Chinchilla Arley, Ricardo.
Mercado de datos: conceptos y metodologías de desarrollo
Tecnología en Marcha, *Vol. 24, N.° 3, Julio-Setiembre 2011, P. 55-66*

Mercado de datos: conceptos y metodologías de desarrollo

Fecha de recepción: 20/09/2010
Fecha de aceptación: 05/10/2010

Ricardo Chinchilla Arley[1]

Palabras clave

Bases de datos, mercados de datos, depósitos de datos, procesamiento analítico en línea.

Resumen

En este artículo se ofrece una descripción sobre la importancia del depósito y del mercado de datos. Además, identifican y formulan las principales definiciones, esquemas y términos básicos relacionados con el mercado de datos. Igualmente, se establecen y desarrollan los modelos de implementación y las metodologías para el desarrollo de un mercado de datos. Por último, se consideran las formas tecnológicas de procesamiento analítico en línea (OLAP) y su utilidad en el análisis del mercado de datos.

Key words

Database, datamart, datawharehouse, online analytical processing.

Abstract

In this article a description offers on the importance of the data warehouse and the data mart. There are identified and formulate the principal definitions, schemes and basic terms related to the data mart. There establish and develop the models of implementation and the methodologies for the development of a data mart. Finally, they are considered to be the technological forms of analytical processing in line (OLAP) and this usefulness in the analysis of the data mart.

Introducción

El mercado de datos constituye una tecnología de bases de datos que ha tomado gran auge debido al crecimiento y muchas veces subutilización de los grandes bancos de datos históricos almacenados en las organizaciones. Es por ello aquí que se presentan algunas precisiones sobre los conceptos y las metodologías utilizadas en el desarrollo del mercado de datos. En la primera parte, se establece la importancia del depósito y del mercado de datos. En este sentido, se deja claro que el mercado de datos (MD) es un elemento sustantivo dentro de la ultima generación de bases de datos, constituida por el repunte de los depósitos de datos (DD). Según Poe (1998),

1. Máster en Computación, Instituto Tecnológico de Costa Rica. Profesor en la Escuela de Bibliotecología y Ciencias de la Información, Universidad de Costa Rica. Correo electrónico: *ricardo.chinchilla@ucr.ac.cr*

DD constituye, *"una base de datos de solo lectura, donde la información extraída de los sistemas operacionales corrientes de la empresa es transformada, integrada y resumida para luego ser usada con efectividad en el soporte de decisiones"*.

En la segunda parte, se desarrollan los principales conceptos, esquemas y términos básicos relacionados con el MD. La exposición se centra, particularmente, en que el modelo del MD está definido por la forma en que el usuario necesita ver la información y cómo quiere que se le presente; en este sentido, el MD posee las mismas características de un DD pero a un nivel más específico, ya que contiene diferentes combinaciones y selecciones de los datos que se encuentran en el DD y ofrece una mayor personalización de los datos del departamento, permitiendo un manejo más eficiente de la información histórica, ejecución de procesamiento independiente del resto de los departamentos y un costo de almacenamiento y procesamiento inferior.

En la tercera parte, se presentan varios modelos de implementación y metodologías para el desarrollo de un DD y un MD, y se aclara que la utilización de alguno de ellos dependerá de los alcances y limitaciones del proyecto que se desee llevar a cabo. En cuanto a los modelos de implementación, existen tres básicos: el de estrella, el de constelación de hechos y el esquema "copo de nieve". El más utilizado es el esquema estrella, que consiste en una tabla central construida con llaves foráneas provenientes de un conjunto de tablas llamadas dimensiones.

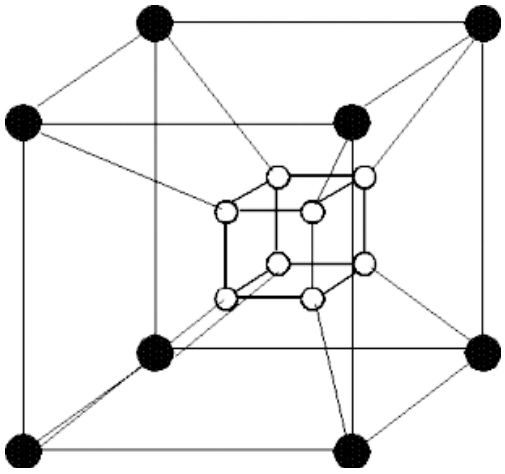

Figura 1. Representación de un hipercubo
Fuente: Elaboración propia a partir de "Construindo um hipercubo" (2004).

En la cuarta parte, se consideran algunas formas de OLAP (Procesamiento Analítico en Línea - Online Analytical Processing) y su utilidad en el análisis del MD; en este sentido, las herramientas OLAP permiten el análisis de datos en forma interactiva.

