



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

Trabajo de Adscripción

## Bases de Datos Multiplataforma Como Soporte Para La Inteligencia de Negocios



Romina Elizabeth Dos Santos - L.U.: 25333

Prof. Director: Mgter. David Luis la Red Martínez

Licenciatura en Sistemas  
Corrientes - Argentina

2005



A mis padres



# Índice General

<b>1</b>	<b>Bases De Datos Para BI</b>	<b>1</b>
1.1	Bases De Datos . . . . .	1
1.1.1	Objetivos De Las Bases De Datos . . . . .	2
1.1.2	Ventajas De Las Bases De Datos . . . . .	2
1.1.3	Estructura De Las Bases De Datos . . . . .	3
1.2	DBMS . . . . .	3
1.3	Arquitectura De Las Bases De Datos . . . . .	5
1.4	DB2 Universal Database . . . . .	6
1.4.1	Características y Funciones . . . . .	6
1.4.2	Valor Estratégico Del DB2 UDB . . . . .	7
1.4.3	Funciones Complementarias Del DB2 UDB . . . . .	9
1.5	Business Intelligence Para DB2 UDB . . . . .	10
1.6	Business Intelligence . . . . .	11
1.6.1	¿Qué es Business Intelligence? . . . . .	11
1.6.2	¿Qué Puede Hacer Business Intelligence? . . . . .	12
1.6.3	Componentes de Business Intelligence . . . . .	12
1.6.4	Arquitectura BI . . . . .	13
<b>2</b>	<b>DB2 E-Business</b>	<b>15</b>
2.1	Conceptos Generales . . . . .	15
2.2	Funciones de Business Intelligence Warehouse Manager . . . . .	15
2.3	Estructura de DB2 Business Intelligence . . . . .	15
2.4	DB2 Business Intelligence - ETL . . . . .	15
2.4.1	¿Qué Es ETL? . . . . .	17
2.4.2	Arquitecturas Típicas De ETL . . . . .	18
2.4.3	DB2 Warehouse Center / Manager (ETL) . . . . .	18
2.5	DB2 BI - Datawarehouse . . . . .	19
2.5.1	Data Warehouse . . . . .	19
2.5.2	DB2 Para Business Intelligence (DW) . . . . .	20

2.5.3	Introducción A Los Almacenes De Datos . . . . .	21
2.5.4	Análisis De Datos Para El Soporte En La Toma De Decisiones. . . . .	21
2.5.5	Problemas . . . . .	22
2.5.6	¿Qué Es Un Data Warehouse? . . . . .	24
2.5.7	Beneficios Económicos . . . . .	26
2.5.8	¿Por Qué Usar Un Data Warehouse? . . . . .	26
2.5.9	Los Requerimientos De Un Data Warehouse . . . . .	27
2.5.10	Arquitectura Datawarehouse . . . . .	28
2.5.11	Componentes y Estructuras . . . . .	28
2.5.12	La Estructura Básica De La Arquitectura DW Incluye: . . . . .	30
2.5.13	¿Cómo Trabaja El Data Warehouse? . . . . .	31
2.5.14	¿En Qué Podemos Usarlo? . . . . .	31
2.5.15	Impactos DW . . . . .	31
2.5.16	Costos v/s Valor De DW . . . . .	33
2.5.17	Costos De Un DW . . . . .	33
2.5.18	Valor Del DW . . . . .	35
2.6	Data Marts . . . . .	35
2.6.1	Definición de Data Marts . . . . .	36
2.6.2	Arquitectura de Data Marts . . . . .	37
2.6.3	Crecimiento De Los Data Marts . . . . .	37
2.6.4	Administración De Los Data Marts . . . . .	38
2.6.5	Paquetes De Data Marts . . . . .	39
2.6.6	Data Marts Virtuales y Meta Vistas . . . . .	39
2.6.7	¿Por Qué Ha Crecido La Popularidad De Los Data Marts? . . . . .	40
<b>3</b>	<b>Toma de Decisiones</b> . . . . .	<b>43</b>
3.1	DB2 Business Intelligence - OLAP . . . . .	43
3.2	“On-Line Analytical Processing” . . . . .	43
3.2.1	¿Qué Es OLAP? . . . . .	43
3.2.2	Las Herramientas OLAP (On-Line Analytical Processing) Son Más Genéricas. . . . .	44
3.2.3	Las Herramientas De OLAP Se Caracterizan Por: . . . . .	45
3.2.4	OLTP v/s OLAP: Dos Mundos Diferentes . . . . .	45
3.2.5	¿Cuál Es La Diferencia Entre OLAP y Minería De Datos? . . . . .	46
3.3	OLTP - “On-line Transaction Processing”. . . . .	46
3.3.1	Diferencia Entre Data Warehouse y Bases De Datos Operacionales (OLTP). . . . .	46
3.3.2	Relación Entre Ambos Esquemas. . . . .	49
3.3.3	¿Una Máquina o Dos? . . . . .	49

<b>4 CUBE VIEWS</b>	<b>51</b>
4.1 DB2 Business Intelligence - DB2 Cube Views . . . . .	51
4.2 ¿Que Es DB2 Cube Views? . . . . .	51
4.3 ¿Que Es DB2 MQT? . . . . .	52
4.3.1 Compartiendo Los Metadatos OLAP Con El DB2 Cube Views. . . . .	52
4.4 ¿Por Qué Cube Views? . . . . .	53
4.4.1 Modelo Del Cubo . . . . .	56
4.4.2 DB2 Cube Views - The OLAP-aware Database (OLAP)	57
4.4.3 DB2 Cube Views. . . . .	58
4.4.4 DB2 Cube Views = An OLAP Accelerator . . . . .	61
4.4.5 Cube Views Performance . . . . .	61
4.4.6 DB2 Cube Views – Beneficios . . . . .	63
<b>Bibliografía</b>	<b>67</b>
<b>Índice de Materias</b>	<b>71</b>



# Índice de Figuras

1.1	Estructura de una Base de Datos. . . . .	4
1.2	Sistemas de Administracion de Bases de Datos (DBMS). . . . .	4
1.3	Arquitectura para Sistemas de Bases de Datos. . . . .	5
1.4	Arquitectura de BI. . . . .	14
2.1	Funciones de BI. . . . .	16
2.2	Estructura de BI. . . . .	16
2.3	ETL - Extraccion, Transferencia y Carga de Datos. . . . .	17
2.4	Arquitectura de ETL . . . . .	18
2.5	Process Model - Build Market Dimension. . . . .	19
2.6	Datawarehouse . . . . .	20
2.7	DB2 para Business Intelligence . . . . .	21
2.8	Almacenes de Datos . . . . .	22
2.9	Ventajas de los Almacenes de Datos . . . . .	23
2.10	Problemas de los Almacenes de Datos . . . . .	23
2.11	Almacenes de Datos II . . . . .	24
2.12	Arquitectura Lógica de un Sistema de Data Warehousing . . . . .	29
2.13	Estructura Básica de la Arquitectura del DW . . . . .	30
2.14	Data Marts . . . . .	37
2.15	Arquitectura de Data Marts . . . . .	38
2.16	Arquitectura . . . . .	41
2.17	Data Warehouse Vs Data Marts . . . . .	42
3.1	Estructura del OLAP dentro de BI. . . . .	44
3.2	Data Warehouse y OLTP. . . . .	48
3.3	Esquema de Relación entre OLTP Y DDW. . . . .	49
4.1	Metadatos Olap y DB2 Cube Views . . . . .	53
4.2	Modelo de cubo . . . . .	55
4.3	The OLAP-aware Database (OLAP) . . . . .	57

4.4	DB2 Cube Views . . . . .	59
4.5	Performance Advisor . . . . .	60
4.6	DB2 and OLAP ACCELERATOR. . . . .	61
4.7	Cube Views Performance . . . . .	62
4.8	DB2 Cube Views . . . . .	63
4.9	Modelo Cube Views . . . . .	65

# Capítulo 1

## Bases De Datos Multiplataforma Como Soporte Para La Inteligencia De Negocios (BI)

### 1.1 Bases De Datos

La necesidad de mejorar la manera de acceder y manejar los datos ha evolucionado desde el concepto de *IMS (Information Management System)* a la nueva generación de sistemas de administración de bases de datos relacionales (*RDBMS*) [35].

En los últimos tiempos se ha escuchado hablar sobre una nueva base de datos, llamada “Universal”, que puede almacenar y hacer búsquedas no solamente de datos alfanuméricos sino también de imágenes, audio, video y otros objetos. [43].

Esta ventaja de las bases de datos universales abre un sin número de oportunidades que permiten mejorar tanto los servicios como las aplicaciones. [9] [10]

*Una base de datos es una colección de datos, lógicamente relacionados, que apoyan el acceso compartido de muchos usuarios y aplicaciones.* Mientras que

un archivo, normalmente, contiene datos acerca de un tipo de entidad (ej.: personal, órdenes, clientes, ventas), una base de datos contiene datos acerca de muchos tipos de entidades e información acerca de cómo las entidades están lógicamente relacionadas entre sí. [4] [30] [37]

Las bases son cualquier conjunto de datos organizados para su almacenamiento en la memoria de un ordenador o computadora, diseñado para facilitar su mantenimiento y acceso de una forma estándar. Los datos suelen aparecer en forma de texto, números o gráficos.

Las bases de datos se crean con el *objetivo* de almacenar grandes cantidades de datos que antes se almacenaba en libros, lo que era lento, costoso y complejo (cualquier actualización a realizar, había que hacerla en cada uno de los libros en los que apareciera dicha información a modificar).

### 1.1.1 Objetivos De Las Bases De Datos

Los *objetivos* de las bases de datos son los de automatizar: [20]

- \* El mantenimiento.
- \* Cualquier generación de información.
- \* Cualquier consulta sobre dicha información.

### 1.1.2 Ventajas De Las Bases De Datos

Las principales *ventajas* se pueden resumir como sigue:

**Ahorro de Espacio:** No hacen falta archivos de papeles que pudieran ocupar mucho espacio.

**Velocidad:** La máquina puede obtener y modificar datos con mucha mayor velocidad que un ser humano. Así es posible satisfacer con rapidez consultas de casos particulares.

**Ahorro de Trabajo:** Se elimina gran parte del tedio de manejar archivos a mano. Las tareas mecánicas siempre serán mejor realizadas por las máquinas.

**Actualización:** Se dispone en cualquier momento de información precisa y al día.

**Comodidad:** Al tener la información en un mismo sitio, se ahorrará tiempo y trabajo.

**Disminución de la Redundancia:** La duplicación de datos genera a su vez una duplicación del trabajo a la hora de mantenerlos y actualizarlos. Por lo tanto las Bases de Datos al reducir la duplicación de datos, disminuyen el trabajo.

**Compartición de Datos:** Se trata de datos actuales, ya que al ser centralizados, se puede tener acceso a los datos con la última actualización en prácticamente tiempo real. [41]

**Posibilidad de Aplicar Restricciones de Seguridad:** Para mantener la seguridad acerca del mantenimiento de los datos, los administradores de la Base de Datos, crean una *jerarquía de acceso*, que permitirá o prohibirá a los usuarios hacer una u otra acción sobre dicha base de datos.

**Posibilidad de Mantener la Integridad:** En una base de datos se debe mantener una coherencia. [20] Esto se controlará mediante:

- Máscaras.
- Reglas de validación.

