improved availability, and improved manageability and flexibility."

As the IT industry transformation continues, we see a handful of corporate IT goals that are greatly influenced, enabled, and supported by a well-conceived DCIM strategy and that put DCIM squarely in the same level of importance as any of the other deployed business management applications:

1. IT transformation including the cloud
2. Actively managing power needs and focus on green IT
3. Asset migrations, management, and capacity planning
4. Operational excellence
5. Lower labor costs
6. Asset tracking
7. Operational data acquisition and analysis
8. Deferring new capital expenditures and in specific data center builds
9. Regulatory and legislation compliance

33.1.3 Common Goals for DCIM

DCIM has become a general purpose "efficiency" category where a wide range of stakeholders have voiced their data center management needs. Some of the most common goals for the introduction of DCIM into operational and strategic plans include the following:

1. Energy management has become the first priority for most data center managers and IT business managers alike. Their goal is to reduce operating costs across the board, with the starting point being much more proactive energy management. DCIM solutions can provide a highly granular understanding about energy usage and a wide range of other physical layer metrics and ultimately help to identify and control inefficiencies.
2. The need for highly accurate and actionable data and views about the current capacity and availability of their data centers and easy access to baselines associated with current operations.
3. Operational best practices are being redefined to accommodate much more streamlined operations and remediation. DCIM is being viewed as the tool best suited to identify, present, and manage the workflow associated with data center physical assets. DCIM essentially captures and enforces the processes associated with change.
4. With such a wide range of traditional tools in use, even within the same organization, there exists a very loose and disconnected source of truth about the data center assets. DCIM promises to coordinate and consolidate disparate sources into a single view of asset knowledge.
5. Resource availability and capacity planning are high on the list of needs. Better predictability for space, power, and cooling capacity means increased useful asset life spans and increased time frames to react to future shortfalls.
6. Keen insight allows large amount of raw data to be transformed into business intelligence. Enhanced understanding of the present and future states of the data center allows for better asset utilization and increased availability.
7. Corporate responsibility goals to assure the latest innovations in IT management are being considered and investigated. Business cases are being created, and as a standard practice major technological advances are not being missed.

33.1.4 Whose Job Is DCIM? IT or Facilities

One of the most interesting aspects of the adoption of DCIM is the audience diversity and their individual driving factors. Traditionally, data centers were built, maintained, and utilized by different distinct organizations: (i) the facilities organization that took care of all space, power, and cooling requirements for any piece of real estate and (ii) the IT organization that took care of the data center physical and logical build and all of the equipment life cycle itself.

IT and facilities now find themselves required to work together for planning and optimization processes. Decisions about equipment and placement are now being made jointly. In many companies, both organizations now report to the CIO (chief information officer). DCIM enables the data center resource capacity planning to be managed over long periods of time. That said, DCIM and all of its capabilities will (in most cases) be driven by the IT organization just as it has for every other types of data center management (systems management, network management, etc.) over the years. Sure there will be benefits across all groups once DCIM is in production, but the IT organization tends to have the largest role and experience in enterprise-class software selection, so it is expected that the deployment and organizational integration of DCIM is best leveraged as an extension to existing software management frameworks driving data center logical operations already.

33.2 TRIGGERS FOR DCIM ACQUISITION AND DEPLOYMENT

While the term "DCIM" has only been used in the vernacular for the past few years, the concept of asset management has been around since the inception of the data center. Traditional approaches to data center asset management were fairly straightforward extensions of the financial bookkeeping tools in use. Earlier asset management methodology simply built upon the accounting systems of the period through the addition of physical attributes and organizational ownerships. In a

TABLE 33.1 Reasons to deploy DCIM

Track asset visibility and management
Increase asset utilization
Enhance change and workflow management
Identify space availability
Improve capacity management
Reduce response time
Improve people productivity
Ensure uptime and availability
Reduce data center costs
Increase energy efficiency
Improve data center customer service
Monitor and ensure safe working environment
Enhance security
Improve compliance
Disaster prevention

Sources: Raritan, Sunbird, Uptime Institute, VXChage

few cases and with a dedicated desire by a handful of IT professionals to innovate, a bit of rack and floor visualization was added to that information. As these minor extensions provided little new business value, the adoption of asset management solutions that embrace energy, visualization, and life cycle capabilities languished. These pre-"DCIM"-type solutions remained a curiosity, a nice to have set of features, rather than a must-have business need.

There are many reasons to deploy DCIM as shown in Table 33.1. Uptime Institute did a survey in 2018 (https://uptimeinstitute.com/uptime_assets/f7bb01a900c060cc9abe42bb084609f63f02e448f5df1ca7ba7fdebb746cd1c4-2018-data-center-industry-survey.pdf) regarding to the adoption of DCIM: "The most common motivation for deploying DCIM was capacity planning (76%) and power monitoring (74%). Other reasons ranged from giving executives and/or customers (of multi-tenant data centers) visibility or reports (52% of respondents) to compliance (35%)." Data center consolidation and migration as well as newly built data centers often cause a DCIM purchase.

33.2.1 Capacity Management Including Power, Cooling, and Floor Space

Power was first major DCIM trigger on everyone's list. The rising cost of power was being seen in all aspects of life, both residential and commercial. Individuals saw the price of gasoline and electricity rise; corporations saw their huge power bills become larger. IT is typically the largest single line item in a corporation, so the abnormal rise in these highly visible costs caused a stir. The CEO (chief executive officer) and CFO (chief financial officer) leaders began asking questions about costs that the CIO and their teams were unable to answer. Power is one of the first quantifiable values that is directly associated with a successful implementation of DCIM.

33.2.2 Business Process Reengineering and Operational Efficiency

The operation of a data center is quickly transforming from individual and disconnected tactical activities with a historical primary goal of "high service levels at any cost" to a planned and predictable approach with the modified metric "service at what cost." Essentially, the cost factor is being added to the equation and being tested at every step of the way. IT organizations are being asked not only to document and then automate their existing practices but also to actually consider their current approaches and determine if they are still valid and optimal. As such, a number of organizations are finding themselves with limited awareness of their existing practices that is impeding their ability to create streamlined new approaches. As baselines are created for existing conditions, IT organizations will begin to author new optimized workflows and deploy new technologies such as DCIM to manage their assets over long periods of time. DCIM promises to be able to capture current business practices and allow optimizations in workflow and labor-related efficiency to be realized.

33.2.3 Data Center Consolidation Projects

Data center consolidation is reality today for most corporations for a variety of reasons. Advances in computing technology, mergers, and acquisitions, DCIM supports the commissioning and decommissioning of vast amounts of computing equipment typically found in data center consolidation projects.

33.2.4 New Capacity and New Data Centers

Many organizations are realizing that their core data center assets are either past their useful life span or are simply not able to support their organization's rapidly increasing demands for processing. The acceleration in the adoption of new business applications was never imagined to be at the current rate.

DCIM promises to address the quantification of current data center capacity and optimization, with a keen eye on capacity management over time. The data center itself provides computing resources, and when a large sample of

time-based usage data is studied in combination with the demands associated with new corporate initiatives, highly accurate data center planning is not only possible but also expected. DCIM quantifies data center capacity and allows it to be planned.

33.2.5 Data Center Cost Reductions and Enhanced Resource Efficiency

With the era of "green computing" came the primary goal of reducing waste. Specifically focused on energy overhead, it has become quite popular to focus the majority of data center optimization projects on their ability to allow the data center to operate at a lower cost per unit of work. Green IT has been used as a catch-all phrase to describe the more efficient usage of power.

33.2.6 Technology Refresh and Architectural Changes

A good number of data centers find themselves with large-scale technology refresh projects. These projects stem from the desire for higher density computing, virtualization, VDI initiatives (virtual desktop infrastructure), or mobility. Entire infrastructures are being redesigned, and when faced with this level of change, IT professionals find themselves looking for innovative ways to manage these new designs more effectively than previously practiced.

33.2.7 Environment and Sustainability Focus

There is a great deal of interest by most major corporations in reducing the impact of IT to the environment. Many organizations use three key metrics proposed by The Green Grid and covered elsewhere in this chapter to describe their efforts toward environmental friendliness: PUE (power usage effectiveness) relating to overall efficiency in the data center; CUE (carbon usage effectiveness), which refers to the carbon footprint associated with energy consumption; and most recently a metric associated with water, WUE (water usage effectiveness), which represents the amount of water consumed in the production of data.

33.2.8 Regulatory and Compliance, Audit, and Documentation

The executive teams within major corporations across the globe have found themselves under new levels of scrutiny regarding IT. IT as the most critical corporate asset is involved in every major company-wide function. The impact of IT has become so great and pervasive that various government and regulatory agencies are striving to provide oversight to assure that data is maintained accurately and that the environmental impact of the data center is considered.

The U.S. government's data center optimization initiative (DCOI) enables CIOs to rapidly become compliant with the U.S. Office of Management and Budget's Federal IT Acquisition Reform Act (FITARA). This FedRAMP compliant dedicated hosting environment will help federal agencies execute their "cloud-first" IT modernization priorities while consolidating and optimizing their facilities, all with the security the U.S. federal government requires.

DCIM becomes a means toward this end. DCIM allows a data center to be documented as a single system, with the intricacies of its components identified and understood. The efficiency of the operation of each component can be seen and over time optimized.

33.2.9 The Cloud

DCIM fits everywhere! Public and private clouds share a common set of characteristics: self-service, quick provisioning and accounting. For a public cloud provider, scalable DCIM solutions are required to help quickly manage assets and dynamically tune supply and demand. A well-conceived DCIM solution is essential for the public cloud providers to understand all capacities (across IT and Facilities) and thus allow quick-turn for remediation, for provisioning, and for decommissioning. DCIM enables the data center to run as a business with all of these costs clearly quantified and optimized. DCIM allows public clouds to exist, be more responsive, more accuracy in their operations, and reduce the overhead required to provide their customers' required levels of service.

Private clouds are just traditional IT infrastructures that have been transformed using the principles pioneered in the public cloud world. DCIM solutions are proving to be one of the most significant enabling technologies for this IT infrastructure reengineering. DCIM will allow this private cloud transformation. Remember that DCIM is all about enabling the data center to be managed like a business. Access to all of the business metrics, costs structures, services, etc. and dynamic management of assets, a comprehensive DCIM solution is essential for the transformation of traditional IT infrastructures to a highly tuned, optimized private cloud.

33.3 WHAT ARE MODULES OF A DCIM SOLUTION

Today's DCIM solutions include all of the necessary functions to allow a fully functional production data center to be streamlined and support all of the required material provisioning, optimization, remediation, and documentation over time. Comprehensive DCIM suites are usually created as a range of functional modules that are intended to work together seamlessly. These modules offer various means to gather static and dynamic data, store this large amount of time-specific data, correlate the associated data, and then pre-

sent and leverage this wealth of data in increasingly meaningful ways. When these tightly integrated modules are driven from a single data repository, the resulting DCIM solution allows highly impactful business decisions to be made.

33.3.1 Asset Life Cycle Requirements and Change Management

DCIM enables life cycle management of the data center and all of its assets. It addresses the physical layer of the data center and includes the same change management and workflow capabilities found in other ERP (Enterprise Resource Planning)-class business management solutions found in the typical enterprise. DCIM is not a monitoring utility. DCIM is an enabler to manage change with a keen eye on the cost structures associated with this change.

Over the course of these years, it is estimated that at least 25–30% of the assets contained within any given data center change each year. Technology refresh cycles due to costs, adoption of dense computing and virtualization, new networking, or storage technologies all account for huge amounts of change.

As you can see in Figure 33.4, changing a single server seems relatively simple, imagine multiplying that effort by a 1,000 or 10,000 times each month! It is staggering. DCIM is the business management platform that keeps all of these add/move/change cycles in order, documents the process at each step, and identifies the tasks needed to be completed in extreme detail to reduce human errors incurred during the execution of these tasks.

33.3.2 Capacity Planning, Analytics and Forecasting

Of specific note to the DCIM opportunity discussion is its ability to consider the data center as a system, with a very specific set of metrics and capacity over time. Data centers have physical attributes and associated limitations. Whether it is space, power, or cooling, each data center has a physical set of limitations that define the limits of a data center's capacity. DCIM has already been shown to be the best way to look at these factors together and then evaluate resource utilization over time. With this ability to consider all resources over time, predictions can be made about when one or more of these critical resources will be exhausted and at what cost it will take to bring new resources online.

The most successful DCIM offerings understand that visibility into the future is extremely valuable. It's quite easy to focus on historical data and present it in various forms, but interpreting historical data and using it to trend into the future is where mature DCIM offerings shine. Worth noting is the recent IDC findings that almost one-third of all data centers are forced to delay the introduction of new business services and more than a quarter of those data centers needed to spend unplanned OpEx budget to maintain a poorly defined data center structure. These unrealized opportunity costs can be huge!

The DCIM model includes a highly granular representation of the data center, which enables it to identify where resources (power, space, cooling, and connectivity) exist and where they are being used. Over the years, many data centers have lost resources due to their inability to identify its exact location. Terms such as "stranded capacity" and "vertical white space" come into discuss when these conditions occur. Essentially the originally designed resources become fragmented and therefore cannot be effectively utilized, or in other cases the availability of one resource is not co-resident with similar capacity of another resource. A great example is a data center that wishes to deploy high density blade chassis systems in an area with plenty of power but limited cooling. That power essentially becomes "stranded." The same types of imbalances occur across all of the data center resources. Modern DCIM solutions help by identifying when resources exist and allow balancing to recapture these resources. In some cases this repositioning of equipment to better balance all available resources may add two or more years of useful life to existing data center structures.

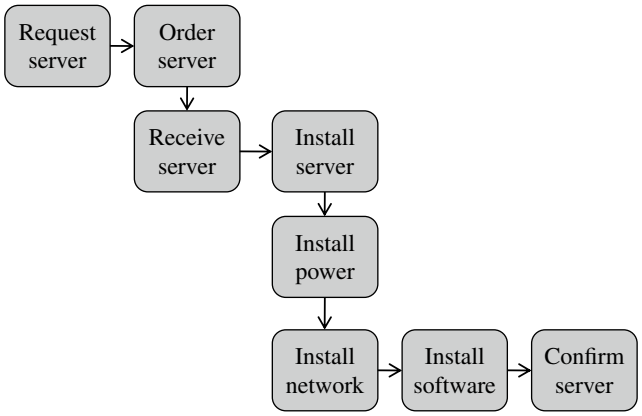


FIGURE 33.4 Leading DCIM suites offer comprehensive workflow capabilities that capture a data center's operational best practices to assure consistency.

33.3.3 Real-Time Data Collection

There are two major types of operational data that must be collected. The first type is the traditional "IT" devices and their virtualization components. These devices provide telemetry by communicating using traditional networking protocols such as SNMP (Simple Network Management Protocol) or modern Web-based API (application programming interface) and include fairly well-defined templates that are embedded by each data collection utility that understands how to interpret the various values provided by the device itself. These devices each report hundreds of data points, so the mapping of just those values needed by DCIM

is critical. Considering the thousands of devices found within a data center, the available telemetry can be overwhelming.

The second type of device important to the DCIM solution is all of the components that form the mechanical and electrical infrastructure. These include power and cooling devices typically found external to the data center or those devices used to provide large volumes of power and cooling for subsequent distribution. This includes generators, battery backup UPS (uninterruptible power supply) systems, large floor-mounted PDUs, cooling chillers, and CRAC/CRAH (computer room air conditioner/computer room air handler) units. These devices typically communicate with more challenging protocols such as MODbus, BACnet, and LonWorks and in some cases older serial command lines via ASCII RS232.

In general data center metrics useful to DCIM solutions are observed every few minutes by polling. In a few cases there are triggered asynchronous events like doors opening, but the vast majority of this "real-time" data in a data center relates to temperature, humidity, pressure, power, and IT equipment (ITE) utilization, and those metrics are measured over longer periods of time, with analytics looking for trends over those same periods. Worth noting is that "real time" in the context of DCIM is not sub-second real time as the manufacturing world might define it, but instead typically deals with metrics observed over minutes or hours.

33.3.4 Integrations with Third-Party and/or Existing Management Frameworks, Web APIs

One of the requirements for DCIM solutions is their ability to connect to existing structures. Most IT and facility organizations have deployed point management solutions over time. These solutions have formed the core of data center management for years. The strongest DCIM solutions will be those that provide connectivity to these solutions as well as a number of traditional business management applications to coordinate workflows and metrics in a meaningful fashion. These systems can provide a wealth of knowledge source, are critical to service desk and ticketing processes, and include all of the control hooks to the existing components. There are dozens of IT and facilities systems that will be found across the many diverse corporate data centers, and DCIM vendors increasingly find their customers asking for these integrations. Integrations range from simple device access using standard protocols like SNMP, WMI (Windows Management Interface), IPMI (Intelligent Platform Management Interface), or Web services to more complex Web-based integrations of workflow and power chain management. The integrations are seemingly endless, and the strongest DCIM are accumulating inventories of these integration "conduits." Figure 33.5 shows just a sampling of major systems that will ultimately be connected over time to perform the DCIM function. Prospective customers should consider these inventories of "off-the-shelf" conduits from each vendor when making their choices.

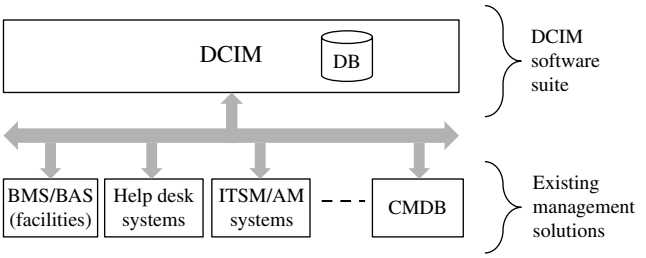


FIGURE 33.5 Modern DCIM deployments will require deep integrations across external enterprise management applications. Ecosystems of supporting technologies are forming.

33.3.5 Discovery Services and Location Services

33.3.5.1 Discovery Services: What Devices Do I Have?

Discovery services can be thought of as the logical discovery of active assets on a network. This active asset discovery can be deployed to identify or confirm the presence of devices upon the network, and then advanced reconciliation techniques can be used to assure that the DCIM asset model matches the reality or what is physical installed and vs. versa. How various DCIM vendors handle this reconciliation between what they model and what they logically discover is based upon their maturity as a solution.

Once logical addressing has been confirmed, active asset identification can occur. Since there is no single approach programmatically to determine the specific make and model and configuration of those devices, various technical approaches must be used determine their specifics and configurations. These approaches leverage a number of protocol interfaces including IPMI, ILO, DRAC, LLDP, SNMP, RSA, Serial, RPC, VNC, WMI, and a multitude of virtualization protocols.

Although a cumbersome process, active asset identification has been done today. A number of DCIM enhancement start-ups have created mission statements based solely on their ability to interrogate active devices and then using a combination of table lookup and metric retrieval to accurately identify each device and its configuration.

33.3.5.2 Location Services: Where Is Each Device Installed?

Sometimes grouped with discovery services, asset location services is part of an important theme of the DCIM segment. While logical detection of devices on a network has always been available using discovery services as mentioned above, there is no easy approach to detect where an asset is located physically. Essentially modern data centers must still rely on mostly manual audit and widely diverse documentation to identify the installed locations of data center assets.

Various vendors have brought forward their versions of physical asset location services, each requiring customized hardware add-ons of various complexities. Some of these systems identify physical asset placement at the high

granular "1U" level in the rack, while others are less specific and can identify regions where assets currently exist. Worth remembering is the various low-tech approaches such as bar code technologies that have been prevalent for the past 25 years and are still in use today to track assets. In some cases these traditional approaches have been adapted to become part of the DCIM solution and tend to offer the granularity needed for asset location tracking purposes.

Over time, there would appear to be a significant need for a standardized approach to determining specific asset location using an agreed industry standard. When available, this would enable all manufacturers of IT gear to release hardware devices that have the ability to identify themselves and their placement in a structure-mounting system for data centers.

33.3.6 Data Import and Export

One of the important features when implementing any DCIM solution is the ability to gather and normalize existing sources of asset and connectivity data. In a complex data center, there may be hundreds of thousands of individual pieces of data that would otherwise have to be manually entered or recreated through some means. In general, the labor cost to establish this knowledge manually without using any data import will exceed the cost of the DCIM software license itself and in some cases may actually be many times the cost of the software license. Hence, the significance of automated data import is a critically important part of the DCIM solution.

In response, most DCIM vendors include some means to import data sources such as spreadsheets and text files. Each vendor takes a different approach to importing and includes varying degrees of intelligence during the import process.

The most mature solutions use advanced field and pattern recognition and will even handle fairly well-defined types of problem resolution during the import process to map to existing source files. These files vary in format, and the error corrections available during the importing process may include missing information lookup, sequential missing data replacement, asset field de-duplication, proper handling of structured cabling conventions, and a general ranking of data fields based on overlapping sources.

Data import is a critical component of any DCIM solution at time of deployment. It is most commonly used once, and then the previous means to track and maintain asset knowledge are abandoned in lieu of the production DCIM solution. The most effective DCIM implementations allow the DCIM suite to become to single source of truth about assets once put into production. Some DCIM solutions provide for automated asset updating through discovery systems and/or integration with third-party CMDBs.

In a related topic, some DCIM solutions also enable the EXPORT of data to industry standard file formats such as CSV or XLS. These exports may include some or all of the DCIM database information and tend to be large files used for ad hoc analysis to feed into other systems transitionally. DCIM solutions that include the export functionality can usually recreate the entire main database using this same file as an import.

33.3.7 Material Catalog and Library

All DCIM solutions are designed to manipulate asset material life cycles, their placements, and their connectivity. In the creation of the physical structure, various types of devices must be selected from this catalog and then used throughout the DCIM modeling process.

Most vendors of DCIM solutions supply material libraries with 15,000 IT devices or more. It is the extent and means to enhance this library that will define the success and ease of use when attempting to articulate the current complexion of the data center faithfully.

The material catalog includes representations of devices and includes the manufacturer's specified parameters for each device. These parameters typically include hi-resolution renderings of the front and back of the devices, power requirements, physical dimensions, weight, connectivity, etc. In the case of complex devices, the material catalog also includes options that may be installed (such as power supplies, interface cards, etc.). All of these materials must be supplied by the DCIM vendor or must be created manually by the end user, which is a huge undertaking. Some vendors offer the ability to request these new devices to be added "on demand" (and typically within a week or two), while other DCIM vendors require the user to create these special new devices themselves. In a few cases, the DCIM vendor provides both mechanisms to enhance the material catalog.

33.3.8 Rack Planning and Design

One of the most visual features of any DCIM solution is the ability to create faithful representations of equipment racks and their installed gear and associated connectivity. In fact it is the visual representations of these rack elevations that typically attract some of the most enthusiastic initial interest by data center managers regarding DCIM. When considering DCIM vendors, a great deal of weight is given to the level of fidelity and absolute accuracy of the racks created by a given DCIM solution, and each offering is judged by its ability to most closely represent their real-world counterparts.

Most DCIM solutions use the aforementioned material catalog as building blocks for rack design as shown in Figure 33.6.

33.3.9 Floor Space Planning

The floor of the data center is essentially an *X–Y* coordinated grid used to identify the actual location of equipment racks and other freestanding data center gears. Floor planning is a

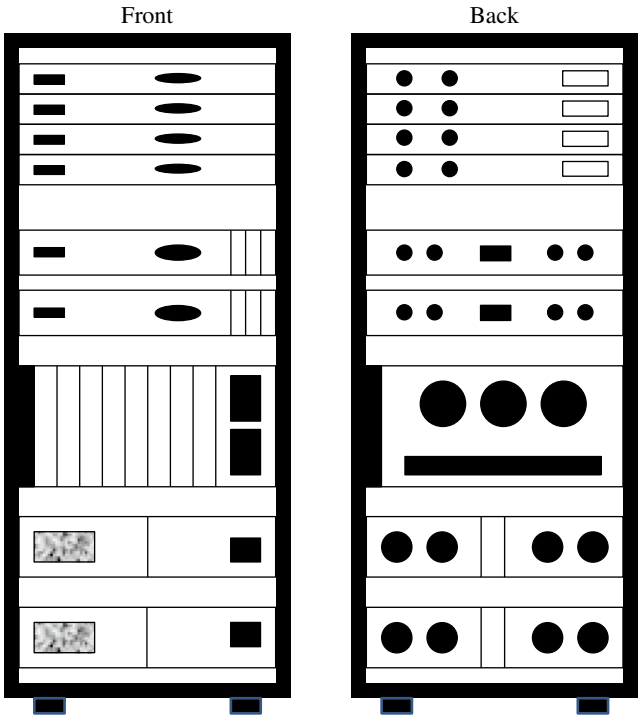


FIGURE 33.6 Rack planning tools allow a highly accurate representation of all installed assets, providing front and back views as well as the cabling interconnection of those devices.

critical reference process as the floor of the data center must be designed to mimic the actual geometries in each data center.

Unfortunately the data centers in use today are not always in simple rectangles. There are many types of construction and various obstacles that influence the placement of equipment and racks with the data center. The DCIM solution's floor planning component must allow these geometries to be captured accurately as they influence nearly all other aspects of modeling and planning within the data center. Accurate positioning for every device and rack is a core requirement to realize the maximum benefit of DCIM.

Most DCIM offerings also include the visual representation of the data center at large using the floor planning component. Shown here in Figure 33.7 is an example of where we see various top-down representations of the data center, with the floor tile systems, racks, CRACs, and other components shown in precise detail. These top-down views are also able to present metric data and aggregations using color-coded scales. For instance, they can represent the number of available rack units or total power consumption in each given rack. Using these visual representations of the data center, capacity can be visualized, and new projects can be created based on the actual complexion of the data center as it sits currently.

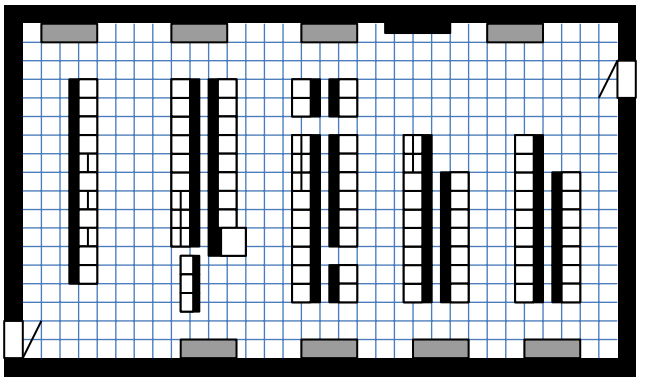


FIGURE 33.7 Data center floor planning enables the efficient placement of racks and cabinets, accounting for service allowances and obstacles.

33.3.10 Reporting: A Critical Part of the DCIM Story

One of the most valued capabilities for any DCIM solution is its reporting. Reporting is the way in which the raw information is correlated and then presented in a business impactful fashion. These reporting systems may also include a library of standard data center management reports and can typically distribute any of these desired reports to specific user(s) in an automated fashion. Other DCIM vendors simply include data store definition schemas and rely on their customers to design their required reports and then use industry leading reporting packages such as Microsoft reporting services, business objects, or SAS to create these desired reports.

33.3.11 Dashboards: A Picture is Worth a Thousand Words

Dashboards tend to a special case of reporting and can be considered the "at-a-glance" report. Dashboards have the ability to present vast amounts of information in easy-to-read displays suitable for desktop or operation "command center" consoles. Even though dashboards could be considered by some to be a "cosmetic" attribute of the DCIM suite, they are one of the first considerations new DCIM prospects look for when selecting a DCIM solution. Remember, the amount of data available within a DCIM solution can be enormous, and the ability to distill large amounts of this raw data into meaningful information that can then be presented using easy-to-understand visual dashboard elements is key to overall success. The ability to quickly access actionable information is a major value for DCIM deployments.

The vendor's included dashboards (Fig. 33.8) are a critical presentation of the DCIM function. With so many stakeholders in the proper operation of the data center, each will have a set of metrics that they hold themselves accountable to. Each has a set of needs that can be measured and derived from data found within the data center. Operations and finance look at costs of equipment, depreciation, warranty,

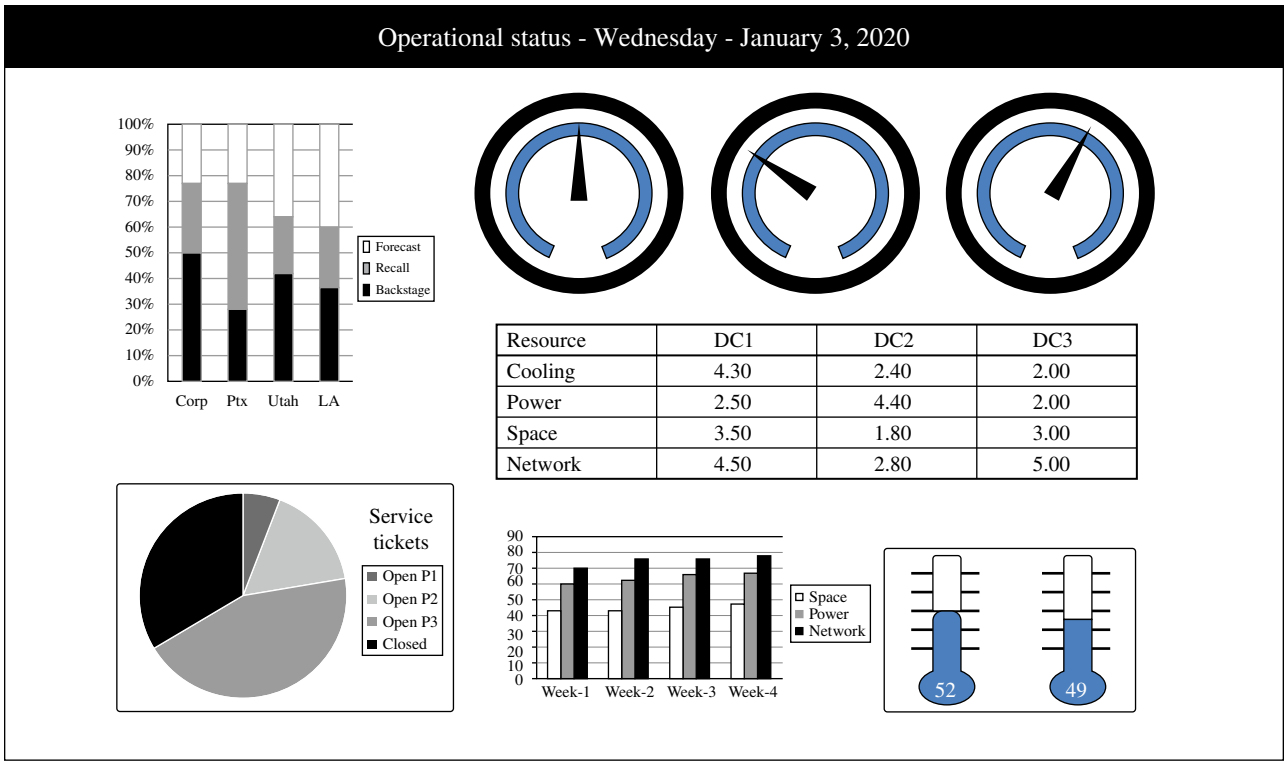


FIGURE 33.8 Critically important to the success of DCIM deployments, dashboards enable vast amounts of data to be visualized easily and in near real time.

etc. Facilities professionals look for trends in power and cooling consumption. Data center operators look for underutilized assets such as zombie servers.

33.4 THE DCIM SYSTEM ITSELF: WHAT TO EXPECT AND PLAN FOR

As we've seen, DCIM implementations provide a comprehensive set of asset management capabilities in the data center. The assets themselves have an enormous set of individual identifiers that are unique to each asset, ranging from physical characteristics and location, business owner attributes, and service information. DCIM technologies allow the data center assets to be organized in a variety of ways, which then allows solid business decisions to be made. In addition to these static attributes, the data center can provide a wealth of dynamic information that is derived from complementary technologies, sometimes referred to as "DCIM-Specialist" solutions. DCIM-Specialist solutions might be offered as integrated solutions by the DCIM vendor. Additional solutions are available from over a hundred vendors today.

33.4.1 The Platform's Architecture

In this section we'll describe the DCIM platform as well as the instrumentation layer that supports it. As stated previously, the DCIM data model is a living 3D model of assets, and as a general rule, the more connected this model is to the wealth of real-world instrumentation available, the higher the realized value from DCIM.

These early single-user developments were not bad choices in the premature time but most of these early DCIM applications are now being augmented or entirely replaced by well-behaved modern Web-based versions that are deployed upon enterprise-class and IT-maintained business servers. "The Web" and all of the advanced communications and presentation technologies have provided a huge opportunity to create complex management applications that can be easily scaled and widely accessed from anywhere on the Internet. These modern DCIM offerings typically scale by spanning across multiple server engines, with each engine serving various users, data collection, storage, and analytics/reporting functions.

33.4.2 The Platform's Data Storage Model

A core component of the DCIM suite is a robust data storage model. Large amounts of data will be sourced from a wide variety of sources, much of it being time series in nature, and all of it being required to be readily accessible for complex analysis and presentation needs. The data model itself must be robust enough to store data in a way that very complex analytics can be used across the data set interactively.

It is critically important for DCIM suites to have data models that are designed for interactive retrieval. Large volumes of

data will be stored over time, and one of the key attributes of a strategic DCIM solution is its ability to present interpretations of vast quantities of raw data into meaningful metrics. While it may sound like a technical detail, the choice of storage approaches will directly affect the usability of the entire DCIM solution. Users will not tolerate the slow performance caused by data retrieval in a DCIM solution. Complex searches across gigabytes of data would take unacceptable amounts of time if the wrong storage technology was chosen. Imagine that every time you wanted to use an application on your smart phone, that there was a 30-second delay before the first screen. You would likely NOT use the smart phone. The DCIM storage model, if chosen poorly, has the potential to have the same effect.

33.4.3 The Platform's User Interface

Modern DCIM suites are most commonly Web-based and utilize the latest Web-based access methods being adopted in common business management applications. The visual presentation of DCIM is complex and varies from vendor to vendor, but each shares a common goal of allowing easy navigation across a vast field of data by multiple users across the Internet.

The GUI (graphical user interface) can be considered one of the key attributes of a DCIM solution, as customer adoption many times is directly related to the intuitive nature of the GUI. A great example of an intuitive GUI interface is Google's Earth application that allows the untrained user to start with a map of the entire planet and within seconds zoom into a view showing the house where they live.

The user interface for DCIM is critical. These applications must be highly intuitive. Large populations of any combination of IT and facilities equipment spanning over thousands of square feet must be quickly accessible. DCIM enables the relationship between components to be clearly articulated in great detail.

33.4.4 Instrumentation: Sensing the Physical Components in Real Time

Modern data centers can provide a wealth of information about the current status of everything from the power chain and cooling status to the performance of the servers and virtualization layers. It all fits into a category that is leveraged by DCIM called "instrumentation." Instrumentation is essential for a DCIM solution to be effective, and it includes a wide range of technologies and protocols, each intended to gather a specific portion of the entire infrastructure. The most mature DCIM suites have instrumentation subsystems, and the DCIM systems themselves deal with the normalization and presentation of this instrumentation data. A point of reference regarding scale is worth noting here as the magnitude of data gathers using various means of instrumentation can be massive. In a typical small data center with 100 racks and 1,000 servers, tens of thousands of data points per minute can be generated!

33.4.4.1 Environment Instrumentation: Temperature, Humidity, and Airflow Sensors

One of the earliest arrivals on the journey to DCIM has been the environmental sensor vendors. For years environmental sensors in the data center were considered merely as a "nice-to-have" tool by data center operators. As such, their usage was limited to a relatively small population of these operators. Some of the reasons given for low adoption included the perception of a relatively high cost solution to an otherwise simple set of needs, nonspecific use cases, installation cabling complexity, and lastly the associated costs and limited pre-DCIM business management value. Environmental sensors were not considered a strategic source of knowledge within the data center.

The "American Society of Heating, Refrigerating, and Air-Conditioning Engineers" (ASHRAE) has published guidelines over the last several years in support of new ways of looking at and optimizing data center cooling. Their recommendations on sensor placement have provided resurgence in sensor innovation, and in fact a handful of start-ups have been formed to meet these ASHRAE-inspired needs for easy-to-use environment sensors to support DCIM deployments. These stand-alone environmental monitoring systems are available in wired and wireless variants. These purpose-built systems fulfill the need to understand the temperature and humidity of a data center and can do so at the granularity recommended by ASHRAE.

Wired environmental systems were the first entrants into the data center market and usually consist of purpose-built micro-PC hardware (controller) with some form of micro-operating system within and all of the necessary analog I/O hardware to monitor temperature and humidity and perhaps read and control dry contacts and relays, listen to or emit sound and alarms, sense light, etc. These controllers are connected to a LAN port anywhere in the data center, and all interactions with these devices are using standard Web- and IP-enabled protocols. This "connected" approach requires significant deployment complexity where cabling can become costly and prohibitive.

A second type of environmental sensing system has emerged, which addresses implementation simplicity through the use of wireless. Wireless systems can either be AC powered commonly using 802.11 (Wi-Fi) or battery powered using 802.15 (e.g., Zigbee) or active RFID technologies. Powered wireless devices operate as long as AC power exists and due to their physical power connectivity tend to behave much like a wired sensor solution, gathering and communicating larger amounts of sensor information much more frequently. Powered wireless devices allow network connection wirelessly, but the requirements for AC power itself make these "wireless" solutions, something less

than truly wireless. Worth noting is a continued concerned by many data center operators prohibiting the use of Wi-Fi in the data center for security reasons, since opening Wi-Fi channels for instrumentation also allows other types of network access using the same Wi-Fi channels via PCs and handhelds alike.

The new generation of battery-powered wireless monitoring devices that can be quite impressive in their ease of deployment and true to their name as wireless. These battery-powered devices significantly limit the amount of data transmitted by employing highly intelligent data manipulation and de-duplication, constantly reporting changing sensor values each period of time. These low-power battery-powered wireless devices tend to be unidirectional, reporting changes upward, but do not receive data of any type. Monitoring can be viewed as an upstream activity, so this approach works perfectly in the majority of DCIM supporting roles. Some of these devices have claimed battery lives in excess of 3 years!

Working together, a data center may include hundreds or thousands of wired and wireless sensors of temperature, humidity, and air pressure or flow. Each of these systems has its pros and cons, and ultimately DCIM installations will likely find use for a combination of these systems to support various portions of their environment.

33.4.4.2 Power Instrumentation: The Rack PDU

The most basic building block in a data center is the rack or cabinet that houses active equipment. Each rack may contain up to 40 or more active devices that require power (and the associated cooling) and connectivity. The number of racks in a data center may range from a small handful to well into the thousands. The scope of each data center varies widely, but what remains constant is the requirement for power in these racks. The means to deliver power to these devices is an appliance referred to as a "rack" PDU (power distribution unit or in some cases referred to as a power strip or plug strip).

While supplying power to the active equipment is mainly a function of the power capacity and number of outlets available, modern energy management focuses data center operators on maximizing the ways in which power is used and the efficiency in doing so with the ultimate goal to reduce costs. Elaborate power distribution strategies have been devised over the past 10 years to move power within a data center more effectively, taking advantage of some of the modern electrical utility's new approaches to supply raw power.

There are two optimization opportunities related to power: (i) distribute power more efficient through higher voltages and higher currents in smaller spaces and shorter distances and (ii) measure and monitor usage at a highly granular level, allowing individual components to be studied and analyzed over time.

DCIM provides the means to visualize these power chains and allows granular business decisions to be made. DCIM allows these power distribution approaches to be deployed, studied, and then actively monitored to assure that loads are properly balanced and demand for available power and cooling is matching available supply.

33.4.4.3 Hidden Instrumentation: Server Intelligence

DCIM solutions have the ability to associate large amounts of operational data for any asset to allow business decisions to be made. Various protocols are used to extract this information from servers, including IPMI, SNMP, WMI, and each of the virtualization vendors' own APIs.

In a typical modern server, storage, or networking device, a wealth of knowledge is being made available today to external applications upon demand. This includes not only the more traditional logical operating parameters and performance metrics (such as CPU and I/O rates) but also typically includes physical device metrics, such as power consumption, power supply status, operational status, internal fan speeds, multiple temperature readings within each device, failure alerts, security lock status, etc. DCIM suites provide the unique opportunity to consider all of this physical and logical information together.

33.4.4.4 Building Instrumentation: Building Management Systems and Mechanical Equipment

Usually referred to as "facilities" equipment, a typical data center has a long list of equipment that becomes part of any well-implemented DCIM solution. This includes the power generation and distribution devices, cooling components, and all of the control systems that have been deployed over the years to control these systems. The devices may be networked already, or they may be stand-alone.

The true promise of DCIM is to join the world of IT and facilities, which allows all of the equipment required to provide computing services to be a useful part of the DCIM structure. Only with a complete picture of the IT and facilities style components can the maximum value be derived from DCIM. For example, a "power chain" consists of many links: the server's power supply, the in-rack PDU, the floor-mounted PDU, the data center UPS, the breaker panels, electrical distribution systems, and the generators. Each of these forms a component of the power structure that is considered when making business decisions in the data center.

