

Análisis del modelo de almacenamiento MOLAP frente al modelo de almacenamiento ROLAP

Comparing the MOLAP the ROLAP storage models

Marysol Tamayo¹ y Francisco Javier Moreno²

RESUMEN

En los últimos años los Data Warehouses (DW) en conjunto con OLAP se han constituido en elementos de ayuda en las organizaciones para la toma de decisiones. Con los tipos de almacenamiento de datos ROLAP y MOLAP se pueden crear los DW. El primero almacena los datos sobre una base de datos relacional. El tipo de almacenamiento MOLAP en matrices multidimensionales. En este artículo se presenta un ejemplo comparativo que permite analizar el rendimiento, las ventajas y desventajas entre estos dos tipos de almacenamiento en un sistema de gestión de base de datos (SGBD) específico. Adicionalmente se ofrece un panorama que permite observar cómo los SGBDs están incorporando hoy en día estas tecnologías dentro de sus principales funcionalidades.

Palabras clave: data warehouse, OLAP, MOLAP, ROLAP.

ABSTRACT

Data Warehouses (DWs), supported by OLAP, have played a key role in helping company decision-making during the last few years. DWs can be stored in ROLAP and/or MOLAP data storage systems. Data is stored in a relational database in ROLAP and in multidimensional matrices in MOLAP. This paper presents a comparative example, analysing the performance and advantages and disadvantages of ROLAP and MOLAP in a specific database management system (DBMS). An overview of DBMS is also given to see how these technologies are being incorporated.

Keywords: data warehouse, OLAP, MOLAP, ROLAP.

Recibido: noviembre 16 de 2005

Aceptado: octubre 9 de 2006

Introducción

En la actualidad muchas organizaciones poseen grandes volúmenes de datos almacenados que requieren ser analizados. El objetivo de crear un DW es el de que una gran cantidad de datos sean transformados en información para que en conjunto con herramientas OLAP, paquetes estadísticos profesionales y herramientas de minería de datos sirva para la toma de decisiones (Elmasri, 2006). Por tal motivo las organizaciones han desarrollado DW.

Un DW debe ir acompañado de un modelo de datos, los dos más utilizados son el relacional y el multidimensional. El modelo relacional es ampliamente soportado en diferentes SGBD. Típicamente un DW puede almacenarse de dos formas: ROLAP y MOLAP. El tipo de almacenamiento de datos ROLAP guarda los datos en una base de datos (BD) relacional. El MOLAP guarda los datos en matrices multidimensionales. Los datos de un DW se pueden representar

por medio de un cubo o hipercubo, el cual representa un conjunto de hechos y otro de dimensiones (véase sección de modelos multidimensionales). El cubo consta de una serie de celdas, cada una representa un hecho que surge a raíz de la combinación de las diferentes dimensiones.

Hoy en día se puede observar la evolución de los SGBD como consecuencia de la necesidad de analizar grandes volúmenes de datos. Uno de estos cambios está enfocado hacia el almacenamiento multidimensional, con el objetivo de recuperar y analizar información compleja rápidamente. En este artículo se presenta un caso de estudio con el fin de analizar el rendimiento de ROLAP y MOLAP.

El cuerpo del artículo es el siguiente: en la sección "El Data Warehouse" se define el concepto de DW las secciones siguientes describen los modelos multidimensionales OLAP

¹ Ingeniera de sistemas, Universidad San Buenaventura, Medellín, Colombia. Candidata a Especialista en Sistemas. Investigadora, Grupo de Ingeniería de Software, Escuela de Sistemas, Universidad Nacional de Colombia, Medellín. mtamayoh@community.com.co

² Ingeniero de sistemas, Universidad de Antioquia, Colombia. Especialista en Gestión y Sistemas, Universidad de Antioquia, Colombia. M.Sc. en ingeniería de sistemas, Universidad Nacional de Colombia. Profesor Asistente de Tiempo Completo, Escuela de Sistemas Universidad Nacional de Colombia, Medellín. fjmoreno@unalmed.edu.co

y los tipos de almacenamiento para OLAP, respectivamente. La sección "SGBD con soporte para ROLAP y MOLAP" presenta SGBD que soportan MOLAP. En la siguiente sección se expone el caso de estudio, base para la comparación realizada, y finalmente se presentan las conclusiones y trabajos futuros.

El Data Warehouse

La mayoría de las decisiones en las organizaciones se basan en información de experiencias pasadas. Generalmente, la información que se quiere investigar sobre un cierto dominio de la organización se encuentra en fuentes internas, BD transaccionales y fuentes externas (documentos de texto, archivos HTML, entre otras.)