Importancia del depósito y del mercado de datos

Un MD es un elemento sustantivo dentro de la última generación de bases de datos, constituida por el repunte de los DD. Un MD proporciona un esquema de almacenamiento que permite realizar consultas complejas, por medio de la interacción e interrelación de múltiples entidades, manteniendo solo un repositorio con la información agregada y ordenada, según las necesidades de los usuarios. Además, provee de una interfase de consulta que permite al usuario la posibilidad de análisis de la información para la toma de decisiones.

Los DD también son conocidos como hipercubos multidimensionales o cubos n-dimensionales, debido a que presentan la información en más de tres dimensiones (>3Dim). Se puede esquematizar de la siguiente manera:

González indica que "*el hipercubo funciona como una representación intuitiva de un evento [generador de datos de más de tres dimensiones] debido a que todas las dimensiones coexisten para cada punto de este y son independientes entre sí" (González,* 1998:8).

La combinación de n-dimensiones y varios niveles o jerarquías por dimensión constituyen la esencia del hipercubo.

Poe define un DD como *"una base de datos de solo lectura donde la información extraída de los sistemas operacionales corrientes de la empresa es transformada, integrada y resumida para luego ser usada con efectividad en el soporte de decisiones"* (Poe, 1998:20). La implementación de este tipo de tecnología implicará una transformación profunda en la forma de manipular los datos. De hecho, el DD por sí mismo traerá consigo el consumo de gran cantidad de recursos y requerirá de un largo tiempo de implementación.

Sin embargo, su utilización es necesaria debido a la necesidad de mejorar la administración general de los datos y de tener una visión más amplia de la información, al disponer de un acceso a esta sin importar su ubicación. Con el DD se pretende solucionar problemas derivados de la dispersión de los datos en las diferentes aplicaciones transaccionales, muchas veces heterogéneas, y proporcionar información para la toma de decisiones.

Según Grill (1996), la idea del DD surge inmediatamente después del nacimiento del usuario como tal (1970-75). Ahora este debe solucionar problemas que antes estaban del lado de los programadores y analistas, al usar soluciones tecnológicas particulares. En la década de los 80 casi todo lo relacionado con datos ha sido automatizado y se hace prioritario el apoyo a la toma de decisiones con base en dichos datos almacenados.

La primera técnica de toma de decisiones se dio con la automatización de reportes. Ya para los 90 se ha dado una gran evolución en cuanto a aplicaciones de usuario final y surge una gran cantidad de sistemas de información que, a diferencia de los sistemas operacionales enfocados en la producción, se especializan en el apoyo a la toma de decisiones, manipulando datos históricos y puntuales optimizados para la consulta.

No obstante, se apoyan en bases de datos transaccionales, las cuales trabajan con datos en tiempo real. Además, muchas empresas tienen diferentes plataformas de bases de datos, por lo que una búsqueda inteligente de información puede tomar gran cantidad de tiempo, degradando la efectividad de la decisión tomada con base en ella. Es aquí donde se hace presente el DD, que ha tenido un desarrollo muy discreto, a la sombra de los sistemas de información, pero actualmente ha tenido un repunte debido principalmente a los avances tecnológicos.

Con la incorporación de los DD, las empresas han tratado de minimizar la herencia negativa de los sistemas transaccionales con que cuentan, aumentar el rendimiento, lograr administrar los datos en el tiempo y brindar un apoyo efectivo a la toma de decisiones, con las ventajas competitivas que esto trae consigo.

Mercado de datos: definiciones y términos básicos

Un DD contiene información detallada y datos integrados e históricos comunes a toda la organización. Cada departamento de una empresa utilizará el DD para sus análisis internos y es probable que una gran cantidad de datos no sea de su interés, ya que solamente querrá trabajar con un rango definido. Por otro lado, en la medida en que se incrementa el volumen de datos, se aumentará el costo de procesamiento.

Corey define un MD o datamart como "bases de datos multidimensionales orientadas a una materia específica" (Corey, 1997:33). González lo define como "una extensión natural" del DD (González, 1998:24), siendo una base de datos con información de interés para un determinado sector de la organización, contribuyendo con los sistemas de apoyo a la toma de decisiones

de los departamentos con una arquitectura fundamentada en la del DD.

