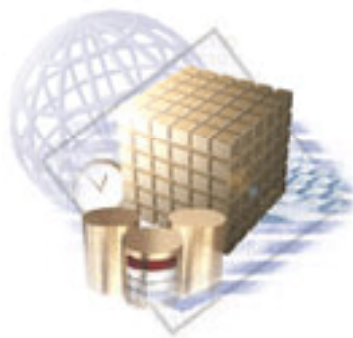




Universidad Nacional del Nordeste
Facultad de Ciencias Exactas, Naturales y Agrimensura

Trabajo de Adscripción
**Procesamiento Analítico en Línea
(OLAP)**



Adscripta: Maria de los Angeles Ibarra - L.U.: 30.505

Materia: Diseño y Administración de Datos

Director: Mgter. David Luis la Red Martínez

Licenciatura en Sistemas de Información
Corrientes - Argentina

2006

Índice General

1	Introducción	1
1.1	Las Bases de Datos	1
1.1.1	Estructura Multidimensional	2
1.2	Data Warehouse	3
1.2.1	Características del Data Warehouse	4
1.2.2	Principales Ventajas y Desventajas de un DW para las Organizaciones	6
1.2.3	Requerimientos Fundamentales de un Data Warehouse	7
1.2.4	Componentes de un Data Warehouse	8
1.2.5	Factores de Exito de un Data Warehouse	9
1.2.6	Problemas en la implementación	10
1.2.7	Caso del Mundo Real	10
1.3	Data Mart	12
1.4	Sistemas OLTP	14
1.4.1	Diferencias entre OLTP y OLAP	14
1.4.2	Diferencias entre OLTP y Data Warehouse	15
1.5	Sistemas Olap	16
1.5.1	Vista de los Datos	17
1.5.2	Modelos de Almacenamientos	18
1.5.3	Utilidades de OLAP	23
1.5.4	Beneficios que Presenta OLAP	24
1.5.5	Casos del Mundo Real	25
2	Introducción a OLAP	27
2.1	Introducción	27
2.2	Operadores OLAP	29
2.3	Tipos de Organizaciones	33
2.3.1	Ventajas e Inconvenientes de los sistemas MOLAP	34
2.3.2	Ventajas e Inconvenientes de los sistemas ROLAP	35

2.4	Conceptos	35
2.4.1	Multidimensionales	35
2.4.2	Esquema en Estrella	36
2.4.3	Medidas	38
2.4.4	Dimensiones	38
2.4.5	Hechos	39
2.4.6	Esquema copo de Nieve	40
2.5	Casos del Mundo Real	41

Índice de Figuras

1.1	Estructura Multidimensional.	2
1.2	Orientado hacia la información relevante de la Organización. . .	5
1.3	Datos Integrados.	5
1.4	Variable en el Tiempo.	6
1.5	No Volátil.	7
1.6	Componentes de un Datawarehouse.	9
1.7	Funcionamiento de un Data Mart Junto con un Data Warehouse. .	13
1.8	Operación Slice and Dice	17
1.9	Vista de los Datos con Planillas de Cálculo.	18
1.10	Vista de los Datos en forma de Cubo.	19
1.11	Rebanadas de Datos de un Cubo.	19
1.12	Ventajas y Desventajas del Modelo MOLAP.	20
1.13	Funcionamiento del Modelos MOLAP.	21
1.14	Ventajas y Desventajas del Modelo ROLAP.	22
1.15	Comparativa del Funcionamiento de los Modelos MOLAP y ROLAP.	24
2.1	Operador Roll Across.	31
2.2	Operador Slice and Dice.	32
2.3	Operador Pivot.	33
2.4	Multidimensionalidad.	36
2.5	Esquema Estrella.	37
2.6	Tabla de Hechos.	39
2.7	Almacenes Formados por 4 Datamarts.	41

Índice de Tablas

1.1	Diferencias entre los Sistemas OLTP y OLAP.	14
-----	---	----

Capítulo 1

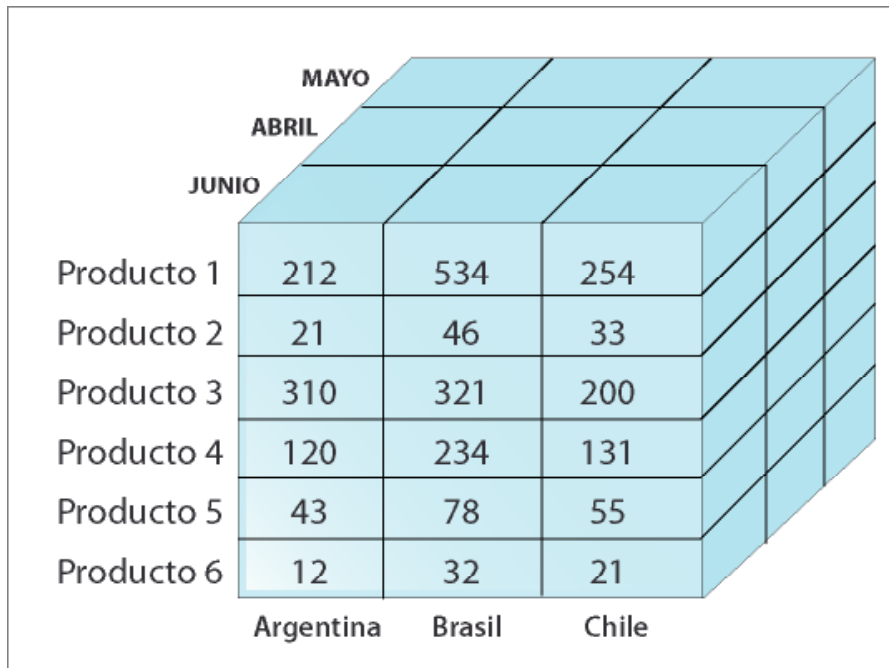
Introducción

1.1 Las Bases de Datos

Imagínese lo difícil que sería obtener cualquier información en un sistema de información si los datos estuviesen almacenados en forma desorganizada, o si no existiese una forma sistemática para recuperarlos. Por lo tanto en todo sistema de información, los recursos de datos deben organizarse y estructurarse en cierta manera lógica de manera que puedan accederse con facilidad, procesarse de manera eficiente, recuperarse con rapidez y manejarse de forma efectiva. Todo esto es hoy en día solucionado gracias a las bases de datos.

Las diferentes estructuras que puede tener una base de datos son la siguientes:

- Jerárquica.
- Red.
- Relacional.
- Multidimensional.
- Orientada a Objetos.



	MAYO	ABRIL	JUNIO
Producto 1	212	534	254
Producto 2	21	46	33
Producto 3	310	321	200
Producto 4	120	234	131
Producto 5	43	78	55
Producto 6	12	32	21
	Argentina	Brasil	Chile

Figura 1.1: Estructura Multidimensional.

1.1.1 Estructura Multidimensional

La estructura multidimensional de bases de datos es una variación del modelo relacional, que utiliza estructuras multidimensionales para organizar datos y expresar las relaciones entre los datos. Se pueden visualizar las estructuras multidimensionales como cubos de datos y cubos dentro de cubos de datos. Cada cara del cubo se considera una dimensión de los datos.

Cada celda dentro de una estructura multidimensional contiene datos agregados que se relacionan con los elementos a lo largo de cada una de sus dimensiones. Por ejemplo una sola celda puede contener las ventas totales de un producto en una región para un canal de ventas específico en un solo mes, como se muestra en la siguiente Figura 1.1 de la Página 2 .

Uno de sus principales beneficios es que constituye una manera compacta y fácil de comprender, para visualizar y manipular elementos de datos que tienen muchas interrelaciones.

Por lo tanto estas se han convertido en la estructura de bases de datos más popular para las bases de datos analíticas que respaldan aplicaciones del *procesamiento analítico en línea (OLAP)* en el cual se esperan respuestas más rápidas a consultas empresariales complejas.

Ya hace un tiempo las organizaciones han usado los datos desde sus sistemas operacionales para atender sus necesidades de información. Algunas proporcionan acceso directo a la información contenida dentro de las aplicaciones operacionales. Otras, han extraído los datos desde sus bases de datos operacionales para combinarlos de varias formas no estructuradas, en su intento por atender a los usuarios en sus necesidades de información.

A través del tiempo estos métodos han mejorado y ahora las organizaciones manejan una datos sobre los cuales la mayoría de las veces se toman decisiones importantes.

1.2 Data Warehouse

Un *data warehouse* o *depósito de datos* es una colección de datos orientado a temas, integrado, no volátil, de tiempo variante, que se usa para el soporte del proceso de toma de decisiones gerenciales.

El data warehouse a adquirido actualmente una gran importancia dentro de las grandes instituciones, debido a que provee un ambiente para que las organizaciones hagan un mejor uso de la información que está siendo administrada por diversas aplicaciones operacionales.

Reunir los elementos de datos apropiados desde diversas fuentes de aplicación en un ambiente integral centralizado, simplifica el problema de acceso a la información y en consecuencia, acelera el proceso de análisis, consultas y el menor tiempo de uso de la información.

Las aplicaciones para soporte de decisiones basadas en un data warehousing, pueden hacer más práctica y fácil la explotación de datos para una mayor eficacia del negocio, que no se logra cuando se usan sólo los datos que provienen de las aplicaciones operacionales (que ayudan en la operación de la empresa en sus operaciones cotidianas), en los que la información se obtiene realizando procesos independientes y muchas veces complejos.

Un data warehouse se crea al extraer datos desde una o más bases de

datos de aplicaciones operacionales. Los datos extraídos se transforman para eliminar inconsistencias y resumir si es necesario y luego son cargados en el data warehouse.

El proceso de transformar, crear el detalle de tiempo variante, resumir y combinar los extractos de datos, ayudan a crear el ambiente para el acceso a la información Institucional. Este nuevo enfoque ayuda a las personas individuales, en todos los niveles de la empresa, a efectuar su toma de decisiones con más responsabilidad.

El data warehouse soporta el procesamiento informático al proveer una plataforma sólida, a partir de los datos históricos para hacer el análisis. Facilita la integración de sistemas de aplicación no integrados. Organiza y almacena los datos que se necesitan para el procesamiento analítico informático sobre una amplia perspectiva de tiempo.

Los datos que ingresan en el data warehouse provienen del ambiente operacional en la mayoría de los casos. El data warehouse es siempre un almacén de datos que se halla transformado y separado físicamente de la aplicación donde se encontraron los datos en el ambiente operacional.

1.2.1 Características del Data Warehouse

Sus principales características son las siguientes:

Orientado hacia la información relevante de la organización: se diseña para consultar eficientemente información relativa a las actividades básicas de la organización, como ser compras ventas, producción, etc y no para soportar los procesos que se realizan en la organización, como ser gestión de pedidos, facturación, etc. En la Figura 1.2 de la Página 5 se ilustra esta característica.