### 1.1.3 Estructura De Las Bases De Datos

Una *base de datos* es una colección de tablas y objetos relacionadas entre sí y organizados como un grupo. La estructura de una base de datos se muestra en la figura 1.1 de la página 4. [37]

## 1.2 Sistema De Administración De Bases De Datos (DBMS)

El *DBMS* es un conjunto de programas que maneja todos los accesos a las *bases de datos*. ver figura 1.2 de la página 4. [37]

*Funciones de un DBMS:*

- \* Definición de datos.

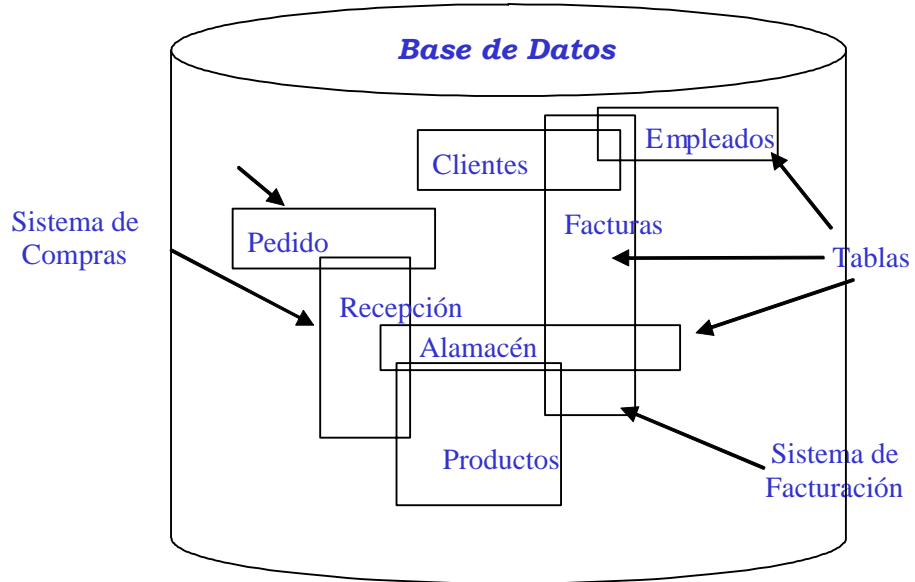


Figura 1.1: Estructura de una Base de Datos.

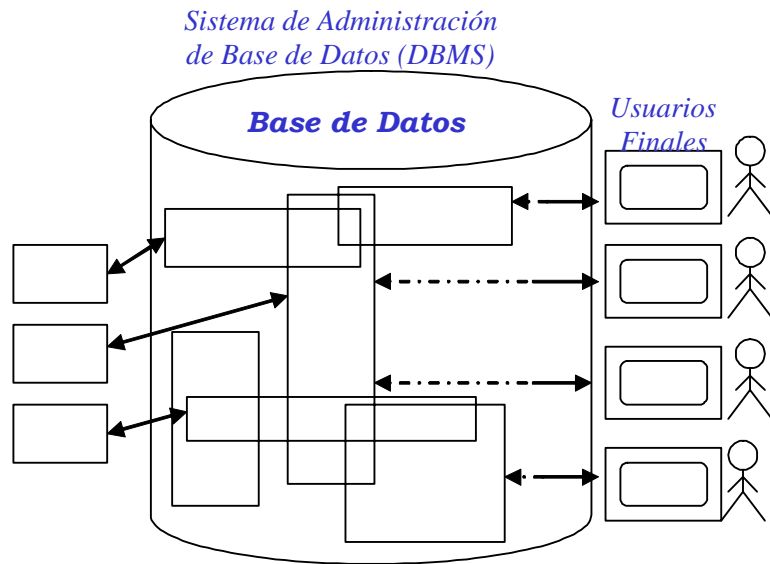


Figura 1.2: Sistemas de Administración de Bases de Datos (DBMS).

- \* Manipulación de Datos.
- \* Seguridad e integridad de los datos.
- \* Recuperación y concurrencia de los datos.
- \* Diccionario de datos.
- \* Desempeño.

### 1.3 Arquitectura De Las Bases De Datos

La *arquitectura* de una Base de Datos se representa en la figura 1.3 de la página 5. [37]

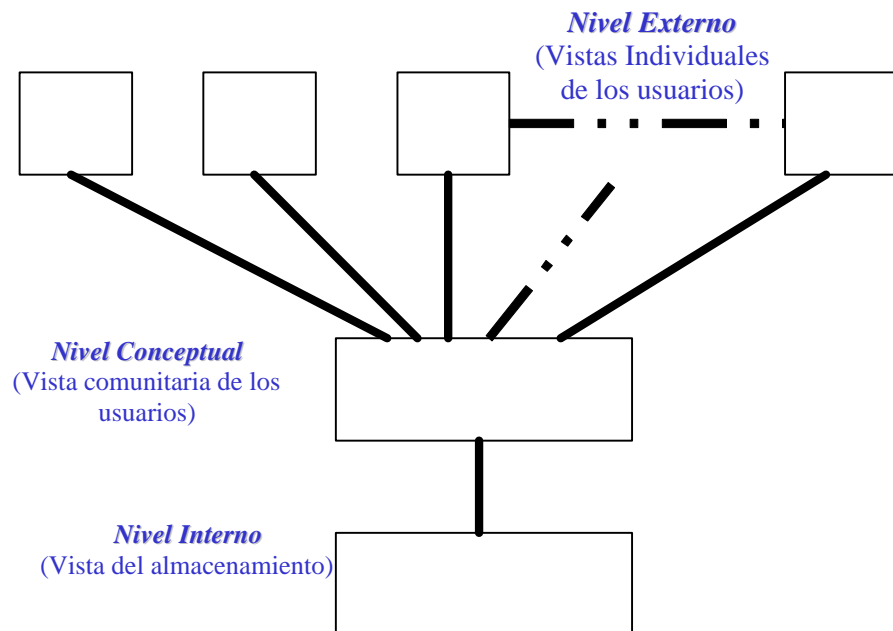


Figura 1.3: Arquitectura para Sistemas de Bases de Datos.

Se consideran los siguientes niveles:

**Nivel Interno:** El nivel más bajo de abstracción describe cómo se almacenan realmente los datos.

**Nivel Conceptual:** Describe qué datos son realmente almacenados en la

base de datos y las relaciones que existen entre los datos.

**Nivel Externo:** Describe sólo parte de la base de datos completa.

## 1.4 DB2 Universal Database

*DB2 Universal Database* es una base de datos universal. Es completamente escalable, veloz y confiable. Corre en modo nativo en casi todas las plataformas, como *Windows NT*, *Sun Solaris*, *HP-UX*, *AIX*, *OS/400* y *OS/2*. [5] [36]

*DB2* es un software de base de datos *relacional*. Es completamente multimedia, disponible para su uso en la Web, muy bueno para satisfacer las demandas de las grandes corporaciones y bastante flexible para servir a los medianos y pequeños negocios.

*DB2 UDB* es un sistema manejador de base de datos relacional fuertemente escalable. Es suficientemente flexible para atender estructuras e inestructuras manejadoras de datos necesarias para usuarios simples de grandes empresas. Es conveniente para una gama amplia de aplicaciones de los cliente, quienes pueden desplegar una variedad de plataformas de hardware y software desde dispositivos manuales a los sistemas multiprocesador paralelos masivos.

### 1.4.1 Características y Funciones

*DB2 UDB* es el sistema de gestión de datos que entrega una plataforma de base de datos flexible y rentable para construir un sistema robusto para aplicaciones de gestión.

*DB2 UDB* libera los recursos con amplio apoyo al *open source* (fuente abierta) y plataformas de desarrollo populares como *J2EE* y *Microsoft .NET*. [9] [19]

La familia *DB2 UDB* también incluye soluciones adaptadas para necesidades específicas como *Business Intelligence* y herramientas avanzadas.

Los productos *Data Management* proveen las habilidades de administración de datos necesarias para administrar, almacenar y acceder a los datos operacionales y no estructurados necesarios de la Web. Estos productos proporcionan soporte a las actividades de reunir, administrar, analizar y compar-

tir informaciones para fines estratégicos.

Las *transacciones de e-business* necesitan una base de información cada vez más heterogénea para sustentar aplicativos que diferencian los procesos de negocio. Las transacciones también deben ser rápidas, seguras y estar siempre disponibles. Sin embargo, las transacciones de e-business deben agregar valor a las informaciones con análisis incorporado para dar soporte a la *toma de decisiones en tiempo real*.

Las transacciones de e-business también necesitan ser “completas en contenido” y administrar formularios de información, como datos *XML*, imágenes y otros formatos que van más allá de simples líneas y columnas de información alfa-numérica. Teniendo el total de sus activos de informaciones digitalizados, las compañías pueden mejorar el servicio a clientes, operar con mayor eficiencia, y ser más competitivas. [9] [19]

#### 1.4.2 Valor Estratégico Del DB2 UDB

*DB2 UDB* es el producto principal de la estrategia de *Data Management de IBM*.

*DB2 UDB* es un sistema para administración de bases de datos relacionales (*RDBMS*) multiplataforma, especialmente diseñado para ambientes distribuidos, permitiendo que los usuarios locales compartan información con los recursos centrales. [11] [9]

#### Historia

*DB2 UDB* no es un producto nuevo. Fue construido en base a dos productos incluidos en el DB2 de AIX en el año 1994: *DB2 Common Server*, que para propósitos generales incluía funciones avanzadas para el mercado de servidores de bases de datos, con soporte de hardware SMP y OLTP; y el *DB2 Parallel Edition*, que fue desarrollado para soportar aplicaciones de gran escala, como *Data Warehousing* y *Data Mining*. [10]

### **Integridad**

*DB2 UDB* incluye características de integridad, asegurando la protección de sus datos aún en caso de que los sistemas sufran un colapso; y de seguridad, permitiendo realizar respaldos en línea con distintos grados de granularidad, sin que esto afecte la disponibilidad de acceso a los datos por parte de los usuarios. [19]

### **Múltiples Usos**

Provee la capacidad de hacer frente a múltiples necesidades, desde procesamiento transaccional de misión crítica (OLTP), hasta análisis exhaustivo de los datos para el soporte a la toma de decisiones (OLAP).

### **Escalabilidad**

Sus características distintivas de escalabilidad le permiten almacenar información en un amplio rango de equipos, desde una PC portátil hasta un complejo ambiente de *mainframes* procesando en paralelo.

### **Web Enabled para E-business**

Incluye tecnología basada en la Web que permite generar aplicaciones en sus *Intranets* y responder a las oportunidades de negocios disponibles en Internet. Además, *DB2 UDB* provee soporte para *Java*.

### **Facilidad de Instalación y Uso**

La primera versión de *DB2* para *NT* fue reconocida en el mercado como una base de datos muy poderosa, pero difícil de instalar y usar.

En esta versión (*DB2 UDB*), *IBM* agregó muchas herramientas gráficas para facilitar el uso para los usuarios, como también para los administradores y desarrolladores. Dicha versión incluye guías para operaciones como instalación, configuración de performance, setup, etc. Además, se agregaron

herramientas para facilitar las tareas de integración con otras bases de datos, tecnologías de *networking* y desarrollo de aplicaciones.