Today, building management systems (BMS) are a mid-point aggregation level and source of metrics for DCIM. Typically installed to control cooling resources, these "BMS" systems can be fairly simplistic in nature or highly complex. In general these systems are rigid in deployment and change very little over time. These systems can become a wealth of great information when integrated into a DCIM solution and in fact will allow DCIM suites to quite easily control cooling resources.

33.4.5 The Rack: The Most Basic Building Block of the Data Center

Data center racks themselves are the physical building block for the IT task. Typically a physical cabinet is made of steel; each rack is typically 6-ft tall, 2-ft wide, and 3-ft deep. Note that racks are available in many different sizes and configurations. Commonly 42 devices may be housed in each rack, with larger or smaller numbers being seen in specific applications.

As standard building blocks, most DCIM offerings understand these mechanical designs and use very accurate templates for the selected rack(s). The size and shape of each rack are well understood and when coupled with floor tile systems allow an extremely accurate representation of the data center to be modeled. DCIM offerings use these building blocks as the basis for their high-fidelity physical topology representations and rely on this ordered approach when depicting location and relative placements.

33.4.6 Remote Access and Power State Management

Related to DCIM has been the notion of remote access to systems. In fact, long before the DCIM marketplace emerged, the concept of managing the infrastructure was left primarily to two constituents: (i) facilities managers who visualized power and cooling using purpose-built BMS and control panels and (ii) systems administrators who used hardware and/or software tools to access their equipment remotely to power cycle or reconfigure operational settings.

The facilities manager's ability to manage their power and cooling infrastructures has become quite mature. Highly advanced and completely customized visualization, dashboard, and control mechanisms have been created by the large building automation vendors. These BMS are tailored individually for each deployment and tend to be quite functional, albeit extremely rigid. BMS and their more advanced counterparts, building automation systems (BAS), have high price tags and must be defined at the time of building construction in extreme detail by the facilities engineers who manage the building power, cooling, security, and lighting systems. Building engineers and mechanical designers work in concert to create these infrastructures, and then BMS/BAS systems are tailored to reveal the inner workings and control capabilities. These systems tend to change very little over time and only when major facilities construction changes occur do these need to get re-evaluated and capabilities updated.

For the IT world, some of these remote management technologies are referred to as "power cycling," "KVM (keyboard video mouse)," or simply "console" and can be seen in Figure 33.9. Essential physical management of IT systems and data center devices was based upon the one-user-to-one-device approach using one or more of these

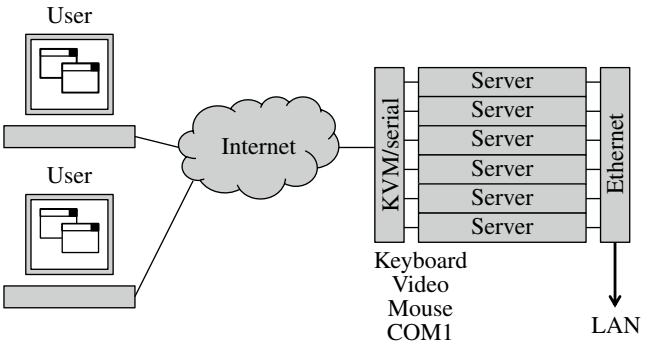


FIGURE 33.9 For years, physical infrastructure management was limited to remote console and KVM access technologies alone.

technologies. This brute force approach to IT device management was directed at a single server, switch, or other type of IT system as a stand-alone management entity. The servers and other devices had no notion of placement and relative location or required resources, and basic metrics for energy consumption and temperature usually did not exist. Remote management technologies can be considered some of the most basic and primitive means of device management used for the last dozen years. The need for this type of remote access has been greatly reduced by hardware and software maturity found in enterprise-class equipment and today is most commonly found where a specific mission-critical device (and single point of failure) is deployed for a specific functions. For these applications, power cycling and associated system reboot are the most common uses of these remote management technologies.

Although not strictly required for a successful implementation of DCIM, these suites can usually take advantage of remote access technologies already deployed. These DCIM systems can allow traditional system administrators to share the user interface found within DCIM suites and navigate to the system where remote operator access is desired, addressing configuration or reboot requirements.

33.5 CRITICAL SUCCESS FACTORS WHEN IMPLEMENTING A DCIM SYSTEM

The approach you use to implement a DCIM can make or break the long-term success of your project. As DCIM is a relatively new category, much of the waters that you and your team will be navigating will be unfamiliar territory, and you'll find a number of surprises along the way. Above all, you'll need to keep reminding yourself about the goals of the project to not let it get away from you.

Here is a list of some of the critical factors that should be considered today to increase the likelihood of success in your DCIM project. While your mileage may vary and every organization is different, there is a common set of steps

found in the most successful DCIM deployments. Your DCIM journey will have many of these same steps:

1. Do your research. Read and talk to your peers at other companies that have invested in DCIM
2. Get buy-in from all of the stakeholders. The four critical organizations to include in this journey are IT, facilities, finance, and the corporate social responsibility team
3. Be realistic with setting scope and timing
4. Document your existing processes and tools
5. Audit and inventory the assets that you already have installed
6. Determine your integration requirements
7. Establish a roster of users and associated security policy
8. Determine each stakeholder's required outputs (dashboards, reports, etc.)

Two important factors worth more discussion are the following.

33.5.1 Selecting the DCIM Vendor

Look for a DCIM vendor with a vision that matches your own. Obviously every data center strategy is unique across the industry, but there will be well-formed thoughts about capacity planning, operational excellence, energy management, disaster recovery, etc. that must be discussed prior to choosing a DCIM vendor.

Take into account how long the vendor's solution has been available and how many installations each vendor has. Obviously more is better as it supports and defends the vendor's approach to DCIM and will ultimately help steer your choice. Each vendor's installed base will have provided a priceless resource of users requirements from similarly situated users who have walked down the same paths you are GOING to walk down. Vendors should be able to share existing customer names and contact details or arrange for discussions with these customers on request.

Consider each vendor's recommended platform, architecture, and integration capability. Will the new solution be able to be integrated with the other systems that you have in place today? Can the vendor cost-effectively deploy it at the scale of your IT structure? How do they handle many users, many assets, and many data centers? How do your geographically dispersed data centers affect the DCIM suite's performance and real-time monitoring capability?

Look for vendors that can provide NEW levels of visibility and analysis, in ways previously not available to you. You are not looking for a prettier way of doing what you can already do, you are looking for new business management insight to allow you to make more informed decisions, respond more quickly, etc. Visibility down to the device level is just the start of a solid selection, and what the vendor's offering does with that level of granular information is where the magic comes from.

Once you have a short list of vendors, you should require out-of-the-box, demonstrable capabilities. The DCIM marketplace is relatively new, and nearly all vendors want to please prospects. When looking for a DCIM vendor, you really want to consider which of these capabilities they can deliver today and avoid the more theoretical discussion about what they could do given enough time and enough money. Engineering projects create orphaned installations and can diverge so much from any commercial offering that customers will be abandoned at the onset, and will not be able to take advantage of the selected vendor's future releases of their software. As a rule of thumb, if they can't show you specific DCIM features, they probably don't have them built yet. Be cautious here as this will determine your long-term costs for DCIM.

Lastly, ask about pricing models. Be specific. Software vendors are notorious about turning what was presented as a product into a time and material project. This can be a costly approach. Choose a mature DCIM vendor that details their cost structure, which pieces are off the shelf and which are custom. They must also articulate ongoing maintenance costs.

33.5.2 Considering the Costs of DCIM

This is one of the most misunderstood topics when discussing DCIM, since the definition of DCIM is so diverse. As mentioned earlier, there are a handful of management software suites that comprise the top-level DCIM functionality. This enables the business management aspects of the deployed solution, and it is the most common interface that users will interact with. All of the rest of the offerings in the DCIM space are actually sub-components of the total solution. Gartner refers to these vendors as "DCIM -Specialists" or simply "enhancements" to the DCIM solution. These enhancements include hardware or software that provides real-time data about power, or environments, and allows deeper analytics or customer presentations, or even the ability to discover and identify various assets and their locations.

Today, there is no single pricing scheme for DCIM. The DCIM software management suites are priced in a wide range of schemes based on size or capacity, with perpetual and subscription licenses further complicating the process. Although different pricing schemes exist, for comparison purposes we can use a unit of measure at the "rack" or cabinet.

DCIM enhancement components on the other hand are much more straightforward in pricing. Sensor vendors, for instance, can tell you exactly what a thousand sensors would cost, and if you plan to use four sensors per rack, you can do the math to determine what these 250 racks would cost to outfit with a DCIM-Specialist vendor's sensors.

So, what does it cost when you are looking to budget DCIM for an upcoming project? As a general rule of thumb in 2020 dollars, DCIM suites and its natural DCIM Specialist (enhancements) should be budgeted at about US$1,100/rack. This will include core DCIM functionality, basic integrations with common systems, real-time sensors, installation, and training. (Intelligent rack-based PDUs will add another $1,000 per rack if intelligent power metrics are desired.)

33.5.3 Other DCIM Considerations

DCIM solutions are at a level of maturity that large and small organizations alike can take advantage of this new area of data center management. It has never been easier or timelier to create an extended view of the data center infrastructure, in the logical views already deployed with physical layer extensions found in modern DCIM offerings.

A few points worth considering as you begin to investigate and then formulate your DCIM plans are as follows:

- What is your adoption time frame, or can you afford to do nothing?
- What are your existing sources of truth and other documentation used in production today?
- What about open source DCIM?
- Once DCIM is deployed, where do all the existing IT support people go?
- Your DCIM needs and capabilities will evolve over time.

33.6 DCIM AND DIGITAL TWIN

Along with rapid development of data centers, management at data centers becomes more complicated than ever. Data centers are evolving every day. Owners not only want current and past information of their data centers but also are very anxious to know its future. When an owner can get reliable prediction on impacts of changes, they can reduce the costs and risks, which is essential to the data center business. Tightly integrated DCIM and digital twin are good solutions to know past, now, and future of a data center.

33.6.1 What Is Digital Twin

Gartner predicts digital twin is one of the top 10 strategic technology trends and IDC also predicts digital twin the top 10 worldwide of IoT 2018 Predictions. What is digital twin? Gartner's definition is "over time, digital representation of virtually every aspect of our world will be connected dynamically with their real-world counterpart and with one another ..." The most important characteristics is dynamically representative. ITE are dynamically changed in data center during its life span. It changes almost every day. It is imperative to build a digital twin model for life-long activities of a data center same as manufactured products.

33.6.2 How Digital Twin Works in a Data Center

Let's look at life cycle of a data center that has a complete different life cycle comparing to a traditional industry infrastructure. A data center has a life span of 10–30 years or longer. It starts from concept design as traditional product, followed by detailed design, architectural, and mechanical and electrical design. At this point, the IT configuration design starts (Fig. 33.10). The most critical stage of its life cycle is operational stage (Fig. 33.11) where the business or cash flow starts.

During the operational stage, IT configuration changes almost every day. It includes plan, deploy, operation, and retire. Generally, ITE life cycle is 3–5 years, which means that IT configuration is diverse from its original design frequently. The owner wants to know every changes it cost and associated risk. That is why digital twin is an important tool to that tracks changes and predicts the results of changes.

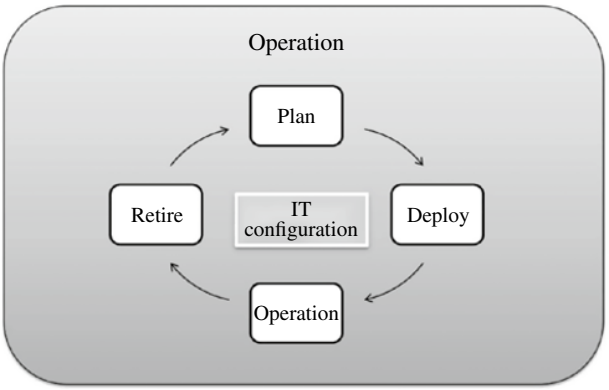


FIGURE 33.11 Data center operational stage.

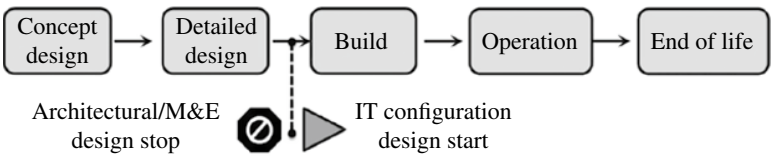


FIGURE 33.10 Data center life cycle.

33.6.3 DCIM and Digital Twin Works Together

DCIM is a great tool to track past and today's information that are relating to a data center. The information collected could be used by operations management to predict and plan maintenance activities. CFD is the best simulation tool to predict best airflow precisely. By integrating DCIM and CFD (Fig. 33.12) help build a complete digital twin model. DCIM can collect and monitor data combining with data from CFD to measure and provide best solution (Fig. 33.13). This will result in the least operating costs to operate a data center. By applying machine learning to this CFD model, it will help calibrate the model and predict the future such as cooling failure scenarios.

DCIM and CFD are complementary in a digital twin system. It can harness prediction to manage by looking forward and removing ambiguity in decision-making process. Combined DCIM and CFD system can increase confidence level in data analytics, reduce troubleshooting turnaround time, and optimize the investment in infrastructure.

33.7 FUTURE TRENDS IN DCIM

The DCIM marketplace is rapidly progressing as a management category. Whereas most efforts underway today allow a highly granular means to maintain and present an accurate representation of the existing IT and facilities infrastructures (complete with real-time metrics), the future of DCIM will include (i) consolidation and/or rationalization of vendors solutions, (ii) new leveraging features including automation and control and support for asset location and auto-discovery technologies, and (iii) ecosystem approach where specific cross-vendor integrations will be formed using more standardized approaches to integration across these related infrastructure management solutions.

33.7.1 Automation and Control

Control is a broad topic that will transform DCIM suites from visibility and analysis solutions into well-orchestrated business and workload management solutions that focus on

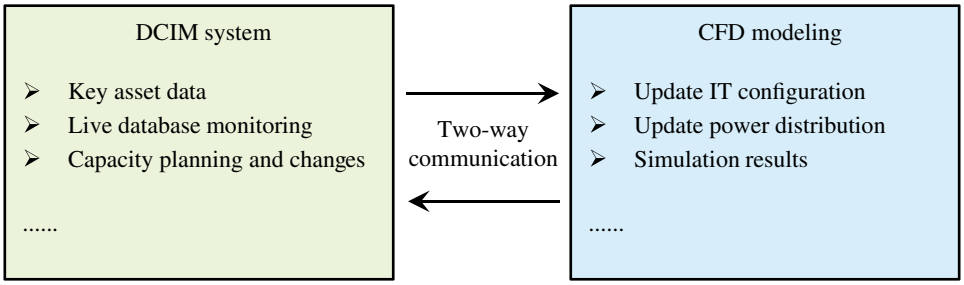


FIGURE 33.12 DCIM and CFD integration.

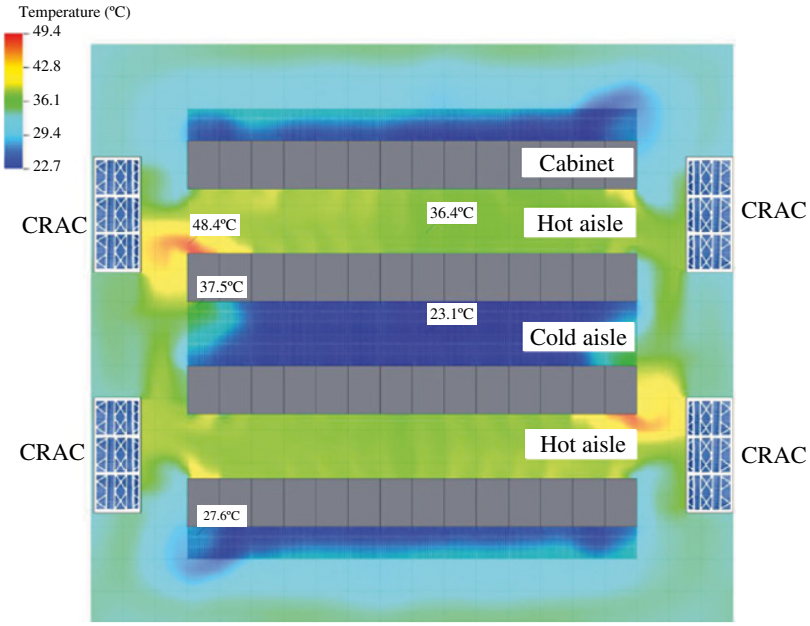


FIGURE 33.13 CFD solution can locate the current hot spot and predict future deployment plan. Source: Courtesy of Rainspur Technology Co., Ltd.

dynamically adjusting ALL resources that are required to meet computing demand. DCIM with artificial intelligence and machine learning will become commonplace over the next few years. While there are a number of companies today that focus on these automated approaches to dynamic capacity management (cooling and processing), the market overall has wide range to learn and expand.

33.7.2 Asset Location and Physical Discovery

Asset management is one of the most important functions in DCIM. There are many ways to determine where an asset is physically located. By applying asset management with smart in-rack PDU, RFID and NFC (near-field communication) tags, assets or devices can be detected and interrogated quite easily to determine where they are on the network, the type of device, and if the server is running.

33.7.3 Ecosystems and Integration "Standards" and Linkages to Other Systems

The DCIM marketplace is poised for strong partnerships to form. Potential customers are looking for the DCIM vendor community to consider all of the strategic pieces required to demonstrate core data center management value and then seek out those portions that they do not make themselves. Prospective customers of DCIM are looking for the "heavy lifting" for integrations to be done by the DCIM vendors involved. It is no longer enough to hide behind standard statements that speak about "protocols" such as "SNMP" or "Web API" as their sole approach to integration. Experienced IT professionals understand that general purpose support for standard interfaces is a far cry from systems that can seamlessly work together. These potential adopters of DCIM are looking for strong ecosystems to form.

33.8 CONCLUSION

Real DCIM is available now. The management of physical aspects of the data center has been a fragmented and poorly understood science for the past decade, and as such this physical layer of management has been addressed historically by over-provisioning ALL resources. The general guideline in the past was to simply create such an abundance of resources that the upper limits would never be tested. Only with the recent and dramatic rising costs of power and the rapid movement to virtualized dense computing has the attention to the gross inefficiency associated with over-provisioning been scrutinized.

With the advent of DCIM, CFO, CIO and even CEO executives are searching for the next phase in cost-effectively managing the data center to include a coordinated IT and facilities costing and service delivery model. While the cloud and virtualization technologies have their own unique impacts and opportunities to their data center strategy, the entire hybrid structure will benefit from DCIM in very tangible ways. DCIM is here to stay, and the most competitive organizations have been executing aggressive plans to leverag these new capabilities today.

ACKNOWLEDGMENT

Our sincere thanks to Mr. Mark Harris at Nlyte Software who prepared and published this chapter in the first edition of the *Data Center Handbook*. His chapter was updated and enhanced by Dongmei Huang, Logan Zhou and TAB members.

FURTHER READING

Azevedo D, Belady C, Patterson M, Pouchet J. Using CUE™ and WUE™ to improve operations in your datacenter. The Green Grid; September 2011.

Avelar V, Azevedo D, French A. PUE™: A Comprehensive Examination of the Metric. https://datacenters.lbl.gov/sites/all/files/WP49-PUE%20A%20Comprehensive%20Examination%20of%20the%20Metric_v6.pdf. accessed 11/20/20.

Belady C. Carbon usage effectiveness (CUE): a green grid data center sustainability metric. WP #32. The Green Grid; 2010. Available https://www.thegreengrid.org/en/resources/library-and-tools/241-Carbon-Usage-Effectiveness-%28CUE%29%3A-A-Green-Grid-Data-Center-Sustainability-Metric. Accessed on May 16, 2020.

Blackburn M. THE GREEN GRID data center compute efficiency metric: DCcE. WP#34. The Green Grid; January 2010. Available at https://www.thegreengrid.org/en/resources/library-and-tools/240-WP. Accessed on 11/20/20.

Cappuccio D. DCIM: going beyond IT. Gartner ID: G00174769; March 2010.

Cappuccio D, Cecci H. Cost containment and a data center space efficiency metric. Gartner ID: G00235289; June 2012.

Cole D. Data center energy efficiency – looking beyond PUE. Nolimits Software; June 2011. Available at https://www.missioncriticalmagazine.com/ext/resources/MC/Home/Files/PDFs/WP_LinkedIN%20DataCenterEnergy.pdf. Accessed on July 10, 2020.

Cole D. Data center knowledge guide to data center infrastructure management; May 2012. Available at https://nswpep.com/wp-content/uploads/white-papers/Data-Center-Knowlede-DCIM-Guide.pdf. Accessed on July 10, 2020.

Data Centre Specialist Group. Data centre fixed to variable energy ratio metric; May 2012. Available at https://www.bcs.org/media/2917/dc_fver_metric_v10.pdf. Accessed on July 10, 2020.

EPA. Annual energy outlook with projections to 2035. DOE/EIA-0383 (2012); June 2012. Available at http://www.eia.gov/forecasts/aeo/pdf/0383(2012).pdf. Accessed on July 10, 2020.

Fichera D, Washburn D, Belanger H. Voice of the customer: the good, the bad, and the unwieldy of DCIM deployments; November 2012.

Fry C. Green data center: myth vs. reality. WWPI; January 2012. Available at http://lanster.home.pl/autoinstalator/joomla/download/Myth-vs-Reality-Green-Data-Center.pdf. Accessed on July 10, 2020.

Harris M. Taxonomy of data center instrumentation. Mission Critical Magazine; 2009. Available at http://www.missioncriticalmagazine.com/ext/resources/MC/Home/Files/PDFs/WP-Taxonomy_of_Data_Center_Instrumentation-Mark_Harris.pdf. Accessed on July 10, 2020.

Howard C. Hybrid IT: how internal and external cloud services are transforming IT. Gartner ID:G00231796; February 2012.

IBM Global Technology Services. Data center operational efficiency best practices. Ref: RLW03007-USEN-01; April 2012.

Kaplan J, Forrest W, Kindler N. Revolutionizing data center energy efficiency; July 2008. Available at https://www.sallan.org/pdf-docs/McKinsey_Data_Center_Efficiency.pdf. Accessed on July 10, 2020.

Kumar R. The six triggers for using data center infrastructure management tools. Gartner ID: G00230904; February 2012.

Mell P, Grance T. The NIST definition of cloud computing. NIST, Pub #800-145; September 2011. Available at https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf. Accessed on July 10, 2020.

Neaves R. Moving the data center from chaos to control. Nlyte Software; 2011. Available at https://dsimg.ubm-us.net/envelope/112192/336762/1258477930136_nlyte_Software_DCPM_White_Paper.pdf. Accessed on July 10, 2020.

Pultz JE. More than half of data center managers polled will likely be using DCIM tools in 2013. Gartner ID: G00231803; March 2012.

Schreck G. Put DCIM into your automation plans. Forrester; December 2009.

Uptime Institute Global Data Center Survey, 2018. https://uptimeinstitute.com/uptime_assets/f7bb01a900c060cc9abe42bb084609f63f02e448f5df1ca7ba7fdebb746cd1c4-2018-data-center-industry-survey.pdf Accessed 11/20/20.

34

DATA CENTER AIR MANAGEMENT

Robert Tozer and Sophia Flucker
Operational Intelligence Ltd, London, United Kingdom

34.1 INTRODUCTION

Data center energy consumption has continued to grow due to the increasing intensity with which businesses and individuals use IT. Current efforts to reduce energy consumption focus on the operation of the IT equipment and cooling systems and include expansion of the operating temperature ranges for IT equipment [1] and the raising of cooling system set points. Liquid cooling is used in some high density applications, however most facilities are air cooled; air management problems, solutions, and metrics are the focus of this chapter. In order to increase operating temperatures and optimize cooling unit fan speeds, it is important to first ensure that data center airflow is properly managed to ensure the continued reliability of IT equipment.

In legacy facility designs, data halls are flooded with cold air without much thought about managing airflow. With low load densities this was not a problem as this did not result in hot spots and energy efficiency was less of a concern, but as densities have increased, the need to manage air and reduce energy consumption has become more apparent. Traditional cooling systems in legacy data centers supply cold air from CRAH (computer room air handling) units located on the perimeter of the room to IT equipment via a raised floor. Air exhausted from the rear of IT equipment returns to the CRAH unit having mixed with the colder air of the data hall. Much energy is wasted when air from cooling units bypasses the IT equipment, returning directly to the cooling unit without cooling any IT equipment. This wastage of supplied air means that IT equipment draws hot exhaust air to fulfill its required airflow volumes. As this results in higher IT inlet temperatures, cooling unit set points are often reduced to compensate, leading to higher energy consumption and inefficiency. Air is also often oversupplied resulting in an overconsumption of energy by the fans in the cooling units. Appropriate air management minimizes bypass and recirculation of air from server exhausts. Through continued monitoring of air performance, operators are able to target areas for improvement and reduce energy consumption and risk.

34.2 COOLING DELIVERY

Cooling units deliver cooled air to the data hall; these are usually chilled water CRAH units, DX CRAC (direct expansion computer room air conditioning) units, or AHUs (air handling units). This air is typically drawn into the front of servers and extracted at the rear before returning to the cooling units. For other IT equipment such as network and storage devices, other airflow paths are common, e.g., top or side discharge. The temperature of the air entering the IT equipment is dependent on the amount of air leaving the cooling units that actually reaches them.

Figure 34.1 illustrates an example of a data hall where the cooling units are controlling to a return temperature of 21°C (69.8°F). A sample of IT equipment inlet temperatures has been measured and is displayed in ascending order. The

Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center, Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

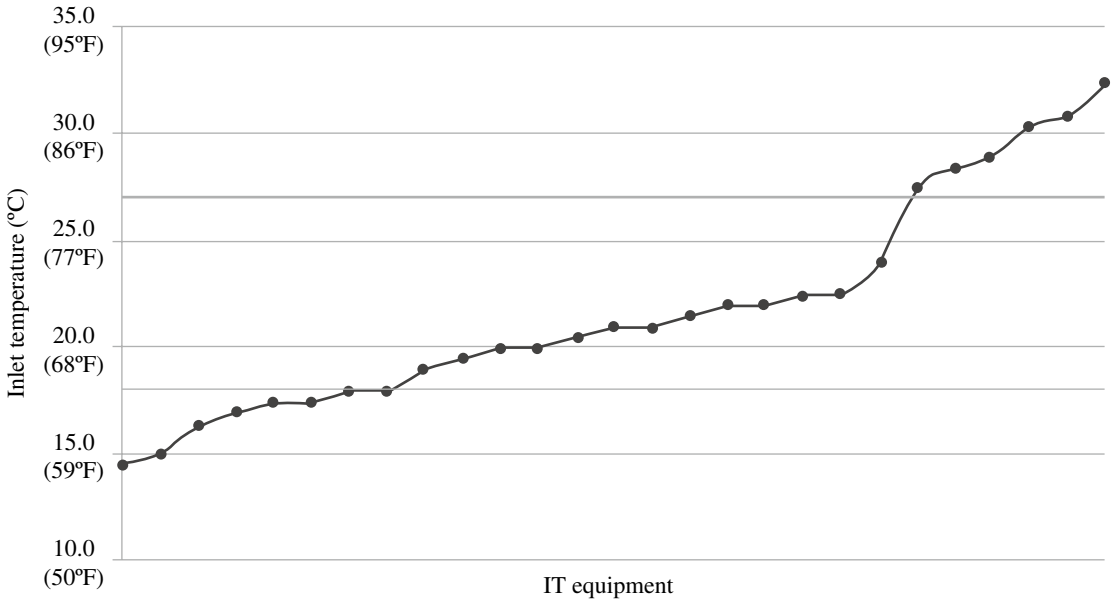


FIGURE 34.1 Example range of measured server inlet temperatures. Source: Courtesy of Operational Intelligence.

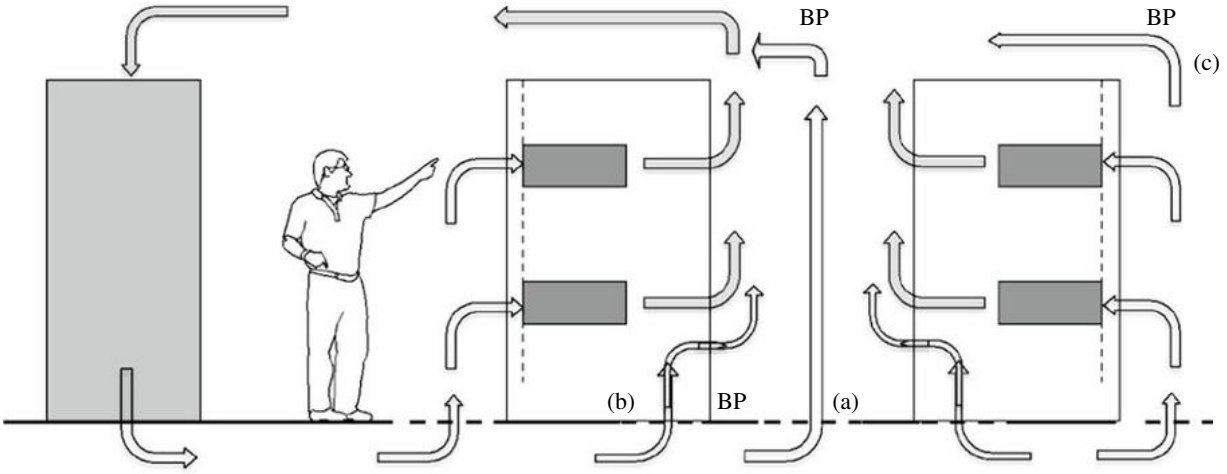


FIGURE 34.2 Bypass flow in the data center. Clear arrows = cold air, dark arrows = hot air. Source: Courtesy of Operational Intelligence.

ASHRAE recommended range lower and upper boundaries are also shown (18–27°C/64.4–80.6°F).

Although the cooling units are working as designed, there is a wide range of air temperatures observed at the inlet to IT equipment. Some of the sample size is receiving air that is colder than required, representing an energy saving opportunity, and some is receiving air hotter than the upper limit, which potentially increases the risk of failures due to overheating/reduces operating lifespan.

Once hot spots have been discovered, where IT equipment receives higher temperature air, a typical solution is to decrease the cooling unit set point. This does not deal with the root cause of the problem and contributes to additional energy wastage; it makes the issue of supplying air at a lower temperature than required worse.

34.2.1 Bypass

Bypass flow (BP), shown in Figure 34.2, refers to cooled air that returns to CRAH units without cooling IT equipment. Bypass is typically found at floor grilles in hot aisles (a), gaps in the raised floor such as cable cutouts at the back of the rack (b), and floor grilles oversupplying air in open cold aisles (c). High bypass can lead to servers starved of cold air, compounding recirculation (discussed next) as the server draws in air whether it is hot or cold.

Bypass air reduces the return temperature at the cooling unit; if bypass was zero, this would be the same temperature as the IT equipment exhaust.

Reducing bypass often allows cooling unit fan speeds to be reduced, which can result in a significant energy reduction (fan affinity laws: theoretical cubic relationship between fan speed and power).

34.2.2 Recirculation

Recirculation flow (*R*) shown in Figures 34.3 and 34.4 refers to air that is drawn from the IT equipment exhaust (its own or from another device), back to the IT equipment inlet to satisfy required air volumes, without returning first to the CRAH for cooling, resulting in elevated server inlet temperatures that may be higher than the desirable range. This is mainly addressed by ensuring sufficient cold air reaches the inlet to all IT equipment and by installing rack blanking plates, and blocking other leakage paths.

If recirculation was zero, the inlet air temperature to IT equipment would be the same as that supplied by cooling units.

Reducing recirculation often allows air and cooling system set points to be increased (increased refrigerant evaporating temperature), which reduces compressor energy requirements; it also increases the opportunity for economizer free cooling (without refrigeration).

34.2.3 Negative Pressure

Negative pressure (NP) flow shown in Figure 34.5 refers to data hall air that is induced into the floor void due to the Venturi effect. This is the result of high velocity pressure (proportional to the square of air velocity), which reduces the static pressure (Bernoulli law of fluid dynamics). Once the velocity pressure is higher than the total pressure (made up of static and dynamic pressure), the static pressure at that point becomes negative. With negative static pressure under the floor, air will be induced into the floor void from the space above. Negative flow and even low static pressure cause considerable problems for racks in the immediate region as these will be starved of cold air. In practice NP flow is low and even negligible but is found near the cooling unit discharge where air velocities are high, i.e., floor grilles closest to cooling units. High velocities from floor tiles can impair airflow to bottom servers.

Unlike BP and *R*, which are almost impossible to eliminate altogether, NP is not present in all data halls and is difficult to measure.

The case of NP applies to raised floor environments; however other examples of velocity pressure issues occur where velocity pressures are higher than static pressure set points (e.g., when controlling delta *P* between hot and cold aisles in contained solutions).

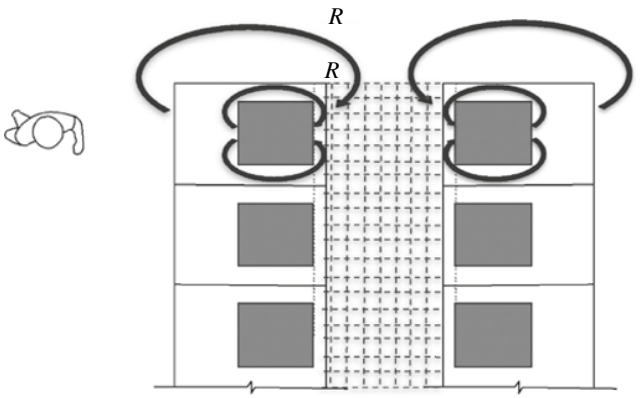


FIGURE 34.4 Recirculation flow in the data center, plan view. Source: Courtesy of Operational Intelligence.

FIGURE 34.3 Recirculation flow in the data center, section view. Source: Courtesy of Operational Intelligence.

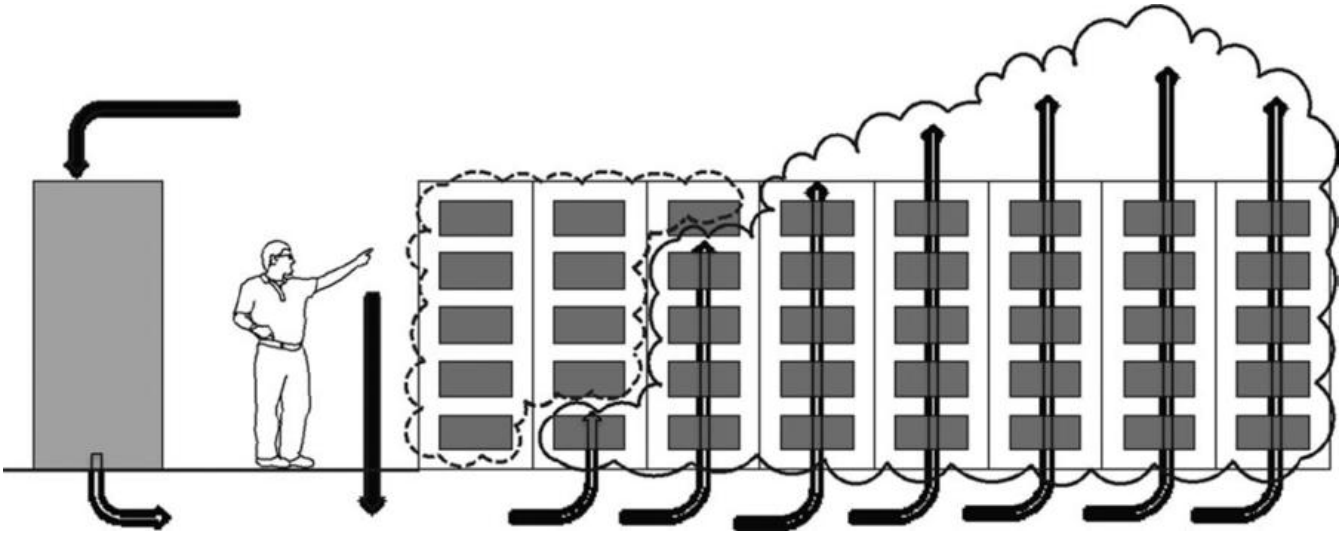


FIGURE 34.5 Cold aisle section showing negative and low pressure in the data center. Note floor grilles are located in front of each cabinet and where the hot air enters the raised floor. Source: Courtesy of Operational Intelligence.

$$P_v = 0.5\rho v^2$$

$$\text{Assuming } \rho = \frac{1.2\,\text{kg}}{\text{m}^3}$$

$$P_v = 0.6v^2$$

where

P_v = velocity pressure (Pa)
ρ = air density, (kg/m^3)
v = velocity (m/sec)

If velocity = 3 m/s, velocity pressure = 5 Pa, which is similar to the amount of static pressure. Therefore lower velocities are recommended.

In contained systems where air comes into the side of the data hall, high velocities may prevent sufficient airflow to be available at end racks in rows.

34.3 METRICS

The authors use four characteristic temperatures in each data hall: cooling unit and IT equipment inlet and outlet, to calculate bypass, recirculation, flow availability, and air segregation by applying mass flow balance equations [2]. Each hall has only one set of these metrics. These characteristic temperatures may be determined by using spot readings for all of the operating cooling units and a representative sample of IT equipment. The weighted average of the temperatures in terms of cold airflow rates, which are proportional to the IT loads, will provide the exact value of flow availability and a good approximation to BP, R, and ASE (air segregation efficiency, defined below). In practice a sufficiently representative approximation of the air performance of a hall can be achieved using average non-weighted temperatures.

Figure 34.6 shows the temperature readings required to model the data hall, NP, bypass, and recirculation.

Note the metrics apply to all types of cooling unit (shown as CRAH unit in the diagram) and both raised floor and non-raised floor environments (raised floor shown in diagram).

$$\Delta T_{max} = T_{io} - T_{co}$$

$$\Delta T_c = T_{ci} - T_{co}$$

$$\Delta T_i = T_{io} - T_{ii}$$

where

T_{ci} = air temperature entering cooling unit
T_{co} = air temperature leaving cooling unit
T_{ii} = air temperature entering IT equipment
T_{io} = air temperature leaving IT equipment
ΔT_c = cooling unit temperature difference
ΔT_i = IT equipment temperature difference
ΔT_{max} = maximum temperature difference

Measuring at the IT equipment inlet and outlet rather than at the rack captures recirculation and bypass issues within the rack as well as within the room. In the example shown in Figure 34.7, there is a 5 K uplift of temperature between the front of the rack and IT equipment inlet and a 7 K reduction between the IT equipment exhaust and back of rack.

The following equations are derived by applying mass flow and energy balance equations:

$$\text{BP} = \frac{T_{io} - T_{ci}}{\Delta T_{max}} = \frac{T_{io} - T_{ci}}{T_{io} - T_{co}}$$

where BP = bypass air flow.

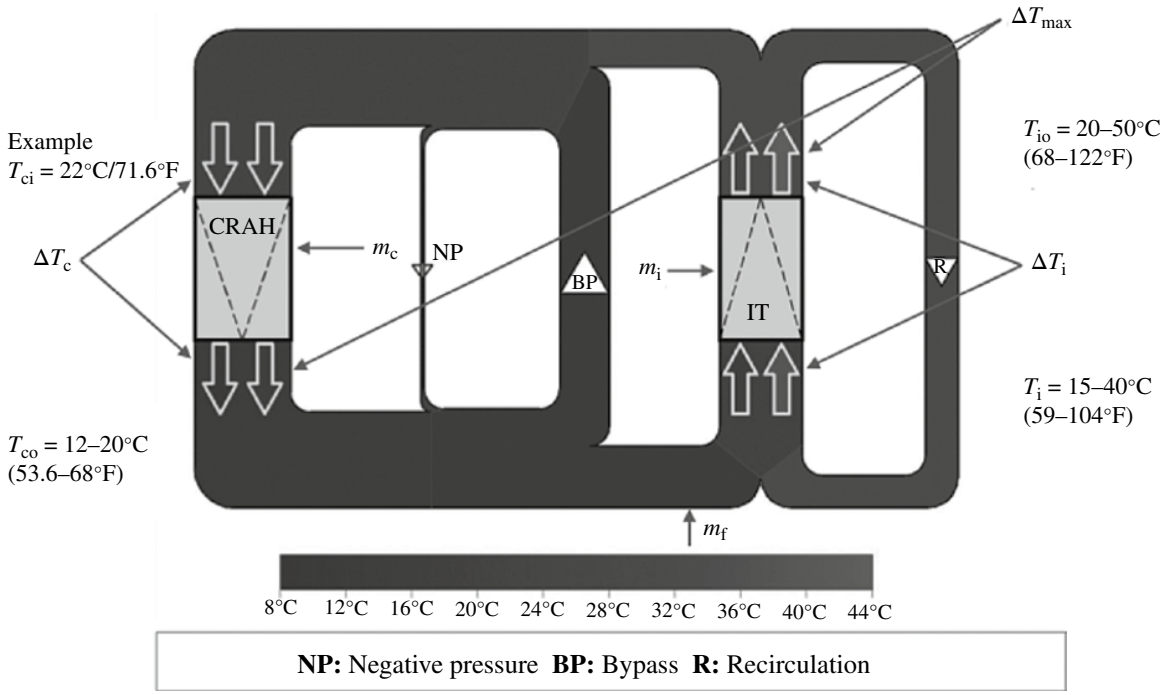


FIGURE 34.6 Data hall airflows in section for control on return air temperature with typical measured temperature ranges. Source: Courtesy of Operational Intelligence.