Datos Integrados: integra datos recolectados de diferentes sistemas operacionales de la organización y/o fuentes externas. En la Figura 1.3 de la Página 5 se puede observar a través de una ilustración esta característica, además se puede observar que cuando se habla de fuentes externas también se considera una gran variedad de formatos como ser archivos txt o html.

Variable en el tiempo: los datos son relativos a un período de tiempo y estos deben ser integrados periódicamente. Estos datos son almacenados como fotos que se corresponden a períodos de tiempo. En la Figura 1.4 de la Página 6 se puede observar a través de una ilustración esta característica.

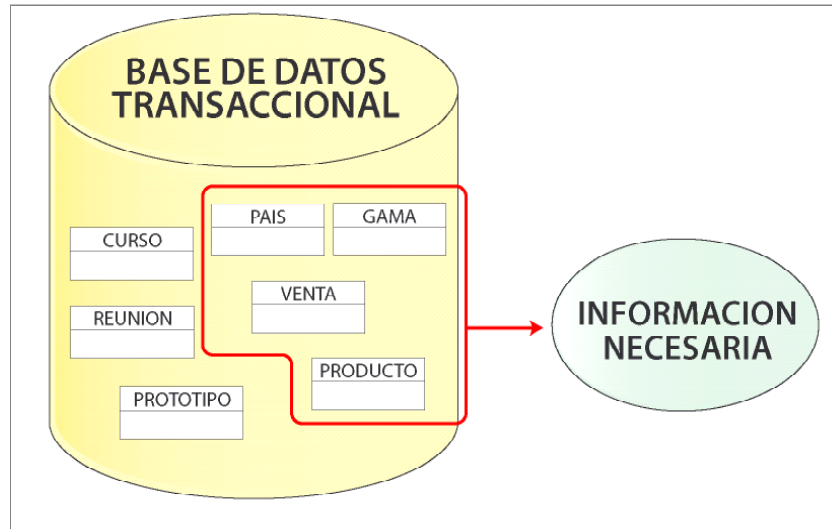


Figura 1.2: Orientado hacia la información relevante de la Organización.

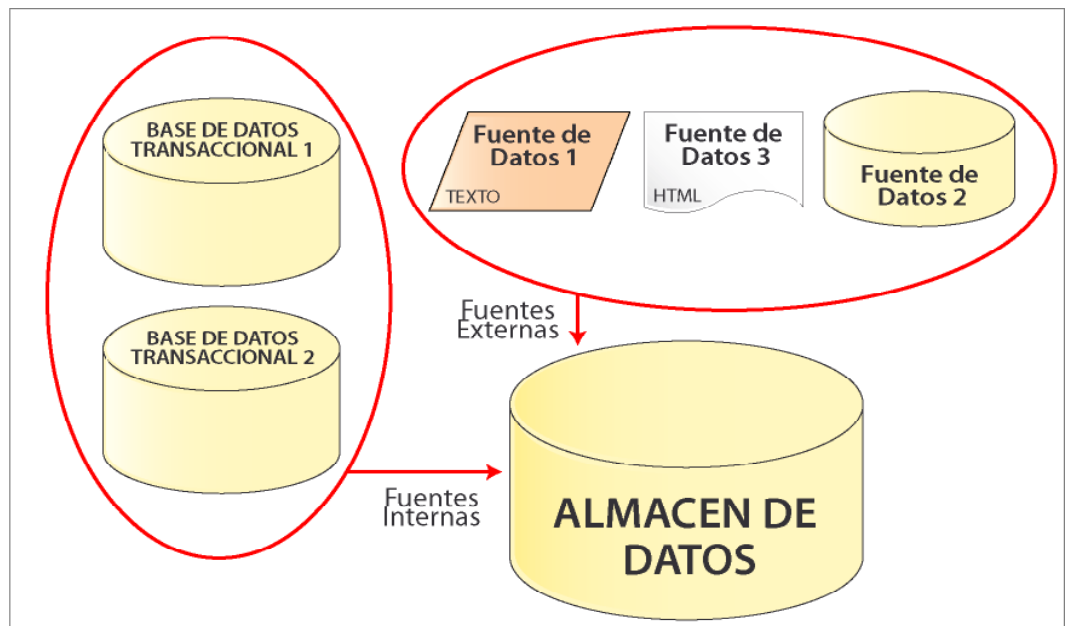


Figura 1.3: Datos Integrados.

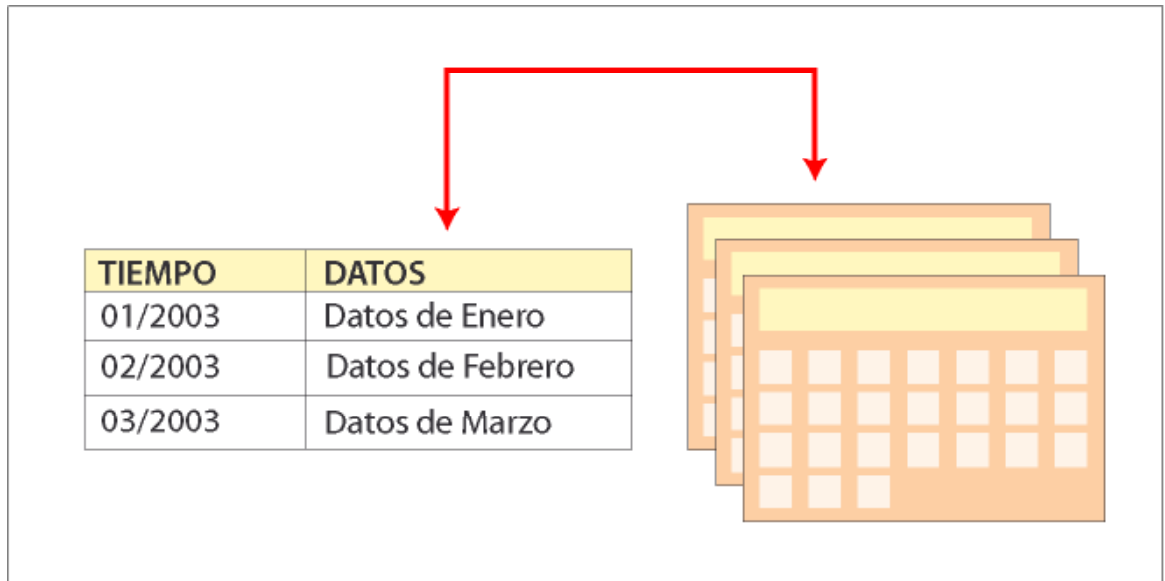


Figura 1.4: Variable en el Tiempo.

No volátil: los datos que son almacenados no sufren de ninguna actualización solo son incrementados. El período de tiempo cubierto por un data warehouse va de 2 a 10 años. En la Figura 1.5 de la Página 7 se puede observar esta característica en relación a las bases de datos operacionales.

1.2.2 Principales Ventajas y Desventajas de un DW para las Organizaciones

Su principal *ventaja* es que presenta una rentabilidad de las inversiones realizadas para su creación, además de producir un aumento de la competitividad en el mercado y favorecer a la productividad de los técnicos de dirección.

Pero a pesar de sus considerables ventajas también se encuentran sus puntos en contra como ser, la infravaloración del esfuerzo necesario para su diseño y creación, además de los recursos necesarios para la captura, carga y almacenamiento de los datos. También se debe considerar el aumento constante de los requisitos de los usuarios y la privacidad de los datos.

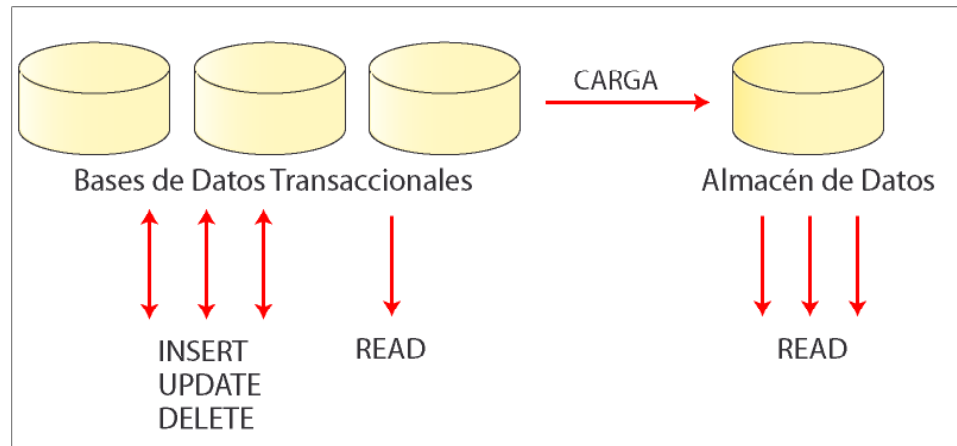


Figura 1.5: No Volátil.

1.2.3 Requerimientos Fundamentales de un Data Warehouse

La infraestructura tecnológica que soporte el Data Warehouse se encuentra separada y, por lo general, será diferente de la que soporta los sistemas operacionales.

En definitiva, los requerimientos fundamentales que determinan la esencia del DW podrían ser los siguientes:

- Acceso universal a los datos, que se crean y se tratan de acuerdo con los requerimientos del usuario y dentro del marco de un modelo de datos.
- Implantación de un sistema abierto, con interfaces a fuentes de datos internas y externas.
- Selección de los datos de acuerdo con el contenido de información y la relevancia para las decisiones.
- Separación lógica y física de las bases de datos de almacén de datos y de metadatos respecto a las bases de datos operacionales.
- Creación de herramientas de consulta para el usuario final, con posibilidades de utilización intuitiva y funciones de unión e interrelación.

1.2.4 Componentes de un Data Warehouse

Los componentes de un data warehouse son los siguientes:

Fuentes de datos: este componente es el que normalmente está presente originariamente en las organizaciones, y a partir del cual se realiza la captura de datos que se contemplará en el DW. Estas fuentes de datos pueden ser sistemas operacionales corporativos (representan el entorno del que se obtienen la mayor parte de los datos significativos de la operativa diaria de la compañía), sistemas operacionales departamentales, fuentes externas, etc.

Extracción y transformación: es responsable de que la información pueda moverse, con las transformaciones que sean necesarias, desde las fuentes de datos antes mencionada, al DW.