Por su parte, Wolff define el mercado de datos como

> *"un conjunto de hechos y datos organizados para soporte decisional basados en la necesidad de un área o departamento específico. Los datos son orientados a satisfacer las necesidades particulares de un departamento dado teniendo solo sentido para el personal de ese departamento y sus datos no tienen por qué tener las mismas fuentes que los de otro datamart" (Wolff, 2000).*

El modelo del MD está definido por la forma en que el usuario necesita ver la información y cómo quiere que se le presente. Posee las mismas características de un DD pero a un nivel más específico, pues contiene diferentes combinaciones y selecciones de los datos que se encuentran en el DD y ofrece una mayor personalización de los datos del departamento, permitiendo un manejo más eficiente de la información histórica, ejecución de procesamiento independiente del resto de los departamentos y un costo de almacenamiento y procesamiento inferior. La figura 2 esquematiza al MD dentro del contexto del DD.

Un conjunto de mercado de datos podría constituirse en un futuro DD, según la posición de Wolff (2000), quien expone que la construcción de mercados de datos constituye la construcción incremental de un DD total. Dicha construcción es una aproximación práctica para construir un DD a escala empresarial de forma evolutiva. Con base en esta posición, se podría desarrollar un DD como lo muestra la figura 3.

Partiendo de esta base, un MD es una excelente plataforma sobre la cual se podrá desarrollar un futuro DD. Cada uno de los incrementos o MD generados en la organización se convertirían en piezas del hipercubo general.

La tecnología de DD presenta una forma de manejo de datos que difiere del modelo relacional tradicional desde las restricciones hasta la propia terminología. Por tal motivo, se incluyen algunos términos básicos para clarificación de conceptos.

Datos multidimensionales

Según Weldon (1995), cualquier dato en teoría puede ser multidimensional. El término se refiere a datos (objetos o eventos) que se pueden describir o clasificar por dos o más atributos.

Bases de datos multidimensionales

Son repositorios grandes de datos, que son descritos de acuerdo con distintas perspectivas o dimensiones. Según Cox (2000), deben cumplir con cuatro características básicas: rapidez, capacidad analítica, soporte multiusuario y multidimensionalidad.

Análisis multidimensional

Podría definirse como la clasificación de la información de acuerdo con categorías. Gill (1997) lo define como el análisis simultáneo de múltiples dimensiones de los datos.

Dimensiones

Son las variables sobre las que se califican los datos, pueden ser descriptivas o de hechos. Las descriptivas proporcionan una explicación de los datos que serán

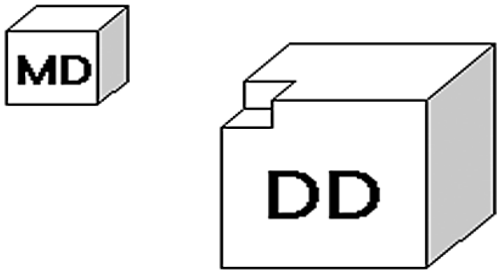

Figura 2. Mercado de datos.
Fuente: Elaboración propia.

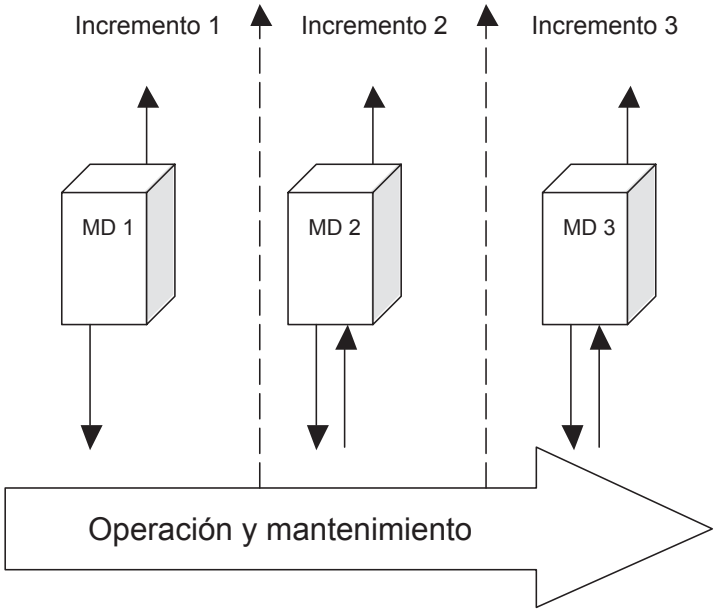

Figura 3. Esquema de desarrollo incremental de un DD.
Fuente: Wolf, (2000).

analizados. En una librería, las dimensiones descriptivas detallarían un producto, como por ejemplo cuadernos, clasificándolos en tamaño, cantidad de hojas, colores, pasta dura o suave... La dimensión de hechos, por su parte, contendría los valores cuantitativos o de medida, como las cantidades, las ventas totales y los precios.