Muchas de estas fuentes se utilizan para el trabajo diario. Actualmente en muchas empresas, el análisis para la toma de decisiones se realiza sobre estas BD de trabajo o BD transaccionales. La situación es esta:

- Se realiza el trabajo diario en las BD transaccionales (proceso conocido como OLTP, On-Line Transactional Processing).
- Se efectúa el análisis de los datos en tiempo real sobre dichas BD (proceso conocido como OLAP, On-Line Analytical Processing).

Tal situación provoca algunos problemas. En primer lugar, perturba el trabajo diario en las BD transaccionales, ya que se realizan consultas muy pesadas sobre ellas. En segundo término, dichas BD están diseñadas para el trabajo transaccional, no para el análisis de los datos.

Como consecuencia de lo anterior surgen los DW y toda su tecnología asociada (Data Warehousing). Los DW recogen los datos de los distintos entornos transaccionales de la compañía, los filtran y procesan para su almacenamiento, proporcionando una plataforma sólida de datos consolidados e históricos para su posterior análisis.

Un DW se define como una colección de datos orientada al tema, integrada, temporal y no volátil, usada principalmente para la toma de decisiones (Inmon, 1996).

Los siguientes términos de la definición merecen ser aclarados (Inmon, 1996):

- Orientación al tema: la información se clasifica en base a los temas o aspectos que son de interés para el analista o usuario final. Por ejemplo: venta, inventario, crimen, etc. Por cada uno de ellos hay sujetos de interés como productos, clientes, proveedores, etcétera.
- La integración de los datos: se refiere a la integración de datos recogidos de diferentes sistemas operacionales de la organiza-

ción o fuentes externas. Se manifiesta en convenciones de nombres (estandarización), en la medida uniforme de las variables, entre otros.

- Temporal: los datos almacenados están referidos a un período de tiempo específico.
- No volátil: una vez almacenados los datos, estos no son modificados.

Modelos multidimensionales

En un modelo de datos multidimensional los datos se organizan alrededor de los temas de la organización. La estructura de datos manejada en este modelo son matrices multidimensionales o hipercubos. Un hipercubo consiste en un conjunto de celdas, cada una se identifica por la combinación de los miembros de las diferentes dimensiones y contiene el valor de la medida analizada para dicha combinación de dimensiones.

- **Hecho:** es el objeto a analizar, posee atributos llamados de hechos o de síntesis, y son de tipo cuantitativo. Sus valores (*medidas*) se obtienen generalmente por la aplicación de una función estadística que resume un conjunto de valores en un único valor. Por ejemplo: ventas en dólares, cantidad de unidades en inventario, cantidad de unidades de producto vendidas, horas trabajadas, promedio de piezas producidas, consumo de combustible de un vehículo, etcétera.
- **Dimensiones:** representan cada uno de los ejes en un espacio multidimensional. Suministran el contexto en el que se obtienen las medidas de un hecho. Algunos ejemplos son: tiempo, producto, cliente, departamento, entre otras. Las dimensiones se utilizan para seleccionar y agrupar los datos en un nivel de detalle deseado. Los componentes de una dimensión se denominan niveles y se organizan en jerarquías, verbigracia, la dimensión tiempo puede tener niveles día, mes y año.

Los hechos se guardan en tablas de hechos y las dimensiones en tablas de dimensiones.

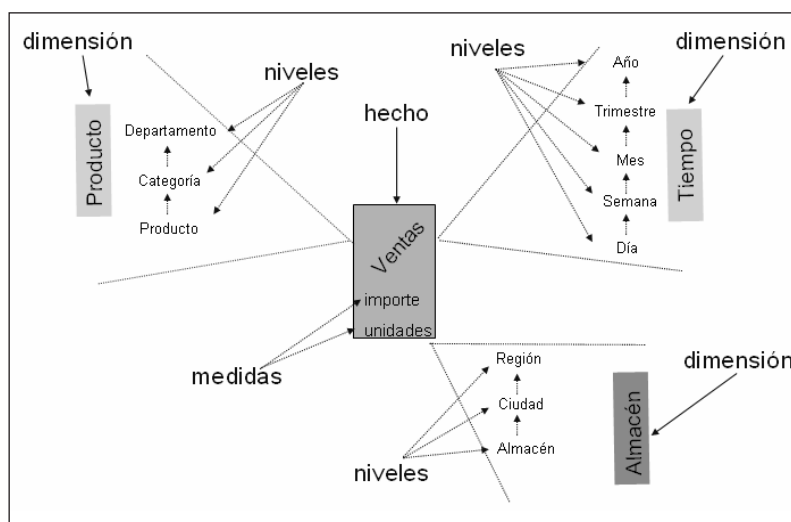


Figura 1. Modelo multidimensional

En la Figura 1 se muestra un modelo multidimensional, donde de hechos es la tabla ventas y las dimensiones son almacén, producto y tiempo.

Un modelo multidimensional se puede representar como un esquema en estrella, copo de nieve (*snowflake*) o constelación de hechos (Chaudhuri, 1997; Kimball, 2002).