Servidor de datos: también podría denominarse componente de gestión. Los servicios que debe ofrecer incluyen un servicio de mantenimiento de datos y un servicio de distribución para exportar datos del DW a servidores de bases de datos descentralizadas y a otros sistemas de soporte de decisiones de usuario. El componente de gestión también ofrece servicios de seguridad (archivo, backup, recuperación) y monitorización. Generalmente estos servicios utilizan los medios suministrados por el software del sistema operativo y de bases de datos subyacente. El componente de SGBD (Sistema de Gestión de Bases de Datos) consiste en el software de base de datos que se utilice para mantener y extraer datos. Hay dos enfoques diferentes para el almacenamiento de la información: las bases de datos relacionales y las multidimensionales. Así, tendremos gestores de bases de datos relacionales (SGBDR) o gestores de bases de datos multidimensionales (SGBDM).

Herramientas de acceso: sin las herramientas adecuadas de acceso y análisis el DW se puede convertir en una amalgama de datos sin ninguna utilidad. Es necesario poseer técnicas que capturen los datos importantes de manera rápida y puedan ser analizados desde diferentes puntos de vista. También deben transformar los datos capturados en información útil para el negocio. Actualmente a este tipo de herramientas se las conocen como business intelligence tool (BIT) y están situadas conceptualmente sobre el DW. Cada usuario final debe seleccionar que herramienta se ajusta mejor a sus necesidades y a su DW. Entre ellas podemos citar las Consultas SQL (Structured Query Language), las Herramientas MDA (Multidimensional Analysis) y OLAP (On-Line Analytical Processing), las Herramientas ROLAP (Relational On Line Analytical Processing) y las herramientas DATA MINIG, de las cuales se trata a

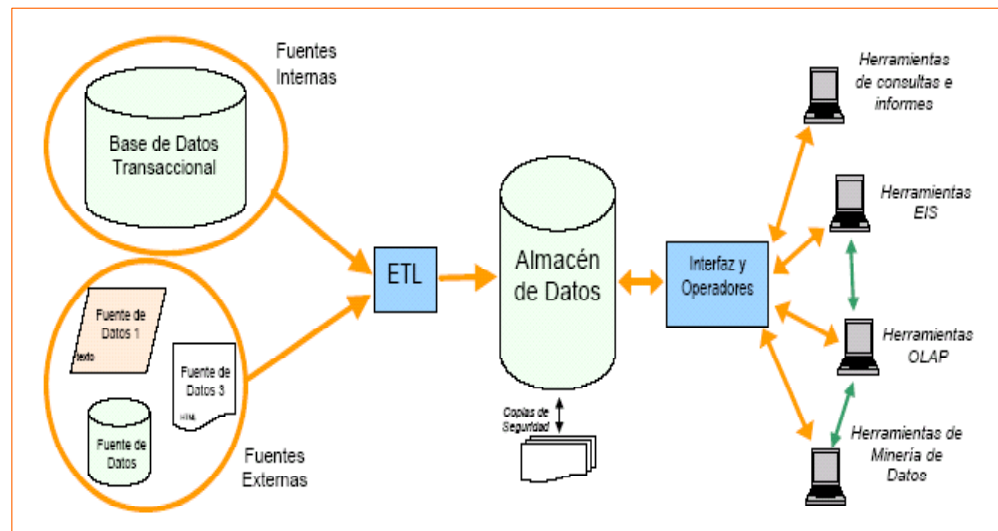


Figura 1.6: Componentes de un Datawarehouse.

continuación.

Repositorio/Metadatos: los metadatos son básicamente datos acerca de los datos contenidos en el DW. Así, uno de los problemas con el que pueden encontrarse los usuarios de un DW es saber lo que hay en él y cómo pueden acceder a lo que quieren. El repositorio les ayuda a conseguirlo. Es sólo una de las utilidades del repositorio, pero éste tiene muchas funcionalidades: catalogar y describir la información disponible; especificar el propósito de la misma; indicar las relaciones entre los distintos datos; establecer quién es el propietario de la información; relacionar las estructuras técnicas de datos con la información de negocio; establecer las relaciones con los datos operacionales y las reglas de transformación; y limitar la validez de la información.

En la siguiente Figura 1.6 de la Página 9 se puede observar la integración de los componentes de un DW.

1.2.5 Factores de Exito de un Data Warehouse

Los principales factores que llevan al éxito al data warehouse son los siguientes:

- Integra datos de producción con datos externos y gestiona historiales.
- Contiene datos útiles..
- Los datos son coherentes, actualizados y documentados (calidad).
- Ofrece acceso directo a los usuarios.
- Aumenta el número de accesos.
- Otorga una flexibilidad que apoya al crecimiento de los usuarios, herramientas así como también al volumen de los datos.

1.2.6 Problemas en la implementación

- Alto coste.
- Inmediata ayuda para la toma de decisiones (Data Mart).
- Alto coste de mantenimiento debido a cambios de necesidades, nuevas fuentes de datos, cambio de la capacidad o de la tecnología.
- Control de calidad de los datos.
- Heterogeneidad e integración de datos.

1.2.7 Caso del Mundo Real

Vtel Corporation: Integración de Aplicaciones empresariales con una Bodega de Datos (DW).

Vtel Corporation, con sede en Austin, Texas elabora sistemas digitales de conferencia multimedia para internet, intranets y extranets corporativas, Vtel se propone convertir su actual conjunto de sistemas empresariales a Oracle Applications, un sistema de software empresarial integrado de Oracle, que incluye una bodega de datos. El movimiento fue motivado por una fusión que tubo lugar el año pasado, de Vtel y Compression Labs, Inc., en San José California, y por que ambas empresas crecieron excesivamente en mucho de sus sistemas operacionales.

“En 1989 estábamos vendiendo 2 millones de dólares anuales; en la actualidad estamos vendiendo 200 millones de dólares anuales” explica Rodney Bond,

presidente ejecutivo financiero de Vtel. “Estamos considerando el crecimiento masivo en el futuro, y este sistema es fundamental para lograr tal cosa.”

Una vez que Vtel decidió operar con Oracle Applications, la empresa desistió de sus planes para mejorar su actual bodega, que ocupa 10 G bytes de una base de datos Oracle residente en un servidor SPARC 20 Sun Microsystems. En cambio, Vtel utilizará la bodega que con su nueva serie de aplicaciones integradas.

“Nuestra intención es aprovechar las lecciones que recibimos de la actual bodega de datos a medida que avanzamos”, dice Steve Cox, vicepresidente y CIO de Vtel. La lección más importante dice Cox es el valor de una bodega y las herramientas analíticas para el negocio. “La bodega nos ha ayudado a comprender nuestro negocio. Nos ha permitido analizar la rentabilidad del mismo”, dice él. Bond explica: “El director regional de ventas puede ingresar a la bodega diariamente y observar las ventas que se hicieron en determinadas áreas en lugar de esperar a que se genere un informe de finanzas”.

Según Cox y Bond, aun cuando la actual bodega de datos reside en una base de datos relacional de Oracle, Vtel la está dejando de lado para evitar tener que mantener dos sistemas. Además la nueva bodega incluirá más tiempo de información. Agregará datos sobre servicio al cliente de operaciones internacionales, por ejemplo, a ítem actuales como datos de venta, de marketing y financieros.

La nueva bodega de datos comprende rutinas de software que extraen y transforman datos provenientes de la base de datos relacional de transacciones de Oracle de Vtel y los trasladan a la bodega, que es una base de datos multidimensional manejada por Oracle Express Server. Los usuarios utilizarán Discover, una herramienta de consulta y elaboración de informes, lo mismo que otras herramientas de la línea de productos Oracle Express, para llegar a los datos.

Cox añade que él espera que a los usuarios empresariales les gusten las nuevas herramientas de elaboración de informes, más que las herramientas hechas sobre pedido que utilizan en la actualidad. Las herramientas Oracle les permitirá estimar los datos organizados en muchas formas, como diagramas de torta y gráficos de barra. La actual herramienta presenta la información solo en cuadrículas.

1.3 Data Mart

Históricamente, los datos de la corporación suelen residir en Bases de datos (BD) que se diseñaron principalmente para introducir y almacenar datos, mediante el llamado Proceso de Transacciones On Line (OLTP). Este método es idóneo para insertar, modificar o borrar registros, pero no lo es tanto para responder a complejas consultas. La relación entre los datos responde, cuando existe, a unas técnicas llamadas de Entidad-Relación (modelo-Relacional).

Los Data Warehouse surgen precisamente en respuesta a los problemas asociados a realizar análisis de datos sobre Bases de Datos del tipo OLTP. La solución propuesta por el Data Warehouse es extraer los datos de una (o más) bases operacionales y moverlos a una Base de Datos independiente y orientada a las consultas.

Pero el problema surge cuando los data warehouse crecen y se tornan más complejos, debido a esto el rendimiento de las consultas decae y el modelo centralizado deja de ser óptimo. En estos casos, la solución propuesta es crear unos almacenes de datos especializados por áreas como Ventas o Compras, que reciben los datos desde el almacén centralizado (DW) y que pueden residir en diferentes máquinas, BD, redes, etc..Estos almacenes se conocen como Data Marts (DM).

Dado que un Data Mart soporta menos usuarios que un Data Warehouse se puede optimizar para recuperar más rápidamente los datos que necesitan los usuarios.

Existen dos tipo de data mart:

- *Data mart dependiente*: los datos son extraídos de data warehouse.
- *Data mart independiente*: los datos son extraídos de los sistemas operacionales.

En la Figura 1.7 de la Página 13 se puede observar el funcionamiento del data warehouse y el data mart.

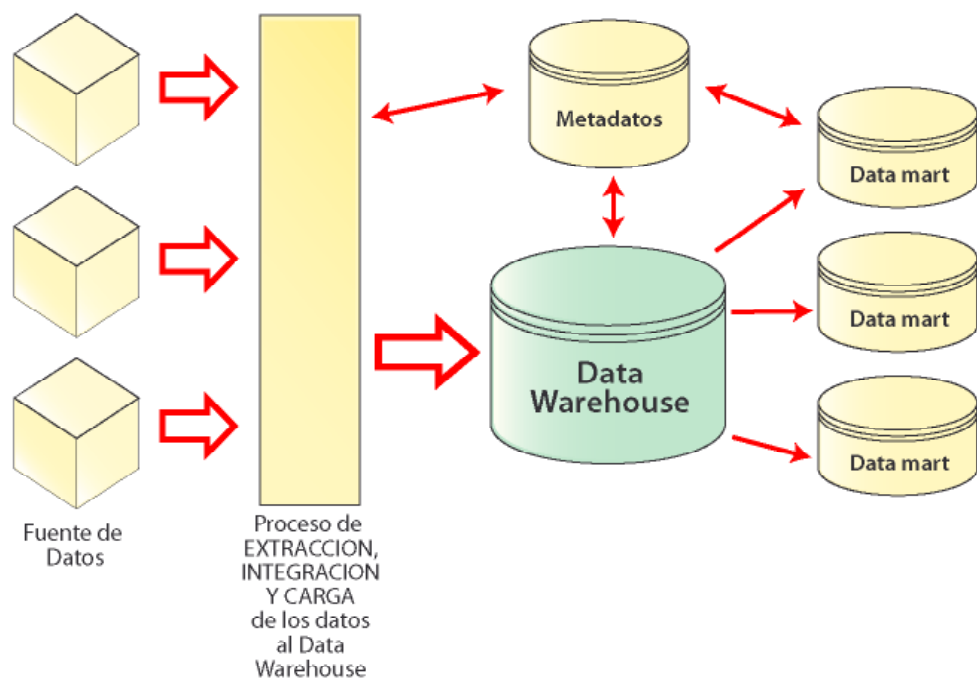


Figura 1.7: Funcionamiento de un Data Mart Junto con un Data Warehouse.