Jerarquías

Thomsen (1997) las describe como una estructura de árbol, con una raíz en la parte superior y las ramas en la parte inferior. Si todas las ramas se encuentran en un mismo nivel, se denomina jerarquía simétrica, pero si esto no ocurre, se denomina jerarquía asimétrica. Por su parte, en Oracle (1996) se define como una colección de miembros de dimensión con series en cascada de relaciones del tipo 1 a muchos, donde miembro es un valor de una dimensión. Por ejemplo, un miembro de la dimensión "tiempo" sería el mes "abril".

Tabla dimensional

Dentro del esquema estrella, corresponde a las tablas que están unidas a la tabla central por medio de sus respectivas llaves. La cantidad de estas tablas le otorgan la característica de multidimensionalidad al modelo.

Rotación

Consiste en cambiar o reacomodar la orientación dimensional de una presentación en pantalla.

Rool-up y Drill-down

Es posible transformar fácilmente la vista dentro de las jerarquías, subiendo o bajando niveles. Aprovechándolas es posible empezar en un nivel de resumen y selectivamente ir obteniendo detalle adicional para explicar observaciones hechas en el nivel de resumen. Cuando se pasa a un nivel más fino, como por ejemplo de mes a semana, se conoce con drill-down. Por el contrario, el pasar de un nivel más fino a otro más global, se conoce como roll-up.

Agregaciones

Para Wolff (2000), es la actividad de combinar datos desde múltiples tablas

para formar una unidad de información más compleja. Para Madsen (1996), las agregaciones consisten en la totalización y almacenamiento de datos disponibles en tablas de hechos para mejorar las búsquedas desarrolladas por el usuario final. Estrictamente hablando, las define como "una totalización precalculada de filas en una tabla de hechos a través de una o más jerarquías de dimensión" (Madsen, 1996:7).

En una base de datos multidimensional se pueden dar tres casos: 1- sin agregaciones, cuando el volumen de datos en pequeño, 2- selectivo, cuando la mayoría de las bases de datos de apoyo a las decisiones es suficientemente grande para requerir alguna agregación selectiva, y 3- completa, cuando se requiere una ejecución óptima, es decir, que una consulta lea un número mínimo de filas para devolver una respuesta.

La manera más eficiente para la implementación de las agregaciones es la construcción de una tabla separada para cada agregación creada, lo cual permite hacer y remover agregaciones independientes y agilizar su mantenimiento.

Modelos y metodologías para el desarrollo

Existen varios modelos de implementación y metodologías para el desarrollo de un DD. La utilización de alguno de ellos dependerá de los alcances y limitaciones del proyecto que se desee llevar a cabo.

En cuanto a los modelos de implementación, existen tres modelos: el de estrella, el de constelación de hechos y el esquema copo de nieve. El esquema estrella presenta el modelo básico de sobre el cual las aplicaciones pueden extenderse hacia los otros dos modelos. Consiste en una tabla central construida con llaves foráneas provenientes de un conjunto de tablas llamadas dimensiones.

Las llaves foráneas en conjunto corresponden a la llave primaria de dicha tabla central, que, conforme a la terminología de base de datos multidimensional también es llamada tabla de hechos. Según Cortéz (1999), cada una de las tablas de dimensión tiene una sola llave primaria que corresponde exactamente a uno de los componentes de la llave primaria compuesta en la tabla de hechos.

La tabla de hechos contiene, además de las llaves foráneas, una o más medidas (hechos numéricos) que ocurren como resultado de combinar las dimensiones que definen cada registro. Las tablas de dimensiones contienen información descriptiva.

En cuanto a las metodologías para el desarrollo de un DD, existen varias alternativas; sin embargo, se destacan dos representativas de tendencias diferentes, las cuales se describen a continuación.

Metodología de Poe

Poe presenta una visión general para la construcción de un DD por medio de una guía de desarrollo, la cual, en primer término, plantea una pregunta a primera vista muy obvia: ¿porqué se quiere un depósito de datos? La respuesta influirá en los alcances del desarrollo, afectará la calidad de los datos y la arquitectura de los sistemas que se elijan para llevarlo a cabo. Lo que esta pregunta busca es tener claro el objetivo que se persigue.

El segundo punto es entender cómo encaja el DD dentro de la empresa. Para cumplir este punto, se debe revisar el entorno automatizado de la empresa mediante la determinación de las redes instaladas de que dispone, las herramientas de consulta que utiliza, el sistema operacional en producción, las plataformas de uso, entre otros. Se debe tener presente que el fin del DD es proporcionar información para la toma de decisiones, por lo cual requerirá una nueva forma de pensar con respecto a la información. En resumen,

es necesaria la identificación exacta de los requerimientos, por medio de una revisión integral de la empresa, tanto en su infraestructura tecnológica como en las necesidades de los usuarios.