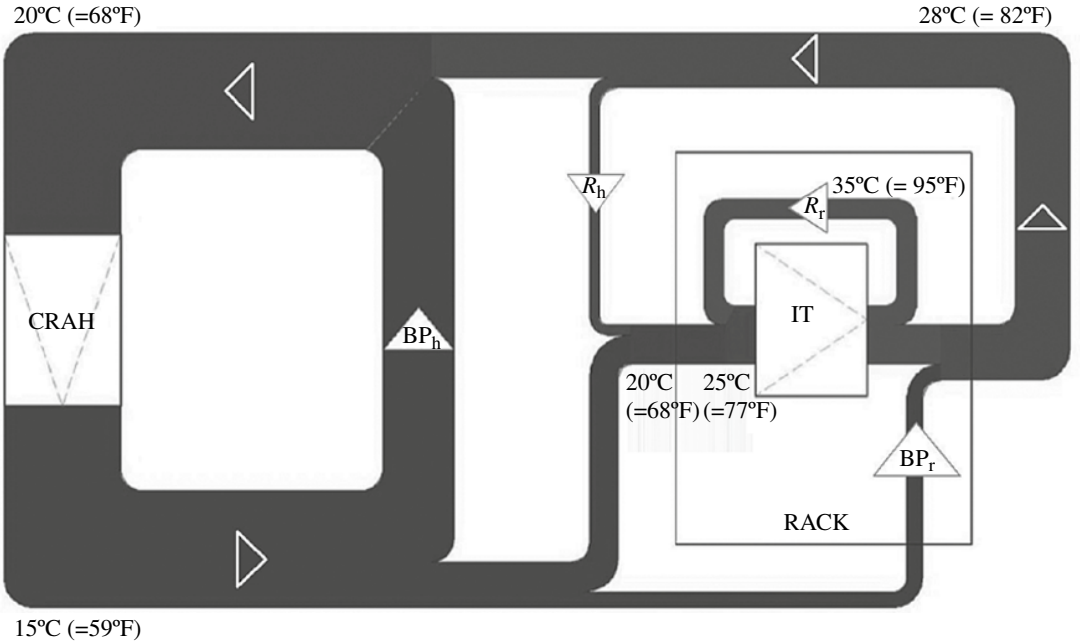


FIGURE 34.7 Data hall airflows in section example rack and IT equipment temperatures. Source: Courtesy of Operational Intelligence.

The amount of air not bypassed is defined as supply performance (η_{supply}); a data hall metric of the ratio of cooled air supplied to IT equipment to air produced by cooling units is calculated as the difference in temperature across the cooling units to the maximum temperature difference in the system, i.e., between the server exhaust air and the air supplied by the cooling units. Ideal supply performance = 1 (no bypass):

$$\eta_{supply} = 1 - \mathrm{BP} = \frac{m_f}{m_c} = \frac{\Delta T_c}{\Delta T_{max}} = \frac{T_{ci} - T_{co}}{T_{io} - T_{co}}$$

$$R = \frac{T_{ii} - T_{co}}{\Delta T_{max}} = \frac{T_{ii} - T_{co}}{T_{io} - T_{co}}$$

where

η_{supply} = supply performance
m_c = mass flow of air from cooling units
m_f = mass flow of cold air to the IT equipment
R = recirculation air flow to IT equipment

The amount of air not recirculated is defined as demand performance (η_{demand}), is a data hall metric of the ratio of cooled air supplied to IT equipment to the total air supplied to IT equipment, and is calculated as the temperature change across the IT equipment to the maximum temperature difference in the system. Ideal demand performance = 1 (no recirculation):

$$\eta_{demand} = 1 - R = \frac{m_f}{m_i} = \frac{\Delta T_i}{\Delta T_{max}} = \frac{T_{io} - T_{ii}}{T_{io} - T_{co}}$$

where η_{demand} = demand performance, m_i = mass flow of air to IT equipment.

Bypass and recirculation affect the amount of cold air that reaches the servers. How well air is being transported in the data center can be determined by comparing the amount of cooling required by the IT equipment and the amount that actually reaches the IT equipment, as opposed to the amount of cooling that is installed. Any air that is bypassed is not available for cooling IT equipment, effectively impacting the cooling capacity. As the supplied cooling is dependent on the amount of bypass and recirculation in the facility, the overall air performance can therefore be defined by the performance of the supply and demand airstreams.

NP is low in comparison and is picked up as a form of recirculation and bypass.

Availability of flow (A_f) is a data hall metric of the ratio of air supplied from the cooling units to the amount of air demanded by the IT equipment and shows whether sufficient air is getting to the IT equipment:

$$A_f = \frac{m_c}{m_i} = \frac{\Delta T_i}{\Delta T_c} = \frac{T_{io} - T_{ii}}{T_{ci} - T_{co}} = \frac{\eta_{demand}}{\eta_{supply}}$$

where A_f = air flow performance.

An A_f greater than 1 indicates an oversupply of air; a value less than 1 is an undersupply. An A_f of 1 represents perfect airflow, although in practice a slight positive pressure (ensuring limited bypass) between cold and hot airstreams is desirable to stop hot air recirculating back into the cold aisle. An A_f of 1.1 is therefore more desirable than an A_f of 0.9, where some servers would run at slightly higher temperatures. With A_f at 1.1, the CRAH fans would use additional energy; at 0.9 there will be more recirculation, and potentially some server fans would be operating at higher speed. Altering fan speeds or adding cooling units can improve A_f, which is often high when a design has not been optimized for part load [3].

ASE is concerned with the effectiveness of hot and cold airstream separation in the data center [4]:

$$\text{ASE} = \sqrt{\frac{(1-\text{BP})^2 + (1-R)^2}{2}}$$

The metric allows operators to monitor air management independently of IT load and cooling unit fan speeds and may be used as a management tool.

These metrics can be plotted on a graph. Figure 34.8 represents the air management of a data center where a facility in the top right hand corner has ideal performance; there is complete separation of the hot and cold airstreams, and bypass and recirculation are both zero.

When improvements are made to air management and bypass and recirculation are reduced, the characteristic point on the air performance diagram moves diagonally toward the top right corner. A_f remains constant unless the flow rates of the IT equipment or cooling units change, when the improved state will move further along the same A_f line.

Conversely, if no changes are made to air management, but the airflow rate of the cooling units (or IT equipment) is changed—by changing the fan speed or changing the number of operating units—the characteristic point will move at 90° to the A_f line in an arc.

There is an assumption that at different BP and R values the differences in the resistances affecting the airflow paths are not sufficient to impact IT equipment or cooling unit fan flow rates.

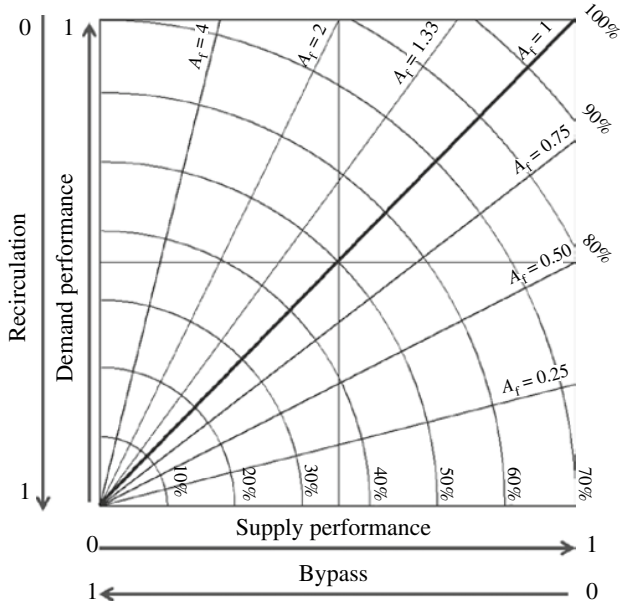


FIGURE 34.8 Air performance graph. Source: Courtesy of Operational Intelligence.

The metrics described above can be calculated using Operational intelligence's air performance tool[1] by inputting the four characteristic temperatures. This also produces a diagram like that shown in Figure 34.6, which can help users to understand the current performance of their facility.

The equations above describe overall performance of the data hall based on averages. Similar metrics have been created by others:

SHI & RHI [5] measure the extent of cold and hot air mixing at the rack and cooling unit due to recirculation based on average rack inlet and outlet and cooling unit outlet temperatures:

$$\mathrm{SHI} = \frac{\sum_{\mathrm{i}}\sum_{\mathrm{j}}\left(T^{\mathrm{r}}_{\mathrm{in}_{i,j}} - T^{\mathrm{C}}_{\mathrm{sup}}\right)}{\sum_{\mathrm{i}}\sum_{\mathrm{j}}\left(T^{\mathrm{r}}_{\mathrm{out}_{i,j}} - T^{\mathrm{C}}_{\mathrm{sup}}\right)}$$

where

$T^{\mathrm{r}}_{\mathrm{in}}$ = rack intake airflow temperature,
$T^{\mathrm{c}}_{\mathrm{sup}}$ = supply airflow temperature from CRAC unit
$T^{\mathrm{r}}_{\mathrm{out}}$ = airflow temperature exhaust from rack.

$$\mathrm{RHI} = 1 - \mathrm{SHI}$$

If the temperatures above related to IT equipment rather than racks, SHI would be the same as R and RHI demand performance.

RTI is a measure of the balance of airflows between cooling units and IT equipment, the result being bypass or recirculation:

$$\mathrm{RTI} = \frac{T^{\mathrm{C}}_{\mathrm{ret}} - T^{\mathrm{C}}_{\mathrm{sup}}}{T^{\mathrm{r}}_{\mathrm{out}} - T^{\mathrm{r}}_{\mathrm{in}}} \cdot 100[\%]$$

where $T^{\mathrm{c}}_{\mathrm{ret}}$ = airflow return temperature to CRAC unit and $T^{\mathrm{r}}_{\mathrm{in}}$ = airflow temperature inlet to rack.

If the temperatures above related to IT equipment rather than racks, RTI would be equal to $\frac{1}{A_{\mathrm{f}}} = \frac{m_{\mathrm{i}}}{m_{\mathrm{c}}}$

Other metrics deal with deviations from acceptable values [6]. RCI is a measure of compliance of rack intake temperatures with respect to recommended limits (ideal = 100%):

$$\mathrm{RCI_{Lo}} = \left[1 - \frac{\sum_{i=1}^{n}\left(T_{\mathrm{LO-rec}} - T^{\mathrm{r}}_{\mathrm{in}_i}\right)}{n \cdot \left(T_{\mathrm{LO-rec}} - T_{\mathrm{LO-allow}}\right)}\right] \cdot 100[\%] \text{ if } T^{\mathrm{r}}_{\mathrm{in}_i} < T_{\mathrm{LO-rec}}$$

$$\mathrm{RCI_{Hi}} = \left[1 - \frac{\sum_{i=1}^{n}\left(T^{\mathrm{r}}_{\mathrm{in}_i} - T_{\mathrm{HI-rec}}\right)}{n \cdot \left(T_{\mathrm{HI-allow}} - T_{\mathrm{HI-rec}}\right)}\right] \cdot 100[\%] \text{ if } T^{\mathrm{r}}_{\mathrm{in}_i} > T_{\mathrm{HI-rec}}$$

[1] https://dc-oi.com/air-performance-tool/.

where
$T_{\mathrm{LO-allow}}$ and $T_{\mathrm{HI-allow}}$ = allowable temperature values
$T_{\mathrm{LO-rec}}$ =recommended temperature values, low
$T_{\mathrm{HI-rec}}$ = recommended temperature values high.

CFD (computational fluid dynamics) can be used as a tool to identify issues and calculate metrics by creating a virtual model of the data hall and running a simulation with the relevant input parameters. Building and calibrating the model and keeping it updated take time and resources; for an existing facility, the same conclusions may be reached quickly by measuring sample temperatures and calculating the metrics described above. For more details please see Chapter 30 on CFD.

34.4 AIR CONTAINMENT AND ITS IMPACT ON AIR PERFORMANCE

Air containment helps to segregate hot and cold airstreams and limits the opportunities for bypass and recirculation, providing opportunity to only supply the necessary airflow and to raise cooling unit set points. Several types of containment exist, described in Figures 34.9–34.12.

Cold aisle containment encloses the cold aisle with a roof and doors, normally fabricated from flame-retardant plastic. The rest of the data hall is the same temperature as the hot aisle. Semi-containment is a variation on this, where curtains (again in flame-retardant material) are fitted above the cold aisle, blocking the air path from the hot aisle. This works well as a retrofit option, particularly where there are cabinets of different heights. Hot aisle containment ducts the air from the hot aisle back to the cooling units, usually by way of a ceiling plenum. The rest of the data hall is the same temperature as the cold aisle. This works best for a new build as it requires coordination with other overhead services, such as cable trunking. Direct rack containment employs special

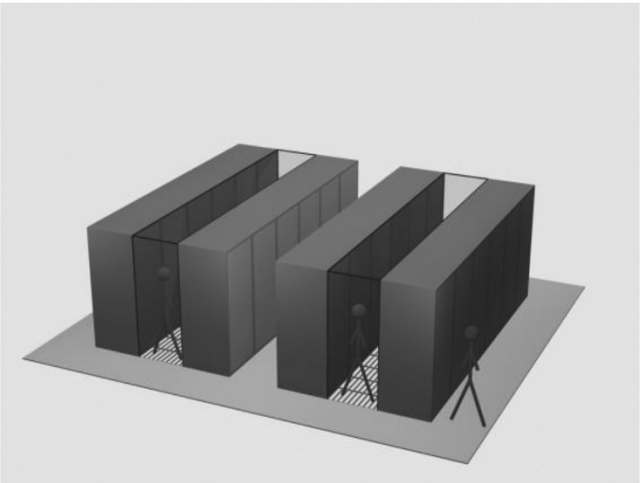


FIGURE 34.9 Cold aisle containment. Source: Courtesy of Operational Intelligence.

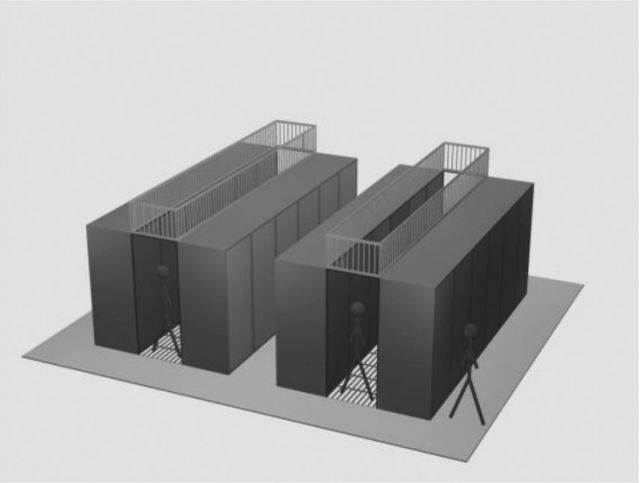


FIGURE 34.10 Cold aisle semi-containment. Source: Courtesy of Operational Intelligence.

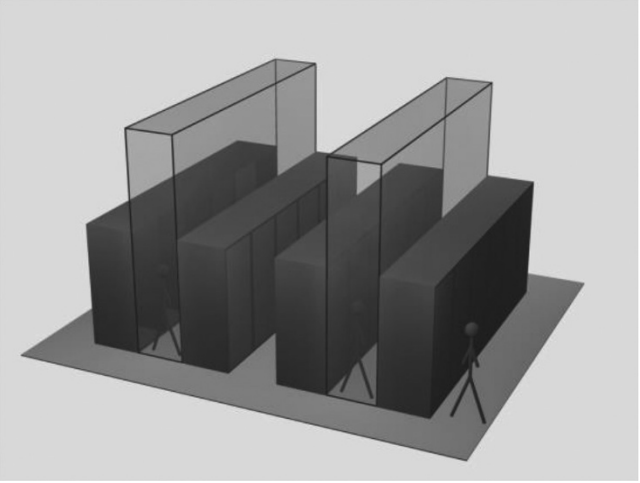


FIGURE 34.11 Hot aisle containment. Source: Courtesy of Operational Intelligence.

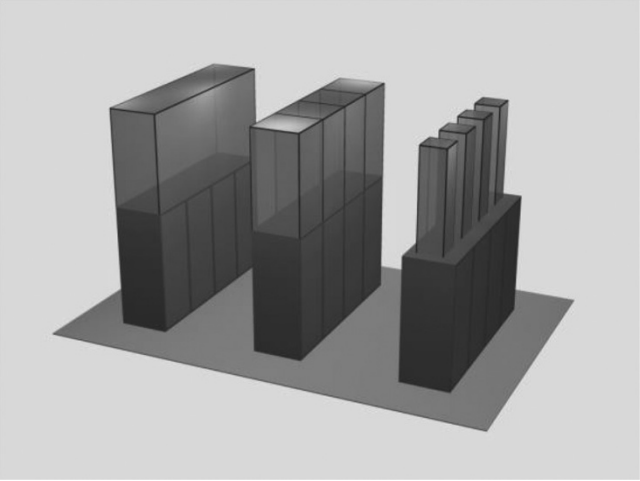


FIGURE 34.12 Direct rack containment. Source: Courtesy of Operational Intelligence.

deeper cabinets that include a chimney at the back to duct hot air back to the cooling units. Variations include ducted rack roofs and row plenums. This method keeps the hot air outside of the room and may become more widely adopted as hot aisle temperatures increase due to increasing IT equipment exhaust temperatures and air supply temperatures.

Each solution has different advantages and disadvantages and is usually chosen based on cost of implementation and facility restrictions (when installed post fit-out) as shown in Table 34.1.

Segregation of hot and cold airflows covers the total area of separation, which extends beyond the containment installation. The efficiency of segregation is dependent on the quality of the containment installation and also how well any other points of leakage are sealed. Figure 34.13 shows ideal segregation of hot and cold airstreams using cold aisle containment, and Figure 34.14 illustrates leakage. There are many potential gaps in a containment system that allow cold and hot air to mix; a large number of small gaps can add up to a large area.

Figure 34.15 shows a scatter of air performance results from a range of facilities [4]. Blue circular points refer to non-contained legacy data halls, and red diamond points to data halls with containment. The yellow triangle point indicates the average result.

Supply performance (1 − BP) differs greatly between the facilities; however, the demand performance $(1-R)$ results are more clustered into the top half of the chart. In general, the contained systems (red diamond points) tend to have less recirculation than the average. In some cases the bypass and flow availability displayed by contained solutions are worse than the average, although recirculation is reduced. The majority of facilities are oversupplying air, as shown by values of $A_f > 1$, but in 15% of cases, air is being undersupplied. Today it is not uncommon to achieve ASE of around 90% with $A_f = 1.1$, particularly in hyperscale facilities.

34.5 IMPROVING AIR PERFORMANCE

Once the performance of air management in the data center has been modeled, the operator is better informed to target reduction of energy and risk. It also provides a basis for benchmarking and comparison of different data centers and improvements made within individual facilities. It can also provide evidence for the effectiveness of measures implemented to improve efficiency and can be used to estimate energy savings and return on investment.

High bypass can be improved by:

- Ensuring floor grilles are located in the cold aisle and not the hot aisle
- Ensuring supplied air speeds and flow are not so high that the air flows over the top of the racks
- Blocking cable cutouts in the floor/in top of rack

TABLE 34.1 Advantages and disadvantages of different containment solutions

Containment type	Advantage	Disadvantage
Cold aisle containment	Easier to fit in existing hot/cold aisle configurations	Most of the hall is exposed to hot aisle temperature Lighting and fire detection/suppression may need modifying Requires raised floor
Cold aisle semi-containment	Inexpensive Can be adapted to suit existing fit-out, e.g., where cabinets have different heights	Bypass air is not fully eliminated Requires raised floor
Hot aisle containment	Only the contained aisle is hot Does not require a raised floor	Harder to coordinate with overhead services in existing facilities High temperatures when working in the hot aisle Requires return plenum
Chimney exhaust containment	No hot air in the data hall Does not require a raised floor	Harder to coordinate with overhead services in existing facilities Limited rack choice—possible densities dependent on size of chimney in relation to rack(s) Requires return plenum

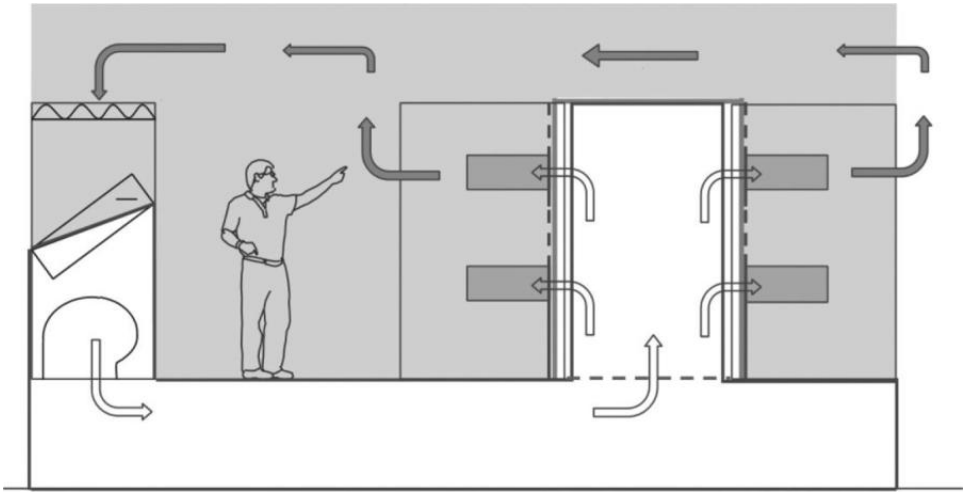


FIGURE 34.13 Segregation of hot and cold air with cold aisle containment. Source: Courtesy of Operational Intelligence.

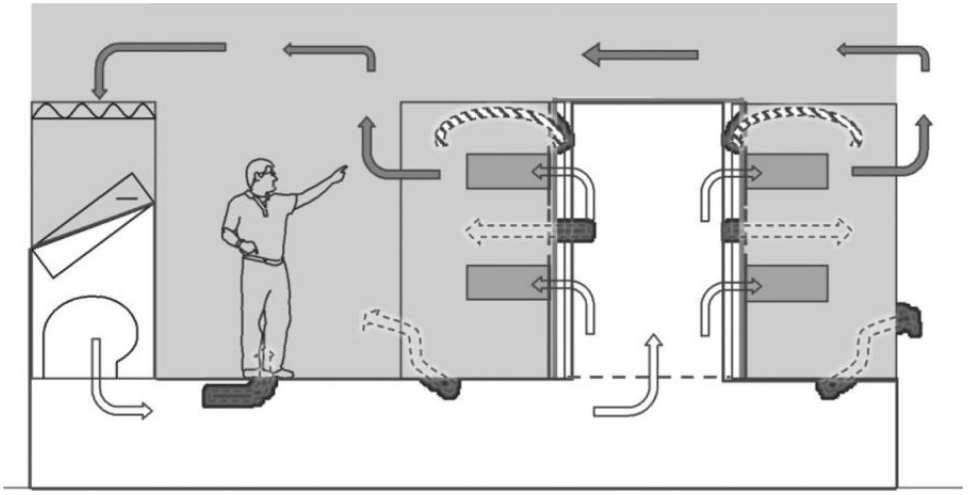


FIGURE 34.14 Leakage between cold and hot airstreams with a cold-aisle-contained system. Source: Courtesy of Operational Intelligence.

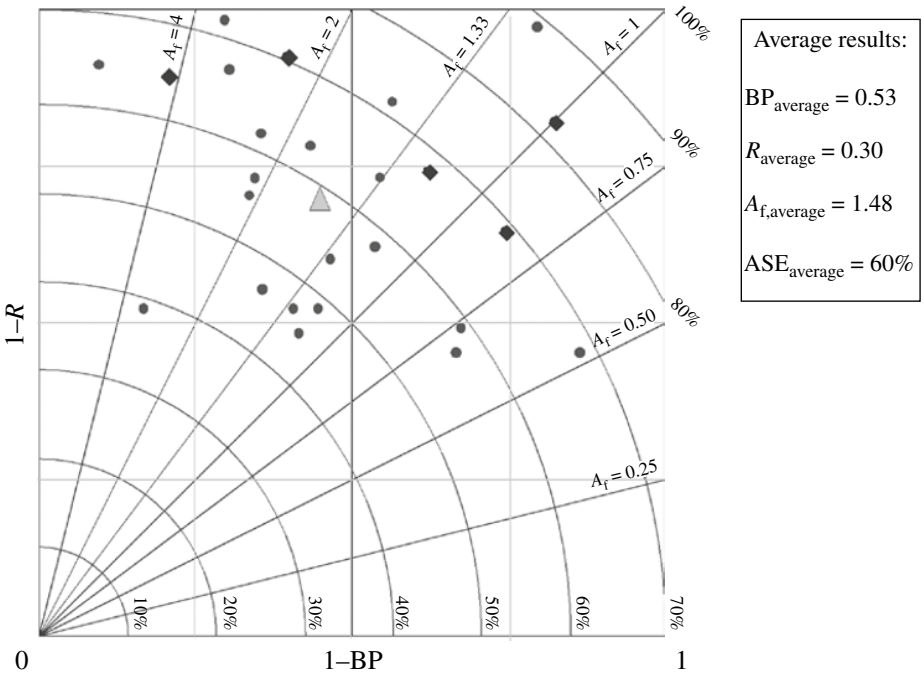


FIGURE 34.15 Air performance in a range of data halls from 2010 to 2014. Source: Courtesy of Operational Intelligence.

- Sealing gaps between racks in contained cold or hot aisle arrangements
- Blocking gaps at bottom of racks
- Replacing servers that are off with blanking panels
- Gaps in ceiling including grilles

High recirculation can be improved by:

- Ensuring sufficient cooled air reaches the IT equipment; for contained systems cold aisle should be slightly pressurized
- Installing rack blanking plates and blocking side gaps in racks
- Ensuring equipment with different airflow requirements are not located in the same racks
- Using rack doors with high free area

NP can be improved by:

- Avoiding high air velocities under floor grilles, for example, through ensuring minimal cable obstructions in the raised floor
- Avoiding high velocities in contained system to maintain velocity pressure below static pressure

Note that in contained systems where there is NP between cold and hot aisles, some of the bypass improvements would improve recirculation due to the reverse airflow direction, hot leaking into cold rather than the other way round.

Many of these best practices and others can be found in the EU Code of Conduct for Data Centers Best Practice Guidelines [7] and EN50600 TR 99-1 [8].

The improvement of air performance can be quantified by conducting a survey collecting sample temperatures before and after these recommendations have been implemented. The characteristic point on the flow and thermal performance plot should move toward the top right of the chart, with flow and thermal performance values of above 0.9 achievable where best practice measures are fully implemented.

Controlling on return air temperature at the cooling unit, even with the measures above in place, can still result in a range of temperatures delivered to the IT equipment. This is due to the nonuniform nature of load distribution in most data halls; each cooling unit will deliver a delta T proportional to its load and thus supply a different temperature. Changing the cooling unit temperature control strategy to supply air control allows this range to be minimized. This can be retrofitted on many cooling units with an additional sensor.

34.5.1 Case Study 1

A global financial services firm made energy efficiency improvements in their legacy data center environments in the United Kingdom [9]. The facility power usage effectiveness (PUE) ratio was 2.3 at the start of the program and was reduced by 34% to 1.49, resulting in significant energy and operational cost reductions and enhanced rack densities and

system capacities. This was achieved through an energy assessment and data hall air temperature survey, operator education workshops, implementation of air management improvements, optimization of cooling unit fan control, a gradual increase in air and chilled water temperature set points (to ensure energy savings did not compromise reliability), and installation of a free cooling circuit. In this case, 80% of the CRAH air was being bypassed, and around 20% of server intake air was recirculated warm air. The flow availability was 4. All server air intake temperatures measured were between 15 and 27°C (59.0 and 80.6°F).

The following measures were implemented to improve air management:

- Installation of blanking plates within cabinets
- Sealing gaps in the raised floor, including cabinet cable cutouts and gaps around power distribution unit (PDU) bases
- Revised placement of floor grilles to where cooling required

To help separate the hot and cold airstreams, a temporary solution was installed in one aisle as proof of concept with doors at the end of the cold aisle and flame-retardant curtains above the racks. Effective blanking within the racks demonstrated significant improvements. The installation of temporary doors resulted in a reduction of approximately 5 K (9°R), and increasing the height of the aisle with curtains resulted in a reduction of about 4 K (7.2°R) at localized server inlets. The success of this trial led to it being deployed throughout the data hall. The total costs of deployment were recovered in less than a year. The air performance metrics are used on an ongoing basis to monitor the performance of the environment and ensure that as equipment is installed and decommissioned, air management best practice is continually employed, and standards are maintained.

The air management improvements meant that the air temperature delivered to servers was controlled within a narrower band and it was possible to modify the CRAH units to reduce fan speeds from 100 to 60% and consider increasing temperature set points. Changing the CRAH unit control strategy from return to supply air control also helps to maintain the air supplied at the server inlet within a narrow range.

The cooling unit set points changed from 24°C (75.2°F) on return air to 23°C (73.4°F) on supply air, with an associated increase in chilled water temperature set points from 6°C/12°C to 17.9°C/23.9°C (42.8°F/53.6°F–64.2°F/75.0°F). Set points were increased by 1°C (1.8°F) at a time over a period of several months. The server inlet temperatures were monitored and recorded before and after every stage to ensure there was no negative impact and allow a rollback if needed. The result was an increase in the overall coefficient of performance (COP) of the chilled water systems and chiller delta *T*. The increased operating temperatures opened up the opportunity for free cooling operation. An economizer circuit allowing free cooling with dry coolers was installed to minimize refrigeration hours, which has reduced further their operating costs. The operator has also since installed a proprietary wireless monitoring and dynamic control system to reduce CRAH fan speeds further via floor pressure control. This adjusts CRAH fan speeds in the zone of influence by using a number of floor pressure sensors to maintain a minimum flow pressure and ensure good airflow distribution and rack inlet sensors to check temperatures. Before/after: BP 80%, *R* 20%, A_f 4/BP 34%, *R* 25%, A_f 1.12. Work is ongoing to reduce bypass and recirculation further.

34.5.2 Case Study 2

A hyperscale operator in Europe with a low energy design discovered that poor air performance was causing a higher than anticipated cooling unit operating cost. Although hot aisle containment was installed, this was not optimized, and a large area of gaps were present. The cooling unit fans control on differential pressure (dp), and the high leakage with the high dp set point meant that the fans did not reduce their operating speed. Improvements to air management reduced the leakage, and the fan speeds automatically reduced as changes were implemented. This led to energy savings of over €1 million/year. Before/after: A_f 4.3 BP 79%, *R* 11%/ASE 90%, A_f 1.1.

Figure 34.16 reflects the results of two case studies.

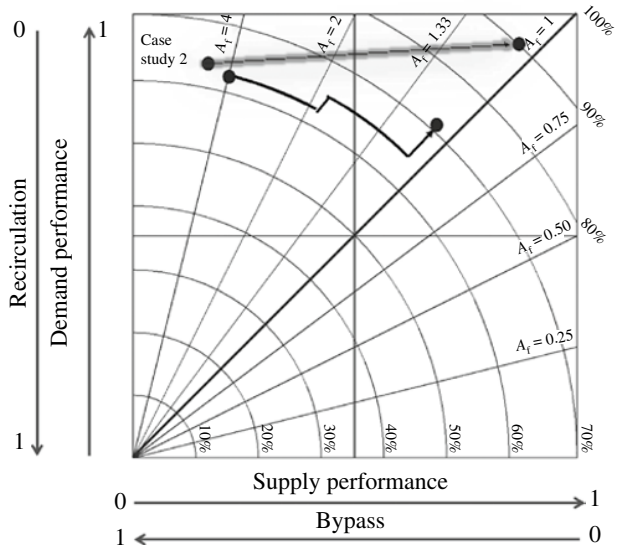


FIGURE 34.16 Results of case studies. Source: Courtesy of Operational Intelligence.

34.6 CONCLUSION

Management of air in the data hall is essential to achieve energy savings and maintain reliable operation. Improving air performance requires measures to optimize bypass, recirculation, availability of flow, and ASE. The metrics described in this paper allow issues to be diagnosed and performance to be quantified, benchmarked, and monitored via a simple low-cost methodology.

REFERENCES

[1] *Thermal Guidelines for Data Processing Environments*. 4th ed. ASHRAE; 2015.

[2] Tozer R, Salim M, Kurkjian C. Air management metrics in data centres. ASHRAE Trans 2009;115, Part 1.

[3] Flucker S, Tozer R. Scalable Data Centre Efficiency. CIBSE Technical Symposium; 2013; Liverpool.

[4] Tozer, R, Whitehead, B, Flucker, S. Data center air segregation efficiency. ASHRAE Trans 2015;121, Part 1.

[5] Sharma et al. Dimensionless parameters for evaluation of thermal design and performance of large scale data centers. 8th ASME/AIAA Joint Thermophysics and Heat Transfer Conference; 2002.

[6] Herrlin M. Airflow and cooling performance of data centers: two performance metrics. ASHRAE Trans 2008;114, Part 2.

[7] European Commission. Best Practices for the EU Code of Conduct on Data Centres; 2020. Available at https://e3p.jrc.ec.europa.eu/communities/data-centres-code-conduct. Accessed 07/01/21.

[8] PD CLC/TR 50600-99-1:2020 Information Technology. Data Centre Facilities and Infrastructures. Recommended Practices for Energy Management, BSI; 2020.

[9] Tozer R, Flucker, S. Data center energy-efficiency improvement case study. ASHRAE Trans 2015;121, Part 1.

35

ENERGY EFFICIENCY ASSESSMENT OF DATA CENTERS USING MEASUREMENT AND MANAGEMENT TECHNOLOGY

Hendrik Hamann, Fernando Marianno and Levente Klein
IBM TJ Watson Research Center, Yorktown Heights, New York, United States of America

35.1 INTRODUCTION

The energy efficiency of a data center (DC) is determined by implementation of *best practices*. Professional organizations like The American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) [1], Uptime Institute [2], and Green Grid [3] are providing recommendations to improve energy efficiency of existing facilities or to integrate energy efficiency in newly designed DCs. One of the major costs to operate data centers is cooling power to maintain safe operating environment for servers and IT equipment (ITE). There are currently few methodologies available for quantitative evaluation of the degree of best practice application to enable data center owners and operators to improve energy efficiency. A comprehensive method to quantify energy efficiency opportunities is presented; the method provides an understandable set of best practice metrics, as well as simple guidance to the data center operators and owners. The methods described can easily empower operators to rapidly translate general advice into actual cost-effective customized solutions.

Specifically, we describe here a new monitoring solution based on spatial and temporal benchmarking of data center using the measurement and management technology (MMT) [4–7] to drive toward quantitative measurement-driven DC best practice implementation. The DC is thermally characterized either by (i) three-dimensional temperature and humidity scans that generate a base map or (ii) using wireless sensor network that captures the spatial–temporal dynamics of temperature and humidity changes.

Through multiple deployment, it has been established that the minimum required key parameters to be measured in a data center are thermal maps, air-conditioning unit (ACU) utilization, and airflow (Table 35.1). These measurements can be assembled in simple metrics that are summarized below.

Once the metrics are measured or estimated, they can be associated with a clear set of solutions, which are readily available to operators for data analytics. By improving on each of these metrics, the data center operator can track systematically the progress toward a more energy-efficient DC.

35.2 ENERGY CONSUMPTION TRENDS IN DATA CENTERS

Power consumption in data centers is largely governed by the ITE power consumption and the power spent to maintain cooling that enable safe operating conditions for ITE. Manufacturer of ITE equipment have recommendations about safe operating ranges that encompass maximum temperature that processors and memory elements can be operated, while avoiding critical situations like condensations or overheating.

A comprehensive study from the Lawrence Berkeley National Laboratory (LBNL) [8, 9] indicates that server-driven power usage amounted to 1.8% (70 billion kWh) of the total U.S. energy consumption. A commonly used metric for energy efficiency called power usage effectiveness (PUE) given by the ratio of IT power (P_{IT}) to the total data center power

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

TABLE 35.1 Minimum required environmental parameters to quantify a data center energy efficiency potential

1. **Server rack's inlet temperature uniformity** (IH) a. Horizontal temperature uniformity/horizontal hotspots (HH) b. Vertical temperature uniformity/vertical hot pots (VH)
2. **Targeted airflow** (TF)
3. **Plenum temperature** (PT)
4. **Air-conditioning unit (ACU) utilization** (UT)
5. **ACU flows** (FL)

(P_{DC}): $\eta_{DC} = P_{IT}/P_{DC}$ is used to compare DCs performances [10]. The efficiencies of DCs can significantly vary [6], which suggest that sizeable energy saving opportunities exist [11, 12]. In newly designed data centers, the PUE can be significantly reduced by optimizing server's rack layout, with PUE metrics reported to go as low as 1.1 [13]. Understanding and improving the energy efficiency of existing data center are critically important from a cost as well as sustainability [14] perspective, and this chapter addresses these challenges.

The flow of electricity from the electric grid to various parts and components in a typical data center is shown in Figure 35.1. The total power for the DC facility (P_{DC}) is split using switch gear equipment (P_{switch}) into a path for IT and other mission-critical equipment and a path for the support of the IT. The supporting path is used for the ACU blowers (P_{ACU}), the power associated with humidity control (P_{hum}), the power for lights and other office equipment in the data center (P_{lights}), the power for pumping chilled water from the ACU to the chiller (P_{AC}) and from the chiller to the cooling tower (P_{CT}), and chiller power ($P_{chiller}$) for the compression cycle. The IT power path is further conditioned via an uninterruptible power supplies (UPS), which supports the mission-critical equipment in the data center such as the IT equipment and other critical equipment (P_{other}). The power from the UPS is distributed within the data center via the power distribution units (PDUs) (P_{PDU}) to the IT equipment (P_{IT}). We note that all the electrical power is eventually converted into heat, which is then rejected to the environment and controlled by using the facilities cooling system. Despite good understanding of data center operations, today most DC managers have only some generic knowledge about the fundamentals of DC best practices, but it is (naturally) a very different challenge to relate this general type of advice to the specific context of their specific environment as every DC is unique [15–18]. Detailed measurable metrics for DC best practices acquired in real time can provide the recommendations for data center operators to implement these solutions in their specific environment.

35.3 COOLING INFRASTRUCTURE IN A DATA CENTER

Figure 35.2 displays a schematic of the cooling energy flow for a data center facility showing the electrical heat energy dissipated by the IT equipment being carried away by

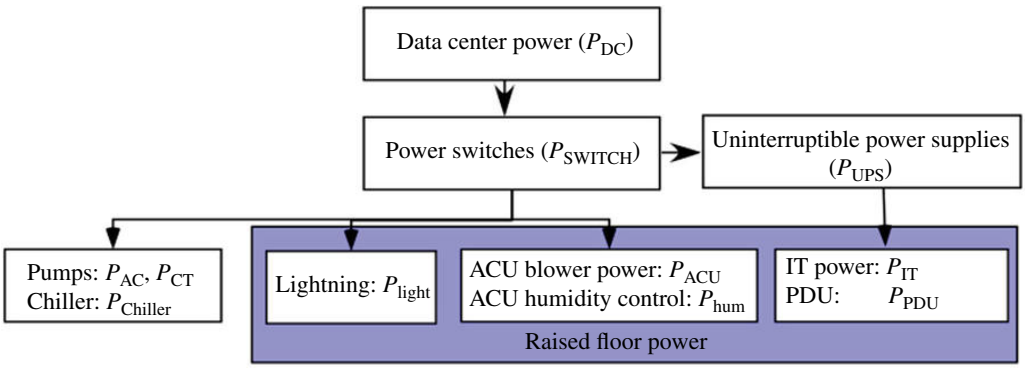


FIGURE 35.1 Electricity flow in a typical data center.

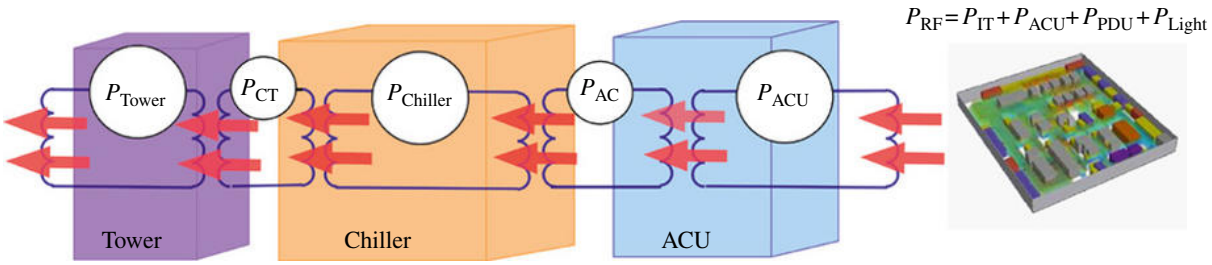


FIGURE 35.2 Heat rejection path via cooling infrastructure in a data center.

successive thermally coupled coolant loops that consume energy either due to pumping (transport) or (thermodynamic) compression work. The cooling system is made up of three elements: the refrigeration chiller plant (including the cooling tower fans and condenser water pumps, in the case of water-cooled condensers), the building chilled water pumps, and the data center floor ACUs.