1.4 Sistemas OLTP

Son los sistemas conocidos como On-Line Transactional Processing. Estos procesan las transacciones de tiempo real de un negocio. Contienen estructuras de datos optimizadas para la introducción y a la adición de los datos.

Su principal desventaja es que proporciona capacidades muy limitas para la toma de decisiones.

Se podría decir que estos sistemas definen el comportamiento operacional de un entorno operacional de gestión:

- Altas, bajas, modificaciones, consultas.
- Consultas rápidas y encuestas.
- Poco volumen de información.
- Transacciones rápidas.
- Gran volumen de concurrencia.

1.4.1 Diferencias entre OLTP y OLAP

Las principales diferencias existentes con los sistemas OLAP se pueden observar en la fig. 1.1 de la pag. 14.

Características	OLTP	OLAP
Tamaño BBDD	Giga Bytes	Giga Bytes a Tera Bytes
Origen de Datos	Interno	Interno y Externo
Actualización	Actual	hISTÓRICO
Consultas	Predecible	Ad Hoc
Atividad	Operacional	Analítica

Tabla 1.1: Diferencias entre los Sistemas OLTP y OLAP.

Las siguientes divergencias provocan que sea imposible la convivencia en una misma BBDD de los entornos OLAP y OLTP:

- Pérdida de rendimiento del entorno OLTP.
- Falta de integración entre distintas aplicaciones OLTP.
- Tecnologías de BBDD sin capacidad para soportar aplicaciones OLTP.
- Incorporación de datos externos difícilmente aplicables a la BBDD OLTP.
- Distribución de datos no adecuada para OLTP.

1.4.2 Diferencias entre OLTP y Data Warehouse

Las principales diferencias existentes entre los sistemas OLTP y Data Warehouse son las siguientes:

- Las aplicaciones OLTP están organizadas para ejecutar las transacciones para las cuales fueron hechas; en cambio el data warehouse está organizado en base a conceptos.
- El número de usuarios de un Data Warehouse en general es menor que el existente en un sistema OLTP.
- Los sistemas OLTP realizan cientos de transacciones por segundos, mientras que una sola consulta en una Data Warehouse podría llevarle varios minutos.
- El diseño de un OLTP es considerablemente normalizado mientras que un DW tiende a ser desnormalizado.
- Por lo general los OLTP son de menor tamaño que los DW, debido a que estos últimos están formados por información de varios OLTPs.
- El OLTP normalmente está formado por un número mayor de tablas, con pocas columnas; mientras que en un DW el número de tablas es menor, pero cada una de éstas tiende a ser mayor en número de columnas.
- Los OLTP son continuamente actualizados por los sistemas operacionales del día, mientras que los DW son actualizados en batch de manera periódica.
- Las estructuras de los OLTP son muy estables, rara vez cambian, mientras las de los DW sufren cambios constantes derivados de su evolución.

1.5 Sistemas Olap

La naturaleza competitiva y dinámica del entorno empresarial de hoy está impulsando la demanda de sistemas de información por parte de gerentes y analistas empresariales, que puedan proporcionar respuestas rápidas a consultas empresariales complejas. La industria de SI respondió a estas demandas con desarrollos como las bases de datos analíticas, las plazas de datos, las bodegas de datos, las técnicas de explotación de datos, las estructuras de bases de datos multidimensionales y con servidores especializados y productos de software que respaldan el procesamiento analítico en línea (OLAP).

Se pueden considerar los sistemas OLAP (On Line Analytical Processing) como pertenecientes a los sistemas de información para ejecutivos, EIS, utilizados para proporcionar al nivel estratégico información útil para la toma de decisiones.

Es un método para buscar en los datos de diferentes maneras. Con OLAP los datos son clasificados en diferentes dimensiones las que pueden ser vistas unas con otras en cualquier combinación para obtener diferentes análisis de los datos que contienen.

En un modelo de datos OLAP, la información es vista como cubos, los cuales consisten de categorías descriptivas (dimensiones) y valores cuantitativos (medidas). El modelo de datos multidimensional simplifica a los usuarios formular consultas complejas, arreglar datos en un reporte, cambiar de datos resumidos a datos detallados y filtrar o rebanar los datos en subconjuntos significativos.

Por ejemplo, las dimensiones típicas de un cubo que contenga información de ventas incluirían: tiempo, región, producto, canal, organización y escenario (planeado o real). Las medidas típicas incluirían: ventas en dólares (u otra moneda), unidades vendidas, número de personas, ingresos y gastos.

OLAP comprende varias operaciones analíticas básicas, incluidas la consolidación, “drill-down” y “slicing and dicing”:

- *Consolidación*: ésta comprende el conjunto de datos. Esto puede involucrar acumulaciones simples o agrupaciones complejas que incluyen datos interrelacionados. Por ejemplo, las oficinas de ventas pueden agruparse respecto a los distritos, y los distritos respecto a las regiones.

The diagram shows a 'Slice and Dice' operation on a sales data cube. The original data is shown in a table with two time periods (Q1 and Q2) and two stores (Store 1 and Store 2). The operation slices the data to focus on Q2 and Store 1, resulting in a smaller table.

VENTAS			
	Productos	Store 1	Store 2
Q1	Electronics	\$5,20	\$5,60
	Toys	\$1,90	\$1,40
	Clothing	\$2,30	\$2,60
Q2	Cosmetics	\$1,10	\$1,10
	Electronics	\$8,90	\$7,20
	Toys	\$0,75	\$0,40
Q2	Clothing	\$4,60	\$4,60
	Cosmetics	\$1,50	\$0,50

VENTAS		
	Productos	Store 1
Q2 Q1	Electronics	\$5,20
	Toys	\$1,90
Q2 Q2	Electronics	\$8,90
	Toys	\$0,75

SLICE y DICE

Figura 1.8: Operación Slice and Dice

- *Drill-down*: OLAP puede moverse en la dirección contraria y presentar automáticamente datos detallados que abarcan datos consolidados. Esto recibe el nombre de drill-down. Por ejemplo, podría accederse fácilmente las ventas por productos individuales o representantes de ventas que conforman las ventas totales de una región.
- *Slicing and dicing*: se refiere a la capacidad de visualizar a las bases de datos desde diferentes puntos de vista. Una parte de la base de datos de ventas podría mostrar todas las ventas del tipo de producto dentro de las regiones. Otra parte podría mostrar las ventas por canal de ventas dentro de cada tipo de producto. El slicing and dicing con frecuencia se realiza a lo largo de un eje de tiempo con el fin de analizar tendencias y encontrar patrones. Esta operación puede ser observada en la Figura 1.8 de la Página 17.

1.5.1 Vista de los Datos

La vista de los datos como un cubo es una extensión natural de como la mayoría de los usuarios de negocios interactúan con los datos.

Ellos ven a un problema de negocios en términos de un cierto número de componentes (dimensiones) tales como productos, tiempo, regiones, fabricantes, o artículos. Los usuarios de negocios desean poder analizar un conjunto de

		REGION		
		Argentina	Brasil	Chile
PRODUCTOS	Producto 1	212	534	254
	Producto 2	21	46	33
	Producto 3	310	321	200
	Producto 4	120	234	131
	Producto 5	43	78	55
	Producto 6	12	32	21

Figura 1.9: Vista de los Datos con Planillas de Cálculo.

números usando cualquier par de estos componentes, como así también poder intercambiarlos para lograr distintas vistas.

Un ejemplo podría ser el siguiente: la mayoría de los usuarios desearían ver como se desarrollan las ventas a lo largo del tiempo. Para ello, se necesitarían varias hojas de la planilla de cálculo como se observa en la Figura 1.9 de la Página 18.

Estas mismas celdas de datos pueden ser visualizadas mediante un cubo, como se puede observar en la Figura 1.10 de la Página 19.

Dado que las celdas de datos pueden ser fácilmente representadas en un cubo, se pueden tomar rebanadas del mismo para responder a una diversidad de preguntas. Esta idea es esquematizada en la Figura 1.11 de la Página 19.

1.5.2 Modelos de Almacenamientos

Existen tres modelos de almacenamiento:

- MOLAP.

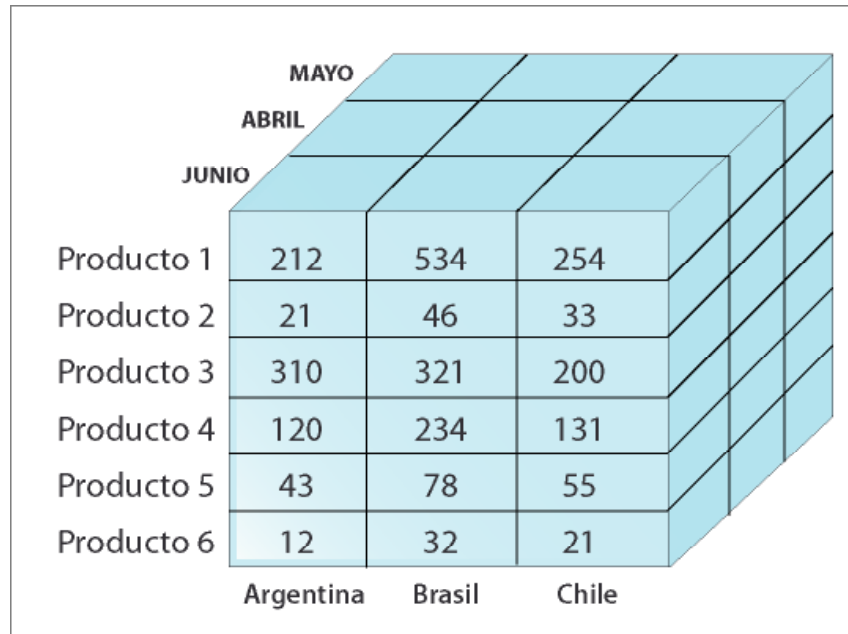


Figura 1.10: Vista de los Datos en forma de Cubo.