- Esquema en estrella: está formado por una tabla de hechos y una tabla para cada dimensión.
- Esquema copo de nieve: es una variante del esquema en estrella que presenta las tablas de dimensión normalizadas.
- Constelación de hechos: son varios esquemas en estrella o copo de nieve que comparten dimensiones.

Ejemplo de cubo: En la Figura 2 se tiene una BD que maneja tres dimensiones: países, productos y períodos de entrega (tiempo). Los datos pueden representarse como un cubo de tres dimensiones; cada valor individual de una celda representa la cantidad total de un producto vendido a un país en una fecha determinada.

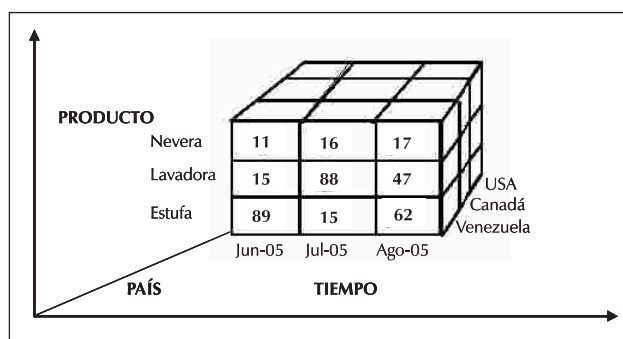


Figura 2. Cubo de datos

OLAP (Procesamiento analítico en línea)

La tecnología OLAP facilita el análisis de datos en línea en un DW, proporcionando respuestas rápidas a consultas analíticas complejas. OLAP es utilizado generalmente para ayuda en la toma de decisiones y presenta los datos a los usuarios a través de un modelo de datos intuitivo y natural. Con este estilo de presentación los usuarios finales pueden ver y entender con mayor facilidad la información de sus BD, lo que permite a las organizaciones reconocer el valor de sus datos.

Generalmente los esquemas de las BD tienen cierta complejidad para el usuario final, debido a ello la concepción de las consultas puede ser una tarea ardua.

OLAP ofrece un conjunto de operadores que facilitan la concepción de consultas, algunos de ellos son Slice & Dice, Swap, Drill Down, Drill Up, Roll-Up, Drill-Across, Drill-Through (Chaudhuri, 1997).

Modos de almacenamiento de OLAP

OLAP puede trabajar con tres tipos de almacenamiento:

Almacenamiento ROLAP (Relational OLAP)

En ROLAP se utiliza una arquitectura de tres niveles. La BD relacional maneja el almacenamiento de datos, el motor OLAP proporciona la funcionalidad analítica, y alguna herramienta especializada es empleada para el nivel de presentación.

El nivel de aplicación es el motor OLAP, que ejecuta las consultas de los usuarios.

El motor OLAP se integra con el nivel de presentación a través del cual los usuarios realizan los análisis OLAP.

Después de que el modelo de datos para el DW se ha definido, los datos se cargan desde los sistemas transaccionales.

Los usuarios finales ejecutan sus análisis multidimensionales, a través del motor OLAP, el cual transforma sus datos a consultas en SQL ejecutadas en las BD relacionales y sus resultados son devueltos a los usuarios.

La arquitectura ROLAP es capaz de usar datos precalculados (si estos están disponibles), o de generar dinámicamente los resultados desde la información elemental (menos resumida). Esta arquitectura accede directamente a los datos del DW y soporta técnicas de optimización para acelerar las consultas como tablas particionadas, soporte a la desnormalización, soporte de múltiples reuniones, precalculado de datos, índices etcétera.

Almacenamiento MOLAP (multidimensional OLAP)

Un sistema MOLAP usa una BD multidimensional (BDMD), en la que la información se almacena multidimensionalmente.

El sistema MOLAP utiliza una arquitectura de dos niveles: la BDMD y el motor analítico.

La BDMD es la encargada del manejo, acceso y obtención de los datos.

El nivel de aplicación es el responsable de la ejecución de las consultas OLAP.

El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través de la cual los usuarios finales visualizan los análisis OLAP.

La información procedente de los sistemas transaccionales se carga en el sistema MOLAP. Una vez cargados los datos en la BDMD, se realiza una serie de cálculos para obtener datos agregados a través de las dimensiones del negocio, poblando la estructura de la BDMD.

Luego de llenar esta estructura, se generan índices y se emplean algoritmos de tablas *hash* para mejorar los tiempos de accesos de las consultas. Una vez que el proceso de poblado ha finalizado, la BDMD está lista para su uso. Los

usuarios solicitan informes a través de la interfaz y la lógica de aplicación de la BDMD obtiene los datos.