Air-Conditioning Units (ACUs): All the power consumption of the raised floor (P_{RF}) used by the IT equipment and the supporting infrastructures (PDU, ACU blowers, lights) is released into the surrounding as heat, which places an enormous burden on the cooling system. Existing cooling technologies utilize air to carry the heat away from the server racks and reject to the ambient. This ambient environment in a typical DC facility is an air-conditioned room, a small section of which is depicted in Figure 35.3. The ACUs take hot air in (typically from the top) and reject cooled air into a plenum. In Figure 35.3 the racks are using front-to-back cooling and are located on the elevated floor with a subfloor underneath. In a well designed data center, the racks are arranged in a hot aisle–cold aisle configuration, thus having alternating directions with inlets and outlets. The cold air is blown through perforated tiles from the subfloor (or underfloor plenum) into the cold aisles. The cold air is then sucked into the racks on the inlet side and dumped on the outlet side of the rack in the hot aisle. The ACUs usually receive chilled water from a refrigeration chiller plant, which is in turn often cooled at its condense using cooling tower cooled water. As such, in most data centers, the ACUs are simple heat exchangers utilizing blower power (P_{ACU}) needed to inject the chilled air into the plenum floor.

The Chiller Plant: The chiller plant is made up of the refrigeration chiller, the cooling tower, the cooling tower pumps and blowers, and the building chilled water pumps. The chiller itself comprises two heat exchangers, connected in a loop that also contains a compressor for refrigerant vapor compression and a throttling valve for refrigerant liquid expansion. One of these heat exchangers condenses the refrigerant vapor, and the other one heats the refrigerant from liquid to vapor phase. The condenser is cooled by water that is circulated through a cooling tower by a pump and loses its heat to the ambient air that is blown through the cooling tower. Heat is exchanged from the warm water to the cooler air primarily by evaporative mass transfer. The evaporator thermally couples the building chilled water loop to the refrigerant loop and allows the exchange of heat from the water to the refrigerant.

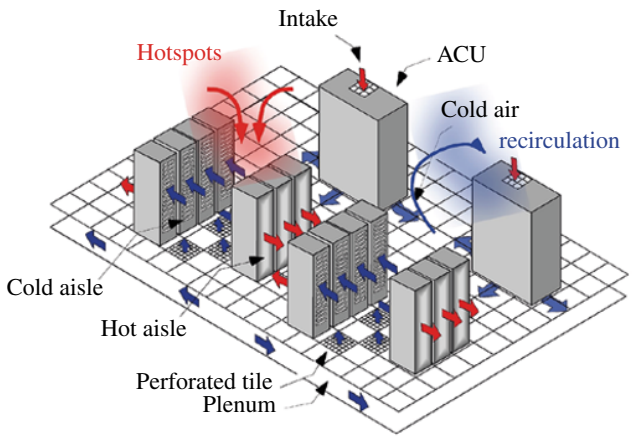


FIGURE 35.3 Thermodynamic and transport contribution of cooling power consumption.

35.4 COOLING ENERGY EFFICIENCY IMPROVEMENTS

To improve the energy efficiency in data centers, *two* types of cooling power need to be considered. (i) The first type is associated with the costs to generate the cooled air (*thermodynamic*), and (ii) the second type associated with the delivery of the cold air (*transport*). To a first order (neglecting the pump power ($P_{AC} = 0$ and $P_{CT} = 0$) and the fan power in the cooling tower), the thermodynamic part of the cooling energy is determined by the chiller ($P_{chiller}$), while the transport term is given by the blower power (P_{ACU}) of the ACUs.

Figure 35.3 illustrates how data center layout can potentially increase energy efficiency as *hotspots* within the raised floor (here caused by intermixing of the cold and hot air) can increase the inlet air temperatures to the respective server racks. An example for such intermixing could be a violation of the hot/cold aisle concept. In order to compensate for these hotspots, data center managers often chose excessive low chiller temperature setpoint, which increases *significantly* the thermodynamic cooling cost at the chiller ($P_{chiller}$). As a result of low temperature setpoint in the data center, the ACUs are just circulating air without reaching the inlets of the racks, in which case ACUs consume blower power (P_{ACU}) without contributing to the cooling of the DC. Simply rearranging the data center layout like placing server's inlets facing each other and separating cold from hot isle can minimize air mixing and reduce the required airflow resulting in significant energy cost reductions.

Chiller Power Optimization: The power consumption of the chiller is governed by four dominant factors: (i) temperature setpoint of chilled water leaving the evaporator to provide cooling for the ACUs, (ii) the loading factor that is the ratio of the operating heat load to the rated heat load, (iii) the temperature of the water entering the condenser from the cooling tower, and (iv) the energy used in pumping water and air in both the building chilled water and the cooling tower loops, respectively [19]. One adjustable parameter is the chilled water setpoint temperature; the temperature setpoint of chilled water can be increased, thereby saving

thermodynamic energy of the chiller. Specifically, it has been reported in the literature [19, 20] that a 0.5°C increase in this setpoint value results in approximately a 0.6–2.5% increase in the chiller efficiency. The typical value of 3.3%/°C [19] can be used to estimate the energy savings by raising DC setpoint.

Air-Conditioning Units: In order to reduce the transport term of the cooling energy in a DC, the ACU blower power can be minimized. If the ACUs are equipped with a variable frequency drive (VFD), blower power can be saved continuously by just simply throttling the blower with significant energy improvements [21]. In most cases, ACU blowers cannot be controlled, and blower power savings come from turning off the ACUs that do not contribute to effective cooling. Automated ACU control by turning on/off based on sensor reading is discussed in Chapter 10.

35.5 MEASUREMENT AND MANAGEMENT TECHNOLOGY (MMT)

A key component of improving the data center efficiency is the ability to rapidly survey and monitor a DC. IBM developed a proprietary MMT for systematic, rapid three-dimensional mapping of a data center collecting the relevant physical parameters and establishing a benchmark measurement of the thermal condition of the data centers. MMT provides a spatially dense three-dimensional thermal snapshot of DCs; however such scans are sporadic in time. A second implementation of MMT is a sensor network, based on a wireless communication protocol where sensors are dispersed to stream data within the data center.

35.5.1 Measurement and Management Technology Scanner

The MMT is one of currently available methods to rapidly measure the full 3D temperature distribution of a DC [22–24]. Figure 35.4a shows a prototype of this technology where a frame with sensors installed at different heights is carried by a mobile robot that moves automatically in the data center and acquires data above each tile that comprise the data center floor layout. The robotized system can create a complete 3D thermal profile for a data center not only around servers but also in the isles that separate the server rows. A 2,000 ft^2 data center can be scanned in 2 hours. The sensing platform has a footprint of the size of a standard tile (2 ft × 2 ft) used to cover the raised floor of data centers. It samples the temperatures at multiple points above each tile, where the thermal and humidity sensors are located at 1 ft height difference starting at 0.5 ft all the way up to 6.5 ft. The collected temperature data are transformed into a 3D thermal map of the data center, which provides the vital information needed to pinpoint trouble spots that indicate cooling inefficiencies, to facilitate better air-conditioning schemes, and to manage the energy consumption of the data center. Although the thermal profile of a data center may change over time due to the changes in IT power level, cooling condition, number of servers and racks, etc., the MMT technique enables a high-density collection of temperature data for basic modeling of the data center thermal profile.

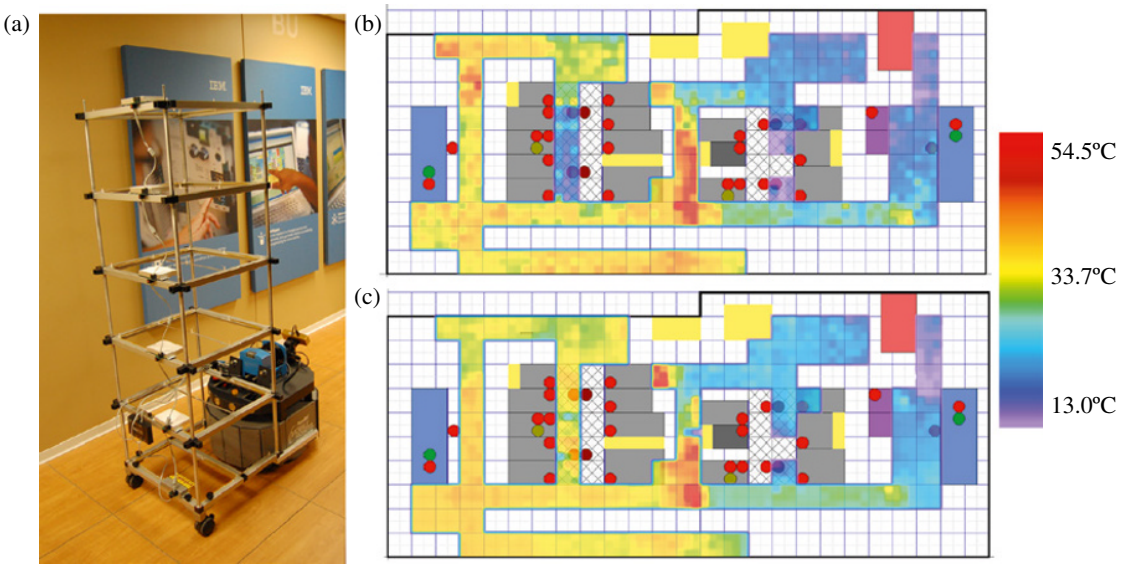


FIGURE 35.4 (a) Robotized measurement and management technology scanner. (b) Example temperature data at 0.5 ft height. (c) Temperature data at 5.5 ft height showing hotspots in a data center.

The data is automatically post-processed in junction with the layout, flow, ACU, and PDU information as discussed further below. The measurement data is essential to (i) diagnose, (ii) understand the problems, and (iii) quantify the degree of best practices. It also provides an excellent mean to communicate to the data center operators the actual issues, thereby empowering the data center owners to implement the respective recommendations. If more precise, dynamic modeling is needed, a sensor network needs to be installed to monitor the temporal variations of the thermal conditions in the data center. In this case, the basic thermal profile generated from the MMT data can be used to identify critical locations for sensor placement.

35.5.2 Wireless Measurement and Management Technology

To capture the dynamic heat signal from servers, that may overheat due to IT load, MMT was extended to measure the dynamic change of temperature and airflow in a DC using wireless sensor network (Fig. 35.5a). The wireless connectivity allows the sensors to be placed both on servers as well as the inlet/outlet of ACUs. Typical sensors can be thermal, relative humidity, airflow, or pressure sensors. Sensors are placed at front of each server rack and positioned at different heights to capture the horizontal and vertical temperature profiles of the DC (Fig. 35.5b). The time series from thermal and other sensors are coupled to DC energy models and thermal approaches, which will allow for "what-ifs" based on the MMT benchmarked model representations of the DC. These "what-if" scenarios can guide DC operators in future growth and planning. Sensor data serve also as an input to computational fluid dynamics (CFD) models [25] that provide similar detailed thermal characterization as the mobile scanner. Details of the MMT can be found in [5], and application of MMT tools for controlling DCs is discussed in Chapter 10.

35.6 MMT-BASED BEST PRACTICES

MMT focuses on improving the energy and space efficiency of the DC by improving *two* aspects of the cooling infrastructure: namely, (i) lowering the chilled water setpoint temperature and thus reducing the compressor refrigeration work done by the chiller (thermodynamic) and (ii) lowering the total chilled airflow supplied by the ACUs on the DC floor and thus reducing the blower pumping work done by the ACU (transport).

Improving energy savings in a data center can be accomplished in four steps. In the first step, there is an initial assessment of the DC cooling efficiency with an estimation of the ACU (P_{ACU}) and chiller power ($P_{chiller}$) associated with the transport and thermodynamic term, respectively. This initial assessment sizes the opportunity because this service targets the reduction of chiller and ACU power and determines a cooling efficiency. The measured cooling efficiency is compared with a benchmark. In the second step, the environmental parameters in the DCs are acquired. The MMT data is compiled to six key metrics (horizontal hotspots (HH), vertical hotspots (VH), non-targeted airflow, sub-plenum air inlet temperature variations, ACU utilization, and ACU flow). This metric gives DC operators a benchmark to understand their DC efficiency and a tractable way to save energy based on low-cost best practices. The first four metrics affect the thermodynamic part of cooling power (chiller), while the last two affect the transport part of the cooling power (ACUs). As it will be discussed further below, based on the measurements, a set of recommendations can be generated. In the third step, data center operators can readily implement these recommendations. In the fourth step, once the recommendations are implemented, the data center is scanned again to quantify the improvement in energy efficiency. Table 35.2 shows a broad overview of the actual metrics compiled from data gathering process. We discuss the details of this metric and the data taking process in the next section.

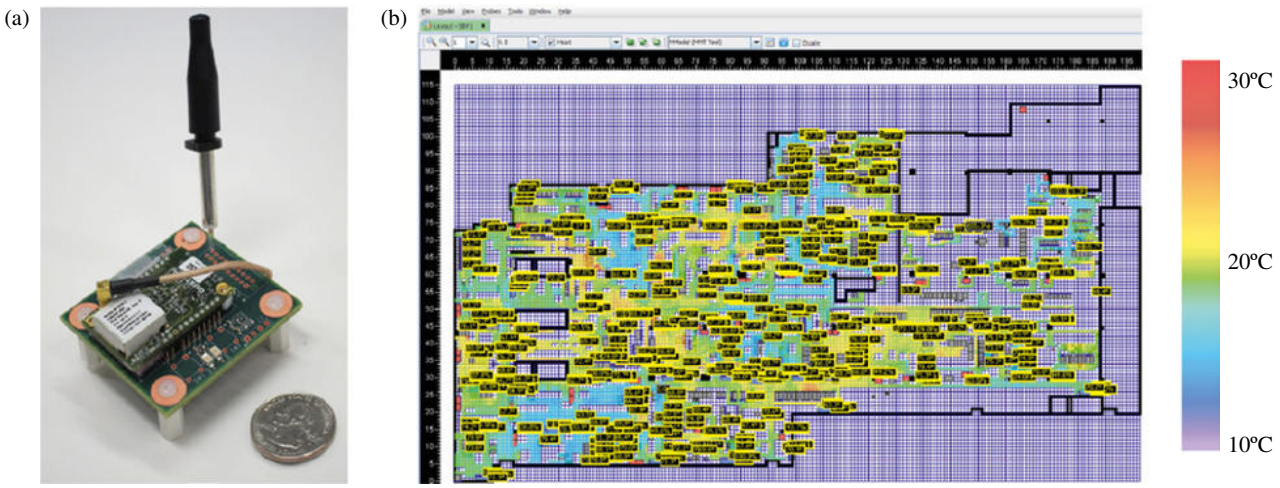


FIGURE 35.5 (a) MMT wireless sensor with temperature and relative humidity sensing and (b) MMT interface for a data center layout with more than 500 wireless sensors streaming real-time data.

TABLE 35.2 Overview of the quantitative values that can be detected and the associated best practice metrics

Metric		Description
DC cooling efficiency	$\eta = P_{IT}/(P_{chiller} + P_{ACU})$	IT power/total power
Chiller power	$P_{Chiller} \approx P_{RF}/\mathrm{COP}$	Chiller power (thermodynamic opportunity)
ACU power	$P_{ACU} \approx \sum_{i=1}^{\#ACU} P^i_{blower}$	ACU power (transport opportunity)
Inlet hotspots (IH) a. Horizontal hotspots (HH) b. Vertical hotspots (VH)	$\mathrm{IH} = T^{max}_{inlet} - T^{95}_{inlet}$ $\mathrm{HH} = T^{95}_{face} - T^{5}_{face}$ $\mathrm{VH} = \Delta T^{max}_{Rack} - \Delta T^{95}_{Rack}$	HH: Airflow provisioning Placement of perforated tiles? VH: Rack airflow Recirculation?
Targeted airflow (TF)	$\mathrm{TF} = \frac{f^{total}_{targeted}}{f^{total}_{ACU}}$	Lost/nondirected Chilled airflow
Plenum air temperature variations	$\mathrm{PT} = T^{avg}_{plenum} = \sum_{i=1}^{\#ACU} w^i_{flow} \cdot T^i_D$	ACU discharge temperatures Which ACUs do not function?
ACU utilization	$\mathrm{UT} = \upsilon^{avg}_{ACU} = P_{RF} / \sum_{i=1}^{\#ACU} P^i_{ACU,capacity}$	Utilization of ACUs Which ACUs can be turned off?
ACU flows	$\mathrm{FL} = \gamma^{avg}_{ACU} = \sum_{i=1}^{\#ACU} \gamma^i_{ACU} / \#\mathrm{ACU}$	Blockage: Which ACUs has low flow?

35.7 MEASUREMENT AND METRICS

35.7.1 Initial Assessment

DC Cooling Efficiency, ACU, and Chiller Power: In the initial phase of this service, the DC cooling efficiency is measured by

$$\eta = \frac{P_{IT}}{P_{chiller} + P_{ACU}} \tag{35.1}$$

where

η = average cooling efficiency
P_{IT} = IT power
P_{ACU} = Air Condition Unit power
$P_{chiller}$ = Chiller power

We note that this DC *cooling efficiency* is different from the more commonly used PUE [7] that uses the ratio between IT power (P_{IT}) and total DC power (P_{DC}). Generally though, Equation (35.1) is preferred here because all the quantitates of (Eq. 35.1) are directly measured (or estimated with very good accuracy).

Specifically, the power for IT (P_{IT}) can be directly measured at the PDUs, which are distributed through the DC raised floor room. Most PDUs have meters, but in some cases current clamps may have to be used to estimate the IT power. For the ACU power (transport), estimates either all the blower powers P^i_{blower} for each ACU are summed or multiply an average blower power P^{avg}_{blower} with the numbers of ACUs (#ACU) on the raised floor neglecting here energy consumption due to dehumidification:

$$P_{ACU} \approx \sum_{i=1}^{\#ACU} P^i_{blower} = \#\mathrm{ACU} \cdot P^{avg}_{blower} \tag{35.2}$$

The thermodynamic chiller power is often available from the facility power monitoring systems. Alternatively, $P_{chiller}$ can be approximately determined by estimating the total raised floor power and an estimated coefficient of performance (COP) for the chiller:

$$P_{chiller} = P_{RF} / \mathrm{COP} \tag{35.3}$$

where P_{RF} = total raised floor power.

The chiller COP as a function of a nominal discharge temperature (T_D) of an ACU can be expressed as

$$\mathrm{COP} \approx 0.216 + 0.0765 \cdot T_D \left[^\circ\mathrm{F}\right] \tag{35.4}$$

In a typical data center, T_D can be easily determined by measuring the discharge temperatures. For a typical discharge temperature $T_D \sim 13°C$, the equations yield a COP of 4.5 (which corresponds to 0.78 kW/tons). While the exact chiller COP depends on particular data center infrastructure, we note that Equation (35.4) does provides a consistent and

reliable way to estimate the energy saving opportunities (by raising the chilled water setpoint from the chiller to the ACU) regardless of details of the chiller.

The total raised floor power is given by

$$P_{RF} = P_{IT} + P_{light} + P_{ACU} + P_{PDU} \tag{35.5}$$

where

P_{light} = power used for lighting,
P_{ACU} = total ACU power on the floor, and
P_{PDU} = losses associated with the power distribution units

P_{IT} is by far the largest term and is known from the PDU measurements for the data center efficiency as described above. The power for lights can be readily estimated by

$$P_{light} \approx A_{DC} \cdot 2 \ \text{W/ft}^2 \tag{35.6}$$

for a typical data center, where A_{DC} is the data center floor area. Typical PDU losses are on the order of 10% of the IT power ($P_{PDU} \approx P_{IT} \times 10\%$).

In several MMT pilots an average cooling efficiency of η = 2.0 is measured. State-of-the-art data centers have cooling efficiencies of 3.0 and larger although DC differ and not all can accomplish cooling efficiencies in the excess of 3.0 depending on the specific details of the data center. The net saving potential can then readily be estimated by η_{target}.

35.7.2 MMT Scans And Best Practice Metrics

35.7.2.1 Metrics Associated with Thermodynamic Energy Savings

3D Temperature Distribution and Hotspots: Hotspots are one of the main sources for energy waste in data centers. In a typical data center, only a *few* racks in specific areas are exceeding recommended operating temperatures. It is quite common that servers in *higher* rack positions are the hottest due to poor air circulations. An energy-costly solution to such hotspots involves compensating for these hotspots by a chosen lower chiller setpoint, which drives up the cooling energy costs. The 3D temperature field in the DC as shown in Figure 35.4b and 35.4c can easily identify these hotspots within the DC. Generally, hotspots come from the fact that certain regions of the DC are under-provisioned, while others are potentially over-provisioned for cooling. Generally, hotspot metrics are focusing on only the hottest server racks, and by mitigating these hotspots, the setpoint of chilled water from the chiller can be increased (in a hotspot) risk neutral manner. Two types of hotspots are identified in a data center: HH and VH, and each of them can be mitigated using different solutions. HH are generally governed by the lateral placement of perforated tiles and possibly the levels of targeted airflow. VH are typically caused by poor air recirculation where hot air from the top gets to higher lying server racks minimizing the cool airflow from the raised floor.

Inlet Temperature Uniformity/Inlet HotSpots (IH): The MMT can measure *all* inlet temperatures across each relevant rack T_{inlet}^{q} in the data center (at a foot increment) (q = 1. . .# INLETS). All inlet temperatures can be readily represented as a simple histogram $h_{IH}(T_{inlet})$ to assess the temperature distribution in the data center. The inlet temperature hotspot factor (IH) is defined as the temperature difference/range from the maximum measured inlet temperature T_{inlet}^{max} to the 95% point T_{inlet}^{95} (95% of all inlet temperatures are below T_{inlet}^{95}, or 5% are above T_{inlet}^{95}):

$$IH = T_{inlet}^{max} - T_{inlet}^{95} \ \text{with} \ \sum_{T_{inlet}^{95}}^{T_{inlet}^{max}} h_{IH}\left(T_{inlet}\right) \Big/ \#\text{INLETS} = 5\% \tag{35.7}$$

We note that the standard deviation (T_{inlet}^{std})

$$T_{inlet}^{std} = \sqrt{\sum_{q=1}^{\#\text{INLETS}} \left(T_{inlet}^{q} - T_{inlet}^{avg}\right)^2 / \#\text{INLET}} \tag{35.8}$$

and the average (T_{inlet}^{avg}) of face temperatures

$$T_{inlet}^{avg} = \sum_{q=1}^{\#\text{INLET}} T_{q}^{face} / \#\text{INLET} \tag{35.9}$$

are other important metrics to understand the hotspots in the data center.

From multiple deployment it has been established that the hottest 5% of inlets can be readily addressed by increasing airflow (lowering temperature in those locations by, e.g., 5°C). Once the hotspots are mitigated and the top 5% of inlets with higher temperature distribution are reduced to safe temperature range, the chilled water setpoint can be increased by 5°C, which increases the chiller COP readily by approximately 15%. We note that this service asset does not initially include the option for raising globally the inlet temperatures (e.g., by increasing the chilled water setpoint), but it rather reduces hotspots before raising it to approach a risk neutral solution for improving the DC energy efficiency. Extensive studies have shown that most hotspots are horizontal, vertical, or both in nature, which we are now describing in more detail.

Horizontal Temperature Uniformity/Horizontal Hotspot (HH): The HH factor is a measure for the lateral temperature "tilt" in the data center:

$$HH = T_{face}^{95} - T_{face}^{5} \ \text{with} \ \sum_{T_{inlet}^{5}}^{T_{face}^{95}} h_{HH}\left(T_{face}\right) \Big/ \#\text{RACK} = 90\% \tag{35.10}$$

where $h_{HH}(T_{face})$ is the frequency distribution of all face temperatures in the DC. Specifically, T^{j}_{face} is the average rack inlet (face) temperatures for each rack in the DC ($j = 1 \ldots \#\text{RACK}$). T^{95}_{face} is the temperature of the distribution with 95% of the face temperatures below, while T^{5}_{face} is the temperature of $h_{HH}(T_{face})$ with only 5% of racks below that setpoint. In a typical report of this service, server racks with face temperatures *above 95%* and below *5%* are indicated, which guides the provisioning/perforated tile layout in the data center. Specifically, for servers that have an input temperature below T^{5}_{face}, the airflow can be reduced (e.g., removing perforated tiles), while for servers that have an input temperature larger than T^{95}_{face}, tiles must be added or higher throughput tiles implemented. We note that the standard deviation (T^{std}_{face})

$$T^{std}_{face} = \sqrt{\sum_{j=1}^{\#\text{RACK}} \left(T^{j}_{face} - T^{avg}_{face}\right)^2 / \#\text{RACK}} \tag{35.11}$$

and the average (T^{avg}_{face}) of face temperatures

$$T^{avg}_{face} = \sum_{j=1}^{\#\text{RACK}} T^{face}_{j} / \#\text{RACK} \tag{35.12}$$

are other important metrics that can guide the choice of optimum perforated tiles that needs to be mounted in each location to mitigate hotspots. In addition, we note that T^{avg}_{face} can be readily used to determine whether the whole DC is not just simply overcooled. While a reduction in the inlet hotspot temperature (IH) naturally leads directly to energy saving opportunities, the reduction in HH does not necessarily translate into savings without addressing VH.

Vertical Temperature Uniformity/Vertical Hotspot (VH): Although the correct allocation of required airflow to each server rack is very important part of best practices (low HH), experience has shown that simply provisioning the right amount of airflow does not always mitigate hotspots and that there are clearly limits to this approach. For example, additional restrictions can inhibit the airflow and create VH. The vertical temperature profiles in a typical data center indicate large temperature gradients between the bottom and the top of the rack. In general, the bottom of the racks is "overcooled," while the servers in the top do not get the required cooled air. In order to quantify, the VH is expressed as the average of the difference between the lowest and highest measurement point ΔT^{j}_{Rack} in each rack j ($j = 1 \ldots \#\text{RACK}$). In analogy to the IH, a VH factor can be defined as

$$\text{VH} = \Delta T^{max}_{Rack} - \Delta T^{95}_{Rack} \quad \text{with} \quad \sum_{\Delta T^{95}_{Rack}}^{\Delta T^{max}_{Rack}} h_{VH}\left(\Delta T_{Rack}\right) \Big/ \#\text{RACK} = 5\% \tag{35.13}$$

where VH is given by the temperature difference/range from the maximum measured vertical temperature ΔT^{max}_{Rack} to the 95% point ΔT^{95}_{Rack} (95% of all vertical temperatures are below ΔT^{95}_{Rack}, or 5% are above ΔT^{95}_{Rack}). The standard deviation (ΔT^{std}_{Rack}) and the mean vertical temperatures (ΔT^{avg}_{Rack}) are an alternative way to gauge the degree of VH.

Targeted Airflow: While the placement of the perforated tiles will mostly affect the HH in a typical DC, a significant fraction of the air is not targeted at all (besides the provisioning issues) coming through cable cutouts and other leaks. Specifically, we define the targeted airflow by

$$\text{TF} = \frac{f^{total}_{targeted}}{f^{total}_{ACU}} \tag{35.14}$$

where f^{total}_{ACU} is the total flow given by the entire flow output from all ACUs within the data center. Because the quantitative measurement of ACU flows is nontrivial, a simple estimate is f^{total}_{ACU}, which uses a combination of balancing the dissipated energy within the DC and allocating this energy between the different ACUs (based on non-calibrated flow measurements) yielding

$$f^{total}_{ACU} \approx \frac{1}{\rho c_p} \sum_{i=1}^{\#\text{ACU}} P^{i}_{cool} \Big/ \Delta T^{i}_{ACU} \tag{35.15}$$

where

P^{i}_{cool} = the *power cooled* by the respective ACU,
ΔT^{i}_{ACU} = the difference between the ACU return (T^{i}_{R}) and discharge (T^{i}_{D}) temperatures
ρ and cp = the density and specific heat of air ($\rho \approx 1.15\,\text{kg/m}^3$, $c_p \approx 1{,}007\,\text{J/kg K}$)

While it is straightforward to measure the actual return and discharge temperatures for each ACU, the respective CRAC cooling power levels P^{i}_{cool} are more difficult. The total power dissipated (P_{RF}) in the raised floor area needs to be equal (to a first order) to the sum of each ACU cooling power for all ACUs:

$$P_{RF} \approx \sum_{i=1}^{\#\text{ACU}} P^{i}_{cool} \tag{35.16}$$

By measuring the (non-calibrated) airflow $f^{i}_{ACU,NC}$ from each ACU as well as the temperature difference between the return and discharge (ΔT^{i}_{ACU}) for each ACU, the relative cooling power contribution ($w^{i}_{cool} = f^{i}_{ACU,NC} \Delta T^{i}_{ACU}$) can be attributed to each ACU and derive the respective power cooled at each ACU by

$$P^{i}_{ACU,cool} \approx P_{RF} w_i \Big/ \sum_{i=1}^{\#\text{ACU}} w_i \tag{35.17}$$

Details about the ACU flow and temperature measurements are discussed elsewhere [26–28]. The targeted airflow can be

readily determined by measuring the airflow from each perforated tile with a standard flow hood [29, 30]. Perforated tiles that are far away from any server or IT equipment are counted toward nontargeted airflow:

$$f_{\text{targeted}}^{\text{total}} = \sum_{j=1}^{\#\text{RACKS}} f_{\text{perf}}^{j} \tag{35.18}$$

If airflow is provisioned adequately in a DC, significant energy savings can be achieved by turning off the underutilized ACUs. The airflow loss due to the reduction in ACUs (if appropriate) will be compensated by the increased targeted airflow.

Plenum Air Temperature: It is common that some ACUs on the raised floor do have excessive discharge temperatures, thereby increasing the average plenum temperature:

$$\text{PT} = T_{\text{plenum}}^{\text{avg}} = \sum_{i=1}^{\#\text{ACU}} w_{\text{flow}}^{i} \cdot T_{\text{D}}^{i} \tag{35.19}$$

with T_{D}^{i} as the discharge temperature for each ACU (i) and w_{flow}^{i} as the relative flow contribution from each active ACU (turned-off ACUs are accounted in the nontargeted airflow section if they leak cold air from the plenum):

$$w_{\text{flow}}^{i} = f_{\text{ACU,NC}}^{i} \Big/ \sum_{i=1}^{\#\text{ACU}} f_{\text{ACU,NC}}^{i} \tag{35.20}$$

There could be various reasons for excessive discharge temperatures such as low utilization, clocked water valves, and overloading. In order to understand plenum air temperature variations better, we investigate the difference δ_i between the respective discharge temperature (T_{D}^{i}) for each ACU (i) and the average discharge temperature ($T_{\text{plenum}}^{\text{avg}}$):

$$\delta_i = T_{\text{D}}^{i} - T_{\text{plenum}}^{\text{avg}} \tag{35.21}$$

and plot δ_i as a function of the utilization levels δ_i for each ACU as shown in Figure 35.6. ACUs in left upper quarter are just underutilized. Consequently, the water valve is closed, and warm air is injected into the plenum. ACUs in the right upper corner would have either capacity issues (overloading) or water supply issues. We note that the frequency distribution of the discharge temperature variation ($h_{\text{PT}}(T_{\text{D}}^{i})$), the standard deviation ($T_{\text{D}}^{\text{std}}$), and the mean vertical temperatures ($T_{\text{D}}^{\text{avg}}$) is an alternative way to gauge the impact of plenum temperature variations.

35.7.2.2 Metrics Associated with Transport Energy Savings

ACU Utilization: In a typical DC more ACUs are running then needed. As such the ACU utilization (UT) is a very important metric to understand possible energy savings

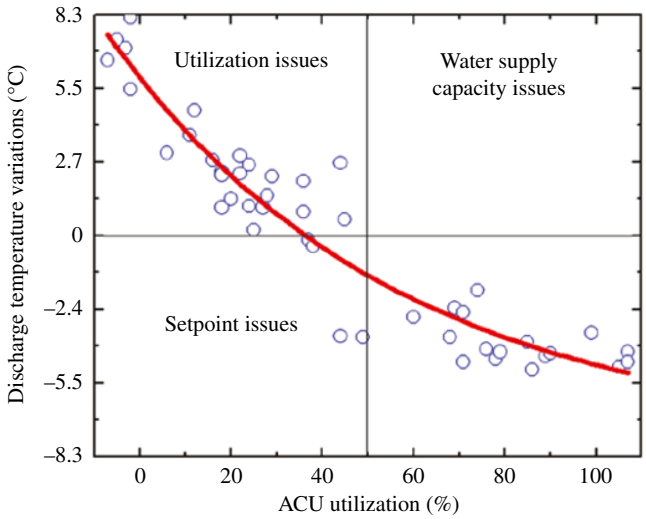


FIGURE 35.6 Discharge temperature variation vs. ACU utilization levels.

associated with transport. An average ACU utilization ($\upsilon_{\text{ACU}}^{\text{avg}}$) within the DC can be readily estimated by

$$\text{UT} = \upsilon_{\text{ACU}}^{\text{avg}} = P_{\text{RF}} \Big/ \sum_{i=1}^{\#\text{ACU}} P_{\text{capacity}}^{i} \tag{35.22}$$

While the average ACU utilization is readily estimated and, in some cases, known by DC operators, MMT provides a *much more detailed look* at the ACU utilization. Specifically, the utilization for *each* individual ACU is quantified as

$$\upsilon_{\text{ACU}}^{i} = P_{\text{cool}}^{i} \Big/ P_{\text{capacity}}^{i} \tag{35.23}$$

and defines an ACU utilization frequency distribution ($h_{\text{UT}}(\upsilon_{\text{ACU}}^{i})$) with its standard deviation ($\upsilon_{\text{ACU}}^{\text{std}}$), which gives the data center operators a detailed way to understand how the heat load is distributed between the different ACUs within the DC and which ACUs can be potentially turned off with the least impact. In addition, the data center operator can understand a "what-if" scenario if a certain ACU would fail and would help planning for emergency situations. Ideally, an energy-efficient data center has a very narrow frequency distribution centered at 100% utilization. Typical DCs have average distributions on the order of 50%. Because most data centers require an $N+1$ solution for the raised floor, it may be recommended to position the average of the frequency distribution not quite to 100% but to ≈(#ACU − 1)/#ACU (e.g., a data center with ACUs would try to target a mean utilization of 90% with a standard deviation of less than 10%). There are basically two ways for improving ACU utilization: (i) turning off ACUs or (ii) raising the raised floor power consumption (i.e., match the IT power bettery to ACU capacity).

ACU Flow: Often blockage, dirty filters, low throughput tiles prevent the blower from delivering the flow to the racks,

which is an additional energy loss term. The average flow capacity ($\gamma_{\mathrm{ACU}}^{\mathrm{avg}}$) is

$$\mathrm{FL} = \gamma_{\mathrm{ACU}}^{\mathrm{avg}} = \sum_{i=1}^{\#\mathrm{ACU}} \gamma_{\mathrm{ACU}}^{i} / \#\mathrm{ACU} \tag{35.24}$$

where $\gamma_{\mathrm{ACU}}^{i}$ is the flow capacity. Further the flow capacity is defined:

$$\gamma_{\mathrm{ACU}}^{i} = f_{\mathrm{ACU}}^{i} \Big/ f_{\mathrm{capacity}}^{i} \tag{35.25}$$

where $f_{\mathrm{capacity}}^{i}$ is the nominal flow specified by the ACU manufacturer and f_{ACU}^{i} the actual (calibrated) measured flow from each ACU. The actual flow from each ACU can be determined from the non-calibrated flow measurements ($f_{\mathrm{ACU,NC}}^{i}$), and the (earlier calculated) total flow in the data center ($f_{\mathrm{ACU}}^{\mathrm{total}}$) by

$$f_{\mathrm{ACU}}^{i} = f_{\mathrm{ACU,NC}}^{i} f_{\mathrm{ACU}}^{\mathrm{total}} / \sum_{i=1}^{\#\mathrm{ACU}} f_{\mathrm{ACU,NC}}^{i} \tag{35.26}$$

The distribution of this flow capacity $h_{\mathrm{FL}}(\gamma_{\mathrm{ACU}}^{i})$ and the respective standard deviation $\gamma_{\mathrm{ACU}}^{\mathrm{std}}$ is a gauge for the degree of blockage and effectiveness of ACU flow delivery.

35.7.3 Practical Solutions for Data Center Thermal/Energy Problems

Data centers can be characterized by a set of metrics extracted from real-time measurements. The performance of each data center is based on multiple factors like data center layout, number of server racks, and ACUs. A set of common metrics can be defined that can improve the data center operations (Table 35.3). Some of these recommendations can be revisited periodically to assess the performance and quantify the achieved energy savings.

There are several additional infrastructure recommendations that can be implemented as part of the MMT service package like:

- Increased airflow
- Curtains
- Increasing the ceiling height

35.7.3.1 Increased Airflow

The perforated tiles supply cold air to server and IT equipment, and the percentage of opening and underflow pressure are controlling the total amount of airflow. The perforated tiles act as vents for the chilled air to enter the room from the underfloor plenum. The smaller the total open area provided for the airflow, the higher the resistance faced by the ACU, and consequently the lower the total flow rate achieved by the blower in the ACU. Typically, the perforated tiles account for 5–15% of the total pressure drop of the airflow loop, i.e., from the suction side of the ACU blowers to the return side. Thus, these perforated tiles play an important role in ensuring that the ACU flow capability metric is high and that the ACU blower is not unduly penalized by highly constricted tiles.

TABLE 35.3 Data center thermal/energy metrics and the recommended measures

Data center metrics	Practical recommendations to improve energy efficiency
Horizontal hotspot	1. Change perforated tile layout 2. Install of higher throughput (HT) open tiles 3. Use of curtains 4. Change in the rack layout
Vertical hotspot	1. Change perforated tile layout 2. Install higher throughput tiles 3. Increase ceiling height
Nontargeted airflow	1. Seal leaks and cable cut out openings 2. Install higher throughput tiles
ACU's operation optimization	1. Incorporate variable frequency drive (VFD) controls 2. Clean heat exchanger coils 3. Replace air filters 4. Improve underfloor airflow

35.7.3.2 Curtains

The curtains that separate inlet and outlet for aligned rows of servers, as displayed in Figure 35.7, can partition the hot exhaust air from the cold intake air in the case of a mixed aisle layout. A mixed aisle layout is basically a violation of the hot aisle–cold aisle arrangement. The curtains will also reduce the VH.

35.7.3.3 Increasing the Ceiling Height

The region above the racks, between the tops of the racks and the ceiling, acts as a quasi-plenum for the hot exhaust air to return from the exhaust (rears of the racks) to the ACU. Thus, the ceiling height places an important role in ensuring that this so-called return plenum is big enough so as to allow the hot exhaust air to make its way to suction side of the ACU without percolating down to the inlets of some of the IT equipment. CFD modeling studies conducted as part of the current project indicate that increasing the ceiling height from 8.75 to 11 in can result in a 2.3°C reduction in the rack inlet temperature in the hot spot region. Figure 35.8 shows temperature contours of a vertical section of the CFD model built and solved to study this important parameter.

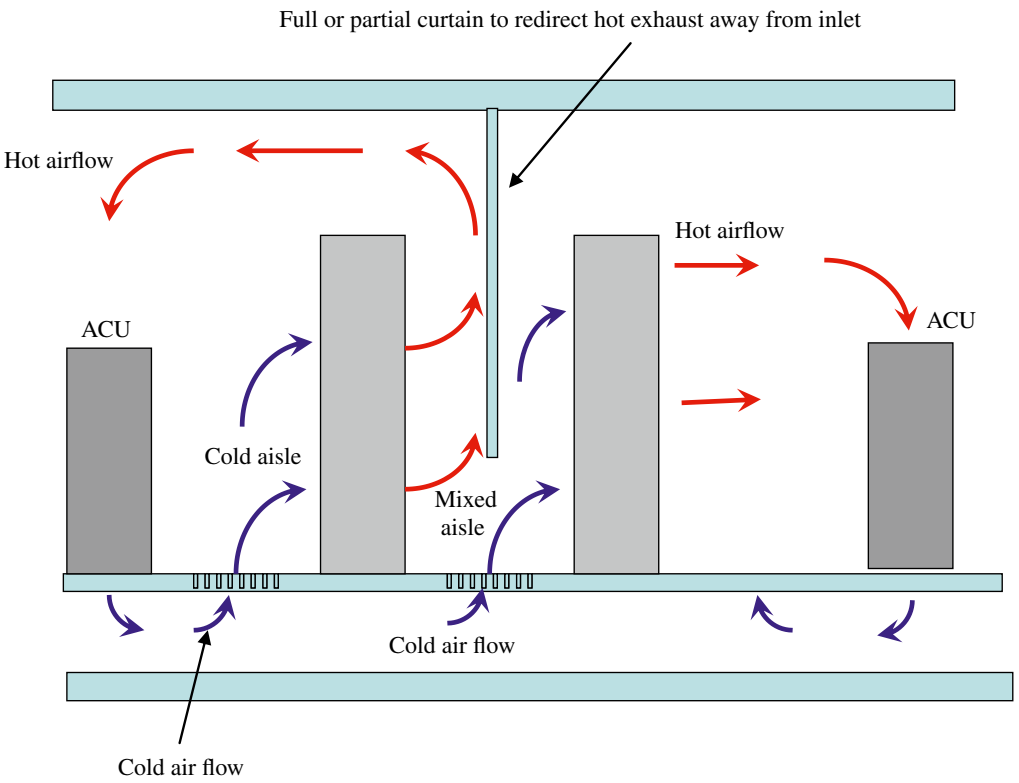


FIGURE 35.7 Schematic depicting the use of curtains.