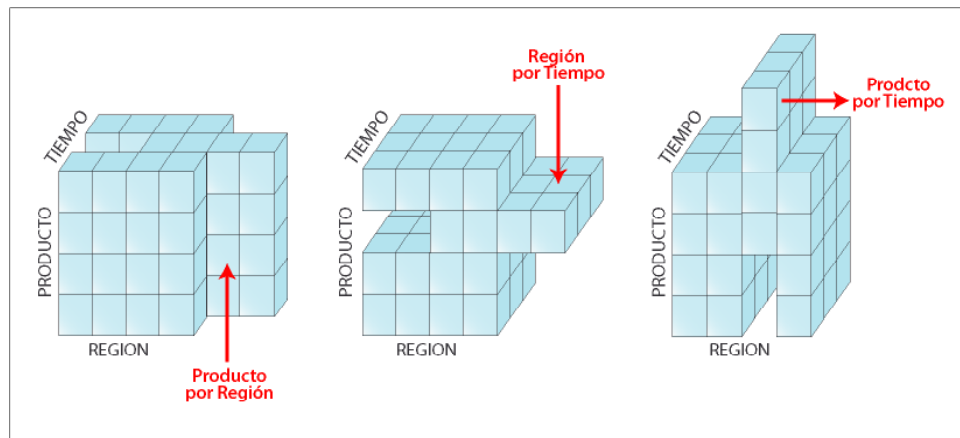


Figura 1.11: Rebanadas de Datos de un Cubo.

VENTAJAS	DESVENTAJAS
❖ Mayor performance en el procesamiento de queries.	❖ Tamaño limitado para la arquitectura del cubo.
❖ Poco tiempo de cálculo realizado en el momento.	❖ No puede acceder a datos que no están en el cubo.
❖ Puede escribir sobre la base de datos.	❖ No puede explotar el paralelismo en de las Bases de Datos.
❖ Posibilita hacer cálculos más complicados.	

Figura 1.12: Ventajas y Desventajas del Modelo MOLAP.

- HOLAP.
- ROLAP.

MOLAP (OLAP multidimensional)

En estos sistemas los datos se encuentran almacenados en una estructura multidimensional.

Para optimizar los tiempos de respuesta, el resumen de la información es usualmente calculado por adelantado. Estos valores precalculados o agregaciones son la base de las ganancias de desempeño de este sistema. Algunos sistemas utilizan técnicas de compresión de datos para disminuir el espacio de almacenamiento en disco debido a los valores precalculados.

Las ventajas y desventajas de este modelo pueden ser apreciadas en la siguiente Figura 1.12 de la Página 20.

En la siguiente Figura 1.13 de la Página 21 se puede observar como funciona este modelo de datos.

ROLAP (OLAP Relacional)

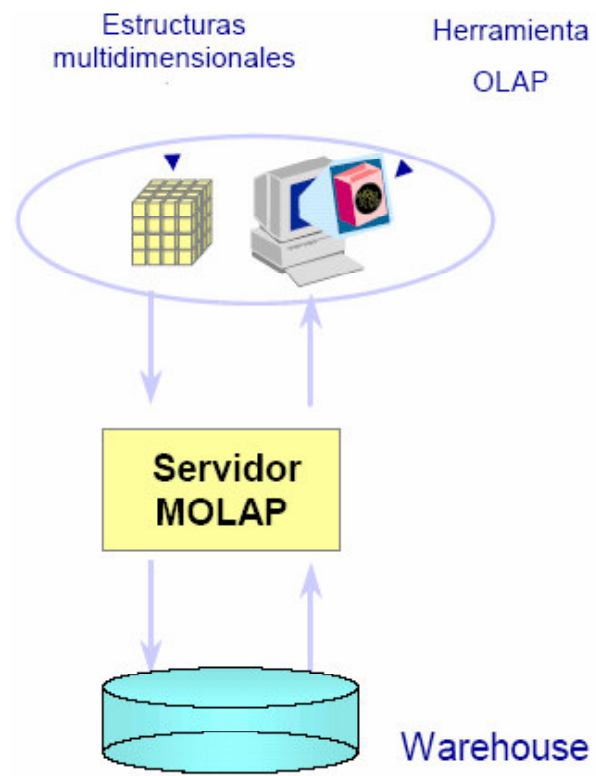


Figura 1.13: Funcionamiento del Modelos MOLAP.

VENTAJAS	DESVANTAJAS
❖ Uso total de la seguridad e integridad de la base de datos .	❖ Queris más lentos.
❖ Escalable para grandes volúmenes.	❖ Construcción cara.
❖ Los datos puede ser compartida con aplicaciones SQL.	❖ Los cálculos están limitados a las funciones de las bases de datos.
❖ Datos y estructura más dinámicos.	

Figura 1.14: Ventajas y Desventajas del Modelo ROLAP.

Son sistemas en los cuales los datos se encuentran almacenados en una base de datos relacional. Típicamente, los datos son detallados, evitando las agregaciones y las tablas se encuentran normalizadas. Los esquemas más comunes sobre los que se trabaja son estrella o copo de nieve, aunque es posible trabajar sobre cualquier base de datos relacional.

Las ventajas y desventajas de este modelo pueden ser observadas en la Figura 1.14 de la Página 22.

HOLAP (OLAP Híbrido)

Estos sistemas mantienen los registros detallados en la base de datos relacional, mientras que los datos resumidos o agregados se almacenan en una base de datos multidimensional separada.

Son conocidos como Híbridos debido a que contiene las características de los sistemas anteriores intentando rescatar lo mejor de cada uno.

COMPARACION

El primer enfoque corresponde a MOLAP. Con este método, los datos son precalculados y luego son almacenados en cubos de datos multidimensionales.

El resultado se traduce en una mejor performance en los tiempos de respuesta debido a que los datos se encuentran disponibles sin necesidad de calcularlos en cada nueva consulta.

La desventaja de este enfoque se debe a que almacenar datos multidimensionalmente ocupa mucho más espacio que con ROLAP.

El segundo enfoque es ROLAP donde los datos son accedidos directamente del Data Warehouse (u otro tipo de fuente relacional) y no son almacenados por separado. Estos datos se calculan en tiempo de consulta (al vuelo). Por lo tanto el tiempo de respuesta sería mayor pero sin usar enormes cantidades de almacenamiento en disco.

Las implementaciones MOLAP normalmente se desempeñan mejor que la tecnología ROLAP, pero tienen problemas de escalabilidad.

Las implementaciones ROLAP son más escalables y son frecuentemente atractivas a los clientes debido a que aprovechan las inversiones en tecnologías de bases de datos relacionales preexistentes.

La mejor solución probablemente esté entre los dos extremos.

En muchos casos se utilizan HOLAP que resultan de una combinación entre ROLAP y MOLAP.

HOLAP mantiene los volúmenes de datos más grandes en la base de datos relacional y las agregaciones en una base de datos MOLAP separada, logrando con esto un balance entre tiempo y espacio.

En la siguiente Figura 1.15 de la Página 24 se puede ver un comparativa de como funcionan los modelos ROLAP y MOLAP.

1.5.3 Utilidades de OLAP

Las utilidades que presentan las aplicaciones OLAP son las siguiente:

- Tienen acceso a grandes cantidades de datos; por ejemplo varios años de datos de ventas en una bodega de datos.
- Analizan las relaciones entre muchos tipos de elementos empresariales, como ventas, productos, regiones y canales.

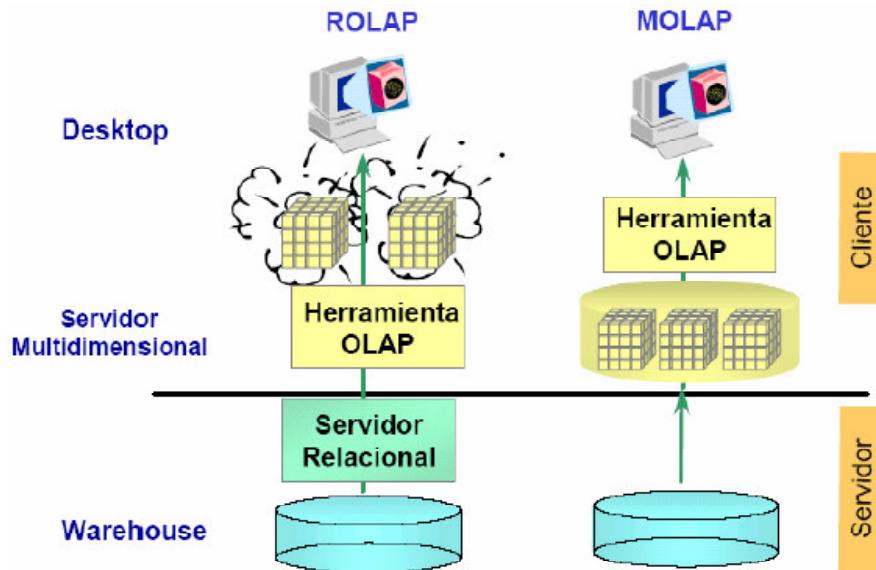


Figura 1.15: Comparativa del Funcionamiento de los Modelos MOLAP y ROLAP.

- Involucran datos agregados; algunos ejemplos son los volúmenes de ventas presupuestados y los dólares que se invierten en una región.
- Comparan datos agregados a través de periodos jerárquicos: mensualmente, trimestralmente, anualmente, etc..
- Presentan los datos en diferentes perspectivas, como ventas por región versus ventas por canal y por producto dentro de cada región.
- Involucran cálculos complejos entre elementos de dato. Por ejemplo; las utilidades esperadas pueden calcularse con una función de ingreso por ventas para cada tipo de canal de venta en una determinada región.
- Puede responder con rapidez a las consultas de los usuarios, de manera que los agentes o analistas puedan seguir un proceso de apreciación analítico o de decisión sin verse impedidos por el sistema

1.5.4 Beneficios que Presenta OLAP

- Es de fácil uso y acceso flexible para el usuario.

- Los datos están organizados en varias dimensiones lo que permite que los usuarios hagan un mejor análisis.
- Ahorro generado por la productividad de personal altamente profesional y caro que usa permanentemente software y sistemas de información.
- Permite encontrar la historia en los datos.
- Genera ciertas ventajas competitivas como ser: aprender más sobre los clientes, permitir usar lo aprendido para aumentar la rentabilidad de la compañía y hacerles a los clientes la vida más fácil además de aprender un poco más sobre los competidores y los proveedores.

1.5.5 Casos del Mundo Real

OLAP en MaterCard International

MasterCard International desarrolló el software OLAP denominado Market Advisor, que permite a los miembros consultar una bodega de datos e indagar información para analizar transacciones y tendencias en línea. El Market Advisor también suministra una base de datos histórica de 13 meses, elaboración de gráficos ampliados de informes y avisos asociados de marketing con base en la actividad, por debajo o por encima del promedio, del titular de la tarjeta o del comerciante.