### Universalidad

*DB2 UDB* es, además, la única base de datos realmente *universal*; es multi-plataforma (16 plataformas - de las cuales 10 no son de IBM), brinda soporte a un amplio rango de clientes, soporta el acceso de los datos desde Internet y permite almacenar todo tipo de datos incluyendo texto, audio, imágenes y video o cualquier otro definido por el usuario.

### 1.4.3 Funciones Complementarias Del DB2 UDB

#### Conectividad

Las herramientas de conectividad permiten acceder a los datos más allá de donde ellos se encuentren. El slogan “*cualquier cliente, a cualquier servidor, en cualquier red*” está completamente sustentado por la funcionalidad que sus herramientas ofrecen.

EL *DB2 Connect* le permiten acceder a los datos de *DB2* en *mainframe* o *AS/400*, desde *Windows NT*, *Windows 95 / 98*, *OS/2* o cualquiera de los *Unix* soportados. Además, el producto *Datajoiner* posibilita acceder de forma única y transparente a los datos residentes en *Oracle*, *Sybase*, *Informix*, *Microsoft SQL Server*, *IMS*, *VSAM* y otros.

#### Data Warehousing

*DB2 UDB* provee la infraestructura necesaria para soportar el proceso de *toma de decisiones* en cualquier tamaño y tipo de organización. *DB2 UDB* es el producto dirigido a resolver la problemática a nivel departamental (*Data Marts*), ya que un único producto provee la capacidad para acceder a datos en *Oracle*, *Sybase*, *Informix*, *Microsoft SQL Server*, *VSAM* o *IMS*, además de la familia *DB2*.

Permite de forma totalmente gráfica acceder, transformar y distribuir los datos automáticamente y sin programar una línea de código.

## Data Mining

*DB2 UDB* posibilita el análisis orientado al descubrimiento de información escondida en los datos, realizando modelización predictiva, segmentación de la base de datos, análisis de vínculos, o detección de desviaciones. Incluye las siguientes técnicas: *clustering (segmentación)*, *clasificación*, *predicción*, *descubrimiento asociativo*, *descubrimiento secuencial de patrones*, y *secuencias temporales*. Todas las técnicas mencionadas permiten realizar *segmentación de clientes*, *detección de fraude*, *retención de clientes*, *ventas cruzadas*, etc.

## 1.5 Business Intelligence Para DB2 UDB

Las ediciones del *DB2 Data Warehouse* proporcionan gran funcionalidad de *BI* dentro de las bases de datos. [38] [43]

Mejor que comprar e integrar diversos productos diferentes en su sistema de *TI*, una única solución proporciona almacenamiento de datos robusto y funcionalidad analítica de fácil utilización y administración.

La tecnología basada en las ediciones del *DB2 UDB Data Warehouse*, permite integrar la información en tiempo real, percepción y toma de decisiones. Estas nuevas ediciones combinan la fuerza del *DB2 UDB* a la infraestructura esencial de *Business Intelligence*.

Las ediciones del *DB2 Data Warehouse* hacen más fácil, la compra de la plataforma completa de *BI* del *DB2*, realizada por los clientes. Proporcionan gran funcionalidad de *BI* dentro de la base de datos.

La Edición Corporativa de *DB2 Data Warehouse* representa el marco más reciente de la evolución de *DB2*.

La funcionalidad de *Business Intelligence* inclusive en el *Data Warehouse Enterprise Edition* incluye:

- Servicios robustos de *ETML (Extract, Transform, Load and Move)* con agentes distribuidos para maximizar el desempeño.
- Aplicaciones eficaces de búsqueda de datos para modelado y visualización de rutinas y resultados de búsqueda, así como para la integración de aplicaciones analíticas

- Optimizaciones integradas a *OLAP* para acelerar el desarrollo e implementación de aplicaciones analíticas
- Soporte para configuraciones de servidores en *cluster*, *MPP* (*Massively Parallel Processing*) en una arquitectura *true shared-nothing*.
- Funcionalidad de administración de consultas y recursos para controlar, administrar y monitorear el ambiente de carga, de consultas y de actividades.
- Y además, todos los recursos de desempeño y funcionalidad de *Business Intelligence* en el *DB2 UDB Enterprise Server Edition* y más...

## 1.6 Business Intelligence

Algo peor que no tener información disponible es tener mucha información y no saber qué hacer con ella. *La Inteligencia de Negocios* o *Business Intelligence* (*BI*) es la solución a ese problema, pues por medio de dicha información puede generar escenarios, pronósticos y reportes que apoyen a la toma de decisiones, lo que se traduce en una ventaja competitiva [38] [22] [28] [8] [1].

La clave para *BI* es la *información* y uno de sus mayores beneficios es la posibilidad de utilizarla en la toma de decisiones. En la actualidad hay una gran variedad de software de *BI* con aplicaciones similares que pueden ser utilizados en las diferentes áreas de la empresa, tales como, ventas, marketing, finanzas, etc.

Son muchas las empresas que se han beneficiado por la implementación de una sistema de *BI*, además se pronostica que con el tiempo se convertirá en una necesidad de toda empresa.

### 1.6.1 ¿Qué es Business Intelligence?

*La Inteligencia de Negocios* o *Business Intelligence* (*BI*) se puede definir como el proceso de analizar los bienes o datos acumulados en la empresa y extraer una cierta inteligencia o conocimiento de ellos [43].

Dentro de la categoría de bienes se incluyen las bases de datos de clientes, información de la cadena de suministro, ventas personales y cualquier actividad de marketing o fuente de información relevante para la empresa. [17] [38]

*BI* apoya a los tomadores de decisiones con la información correcta, en el momento y lugar correcto, lo que les permite tomar mejores decisiones de negocios.

La información adecuada en el lugar y momento adecuado incrementa efectividad de cualquier empresa.

### 1.6.2 ¿Qué Puede Hacer Business Intelligence?

Con *BI* se puede:

- Generar reportes globales o por secciones.
- Crear una base de datos de clientes.
- Crear escenarios con respecto a una decisión.
- Hacer pronósticos de ventas y devoluciones.
- Compartir información entre departamentos.
- Análisis multidimensionales.
- Generar y procesar datos.
- Cambiar la estructura de toma de decisiones.
- Mejorar el servicio al cliente.

### 1.6.3 Componentes de Business Intelligence

#### **Multidimensionalidad.**

La información multidimensional se puede encontrar en hojas de cálculo, bases de datos, etc. [38] [17] [43]

Una herramienta de *BI* debe de ser capaz de reunir información dispersa en toda la empresa e incluso en diferentes fuentes para así proporcionar a los departamentos la accesibilidad, poder y flexibilidad que necesitan para analizar la información. Por ejemplo, un pronóstico de ventas de un nuevo producto en varias regiones no está completo si no se toma en cuenta también el comportamiento histórico de las ventas de cada región y la forma en que la introducción de nuevos productos se ha desarrollado en cada región en cuestión.

### **Data Mining**

Las empresas suelen generar grandes cantidades de información sobre sus procesos productivos, desempeño operacional, mercados y clientes. Pero el éxito de los negocios depende por lo general de la habilidad para ver nuevas tendencias o cambios en las tendencias.

Las aplicaciones de *Data Mining* pueden identificar tendencias y comportamientos, no sólo para extraer información, sino también para descubrir las relaciones en bases de datos que pueden identificar comportamientos que no son muy evidentes.

### **Agentes**

Los agentes son programas que “*piensan*”. Ellos pueden realizar tareas a un nivel muy básico sin necesidad de intervención humana. Por ejemplo, un agente pueden realizar tareas un poco complejas, como elaborar documentos, establecer diagramas de flujo, etc.

### **Data Warehouse**

Es la respuesta de la tecnología de información a la descentralización en la toma de decisiones. Coloca información de todas las áreas funcionales de la organización en manos de quien toma las decisiones. También proporciona herramientas para búsqueda y análisis.

#### **1.6.4 Arquitectura BI**

La arquitectura de BI se muestra en la figura 1.4 de la pág. 14.

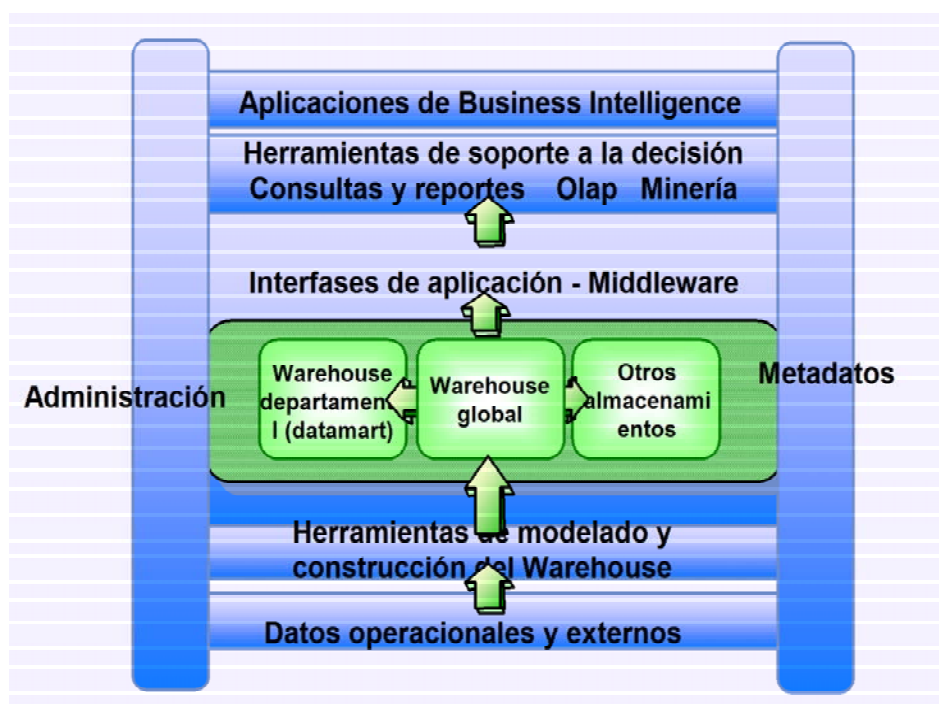


Figura 1.4: Arquitectura de BI.

## Capítulo 2

# Introducción al DB2 Para E-Business

### 2.1 Conceptos Generales

### 2.2 Funciones de Business Intelligence Warehouse Manager

Las Funciones de BI Warehouse Manager, se representan en la figura 2.1 de la pagina 16.

### 2.3 Estructura de DB2 Business Intelligence

La Estructura de *BI* se muestra en la figura 2.2 de la página 16 [33].

### 2.4 DB2 Business Intelligence - ETL

La estructura del *ETL* (Extracción - Transformación - Limpieza - Carga) se representa en la figura 2.3 de la pagina 17. [37] [38]



Figura 2.1: Funciones de BI.