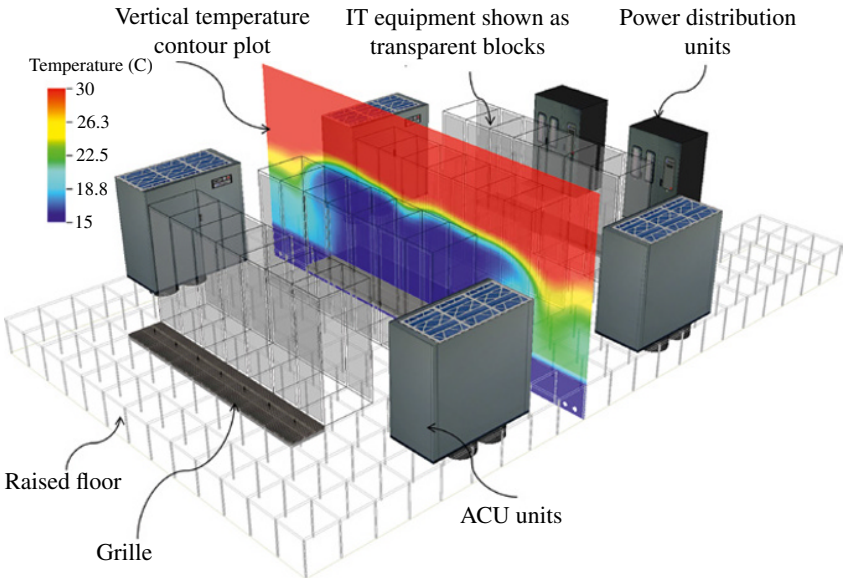


FIGURE 35.8 Temperature changes from floor to ceiling based on CFD study. Source: Courtesy of Rainspur Technology Co., Ltd.

35.8 CONCLUSIONS

Simple energy efficiency metrics can be calculated from high spatial–temporal measurements that can be easily translated into best operating practices to improve energy efficiency recommendations. Here we present a method called measurement and management technology to quantitatively measure environmental parameters in a data center and extract from the measurement quantitative metrics that can provide recommendations to be followed by DC operators. The measurement and recommendation methods apply both to legacy and newly designed data centers.

REFERENCES

[1] T. ASHRAE. 9: best practices for datacom facility energy efficiency. ASHRAE, Atlanta; 2008. p 64–66.

[2] Brill KG. Datra center energy efficiency and productivity. The Uptime Institute-White Paper; 2007.

[3] Jaureguialzo E. PUE: the Green Grid metric for evaluating the energy efficiency in DC (data center). Measurement method using the power demand. 2011 IEEE 33rd International Telecommunications Energy Conference (INTELEC); 2011. p 1–8.

[4] Hamann HF, Schappert M, Iyengar M, van Kessel T, Claassen A. Methods and techniques for measuring and improving data center best practices. 2008 11th Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems; 2008. p 1146–1152.

[5] Hamann HF, et al. Uncovering energy-efficiency opportunities in data centers. IBM J Res Dev 2009;53:10:1–10:12.

[6] Greenberg S, Mills E, Tschudi B, Rumsey P, Myatt B. Best practices for data centers: lessons learned from benchmarking 22 data centers. Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings in Asilomar, CA. ACEEE; August, 2006, vol. 3. p 76–87.

[7] Ebrahimi K, Jones GF, Fleischer AS. A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities. Renew Sustain Energy Rev 2014;31:622–638.

[8] Shehabi A, et al. *United States Data Center Energy Usage Report*. Berkeley, CA: Lawrence Berkeley National Lab (LBNL); 2016.

[9] Koomey JG. Worldwide electricity used in data centers. Environ Res Lett 2008;3:034008.

[10] Belady C, Rawson A, Pfleuger J, Cader T. Green grid data center power efficiency metrics: PUE and DCIE. Technical Report, Green Grid; 2008.

[11] Shuja J, et al. Survey of techniques and architectures for designing energy-efficient data centers. IEEE Syst J 2014;10:507–519.

[12] Ascierto R. Uptime institute global data center survey. Technical Report, Uptime Institute; 2018.

[13] Gao J. Machine learning applications for data center optimization; 2014.

[14] Klein LJ, Bermudez S, Wehle H-D, Barabasi S, Hamann HF. Sustainable data centers powered by renewable energy. 2012 28th Annual IEEE Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM); 2012. p 362–367.

[15] Sharma M, Arunachalam K, Sharma D. Analyzing the data center efficiency by using PUE to make data centers more energy efficient by reducing the electrical consumption and exploring new strategies. Procedia Comput Sci 2015;48:142–148.

[16] Pacific Gas and Electric. *High Performance Data Centers – A Design Guidelines Sourcebook*. California: Pacific Gas and Electric Co.; 2006.

[17] Schmidt RR, Iyengar M, van der Mersch PL. Best practices for data center thermal and energy management-review of literature/discussion. ASHRAE Trans 2007;113:206.

[18] Kurkjian C, Glass J. Air-conditioning design for data centers accommodating current loads and planning for the future. ASHRAE Trans 2005;111.

[19] Dubin FS, Long Jr CG. Eneregy conservation standards; 1982.

[20] Patterson MK. The effect of data center temperature on energy efficiency. 2008 11th Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems; 2008. p 1167–1174.

[21] Bash C, Patel CD, Sharma RK. Dynamic thermal management of air cooled data centers. Thermal and Thermomechanical Proceedings 10th Intersociety Conference on Phenomena in Electronics Systems, 2006. ITHERM 2006; 2006. p 8–452.

[22] Zhou Y, Lu X, Zhong X, Klein LJ, Schappert M, Hamann HF. A tele-operative RMMT system facilitating the management of cooling and energy in data centers. 2010 IEEE International Conference on Robotics and Biomimetics; 2010. p 822–827.

[23] Connell IJH, et al. Detecting energy and environmental leaks in indoor environments using a mobile robot. Google Patents; 2012.

[24] Mansley C, et al. Robotic mapping and monitoring of data centers. 2011 IEEE International Conference on Robotics and Automation; 2011. p 5905–5910.

[25] Lopez V, Hamann HF. Heat transfer modeling in data centers. Int J Heat Mass Transf 2011;54:5306–5318.

[26] Liu M, Claridge D, Turner W. *Continuous Commissioning Guidebook for Federal Energy Managers, Maximizing Building Energy Efficiency and Comfort*. Federal Energy Management Program, United State Department of Energy; 2002.

[27] Pugh M. Benefits of water-cooled systems vs air-cooled systems for air-conditioning applications. AHR/ASHRAE/ARI Expo; 2005.

[28] Claassen A, Hamann HF, Iyengar MK, O'boyle MP, Schappert MA, Van Kessel TG. Techniques for analyzing data center energy utilization practices. Google Patents; 2008.

[29] Shrivastava S, Sammakia B, Schmidt R, Iyengar M. Comparative analysis of different data center airflow management configurations. ASME 2005 Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS, and Electronic Systems Collocated with the ASME 2005 Heat Transfer Summer Conference; 2005. p 329–336.

[30] Pardey ZM. Measurement of perforated tile airflow in data centers. ASHRAE Trans 2016;122:88.

36

DRIVE DATA CENTER MANAGEMENT AND BUILD BETTER AI WITH IT DEVICES AS SENSORS

Ajay Garg[1] and Dror Shenkar[2]

[1] *Intel Corporation, Hillsboro, Oregon, United States of America*
[2] *Intel Corporation, Haifa, Israel*

36.1 INTRODUCTION

The energy costs of data centers continue to rise along with the expansion; Gartner estimates[1] that ongoing power costs are increasing at least 10%/year due to cost per kilowatt-hour (kWh) increases in underlying demand. It is important that an appropriate data center management strategy is adopted to operate at optimum efficiency, considering the workloads running in a typical data center vary over time.

36.2 CURRENT SITUATION OF DATA CENTER MANAGEMENT

Today data centers are divided into two major silos:

The first silo looks at the data center as a commercial building and is tasked with the management of power and cooling needed by the data center. Typically called facility operations, this group ensures there is a good building management system in place that power and cooling requirements for each rack are handled appropriately.

The second silo or group looks at the IT devices placed in the racks and are typically called IT operations. Their management is confined to ensuring sufficient compute, storage, and networking to support the needs of their data center users.

[1] https://www.gartner.com/smarterwithgartner/5-steps-to-maximize-data-center-efficiency/.

These two silos can be categorized as provider and consumer where facility operations group providing power and cooling requirements to the IT operations group.

36.2.1 Facility Operations Management

Typically facility operators do not track what is happening at the IT operation level, but the consumer is very dynamic since workloads change rapidly and power cycles frequently go up or down. However, the facility operators, for example, don't make their chillers dynamic to their IT operator needs.

Facilities managers typically have a good handle on how to manage devices like UPS, power bus, generators, and chillers and utilize building management systems. But IT devices are usually unknown territory for them, so in order to get a better handle on these devices, they deploy an IoT solution that is essentially a system equipped with sensors to track IT infrastructure. These systems are typically expensive and require maintenance, updates, and upgrades over time.

36.2.2 IT Operations Management

Typically, there are unused capacity buffers in the data centers to ensure operations run without issue. For example, if the rack capacity is 5 kW, IT operations will place devices that consume no more than 3–4 kW during the full load. In addition, there is typically no real-time monitoring of IT device power consumption. These loads are calculated using specifications provided by the manufacturer or running a workload in the test lab and using that number for that particular IT make and model.

Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center, Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

In addition, lot of data centers keep track of IT infrastructure in an Excel sheet or some other database as an asset placed in the data center without knowing the real-time utilization of the device. There are old IT devices that may still be present in the data center consuming power while the actual workloads are shifted to newer systems, and the data center manager may not even know it.

Also, in most cases, there is no real-time monitoring and management of the IT infrastructure by data center managers; data centers can go through major unplanned shutdown or outages and costs businesses millions of dollars.

36.3 AI INTRODUCED IN DATA CENTER MANAGEMENT

IDC projects half of IT assets in data centers will be able to run autonomously because of embedded AI functionality by 2022. However, data center teams must work to fully understand the potential of this feature in order to leverage AI capabilities to its fullest extent. By integrating AI capabilities, IT teams are creating smarter data centers that have the flexibility to thrive in hybrid environments while optimizing every facet of future operations.

36.3.1 How AI Benefits Data Center Operations

The idea of AI in data centers is to make extensive data-driven decisions. AI can help data center managers collect analytics and determine the best device or set of devices for any given workload.

36.3.2 How to Introduce AI in Data Center Management

Legacy data centers are incapable of meeting the demands of new and emerging technologies, including AI, IoT, and machine learning, These technologies involve enormous volumes of raw and processed data. To support the pace of AI and deep learning application development required by businesses, the software-defined data center provides the agility, flexibility, management controls, automation, and cost reduction necessary in today's competitive environment.

One major benefit of a data center management solution is the monitoring component that tracks daily operations, identifies any efficiency gaps, and flags potential vulnerabilities. By tapping into next-generation tools like AI and machine learning, data center managers can gain real-time visibility into infrastructure performance and further reduce the risk of an equipment outage. If one happens to occur, data center managers are equipped with the right tools and insights to quickly remedy the vulnerability and ensure operations continue without interruption. Under any conditions, a highly automated, software-defined resource model and orchestration layer can protect data centers against potential equipment failures and otherwise undetected cooling anomalies that can cause outages and revenue loss.

36.3.3 AI in Software-Defined Data Centers

Software-defined data centers are a radical paradigm shift in data center deployment. In a software-defined data center, software abstraction virtualizes the majority of data center components, such as computing, storage, and networking, while policy-driven software controls infrastructure management tasks. By separating service management from the physical infrastructure, data centers become more agile than ever before.

But AI and machine learning go far beyond these consumer applications and are also being used to tackle some of the world's most challenging business and industrial problems, such as rampant energy consumption that adversely impacts the environment. One organization that utilized AI to manage power in its software-defined data centers achieved a 40% reduction in the electricity needed for cooling across its facilities.

36.4 CAPABILITIES OF IT DEVICES USED FOR DATA CENTER MANAGEMENT

There has been a lot of research and development done by Intel and HW manufacturers to build sensors that provide important parameters about the real-time status of the device. For example, a system can provide real-time power consumption, inlet and outlet temperature, and the health status of various subcomponents within the system. Another example is the status of the fan responsible to cool the system processor. One of the major problems for data center managers is collecting this data, because it requires awareness of a wide variety of protocols that device manufacturers use to expose the data, the good news is there are various tools offered by system manufacturers, DCIM vendors, and Intel® data center manager that utilize capabilities of IT devices for overall data center management.

36.5 USAGE MODELS

Previous sections introduced the current status of data center management, how AI is used in data center management and how it plays a critical role driving data center efficiency and improving data center uptime, and how granular real-time device-level temperature monitoring is important for an automated data center management solution. The monitored data, along with the analytics and AI on top of it, can help the data center manager and IT operations drive efficiency and reduce operation costs. These gains are found through

driving higher temperatures in the data center, solving thermal issues, increasing rack density, better capacity planning, locating ghost and underutilized servers, and more.

36.5.1 Drive Higher Temperatures in Data Centers

In last several years, there has been an ongoing industry effort to raise operating temperature in data centers. It is aligned with ASHARE standard and initiative to lower data center energy consumption in the data center. Despite this effort, we still see many data centers overcooling and operating in temperatures lower than they can potentially operate. One of the main reasons for this is because the IT industry is relatively conservative; often they are not willing to take unnecessary risks that could potentially impact reliability. This group prefers to maintain their current environment and processes, especially if they're not forced to do so by management.[2] Things are changing now, and we see a stronger push from management to reduce data center operating costs and thus more openness from IT and data center operations to consider raising temperatures in the data centers, assuming they don't impact data center uptime and reliability.

Today, in many data centers, operating temperature is monitored by thermal sensors that are positioned in various locations in the data center. This approach can provide high level indication of the temperatures in the data center, but it is not sufficient if operators want to increase the set point and ensure it doesn't impact the environment. Depending on the sensor count and exact location, some areas of the data center can remain unmonitored.

In order to increase the set point in the data center, IT needs to see and understand impact across all areas of the data center; hence device-level temperature monitoring is required. As we explained earlier, most IT devices today, including servers, network, and storage, can report their temperature. When monitoring and analyzing temperature for all devices in the data center, IT can create very granular thermal maps at rack and device level and have all the data needed in order to identify overcooling scenarios and in order to safely increase the set point in the data center.

36.5.2 Solve Thermal Issues in the Data Center

While cooling management has improved in last years, many facilities still face cooling issues causing them to waste energy or prevent them from achieving full capacity. In addition, thermal issues, if not identified and addressed quickly, can lead to equipment failure and system downtime.

Some of the common thermal issues in data centers include the following[3]:

- Insufficient cool air: Not enough cold air is sent to the cold aisle. The result is that servers at the bottom of the racks consume the cold air and then, when fans in the upper servers draw in air from the room, it is actually hot air, and hence those servers see relatively high temperatures that can impact their reliability and performance.
- Lack of perforated tiles: It might be that there's enough cool air; however, it is not being distributed to the racks due to lack of perforated tiles or having perforated tiles in some locations in the data center.
- Too many perforated tiles: Putting perforated tiles in hot aisles and white spaces or even putting too much perforated tiles in the cold aisle can lead to waste of cooling capacity.
- Empty rack spaces: When some rack space is left empty, the airflow balance gets skewed, leading to recirculation of exhaust air into the cold aisle or loss of cool air from the cold aisle. While blanking panels can easily solve most of those issues, not everyone does it.
- Poor rack sealing: Some racks are not designed where space between mounting rails and sides of racks are sealed.
- Unsealed raised floor openings: In some cases, cable openings and other holes in the raised floor remain unsealed, and cool air can go to these areas, where it is really unneeded.

Granular device-level temperature monitoring can be a key in identifying and solving those potential thermal issues. Monitoring and analyzing the temperature data for all the devices in the data center can help the operators to easily identify such issues. AI can be built in order to analyze all the temperature data coming from the devices, identifying thermal issues, categorizing the issue, and even recommending a solution.

Figure 36.1 is an example of impact of empty rack spaces and lack of blanking panels on the servers.

Figure 36.2 is an example of an issue where fans in one chassis were not working. Monitoring and analysis of the temperature data helped identifying this issue and solved it by configuring the chassis and turning on the fans.

36.5.3 Capacity Planning and Rack Density Increase

Over the last several years, we see an increased reliance on compute to deliver mission-critical applications, while at the

[2] https://www.apc.com/salestools/VAVR-9SZM5D/VAVR-9SZM5D_R1_EN.pdf.

[3] https://www.datacenterknowledge.com/archives/2014/08/20/ten-common-cooling-mistakes-data-center-operators-make.

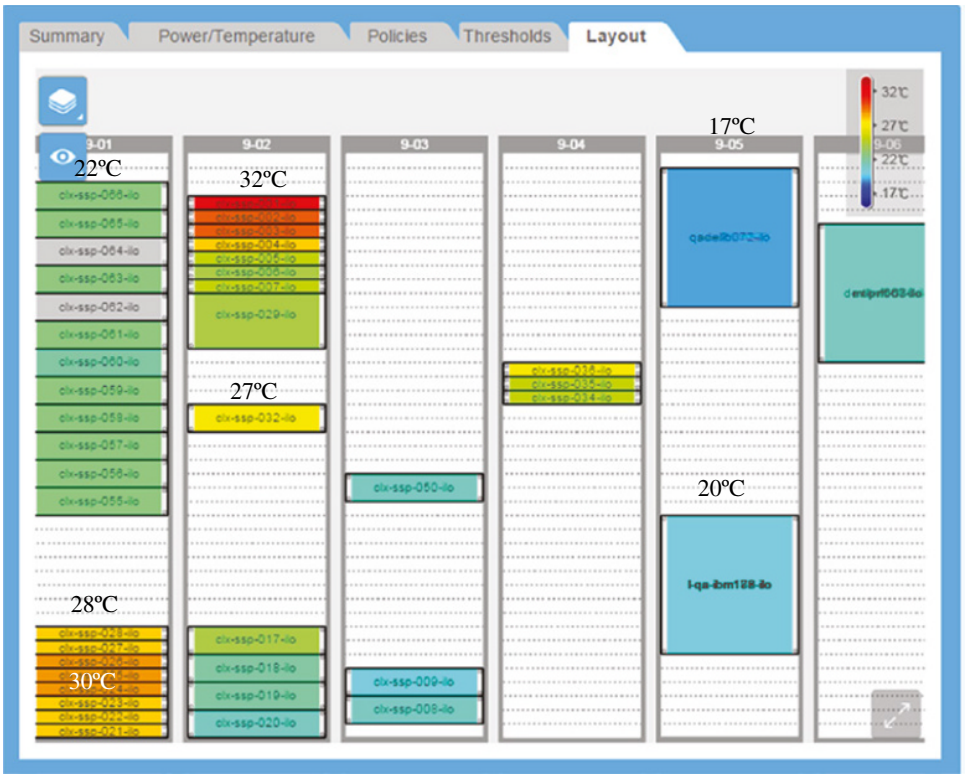


FIGURE 36.1 Front view temperature layout of a row in a data center. Source: Intel® Data Center Manager Console screenshot of undisclosed customer.

same time, we see a dramatic rise of power costs. Hence, we see more and more organizations put greater focus on power and thermal management and on conserving energy, especially in their data centers, which are one of the most heavy power consumers for those organizations.[4]

When energy management was a low priority, data center managers and IT managers could rely on nameplate values for capacity planning, even though relying on those nameplates could lead to a very inefficient data center. As noted above, today energy management priority is pretty high, and IT managers cannot rely on those nameplates anymore and need more accurate ways to measure IT power consumption.

OEMs indicate nameplate value on each server. However, this number represents a worst-case scenario that might be relevant only for a certain server configuration and a certain workload running on the server. In real life, power consumption will likely never reach those nameplate values, and hence, relying on those numbers when doing capacity planning and rack/server provisioning will result in a very inefficient design. One simple way to increase server density is by derating the nameplate value by a certain percentage.

[4] https://www.raritan.com/assets/re/resources/white_papers/White_Paper_-_Data_Center_Power_Distribution_Plan.pdf.

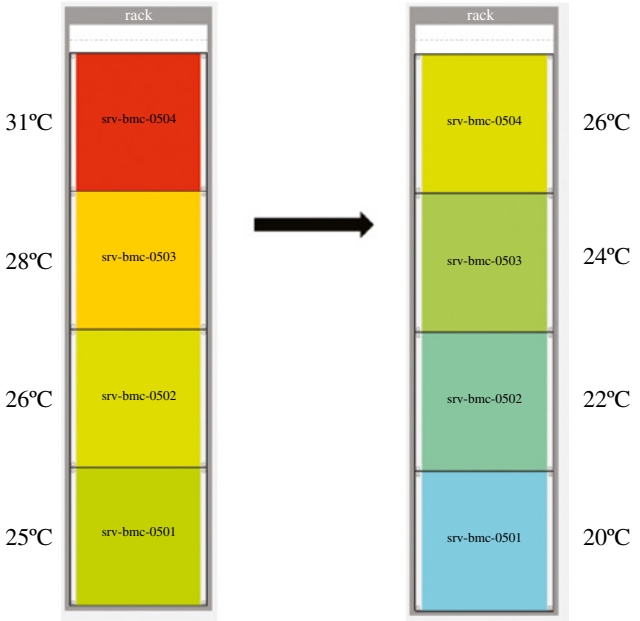


FIGURE 36.2 Front view temperature layout of a rack in a data center, before and after solving thermal-related issues. Source: Intel® Data Center Manager Console screenshot of undisclosed customer.

However, this method still has some problems as data center manager doesn't know the percentage they should derate by. If they derate too little, the efficiency won't get improved a lot, and if they derate too much, it might lead to power issues.

With server power instrumentation, management software can monitor and analyze power consumption at device level and get real-time power consumption for every server and device in the data center. Then, IT managers can do better capacity planning and improve efficiency and uptime in the data centers they're managing.

36.5.4 Identify Underutilized and Ghost Servers

One factor that creates inefficiencies in data centers is unnecessary heat generation; consumption of unneeded power leads to server underutilization.

According to analysts, ~15% of servers[5] in a typical data center are ghost servers, meaning they're not used at all and are not doing any useful work. These ghost servers can still consume 70–85% of power compared with fully utilized servers, which means that they're taking away needed resources from operational devices, including power, cooling, and space, plus increasing inefficiency in the data center.[6]

Ghost servers can also introduce security vulnerabilities. In 2006, Ohio University reported that someone hacked into one of their database servers and stole personal information on more than 300,000 people and organizations. When analyzing the logs, the university IT staff found out that the server was compromised for more than a year and that it was used for a denial-of-service attack against an external target. Apparently, this server was supposed to have been decommissioned, and thus it didn't get latest security updates and patches that could have prevent the attack.

In addition to the 15% of ghost servers, there are typically additional servers that are low utilized. These servers, like ghost servers, consume unnecessary power, cooling, and space and lead to increased energy consumption and data center inefficiencies.

In order to identify ghost and underutilized servers, IT admins need to monitor server utilization, power consumption, and other telemetries. With this data operators can determine which servers will be decommissioned and what servers are underutilized, can be virtualized, or consolidated.

36.6 SUMMARY AND FUTURE PERSPECTIVES

As noted before, data center operation costs are still rising and are today a huge burden on data center costs. One cannot ignore this burden and needs to pay more and more attention to improving efficiency and reducing those costs. In order to do that, granular data and analytics are required. This data and AI on top of it can help the data center manager and IT operators drive efficiency in the data center and reduce operation costs by eliminating thermal issues, raising temperatures in the data center, increased rack density, and better capacity planning.

Looking forward, we already see OEMs adding more telemetries to IT devices, and we believe that those new telemetries, in addition to further AI and analytics, will help the data center manager and IT operations drive even further efficiency in the data center, reduce OpEx/CapEx, and even improve data center health and uptime.

FURTHER READING

https://www.intel.com/content/www/us/en/software/cutting-man-hours-and-optimizing-server-operation.html

https://www.intel.com/content/www/us/en/software/improving-cooling-energy-consumption-server-visibility.html

https://www.intel.com/content/www/us/en/software/helping-solve-mysteries-universe-case-study.html

www.intel.com/dcm

[5] http://www.itwatchdogs.com/environmental-monitoring-news/data-center/ghost-servers-affect-data-center-cooling-and-power-consumption-533360.

[6] https://www.computerworld.com/article/2555157/ghost-server.html.

37

PREPARING DATA CENTERS FOR NATURAL DISASTERS AND PANDEMICS

Hwaiyu Geng[1] and Masatoshi Kajimoto[2]
[1]*Amica Research, Palo Alto, California, United States of America*
[2]*ISACA, Tokyo, Japan*

37.1 INTRODUCTION

The best way to prevent deadly effect of natural disasters is to anticipate, be educated, and informed [1]. By applying this process data center operators can effectively prepare, prevent, and control an event.

In addition to natural disasters, technological disasters such as system broken down, power outages, natural gas explosions, and man-made disasters from terror attacks to race riots could be devastating.

In May 2020 amid pandemic COVID-19, two dams failed in Michigan during a 500-year flood event where 11,000 people were told to evacuate from their homes. Pandemic COVID-19 results in a new normal and a pandemic contingency binder must be available.

It is imperative to have a master plan covering business continuity (BC) and disaster recovery (DR) to anticipate, prepare, and contain, with minimal mitigation effort. Each crisis calls for the following stages: resolve, resilience, return, reimagination, and reform [2]. The more we learn from past experience to augment BC and DR, the better we will be prepared for preventing the inevitable and less mitigation. BC and DR plan must be prepared with the consideration of localization in the forms of government, democracy or authoritarianism, culture, and discipline differences among the countries.

The goal of this chapter is to broaden the awareness, prevention, and preparedness of data center owners toward natural, technological, and man-made disasters as well as pandemics.

37.2 DESIGN FOR BUSINESS CONTINUITY AND DISASTER RECOVERY

An ounce of prevention is better than a pound of cure. Data center infrastructure must be built robustly with consideration of BC and DR requirements. These requirements are normally beyond jurisdictional building codes and standards. International Building Code (IBC) or local building code such as California Building Code generally addresses life safety of occupants relating to predicted seismic or hurricane activities in the region. Those codes provide with little or no regard to property or functional losses. To sustain data center operations after a natural or man-made disaster, the design of a data center must consider system redundancy. Building structural and nonstructural components must be hardened and fortified considering the appropriate level of BC [3].

There are different levels of enforcement and redundancy, thus robustness, to fortify data centers that have also been addressed in other chapters of this handbook. To share some highlights, the Chapter 13 of ASCE 7-16 (*Minimum Design Loads and Associated Criteria for Buildings and Other Structure*) contains extensive coverage on the seismic design

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.
© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

and anchorage of mechanical and electrical nonstructural components. The U.S. Federal Emergency Management Agency (FEMA) developed the following self-explanatory design guidelines to fortify building nonstructural components in battling a seismic event:

- "Installing Seismic Restraints for Mechanical Equipment," FEMA, December 2002 [4]
- "Installing Seismic Restraints for Electrical Equipment," FEMA, January 2004 [5]
- "Installing Seismic Restraints for Duct and Pipe," FEMA, January 2004 [6]

Costs of designing and installing seismic restraints are minimal in particular for a new construction project. American Society for Testing and Materials (ASTM) has issued guidelines on how to estimate seismic loss. Probable maximum loss (PML) is a term that defines the value of maximum loss of property expected after a disaster. In addition to PML, loss of sales and market share should also be considered. With total extended loss, additional seismic hardening costs could be easily justified. One might also consider scenario expected loss (SEL), scenario upper loss (SUL), and probable loss (PL) beyond PML.

Electrical power outage commonly happened along with natural disasters. Alternative power sources such as using fuel cell technology or hydrogen fuel technology should be considered. Relating to Internet service provider (ISP), check if the services have multiple Internet connections such as fiber cable, T1 lines, satellite, or digital subscriber line (DSL) in addition to providing baseline service.

37.3 NATURAL DISASTERS

Different natural hazards can be categorized as follows:

- Geophysical: Earthquake, volcanic eruption, and tsunamis
- Meteorological: Typhoons, hurricanes, and cyclones
- Hydrological: Floods, landslides, and avalanches,
- Climatological: Tornadoes, droughts, temperature extremes, lightning, etc.

Any event could impact operations of data centers. This chapter will concentrate in lessons learned from the Great East Japan Earthquake combined with tsunami and the Eastern U.S. Superstorm Sandy. An earthquake or man-made disaster is unpredictable. Hurricanes or storms are predictable and allow advanced time for preparedness.

Tsunamis and volcanic activities are closely related to earthquakes. In an order of magnitude, 100% of Japan is in the seismic zone, 60% for China, and 30% for the U.S. with "Ring of Fire" [7] from Australia, Asia to Americas (Fig. 37.1). Earthquakes are caused by displacement of ground plates due to a "divergent, convergent, or transform fault plate boundary."[1] Earthquakes could trigger tsunamis [8], landslides, or volcanic activities.

"A tsunami is really a series of waves. A tide gauge measures the water level every minute, effectively measuring the height of each wave as it passes the gauge. After the predicted normal water level, including tides, is removed, the result shows the deviation from normal water levels due to a tsunami" (Fig. 37.2).

37.4 THE 2011 GREAT EAST JAPAN EARTHQUAKE

On March 11, 2011, a subduction zone earthquake of magnitude 9.0 occurred off the Pacific coast of Tohoku, often referred to as the Great East Japan Earthquake in Japan. This was the largest earthquake ever to have hit Japan since instrumental recordings began in 1900 and the fourth largest ever recorded according to the U.S. Geological Survey. Thirty minutes later, a devastating tsunami, at speeds up to 500 miles/h (800 km/h), struck the great east coastal line (406 miles or 650 km) of Japan and flooded 59,000 acres (24,000 ha) of agricultural land. The event left 18,884 dead, 2,636 missing, and thousands more injured. It leveled 127,290 houses. This mega disaster included an earthquake, a tsunami, a nuclear power plant shutdown, and disruption of global supply chains. It was the costliest natural disaster in world history. The economic cost of damages to buildings, roads, ports, and others was estimated at about US$235 billion. The loss of life and properties could have been far greater, where if not for the fact that Japan had an advanced disaster risk management system, built up from lessons learned in nearly 2,000 years of her history. Lessons from the Great East Japan Earthquake could be learned by the United States that has a similar pattern in the U.S. northwest coast from the state of Washington to Oregon. (Fig. 37.3).

37.4.1 Lessons Learned from Japan Earthquake and Tsunami

First, let us look at what happened **[10]**:

1. First three days:
   - Many enterprises lost key personnel (decision makers who were responsible for disaster response etc.) as a result of the tsunami.
   - The telephone/communication networks were congested as a result of the disaster.
   - The electricity supply was stopped.

[1] https://pubs.usgs.gov/gip/117/gip117_ebook.pdf.

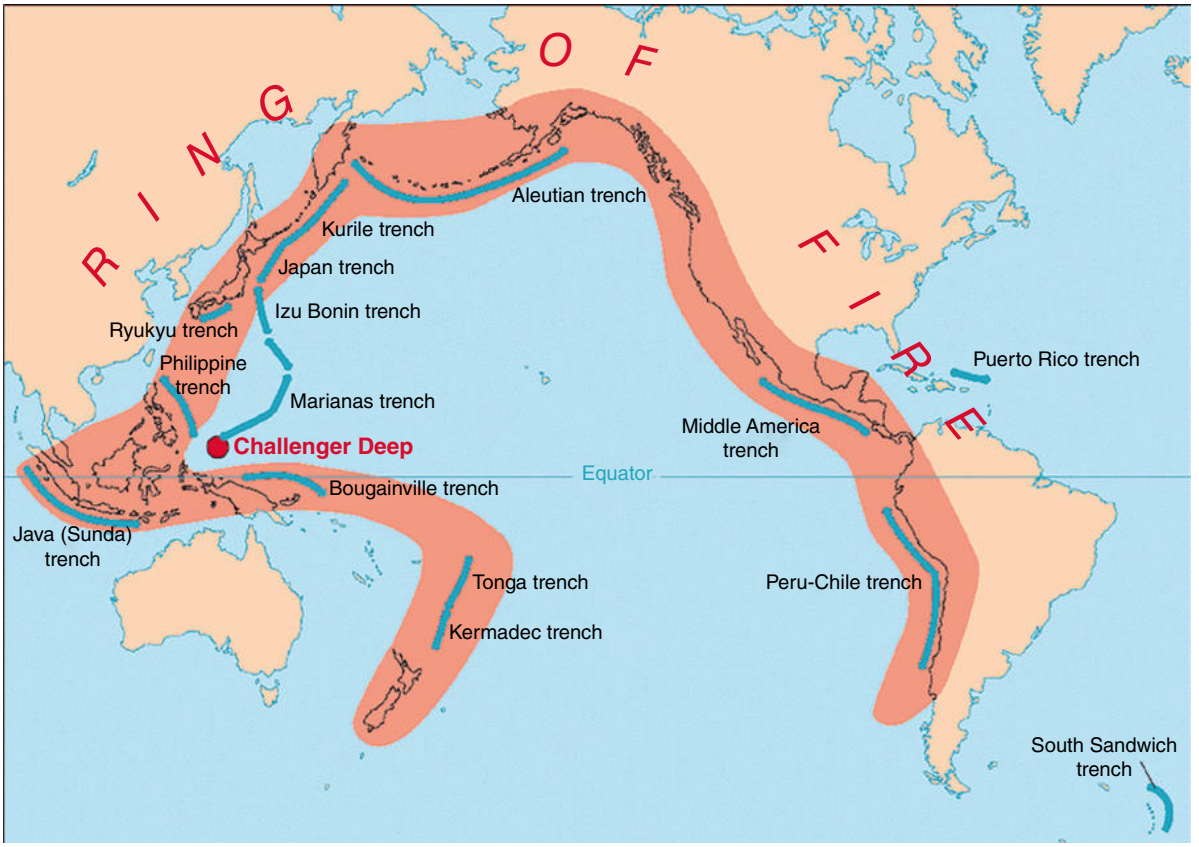


FIGURE 37.1 "Ring of Fire" where the Pacific Plate meets many surrounding tectonic plates. Source: USGS Natural Hazards. Available at https://pubs.usgs.gov/gip/dynamic/fire.html.

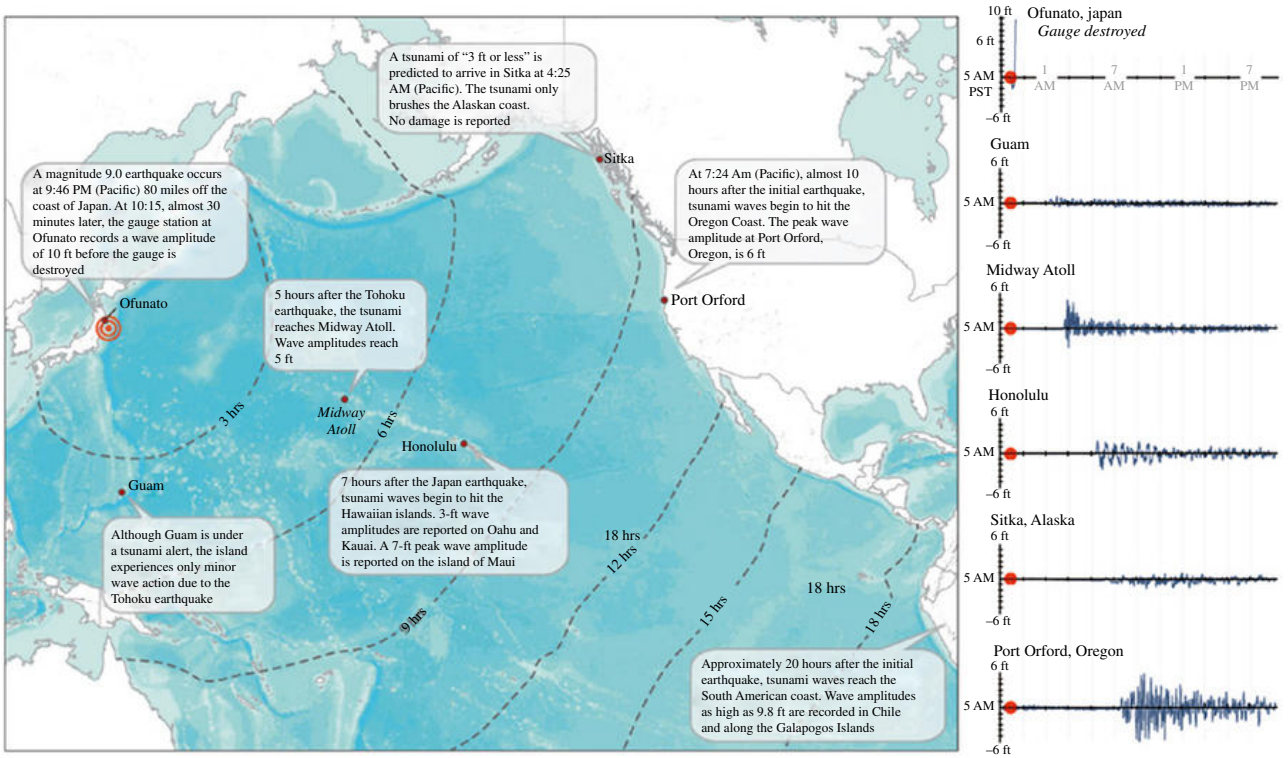


FIGURE 37.2 Progression of the Tohoku tsunami across the Pacific Ocean with tide gauge water level measurements showing deviation from predicted tide levels [9].

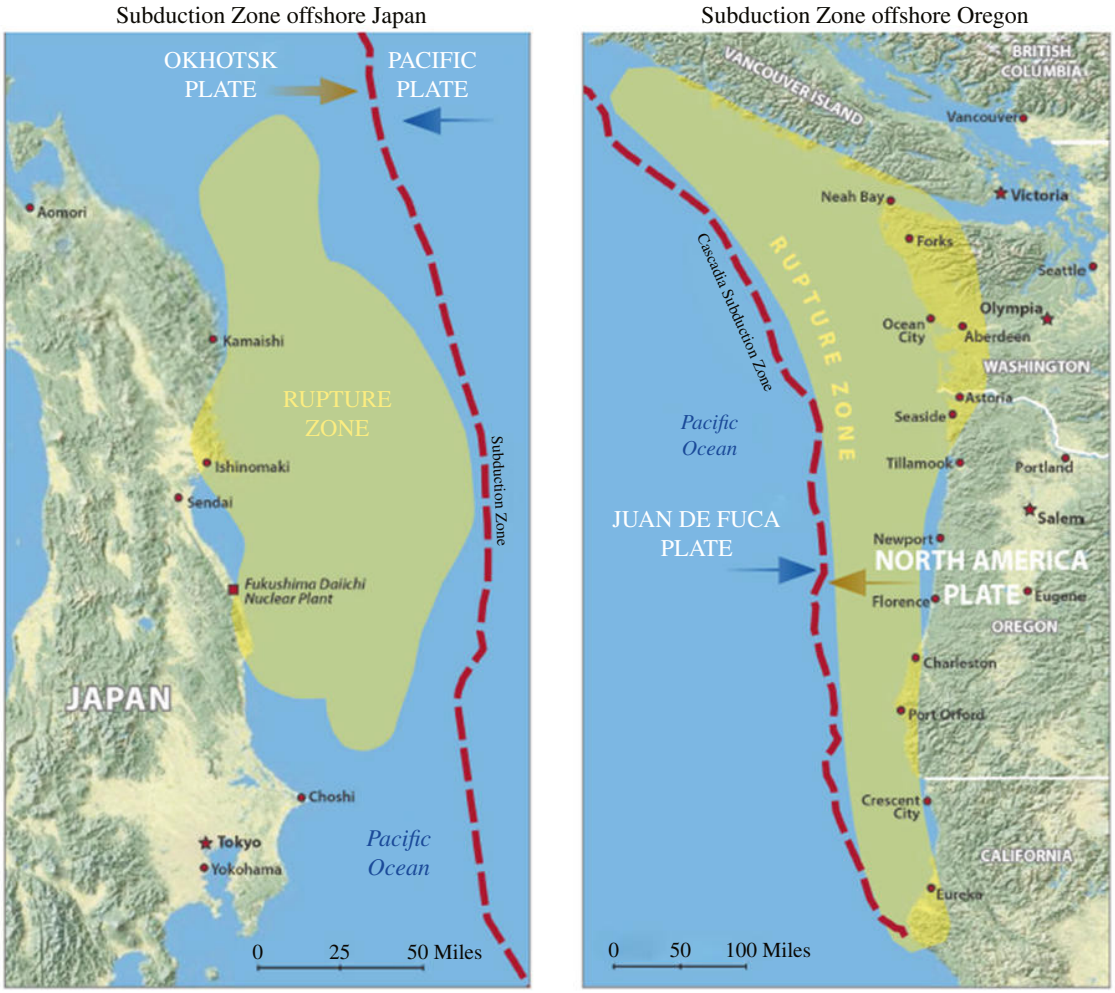


FIGURE 37.3 Lessons learned: Oregon's tectonic setting: a mirror image of Japan. (Left) Light gray zone is the exact of the Tohoku rupture zone. (Right) Light gray zone indicates a region where earthquakes can occur in the Pacific Northwest. Source: The Oregon Department of Geology and Mineral Industries. The 2011 Japan Earthquake and Tsunami: Lessons for the Oregon Coast; Winter 2012.