En su tercer punto, Poe hace una llamada de atención bastante fuerte: según su criterio, se debe eliminar la idea de que un DD puede implementarse en uno o dos meses, subestimando la cantidad de análisis que se requiere y la potencial complejidad de trabajar con tecnologías nuevas a la vez. El tiempo de desarrollo dependerá de los recursos disponibles, las plataformas sobre las que se trabajará y de la experiencia que se tenga.

El cuarto punto se refiere a la responsabilidad del equipo desarrollador con respecto a las actividades que se deben completar para llevar a cabo un proyecto exitoso. Dichas actividades son: 1- el análisis de los datos y su modelaje, 2- la definición de los requerimientos de los datos, 3- hallar y correlacionar las fuentes de información que se requerirán, 4- reconciliar los conflictos existentes entre los datos de diferentes fuentes, 5- transformar, integrar y calcular, según sean las especificaciones del proyecto y 6- diseñar la base de datos para acceso de solo lectura.

Un quinto punto se refiere a la educación del usuario con respecto a la diferencia entre los datos transaccionales y los de soporte a la toma de decisiones. En este sentido, en la mayoría de las organizaciones que manejan datos para la toma de decisiones aún no se tiene clara la diferencia entre los dos tipos de datos mencionados.

El sexto punto de la metodología tiene que ver con el presupuesto para el proyecto. Este debe contemplar la adquisición de herramientas costosas para la automatización de los procesos de transformación, extracción y generación de código, la contratación de personal especializado y experimentado, así como la compra de bases de datos diseñadas para el procesamiento de los datos.

El desarrollo o compra de las herramientas Front-End y de acceso a los datos con base en las habilidades y necesidades de los usuarios es el sétimo punto. La elección de la herramienta de consulta puede afectar significativamente la aceptación del DD por parte de los usuarios hacia los cuales va dirigido. Es de vital importancia ser realista sobre las necesidades y condiciones de los usuarios con respecto a la parte del sistema que está en contacto con ellos.

En el octavo y último punto, Poe recomienda buscar toda la experiencia necesaria, ya que se debe lidiar con tecnología nueva y de rápida evolución. A fin de contar con experiencia de primera mano, es importante tener contactos con empresas que ya hayan implantado un DD con éxito. Indica que el equipo de expertos debe estar compuesto por: 1- analistas de datos, diseño y administración de base de datos, 2- analistas de negocios con experiencia en trabajo con usuarios, 3- personas con experiencia en los actuales sistemas de información de la empresa, 4- administradores de proyectos con experiencia en el desarrollo de sistemas de apoyo a las decisiones y 5- programadores y diseñadores de interfaces gráficas de usuario (GUI).

Metodología de Corey

Una metodología novedosa la plantea Corey (1997) con su ciclo de vida del desarrollo de sistemas hacia atrás, o "la teoría de la evolución". Para este autor, no se debe realizar una especificación de requerimientos, en contraposición con muchos autores, pues, según su postura, los usuarios finales todavía no conocen cómo analizarán los datos. Es necesario que usuarios clave vean los datos sin limpiar. Se trata de un proceso de crecimiento, donde desarrolladores y usuarios trabajan conjuntamente determinando las herramientas apropiadas tanto para el

desarrollo como para la puesta en marcha del DD.

Para este autor, lo primero que se debe hacer es crear el equipo del proyecto, el cual debe estar en un continuo diálogo con el usuario, a fin de identificar sus preferencias acerca de las formas en que quiere representar los datos, tal y como se muestra en la figura 4.

MD y formas de procesamiento analítico en línea: MOLAP, ROLAP, HOLAP

Los DD y los MD almacenan información para su posterior búsqueda y análisis, a diferencia de los sistemas transaccionales, que la almacenan para operación y control. Es por ello que se requiere de la tecnología de OLAP encargada de realizar dicho análisis.

Las herramientas OLAP permiten el análisis de datos en forma interactiva. Cox lo define como la *"tecnología que permite que la información sea analizada de forma rápida, amigable e interactiva y que facilite la construcción de modelos cuantitativos complejos" (Cox, 2000:30).*

Es una característica del *software* que permite al usuario organizar, visualizar y resumir información desde distintas perspectivas, de manera rápida e interactiva.

El OLAP incluye, principalmente, la incorporación de grandes cantidades de datos dispersos y puede abarcar millones de datos con relaciones complejas. Su objetivo es contribuir al análisis de estas relaciones mediante la facilitación de la búsqueda de patrones, tendencias y excepciones. Debe soportar procesamiento lógico y estadístico de los resultados sin que el usuario tenga que programar, además de incluir los requerimientos de seguridad para la confidencialidad y las actualizaciones concurrentes.