Almacenamiento HOLAP (Hybrid OLAP)

Se han desarrollado soluciones de OLAP híbridas que combinan el uso de las arquitecturas ROLAP y MOLAP. En una solución con HOLAP, los registros detallados (los volúmenes más grandes) se mantienen en la BD relacional, mientras que los agregados lo hacen en un almacén MOLAP independiente (Ibarzábal, 2003).

Estrategias de agregación y almacenamiento

Los servidores OLAP se clasifican de acuerdo a como se almacenan los datos:

- Un servidor MOLAP almacena los datos en disco en estructuras optimizadas para acceso multidimensional. Típicamente, los datos son almacenados en arreglos densos, los cuales requieren cuatro u ocho bytes por celda.
- Un servidor ROLAP almacena sus datos en una BD relacional. Cada fila de una tabla de hechos tiene una columna para cada dimensión y otra para cada medida.

Es necesario almacenar tres tipos de datos: hechos, agregados y dimensiones.

Una de las características distintivas de MOLAP es la preconsolidación de los datos. En una BD relacional para responder a una consulta del tipo ¿cuánta cantidad del producto X se vendió en el último trimestre? normalmente se tiene que hacer una búsqueda de todos los registros relevantes y totalizar los datos. En una BDMD, en cambio, estos totales se calculan rápidamente usando operaciones sobre arreglos. Una vez calculados, los totales se pueden almacenar en estructuras de la misma BDMD. Las BDMD pueden preconsolidar agregados en los diferentes niveles de las dimensiones, por ejemplo: totales por semana, totales por mes, gran total. El preconsolidado de estos agregados puede requerir mucho espacio y tiempo de carga. Una alternativa consiste en preconsolidar sólo los totales más usados y calcular el resto en el momento en el que se consultan.

Otra característica importante en MOLAP son los datos dispersos. La dispersión de datos surge en casos donde no todas las combinaciones de miembros de las dimensiones van a tener su valor correspondiente (Lehner, 1998), como en el caso de una organización con varias sucursales, que puede vender cientos de productos por día en cada una, pero no todos ellos necesariamente se van a vender todos los días en todas las sucursales. Si se analizan estas ventas en períodos diarios y por sucursal creando un cubo, con las ventas como medida y sucursales, productos y días como dimensiones, el cubo contendrá algunas celdas vacías.

Cada herramienta MOLAP tiene su propio mecanismo para evitar guardar explícitamente este tipo de celdas. En general se comprime la BD, con el consiguiente costo de descomprimirla cuando se accede a los datos.

Un sistema HOLAP resuelve el problema de dispersión, dejando los datos más granulares (menos agregados) en la BD relacional,³ pero almacena los agregados en un formato multidimensional, minimizando así la presencia de celdas vacías.

Es necesario el precálculo de agregados cuando el conjunto de datos es muy grande, de otra forma ciertas consultas podrían no ser resueltas sin leer toda la tabla de hechos.

Los agregados en ROLAP son almacenados en tablas. En algunos sistemas ROLAP los agregados son manejados explícitamente por el servidor OLAP en otros sistemas como en Oracle, las tablas son declaradas como vistas materializadas (Gupta, 1999) y son usadas implícitamente cuando el servidor OLAP lanza una consulta que se corresponde con la definición de la vista (Oracle Corporation, 2005).

El componente final de la estrategia de agregación es la memoria caché. Esta guarda agregados precalculados en memoria de tal forma que las consultas futuras puedan acceder a los valores de las celdas sin ir al disco. Si la memoria caché almacena los datos en un nivel bajo de agregación entonces podrá calcular agregados a un nivel más alto si son requeridos.

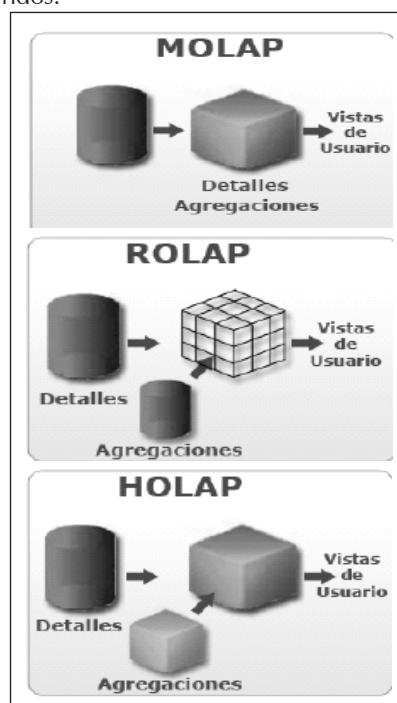


Figura 3. Tipos de almacenamiento

La memoria caché es una de las partes más importantes de la estrategia de agregación porque es adaptativa. En general es

³ Nótese que en una BD relacional, en la tabla de hechos sólo se guardan las combinaciones de dimensiones que generaron "en verdad" una medida (Date, 2001).

difícil elegir el conjunto de agregados a precalcular, los cuales le den velocidad al sistema sin usar grandes cantidades de espacio, particularmente cuando hay muchas dimensiones o cuando los usuarios están emitiendo consultas impredecibles constantemente. En un sistema donde los datos están cambiando en tiempo real, es impráctico mantener los agregados precalculados. Una memoria caché de tamaño razonable puede permitir que un sistema se desempeñe adecuadamente al enfrentar consultas impredecibles, con pocos o sin agregados precalculados.