En una aplicación típica, los analistas de marketing pueden examinar una tendencia en el gasto a niveles agregados para una determinada categoría de comerciantes, como ferretería restaurante, agencia de alquiler de automóviles o gasolinera. Mediante el uso del Market Advisor, los analistas pueden determinar que estados o provincias explican el volumen y pueden identificar que comerciantes constituyen el mayor volumen. Un analista puede incluso indagar los datos para encontrar cuáles cuentas de los titulares de tarjetas se utilizaron en una determinada tienda durante un periodo. Posteriormente, el analista puede encontrar patrones de gastos comunes entre ciertas categorías de titulares de tarjetas, y ajustar las promociones de marketing en forma apropiada.

Capítulo 2

Introducción a OLAP

2.1 Introducción

La información nos permite reducir la incertidumbre (sobre algún aspecto en particular) permitiéndonos tomar mejores decisiones. En un principio los sistemas de información tenían como objetivo recopilar información sobre una porción del mundo para ayudar en la toma de decisiones.

Actualmente con el advenimiento de la informatización de las organizaciones y además la aparición de software operacionales su finalidad paso a ser la de dar soporte a los procesos básicos de la organización: ventas, productos, personal, etc.

Primero se busca satisfacer la necesidad de tener un soporte informático para los procesos básicos de la organización como ser, los sistemas de información para la gestión.

Luego las organizaciones terminan exigiendo nuevas prestaciones para los sistemas de información, como ser los sistemas de información para la toma de decisiones.

En la actualidad han aparecido herramientas de negocios o DSS que co-existen: EIS, OLAP, Minería de Datos, etc. Lo cual nos plantea una serie de interrogantes como ser:

Diferencias entre un EIS y OLAP

Los EIS son sistemas de información junto con un conjunto de herramientas asociadas que permiten:

- Acceso a los directivos a la información de estado y sus actividades de gestión.
- Especializado en análisis del estado diario de la organización (mediante indicadores claves) para informar rápidamente sobre cambios a los directivos.
- La información solicitada por lo general suele ser de naturaleza numérica (ventas semanales, niveles de stock, balances, etc.) con una representación gráfico al estilo hoja de cálculo.

En cambio las herramientas OLAP son genéricas:

- Funcionan sobre un sistema de información como almacenes de datos.
- Permiten realizar agregaciones y combinaciones de los datos de diversas manera con el objetivo de realizar análisis más estratégicos de datos.

Diferencia entre informes avanzados y OLAP

Los sistemas de informes avanzados se caracterizan por lo siguiente:

- Están basados generalmente en sistemas relacionales u objeto-relacionales.
- Utilizan operadores clásicos, como ser concatenación, proyección, selección etc. (tanto en SQL como en extensiones).
- El resultado es presentado de manera tabular..

En cambio las herramientas OLAP:

- No solo utilizan los operadores clásicos si no que además utilizan operadores específicos como por ejemplo Drill, Roll, Pivot, Slice & Dice, entre otros.
- El resultado es expresado en manera matricial o híbrida.

Diferencia esencial entre OLAP y el Data Mining

Como, ciertas cosas en las bases de datos son verdaderas, el usuario plantea una hipótesis (patrón hipotético) y la verifica con una herramienta OLAP. Surgen problemas cuando crece el número de variables.

Mientras que el Data Mining usa los datos para describir patrones. Encuentra patrones en los cuales el usuario no había pensado.

Relación existente entre estas herramientas

Algunas de estas herramientas han hecho cambiar la manera de trabajar de otras. La manera en que trabajan cada una de ellas es la siguiente:

- *Censar el ambiente*: OLTP.
- *Recordar lo censado*: data warehouse y data marts.
- *Razonar*: OLAP.
- *Aprender*: Data mining, KDD.

2.2 Operadores OLAP

Estas herramientas permiten al usuario obtener una visión multidimensional de los datos, para cada actividad que es objeto de análisis.

Con este tipo de herramientas el usuario le formula consultas seleccionando atributos de esta visión o esquema multidimensional sin tener este conocimiento de la estructura interna del almacén de datos. La herramienta OLAP genera la consulta y la envía al gestor de consultas del sistemas por ejemplo esto podría hacerse por medio de una sentencia SELECT.

Una consulta a un almacén de datos consiste básicamente en la obtención de medidas sobre los hechos parametrizadas por los atributos de las dimensiones y limitadas por las condiciones impuestas sobre las dimensiones.

Un ejemplo típico es el caso de un supermercado supongamos que se quisiera saber cuál es el importe de las ventas durante el año 2002 de los productos del departamento bebidas, por trimestre y por categoría. En este caso :

- La medida sería : *importe*.

- El hecho: *ventas*.
- Las restricciones: *productos del departamento bebidas, ventas durante el año 2002*.
- El parámetro de la consulta: *por categoría de producto y por trimestre*.

Lo interesante de las herramientas OLAP son sus operadores de refinamiento o manipulación de consultas:

- Drill.
- Roll.
- Slice & Dice.
- Pivot.

En el agregado de las consultas en el análisis de datos, es aconsejable definir nuevos operadores que faciliten la agregación (consolidación) y la disgregación (división) de los datos:

Disgregación (DRILL): permite introducir un nuevo criterio de agrupación en el análisis, disgregando los grupos actuales. Este operador actúa sobre el informe original, no es necesario realizar un nuevo informe.

Por ejemplo supongamos que se quisiera saber las ventas durante el año 2002 de los productos del departamento bebidas, por trimestre, por categoría y por ciudad de el almacén. En este caso:

- La restricción sería: *producto del departamento bebidas, ventas durante el año 2002*.
- El parámetro para la consulta: *por categoría de producto, por trimestre y por ciudad del almacén*.

Agregación (ROLL): permite eliminar un criterio de agrupación en el análisis, agregando los grupos actuales. Este operador también, actúa sobre el informe original y no requiere que el usuario realice uno nuevo. Por ejemplo si se desea eliminar el criterio de agregación sobre la dimensión tiempo en la

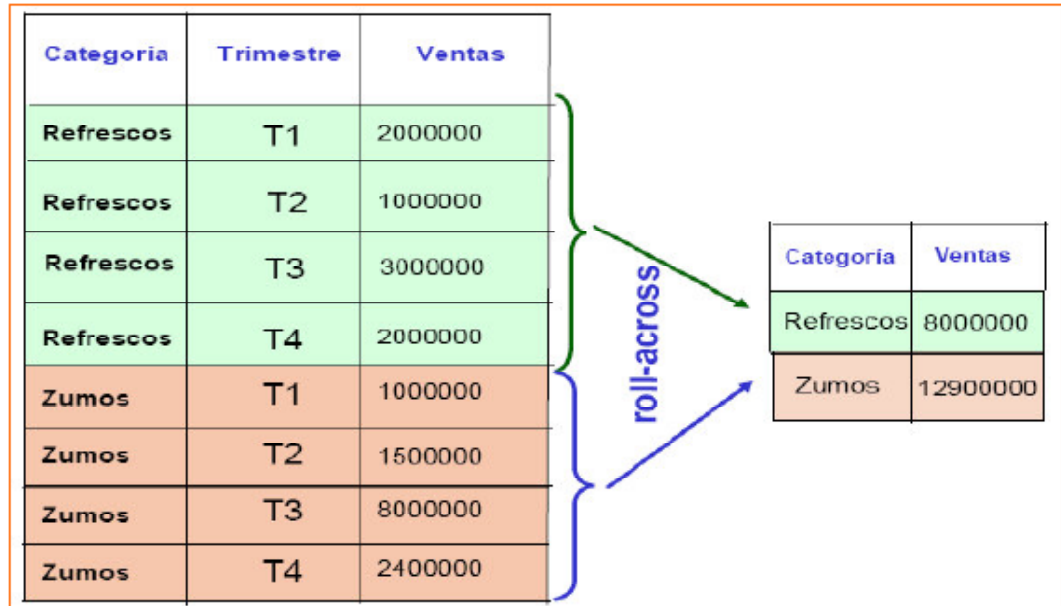


Figura 2.1: Operador Roll Across.

consulta original, como ser saber el importe total de las ventas durante el año 2002 de los productos del departamento bebidas, por categoría.

En la Figura 2.1 de la Página 31 puede observarse el funcionamiento de este operador.

Estas operaciones se pueden hacer sobre:

- Atributos de una dimensión sobre los que se ha definido una jerarquía: Drill-Down, Roll-Up.
- Sobre dimensiones independientes: Drill-Across, Roll-Across.

Otras operaciones de OLAP son las siguientes:

Slice & Dice: este operador permite seleccionar y proyectar datos en el informe. La siguiente Figura 2.2 de la Página 32 ilustra el funcionamiento de este operador.

Pivot: permite la reorientación de las dimensiones en el informe. En la

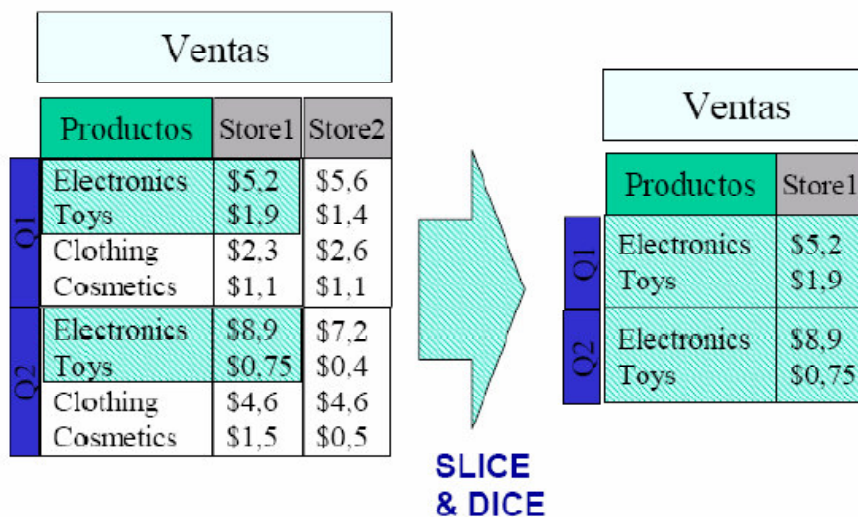


Figura 2.2: Operador Slice and Dice.

Figura 2.3 de la Página 33 se puede observar este operador.

Las herramientas OLAP se caracterizan por lo siguiente:

- Permiten una visión multidimensional de los datos a los usuarios (matricial).
- No impone ningún tipo de restricciones sobre el número de dimensiones.
- Ofrece simetría para las dimensiones.
- No ofrece ningún tipo de limitación a la hora de definir las dimensiones: *restricción, agregación y jerarquía entre ellas.*
- Ofrece operadores intuitivos de manipulación: *drill-down, roll-up, slice & dice, pivot.*
- Es transparente al tipo de tecnología que soporta el almacén de datos (Rolap, Molap).