El OLAP logra proveer una visión multidimensional de los datos por medio de la consulta interactiva y el análisis de datos. Permite la profundización sobre niveles cada vez más detallados o el ascenso a distintos niveles de resumen. Ofrece opciones de medición de datos numéricos mediante múltiples dimensiones, creando resúmenes y consolidaciones. Debe responder con rapidez a las consultas, de forma tal que el proceso de análisis no sea interrumpido y la información para la toma de decisiones no pierda valor.

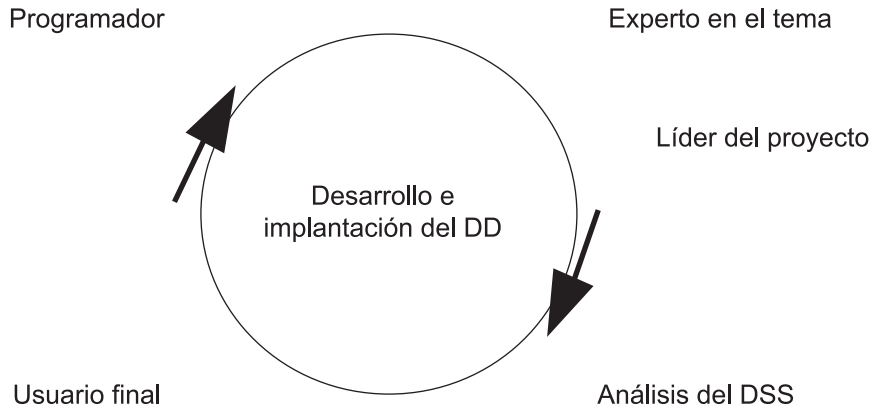

Figura 4. Equipo de proyecto de DD.

Al estar vinculadas las distintas dimensiones, una instancia de su intersección produce una celda, la cual contiene los valores de medida intersecados de las dimensiones involucradas. Una celda es un punto que corresponde a la intersección definida por el valor de cada una de las dimensiones del vector multidimensional.

En general, estos sistemas OLAP deben:

- Soportar requerimientos complejos de análisis.
- Analizar datos desde diferentes perspectivas.
- Soportar análisis complejos contra un volumen ingente de datos.

Los términos MOLAP, ROLAP y HOLAP, se refieren a las formas de almacenamiento de los datos. Ahora bien, la decisión se debe tomar dependiendo de si se quiere que el almacenamiento refleje los análisis que se harán (multidimensional), o los datos se almacenarán sin tomar en cuenta para qué serán utilizados (de forma relacional).

El primero de los casos se denomina MOLAP (Multidimensional On-Line Analytical Processing) y los datos se almacenan en forma lógica en arreglos matriciales con estructura multidimensional.

El sistema MOLAP utiliza una arquitectura de dos niveles: la base de datos multidimensional y el motor analítico (ver figura 5).

- La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato.
- El nivel de aplicación es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona una interfaz mediante la cual los usuarios finales visualizan los análisis OLAP. Una arquitectura cliente/servidor permite a varios usuarios acceder a la misma base de datos multidimensional.

Aquí los datos son redundantes, por lo que se requiere mucha capacidad de almacenamiento y puede contener totales precalculados. La estructura permite la reservación de espacio para todas las combinaciones de los posibles valores de las dimensiones de cada una de las variables, incluyendo los valores de dimensión que representan acumulados. Es decir, un