- Many organizations suffered simultaneously, meaning that not only was an enterprise suffering and/or unable to function but also were its vendors in many cases.
- Information technology (IT) resources were heavily damaged or lost.
- Transportation routes were heavily damaged.
- Earthquake aftershocks continued to occur.

2. First three months:
   - Rotational (but not well planned) blackouts were enforced.
   - Electricity shortages continued in the disaster area.
   - Legal "saving electricity" measures were enforced in the Kanto area (Tokyo and surrounding prefectures).
3. Four months and beyond:
   - Electricity shortage in the western side of Japan became severe. Because of the shutdown of nuclear power plants, many factories, data centers, and others from the Kanto area migrated to the western side of Japan.

Next, let us look at the impacts to IT-related businesses:

1. First:
   - Chains of command were lost. Almost nobody could decide appropriate measures for IT infrastructure recoveries.
   - Communication channels were lost. It was nearly impossible to get correct information such as the following: "Who is still living?" "Who is in charge?" "What happened?" "What is the current status?"

- Stock of fuel for emergency power supply was very limited (one or two days).
- Server rooms were strictly protected by electronic security systems, so without enough electricity, these security systems became obstacles for emergency responses as the aftershocks kept coming. At some organizations, they kept the door of their server room open.

2. Second:
   - Replacement facilities or equipment (servers, PCs, etc.) were not supplied quickly from the vendors, because so many organizations had suffered from the disaster.
   - Backup centers also suffered in the disaster. Therefore, quick recovery was almost impossible at many organizations.
   - Because of rotational blackouts, companies could not access their servers from remote offices.
   - Data recovery was a very heavy task. If a company's backup schedule was once per week, they lost almost one week's worth of data. In some cases, both electronic and paper-based backups of data were lost.
   - In the areas evacuated as a result of the nuclear power plant accidents, nobody could enter their own offices.
   - Many IT-related devices were washed away by the tsunami. Some fell into the wrong hands.
   - Diesel fuel supply was interrupted in many areas because transportation routes were not repaired quickly.
   - Many organizations moved their data centers and factories to the western side of Japan.
   - The emergency power supply could not operate for long periods of time. They were designed for short-term operation.
   - Monthly data processing was impossible in many organizations resulting in much delay.
3. Third:
   - At the western side of Japan, many organizations confronted electricity shortages and could not operate in full operations.

Now that we have a full picture of the devastation that occurred, let us look at the lessons we've learned:

1. Bad situations can continue for a long time. Quick recovery is sometimes impossible. Be prepared for this.
2. Prepare as many people as you possibly can who can respond to disasters. Having a fixed definition of roles and responsibilities may be non-operable.
3. Data encryption is indispensable.
4. A cloud computing-like environment can be very helpful in situations like this.
5. Uncertainty-based risk management is necessary.
   - In Japanese history, many huge earthquakes and tsunamis were recorded. We must study our history carefully and be prepared for those things that will happen.
   - The "Sumatra disasters," from 2004 to 2010, caused major earthquakes and tsunamis, including a magnitude 9.1 earthquake in 2004. We must learn from these disasters. And we must take account of the fact that a similar-size disaster can occur anytime in the same seismic zone.
   - Although we cannot predict exactly when, where, and how, we can prepare for the uncertainties.
6. Preparation of many risk scenarios may be useless. Too many risk response manuals will serve as a "tranquilizer" for the organization. Instead, implement a risk management framework that can serve you well in preparing and responding to a disaster.

Natural disasters can occur anytime and anywhere. Sit down with your colleagues and make a plan now.

37.5 THE 2012 EASTERN U.S. COAST SUPERSTORM SANDY

On October 29, 2012, Sandy, a category 2 superstorm, made landfall in New Jersey, pummeled the U.S. east coast from Florida to Maine with effect felt across 24 eastern states. Sandy forced winds extend 175 miles out its eye making it much larger than most storms of its type (Fig. 37.4). It drove a catastrophic "storm surge" into the New Jersey, New York, and Connecticut regions with 80 miles/h (129 km/h) sustained winds, and the heavy rain battered the densely populated states of New York and New Jersey. The storm surge (high winds pushing on the ocean's surface) reached 13.9 ft (4.2m) at Battery Park, New York, surpassing the old record of 10.02 ft by Hurricane Donna in 1960. It is the second largest Atlantic storm on record. The storm produced severe flooding along the Atlantic Coast, contributed to fuel shortage across the New York metropolitan area, and two feet of snow in areas of West Virginia, Virginia, Maryland, and North Carolina. The storm killed 117 people in the United States and 69 more in Canada and Caribbean, caused US$50 billion in damages, and left 8.5 million residents without electrical power. Sandy ranks as the second costliest tropical cyclone on record, after Hurricane Katrina of 2005. Detailed chronological events

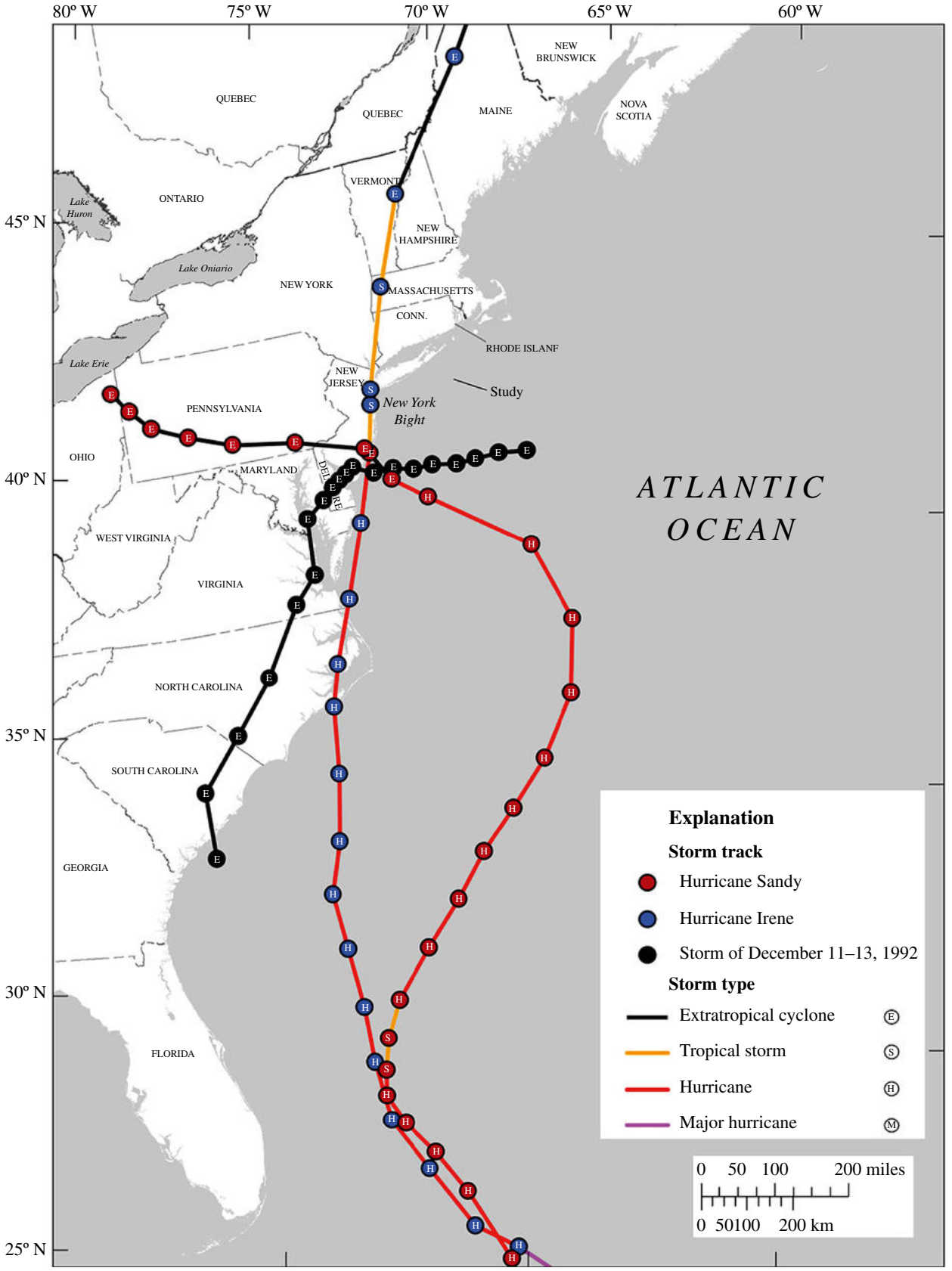


FIGURE 37.4 Sandy made landfall near Atlantic City, New Jersey, on October 29, 2012, as a post-tropical cyclone after traveling up the southeastern U.S. coast as a category 1 hurricane [11].

can be found at CNN's "Hurricane Sandy Fast Facts" [12]. Here are some highlights relating to data center's preparedness:

- Authorities suspended train, subway, commuter rail, and bus services. Nearly 11 million commuters lost transportation services. Airlines canceled flights.
- Three reactors experienced trips, or shutdowns, during the storm, according to a Nuclear Regulatory Commission statement.
- 7.9 million businesses and households were out of electric power in 15 states and the Washington D.C. It lasted 10 days until November 7 with still 600,000 people without power.
- Areas hit by Sandy experienced gasoline shortage due to loss of electrical power at gas stations.
- The U.S. Energy Information Administration reported that approximately 67% of gas stations in metropolitan New York did not have fuel for sale.
- New York City public schools announced via their official feed that schools would begin to open on November 5.
- A strong low pressure system with powerful northeasterly winds coming from the ocean ahead of a storm hits the areas already damaged by Sandy.

Modern weather forecasts provided warning of Sandy well before the storm hit. The devastation in areas such as Manhattan during Sandy involved flooded water (Fig. 37.5) in such an exceptional magnitude.

For installed data centers before the Sandy, most companies had done a great job having the data center out of the basement or even above the first floor. However, due to the need of rack-level cooling, there was a tendency of designing lower level as gray space to house supporting infrastructure including electrical switchgear, mechanical equipment (chillers and pumps), and uninterruptible power supply (UPS). This was still a good design so long as the gray space was not in the basement and was above ground.

For those cities' fire and building codes require fuel tank to be installed at the bottom of a building, the basement must equip with high capacity sump pumps to protect the basement that contains fuel tank and possibly the switchgear and mechanical room. To avoid flood damage to fuel tank delivery pumps, encase pump in a waterproof box or use a submersible fuel pump in the fuel tank with watertight power feed and tall vents. To hold water back, build a watertight fuel tank room so water couldn't get in. If flood water fills up the basement, the fuel tank could be lifted off the foundation and get disconnected. Provision should be made for a fuel truck to physically connect the fuel oil risers in a building to facilitate refill.

FIGURE 37.5 Hurricane Sandy caused flooding in New York City subway stations. Source: https://www.climate.gov/news-features/climate-case-studies/how-sandy-affected-new-york-city%E2%80%99s-long-term-planning.

Here are some lessons learned from Sandy that could be considered in preparing your BC and DR binders:

BC and DR Planning:

- Firms that heed a storm warning, communicate early and often, are better prepared and fared well.
- Invoke the BC/DR as soon as you hear a storm warning and follow the procedure previously established.
- Planning period for 48, 96, or 144 hour's duration.
- Provide customers with ample information in time for them to make better decisions on their DR plan.
- Work with your customers or business partners during regular DR drill. Work out a DR solution for the worst-case scenario.
- Review noncritical tasks and agree on when to bring them online during an emergency recovery process.
- Re-evaluate overall BC/DR preparedness after an event, including the performance of service providers, on what worked well, what did not.

Communications:

- Harden communication system (equip engine generator) for key staff and decision makers who are critical to incident management process.
- Communicate status to decision makers, staff, and customers by email; post status on a website with dashboard.
- Communication tools include email, direct phone call, website posting, and instant message.
- Use social media and GPS to communicate and locate employees.
- The Federal Communications Commission reported 25% of cell phone towers lost power rendering mobile phones are useless. Diversity of telecom providers includes 4G/5G and satellite phone.
- Provision to charge mobile phones.
- Staff phone line 24/7.
- Redundancy of voice and data infrastructures.
- Employees who were unable or unsafe to leave home due to storm damage, lack of transportation, or unwilling to leave their families could work through telecommuting and videoconferencing systems and can do what needs to be done remotely.

Emergency Power/Backup Generators:

- 12 hour on-site fuel storage minimum.
- Regular testing of switchover.
- Consider multiple fuel delivery contracts with diesel fuel suppliers from diverse locations.
- Check your backup generators for maintenance requirements and limitations to support 48, 96, or 144-hour plans.

Logistics:

- Obtain roll-up generators, extension cords, and needed spare parts delivered before storms.
- Acquire survival resources including food, drinking water, flashlights, sleeping bags, clothing, medication, and personal medical needs.
- Understand the BC plan of your suppliers and third parties during supply chain disruption.
- Stock gasoline for motor vehicles.
- Pre-arrange to ensure purchases during the crisis with limited credit line and a delay internal procurement system.

Preventive Maintenance:

- Conduct BC/DR training drills quarterly, semiannually, and annually.
- Regular BC/DR testing of systems, infrastructure with incident response team members.
- Fully test the mission-critical infrastructure that supports data center before a storm.
- Regularly operate diesel generators for 6–8 hours a time.
- Regularly conduct pull-the-plug test. The load goes on UPS and generators run for 10 hours.
- Before a disaster event, top off the fuel tank and test run generators.

Human Resources:

- Pre-plan incident support teams that can deploy staff from non-affected data centers at a short notice so staff can join before the road or airport is closed. (Terrorist attack may be an exception.)

Information Technology:

- Move IT loads to other sites prior to an event.
- Regular backup storage and mirroring processes may need to be changed due to vulnerable power loss.
- Move email system, documentation, and storage to an external cloud provider to ensure uninterrupted communication.
- Moving communication systems to cloud-based hosting may prove very valuable.
- Prioritize critical tasks to streamline recovery.
- Oftentimes, IT securities are more lax during and after a disaster using weak links in security and business. It is important to remain vigilant and apply security best practices during the BC/DR process.

There are many BC/DR practices relating to data centers and IT areas that could be considered to ensure resilience. These

include deploying data in redundant facilities, add-on modular data center containers, co-location, cloud, etc.

37.6 THE 2019 CORONAVIRUS DISEASE (COVID-19) PANDEMIC

The most threatened and worst scenario is a combination of natural disaster and nuclear power plant shut-down or pandemic happening at the same time. The examples include the 2011 earthquake and the Great East tsunami in Japan, April 13, 2020 disaster of Southeast tornado, during the COVID-19 pandemic.

A pandemic is a global outbreak of disease when a new virus emerges to infect people. There were pandemics in 1918 (Spanish flu, H1N1 that killed an estimated 50 million people), 1957 (H2N2), 1968 (H3N2), 2003 (SARS), 2009 (H1N1 pdm09 virus), and 2015 (MERS-CoV) with different levels of case fatality rate. Coronavirus disease 2019, abbreviated COVID-19, is a respiratory disease spreading from person to person caused by a novel coronavirus. COVID-19 can cause mild to severe illness especially susceptible to adults above 65 years old and people of any age who have underlying medical conditions. Pandemic is not like a hurricane or earthquake in a sudden burst but has far large and broad depressions and prolonged recovery weighing on the global economy.

37.6.1 New Normal during COVID-19 Outbreak

The COVID-19 outbreak results in new normal that includes closing border, severing supply chains, banning nonessential travel, discouraging presenteeism or coming to work sick instead of self-quarantine and working from home, cleaning hands against infection, practicing sensible social distancing to avoid contagion, and wearing personal protection while at workplaces.

Data shown from California, New York, and the United Kingdom with lockdown experienced 45–79% Internet traffic increase [13]. Accommodating people working from home contributed to rise in Internet traffic growth. Zoom meeting participants were 10 million in December 2019 increased to 200 million in March 2020. Microsoft Teams usage soared in one month by 775% in Italy. Demands for virtual private network (VPN) AT&T offering skyrocketed 700% during the crisis. Verizon monitored and adjusted resources during peak hours. Amazon, Netflix, and YouTube were controlling streaming bit rates in Europe to conserve bandwidth [14].

37.6.2 Lessons Learned from COVID-19 Outbreak

Lessons learned from COVID-19 pandemic encompass ignored early warning, ill-prepared for the event, not applying data-driven decision, not practicing sensible social distancing, not wearing masks to stop respiratory droplets (>5–10 μm) and droplet nuclei (<5 μm) virus from transmitting by air, and no orchestrated effort to obtain required personal protective equipment (PPE) for people who need them to perform their work safely.

A pandemic contingency binder must be prepared to cover necessary processes and procedures. Stocking PPE for pandemic is essential. PPE includes surgical masks, N95 respirators, protective eyewear, face shields, disposable gloves, and medical coverall suits. It is important to have areas for proper donning and doffing as well as procedures and waste management to properly dispose of used and contaminated materials.

In addition, non-contact infrared thermometers must be equipped, used, and calibrated daily for accuracy. Soap and hand sanitizer must be provided to keep the workplace clean and safe. Antibacterial or alcohol wipes to clean hands, tools, and bleach spray to sterilize equipment and the workplace must be equipped.

A clear chain of command and whereabouts of your data center staff and their substitutes is crucial to have a rapid centralized response to crisis and shift into "war-fighting mode."

37.7 CONCLUSIONS

Data center BC and DR to contain, manage, and mitigate natural or man-made disasters must be well thought out and prepared before an event. A good DR plan should be concise and succinct. Staff may not have time to read a long and complicated plan during an event. Investments must be considered to ensure the integrity of data center building structural and nonstructural components as well as IT equipment and infrastructure.

Disaster preparedness including incident response team, human resources, policy and procedure, communication protocol, training, and drill must be systematically reviewed and practiced at local and company-wide levels. Engage your customers or business partners with an emergency recovery and operations plan. Coordinate with various groups such as diesel oil suppliers, utility companies, and governments to ensure needed supplies and electrical power during rollbacks.

Building a culture of BC/DR preparedness for a sustainable data center is everyone's business.

Practice makes perfect. Involve personnel at data center operations, IT, and customers with planning and systematic training and drill. Pay special attention to communication and coordination.

For those who work at data center during an event, prepare and harden workers' home with emergency power set up to ensure the family members have heat, electricity, and needed supplies.

Lessons learned after each event must be discussed to see what worked well and what didn't. After-the-event studies should include potential vulnerabilities of key service

providers, issues on cloud services relating to emails, DR, adequate insurance policies for loss of use, flood coverage, etc. Incorporate the lessons to continuously improve BC/DR technology, policy, and procedure, and reach a new level of preparedness so not to repeat the same mistakes in the future.

During 2020 COVID-19 pandemic, Internet traffic surged due to work from home and online study. This traffic surge can overload a data center if provision is not properly planned.

It is better to be prepared than to be sorry—take extra steps when there is a natural disaster or pandemic warning. By doing our homework, data center downtime could be significantly reduced, and recovery time be shortened.

REFERENCES

[1] UNESCO. *Natural Disaster Preparedness and Education for Sustainable Development*. Bangkok; 2007.

[2] Craven M, et al. COVID-19 Implications for Business, McKinsey; March 30, 2020. Available at https://www.mckinsey.com/business-functions/risk/our-insights/covid-19-implications-for-business. Accessed on June 15, 2020.

[3] Braguet OS, Duggan DC. Eliminating the confusion from seismic codes and standards plus design and installation instruction. 2019 BICSI Fall Conference. Available at https://www.bicsi.org/uploadedfiles/PDFs/conference/2019/fall/PRECON_3C.pdf. Accessed on June 15, 2020.

[4] FEMA, Society of Civil Engineers, and the Vibration Isolation and Seismic Control Manufacturers Association. Installing Seismic Restraints for Mechanical Equipment. Available at https://kineticsnoise.com/seismic/pdf/412.pdf. Accessed on February 22, 2020.

[5] FEMA. Installing Seismic Restraints for Electrical Equipment; January 2004. Available at https://www.fema.gov/media-library-data/20130726-1444-20490-4230/FEMA-413.pdf. Accessed on February 22, 2020.

[6] Installing Seismic Restraints for Duct and Pipe, FEMA P414; January 2004. Available at https://www.fema.gov/media-library-data/20130726-1445-20490-3498/fema_p_414_web.pdf. Accessed February 22, 2020.

[7] The Oregon Department of Geology and Mineral Industries. The 2011 Japan Earthquake and Tsunami: Lessons for the Oregon Coast; Winter 2012.

[8] Progression of the Tohoku tsunami across the Pacific Ocean with tide gauge water level measurements showing deviation from predicted tide levels. Available at http://tidesandcurrents.noaa.gov/. Accessed on June 15, 2020.

[9] The 2011 Japan earthquake and tsunami: Lessons for the Oregon Coast, The Oregon Department of Geology and Mineral Industries: News & Information; Winter 2012. Available at https://www.oregongeology.org/pubs/cascadia/CascadiaWinter2012.pdf. Accessed on July 6, 2020.

[10] Kajimoto M. *One Year Later: Lessons Learned from the Japanese Tsunami*. ISACA; March 2012.

[11] USGS. Available at https://archive.usgs.gov/archive/sites/soundwaves.usgs.gov/2015/09/images/sandy-fig01.jpg. Accessed on July 6, 2020.

[12] Cable News Network. Available at http://www.cnn.com/2013/07/13/world/americas/hurricane-sandy-fast-facts/. Accessed on July 6, 2020.

[13] Bergman A, Iyengar J. How COVID-19 is Affecting Internet Performance. Fastly, San Jose; April 8, 2020. Available at https://www.fastly.com/blog/how-covid-19-is-affecting-internet-performance. Accessed on July 6, 2020.

[14] Miller R. *Data Network Traffic Impact from COVID-19*. Data Center Frontier; April 6, 2020. https://datacenterfrontier.com/covid-19-data-network-traffic-impact/. Accessed on July 6, 2020.

FURTHER READING

Blake E, Kimberlain T, Berg R, Cangialosi J, Beven J. *Tropical Cyclone Report Hurricane Sandy*. National Hurricane Center, National Oceanic and Atmosphere Administration; February 12, 2013.

Bosco ML. Hurricane Sandy Prompts New Way of Thinking about Data Center Resiliency. FacilitiesNet; May 2013. Available at http://www.facilitiesnet.com/datacenters/article/Hurricane-Sandy-Prompts-New-Way-Of-Thinking-About-Data-Center-Resiliency--14015. Accessed on July 6, 2020.

Deloitte. Disaster Recovery: 10 Lessons from Hurricane Sandy; November 2012. Available at http://deloitte.wsj.com/cio/2012/11/29/disaster-recovery-planning-10-lessons-learned-from-hurricane-sandy/. Accessed on July 6, 2020.

Eruptions of Hawaiian volcanoes-past, present, and future. U.S. Geological Survey; 2010.

Japan Metrological Agency. Lessons Learned from the Tsunami Disaster Caused By the 2011 Great East Japan Earthquake and Improvements in JMA's Tsunami Warning System; October, 2013. Available at https://docplayer.net/20934619-Lessons-learned-from-the-tsunami-disaster-caused-by-the-2011-great-east-japan-earthquake-and-improvements-in-jma-s-tsunami-warning-system.html. Accessed on July 6, 2020.

Kidman A. Top 10 Data Center Management Lessons from Hurricane Sandy. Lifehacker; April 2013. Available at http://www.lifehacker.com.au/2013/04/top-10-data-centre-management-lessons-from-hurricane-sandy/. Accessed on June 15, 2020.

Learning from the Great East Japan earthquake and tsunami policy perspectives—summary statement. Japan-UNESCO-UNU Symposium; February 2012. Available at http://ioc-tsunami.org/index.php?option=com_oe&task=viewEventRecord&eventID=1035. Accessed on June 15, 2020.

Lessons Learned Information Sharing System (LLIS.gov).

Rimler B. *Lessons Learned from Hurricane Sandy*. Schneider Electric Data Center; November 29, 2012. Available at http://blog.schneider-electric.com/datacenter/2012/11/29/lessons-learned-from-hurricane-sandy/. Accessed on June 15, 2020.

Supprasri A, Shuto N, Imamura F, Koshimura S, Mas E, Yalciner A. Lessons learned from the 2011 Great East Japan tsunami: performance of tsunami countermeasure, coastal building, and tsunami evacuation in Japan. Pure Appl Geophys July 2012.

Sverdlik Y. *Japan Data Centers Face Rolling Blackouts*. Datacenter Dynamics; March 2011.

Tellefsen and Company. Industry Impact and Lessons Learned from Hurricane Sandy; January 2013. Available at https://secure.fia.org/downloads/Industry-Impact-and-Lessons-Learned-From-Hurricane-Sandy_Summary-Report.pdf. Accessed on July 6, 2020.

The World Bank. The Great East Japan Earthquake Learning from Megadisasters; 2012.

Tilling R, Heliker C, Swanson D. Eruptions of Hawaiian volcanoes—past, present, and future. U.S. Geological Survey; 2010. Available at http://pubs.usgs.gov/gip/117/gip117_ebook.pdf. Accessed on July 6, 2020.

Yamanaka A, Kishimoto Z. *The Realities of Disaster Recovery: How the Japan Data Center Council is Successfully Operating in the aftermath of the Earthquake*. JDCC, Alta Terra Research; June 2011.

Water mitigation in data centers. Cabling Installation and Maintenance Magazine; April 1, 2019. Available at https://www.cablinginstall.com/data-center/article/16466567/water-mitigation-in-data-centers. Accessed on July 6, 2020.

INDEX

General Rules:
– *Italic* numbers refer to figures
– **Bolded** numbers refer to tables
– Use "*see*" or "*see also*" to recommend related entry
– 168n denotes entry is a footnote in page 168
– Numeric entry is alphabetized as though spelled out (e.g. "5G" as "fifth generation")

absolute humidity, 177–178, 180
active state efficiency (ASE), 325
  thresholds, *325*
actuators, 273
  applications, 3, 165, 168–170
  damper, 426, 431
  design, 574
  fire protection, 419
adiabatic evaporate cooler, *214*, *428*
adsorbent-loaded nonwoven fiber media (ALNF), *245*
  applications, 242, 246, *248*, *250*, 245, 251, 255, 260
adsorbent media, 243
advanced technologies, 1, 2
  air cleaning technologies, 242–*243*, 245, 250–251
  air filtration, 242, 243 (def.)
  anatomy, 2
  ecosystem, 2
  taxonomy, 2
air cleaning system design, 249–250
  closed systems, 249
  outside air no pressurization, *250*
  outside air with pressurization, *250*
Air Conditioning, Heating, and Refrigeration Institute (AHRI), 624
  standards 340, 61
air conditioning optimization, 659
air containment, 45–46, 177, 205, 207, 228, *230*, 435, 651
air cooling equipment spec, **180**
air exchanges, 244, 245
air management, 188 (def.), 419–421, 426, 645–656
  ASHRAE recommended temperatures, 646
  Bernoulli law of fluid dynamics, 647
  best practices, 9, 36, 45, 183
  bypass, 9, 207, 246, 409, 433, 599–600, 606–607, *646*
  CFD, 609, 651 *see also* computation fluid dynamics
  containment, 230, 651
    cold aisle containment, *651*, *653*
    cold aisle containment leaks, *653*
    cold aisle semi-containment, *651*
    direct aisle containment, *652*
    hot aisle containment, 412, *652*
    pro-con among containments, **654**
  cooling delivery issues, 645
  design, 416–417, 420
    CRAC CRAH, 429, 645, 648, 650, 655 *see also* CRAC and CRAH
    hot cold aisle, 436
    economization, 437
air management, design (*cont'd*)
    in row cooling, 45, 230
    negative pressure, 243
    non raised floor, 415
    overhead plenum, 432
    rack cooling, 605
    safety, 405, 412
    through the space, 432–433
  energy costs hours, 108, *110*, 111, 169
  energy efficiency, 29
  fire protection, 419
  hybrid, 415, 435
  IT layout, 420
  infrastructure systems, 23
  improve performance, 652–655
    case studies, 654–655
    Code of Conduct best practices, 654
    EN50600 TR 99–1, 654
  liquid cooling, 645
  measurement, **57**
  metrics, 648
    air segregation efficiency (ASE), 648, *650*, 655
    demand performance, 650, 652, *655*
    RCI performance, 621, 650, *654*
    tier III IV, 188
    return heat index (RHI), 605, 651
    supply heat index (SHI), 605, 651

Data Center Handbook: *Plan, Design, Build, and Operations of a Smart Data Center,* Second Edition. Edited by Hwaiyu Geng.

© 2021 John Wiley & Sons, Inc. Published 2021 by John Wiley & Sons, Inc.

air management, metrics (*cont'd*)
supply performance, 649–650, *650*, 652, *655*
strategies, 7, 44
PUE, 108, **109**, 654
safety, 405, 412
server design, 329
strategies, 44
through the space, 432
work environment, 405
air movement, 183
baffles, 207
air negative pressure, *648*
case study, 655
CFD, 9
definition, 647
design, 243, 252
human unawareness, 128
temperature controls, *649*
air quality
control, 239
design, 212, 248
energy savings, 171
measurement, 172–173
monitoring, 164–165, 257
services, 252
sensors, 165, 167
cleaning technologies, 242, 247
air recirculation (air mixing/management) 645–656
best practices, 9, 671
CFD, 594, 598, 607
design, 110, 229
hotspots, 663
improve performance, **60**
measurement, 663
metrics, **662**
overhead cooling, 229
theory, 659
air recirculation (energy saving), 208
best practices,
CFD, 601
definition, 246
design, 243–244, 249, 609
improve performance, **60**
air side economizer (ASE), 169, 211–225, 248 (def.), 425, 426, 430 *see also* economizer
direct (DASE), 212, *213*, 214, *222*, 224
direct evaporative cooler (DEC), 212
dry bulb temperature, 213
wet-bulb temperature, 213
energy saving, *171*
filter, 179
indirect (IASE), 212, 215, *216*, *222*
air-to-air heat exchanger (AHX), 213, 215
indirect evaporative cooled heat exchanger (IECX), *218*, *219*, 220, *221*, 224
indirect evaporative cooling (IEC), 217, 224
Air Tightness Testing and Measurement Association (ATTMA), 43
air to air energy recovery equipment, 215
airborne contaminants, 240 *see also* contamination
airflow distribution, 183, *333*, 387, 589, 594, 655
front to back, *388*
right to left, *388*
airflow management *see* air management
airlock entry, 251
alarm response procedures (ARP), 130, 137
alternative compliance path, 188
Amazon Web Services (AWS), 21, 77, 145, 158–159, 345, 268
American National Standards Institute (ANSI), 175, 188, 189, 195, 291
ANSI 50/51 protection codes, 463–464
ANSI/ASHRAE Std 90.1 Energy Standard for Building, *see also* ASHRAE
background, 184, 188–190, 622
balanced distribution, 433
best practice, 184
building envelope info, 43
design, 8, 407, 433, 622–623
energy efficiency, 410, 426, 438
energy standard for building, 8, 43
measurement, 61 *see also* ASHRAE guideline 14
waste heat reuse, 405
ANSI/ASHRAE Std 90.4 Energy Standard for Data Centers, *see also* ASHRAE
background, 184, 188–190, 622
balanced distribution, 433
design, 187–190, 223, 406–407, 433, 622–624
electrical loss component, 617–619, 623
energy consumption, 623
energy efficiency, 410, 426, 438
maintenance, 190
mechanical load component, 617–618, 623–624
performance based standards, 189, 429
PUE, 617
waste heat reuse, 405
ANSI/ASHRAE Std 127 Method of Testing for Rating Aircon Serving DC, *see also* ASHRAE
particulate gaseous contamination, **252**
precision air conditioning, 187, 420
ANSI/ATIS-0600404, 203
ANSI/ISA-71.04, 240, 252, *255*, 257
ANSI/TIA 568, 10, 195
ANSI/TIA 569, 195
ANSI/TIA-606, 195, 209
ANSI/TIA-607, 195
ANSI/TIA-758, 195
ANSI/TIA-862, 195
ANSI/TIA-942, 8, 10, 195
ANSI/TIA-5017, 195
ASHRAE (American Society of Heating, Refrigeration and Air Conditioning Engineers), 175–191 *see also* other ASHRAE-entries
air cool ITE categories, 437
attenuation of telecommunication cabling, 207
BACnet, 351
balanced distribution, 433
building audit, 55
building envelope leakage, **43**
chiller minimum energy requirement, 33
contamination, 184, 245
cooling guidelines, 212, 438
copper silver coupons, 249
Data Center Infrastructure Management, 186
design guidelines, 44, 425, 439
dew point absolute humidity, 178, 410
electrical loss component, 617–619, 623
energy consumption measure, 55
energy efficiency, 438, 657
energy measure report, 61 *see also* ASHRAE guideline 14
environmental guidelines
equipment specifications, 180
psychrometric chart, *177*, *214*
environment instrumentation, 637
environment monitoring, 637
equivalent organizations, 108
gaseous particulate contamination, 8, 184, **252**
green tips for datacenters, 58
human comfort, 435
humidity control, 44, 51, 425. 431
humidity sensor, 276
humidity temperature ranges, 9, 51, **60**, 183, 206
IT OT connectivity protocol, 350
ITE operating condition, 409
inlet air temperature, 9, **60**, 176, 621
legionnaires disease, 224
measure report PUE, 57 *see also* ASHRAE guideline 14
mechanical load component, 617–618, 623–624
particulate gaseous contamination, **252**
particulate removal efficiency, *221*, 242, 248, 252 *see also* MERV
power consumption trends, 18, 409
silver corrosion, 251
small office building, 43
standards and codes, 187
standards and practices, 175–191
supply air temperature moisture level, *214*
temperature envelope, 430, 646
temperature sensor placement, 275, *276*
thermal envelope, **24**
thermal guidelines, 35, 229, *244*, 605–607