La Figura 3 muestra los tres tipos de almacenamiento: MOLAP, ROLAP y HOLAP.

Diferencias entre ROLAP y MOLAP

En los últimos años se han producido debates alrededor de los tipos de almacenamiento MOLAP y ROLAP. Por lo general, las implementaciones de MOLAP presentan mejor rendimiento que la tecnología relacional; sin embargo, tienen problemas de escalabilidad, por ejemplo, la adición de dimensiones a un esquema ya existente. Por otra parte, las implementaciones de ROLAP son más escalables y a menudo son más atractivas debido a que aprovechan las inversiones efectuadas en tecnología de BD relacionales.

La Tabla 1 resume las diferencias entre ambas tecnologías.

Tabla 1. Diferencias entre ROLAP y MOLAP

	<i>Multidimensional</i>	<i>Relacional</i>
Datos	Detalle y precalculados (agregados)	Detalle y agregados
Estructura	Matrices comprimidas	Tablas relacionales
Administración	Especialista en BDMD	Administrador BD
Acceso	Lenguaje especializado	SQL

Las Tablas 2 y 3 detallan las ventajas y desventajas de cada tipo de almacenamiento.

Tabla 2. Ventajas y desventajas de ROLAP

<i>Ventajas</i>	<i>Desventajas</i>
<ul style="list-style-type: none"> -Se puede aprovechar la tecnología relacional, ya que se facilita aprovechar las inversiones realizadas en <i>hardware</i> y en SGBD relacionales -Uso de la seguridad e integridad de los SGBD relacionales -Capaz de manejar conjuntos de datos muy grandes, por encima de un terabyte -Pueden utilizarse SGBD relacionales gratuitos -Puede soportar un gran número de dimensiones -s escalable (adición de dimensiones a un esquema existente). 	<ul style="list-style-type: none"> -Es menor en rendimiento frente a MOLAP en BD pequeñas -Limitación para consultas complejas, debido a que se pueden requerir muchas reuniones para obtener la consulta deseada -Utilización de mucho almacenamiento en disco (NewTec Ediciones, 2002).

SGBD con soporte para ROLAP y MOLAP

Entre los SGBD que permiten utilizar almacenamiento de datos de tipo ROLAP y que han incorporado características adicionales (Gray, 1997) para su manejo están Oracle, DB2 y SQL Server.

Por otro lado, entre los SGBD que permiten utilizar almacenamiento de datos de tipo MOLAP están:

- **SQL Server - Microsoft Analysis Services:** soporta la construcción y gestión de cubos multidimensionales, permite flexibilidad en los modos de almacenamiento, ya que también soporta ROLAP (NewTec Ediciones, 2002)
- **Hyperion:** fabricante de herramientas analíticas que se apoyan en OLAP. Hyperion Essbase OLAP Server es la plataforma empresarial para la elaboración de informes, análisis, modelos y presupuestos. Permite el acceso de lectura/escritura de múltiples usuarios, capacidad de almacenamiento de grandes volúmenes de datos, realización de cálculos analíticos complejos y consultas OLAP sofisticadas (Hyperion, 2002).
- **Oracle Express:** contiene herramientas y aplicaciones que se apoyan en Oracle Express Server, un motor de cálculo y gestor de memoria caché de datos. Las herramientas Oracle OLAP toman en consideración todo lo referente a las necesidades de los usuarios, desde consultas y análisis simples de los datos contenidos en un DW, hasta análisis, presupuestación y modelaje sofisticados y desarrollo de aplicaciones OLAP orientados a objetos (Audifilm Grupo Brime, 2003).

<i>Ventajas</i>	<i>Desventajas</i>
<ul style="list-style-type: none"> -Ofrece buen rendimiento cuando se trabaja sobre datos agregados, totales, subtotales, series temporales y diversos grados de detalle de los datos -Facilita el estudio a alto nivel de los datos, al ofrecer una mayor flexibilidad y rapidez de acceso para el análisis de los mismos -Almacenamiento de datos y consultas bastante eficientes -La complejidad de la BD se oculta a los usuarios -El análisis se hace sobre datos agregados y métricas o indicadores precalculados -Mayor rendimiento frente a ROLAP en el procesamiento de consultas en BD pequeñas -Almacena agregados para facilitar un acceso rápido. 	<ul style="list-style-type: none"> -La asimilación de los conceptos multidimensionales, en especial, cuando se tienen hipercubos de muchas dimensiones -La construcción y poblado de las estructuras multidimensionales puede demandar mucho tiempo -Están limitados a tener diez o menos dimensiones debido a la complejidad para el manejo de las mismas -No se puede acceder a datos que no están en el cubo (Ibarra, 2005) -Debe trabajar con volúmenes de datos limitados, menos de 5 GB (Nader, 2003) -Existen pocas herramientas gratuitas que lo soporten.