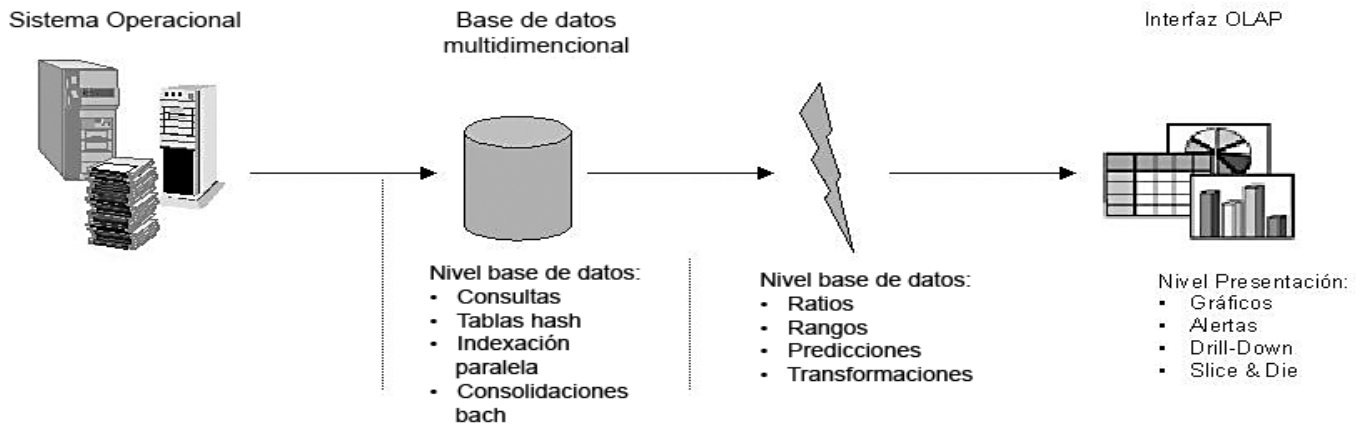

Figura 5. Multidimensional On-Line Analytical Processing (MOLAP).

sistema MOLAP contiene precalculados (almacenados) los resultados de todas las posibles consultas a la base de datos.

MOLAP consigue consultas muy rápidas a costa de mayores necesidades de almacenamiento, retardos en las modificaciones (que no deberían producirse salvo excepcionalmente), y largos procesos *batch* de carga y cálculo de acumulados.

El segundo caso denominado ROLAP (Relational On-Line Analytical Processing) es la arquitectura de base de datos multidimensional en la que la información se encuentra almacenada en una base de datos relacional, la cual tiene forma de estrella. En ROLAP, en principio, la base de datos solo almacena información relativa a los datos en detalle, evitando acumulados (redundancia). Aquí se crea una capa semántica de metadatos en una base de datos relacional, que permite presentar los datos en término de dimensiones. Su arquitectura es de tres niveles: la base de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica (ver figura 6).

- El nivel de base de datos usa bases de datos relacionales para el manejo, acceso y obtención de los datos.
- El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.
- El motor ROLAP se integra con niveles de presentación, a través de los cuales los usuarios realizan los análisis OLAP.

En ROLAP, al contener solo las combinaciones de valores de dimensión que representan detalle, es decir, al no haber redundancia, el archivo de base de datos es pequeño. Los procesos *batch* de carga son rápidos (ya que no se requiere agregación); sin embargo, las consultas pueden ser muy lentas, por lo que se aplica

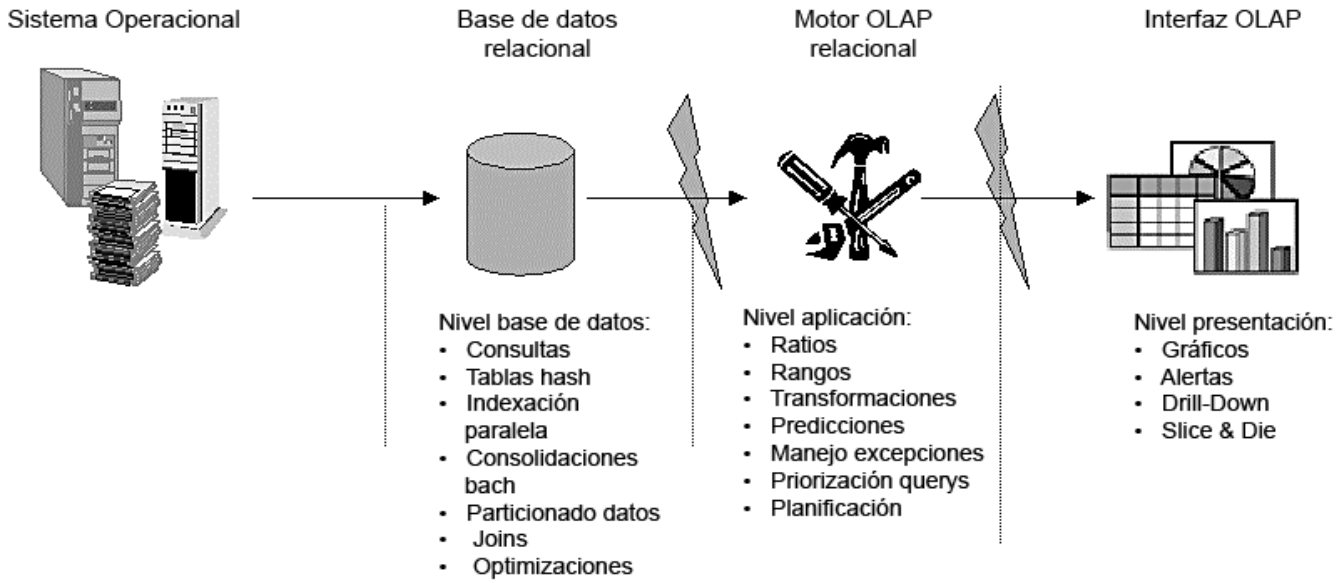

Figura 6. Esquema del procesamiento analítico en línea relacional (ROLAP)
Fuente: http://www.csae.map.es/csi/silice/DW2251.html

la solución de tener al menos algunas consultas precalculadas.