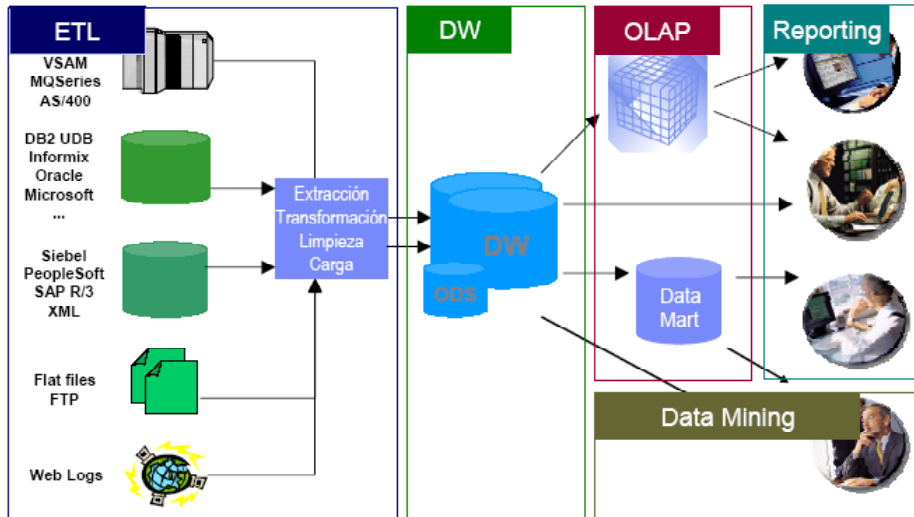


Figura 2.2: Estructura de BI.

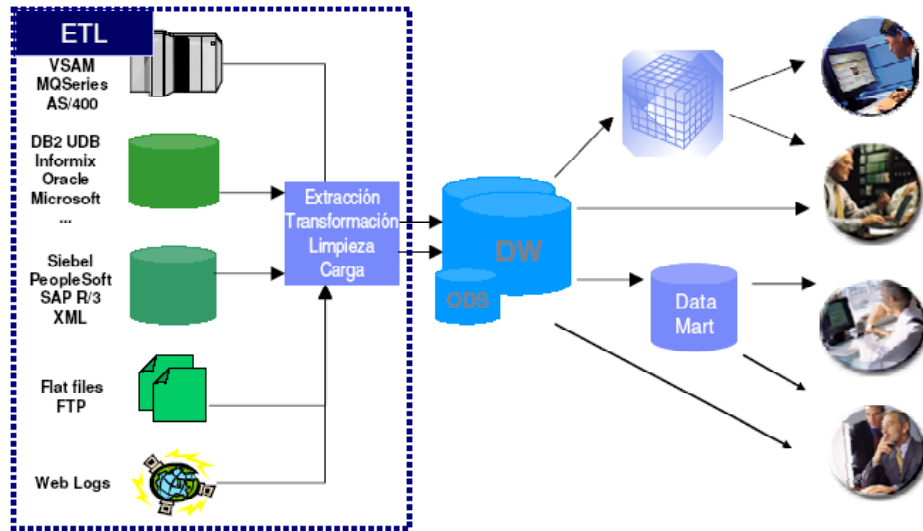


Figura 2.3: ETL - Extracción, Transferencia y Carga de Datos.

### 2.4.1 ¿Qué Es ETL?

- El proceso de extraer datos de una o más fuentes de datos (Extraction), aplicar reglas de negocio o técnicas (Transformación) y cargar los datos en la base de datos de destino (Load). [37] [38] [5]
  - Una solución completa de *Data Warehouse* tiene varios subsistemas de *ETL*.
    - \* Operational System(s) a Operational Data Store (ODS).
    - \* Operational System(s) y ODS a Warehouse.
    - \* Carga inicial de ETL (o refresco completo).
    - \* ETL con carga incremental.
    - \* Warehouse a Data Mart (Relacional y non-Relacional).
  - Típicamente uno por cada *Data Mart*.
- Periodicidad y latencia.
  - Periodicidad es cuán frecuentemente se invocan procesos *ETL*:
    - \* Mensual, semanal, diario, horario, minuto y casi tiempo real.

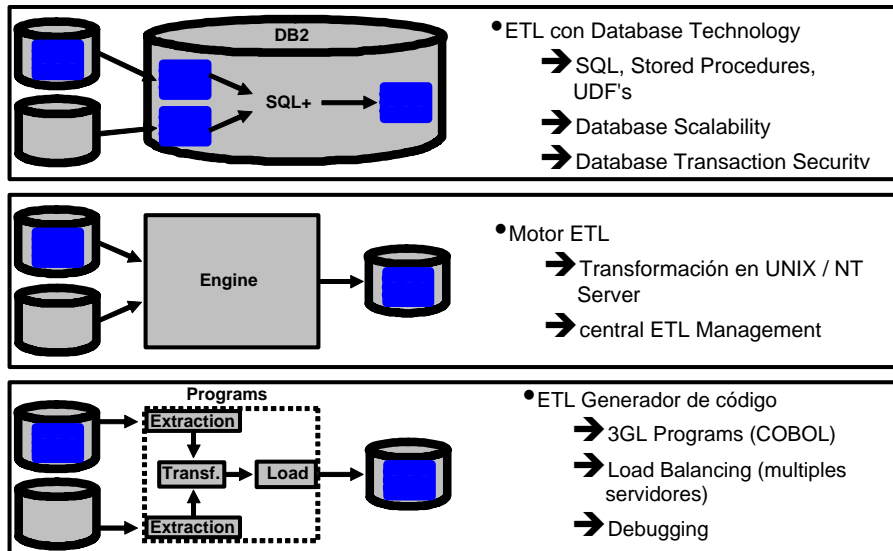


Figura 2.4: Arquitectura de ETL

- Latencia es cuán rápido cambios operacionales son reflejados en el *Warehouse*.
- Relacionada de alguna manera a la Periodicidad.
  - \* Para Periodicidades grandes se refiere a la ventana batch.
  - \* Para Periodicidades cortas o casi tiempo real se refiere a la demora con que un cambio de datos operacionales llega al *Warehouse*.

### 2.4.2 Arquitecturas Típicas De ETL

La arquitectura interna del *ETL* se representa en la figura 2.4 de la página 18.

### 2.4.3 DB2 Warehouse Center / Manager (ETL)

- Consola de administración gráfica. [32]
- Acceso a la mayoría *RDBMS's* y archivos planos.
- Monitorea las tareas de la base de datos.

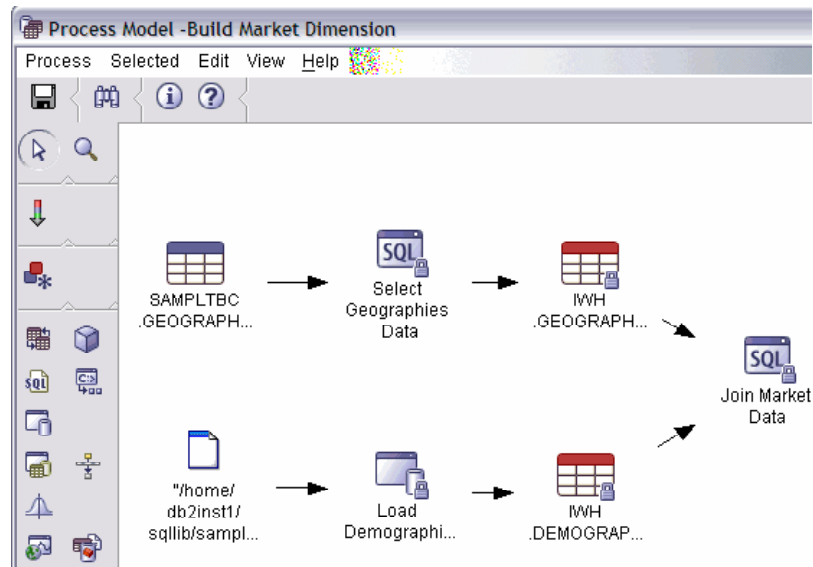


Figura 2.5: Process Model - Build Market Dimension.

- Más de 150 transformaciones de datos. [37]
- Cargas masivas el *DB2 data warehouses*.
- Adhiere al estándar *CWMI*.
- Agentes de extracción y transformación en servers remotos. [42]
- Librerías de transformación.

## 2.5 DB2 Business Intelligence - Datawarehouse

### 2.5.1 Data Warehouse

La Estructura del *Datawarehouse* dentro del *BI* se representa en la figura 2.6 de la pagina 20. [37] [42]

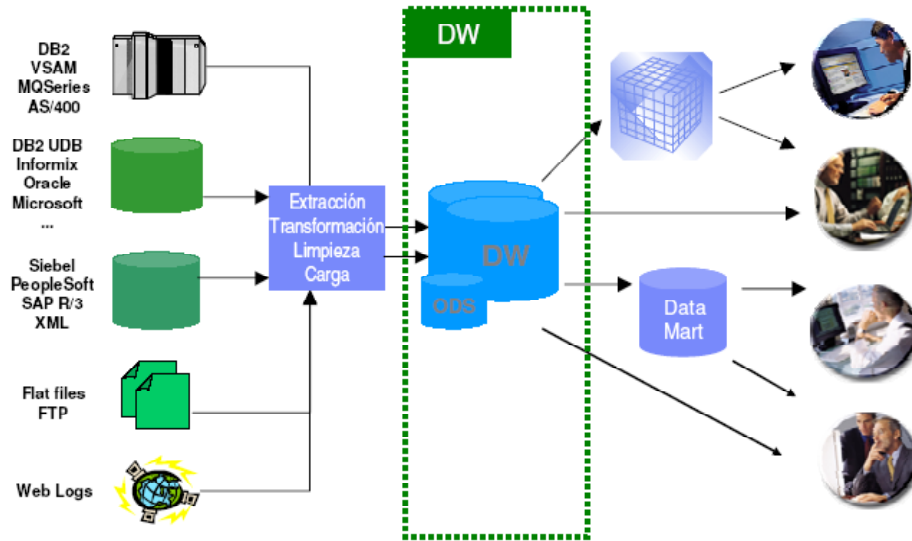


Figura 2.6: Datawarehouse

### 2.5.2 DB2 Para Business Intelligence (DW)

- Líder Tecnológico.
- Multi-Plataforma.
- Procesamiento en Paralelo.
- Optimizador Avanzado de Consultas.
- *Cube Views*.
- Funciones Autonómicas.
- Administración, Alta Disponibilidad.
- T.C.O.

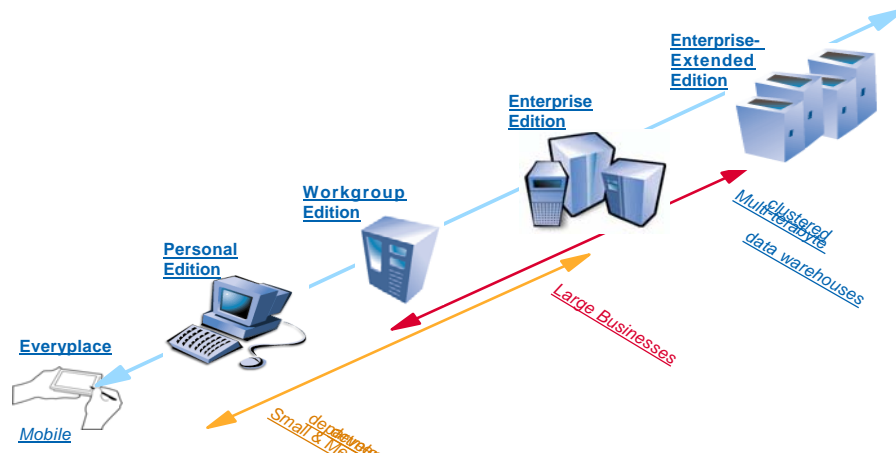


Figura 2.7: DB2 para Business Intelligence

### 2.5.3 Introducción A Los Almacenes De Datos

### 2.5.4 Análisis De Datos Para El Soporte En La Toma De Decisiones.

Generalmente, la información que se quiere investigar sobre un cierto dominio de la organización se encuentra en bases de datos y otras fuentes muy diversas, tanto internas como externas. [21] [29]

Muchas de estas fuentes son las que se utilizan para el trabajo diario (*Bases de Datos Operacionales*). [14] [2]

Sobre estas mismas bases de datos de trabajo ya se puede extraer conocimiento (Visión Tradicional).