Tabla 3. Ventajas y desventajas de MOLAP

Desarrollo de un caso de estudio

Para la comparación y el análisis del rendimiento de los tipos de almacenamiento ROLAP y MOLAP se realizó un caso de estudio en el SGBD SQL Server con su componente Microsoft Analysis Services.

Caso de estudio: ventas

El caso de estudio realizado consiste en las ventas de productos de una cadena de almacenes que se encuentran en varias ciudades del país.

Antes de diseñar los cubos es necesario establecer una BD OLAP. Esta es similar a una BD SQL Server relacional, sin embargo la última contiene tablas relacionales y vistas, mientras que una BD OLAP contiene cubos multidimensionales, dimensiones, orígenes de datos y otros objetos.

En la BD OLAP seleccionada se crearon seis dimensiones: almacén, cliente, geografía, producto, vendedor y tiempo.

Se crearon ocho cubos, cuatro con tipo de almacenamiento MOLAP y cuatro con tipo de almacenamiento ROLAP en forma correspondiente. Estos cubos se describen en las próximas secciones.

Cubo comportamiento clientes

La tabla de hechos es la de ventas, y las dimensiones son clientes, producto y tiempo, como se muestra en la Figura 4. Este cubo muestra la información sobre el comportamiento de los clientes a través del tiempo en cuanto a la compra de productos.

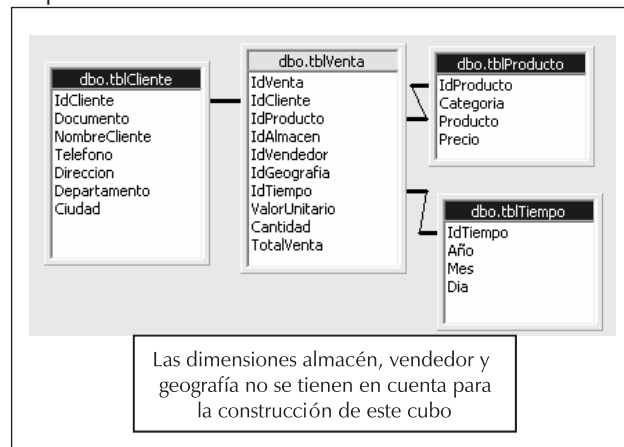


Figura 4. Cubo comportamiento clientes

Cubo desempeño vendedores

La tabla de hechos es la de ventas y las dimensiones son vendedor, producto y tiempo, como se muestra en la Figura 5. Este cubo permite estudiar y evaluar cómo ha sido el desempeño de cada vendedor a través del tiempo en cuanto a la venta de productos.

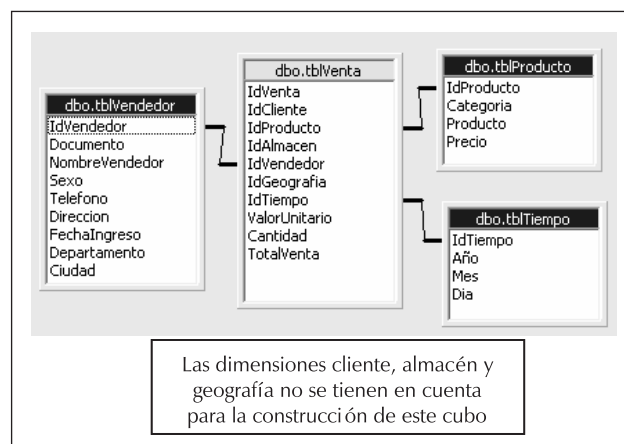


Figura 5. Cubo desempeño vendedores

Cubo ventas por almacén

La tabla de hechos es la de ventas y las dimensiones son almacén, producto y tiempo, como se muestra en la Figura 6. Este cubo se creó con el fin de analizar las ventas de productos de cada almacén a través del tiempo.

Cubo ventas por geografía

La tabla de hechos es la de ventas y las dimensiones son geografía, producto y tiempo, como se muestra en la Figura 7. Este cubo permite un análisis a través del tiempo acerca de las variaciones registradas en las ventas de productos en los distintos niveles geográficos (ciudad, departamento, país) donde hay almacenes.