ASHRAE (*cont'd*)
typical meteorological year (TMY) data, 216, *222*, *223*
waste heat reuse, 405
ASHRAE communication protocols, 350–351
ASHRAE cooling guideline, 43, 212, *214*, 426
ASHRAE datacom series, 33, 58, 179
Advancing DCIM with IT Equipment Integration, 186, 637
Best Practices for Datacom Facility Energy Efficiency, 182
Design Considerations for Datacom Equipment Centers, 181
Green Tips for Data Centers, 58, 184
High Density Data Centers—Case Studies and Best Practices, 183
IT Equipment Design Impact on Data Center Solutions, 186
IT Equipment Power Trends, 180
Liquid Cooling Guidelines for Datacom Equipment Centers, 181, 439
Particulate and Gaseous Contamination Guidelines for Data Centers, 8, 183, 248–249, 252 *see also* contamination
PUE™: A Comprehensive Examination of the Metric, 185, 406
Real-Time Energy Consumption Measurements in Data Centers, 184
Server Efficiency-Metrics for Computer Servers and Storage, 185
Structural and Vibration Guidelines for Datacom Equipment Centers, 182
Thermal Guidelines for Data Processing Environments, 179, **244**, 229, 437, 438, 606
ASHRAE guideline 12 on Legionella, 224, 251
ASHRAE guideline 14 energy demand measurement, 55, 61
ASHRAE guidelines green tips, 58, **60**, 207, 224
ASHRAE Handbook, 175, 187, 223–224, 252
building air leakage, 43
particulate gaseous contamination, 252
ASHRAE Humidity Control Design Guide, **24**, 35
ASHRAE Journal, 175, 190
ASHRAE rated particulate removal efficiency, 248, 252 *see also* MERV
ASHRAE sensor placement diagram, 275, *276*
ASHRAE standards and practices, 175–191
ASHRAE TC 9.9
contamination, 178, 245
copper silver coupons, 249
datacom series, 179
design standards guidelines, 407
consumption, 439
ASHRAE TC 9.9 (*cont'd*)
energy savings, 184
gaseous contamination monitor, 249
Green Grid, the, 185, 406
green tips, 58
GX- SEVERE, 257
humidity control, 178, 206, 419
ITE operating conditions, 409
introduction, 175
liquid cooling, 437
particulate filtration, 242
static discharge, 178, 206
temperature limits, 9, 51, 207, 435, 439, 605–606, 621, 646
temperature humidity, 198
thermal guidelines, 39, 175–176, *244*, 255, 439
network equipment, 8
power requirement, 8
ASHRAE thermal envelope standards, 24, *214*, 229, 430, 437, 439
anchor nonstructural components, 522, 525, 527, 531
application-based approaches, 151
application-specific integrated circuits (ASIC), 83, 144, 324, 326, 329–330
architecture design, 8, 381–402, 411, 562, 577 *see also* rack floor plan
CFD modeling, 400 *see also* Computational Fluid Dynamics
computer room design, 395
coordinating systems, 392
aisle containment, 395
CRAC and CRAH, 393 *see also* CRAC, CRAH
fire protection, 394
lighting, 394
power distribution, 393
raised floor *vs*. ceiling, 394 *see also* raised non-raised floor
fiber optic network design, 381
pathways, 390
overhead, 391
underfloor, 391
rack and cabinet design, 307, 386
scalable design, 398
scalable *vs*. reliability, 399
space and power design, 389, 437
space planning, 400
battery room, 401
burn-in room, UPS, 400
circulation, 400
entrance way, 400
loading dock, 401
MEP rooms, 401
network operations center (NOC), 400–401
receiving, 401
repair room, 400
security CCTV, 402
staging, 402
architecture design, space planning (*cont'd*)
storage room, 401
support space, 400, 401
Art of War, 6
artificial intelligence (AI), 143, 323, 363
artificial internet of things (AIoT), 3
chatbot, 145
DCIM, 186, 643
digital twin, 608
dynamic server management, 341
emerging technologies, 1, 15, 82
overview, 5
Q-learning network architecture, *344*
reinforcement learning (RL), 343
augmented reality, 5, 85
Authority Having Jurisdiction (AHJ), 187, 188, 394, 533
automatic investigation and remediation, 150
automatic transfer switch (ATS), 281
autonomous vehicle, 1
availability, *132 see also* reliability engineering
mean time between failures (MTBF), 132 *see also* MTBF
mean time to fail (MTTF), 132 *see also* MTTF
mean time to repair (MTTR), 132 *see also* MTTR
reliability, 132, **133**
six 9s (99.9999%), 132
two 9s (99%), 132

balanced distribution, 433
base of design (BoD), 131, 136, 138
battery types
lithium-ion (Li-ion), 472
valve-regulated lead–acid (VRLA), 472
Beaty, Don, 175
behavior modeling, 150
Bellcore/Telcordia data, 175, 254
benchmark metrics, 9, 617–624
electrical loss, 620
energy uses in data center, *619*
green grid, the, 10, 55, 57, 176, 286, 406, 605, 617, 620
metrics assessments, 620
carbon usage effectiveness (CRE), 621
energy reuse effectiveness (ERE), 620, 631, 657
floating point operations per second (FLOPS), 338, **622**, *623*
inverse PUE (DCiM), 287
partial PUE (pPUE), 224, 618, 623
Power Usage Effectiveness (PUE) equations, 617 *see also* PUE
Rack Cooling Index (RCI), 605–606, 620, 621, 651
Return Temperature Index (RTI), 605, 621
water usage effectiveness (WUE), 620–621

benchmark metrics (*cont'd*)
metrics tools
ASHRAE 90.1, 622–623 *see also* ANSI/ASHRAE Std 90.1
ASHRAE 90.4, 617–619, 622–624 *see also* ANSI/ASHRAE Std 90.4
electrical loss component (ELC), 189, 617–619, 622–624
European Code of Conduct for Data Centers, 9, 13, 59, 654
linear equations software package (LINPACK), 622
mechanical load component (MLC), 617–619, 622–624
Standard Performance Evaluation Corporation (SPEC), 622
TOP500/Green500™, 622
PUE applies in facility, 619
UPS load evaluation, 619
best practices
Energy Star, 9, 323
Federal Energy Management Program (FEMP), 9
guide for energy efficient data center design, 8
structural design, 182, 183
bidirectional (BiDi), 297
big data, 86
analytics, 3
anatomy, 4
characteristics, 4
blanking panel, 60, 207, 654, 671
brace nonstructural components, 522, 525, 527, 531
breakeven analysis, 100, **101**, 118, ***119***
building automation system (BAS), 633, 639
building envelope cooling and cooling load, **43**
building envelope effect and energy use, 43
building envelope energy model, 44
Building Industry Consulting Services International (BICSI)
access space, 208
best practices, 8, 196, 295
cabling standards, 294
class F2, 393
earthing, 386–387
entrance room routing, 390
UPS power distribution, 397
building leakage and relative humidity, *44*
building management system (BMS), 10, 135–136, 139, 350, 357, 401, 628, 633, 638–639
building signaling systems, 194
Building Technology Office (BTO), 49
build-to-order/engineered-to-order, 264
build, upgrade, lease or rent, 124
build *versus* buy, 72, 73
bulk fill media module, 245, 250
business as usual (BAU), 169
business continuity (BC) and disaster recovery (DR), 675–685
architecture, 6
building codes, 521
COVID 19, 675, 683
lessons learned, 683
personal protective equipment (PPE), 683
considerations, 376
design for BC and DR, 675
ASCE 7–16, 675
ASTM, 676
California Building Code (CBD), 675
FEMA seismic MEP restraints, 676
insurance, 535
International Building Code (IBC), 675
probable loss (PL), 676
probable maximum loss (PML), 676
scenario expected loss (SEL), 676
scenario upper loss (SUL), 676
Japan earthquake tsunami, 676–679
lessons learned, 676–679
Oregon tectonic setting, 678
progression of tsunami, *677*
master plan, 675
natural disaster, 676
requirements, 254
ring of fire, 676, *677*
risk, 549
Superstorm Sandy, 679–683, *680*
hurricane Sandy fast facts, 681, *681*
lessons learned, 682–683
uninterruptible power supply (UPS), 681 *see also* UPS
UPS, 483
virtualization integration, 518
business driver, 6–7
business intelligence, 4–5, 145, 629
bypass air
air management, 9, 409, 433
containment type, 653
design, 246, 599–600
improve cooling effectiveness, 45, 207
measure RTI, 606–607
metrics, 648
return air temperature, 647

cabinet density, 307
cabinet design, 386–388
attenuation (dB), *318*, *319*
air pressure, 207
fiber, 313, *314*, 315, 318
link loss, 301, 309
reduce avoid, 204, 208, 235, 308
reduction, 204
temperature factors, 207
signal, 317
cabinet design (*cont'd*)
bundling in basket ladder, *306*
cable *see* fiber cabling fundamentals
connect switch ports (LAN/SAN), *296*
design and installation, 205–208
duplex zipcord cable, 293
EIA/CEA-310-E, 387
end of row (EoR), 210, 291, 298, 388
end to end link, 298, *299*, 313, 317
fiber transport services (FTS), *292*
IBM Global Services (IGS), 292
jumper storage, 317
LAN/SAN cabling, 296
maximum length, 206
methodologies, 292
Open Compute Project, 387 *see also* OCP
patch panel, 292, 297, *298*, 308–309, 311, 316, 318
central patching locations (CPL), 292, 295, 311, 316
zone patching locations (ZPL), 292
pathways, 208, 390
computer room, 391design, 307
entrance network BICSI class 1/2/3/4, 390
overhead fiber cabling pathway, *307*
overhead pathway, 207, 208, 391
under-floor pathway, 207, 208, 391
point to point (PtP), 291
structured cabling, 291, *293*, 317
top of rack (ToR), 291, 298
tray, 208
cable depth inside tray, 208
optical fiber, 208
type, 201
balanced twisted pair, **203, 204**
coaxial, 203, **204**
optical fiber, 204
unstructured cabling, *292*, *293*
unstructured nonstandard cabling, *194*
cabling ANSI/BICSI 002, 8, 196, 295, 386–387, 397
cable infrastructure, 201, *299*
permanent links, *299*
redundancy, 201, *202 see also* redundancy
topology, *201*
cable management
best practices, 9–10, 305–307, 590
cabling standards, 294–295
cord retention, 274
design installation, 208–209
horizontal, 209, 317
jumper cable, 317
network, 387–389
operations management, 9, 10
planning, 304
power cord, 388
rack PDU, 272

cable management (*cont'd*)
raised floor CFD, 596
safety, 206
vertical, 209, 317, 386
cable reliability, 209
cabling standards, 200, 203, 210, 294, 300, 310–312, 315
CENELEC EN 50173-5 Information Technology, 295
ISO/IEC 24764 Information Technology, 295
National Electrical Code (NEC), 306
TIA-526, 315
TIA-568, 312
TIA-942, 295, 307–308
TIA/EIA 569, 308
TIA/EIA 606, 308
TIA-FOTP 171, 312, 316
$T_NE_NA_NM_N$, 209
cabling redundancy, 201
cabling system commissioning, 313
light source power meters (LSPM/PMLS), 312, *314*, 315, 317
Tier I-II testing, 312, *314*
cabling system design, 295, *299*
entrance room, 295
equipment distribution area (EDA), 295–296
horizontal distribution area (HDA), 295–296
main cross connect (MC), 295
main distribution area (MDA), 295–296
mixed fiber grade network, 320
zone distribution area (ZDA), 295, *298*
cabling system troubleshooting, 316
cabling testing
optical time domain reflectometer (OTDR), 312–313, 316, *317*
OTDR signal, *318*
reflective loss, 319
tier I and II, 312–313, *314*, 317
Canadian Standards Association (CSA), 195
capability-based resource abstraction, *147*, 148, 150
capacity kilowatts, 123
capacity own, cloud, lease, or color, 125
capacity planning, *7*, 186, 389
capacity space, 123
refresh, 389
capacity upgrade, 112, *113*
Monte Carlo analysis, *113*, *114*
capital expenditure (CapEx), 6, *7*, 484, 472, 477, 673
Carbon Disclosure Project (CDP), 31
cause and effect diagram, *5*, 140
central processing unit (CPU), 323
chaos engineering, 150
channel, 298 (def.)
channel insertion loss (CIL), 310, *312*
Chartered Institution of Building Services Engineers (CIBSE) Guide A, 43
chatbot, 145
chiller energy by economizer hours, **110**
chiller power optimization, 659
chiller temperature PUE, 109
churn rate, 184
cleanliness class, 183, 244
climate change, 1, 21 *see also* sustainability
climate data (ASHRAE), 107, 108, 216, 223
climate sensitivity, 108, *109*
cloud and commoditization, **119**
cloud computing, 1, 77, 79, 155
architectural principles, 82, *158*
benefits, 81
definition, 78
multi-cloud architecture, *159*, 162
on-demand computing, 27
pay-as-you-go, 81
region and zone, 83
transformation, 81
cloud deployment model
hybrid, 80
multi-cloud, 80, *160*
private, 80
public, 80
single cloud environment, *159*
cloud service model
IaaS, 80
PaaS, 80
SaaS, 80
coaxial cable and circuits, **204**
codes definition, 187
coefficient of performance (CoP), 108, 169, 172, 655
cogeneration, 428
cold aisle containment, 45, 46
cold plate, *234*, *235*
collaboration computing, 145
colocation (wholesale), 66, 368, 399
comatose servers, 22
CENELEC EN 50173-1, 207
commissioning, 135, *136*
communication protocols, 165
competitive forces strategy, 6
Compliance for IT Equipment (CITE), 186 (def.)
composable architecture, 150
composable disaggregated data centers, 150
computational fluid dynamics (CFD), 579–609
CFD applications, 588
airflow management, 588
airflow perforated tiles, 602
airflow underfloor, 432
artificial intelligence, 608
assessment, 588
best practices, 9
cable penetrations, 600
CFD, CFD applications (*cont'd*)
ceiling height, 666
CRAC/CRAH, 243n, 589–590, 592, 594 *see also* CRAC and CRAH
control room, 600
cooling pipe obstruction, 596
data cable obstruction, 597
IT equipment (ITE), 592
ITE heat pattern, 598
ITE inlet temperature, 592
internal cabinet recirculation, *599*
machine learning, 608
obstructions to airflow, 596
power cable obstruction, 596
power distribution unit, 592, 595, 599
staggered array of obstructions, 596
temperature changes floor to ceiling, *667*
troubleshooting, 588, 590
typical applications, 588
UPS room cooling, 588
CFD DCIM integration, 642
CFD design
airflow CRAC CRAH, 429
airflow modeling, 334, 400, 583, 590
airflow patterns, 586, *595*
ASHRAE, 605–606
architecture, 592
cable routing, 608
calibration, 591
costs capitalization, 116
conceptual design, 588–589
conjugate heat transfer, 583
controller response, 601
data center design, 171, 588
data center modeling, 602
data hall specific analysis, 605
detailed design, 589–589
facility behavior modeling, 601
heat inertia, 604
heat transfer, 583
heat transfer coefficient (U-value), 593
ITE power draw, 592
internal architecture, 593
large eddy simulation (LES), 590
maximum inlet temperature, *605*
modeling, 592
numerical discretization, 602
planning, 7, 9
post-processed data, 587
proportional–integral–derivative (PID) vs. CFD, 608
rack cooling index (RCI), 605–606
return heat index (RHI), 605, 651
return temperature index (RTI), 605–607
sensors readings modeling, 165
simulation, *233*
space planning, 400
supply heat index (SHI), 605, 651

CFD, CFD design (*cont'd*)
time dependent simulation, 603
total cooling failure, 604
water mist simulation, 601
white space modeling, 598
CFD fundamentals principles, 580, 580–587
finite element method, 581–582
finite volume method, 581
Lattice Boltzmann method (LBM), 582
low energy data center design, 601
magnitude of velocity (flux), 586
Navier–Stokes differential equations, 580
redundancy failure scenario, 603 *see also* redundancy
result plane, 586, *587*
streamlines, 586, *588*
structured Cartesian grid, 584
surface plots, 587
symbol library, 592
CFD operational management, 588
airflow analysis, 229, 231
costs, 116, **117**, 118
cold aisle capture index (CACI), 607
cooling unit load, *606*
DCIM, 10, 591, 607–608, 628, 642 *see also* Data Center Infrastructure Management
digital train, 10, 579, 607–608
hot aisle capture index (HACI), 607
hot spots, 642
liquid cooling system, 609
metric RCI, 621
metric RTI, 621
MMT, 661
overheat risk, *606*
PUE, 607
Predict DCIM, 591, 607
recirculation, 609 *see also* air management
simulation, 607
computerized maintenance management system (CMMS), 5
computer processor history, **212**
computer room air conditioning (CRAC) *see also* HVAC
air management, 433, 594
airflow temperature, 651
case study, 114
commissioning, 136
control system fight, 427
CFD, 9, 589–590, 592, 594 *see also* Computational Fluid Dynamics
corrosion
filter, 260
sensors, 166
CRAC vs CRAH, 420, 429
control, 163, *168*, 170, *173*, 419
computer room air conditioning (CRAC) (*cont'd*)
definition, 243n, 425, 425n
deployment, 393, 589–590, 592, 642
discharge top return bottom, 432, 439
efficiency measurement, **56, 60**
financial justification, 106–116
case study, 114
upgrade, 107, 114
location sensitivity, 108
free cooling, 251
in row unit, 429
installation, 433
life, 94
maintenance, 137
monitoring, 592
operations, 437
operating costs, 425
PUE, 219
quantify energy savings, 171
rack level cooling, 227, 230
sensor deployment, 164–165, *167*
set temperature, 437
system selection, 430, 433, 589, 645
software control, 170
UPS connections, 620, 633
usable capacity, 415n
water cool CRAC, 41
computer room air handler (CRAH) *see also* HVAC
airflow management, 594, 655
CFD, 589–590, 592, 594 *see also* CFD
corrosion sensors, 166
definition, 243n, 425, 425n
deployment, 589, 392–393
energy use, 650
metrics, 648
operating costs, 425
PUE, 219
rack level cooling, 227, 230
sensor deployment, 164–165
system selection, 589, 645
UPS connections, 620, 633
computer room design, 395
computer vision, 145
condenser water return (CWR), 35
condenser water supply (CWS), 35
connectivity design solutions
fiber channel PODS, *384*, *385*
main cross connect (MC), 295
modular panel-mount, 292
MPO (multi-fiber push-on), 292–295, 297
MPO to MPO, *298*
MTP, 292–296
small form factor (SFF), 292
TIA-526-14-B, 315
ZDA extensible fiber distribution, 298
ZDA interconnect, *298*
connector insertion loss (dB), 313, *315*
connector inspect lean connect, 319, *320*
contaminant, 179
contamination
best practices, 8
computer room, 198
control design, 243
ECC filters, *245*
free cooling, 248
gaseous, 183
guideline
ASHRAE, 178**, 252**
ISO 14644 class 8**, 252**
prevention, 183
TC 9.9 datacom books, 178 *see also* ASHRAE TC 9.9
continuous optimization, **148**
coolant, water and liquid, **235**
cooling load factors (CLF), 235
copper silver coupon, 183, 240
corporate real estate (CRE), 91
corrosion and contamination control, 239–262
Classification Coupon (CCC), *248*, *249*, 257, **259**
control process flow, *255*
corrosive gases, 240, 257
active sulfur compounds (H_2S), 257
ammonia and derivatives, 258
hydrogen fluoride (HF), 258
inorganic chlorine compounds, 257
nitrogen oxides (NOX), 257
photochemical species, 258
strong oxidants, 258
sulfur oxides, 257
monitoring system, *248*
rate, 165, 167, *168*
reacitivity monitoring, *261*
sensors, 165, *167*
cost and risk, 112
cost of capital, 93
COVID -19 pandemic, 186
creativity, 12
creativity for invention and innovation, *12*
curtain system & NFPA 75, 405n
customers relationship management (CRM), 256
cyber physical system (CPS), 1
cybersecurity in data center, 349–358
cyberattack SolarWinds, 355
IT and OT, 349
operation equipment (IoT, PDU, UPS), 353, 354
operation equipment protocol
BACnet, 350–351, 353
Modbus/TCP (Transmission Control Protocol), 350–351, *352*
Simple Network Management Protocol (SNMP), 350–355
SNMP network management system, 355

cybersecurity in data center (*cont'd*)
operational technology (OT), *350*
OT attacks, **351, 354**
OT network segment, 356
OT software (DCIM, BMS), 357
perimeter firewall system, 356
protect OT systems, 355
2-factor authentication, 357
virtual private network (VPN), *357*, 358

data analytics, 3, 4, 5
anatomy, 4
augmented, 5
descriptive analysis, 5
diagnostic analysis, 5
predictive analysis, 5
prescriptive analysis, 5
process, *4*
data center
alternatives, 90, 124
characteristics, 148
classifications, 256
ANSI/BICSI 002 rated 1/2/3/4 levels, 254
Chinese national standards A/B/C levels, 8–9
Japan Data Center Council tiers, 8–9, 13, 685
TIA 942 rated 1/2/3/4, 9, 209, 210, 254, 291, 534–535, 546, 578
Uptime Institute tiers 1/2/3/4, 9, 25, 72, 72n, 73, 188, 380, 407, 505, 535, 546, 549 *see also* Uptime Institute
cooling systems, *227*
definition, 2, 239
energy consumption trend, 2
energy efficient environment responsible, 29
energy efficient measure and report, **56, 57, 324**
energy metrics standards, 55
energy use and PUE, *37*, *38*
environmental assessments, 240, 253
environmental control, 253
environmental testing, 253
history, 253
hyperscale, 2
infrastructure architecture, 6
life cycle, 131, 321
network infrastructure, 386
no chiller, 109
operations
cost, 91n
efficiency, 49, **60**
performance assessment, 54
planning errors, 37
reduce energy use, 21, 27
redundancy, 73 *see also* redundancy

data center (*cont'd*)
scaling, 72
containerized, 68
monolithic modular, 67
stand alone,71
Tier I/II/III/IV types, 9, 73
traditional design, 66
data center commoditization, **119**
data center design *see also* architecture design
air cleaning, 243
air-side economizer, 248, 426
battery plants, 181
burn-in rooms, 181
cabinet placement, 206
cabling, 206
ceiling height, 197
design and construction, 8
design guidelines, 8
emergency generator rooms, 181
floor layout design, 196, 254
grid measurement, 392
multiplatform computer room, *390*, *392*
floor loading, 197
gaseous contamination control, 245 *see also* contamination
guidelines, 198
particulate control, 244
proper sealing protected space, 244
requirements, 254
room air pressurization, 243
room air recirculation, 243
spare part storage, 181
space planning CFD, 400
temperature and humidity control, 244
temperature and humidity recommendations, **244**
temperature, dew point and humidity, 206, *223*
test lab, 181
virtual desktop infrastructure (VDI), 331
data center energy usage, 28
bottom up model, 16, 19–20
characteristics, 18
extrapolation-based model, 17, 19
forward looking analysis, 19, *20*
strategies, 21
Data Center Infrastructure Management (DCIM), 627–643
ASHRAE, 184, 637 *see also* ASHRAE datacom series
best practice, 10
characteristics, *628*
business applications integration, 633
cloud based, 631
graphical user interface, 637
IT OT software integration, *633*
real time data collection, 632
server location and services, 633

Data Center Infrastructure Management (DCIM) (*cont'd*)
CFD integration, *642 see also* Computational Fluid Dynamics
cybersecurity, 357
definition, 627, *628*
dash board, *279*
digital twins and DC life cycle, 641
Federal IT Acquisition Reform Act (FITARA), 631
FedRAMP compliance, 631
fundamentals
DCIM (def.), 627, *628*
goals, 629, *630*, 630
ownership, 629
implementation, 639
considerations, 640–641
costs, 640
selections, 640
instrumentation interface
building management system, 638–639
environment sensors, 637
mechanical system, 638
rack PDU, 638
rack PDU power management, 637
remote access and power management, 639
sensing temperature, humidity, airflow, 637
servers, 638
uncover server intelligence, 637
integration other systems, 643
management process maturity, *628*
modules, 186, 631
asset management, 632–633, 643
capacity planning, 632
change management, 632
dashboard and reporting, 635, *636*
data collection, storage, import, export, 633–634, 636
equipment catalog library, 634
floor space planning, *634*
instrumentation interface, 637
location control, 386
platform interface, 637
rack planning designing, 634, *634*
operations, 641
platform architecture, 636
Sarbanes–Oxley (SOX) Act, 628
wireless interface, 637–638
work flow management, *632*
Data Center Energy Practitioner (DCEP), 11
Data Center Energy Profile (DC Pro), 11
Data Center Infrastructure Efficiency (DCiE), 258
data center life cycle, 641
data center management (DCM), 669–674
AI in DCM, 670
AI in SDDC, 670

data center management (DCM) (*cont'd*)
capacity planning, 671
facility operations management, 669
identify underutilized servers, 673
IT operations management, 669
thermal issues, 671, *672*
data center planning design implementation operations, 1, 29, 39
data center service demand on macro-level indicators, 15
data center strategic planning forces, *7*
datacom equipment, 183
deep-bed bulk media air filtration systems (DBSs), *250*, 251
deep bed scrubber (DBS), *247*, 250
deep learning, 145, 331
Delta T (ΔT), 214, 224, 599, 648
demographic trend, 1
denial of service (DOS), 277
dependability engineering, 5 *see also* reliability engineering
availability, 5
maintainability, 5
reliability, 5
descriptive analytics, 5
design collaborative, 48
design complexity, 134, *135*
design cooling system, 212
dew point temperature, *177 (def.)*, **180**, 206, 213
diagnostic analytics, 5
differential mode delay (DMD), 301
differential pressure controllers,251
digital twin (DT), 10, 641
digital twin and data center life cycle, *641*, 642
direct expansion (DX), 34, 218–220, *222*
direct expansion trim, 219
disaster management, 10 *see also* business continuity
discount rate, **99**
real and nominal, 105
distributed antenna systems (DAS), 194
Distributed Energy Resources (DER), 365, 497
Docker container, 150
Docker Kata Container, *346*
dot-com boom, 376
dry-bulb temperature, *177 (def.)*, **180**, 212
dynamic binding, 149

economic development officials, 367, 369, *371*, 372, 375
economizer, 188, 404, 430–432, 436–438
air management, 46
air-side economizer, 23, 41, 169, 211–225, 425
applications, 23, 45, 51
best practices, 9, 36, 38
combined benefits, 104–105
economizer (*cont'd*)
data collection, **57**
definition, 41
direct indirect economizer, 42
free cooling considerations, 43
indirect refrigerant-side economizer (IRSE), 221
indirect water-side economizer (IWSE), 221
temperature binned hours, 107
thermodynamic process, 213
water side economizer, 23, **35**, 42, 426
economizer cycles, 183
edge computing, 1, 3, 12, 77, 84
edge data center, 86
edge device compute concepts, *84*
effective modal bandwidth (EBM), 301
entry way, 401
80 PLUS® certification, **328**
Einstein, Albert, 12
electrical design in data centers, 441–481
automatic transfer switch changeover
close transition, *457*
open transition, *457*
backup technologies, 445
design considerations, 442
active power, 450
ANSI 50/51, 464
architecture resilience, 443
CapEx OpEx improvements, 444
design and installed capacity, *443*
facility up time, 441
grid reliability, **444** *see also* reliability engineering
grid reliability performance, 443–444
load balance calculation, 450
modularity and scalability, 443
options to size HV/MV transformers, ***455***
power factor, *443*
reactive power, 450
reliability, 442 *see also* reliability engineering
static *vs*. rotary UPS, **445**
tier classification, 446–447
UPS backup to grid interruption, 444–445
uptime level, 441
design inputs, 441
backbone requirements, 441
cooling load, 442
IT load calculations, 442
loads, 442
power system questions, *442*
strategy, 441
Facebook electrical design, 477, *479*
electrical planned operations, **478**
electrical room, 480
UPS backup scheme, *480*
UPS battery cabinet, *480*
electrical design in data centers (*cont'd*)
generators, 454–455
alternator voltage regulator, 458
generator redundancy and reliability, 455–456 *see also* reliability engineering
generator starting sequence (LV MV), 457
generator to LV, *456*
generator to LV vs. MV, **456**, *457*
generator to MV, 455, *456*
LV starting sequence, 457
MV starting sequence, 457
on-load test, 457
overcurrent undervoltage protections, 458
power rating, 458
HV grid connection, 452
air-insulated switchgear (AIS), 454
gas-insulated metal-enclosed switchgear (GIS), 454
on-load tap changer (OLTC), 454
2 incomers transformers double bus bar, *453*
2 incomers transformers single bus bar redundancy, *453*
2 incomers transformers single bus bar with a tie redundancy, *453*
HV redundancy *see also* redundancy
substation double bar redundancy, *454*
substation single bar redundancy, *453*
HV/MV grid architecture, 443
fault at HV lines substation, 444
fault at MV feeder, 444
LV radial architecture, 444
MV/LV open loop topology, 444
IEC62271–1 standard voltages, 455
ISO 8528 generator set ratings
continuous operating power (COP), *459*
emergency standby power (ESP), *459*
limited-time power (LTP), *459*
prime rated power (PRP), *459*
key performance indicators, 443
LV breakers
air circuit breakers (ACB), 468
miniature circuit breakers (MCB), 468
molded case circuit breaker (MCCB), 468
LV power system design, 464
ATS, 466–467
circuit breaker capacity, 476
cooling loads, 464–465
design optimization, 476
distribution systems, 465, *467*, *468*
electromagnetic compatibility (EMC), 465, *466*
IEC60529, 469
IEC60947, 468

electrical design in data centers, LV power system design (*cont'd*)
IEC61439, 468–469
IT racks PDU, 464, *465*, *466*, 638
load separation principle, 465, *466*
OCP design, 464, *466*, 477 *see also* Open Compute Project
power monitoring system, 476
switchboard architecture, 467, *469*
switchboard with UPS, 469
Terra Neutral – Separate (TNS), 465, 467
variable frequency drives (VFD), 465
LV protection, *474*
circuit breaker liming capacity, *476*
Discrimination vs. Switchboard Safety, 475, *475*
discrimination with static UPS, 474
Use Circuit Breaker Limiting Effect, 475
MV grid connection, 452
protection selectivity, 452, *453*
single MV grid substation single MV incomer, *452*
single MV grid substation two MV incomers, *452*
substation, 444, 450, 451, *452*
two MV grid substations single MV incomer, *452*
two MV grid substations two MV incomers, *452*
MV power system design
automatic transfer switch (ATS), 464
distribution topologies, 460, *462*
power monitoring, 464
power system, *463*
protection system, 463
switchboards, 461
MV switchboards
applications, 461
installation types, 461
insulation technology, 461
loss of service continuity (LSC), 461
MV/LV transformers, 462, 464
planned maintenance, 447
power system failures, 441
programmable logic control (PLC), 464
redundancy, *452 see also* redundancy
double cored server, 447
full redundant MV generator power plant double feeder, *462*
full redundant MV generator power plant using ring architecture, *461*
fault tolerance architecture, 446–447, *450*
iso-paralleling bus with DRUPS, *449*
N, 445, *447*
N+1 architecture with LV generators, *459*
N+1 architecture with LV generators redundancy MV, 460
electrical design in data centers, redundancy (*cont'd*)
N+1 architecture with MV generators, *461*
N+1 block, *448–449*
N+1 catcher system, 445
N+1 diesel rotary, *449*
N+1 distributed, 446, *448*
redundancy topology, 445, *459*
topology comparison, **451**
two redundant grid substations, 445
2N, 445, *447*
2N architecture with LV generators, *459*
2*N* architecture with LV generators single MV, *460*
2*N* architecture with MV generators, *460*
reliability and availability *see also* reliability engineering
generator, 455
generator connection, *456*
generator power rating, *459*
grid interruptions, 444, **445**
LV connections, 456
LV MV generators distribution, 459
MV connections, 456
optimization, 476–477
selectivity time-graded, *475*
short-circuit current value, *462*
TCO optimization, 477 *see also* financial analysis
topology comparison, **451**, *479*
topology of utility grid, *444*
transformer inrush current, *458*
transmission system, 444
UPS design
active filter modes, 472
battery assembly, *473*, *480*
battery system design, 472
battery VRLA or Li-ion selection, 472
conversion mode, *471*
design parameters, 470
economical (ECO) mode, 471, *472*
ECO mode risks, 471
failures, 470–471
power system efficiency, 471
scalability, 470
static bypass switch, *471*
static transfer switch, 473, *474*
static UPS system, 470
UPS backup scheme, *480*
UPS downstream short circuit, *474*, *475*
UPS fault scenarios, *475*
UPS internal modularity, *470*
electrical distribution system, 183
electrical loss component (ELC), 189, 617–619, 622–624
electrical rate, 498
electrical room, *480*
electrical system efficiency, 46
electricity expenditure reduction, 361
electricity green profile, 361
electricity grid disruption, 360
electricity supply resilience, 361
electromagnetic, **270**
electromagnetic compatibility (EMC), 36
electromagnetic induction, 486
electromagnetic interference (EMI), 10
Electronic Industries Association (EIA), 10, 291
EIA/CEA-310 rack cabinet standards, 386–387, 389
electrostatic discharge (ESD), 23, 24, 35, 178, 198, 206, 207
11th-hour design change, 405
emergency operating procedures (EOP), 137
emerging technologies, 2, 183
emerging technologies applications, 145
enclosure labeling, 308
end to end link, *299*
energy consumed, 17, 35, 38, 110–111, 163, 169, 185, 279
energy consumption, 2
cooling system, 40
server inlet air temperature, 40
strategies, 163
energy efficiency decision making time, *30*
energy efficiency equipment, 23, 323
servers, 22
storage devices, 22
energy efficiency equipment benchmarks, 324
LINPACK, 324
PCMark, 324
SPEC, 324
SPEC GEOMEAN, 326
energy loss, 46
energy monitoring system (EMS), 357
energy reduction by server management, *51*, *52*
energy reduction impacts analysis, **51**
Energy Star, 22, 48, 49
energy usage *vs*. data center air temperature, *39*
engineered link, 311
engineering channel, *311*
enterprise resource planning (ERP), 145, 186
entrance pathway, 390
BICSI class 1/2/3/4, 390
entrance room *see* TER
environment specification equipment, 244
environmental costs, 95
Environmental Protection Agency (EPA), 617
EPA report energy use and costs, 28, 409
environmental reliability monitor (ERM), *248*, 249, 251, *255*, 257, 260, *261*
environmental sensors, 164

equipment distribution areas (EDAs), 199, 200, 201 *see also* telecommunication cabling standards
Eucalyptus, 159
European Commision Code of Conduct, 96, 107, 120
best practice guidelines, 8–9, 624, 654, 656
energy efficiency
measure, 59
UPS, 33
metrics, 55
European Committee for Electrotechnical Standardization (CENELEC), 195
European Telecommunications Standards Institute (ETSI), 84
exascale computing, 2
extensible markup language (XML), 144
external network interface (ENI), 198
externality costs, 96n
extracted transformed loaded (ETL), 5
extruded carbon composite media (ECC), 243, *245*, 246–248, 250, 251

Facebook, 8, 21, 23–24, 83, 145, 288, 328, 332, 337, 341–342, 346, 379, 477, *479–480*
facility commissioning operations maintenance, 183
failure discover, 150
failure mode and effect analysis (FMEA), 132, 134
failure mode and effect criticality analysis (FMECA), 132, 134, **135**
failure, pseudo-random, 150
failure rate, 134
fan gallery, *392*
fan powered chimney IT cabinets, 45
fault tolerant, 9
  electrical design, 446–447, *450*
  cloud computing regions, 83
  multi-cloud architecture, 159
  reliability, 578
  risk management, **131**
  TIA-942 rated, 254
  UPS, 470
fault tree analysis (FTA), 132, 134
Federal Emergency Management Agency (FEMA), 529
  duct and pipe restraints guidelines, 8
  electrical equipment restraints guidelines, 8
  mechanical equipment restraints guidelines, 8
Federal Energy Management Program (FEMP), 9
Federal IT Acquisition Reform Act (FITARA), 631
FedRAMP compliance, 631
ferroresonant transformer, 494
fiber cabling fundamental, 291–321
  cable details, *303*
  cable types, 302
  connectors, 302
    field pre-terminated, *302*
  management, 305
  modular patch
    any to any, 297
    switch to host, 297
    switch to switch, 297
  multimode fibers, 296
    OM3 OM4, 296, 296
  plant labeling identification, 308
    selection criteria, 300
fiber channel reach, 310
fiber enclosures (rack mounted), 305
fiber interfaces
  multimode, **382**
  single mode, **382**
fiber optic connector intermateability standard (FOCIS), 294, 302
  charcoal, 179
  desiccant, 179
  effectiveness testing, 248
  filter life, 248
  strands, 381
  trunk cables, 381
fiber optic network design, 381
fiber types, **300**, 301
financial analysis, 89, 97, 89–127
  Activity-Based Costing (ABC), 89, 121, 123, 124
  airflow upgrade project, 114
  brand value, 95
  charge back model, 122
  cloud computing, 90
  colocation capacity, 125
  cost of carbon, 97
  cost per delivered IT kWh, 89
  end of life costs, 95
  energy costs, 108
  environmental, 96
  goodwill costs, 95
  initial capital investment, 94
  operating costs, 95
  renewable energy
  return analysis, 117
  Return on Investment (ROI), 29, 89, 91, 92, **93**, 93–95, 97, 100, 101, 104–105, 107, 109, 112, 114, 116, 498, 501, 516
  taxes, 95, 96, 105
  Total Cost of Ownership (TCO), 6, 89, **92**, 94–95, 97, 121, 137, 180, 462, 477, 501, 513, 515
  upgrade projects
    capitalized, **114**
    costs, 95
    IRR analysis, **118**
financial analysis, upgrade projects (*cont'd*)
    NPV analysis, **117**
    operating cost analysis,**117**
    PUE analysis, **114**
    return analysis, **118**
field programmable gate array (FPGA), 82–83, 144, 151, 330
5G, 3, 6, 86
fine grain isolation, 150
fire protection, 533–549
  active and suppression, 537
  agent and gaseous fire suppression, 537
    carbon dioxide, 543
    comparison, *545*, **545**
    FE25, *545*
    FM200, 541, *545*
    halon, 537, 541, *545*
    hydrofluorocarbon (HFC), 541–542
    hypoxic air, 543
    inert gas, 541
    Ingergen, *545*
    Novec 1230, 542, *545*
  automatic sprinkler systems, *538*
    deluge, 539
    dry pipe, 538
    pre-action, 539
    wet pipe 538
  codes and standards, 534
    FM Global (Factory Mutual), 534–534, 540–541
    International Building Code (IBC), 534, 536
    International Fire Code (IFC), 534
    NFPA codes, 534–538, 540–542, 544, 546 *see also* NFPA
    Underwriter Laboratory (UL), 537
  corrosive gas, 533
  design, 549
    aisle, 537
    egress, 536
    exit, 536
    galvanized piping, 539
    occupancy, 535
    occupant load, 536
    stairway, 536
    travel distance, 536
  detection, 546
    early warning fire detection (EWFD), 547
    gas detection, 547
    heat detection, 546
    smoke detection, 546
    standard fire detection (SFD), 547
    very early warning fire detection (VEWFD), 547
  electrical fire growth, **534**
  fire tetrahedron, *540*, 544
  fixed aerosol extinguisher, 543
  fundamentals, 533

fire protection (*cont'd*)
global warming potential (GWP), 542
HFC, FE, FM-200, 542–543
hot cold aisle ventilation, 544
hybrid fire passive protection, 544
in-cabinet fire suppression, 543
life safety code, 535, 538 *see also* structural design
passive fire protection, *537*
ASTM E119, 537
UL design U465 wall, 537
portable fire extinguisher, 544
pre-action sprinkler systems, 539
double interlock, 539
non-interlock, 539
single interlock, 539
sequence of operations, 547
signaling, 546
fire alarm matrix, **548**
TIA-942 rated standards, 535, 537, 542 *see also* TIA
Uptime Institute tiered system, 535
water mist systems, 540
fit-for-purpose systems, 144
FIT4Green, 59
five senses, 3
five V's, 4
floating-point operations per second (FLOPS), 50, 338
floor layout, 392
flux, 586
fog and mist computing, 1, 12, 78
forces of nature, 377
Foreign Account Tax Compliance Act (FATCA), 145
form factors, xvii, 182
42-pole panel, 393
free cooling, *110*, 176, 178, 188, *237* *see also* economizer
full load versus active idle, *50*
full self-driving, 86

Gartner hype cycle, 5, 7, 143, 627, *628*, 640, 669
gas phase filtration, 240
gaseous contamination, 183 *see also* contamination
control, 240, 246
guidelines, 183, 241, **252**
limits, **242**
gateway, 164
gauge repeatability and reproducibility (GR&R), 316
General Data Protection Regulations (GDPR), 352–353
global data center energy use, 17, *18*
forward-looking analyses, 19, *20*
Global e-Sustainability Initiative (GeSI), 58 *see also* sustainability
global strategic locations, 379
global warming, 11, 96, 183, 542, 545
globalization, 1, 368, 379
Google cloud platform (GCP), 80, 86, 145, 346
GR-326 standard, 316
GR-1275-CORE, 306
GR-3160 Telcordia NEBS Requirements, 254
graphics processing unit (GPU), 83, 144, 329, 582
green field development, 6, 408, 621
Green Grid, the, 10, 55, 57, 176, 185, 286, 406, 605, 617, 620
green tips, 284
Green Virtual Machine Migration Controller (GVMC), 157
greenhouse gas (GHG) emissions, 2, 29, 31, 35, 96–97
grounding, 24
ANSI/BICSI 002, 386
ANSI/NECA/NICSI 607, 210
ANSI/TIA 607C, 195
ASHRAE, 178, 409
bonding point, 386
design, 308, 395, 403
IEEE 1100, 386
NEC, 270, 501
TIA TSB 153, 207
Guo Biao GB 50174 Code, 8

Health and Human Services, 352
Health Information Technology for Economic and Clinical Health Act (HITECH), 352
Health Insurance Portability and Accountability Act (HIPAA), 352–353
heat density, 183, 434, 439
heat dissipation, 32, 84, 234–236, **236**, 590, 598, 603, 605, 607–608
heating ventilation air conditioning (HVAC) system
ASHRAE, 175, 187
balance, 276
contamination, 245
control, 183, 194, 251, 285, 402, 534, 537
cybersecurity, 351
data collection, 41–42
design
ASHRAE, 43, 179, 181
collabroation, 48
cost performance, 498
customer design, 407
dedicated system, 198, 249
fire protection, 547, 549
innovation, 624
rating level, 254
heating ventilation air conditioning (HVAC) system, design (*cont'd*)
shutdown, 541
value engineering, 413
discharge air temperature, 36
energy management, *37*, 40, 183, 288–289, 411
maintenance, 577
natural disaster, 11
outside air, 251
parametric analysis of different cooling system, *35*
power consumption, 619
reliability, *575*, 578
shutdown, 547, 549
types, **35**, 41
water-cooled, 34, 621
heat rejection, 31, 34, 46, 111, 219–220, 411, 414, 418, 429–431, 580, *658*
high-density switch port mapping, *297*
high-efficiency particulate air (HEPA) filtration, *245 see also* MERV
high performance computing (HPC), 294, 324, 333, 389, 397, 622
horizontal distribution area (HDA), 193, *194*, **197**, 198, 199–200, *201*, *202*, 208, 210, 295–296, 298, *299*, 304, *384*, *385*
hosting, 2, 22, 33, 65–75, 83, 85, 239, 439, 615, 631, 682
hosting colocation data centers, 65–75
hot air solder leveling (HASL), 240
hot cold aisle containment, 45, 46, 207, *434*
hot cold aisle design, 205, 213, *217*, *230*, 395
air management, 23, 395, 436, 666
design, 46, 207, *229*, *230*, 308, 659
hot spot, 169, 659
hourly average data form, 107
humidity ratio vs. dew point temperature, 34
hype cycle Gartner, *628*
hyperscale, 2, 12, 16, **17**, 18–19, 22, 83, 346, 382, 497
convergence, 186
hypervisor, 78, 150–151, 156–157, 330–331, 345, 452, 518, 656
hypervisor system based, 151

immersion liquid cooling, 9, 46, 182, 198, 234, *235*, 236, 240, 395
Indian Society of Heating Refrigerating and Air Conditioning Engineers, 108n
Industrial Internet of Things (IIoT), 3
industrial revolution (IR), 2
industry standard architecture (ISA), 327
information and communication technologies (ICT), 2, 8, 12, 21, 58–59, 90, 118, 178, 337, *346*
standards, 195
strategies, 90

IT design loads, 189
IT OT
  communication, 184
  cost and efficiency, 51, 53, 91
IT power design, 189, 484, 499, 572–573, *574*, *575*
IT power management, 39, 67, 111, 286, 288, 657–658, 662, 672
IT power metering, 122, 123
IT powerload, 189, 286, 406, 489, 657, 660, 665
IT technology stack, 78
Infrastructure-as-a-Service (IaaS), 79–81, 86–87, 143, 158–159
inlet temperature, 40, 177, 275, 430, 435, 590, 605, 621, 645, *646*, 647, 655, 663
innovation, 1, 5, 12, 29, 49, 59, 74, 81, 83, 84, 86, 145, 179, 247, 287, 403, 628–629, 637
inorganic chlorine compounds, 241
inrush current, 271, 273, 280, 283, 285–286, 442, 458, 473
insertion loss (IL), 302
  channel (CIL), 310, *312*
  connection IL, 313
  splice IL, 313
Institute of Electrical and Electronics Engineers (IEEE), 187, 309
  81 standards, 386–387
  493 standards, 576
  802.3 PMD, **300**, *301*, 311, 314
  1100, 386
  1366, 556
  insertion loss value, 312
  optical fiber distance, 205
  power efficiency, 343
  UPS, 492–493, 500
  VM allocation, 345
instruction set architecture (ISA), 144
insulated-gate bipolar transistor (IGBT), 485
intelligent orchestration, 150
inter-symbol interface (ISI), *310*
Intergovernmental Panel on Climate Change (IPCC), 1, 11
intermediate cross-connect (IC), 197, 199
intermediate distribution area (IDA), **197**, 198199, *201*, *202*, 208, *384*, *385*
intermediate distribution frame (IDF), 199, 317, 386 *see also* IDA
Internal Rate of Return (IRR), 94, 101, **102**, **103**, **106**, 117 *see also* financial analysis
International Building Code (IBC), 10, 521, 534–537, 675
International Electrotechnical Commission (IEC), 195, 267, 298 *see also* ISO/IEC and TIA/IEC
  IEC 61300, *319*
  IEC 61754, 294
International Engineering Task Force (IETF), 350
International Organization for Standardization (ISO), 195, 198, 298
  ISO 14644-1 Class 8, 252
  ISO/IEC 11801-1, 207
  ISO/IEC 14763-2, 206
  ISO/IEC 24764 Information Technology, 291, 295
International Society of Automation (ISA), **240**, 240, 253, *255*, *258*, 327
  class G1/G3/GX, 251–252, 257, **258**, 260
  ISA std 71, 240–241, **242**, 251, 257
Internet of Thing (IoT), 12, 77, 84–85, 137, 186, 323–324, 353, 641, 669–670
  anatomy, 3
  ecosystem, *3*
  taxonomy, 3
  users, 3
internet protocol (IP), 150
  IPv4 and IPv6, 3
introspection deep, 149