#### Uso De La Base De Datos Transaccional Para Varios Cometidos:

- Se mantiene el trabajo transaccional diario de los sistemas de información originales (conocido como *OLTP*, *On-Line Transactional Processing*).
- Se hace *análisis de los datos en tiempo real* sobre la misma base de datos (conocido como *OLAP*, *On-Line Analytical Processing*).

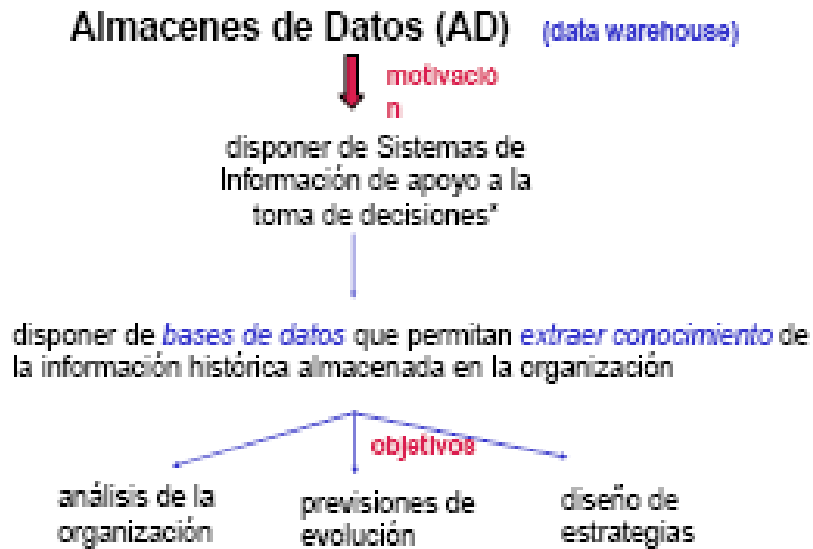


Figura 2.8: Almacenes de Datos

### 2.5.5 Problemas

- Perturba el trabajo transaccional diario de los sistemas de información originales ("killer queries"). Se debe hacer por la noche o en fines de semana.
- La base de datos está diseñada para el trabajo transaccional, no para el análisis de los datos. [13] [29]
- Generalmente no puede ser en tiempo real (era AP pero no OLAP).
- Se desea operar eficientemente con esos datos.
- Los costos de almacenamiento masivo y conectividad se han reducido drásticamente en los últimos años.
- Parece razonable recoger los datos (información histórica) en un sistema separado y específico.

**Nace El Data-Warehousing**



Figura 2.9: Ventajas de los Almacenes de Datos

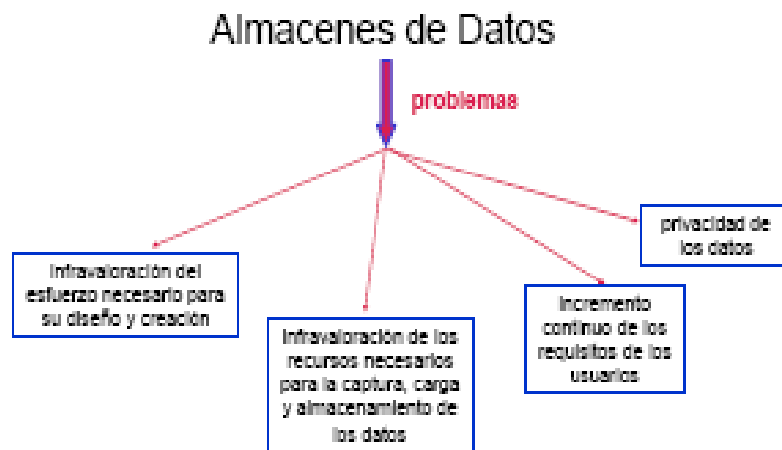


Figura 2.10: Problemas de los Almacenes de Datos

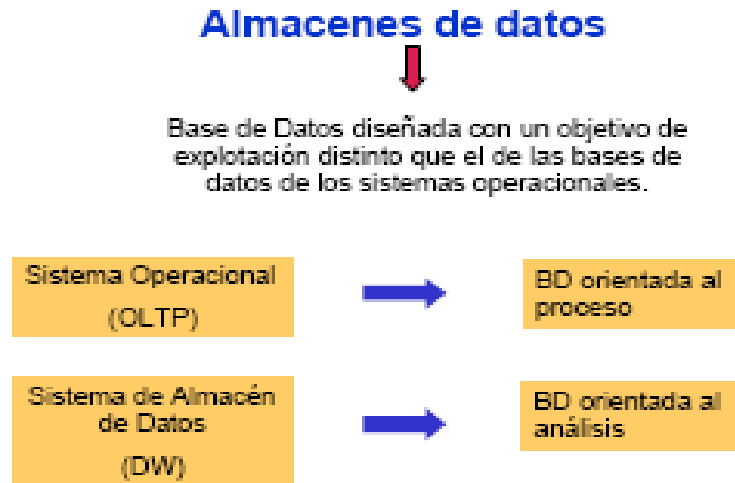


Figura 2.11: Almacenes de Datos II

Sistemas Operacionales (OLTP)	Almacen De Datos (DW)
- Almacena datos actuales.	- Almacena datos históricos.
- Almacena datos de detalle.	- Almacena datos de detalle.
- Bases de datos medianas.	- Bases de datos grandes.
- Los datos son dinámicos.	- Los datos son estáticos.
- Los procesos son repetitivos.	- Los proc. no son previsible.
- Elevado N° de transacciones.	- Bajo N° de transacciones.
- Tiempo de respuesta es pequeño.	- Tiempo de respuesta es variable.
- Dedicado al proc. de transacc.	- Dedicado al análisis de datos.
- Orientado al proc. de la organiz.	- Orientado a informac. relevante.
- Soporta decisiones diarias.	- Soporta decisiones estratégicas.
- Sirve a muchos usuarios.	- Sirve a técnicos de dirección.

### 2.5.6 ¿Qué Es Un Data Warehouse?

Los sistemas de *Data Warehousing* son el centro de la arquitectura de los Sistemas de Información de los 90's. Han surgido como respuesta a la problemática de extraer información sintética a partir de datos atómicos almacenados en bases de datos de producción. Uno de los objetivos principales de este tipo

de sistemas es servir como base de información para la toma de decisiones. Los beneficios obtenidos por la utilización de este tipo de sistemas se basan en el acceso interactivo e inmediato a información estratégica de un área de negocios. [42] [24] [13]

Un *Data Warehouse* es una colección de datos que es:

**Orientada a Sujetos:** Un primer aspecto de un *Data Warehousing* es que esta orientado a los mayores sujetos de la empresa. Un *Data Warehousing* está enfocado en la modelización de los datos y el diseño de la base de datos, exclusivamente.

Los datos se organizan de acuerdo al sujeto en vez de la aplicación, por ejemplo, una compañía de seguros usando un almacén de datos podría organizar sus datos por cliente, premios, y reclamaciones, en lugar de por diferentes productos (automóviles, vida, etc.). Los datos organizados por sujetos contienen solo la información necesaria para los procesos de soporte para la toma de decisiones.

**Los Datos son Integrados:** El aspecto más importante del ambiente de un *Data Warehouse* es que sus datos están *integrados*. Cuando los datos residen en muchas aplicaciones separados por los distintos entornos operacionales, la decodificación de los datos es a menudo inconsistente. Por ejemplo, en una aplicación, la palabra gender podría codificarse como "m" y "f" en otra como "0" y "1". cuando los datos fluyen de un entorno operacional a un entorno de almacen de datos o de data warehouse, ellos asumen una codificación consistente, por ejemplo gender siempre se transformaría a "m" y "f".

**Variante en el Tiempo:** Los datos en el *Warehouse* son precisos para un cierto momento, no necesariamente ahora; por eso se dice que los *datos en el Warehouse son variantes en el tiempo*. El *almacén de datos* contiene un lugar para guardar datos con una antigüedad de 5 a 10 años, o incluso más antiguos, para poder ser usados en comparaciones, tendencias y previsiones. Estos datos no se modificarán. Toda estructura clave en un *Warehouse* contiene implícita o explícitamente un elemento del tiempo. Esto no necesariamente pasa en el ambiente operacional.

**No Volátil:** Solo ocurren dos operaciones, la carga inicial, y el acceso a los datos. *No hay necesidad de updates* (en su sentido general). Los datos no serán modificados o cambiados de ninguna manera una vez ellos han sido introducidos en el almacén de datos, solamente podrán ser cargados, leídos y/o accedidos.

Un *Data Warehouse* soporta procesamiento informático, brindando una sólida plataforma de datos históricos, integrados, de los cuales hacer análisis.

El *Data Warehouse* con las herramientas adecuadas nos permitirá obtener o realizar análisis, reporting, extracción y exploración de los datos para, en suma, transformar los datos en información útil para nuestra organización.

### 2.5.7 Beneficios Económicos

Normalmente los beneficios económicos que podemos obtener de un *Data Warehouse* no tienen la inmediatez de los que pueden obtenerse mediante un eficiente sistema de información operacional, por lo general mediante los *Data Warehouse* o *Almacenes de datos* hemos de esperar el ahorro de gastos motivados por los cambios que puedan sugerirse en la gestión de nuestra empresa en el medio y largo plazo. [18]

### 2.5.8 ¿Por Qué Usar Un Data Warehouse?

- La información sumariada es almanezada en el D.W.
- Obtiene respuestas en tiempos razonables.
- Analiza desde una perspectiva en el tiempo con la información histórica que se brinde.
- Nos permite tener fuentes externas para ayudar a nuestra información.

La información proveniente de fuentes operacionales es transformada y limpiada para lograr consistencia.

Un Sistemas de *Data Warehousing* incluye funcionalidades tales como:

1. **Integración de Bases De Datos Heterogéneas** (relacionales, documentales, geográficas, archivos, etc.). [27]

2. **Ejecución de Consultas Complejas no Predefinidas** visualizando el resultado en forma de gráfica y en diferentes niveles de agrupamiento y totalización de datos.

3. **Agrupamiento y Desagrupamiento de Datos en Forma Interactiva.**

4. **Análisis de Problemas en Términos de Dimensiones.** Por ejemplo, permite analizar datos históricos a través de una dimensión de tiempo.

5. **Control de Calidad de Datos** para asegurar, no solo la consistencia de la base, sino también la relevancia de los datos en base a los cuales se toman las decisiones.

En primer lugar, *DW* no es un producto que pueda ser comprado en el mercado, sino más bien un concepto que debe ser construido.

El objetivo principal es satisfacer los requerimientos de información internos de la empresa para una mejor gestión, con eficiencia y facilidad de acceso. [38] [24] [27]

*Data Warehousing*, o *Almacenamiento de Datos* es el proceso que *facilita la creación y explotación* de un *Data Warehouse*. *Data Warehousing*, es un proceso complejo, que *hace real la Gestión del Conocimiento*.