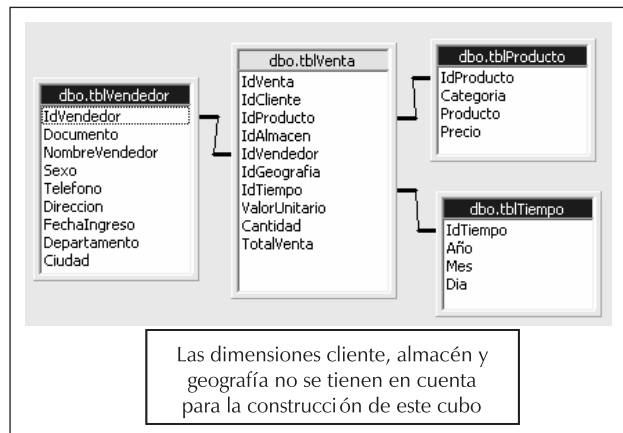


Figura 6. Cubo Ventas por almacén

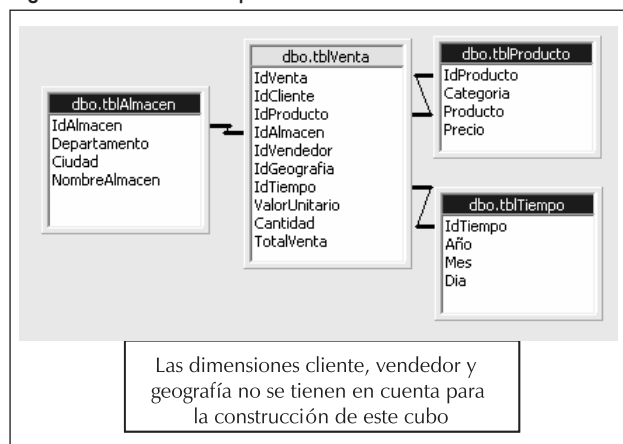


Figura 7. Cubo ventas por geografía

Generación de los informes y análisis de los datos

Para la presentación y análisis de los datos se utilizaron tres herramientas, las cuales permiten filtrar los datos según las dimensiones, aplicar operaciones para ver los datos más agregados (*Drill Up*) o para ver los datos más detallados (*Drill Down*):

- Cube Browser incluido en Analysis Services
- Excel 2003, junto con un servidor SQL Server OLAP Server y un cliente motor de cálculo y almacenamiento en caché denominado Microsoft PivotTable Service se logra el análisis de los cubos (Microsoft Corporation, 2003)

- MDX (Multidimensional Expressions), el cual es un lenguaje usado para consultar una BD OLAP en SQL Server 2000 Analysis Services (Microsoft Corporation 2, 2004), de la misma forma en que se usa SQL para consultar una BD SQL Server. Comparte con SQL la misma estructura básica, pero tiene algunas características adicionales creadas especialmente para manipular este tipo de BD.

Tamaño de la BD OLAP utilizada

La tabla de hechos y las dimensiones de la BD OLAP utilizada tienen la siguiente cantidad de registros. Tabla de hechos Venta: 124.049, Dimensión Cliente: 65.000, Dimensión Producto: 28, Dimensión Vendedor: 39, Dimensión Almacén: 20, Dimensión Geografía: 11 y Dimensión Tiempo: 365.

Informes

Se generó un informe por cada cubo. A cada informe se le aplicaron las diferentes operaciones como: *Slice & Dice*, *Drill Down*, *Drill Up*, entre otros, para observar el comportamiento de los datos y analizar el tiempo de respuesta de cada informe.

Informe 1 de comportamiento de clientes

Se filtraron los clientes (dimensión cliente) de un departamento específico agrupados por categoría (dimensión producto) y por día (dimensión tiempo).

Informe 2 de comportamiento de clientes

Se filtraron los clientes (dimensión cliente) de dos departamentos específicos agrupados por categoría (dimensión producto) y por año (dimensión tiempo).

Informe de desempeño de vendedores

Informe de todos los vendedores (dimensión vendedor) agrupados por producto (dimensión producto) y por día (dimensión tiempo).

Informe de ventas por almacén

Informe de todos los almacenes (dimensión almacén) agrupados por producto (dimensión producto) y por día (dimensión tiempo).

Informe de ventas por geografía

Informe de todas las ciudades (dimensión geografía) agrupadas por producto (dimensión producto) y por día (dimensión tiempo).

La Tabla 4 resume los resultados obtenidos de los cinco informes.

Tabla 4. Tiempos de respuesta de los informes

Nombre del Informe	Tiempo de respuesta MOLAP	Tiempo de respuesta ROLAP
Informe 1 de comportamiento de clientes	14 s	20 s
Informe 2 de comportamiento de clientes	7 s	11 s
Informe de desempeño de vendedores	2 s	6 s
Informe de ventas por almacén	5 s	10 s
Informe de ventas por geografía	4 s	7 s

Otros informes

Cuando se generaron informes con las dimensiones en su nivel más alto de agregación, los tiempos de respuestas eran de dos y tres segundos en los dos tipos de almacenamientos.