VENTAS			
	Productos	Store 1	Store 2
Q1	Electronics	\$5,20	\$5,60
	Toys	\$1,90	\$1,40
	Clothing	\$2,30	\$2,60
	Cosmetics	\$1,10	\$1,10
Q2	Electronics	\$8,90	\$7,20
	Toys	\$0,75	\$0,40
	Clothing	\$4,60	\$4,60
	Cosmetics	\$1,50	\$0,50

PIVOT

VENTAS			
	Productos	Q1	Q2
Store 1	Electronics	\$5,20	\$8,90
	Toys	\$1,90	\$0,75
	Clothing	\$2,30	\$4,60
	Cosmetics	\$1,10	\$1,50
Store 2	Electronics	\$5,60	\$7,20
	Toys	\$1,40	\$0,40
	Clothing	\$2,60	\$4,60
	Cosmetics	\$1,10	\$0,50

Figura 2.3: Operador Pivot.

2.3 Tipos de Organizaciones

Como se había comentado en el capítulo anterior los almacenes de datos y las herramientas OLAP pueden basarse físicamente en varias organizaciones:

Sistemas ROLAP: se los implementa sobre tecnología relacional, pero posee facilidades que mejoran el rendimiento. En estos sistemas el almacén de datos se construye sobre SGBD relacional. Los fabricantes de estos productos ofrecen extensiones y herramientas para poder utilizar el SGBD como un sistema gestor de almacenes de datos. Las extensiones de los SGBD son las siguientes:

- Índice de mapas de bits.
- Índice de join.
- Técnicas de particionamiento de los datos.
- Optimizadores de consultas.
- Extensiones de SQL.

Sistemas MOLAP: poseen estructuras de almacenamiento específicas y técnica de compactación de datos que favorecen el rendimiento del almacén.

El objetivo de estos sistemas es almacenar físicamente los datos en estructuras multidimensionales de forma que la representación externa y la representación interna coincidan.

El servidor OLAP construye y almacena datos en estructuras multidimensionales y la herramienta OLAP presenta estas estructuras multidimensionales.

Características principales:

- Los datos son :
 - Arrays.
 - Extraídos del almacén de datos.
- El almacenamiento y los procesos son considerablemente eficientes.
- Se oculta al usuario la complejidad de las bases de datos.
- El análisis se realiza sobre datos agregados y métrica o indicadores precalculados.

Sistemas HOLAP: estos sistemas son considerados híbridos debido a que posee características de los dos sistemas explicados antes.

2.3.1 Ventajas e Inconvenientes de los sistemas MOLAP

Beneficios

- Mayor performance en la ejecución de queries.
- Poco tiempo de cálculos realizados en el momento.
- Puede escribir sobre la base de datos.
- Posibilita hacer cálculos más sofisticados.

Desventajas

- Tamaño limitado por la arquitectura del cubo.
- No puede acceder a datos que no están en cubos.
- Housekeeping/Backups limitados.
- No puede explotar el paralelismo de las bases de datos.

2.3.2 Ventajas e Inconvenientes de los sistemas ROLAP

Beneficios

- Uso total de la seguridad e integridad de la base de datos.
- Escalable para grandes volúmenes.
- La data puede ser compartida con otras aplicaciones SQL.
- Data y estructura más dinámicas.

Desventajas

- Queries más lentos.
- Construcción cara.
- Los índices y sumarios no son mantenidos automáticamente.
- Los cálculos están limitados a las funciones de la base de datos.
- Menos abiertos, ya que tiene clientes propietarios.

2.4 Conceptos

2.4.1 Multidimensionales

Multidimensionalidad es convertir los datos de varias fuentes, tablas relacionales o archivos planos en un estructura donde los datos estén agrupada en dimensiones separadas y heterogéneas , que generalmente son llamadas “cubo”.

Las dimensiones son perspectivas de alto nivel de los datos que representan la información más importante de un negocio. En un banco se tendrán Cuentas, Clientes, Tiempo, Productos, Agencias, Regiones, etc.

En una aplicación OLAP estas dimensiones tienden a no cambiar durante el tiempo.

Cada dimensión tiene componentes que son llamados “miembros”. Por ejemplo el primer trimestre del año es un miembro de la dimensión Tiempo. Cada dimensión puede tener jerarquías entre sus miembros, por ejemplo un

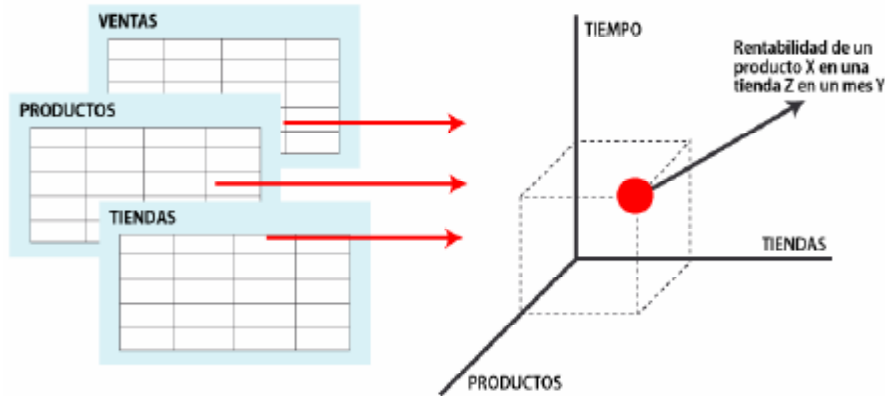


Figura 2.4: Multidimensionalidad.

mes se puede considerar dentro de un trimestre. En la Figura 2.4 de la Página 36 se ejemplifica mediante un dibujo este concepto.

El esquema multidimensional puede ser implementado por un esquema relacional. Dos tipos de esquemas relacionales pueden implementar un esquema multidimensional: esquema en estrella (star schema) y esquema copo de nieve (snowflake schema).

2.4.2 Esquema en Estrella

Un modelo estrella es un modo de representar datos multidimensionales en una base de datos relacional. Las tablas de dimensión guardan información descriptiva acerca de sus miembros y sus relaciones, mientras que las tablas de hechos almacenan datos de negocios.

Los queries multidimensionales son hechos haciendo joins entre las tablas de hechos y de dimensión.

Los sistemas OLAP que utilizan este tipo de modelo de base de datos son llamados sistemas ROLAP.

En la Figura 2.5 de la Página 37 se puede observar lo descrito anteriormente.

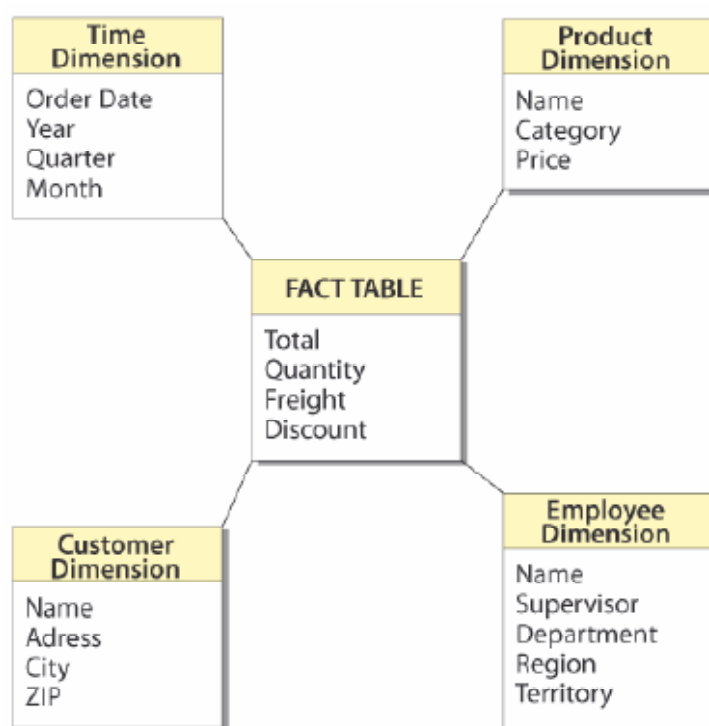


Figura 2.5: Esquema Estrella.

2.4.3 Medidas

Una medida es un tipo de dato cuya información es usada por los analistas (usuarios) en sus queries para medir la performance del comportamiento de un proceso o un objeto del negocio. Por ejemplo: Cantidades, tamaños, montos, duraciones, etc.

Las medidas candidatas son los datos numéricos, pero no cada atributo numérico es una medida candidata.

Son parte de dominio de valor continuos, se deben distinguir las medidas de atributos discretos que son parte de las dimensiones.

Se podría decir que son las que están involucradas en cálculos de resúmenes.

2.4.4 Dimensiones

Una dimensión es una entidad o una colección de entidades relacionadas, usadas por los analistas para identificar el contexto de las medidas con las que trabajan, estas determinan el contexto para las medidas.

El número y tipo de dimensiones para cada medida del modelo debe ser determinada cuidadosamente. El significado de una medida está influenciado por las definiciones de los tipos de medidas que tiene. Cuando se trata de definir la dimensiones, el añadir, eliminar o cambiar propiedades particulares de las dimensiones candidatas cambia el contexto y en consecuencia el significado de la medida candidata.

Las dimensiones son referenciadas por las llamadas “llaves de dimensión”.

Las dimensiones contienen:

- Entidades de dimensión.
- Atributos de dimensión.
- Jerarquías de dimensión.
- Niveles de agregación.

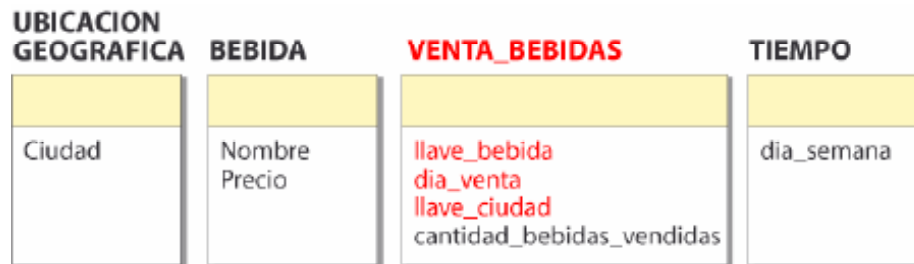


Figura 2.6: Tabla de Hechos.

2.4.5 Hechos

Un hecho es una colección de medidas relacionadas con sus dimensiones relacionadas, representadas por las llaves de dimensión.

Un hecho puede representar un objeto de negocio, una transacción o un evento que es utilizado por el analista de información.

En la Figura 2.6 de la Página 39 se puede observar ejemplos de hechos.