*Data Warehousing* esta compuesto por un conjunto de procesos. Por citar los más conocidos, formarían parte de este proceso la creación del *Data Warehouse*, su explotación analítica mediante técnicas de *OLAP* (*On-line analytical processing*) o técnicas de *Data Mining*.

*OLAP* o procesamiento analítico en línea, es el paradigma de análisis multidimensional de un *Data Warehouse*. Este concepto se contraponen al de *OLTP* o procesamiento transaccional en línea, que es el empleado por los sistemas fuentes para optimizar la recogida de información. En pocas palabras, mientras *OLTP* se encarga de procesar óptimamente multitud de pequeñas transacciones de captura de información (su entrada, modificación o eliminación), *OLAP* se dedica al análisis de enormes cantidades de información; por ende realizará pocas transacciones, pero éstas procesaran un volumen muy superior (cientos de miles de registros).

### 2.5.9 Los Requerimientos De Un Data Warehouse

Existen muchas formas de considerar un *Data Warehouse*. Las siguientes son algunos ejemplos:

- El *Data Warehouse* ofrece una serie de características y funciones para implementar procesos empresariales y enlazarlos con otros procesos fuera del ámbito del *Data Warehouse*. En forma muy similar a otros sistemas

de aplicaciones empresariales, se requiere que el *Data Warehouse* proporcione al usuario final estos conjuntos prescritos de características y funciones del modo más eficiente posible.

- Un Data Warehouse. Almacena información resumida que se organiza de acuerdo con temas empresariales, para analizar la información con más facilidad. La carga de mostrar, organizar y reportar la información que guarda el Data Warehouse corresponde a las herramientas que deben incorporarse en el Data Warehouse. En esta visión, *un Data Warehouse tiene una capacidad latente* que sólo se vuelve útil cuando las herramientas de análisis y reporte se aplican con inteligencia a los datos que conserva el Data Warehouse.
- Es necesario que el *Data Warehouse organice grandes cantidades de información de manera compacta y eficiente*. También se requiere que proporcione técnicas para resumir, a fin de que los usuarios finales comprendan las lecciones de los antecedentes con más facilidad
- En ocasiones. *Entrega información operacional copiando información de los sistemas operacionales de bases de datos*. En este caso, se requiere que el Data Warehouse distribuya información operacional de manera eficiente.

Es evidente que *los requerimientos de un Data Warehouse son tan variados y diversos* como las clases de usuario que lo utilizan para obtener beneficios empresariales. Por lo tanto, es necesario clasificar los requerimientos del Data Warehouse utilizando técnicas clásicas. [7] [23]

### 2.5.10 Arquitectura Datawarehouse

#### 2.5.11 Componentes y Estructuras

La arquitectura lógica de un sistema de Data Warehousing es del tipo mostrado en la figura 2.12 de la pagina 29. [13] [25] Un Sistema de Data Warehousing consta de tres niveles: (1) *Bases De Datos Fuentes* (de producción e históricos), (2) *Una Base De Datos con Datos Resumidos* extraídos de las bases de producción (el Data Warehouse), e (3) *Interfaces Orientadas a Usuarios que Extraen Información Para La Toma De Decisiones*. Las clásicas son: Análisis Multidimensional, Consultas y Reportes y Data Mining. [42] [24]

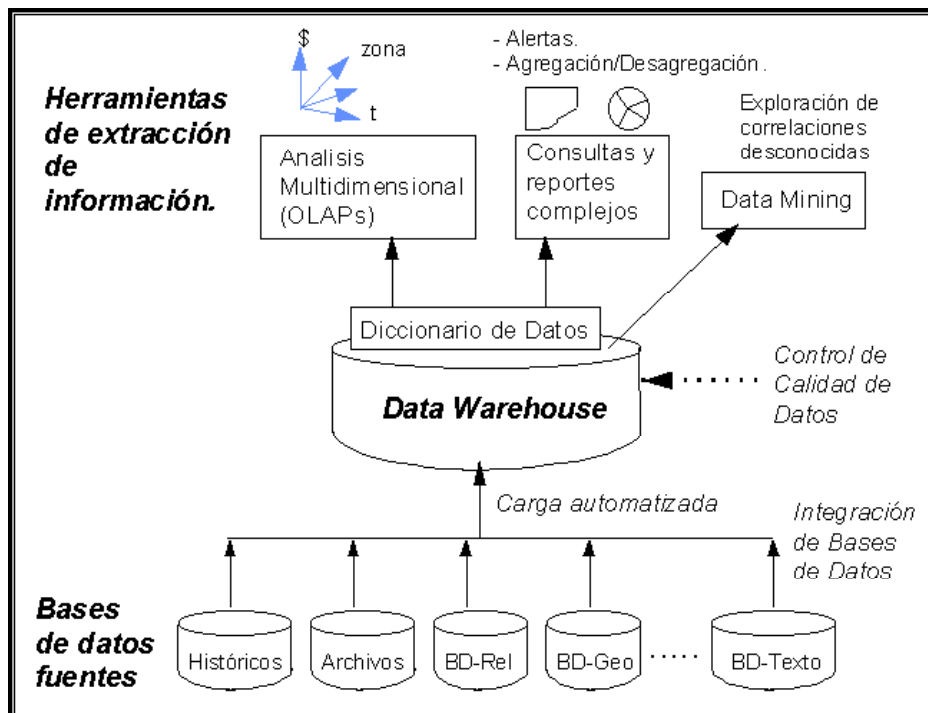


Figura 2.12: Arquitectura Lógica de un Sistema de Data Warehousing

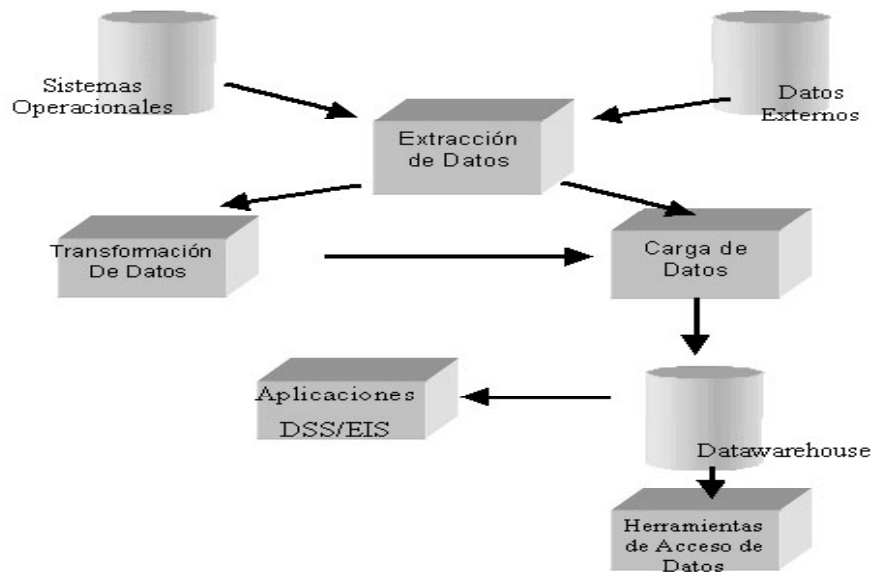


Figura 2.13: Estructura Básica de la Arquitectura del DW

### 2.5.12 La Estructura Básica De La Arquitectura DW Incluye:

1. **Datos Operacionales:** Un origen de datos para el componente de almacenamiento físico DW. [13]

2. **Extracción De Datos:** Selección sistemática de datos operacionales usados para poblar el componente de almacenamiento físico DW.

3. **Transformación De Datos:** Procesos para sumarizar y realizar otros cambios en los datos operacionales para reunir los objetivos de orientación a temas e integración principalmente. [24]

4. **Carga De Datos:** Inserción sistemática de datos en el componente de almacenamiento físico DW. [42]

5. **Datawarehouse:** Almacenamiento físico de datos de la arquitectura DW.

6. **Herramientas De Acceso Al Componente De Almacenamiento Físico DW:** Herramientas que proveen acceso a los datos.

Los pasos 2, 3 y 4 considerados en la figura anterior, conforman el proceso conocido como *ETT* (*Extracción, Transformación y Transporte*).

### 2.5.13 ¿Cómo Trabaja El Data Warehouse?

- Extrae la información operacional.
- Transforma la operación a formatos consistentes.
- Automatiza las tareas de la información para prepararla a un análisis eficiente.

### 2.5.14 ¿En Qué Podemos Usarlo?

- Manejo de relaciones de marketing.
- Análisis de rentabilidad.
- Reducción de costos.

### 2.5.15 Impactos DW

El éxito de *DW* no está en su construcción, sino en usarlo para mejorar procesos empresariales, operaciones y decisiones. [42]

Posicionar un *DW* para que sea usado efectivamente, requiere entender los impactos de implementación en los siguientes ámbitos:

#### Impactos Técnicos De DW.

Considerando las etapas de construcción, soporte del *DW* y soporte de sistemas operacionales, se tienen los siguientes impactos técnicos:

**Nuevas Destrezas De Desarrollo:** Cuando se construye el *DW*, el impacto más grande sobre la gente técnica está dada por la curva de aprendizaje, muchas destrezas nuevas se deben aprender, incluyendo:

- a) Conceptos y estructura DW.

b) *El DW introduce muchas tecnologías nuevas* (ETT, Carga, Acceso de Datos, Catálogo de Metadatos, Implementación de DSS/EIS ), y cambia la manera que nosotros usamos la tecnología existente. Nuevas responsabilidades de soporte, nuevas demandas de recursos y nuevas expectativas, son los efectos de estos cambios.

c) Destrezas de diseño y análisis donde los requerimientos empresariales no son posibles de definir de una forma estable a través del tiempo.

d) Técnicas de desarrollo incremental y evolutivo. [24]

e) Trabajo en equipo cooperativo con gente de negocios como participantes activos en el desarrollo del proyecto.

### **Impactos Empresariales.**

- Procesos Empresariales y Decisiones Empresariales.

Se deben considerar los beneficios empresariales potenciales de los siguientes impactos:

a) *Los Procesos de Toma de Decisiones pueden ser mejorados* mediante la disponibilidad de información. Decisiones empresariales se hacen más rápidas por gente más informada. [13]

b) *Los procesos empresariales pueden ser optimizados.* El tiempo perdido esperando por información que finalmente es incorrecta o no encontrada, es eliminada.

c) Conexiones y dependencia entre procesos empresariales se vuelven más claros y entendibles. Secuencias de procesos empresariales pueden ser optimizados para ganar eficiencia y reducir costos.

d) Procesos y datos de los sistemas operacionales, así como los datos en el *DW*, son usados y examinados. Cuando los datos son organizados y estructurados para tener significado empresarial, la gente aprende mucho de los sistemas de información. Pueden quedar expuestos posibles defectos en aplicaciones actuales, siendo posible entonces mejorar la calidad de nuevas aplicaciones.

- Comunicación e Impactos Organizacionales.