Conclusiones y trabajos futuros

Con el caso de estudio se pudo corroborar que es mejor el rendimiento de las consultas cuando se usó MOLAP en BD pequeñas (al menos en SQL Server).

Se pudo observar que MOLAP tiene mejor o igual rendimiento que ROLAP, aunque la diferencia en los tiempos de respuesta obtenidos en la generación de los informes fue poco significativa.

El caso permitió comprender mejor los conceptos multidimensionales, como las medidas y el trabajo con datos agregados. Además posibilita la identificación de patrones y tendencias

También permitió observar el estado del arte de las herramientas comerciales en cuanto al manejo de modelos multidimensionales, se pudo comprobar que las utilizadas facilitan la creación de una BD OLAP, sus dimensiones, cubos y todo lo que implica su procesamiento, al igual que el uso de las herramientas OLAP para el análisis y manipulación de los datos.

La respuesta a la pregunta ¿cuál es la mejor opción entre ROLAP y MOLAP?, es que cada alternativa tiene sus ventajas y desventajas. En lugar de discutir cuál de las dos es mejor hay que definir criterios para optar por una u otra. También se debe considerar la opción de HOLAP, que intenta combinar lo mejor de ambas opciones. Algunos criterios son: tamaño actual y futuro de la BD, número de dimensiones y complejidad de las consultas.

En forma general los productos MOLAP proporcionan cálculos más rápidos, con la desventaja de que soportan cantidades de datos más pequeñas que las soluciones basadas en ROLAP, que ofrecen características de escalabilidad, concurrencia y administración más maduras.

MOLAP es una solución adecuada para soluciones departamentales con unos volúmenes de información y número de dimensiones limitados.

Como trabajos futuros se tiene contemplado realizar pruebas con BD más grandes, comparaciones entre diferentes SGBD que soporten MOLAP, e igualmente confrontar con sus opciones en ROLAP. También es de interés explorar soluciones tipo HOLAP para determinar su desempeño, ventajas y desventajas.

Bibliografía

Audifilm Grupo Brime., Oracle Express technology., Reporte técnico, 2003.

Chaudhuri, S. and Dayal, U., An overview of data warehousing and OLAP technology., SIGMOD Record, 1997.

Date, C. J., Introducción a los sistemas de bases de datos., Prentice Hall, 2001.

Elmasri, R. and Navathe, S., Fundamentals of database systems., Addison Wesley, 2006.

Gray, J., Chaudhuri, S., Bosworth, A., Layman, A., Reichart, D., Venkatrao, M., Data cube: a relational aggregation operator generalizing group-by, cross-tab, and sub-totals., Data Mining and Knowledge Discovery, 1997.

Gupta, A. and Mumick, I., Materialized views: techniques, implementations, and applications., The MIT Press, Massachusetts, 1999.

Hyperion., Hyperion Essbase OLAPServer., Reporte técnico, 2002.

Ibarra, M., Procesamiento analítico en línea., Universidad Nacional del Nordeste Corrientes Argentina, Reporte técnico, 2005.

Ibarzabal, J., Estrategia de reporting., Cedyc S.Coop., Sangroniz, 2003.

Inmon, W. H., Building the data warehouse., John Wiley & Sons, New York, 1996.

Kimball, R. and Ross, M., The data warehouse toolkit: the complete guide to dimensional modelling., John Wiley & Sons, New York, 2002.

Lehner, W., Albrecht, H. and Wedekind, H., Normal forms for multidimensional databases., Proc. 10th International Conference on Scientific and Statistical Database Management (SSDBM), 1998.

Microsoft Corporation., Servicios de ayuda a la toma de decisiones de Microsoft SQL Server 7.0., Reporte técnico, 2003.

Microsoft Corporation 2., MSDN Library: MDX Language., Reporte técnico, 2004.

Nader, J., Sistema de apoyo gerencial universitario., Tesis presentada al Instituto Tecnológico de Buenos Aires, para optar al grado de magíster en Ingeniería del Software, 2003.

NewTec Ediciones., SQL Server 7.0 y OLAP Server., Reporte técnico, Barcelona, 2002.

Oracle Corporation., Oracle materialized views and query rewrite., Informe Técnico Oracle Corporation, 2005.

Grupo de Ingeniería de Software Universidad Nacional de Colombia -Sede Medellín

Líneas de trabajo:

- Bases de Datos Multidimensionales
 - Ingeniería de Requisitos
 - Modelado con UML

Grupo de apoyo a la Maestría y al Doctorado en Ingeniería de Sistemas de la Universidad Nacional Sede Medellín



Mayor información: Francisco Javier Moreno,
fjmoreno@unalmed.edu.co