Japan data center council (JDCC), 8–9
Japan Green IT Promotion Council, 55
jumper cables, 291–292, 295, 298, 304–305, 307, 315–320

Kalman filter, 167
key performance indicator (KPI), 10, 148
keyboard, video, mouse (KVM) switches, 155–156, 158–159, 282, 284, 295, *639*
Kinetic Edge Alliance, 85
knowledge of transfer, 131, 136
Koomey, Jonathan, xvii, 2, 341

labeling, (enclosure, port, trunk), 309
latency, 368
Lawrence Berkeley National Laboratory (LBNL), 2, 56, 248, 657
Leadership in Energy Environmental Design (LEED), 8, 30, 55, 501
Legionella bacteria, 224, 404
Legionnaire's disease, 412
life cycle costs, 6, 183, 321 *see also* total cost of ownership
life safety, 386, 405, 521, 523–527, 530–531, 533–535, 546, 549, 675
Lightweight Directory Access Protocol (LDAP), 285
liquid cooling, 437
liquid immersion system, 46, 182
local area networks (LAN), 193
local exchange carrier (LEC), 375, 378
location plan, strategic, 8, 379
location sensitivity, 108
Lucent connector (LC), 297–298, 302, 304–305, 320

machine learning (ML), 4, 5, 83, 87, 145, 151, 331, 438, 608, 642–643, 670
macro med micro level, 7, 15–16, 20
magnetic field, 198
main cross connect (MC), 295 *see also* cabling system design
main distribution area (MDA), 198, *293*, 297 *see also* cabling standards
maintainability
  cabling methodology, 292
  cause effect, 5
  cost benefit, 32, 406, 414
  design, 436
  efficiency, 46
  failure analysis, 128, **139**
  maintainability analysis,551
  operations, 73, 418
  site system selection, 35
  strategies, 7
makeup air, 220
makeup air handler, 245
Markov approximation technique, 157, 567, *568*, 569, **570**
McFarlane, R. E., xi, xvii, 175–190, 403–440
mean down time (MDT), 552, *553*
mean time between failures (MTBF), 132, *133*, 502, 552, *553*
mean time to fail (MTTF),132, 552, *553*
mean time to repair (MTTR), 132, 502, 552, *553*
mean up time (MUT), 552, *553*
measurement and management technologies (MMT), 657–668
  air conditioning optimization, 659
  best practices, 661
  best practice quantification, 657
  CFD and MMT, 661 *see also* CFD
  chiller power optimization, 659
  cooling energy efficiency optimization, 659
  cooling infrastructure, 658
    energy ejection path, *658*
    energy transfer, 659
    energy transport, 659
    thermodynamic, *659*
  discharge temperature variation, *665*
  metrics, 662
    ACU utilization, 665
    DC cooling efficiency, 662
    horizontal hotspots, 663
    horizontal temperature uniformity, 663
    inlet hotspots, 663
    inlet temperature uniformity, 663
    thermo metrics measures, **666**
    vertical hotspots, 663
    vertical temperature uniformity, 663
  overview energy parameter values, **662**

measurement and management technologies (MMT) (*cont'd*)
power consumption in data center, 657
air conditioning units (ACU), 659
chiller plant, 659
electricity flow, *658*
energy efficiency control parameters, **658**
robotized scanner, *660*
temperature at different heights, *660*
thermal energy problems, 666
curtains, 666, *667*
increase airflow, 666
increase ceiling height, 666, 667
wireless MMT, *661*
mechanical design in data center, 403–440
above finished floor (AFF), 418, 422
adiabatic humidifier, 428
air delivery distribution, 415, 431
overhead, 432
overheat plenum, 432
through the space, 432
underfloor, 431
air- or water- side economization, 9, 23, 104, 211–212, *213*, 215, *216*, 221, 404, 425–428, 430–432, 436–438 *see also* economizer
airflow, 415
front to back, 415
front to rear, 415
side to side, 415
airflow management *see also* air management
balanced distribution, 433
containments, 434–435
efficiency, 426
fire protection, 419
fully mixed, 433
hybrid, 435
investigate, 415
lower temperature, 409
rack level, 435
architecture design, 411 *see also* rack floor plan
best practices, 436
central cooling plant, 41
air cooled, 41
direct economization, 42
direct expansion (DX), 41
evaporative cooling, 41
free cooling, 42
heat exchanger operation, 42
indirect economization, 42
water cooled, 41
water economization, ***42***
chiller plant, 41, 188, 425–427, 659
clearance interference, 418
cogeneration, 428
CFD, 432 *see also* Computational Fluid Dynamics
mechanical design in data center (*cont'd*)
CRAC CRAH, 419–420, 425, 427, 429–430, 432–433, 437, 439 *see also* CRAC and CRAH
controls, 416, 418–419, 421–422
direct digital control (DDC), 416, 418–419, 438
cooling load, 410
cooling medium, 429
air handler, 429
liquid, 429, 437
cooling system, 40
coordination A/MEP/IT, 404–405, 411–413, 415–423
cost estimation, 416
DCIM interface, 417, 438
dehumidification, 428
design changes, 423
design code, standards and guidelines
ASHRAE 90.1 90.4, 407, 410, 426, 432 *see also* ANSI/ASHRAE
ASHRAE std 127, 420, 430 *see also* ASHRAE
ASHRAE TC9.9, 406–407, 419, 437, 439 *see also* ASHRAE
code evaluation, 412–413, 416
NFPA 70, 419 *see also* NFPA
NFPA 75, 434
PUE, 406
standards guidelines, 407
design criteria, 403
aesthetics, 405
efficiency, 406
expansion, 408, 414
flexibility, 405
free cooling, 409
profitability, 406
reliability, 404 *see also* reliability engineering
design options, 429
design process, 407
commissioning, 424
construction administration (CA), 417, 423
construction documents (CD), 413, 417, 421
cost control, 416
design development (DD), 412–413, 417
final approval, 424
post construction support, 424
pre-design, 408–409
punch list, 424
schematic design (SD), 409, 413, 417
site inspections visit, 423
design review, 423
dew point temperature, 409, 419, 428
direct expansion (DX), 427–428
drawings and deliverables, 409, 412, 415, 417, 421
mechanical design in data center, drawings and deliverables (*cont'd*)
as-built, 424
bid package, 417, 422
build-out, 424
permit, 422
efficiency, 414, 438
emergency power off (EPO), 419
environment considerations
ASHRAE 90.1 90.4, 405 *see also* ANSI/ASHRAE
rejecting heat, 414
waste heat reuse, 405, 408
equipment layout, 413
equipment selection, 413–415, 417–418, 424–428
fan, 428
fire protection, 419
footprint evaluation, 411
heat rejection, 430
air side economizer, 430
cooling towers, 430
geothermal, 431
hot cold aisle, *434*, 436
humidification, 425, 427, 438
humidity control, 419, 425
interpretation, 423
layout, 413
preliminary, 416
liquid cooling, 429–430, 437
load estimate, 413
plenums
overhead, 415
raised floor, 415 *see also* raised non-raised floor
precision air conditioning (ASHRAE Std 127), 420 *see also* ANSI/ASHRAE
project types
convert existing building, 408
expansion, 408
green field, 408
remodeling, 408
relative humidity (RH), 409–410, 431, 438
reliability redundancy, 418, 404, 436 *see also* reliability engineering
request for information (RFI), 423
returned air temperature by models, *433*
safety, 404–405
security (bar, doors, curb), 404
size and locate MEP distribution, 415
space requirement evaluation, 409, 411, 437
specifications, 420, 422
system evaluation efficiency, 411
system selection, 411, 413
2-D plans, 418
3-D model, 418, 421
value engineering and estimation, 413, 416
water cooling, 408, 439

mechanical, ingress, climate, electromagnetic (MICE), 198
mechanical load component (MLC), 189–190, 617–619, 622–624
mechanical subfloor, 392, 432
megatrend, 1–2
memcached experiments, 151
memory resource requirements, 151
mesh network,86, 165
meter IT power, 122, 123
metrics sustainable data centers, 10
micro bends optical fiber, 208
micro service architecture, 150
microgrids, 359–365
  capital expenditure, 364
  definition and characteristics, 360
  design microgrids, 362
  engineering construction, 364
  financial sponsor, 364
  islanded mode, 361–362
  market, 359, 364–365
  operating modes, 361
  operation and maintenance, 363, 365
  overview, *360*
  value benefit, 361
  value chain, 364, *365*
Microsoft Azure, 80, 82, 86, 145, 346
middleware system based, 78–80, 150–151, 158, 256
minimum efficiency reporting value (MERV), *221*, 248, 252
mission critical design, 179, 188
Missouri University of Science and Technology, 178
mixed reality, 5
modular data center (MDC), 32, 232
modular design, 32–34, 46, 193, 207, 332, 334
modular *vs*. monolithic design, *33*
moral law, 6
motes, 164–165, 171
moves adds changes (MACs), 293–295, 297, 304, 306, 308, 311, 315–317
multi-access edge computing (MEC), 84
multi-floor computer room HVAC, 392
multimode optical fiber
  Ethernet, 205
  interface, *382*
  OM1~5, 204
  OM3 effective length, **321**
  SR4, *311*, *312*, *382*
Murphy's law, 5

nameplate data, 269, 288
National Aeronautics and Space Administration (NASA), 11, 158, 354
National Electrical Code (NEC), 187, 267, 270–272, 281, 284, 306, 501, 535
National Electrical Manufacturers Association (NEMA), 267, *269*, 272, 284, 287
  NEMA 5-15R, 284
  NEMA L21-30R, 284
National Fire Protection Association (NFPA), 70, 187, 208, 271, 419, 434, 534–538, 540–542, 544, 546
National Oceanic and Atmospheric Administration (NOAA), 11, 30, 529, 677
National Renewable Energy Laboratory (NREL), 34, 37
natural language processing, 145
negative air pressure, 9, 128, 243, 252, 647, *647*, *648*, 649, 655
net present value (NPV), 92, 94, 97, **100**, **101**, **102**, **103**, 104, **105**, **106**, 114
network architecture, 155, 163–164, 304, 382, *383*, 396, 400, 402
network automation, 162
Network Equipment Building System (NEBS), 254, 256, 261, 606, 651
network interface cards (NICs), 86, 326, 327
network management system (NMS), 164, 350, 355
network overhead, 391
network star, 165
network throughput, 381–382, 391
network topology, 10, 165, 381, 386
network traffic anomaly, 150, 278, 330
network underfloor, 392
Nimbus, 159
nitrogen oxides (NOX), 241
nodes, 164
non-governmental organizations (NGO), 96
non-structural restraints, 8, 182, 676, 684 *see also* structural design
NoSQL database (Cassandra), 151
Nova cell architecture, *161*
  Nova API, 161
  Nova conductor, 161
  Nova scheduler, 161
noise emission, 111, 137, 151, 181, 235, 236, 292, 331, 386, 405, 414, 432, 485, 572–573
noise (modal, mode, reflection, relative), 301, 309, *310*

Occupational Safety and Health Administration (OSHA), 405, 435, 543
ochestration engine, 150
off-grid, 361–362
online analytic processing (OLAP), 145
online transaction processing (OLTP), 145
Open Compute Project (OCP), 83, 87, 329, 332, 341, 386–387, 440, 464, *466*, 477
Open Edge Computing, 84
open source approach, 155
OpenFog, 84–85
OpenFog Consortium, 84, 87
OpenNebula, 158
OpenStack, 82, 155, 158–161, *162*, 346
  cascading, *162*
  multi-cell, *161*
operating expenditure (OpEx), 6, 183, **373**, 6, *7*, 74, 117, 183, 444, 477, 497, 632, 673
operating system, 78, 156, 342, 370, 518
operations optimization, 59, **60**
operations technology (OT), 9, 11
  cyber attacks, **351**
optical channel budget, *310*
optical return loss, 317
Organization for Economic Co-operation and Development (OECD), 1, 12
out-of-band management connections, 194
ozone (O_3), 241–243, 257, 260, 541–542, 545

parameterized model, **112**
paravirtualization, *156*
particulate contamination, 8, 12, 183, 239, 244–245, 252 *see also* contamination
  control, 320
  guidelines, 183, 249, **252**
  monitoring systems, *248*
particulate filtration, 12, 240, 242–*243*, 245, 248, 250–251, 431, 547
passive chimney IT cabinets, 45
patch panel and management, 199–200, *202*, 204, 209, *292*, 295, 297–298, 308–309, *311 see also* fiber cabling fundamentals
peak shaving, *497*, 498 *see also* UPS
performance configuration, 324
performance per watt, 27, *49*, 185, 324, 326, 329–332, *622*
permanent link, 204, 293, 299 (def.), 310, 312–316, 320
perverse incentive, 91n
Pfister framework, 147
photochemical oxidation, 241
photovoltaics (PV), 360, 363
physical media device (PMD), 300
platform dependent, 389
plugs and receptacles
  ISA, *268*
  NEMA, *269*
point of delivery (POD), 148
point to point cabling, 200, 291, **292**
policy based adjudication, 150

Polya, G. How to Solve It, xxi
port labeling, 309
portfolio manager EPS's, 49–50, 58
positive pressurization unit (PPU), *246*, *248*, 250
power demand density, 389
power density (w/sf or w/rack), 180–181, 183, 239, 389, 579, 609
power distribution unit (PDU), 29, 38, 47, 198, 337, 392–393, 399, 445, 592, 595, 620, 655, 658, 663 *see also* rack PDU
  Delta PDUs, 284
  wye PDUs, 284
power purchase agreement (PPA), 21, 346, 361–362
power supply units (PSUs), 293, 304, 324, 334, 340, 444–445
power usage effectiveness (PUE), 10, 16, 19, 176, 184–185, 189, 219, 220, 235, 286, 362, 607, 654, 657 *see also* benchmark metrics
  definition, 406
  measure report PUE, 57
PUE and IT loads in different redundancy, *39*
PUE and IT power savings, 111
PUE chiller temperature, *109*
PUE to estimate energy costs, 122
predictive analytics, 5, 145
predictive centered maintenance (PCM), 137
prescriptive analytics, 5, 145
present value (PV), 93, 98, 99 *see also* financial analysis
pressurize room for cleanliness, 243
preventive maintenance (PM), 137, 376, 426, 436, **478**, 515–517, 555, *558*, **561**, 576, 578, 682
printed circuit boards (PCBs), 165, 239, 533
processor socket power (PSP), 324
procure-to-pay (PTP), 148
profitability index, 100, **101, 103, 118**
project management (PM), 368, 611–616 *see also* mechanical design in data center
  closeout
    closeout tasks, 616
    lessons learned, 616
  costs, 615
    control, 615
    estimates and budgets, 615
    rough order of magnitude (ROM), 615, **615**
  kickoff, 611
    chain of command, 611
    document storage and sharing, 615
    prioritize scope/schedule/cost, 611, *612*
PM, kickoff (*cont'd*)
    project objective, 611–622
    project tools, 611
    roles responsibilities, 611
    scheduling tools, 615
    scope of work, 611–612
    virtual meeting, 615
  organization, 611–612
    responsible, accountable, consulted, and informed (RACI), 135, **612**
    team building, 612
    team organization, 611–612
  process monitoring control, 616
  program evaluation review technique (PERT), 613, *614*
  project failure, 369
  schedule, 613–615
    critical path and float, 613, *614*
    Gantt chart, 613, *614*
    list of tasks, 613
    PERT, 613
    schedule control and updating, 614–615
    task relationship, 613–614
Project Olympus, 83
pulse width modulation (PWM), 485, **486, 487, 488**, **495** *see also* UPS
psychrometric chart, 176, *177*, 178, 179, *214*, *217*

quality of life, 5, 7, *371*, **373**, 374–375, **378**
quality of service (QoS), 148
quad (4-channel) small form-factor pluggable (QSFP), 297
quantified ranking and selection, **378**

Rack Cooling Index (RCI), 605–606, 620, 621, 651
rack design, 386–387
  four post racks, 387
  network, 387
  OCP, 387 *see also* Open Compute Project
  rack unit (RU), 386, 387
  server mounting, 388
  storage mounting, 388
  two post racks, 386
rack floor plan, 381–402
  design
    computer room, 382
    fiber optic network, 381
    fire protection and sprinkler, 394
    grounding bonding point, 386
    IT design, 381
    lighting fixture, 394
    MEP design, 381
    multi-floor, 392
    network architecture, *383*
    network configuration, *384*
    raised non-raised floor, 394
rack-level cooling, 227–228, 435
  advantages disadvantages, 228–233
  enclosed, *230*
    cooling distribution unit (CDU), 230
  In-Row™, 230
  micro module, 232
  overhead, *229*
  rear door, 231
rack power distribution unit (rack PDU), 263–290, 638
  anatomy
    cord and retention, 274
    circuit breaker metering, 274
    circuit breakers, 273
    circuit protection, 273
    external connectors, 272
    graphic management, *275*
    interface security, 274
    outlet retention, *274*
    outlets, 272
    remote interface, 274
    single or 3-phase, 271
    xU form, 272
  approval agencies, 270
  branch circuits, 265, **267**
    load capacity, 265
    nameplate data, 269
    protections, 273
    rated current, 267
    rated voltage, 265, **266**
  configurations, 264–265
  efficiency PUE, 286
  environment management, 275, 277
    air pressure, 277
    airflow, 276
    contact close, 277
    humidity, 276
    temperature sensors, 275, *276*
    vibration sensors *see* vibration
  fundamentals, 264
  future trends, 287
  load capacity, 269
  management, 278
    analytics and reporting, 279
    DCIM, *279*, 280
    data collection, 278
    element, 280
    fault, 280
    graphic user interface, 279
    measure and report, 279
    open point integration, 280
    remote switching, 280
    security, 280
    topology, *278*
  planning selection, 280
  power distribution, 281
    dual feed to racks, 282
    inrush power, 283
    load balancing, 283

rack power distribution unit (rack PDU), power distribution (*cont'd*)
multiple power supply, 282
redundancy, 282
single and dual feed, *281*
power rating, 269
power requirement, 283
dual feed to racks, 282
208 V Single- and 3-Phase, 284
400 V 3-Phase, 284
safety and electromagnetic, **270**
approval, **270**
grounding, 178, *195*, 270 *see also* grounding
overload protection, 270
selection, 285
application requirements, 285
functionality, 285
system communication protocol, 277
system management, 277
system topology, 277
three-phase wiring, 266
type, *264*, 265
intelligent, 265, 286
non-intelligent, 265, 285
switched, 265
switched with metering, 265
rack unit (RU, rack mount equipment size), 255n, 332, 380, 386 *see also* rack PDU
radio-frequency emissions, 198
raised non-raised floor, 244, *247*, 394
cooling, *228*
design, 415
random access method of accounting and control (RAMAC), 380
reach (maximum) per bandwidth value, **300**, 301
reactive environments classification, **240, 257**
reactive monitoring, 240, *261*
reactivity monitoring, *240*, 241, 248, 252, 256–258, *258*, 260–261
acceptance criteria, **257**
rear door heat exchangers (RDhX), 182
recirculating air filtration unit (RAU), *247*, 260
recirculating air handlers (RAHs), 246, 253
recirculating room air, 243 *see also* air management
rectifier, 471, 474–476, 488–490, 492, 494–497, *503*, 505
recycle energy, 331
redundancy (N+1, N+2, 2N), 8–9, 39, 46, 73, 130–131, 134, 163, 188–189, 193, 201, *221*, 282, 349, 404, 418, 436, 452–456, 502–505, 556–557, 574, 603 *see also* UPS and Electrical Design chapters
refractive index, 300, *301*, 302
relative humidity, 24, 34, *44*, **57**, 163–169, 173–175, *177*, 409, 427, 431, 601, *661*
reliability and availability, 6, 48, 189, 281, 443, 470, 476, 477, 551, 553 *see also* reliability engineering
reliability and maintainability, 46, 418
reliability and redundancy, 8, 41, 404, 502
reliability block diagrams, 132
reliability-centered maintenance (RCM), 137
reliability engineering, 551–578
dependability, 551, *552*
active redundancy, 556, *557 see also* redundancy
assessments, 569, *571–572*, *575*
analysis, 558
availability, 551–552
data collection and assessment, 560, **561**, **562**
management, 576, *577*
methodology, *559*
modeling, *556*
partial redundancy, *557 see also* redundancy
passive redundancy, *557*
redundancy, 574
safety targets, 572
design, 573–578
data center infrastructure, *573*
functional analysis, *573*
IEC 62380, 576
IEEE, 1366, 556
MIL-HDBK 217F, 576
process, 576–577
tier classification, **578**
failure, 563–576
combination analysis, 567
detection index table**, 564**, 576
discrete-time Markov chain, *568*, 569
dysfunctional analysis, 558, **563**, 569, **570**
environment, 557
event tree, *568*
failure behavior, *560*, *567*
Failure Mode and Effects Analysis (FMEA), 551, 563–568, 576
FMEA assessment principle, *565*
FMEA IEC 60812, 563, **564**
FMECA tables, 565, **566**
failure rate, 576
fault tree, 551, *564*, *567*
frequency index table**, 564**
gravity index table**, 564**
hidden, *558*
human error, 557
model, *558*
risk acceptance matrix, **565**
time-sequential stochastic simulation, *569*
reliability engineering (*cont'd*)
maintenance, 521–558
external functions, *552*
maintainability, 551–552
preventive, 555, *558*
preventive replacement, *558*
operations
degraded mode, 574
loss of IT equipment, *574*
operating mode, 574
reliability, 551, 552
accelerated life testing, 551
block diagram, *567*
common mode failure, 576
curative maintenance data, 554, *555*
failure mode, **554** (def.)
failure rate, 552, *553*, *556*
field experience statistics, 551
mean down time (MDT), 552, *553 see also* MDT
mean time between failure (MTBF), 552, *553 see also* MTBF
mean time to failure (MTTF), 552, *553*, **554**, 567–568 *see also* MTTF
mean time to repair (MTTR), 552, *553*, 563, 567–568 *see also* MTTR
mean up time (MUT), 552, *553 see also* MUT
maintenance time distribution, *555*
reliability data, 576
system average interruption frequency index (SAIDA), 556
testing, 551
risk analysis
assessment principle, *565*
examples, 562
hazards risk analysis, 551
risk acceptance, *559*, 565
undesired event (UE), *559*, 567–569, 572, 574
system reliability and availability
integrated logic support, 551
maintainability analysis, 551
reliability prediction, 551
statistical analysis, 551
reliability ICT equipment, 178 *see also* reliability engineering
remote direct memory access (RDMA), 147
remote power panels (RPP), 393, 476
request for proposal (RFP), 75, 146, 148
request for quote (RFQ), 146, 148, 334
resilience assurance, *149*
resiliency of service measurement, 149
resource abstraction mask vulnerability, 149
resource management, 155–163
architecture, 160
cascading mechanism, 160
multi-cell, 160, 161
multi-region, 160

resource management (*cont'd*)
cloud platform, 157
intelligence, 156
resource virtualization, 155, *156*, 157
scheduling, 157
single or multi cloud, *159*, *160*
virtual, 156
responsible, accountable, consulted, informed (RACI), 135, **612**
Restriction of Hazardous Substances Directive (RoHS), 179, 183, 190, 239–240, 248, 501
return air temperature, *433*
return on investment (ROI), 92 (def.), **93**, 94, 104 *see also* financial analysis
return temperature index (RTI), 605, 621
Revit (Autodeck) 3-D building design, 405
ring of fire, 524, 676, *677*
risk-adjusted cost performance, 149
man-made risks, 377
risk management operating data center, 127–141
cost of failure, 129
human error analysis, 139
human learning, 129, 139
human unawareness, 128
Kolb cycle, *129*, *130*
learning curve, 127, *128*, 139
lessons learned, 130
risk reduction plan, 138
site selection, 130 *see also* site selection
TESEO (empirical technique to estimate operator failure), 139
robotic surgery, 6
room level cooling, 228
root cause analysis (RCA), 5, 140
Royal Institute of British Architects (RIBA), 131, 136, 138

scavenger air (ScA) fan, 42, *216*, 217–220, 223–224
Schmidt, Roger, xv, 175, 176, 181
seismic restraints, 8, 182, 676, 684 *see also* structural design
Senate Bill 327 (information privacy), 352–353
sensing temperature, humidity, airflow, 637
sensitivity analysis, 8, 104, *109*
sensor communication, 171
sensor failure, 171
sensor measurements, 171
sensor network schematics, *167*
sensor operating events, *172*
sensor parameters, 170
sensors analytics, 166
sensors actuators, 165
sensors network, 165, *166 see also* temperature sensor, corrosion sensor
serial AT attachment (SATA), 151, 326, 328
server design energy efficiency, 53, 323–337
concepts of design, 329
design process, *331*
design verification test (DVT), *331*, 335
engineering verification test (EVT), *331*, 335
mass production (MP), *331*, 335
production verification test (PVT), *331*, 335
energy-efficient utilization indicator (EEUI), 338
enterprise & datacenter SSD Form Factor (EDSFF), 329
IEC 60297 structure, 332
mechanical design, 332 *see also* mechanical design in data center
product life, 332
server anatomy, 325
application-specific integrated circuits (ASIC), 144, 324, 326, 329–330
computer processing unit (CPU), 326
FPGA accelerate, 82–83, 144, 151, 330
graphics processing unit (GPU), 144, 329, 582
high-speed fiber optic network card, 330
motherboard, 326, *333*
network interface card (NIC), 326–327, 332
power supply unit (PSU), 326–327
server efficiency rating tool (SERT), 325, 334, *335*
server energy consumption modeling, *340*
Intelligent Platform Management Interface (IPMI), 339, 340
performance monitor counter (PMC) based, 339
resource-utilization-based, 339
roaming agreement preference list (RAPL), 339, 340
workflow, *339*
server power efficiency, 338
server room, 254n
server types
ASIC server, 330
expandable server, 330
GPU server, 329
network server, 330
storage server, 329
storage media, 326
hard disk drive (HDD), **328**
solid state drive (SSD), **328**
system design, 332
thermal design, 333
thermal design power (TDP), 329, 333–334
server design energy efficiency (*cont'd*)
voltage regulator module (VRM), 329, 331–332
server energy saving technologies, 341
adaptive voltage and frequency scaling(AVFS), 342
dynamic voltage frequency scaling (DVFS), 342
load scheduling, 346
power efficiency at CPU utilization, *327*, 343
proactive load control, 342
renewable energy, 346, *347*
scale in and out, 342
server management, 341
work mode switching, 342
server level cooling, 227
advantages disadvantages, 234–236
cold plate, 234
immersion, 234
server power models, *340*
server room, 254n
service level agreement (SLA), 91, 130, 141, 149, 151, 401
7x24 Exchange, 57
shield un-shield power, 391
CENELEC EN 50174, 391
IEC 14763–2, 391
TIA 690, 391
shock and vibration testing, 182
short list, 367, 377
side access system (SAS), *246*, 250, 260
Silicon Valley Leadership Group (SVLG), 57
Simple Network Management Protocol (SNMP), 272, 275, 277, 285, 350–355, 633, 638, 643
Singapore Green Data Centre Technology Roadmap, 58
single mode fiber
OS1, 204
OS2, 204
single point of failure (SPOF) analysis, 132, 133
site selection, 130, 367–380
brown field site, 367
business environment, 374
business requirements, 370, 372
considerations, **372**
criteria, 374
infrastructure and services, 375
macro, mid, micro levels, 368
natural disaster, 377
operating expenses, **373**
political environment, 374
process, 369–370, *371*
quality of life, 374
quantify selections, **378**
secrecy, 369
team, 368, **372**

6S pillars, 10
smart rack, 288
social computing, 145
software-defined compute (SDC), *144*
software-defined data center (SDDC), *144*, 330
software-defined environment (SDE), 143–153
  architecture, *144*
software-defined infrastructure (SDI), 148
software-defined network (SDN), 143
Software-defined storage (SDS), 143
solar panel *see* photovoltaics
SolarWinds cyberattack, 355
space design, 389 *see also* rack floor design
space planning, 400, 634
speech recognition, 145
spine-leaf model, 150
splice insertion loss (dB), 313
standard operating procedures (SOP), 137, 335
Standard Performance Evaluation Corporation (SPEC), 48, 50, 54
static transfer switch (STS), 281
storage area network (SAN), 148, 193, 199, 208–209, 287, 292, 294–297, *296*, *299*, 382–385
storage devices energy efficient, 22
storage drive array, *385*
strategic conserve energy, 15
strategic location plan, 7
strategic planning, 6
  forces, 6
strong oxidants, 241
structured cabling system, 193, 316
  cabling redundancy, 201
  topology, 200, *201*
structured standards-based cabling, *194*
structural design, 521–532 *see also* Fire Protection
  Building Operation Resumption Program (BORP), 531
  building post earthquake performance state, 525, *526*
  design, 521–529
    ASCE/SEI 7, 521, 525, 527, 529
    ASTM E1966, 528
    anchor brace nonstructural components, 522, 525, 527, 531
    code-based design, 523
    considerations, 525
    earthquake ground motion, 525
    FEMA 750, 525
    FEMA P-58, 523
    new existing building design considerations, 524, 525
    performance-based design, 523
    risk categories, 523–529
  hazard effects on building, 525, 527, 528
structural design (*cont'd*)
  life safety standards, 521, 523–*526*, 527, 530–531, 533
  mean recurrence interval, 527
  mitigation strategies, 527, 528–530
    ASCI 41, 527
    FEMA, 531
    structural performance levels, 527
  natural disaster hazards, 522, 524
    earthquake effects, 522, 524
    flooding, 522
    rain effects, 522
    snow effects, 522
    tsunami effects, 522
    wind effects, 522
  post-disaster planning, 531
  pre-disaster planning, 530
  resiliency strategies, 530
  seismic bracing anchoring, 527
Structural Engineers Association of Northern California, 531
structural loads, 179
subject matter experts (SMEs), 331, 369
sulfur compound, active, 241
sulfur dioxide, 183, 241
sulfur oxide (SO_2, SO_3), 241
sulfuric acid (H_2SO_4), 241
sulfurous acid (H_2SO_3), 241
supervisory control and data acquisition (SCADA), 135, 514, 554, 574
supply air temperature and energy use, 51
sustainability, 1, 11, 21, 58, 501
switch fabric, *210*
synchronous digital hierarchy (SDH), 375
synchronous optical network (SONET), 375
system of engagement, 145
system of insight, 145 *see also* OLAP
system of record, 145

task scheduling, *344*
taxes, 91, 93, 95–96, 105, 122, 124, 131, 145, 368, *371*, **372**, **373**, 375, 379
  abatement, 374, 376
  green, 96
  incentive, 7, 369
  law, 94
  ROI and TCO, 95
technical resource manual (TRM), 60, 61
technological changes, 1
telecommunications cabling standards, 193, 208, 298
  ANSI/TIA, *195*
  ANSI/TIA 568, 10, 195, 198, 203, 207, 294, 296, 306, 313
    $M_1I_1C_1E_1$ classification, 198
  ANSI/TIA 569, 195
  ANSI/TIA 606, 195
  ANSI/TIA 607, 195
  ANSI/TIA 758, 195
telecommunications cabling std (*cont'd*)
  ANSI/TIA 862, 195
  ANSI/TIA 942, 10, 535
    2015 ASHRAE guidelines, 206, 207
    cable length, 203
    cabling standard, 294
    design and instllation, 195–196
    fiber connectivity system, 292
    HDA MDA, 296
    infrastructure rating, 209
    rate 1/2/3/4, 254, 535
    structure cabling, 291
    switch fibric, 209, *210*
    telecommunications infrastructure standards, 8
  ANSI/TIA 5017, 195
  BICSI, 196
  cabling topology, 200
  CENELEC, 195, *196*, 198, 207
  CSA, 195
  EIA/TIA 455, 291, 301
  organization, *195*, *196*
  structured cabling, 193, *201*
  terminology cross reference, **197**
Telecommunications, Electrical, Architectural, Mechanical (TEAM), 209
  $T_NE_NA_NM_N$, 209
telecommunications entrance room (TER) design, 198
Telecommunications Industry Association (TIA), 194, 253
  EDA, 200
  ENI, 200
  HDA, 199
  IDA, **199**
  IEC, 195, 196, 294, 298, 332
  ISO/IEC, 195, *196*, 298
  LAN, 199, 208
  MDA, 198–199
  SAN, 199, 208
  TIA TSB 153 ESD, 207
  TIA TSB 155, 204
  WAN, 208
  ZDA, *199*
TIA redundancy, 188, 201
telecommunications room (TR), 200
telecommunications space requirement, 196
Telecordia GR-1275-CORE, 306
Telecordia GR-3160, NEBS Requirements, 254
temperature binned hours, 107
tensor processing unit (TPU), 83, 144
thermal envelope, 176 *see also* ASHRAE envelope standard
thermal guidelines, 175, 177 *see also* ASHRAE envelope standard
thermal management, 182
thermodynamic properties, **235**
test adjust balance (TAB), 253

transceiver options, 300
top-of-the-rack (TOR) switches, 148, 150
topology, 146, 149, 196, 296, 441, 444, 450, 459
  ac-dc conversion, 363, 444
  cabling, 200, 386
  comparison, **451**
  distribution, 619
  hierarchical, 295
  linear, 292
  mesh network, 165, 443–444
  modularity, 47
  N, 445, *447*
  N+1, *39*, 445–446, *448*, 449
  N+2, 32
  open looped, 444
  rack PDU management, 278
  network, 10, 381, 386, 396, 402
  reliability, 37, 46, 130–131, 254, *442*, 546, 639
  ring, 292, 390
  selection, 465
  star, 200, *201*
  system connectivity, 277
  2N, *39*, 445
total cost of ownership (TOC), 6, 89, **92**, 94–95, 97, 121, 137, 180, 462, 477, 501, 513, 515
  definition, 92
Total Harmonic Distortion current (THDi), 442, 488, 495
trap air and p-trap, 220
transactive energy, 364
trunk engineered limits, 311, *312*
trunk labeling, 309
2-factor authentication, *357*
Typical Meteorological Year (TMY), 107, 108, 216, *222*, *223*

U-value, 593
Ubuntu-based software, 156
under-floor air filtration, 247
Underwriters Laboratories (UL), 267, 271, 298, 501, 537
uninterruptible power supply (UPS), 483–520 *see also* electrical design in DC
  absorbed glass mat (AGM), 491
  alternative energy source, 513
    fuel cells, 515
  backup power, 359
  battery
    flooded electrolyte, 501
    lead acid, 491
    lithium-ion, 515
    nickel–sodium, 515
    ultra capacitors, 514
    valve-regulated lead acid (VRLA), 491
  best practices, 519–520
uninterruptible power supply (UPS) (*cont'd*)
  catcher UPS arrangement, 511, *512*, *514*
  centralized bypass system, *506*
  components and subsystem, 484
    converter, 489, 494
    I-T-PWM inverter designs, *486*, *487*
    IGBT based PWM inverter, *486*
    inverter, 485, 492 (def.)
    inverter PWM switching, *495*
    logic control, 489
    momentary static switch, 489
    multi lever conversion improves efficient, 486
    pulse width modulation (PWM), *485*
    PWM switching pattern, *485*
    rectifier, 486, 488, *489*, 494
    silicon-controlled rectifiers (SCR), 489
    static bypass switch, 489, *494*
    T-type PWM switching, *486*, *487*
  conduction and switching loss, 485, *488*
  conversion, 483 (def.), *493*, 494
    double conversion, 54, *494*, *495*
    single conversion, 492
  costs, 485–486
    CapEx OpEx, 497
    dual-bus architecture, 502
    electrical costs, 498
    peak shaving, *497*
    ROI, 498, 501 *see also* financial analysis
    total cost of ownership, 501–502
  distributed bypass parallel system, *506*
  distributed energy resources (DER), 365, 497
  efficiency, 23, 46, *488*, 499, *500*
  electromagnetic induction (EMI), 7, 10, 36, 198, 203, **270**, 465, 486, 558
  energy aware UPS system, 59, 185, *344*, *347*, 497, *498*
  energy use, 618
  environment health safety, 501
  feed(s) to IT equipment, *485*
  flywheel, 514
  inrush current, 442, 458, 473
  load performance, *47*
  maintenance, 515
    bypass, 489
    bypass breaker schematic, *490*, *491*
    hot tie cabinet (HTC) maintenance, 509, *511*
    independent service provider, 516
    internal switch, 489
    load bank breaker, 490
    manufacturer's service, 516
    multimode parallel system, 489
    self-maintenance, 517
    serviceability, 501
    solenoid key release unit (SKRU), 490
UPS, maintenance (*cont'd*)
    switch bypass, 590
    time and materials, 517
    wrap around switch, 489
  management and control, 518
    individual server level, *518*
  multimode multi-paths high eff. UPS, *496*
  power flow diagram, *492*, *506*
  principal and applications, 484
  reliability and redundancy, 502, *503*, *504*, *505 see also* reliability and redundancy
    availability, 502, 503
    dual corded IT equipment, *509*
    mean time between failures (MTBF), 502 *see also* MTBF
    mean time to repair (MTTR), 502 *see also* MTTR
    parallel redundancy, 503, *505*
    power delivery path strategies, 502, *503*
    power synchronization control (PSC) unit, *509*, *510*
    subsystem redundancy, 503, *505*
    system bypass module (SBM), *510*
    2N architecture, 189, 507, *508*
  response time, 489, 498, *499*, *500*
    IEC guidelines, 499, *500*
    ITIC/CBEMA, 499
  rotary system, 484, 489, 492
    generator feed to flywheel, *493*
    total harmonic distortion (vTHD, iTHD), 488
  selection considerations 498, 516
  serviceability
    dead front, 501
    field replaceable units (FRUs), 501
  SNMP, 354
  sustainability and safety, 501 *see also* sustainability
    battery hazards, 491, 501, 537
    IEC62040, 501
    NEC, 501
    RoHS, 501
    UL, 501
  technology types
    line-interactive (LI), 492–494, 499
    rotary UPS systems, 492, *493*, 494
    single conversion, 492
    standby, 493
    static UPS system, 491, *492*
  three-phase wye transformer, *484*
  topologies *see also* topology
    double conversion, *493*, 494
    single conversion, 489, 492, *493*, 494
    transformer based double conversion, 494
    transformer-less double conversion, *495*

unshielded twisted-pair (UTP), 203, 388, 392
upgrade projects, 29, 101, 112, *113*, 114, 115
Uptime Institute
  DCIM, 630
  data center tiers, 9, 188, 407, 535
  costs, 380
  energy efficiency, 657
  managing risk, 127
  measure report efficiency, 57
  redundancy, 73, 505
urbanization, 1, 360, 454
U.S. Environmental Protection Agency (EPA)
  data center definition, 2
  Energy Star, 323
  energy use 2007 report, 28, 49, 409, 617
  global warming potential HFC, 542
U.S. EPA 2007 Report to Congress on Server and Data Center Energy Efficiency, 28, 49, 96, 409, 617
U.S. Green Building Council (USGBC), 8, 55, 57
U.S. LBNL 2016 Data Center Energy Usage Report, 2
U.S. Library of Congress, 4
user interface (UI), 79, 274, *275*, 519, 637, 639
utility grid, 67, 361–365, 444–445, 461

value engineering, 413, 416
variable frequency drive (VFD), 95, 168, 425, 465, 660, 666
vibration and monitoring, 36, 182, 198, 236, 275–276, *277*, 335, 454
virtual assistant, 331
virtual machine, 5, 22, 84, 123, 143, 148, 150, 156, 288, 330, 340, *345*, 518
  allocation and consolidation, *345*
virtual machine deployment, 157
virtual machine monitor (VMM), 156 *see also* hypervisor
virtualization, 18, 22, 50, **51**, *78*, 79, *80*, 82, 86, 111, 119, 122, 143–144, 149, 155, *156*, 263, 271, 279, 286, 330, 340, 345, 399, *518*, 627, 631–633, 637, 643
virtualization obfuscates monitoring, 149
virtuous cycle of data analytics process, *4*
volatile organic compounds (VOCs), 260
vulnerabilities analysis, **138**

warehouse scale computer (WSC), 83
water evaporation, 227
water quality conditioning, 182
water side economizer, 9, 23, 104, 221, 426–427, 430, 438 *see also* economizer
water use, 34
water usage effectiveness (WUE), 55, 620, 621, 624, 631
wavelength division multiplexer (WDM), 300
web-based collaboration, 372, 615
wet-bulb temperature, 177, **180**, *212* *see also* dew point temperature
Wheatstone bridge circuit, 165
wide area network (WAN), 193, 200, 208, 285
wide bandgap device, *486*
wireless sensor networks, 163, *164*, 170, 174, 657, 661
  accuracy, 165
  deployment, 165, *166*
workforce development, 12
workload-optimized systems, 144
wye transformer three-phase, *484*

X-as-a-Service, 6

zone distribution area (ZDA)*see also* fiber cabling fundamentals
  architecture, *383*
  cable consolidation points, *199*, 295, *298*
  fiber channel illustrations, *384*, *385*
  standards cross reference, **197**
  structural cabling, 200–201

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.