Apenas el DW comienza a ser fuente primaria de información empresarial consistente, los siguientes impactos pueden comenzar a presentarse:

a) *La gente tiene mayor confianza en las decisiones empresariales que se toman.* Quienes toman las decisiones como los afectados conocen que está basada en buena información.

b) *Las organizaciones empresariales y la gente de la cual ella se compone queda determinada por el acceso a la información.* De esta manera, la gente queda mejor habilitada para entender su propio rol y responsabilidades como también los efectos de sus contribuciones; a la vez, desarrollan un mejor entendimiento y apreciación con las contribuciones de otros.

c) *La información compartida conduce a un lenguaje común, conocimiento común, y mejoramiento de la comunicación en la empresa.* Se mejora la confianza y cooperación entre distintos sectores de la empresa, viéndose reducida la sectorización de funciones.

d) *Visibilidad, accesibilidad, y conocimiento de los datos producen mayor confianza en los sistemas operacionales.*

### 2.5.16 Costos v/s Valor De DW

En todo proyecto es importante e inevitable realizar un análisis desde la perspectiva Costo/Valor [25].

A grandes rasgos, los costos asociados a un proyecto *DW* incluyen el costo de construcción y, la mantención y operación una vez que está construido. En cuanto al valor, éste considera, el valor de mejorar la entrega de información, el valor de mejorar el proceso de toma de decisiones y el valor agregado para los procesos empresariales [27].

### 2.5.17 Costos De Un DW

#### Costos De Construcción

Los costos de construir un DW son similares para cualquier proyecto de tecnología de información. Estos pueden ser clasificados en tres categorías:

**RRHH:** La gente necesita contar con un enfoque fuerte sobre el conoci-

miento del área de la empresa y de los procesos empresariales. Además es muy importante considerar las cualidades de la gente, ya que *el desarrollo del DW requiere participación de la gente de negocios* como de los especialistas tecnológicos; estos dos grupos de gente deben trabajar juntos, compartiendo su conocimiento y destrezas en un espíritu de equipo de trabajo, para enfrentar los desafíos de desarrollo del DW.

**Tiempo:** Se debe establecer el tiempo no solo para la construcción y entrega de resultados del DW, sino también para la planeación del proyecto y la definición de la arquitectura. La planeación y la arquitectura, establecen un marco de referencia y un conjunto de estándares que son críticos para la eficacia del DW.

**Tecnología:** *Muchas tecnologías nuevas son introducidas por el DW.* El costo de la nueva tecnología puede ser tan sólo la inversión inicial del proyecto.

### Costos De Operación

Una vez que está construido y entregado un DW debe ser soportado para que tenga valor empresarial. Son justamente estas actividades de soporte, la fuente de continuos costos operacionales para un DW. Se pueden distinguir tres tipos de costos de operación:

**Evolutivos:** Ajustes continuos del DW a través del tiempo, como cambios de expectativas y, cambios producto del aprendizaje del RRHH del proyecto mediante su experiencia usando el *DW*.

**Crecimiento:** Incrementos en el tiempo en volúmenes de datos, del número de usuarios del DW, lo cual conllevará a un incremento de los recursos necesarios como la demanda de monitoreo, administración y sintonización del DW (evitando así, un incremento en los tiempos de respuesta y de recuperación de datos, principalmente).

**Cambios:** El *DW* requiere soportar cambios que ocurren tanto en el origen de datos que éste usa, como en las necesidades de la información que éste soporta.

Los dos primeros tipos de costos de operación, son básicos en la mantención de cualquier sistema de información, por lo cual no nos resultan ajenos; sin embargo, se debe tener especial cuidado con los costos de operación por cambios, ya que ellos consideran el impacto producto de la relación del *OLTP*

y del Ambiente Empresarial, con el *DW*.

Resulta esencial para llevar a cabo un proyecto *DW*, tener claridad en la forma que éste se ve afectado por medio de cambios a nivel de *OLTP* como del Ambiente Empresarial; por ello entonces, a continuación se analiza más en detalle este tipo de costos de operación.

### 2.5.18 Valor Del *DW*

El valor de un *DW* queda descrito en tres dimensiones:

1. *Mejorar la Entrega de Información*: Información completa, correcta, consistente, oportuna y accesible. Información que la gente necesita, en el tiempo que la necesita y en el formato que la necesita. [25] [7]

2. *Mejorar el Proceso de Toma de Decisiones*: Con un mayor soporte de información se obtienen decisiones más rápidas; así también, la gente de negocios adquiere mayor confianza en sus propias decisiones y las del resto, y logra un mayor entendimiento de los impactos de sus decisiones. [27]

3. *Impacto Positivo sobre los Procesos Empresariales*: Cuando a la gente se le da acceso a una mejor calidad de información, la empresa puede lograr por sí sola:

- *Eliminar los retardos de los procesos empresariales* que resultan de información incorrecta, inconsistente y/o no existente.
- *Integrar y optimizar procesos empresariales* a través del uso compartido e integrado de las fuentes de información.
- *Eliminar la producción y el procesamiento de datos que no son usados ni necesarios*, producto de aplicaciones mal diseñados o ya no utilizados.

## 2.6 Data Marts

El conocimiento de los meta datos es tan esencial como el conocimiento de los datos del Data Warehouse. Deben incluir dominio, reglas de validación, derivación y transformación de los datos extraídos.

También describen las bases de datos del *Warehouse*, incluyendo reglas de distribución y control de la migración hacia los *Data Marts*. Los procesos que monitorean los procesos del *Warehouse* (como extracción, carga, y uso) crean meta datos que son usados para determinar que tan bien se comporta el sistema [26].

Los meta datos, deberían estar disponibles para los usuarios, para ser usados en sus análisis. Los administradores pueden manejar y proveer el acceso a través de los servicios del repositorio.

El uso efectivo de los *Data Marts* en un ambiente de *Data Warehousing*, es un factor importante para la efectividad del *Warehouse*. *Los Data Marts son diseñados para satisfacer las necesidades específicas de grupos comunes de usuarios* (divisiones geográficas, divisiones organizacionales, etc.).

*Los Data Marts son generalmente, subconjuntos del Data Warehouse*, pero pueden también integrar un número de fuentes heterogéneas, e inclusive ser más grandes, en volumen de datos, que el propio *Warehouse* central.

El concepto *DataMart* es una extensión natural del *Data Warehouse*, y está enfocado a un departamento o área específica, como por ejemplo los departamentos de Finanzas o Marketing. Permitiendo así un mejor control de la información que se está abarcando.

### 2.6.1 Definición de Data Marts

Es un pequeños *Data Warehouse*, para un determinado numero de usuarios, para un area funcional, específica de la compañía. También podemos definir que un *Data Marts* es un subconjunto de una bodega de datos para un propósito específico [26].

Su función es apoyar a otros sistemas para la toma de decisiones.

Los procesos que conforma el datawarehouse son:

1-Extracción.

2- Elaboración.

3-Carga.

4-Explotación.

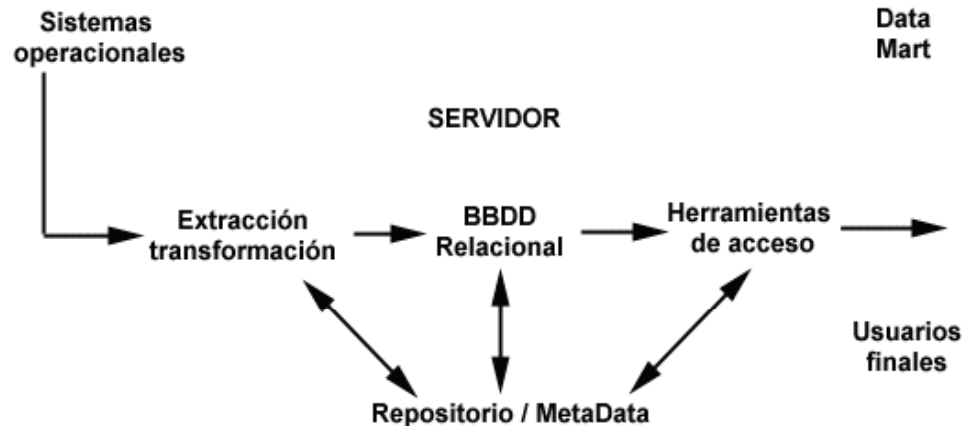


Figura 2.14: Data Marts

### 2.6.2 Arquitectura de Data Marts

### 2.6.3 Crecimiento De Los Data Marts

Los Data Marts están creciendo, llegando a tener tamaños semejantes a los Data Warehouse algunas distinciones entre ellos son todavía importantes:

- \* Un Data Mart está enfocado a una sola área o grupo de usuarios.
- \* Una organización puede tener un sólo Data Warehouse, pero varios Data Marts.
- \* Los Data Marts no contienen información almacenada como datos operacionales.
- \* Como los Data Marts contienen menos información, son más fáciles de entender y navegar.

El reciente crecimiento de los Data Marts, ha generado también, muchos problemas a los usuarios, para acceder a la información de la organización.

- Se pierde performance a medida que aumenta el tamaño de los Data Marts.
- Los usuarios requieren acceso a datos de muchos Data Marts. Los datos

▪ Fuente: Connolly & Begg, 2002

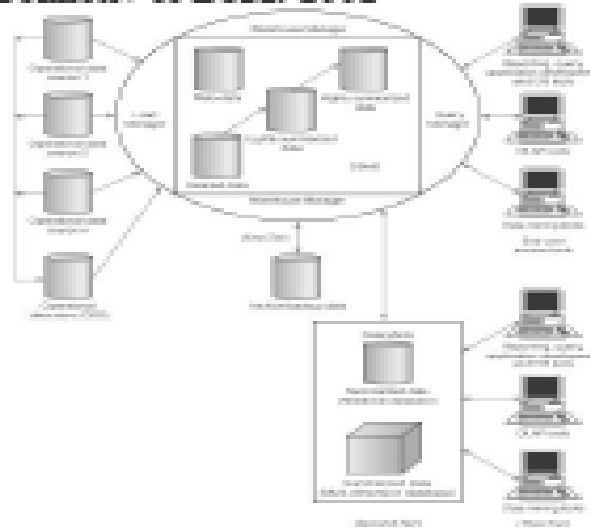


Figura 2.15: Arquitectura de Data Marts

pueden ser replicados entre los Data Marts, pero se requieren mejores soluciones.

- Las compañías no pueden administrar fácilmente muchos Data Marts. Mientras sólo se tiene un Data Warehouse, se pueden tener muchísimos Data Marts.
- Las organizaciones tienen dificultades para construir los Data Marts. Los Data Marts requieren un ciclo de desarrollo muy corto, para una inversión moderada.

Las soluciones de Data Marts, requieren una arquitectura de 3 capas: Los Data Warehouses son la primera capa (opcional), los Data Marts son la segunda capa, y las estaciones de trabajo de usuarios son la tercera.

#### 2.6.4 Administración De Los Data Marts

A medida que el número de Data Marts va creciendo, crece también la necesidad de administración y coordinación central, de actividades. La coordinación

y administración de toda la colección de Data Marts, debe tener un enfoque centralizado.

La administración de los Data Marts, es un área con crecientes requerimientos, como la coordinación, la extracción de los datos, la lectura, los procedimientos de replicación, los procedimientos de backup y recuperación, el manejo de metadatos, la seguridad, y la performance [26].

### 2.6.5 Paquetes De Data Marts

Muchos vendedores han reconocido la necesidad de hacer que los Data Marts sean más fáciles de instalar e implementar que un Data Warehouse corporativo. Los paquetes de Data Marts pueden proveer herramientas convenientes, y de relativamente bajo costo, que pueden ser el puntapié inicial para el desarrollo de los Data Marts. Aunque un Data Mart es relativamente fácil de instalar, hay que tener en cuenta otros aspectos como la lógica de los datos operacionales extraídos, la consistencia en la definición de los datos, y el diseño del Data Mart, para lograr una óptima performance.