Los hechos contienen:

- Un identificador de hechos.
- Llaves de dimensión, que lo enlaza con las dimensiones.
- Medidas.
- Varios tipos de atributos, los que usualmente se derivan de otros datos en el modelo.

Cada hecho debería tener un equivalente en el mundo real de los negocios. Los hechos relacionados al negocio representan una de las siguientes cosas:

- Objetos del negocio cuyo estado es de interés del analista de información.
- Objetos del negocio cuyos cambios de estado son de interés del analista de información.
- Transacciones o eventos del negocio.

2.4.6 Esquema copo de Nieve

Las bases de datos relacionales a menudo emplean esquemas de copo de nieve para proporcionar los mejores tiempos de respuesta posibles a las consultas complejas. Los esquemas de copos de nieve contienen una tabla de hechos central sin normalizar para el tema y numerosas tablas de dimensión para la información descriptiva sobre las dimensiones del tema. La tabla de hechos puede contener varios millones de filas. La información a la que se tiene acceso con más frecuencia se agrega previamente y se resume para mejorar aún más el rendimiento.

Si bien el esquema de copo de nieve se considera fundamentalmente una herramienta con la que el administrador de bases de datos puede aumentar el rendimiento y simplificar el diseño del almacén de datos, también se utiliza para representar la información del almacén de datos de forma que tenga más sentido para los usuarios finales.

Suele aplicarse cuando muchos atributos caracterizan a los niveles más altos de jerarquías.

Se puede elegir el normalizar sólo algunas dimensiones y otras no, aumentando así la complejidad del diseño y metadatos.

El inconveniente es que al aumentar el número de tablas algunos requisitos pueden demorarse en exceso.

En resumen:

Ventajas:

- No se necesita el atributo de nivel
- Fácil para definir jerarquías
- Mejora considerablemente el rendimiento cuando un gran número de requisitos solicita datos agregados
- Los requerimientos escanean un reducido número de filas

Desventajas:

- Aumenta la complejidad de mantener la meta información debido al aumento del número de tablas

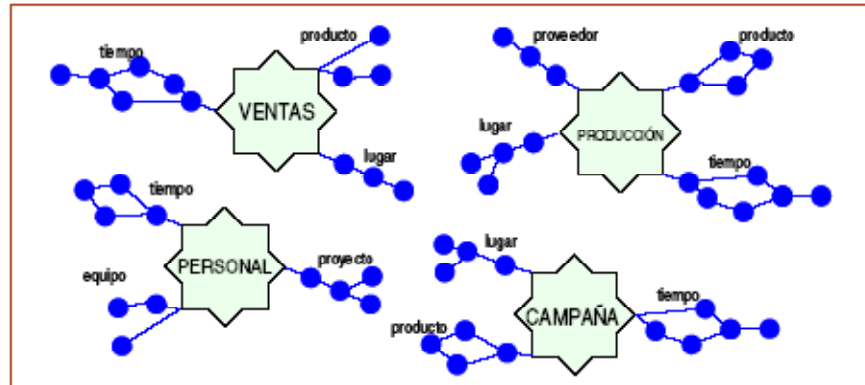


Figura 2.7: Almacenes Formados por 4 Datamarts.

- Si no se dispone de la suficiente cantidad de tablas de agregados, el rendimiento general podría disminuir

No es posible recopilar toda la información necesaria en un único esquema estrella o como de nieve, por ello se hace necesario utilizar varios esquemas. Cada uno de estos esquemas se denomina Datamart. En la Figura 2.7 de la Página 41 se ilustra lo antes mencionado.

2.5 Casos del Mundo Real

Cuando el año pasado fracasaron los planes de fusión de Office Depot con Staples, Inc., analistas del Wall Street como Donald Spindel dijeron que a la compañía le tomaría “varios trimestres recuperar el impulso perdido”. Después de todo, el líder del mercado de suministro de oficina había perdido a muchos de sus altos gerentes durante los debates de la infortunada fusión. También había perdido docenas de empleados de oficina que solían ayudar a generar informes sobre ventas de productos en sus 600 tiendas de Estados Unidos.

Pero gracias a su incremento en el uso de software de procesamiento analítico en línea (OLAP) por parte de sus 200 comercializadores y ejecutivos de finanzas, Office Depot se las arregló para generar un crecimiento “respetable” de 4% en las ventas durante el segundo semestre del año pasado. Al equipar a sus comercializadores con sus propias herramientas analíticas, la empresa con

sede en Delay, Florida, pudo “operar en forma más eficiente, sin las personas que perdió”, dijo Spindel.

Office Depot comenzó a utilizar y Wired para software OLAP a comienzos de 1996. Esto ha permitido centenares de comercializadores, vendedores y ejecutivos iniciar sus propias consultas y “dar un nuevo enfoque a la empresa”, dice Bob Nickerson, director de planeación de mercancía e información en Office Depot,

Por ejemplo, en el otoño de 1996, Nickerson y otros 100 comercializadores de Office Depot utilizaron Essbase para revisar el negocio de los computadores personales del minorista, generando análisis detallados del margen bruto del retorno sobre las inversiones, por tienda y por tipo de producto.

“Encontramos que teníamos demasiadas existencias marginales en las tiendas equivocadas”, dijo Nickerson.

Por tanto, el minorista redujo su surtido de computadores personales de 22 a 112 productos. Esto ayudó a la empresa a eliminar el inventario innecesario y a evitar costosas reducciones de precios en los equipos que se estaban cubriendo de polvo.

A su vez, el retorno sobre activos de Office Depot en su negocio de los computadores “ha mejorado sustancialmente”, dice Nickerson.. Él se reusó a cuantificar esas ganancias específicas, pero dijo que el uso, por parte de la empresa, del Essbase y Wired para herramientas OLAP “se ha traducido en millones de dólares en rentabilidad”. El año pasado, Office Depot ocupó el tercer lugar entre los minoristas de computadores, con 2,590 millones de dólares en ventas, después de las tiendas CompUSA y Best Buy.

Incluso Dilbert, al que Office Depot ha otorgado licencia de uso de su actual campaña publicitaria, pasaría por un momento difícil al referirse únicamente a los retornos que Office Depot ha logrado sobre su inversión inferior a 5 millones de dólares en software Essbase y servidores Compaq 7000. Para allanar el camino se realizó un esfuerzo de cinco años, ya terminado, con el fin de reemplazar unos cuantos computadores AS/400 de rango intermedio de IBM por un mainframe 900-MIPS de IBM que corre con DB2.

Es posible que un desplazamiento hacia el procesamiento en mainframe no suene muy seductor, pero Office Depot acabó con sistemas de buena arquitectura que se desempeñaban bien en el mantenimiento de datos detallados. Esto ayudó a Office Depot a establecer un ambiente limpio de apoyo a las

decisiones que comprende 4 terabytes, principalmente de datos DB2 y 300 gigabytes de información Essbase que se almacena en plazas de datos o “cubos”. Paul Gaffney, vicepresidente senior de desarrollo de sistemas de Office Depot, dice: “Habría sido imposible entregar tales datos si no tuviéramos una buena arquitectura de éstos”.

Pero el éxito no se ha producido fácilmente. Un solo cubo puede contener hasta 30.000 productos y hasta 53 semanas de datos financieros para cada una de sus 600 tiendas, dice Nickerson.. Y cuando los cubos crecen demasiado, los servidores de disco de Office Depot se atascan.

Para deshacer estos cuellos de botella, Office Depot adquirió unidades de disco magnético RAID de múltiples canales de IBM. Esto ayudó a que las plazas de datos generaran respuestas en fracciones de segundos “a cualquier solicitud de indagación” en cubos diseñados previamente , dice Gaffney.

Office Depot esta considerando la posibilidad de dar a los proveedores acceso extranet a sus plazas de datos utilizando la actividad de ventas con dos proveedores clave a través de series de transacciones de intercambio electrónico de datos, de manera que “tenemos que decidir si realmente necesitamos enviarles dicha información por medio de Essbase”, dijo Nickerson

Índice de Materias

- Bases de Datos
 - Analíticas, 3
 - Cubo de Datos, 2
 - Estructuras, 1
 - Multidimensional, 2
 - Operacionales, 3
- Consultas, 29
 - en un Almacén de Datos, 29
- Data Mart, 12
 - Categorías, 12
 - Data Warehouse, 12
 - OLTP, 12
- Data Warehouse
 - Características, 4
 - Componentes, 8
 - Creación, 4
 - Definición, 3
 - Factores de Éxito, 9
 - Problemas, 10
 - Requerimientos, 7
 - Ventajas y Desventajas, 6
- Db2, 42
- Dimensión, 38
 - Atributo de la, 38
 - Entidad de la, 38
 - Jerarquía de la, 38
 - Llave de, 39
 - Niveles de Agregación, 38
- Dimensión
 - Concepto, 38
- Dss, 27
- Esquema
 - Copo de Nieve, 36
 - Concepto, 40
 - Desventajas, 40
 - Ventajas, 40
 - Estrella, 36
 - Concepto, 36
- Essbase, 42
- Hechos, 39
- Medidas, 38
 - Candidatas, 38
- Multidimensionalidad
 - Concepto, 35
- Procesamiento Analítico en Línea
 - Beneficios, 24
 - Casos del Mundo Real, 25
 - Diferencias
 - entre EIS y OLAP, 27
 - entre informes avanzados y OLAP, 28
 - entre OLAP y Data Mining, 29
- HOLAP, 22
- Modelos de Datos, 18
 - Comparativa, 22
- MOLAP, 20
 - Ventajas y Desventajas, 20
- OLAP

- Definición, 16
- Modelo de Datos, 16
- Operaciones Básicas, 16
 - Consolidación, 16
 - Drill-down, 17
 - Slicing and dicing, 17
- Operadores, 29
- Relaciones entre las herramientas, 29
- ROLAP, 20
 - Ventajas y Desventajas, 22
- Utilidades, 23
- Procesamiento Analítico en Línea
 - Operadores
 - DRILL, 30
 - PIVOT, 31
 - ROLL, 30
 - SLICE AND DICE, 31
 - Relaciones entre las Herramientas
 - Data Mining, KDD, 29
 - Data Warehouse y Data Mart, 29
 - OLAP, 29
 - OLTP, 29
- Queries, 35
 - Multidimensionales, 36
- Sistemas
 - de Información, 27
 - para la Gestión, 27
 - EIS, 28
 - OLAP, 16
 - OLTP, 14
 - OLTP Diferencias
 - con Data Warehouse, 15
 - con OLAP, 14
 - Operacionales, 3
 - para Toma de Decisiones, 27
- Tablas
 - de Dimensión, 36
 - de hecho, 36
- Vista de los Datos, 17
 - Como Cubo de Datos, 17
 - Dimensiones, 17