Por último, el HOLAP se refiere al sistema mixto o híbrido, donde la configuración típica es un motor multidimensional basado en una tecnología de almacenamiento relacional. Gill (1996) dice que en esta configuración las consultas e información frecuentes o de cómputo intensivo se calculan a priori y se almacenan en el depósito multidimensional.

Conclusiones

La tecnología de DD conlleva una nueva forma de pensar por parte de los desarrolladores, acostumbrados a trabajar sobre sistemas relacionales transaccionales. La utilización de modelos estrella, el manejo de redundancia y la falta de normalización muchas veces chocan con los esquemas de desarrollo de bases de datos y requieren de tiempo para ser asimilados. Pero para el manejo de bases de datos con información masiva para la toma de decisiones no es posible con los sistemas convencionales de administración transaccional.

Los DD y los MD vienen a llenar el vacío existente en cuanto a la utilización de esa valiosa veta de conocimientos, almacenada en los datos históricos.

Existen muchas metodologías de desarrollo de los MD; sin embargo, están enfocados en el modelo tradicional de cascada o espiral. Por el contrario, las metodologías expuestas en este artículo, se enfocan en brindar consejos prácticos, en el caso de Poe, y en escuchar siempre al usuario como elemento esencial para el éxito del proyecto. Sin embargo, la utilización de uno u otro dependerá siempre del tipo de proyecto que se emprenderá y las capacidades económicas y técnicas con que se cuenten.

Pero un MD no podría funcionar como tal sin un OLAP eficiente, capaz de procesar las consultas ad-hoc requeridas por los usuarios meta. La creación de OLAP constituye todo un reto para los desarrolladores, debido a la complejidad que representa. Un buen programa de consulta podrá explotar las capacidades de los MD y constituye la base para la explotación eficiente de la información.

Bibliografía

"Construindo um hipercubo", consulta del 10 de agosto del 2004, http://inga.ufu.br/~silvestr/artigos/bigbang/11/

Corey, Michael; Abbey, Michael. (1997) *Oracle Data Warehousing: guía práctica para analizar, construir e implantar con éxito un sistema data warehouse*. Madrid: McGraw-Hill.

Cortez Araniva, Sonia Elizabeth. (1999) *Desarrollo de un Modelo para el Diseño de Bases de Datos Multidimensionales*. Cartago: Instituto Tecnológico de Costa Rica.

Cox Alvarado, Alexander. (2000) *Propuesta de un álgebra para modelar bases de datos multidimensionales*. Cartago, Instituto Tecnológico de Costa Rica.

Chinchilla Arley, Ricardo; Vargas Murillo, José Luis. (2002) *Prototipo de mercado de datos para la División de Control, y Fiscalización de la Dirección General de Aduanas*. Cartago: Instituto Tecnológico de Costa Rica.

Devlin, Barry. (1996) *Data warehouse: from architecture to implementation*. New York: Addison-Wesley.

Ebel, Doug. (1998) *Data Warehousing: start to start*. NCR Business Solution Architect. S.l.: NCR Corporation. http://www.sdw.bull.com/ncr/dwstart.pdf

Gill, H.S; Rao, P.C. (1997) *Data Warehousing: la integración de información para la mejor toma de decisiones*. Prentice-Hall.

González Alvarado, Carlos. (1996) *Sistemas de bases de datos*. Cartago: Editorial Tecnológica de Costa Rica.

González Alvarado, Carlos. (1998) *Depósito de datos*. Cartago, Instituto Tecnológico de Costa Rica.

González Alvarado, Carlos. (1998) *Procesamiento analítico en línea*. Cartago, Instituto Tecnológico de Costa Rica.

Madsen, Mark. (1996) *Warehouse Design in the Aggregate. Database programming and design*.

Oracle Corporation. (1996) *Design a database for OLAP*. March/April. http://www.oramag.com/oracle/96-mar/26meth.html

Poe, Vidette. (1998) *Building a data warehouse for decision support.* New Jersey: Prentice-Hall.

Simon, Alan. (1996) *Beyond the warehouse. Database programming desing.*

Weldon, Jay; Weldon, Louise. (1995) *Managing multidimensional data*. Database programming and design.

Thomsen, Erik. (1997) *OLAP solution: building multidimensional information system.* New York: John Wiley & Sons.

Wolff, Carmen. [2000] *Implementando un DataWarehouse*. Revista Ingeniería Informática: revista electrónica del DIICC. Edición 5, año 3.