### 2.6.6 Data Marts Virtuales y Meta Vistas

Los vendedores están desarrollando el concepto de Data Marts Virtuales para satisfacer la necesidad de los usuarios de acceder a muchos Data Marts, sin necesidad de excesivas replicaciones entre ellos. Los Data Marts Virtuales son vistas de varios Data Marts Físicos, o del Data Warehouse corporativo, brindadas a grupos específicos de usuarios.

Otros vendedores, como Sagent Data Mart Solution, de Sagent Technology Inc., proveen los conceptos de Vista Básica y Meta Vistas. *Una Vista Básica es una representación gráfica de una base de datos que incluye tablas, columnas y joins.* Una vez que una Vista Básica es creada, múltiples Meta Vistas se pueden derivar de ella. *Una Meta Vista es una representación lógica de partes, de una o más Vistas Básicas.* Inicialmente las tablas son desplegadas como categorías, y los campos como partes. Se pueden renombrar o remover categorías o partes de una Meta Vista. Esos cambios no afectan a las Vistas Básicas que la soportan. La Meta Vistas permiten usar una única Vista Básica para presentar diferentes partes de la información a diferentes grupos de usuarios.

### 2.6.7 ¿Por Qué Ha Crecido La Popularidad De Los Data Marts?

Los data marts han reducido drásticamente el costo implícito en la creación y operación de un sistema de soporte a las decisiones. El concepto del Data Marts ha logrado situar la instalación de la tecnología de soporte a las decisiones dentro del rango de posibilidades económicas de un número mucho mayor de usuarios [26].

Wayne W. Eckerson, del Patricia Seybold Group, Inc., indicó en su informe, "Understanding Data Marts," ("Entendiendo a los Data marts"), Agosto de 1996, que los data marts son los preferidos por los departamentos autónomos y las pequeñas unidades comerciales que los emplean para crear sus propios sistemas de soporte de decisiones. Pero los data marts también se han convertido en los favoritos de la mayoría de los departamentos de Sistemas de Información (IS), para crear grandes almacenes de datos centrales. En comparación con los data warehouses, los data marts son más limitados en cuanto a alcance, y se concentran más en un grupo específico de necesidades del usuario.

Los Data Marts permiten una prototificación más rápida para la captura de los requisitos del sistema de soporte a decisiones.

Proyectos que comenzaron originalmente como Data Warehouses evolucionan hasta convertirse en Data Marts.

Este descubrimiento ha permitido reducir el almacenamiento, o archivar información basándose en algunos criterios definidos (como la fecha, por ejemplo), y por lo tanto reducir sus Data Warehouses a Data Marts más concentrados. O la organización puede dividir su Data Warehouse en Data Marts separados para ofrecer:

- Data Marts más pequeños con tiempo de respuesta más rápido.
- Accesos menos complejos para los usuarios a los Data Marts.
- Data Marts diseñados para grupos de usuarios específicos.

El resultado es una arquitectura de Data Warehouse que se visualiza en la figura 2.16 de la pagina 41 de la siguiente manera:

El administrador del Data Warehouse puede observar patrones de acceso a consultas. Las respuestas a las consultas se pueden precomputarizar y almacenar en Data Marts de menor costo, quizás instalados en grupos de trabajo

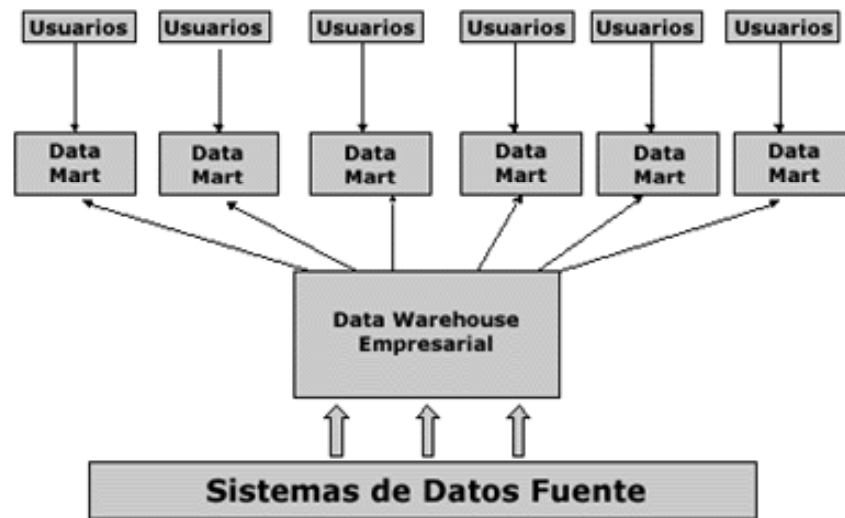


Figura 2.16: Arquitectura

para simplificar el acceso a la red y mejorar todavía más los tiempos de respuesta. Esto mejora el rendimiento y reduce la complejidad y el tamaño de los Data Warehouses/Data Marts. A menudo, los Data Warehouses se convierten en Data Marts en este escenario.

El *Microsoft SQL Server* en Windows NT es una plataforma especialmente popular en la instalación de Data marts. El precio/rendimiento y las capacidades de la plataforma de Microsoft se adecúan perfectamente a las prioridades de selección de los clientes que se encuentran implementando Data Marts.

Intelligent Solutions, Inc., compara los Data Warehouses con los Data Marts de la siguiente manera:

DATA WAREHOUSE	DATA MART
Construido para satisfacer las necesidades de información de toda la empresa.	Construido para satisfacer las necesidades de una función o unidad comercial específica.
Diseñado para optimizar la integración y la administración de los datos fuente.	Diseñado para optimizar la entrega de información de soporte a decisiones.
Administra grandes cantidades de historia a nivel atómico.	Primordialmente se concentra en administrar resúmenes y/o datos de muestreo.
Pertenece a, y se administra por, las organizaciones de Sistemas de Información (SI) de la empresa.	Puede ser propiedad de, y administrado por, el grupo de Sistema de Información (SI) en la Línea del Negocio.

Figura 2.17: Data Warehouse Vs Data Marts

## Capítulo 3

# Herramientas Para La Toma De Decisiones

### 3.1 DB2 Business Intelligence - OLAP

### 3.2 “On-Line Analytical Processing”

#### 3.2.1 ¿Qué Es OLAP?

OLAP es un estándar para ”On-Line Analytical Processing”. En contraste al más familiar OLTP (”On-line Transaction Processing”), OLAP describe la tecnología asociada al acceso y análisis de datos en línea, OLAP se ha convertido en un sinónimo de bases de datos multidimensionales mediante las cuales se provee de una tecnología para el cálculo y análisis requerido por las aplicaciones analíticas para el ”*Business Intelligence*”.

El OLAP describe la tecnología asociada al acceso y análisis de datos en líneas. [14]

Podemos decir que:

- Los servidores OLAP son una tecnología superior para las aplicaciones ”*Business Intelligence*”.
- Las aplicaciones OLAP permiten realizar consultas a un nivel agregado

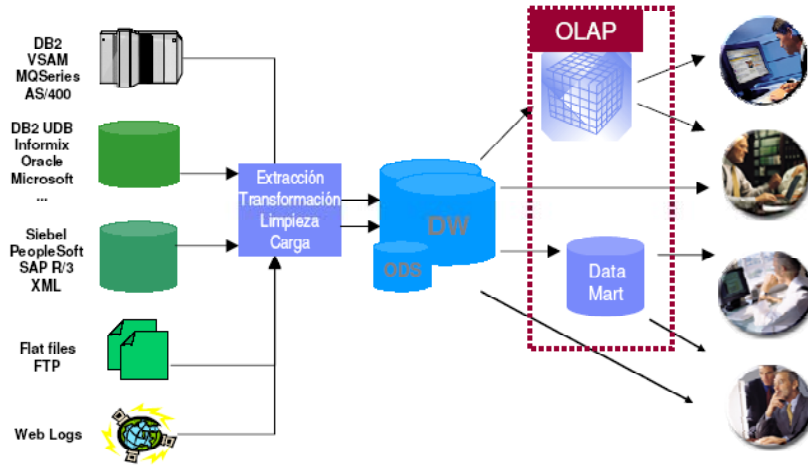


Figura 3.1: Estructura del OLAP dentro de BI.

de la información. Como totales de ventas por línea de producto, región y vendedor.

- Las bases de datos OLAP son optimizadas para el análisis.
- Consultas rápidas y consistentes a cualquier nivel de sumarización de la información.
- Las bases de datos OLAP son alimentadas de distintas fuentes de información.

### 3.2.2 Las Herramientas OLAP (On-Line Analytical Processing) Son Más Genéricas.

\* Funcionan sobre un sistema de información (transaccional o almacén de datos).

\* Permiten realizar agregaciones y combinaciones de los datos de maneras mucho más complejas y ambiciosas, con objetivos de análisis más estratégicos.

\* Están basadas, generalmente, en sistemas o interfaces multidimensionales,

\* Utilizando operadores específicos (además de los clásicos): drill, roll, pivot, slice & dice.

\* El resultado se presenta de una manera matricial o híbrida.

Las herramientas de OLAP presentan al usuario una visión multidimensional de los datos (esquema multidimensional) para cada actividad que es objeto de análisis. [3] [2]

El usuario formula consultas a la herramienta OLAP seleccionando atributos de este esquema multidimensional sin conocer la estructura interna (esquema físico) del almacén de datos.

La herramienta OLAP genera la correspondiente consulta y la envía al gestor de consultas del sistema (p.ej. mediante una sentencia SELECT).

### 3.2.3 Las Herramientas De OLAP Se Caracterizan Por:

- \* Ofrecer una visión multidimensional de los datos (matricial).
- \* No imponer restricciones sobre el número de dimensiones.
- \* Ofrecer simetría para las dimensiones.
- \* Permitir definir de forma flexible (sin limitaciones) sobre las dimensiones: restricciones, agregaciones y jerarquías entre ellas.
- \* Ofrecer operadores intuitivos de manipulación: drill-down, roll-up, slice-and-dice, pivot.
- \* Ser transparente al tipo de tecnología que soporta el almacén de datos (ROLAP o MOLAP).

### 3.2.4 OLTP v/s OLAP: Dos Mundos Diferentes

De acuerdo como se entiendan las diferencias entre estos dos tipos de sistemas uno gana un mejor entendimiento de OLAP. Esto es muy importante en especial para diseñadores, ya que ellos necesitan ver estas diferencias para poder llevar a cabo de mejor manera un proyecto de esta naturaleza. [2] [16]

OLAP o Procesamiento Analítico en Línea, es el paradigma de análisis multidimensional de un Data Warehouse. Este concepto se contrapone al de OLTP o Procesamiento Transaccional en Línea, que es el empleado por los sistemas fuentes para optimizar la recogida de información. En pocas palabras, mientras OLTP se encarga de procesar óptimamente multitud de pequeñas

transacciones de captura de información (su entrada, modificación o eliminación), OLAP se dedica al análisis de enormes cantidades de información; por ende realizará pocas transacciones, pero éstas procesaran un volumen muy superior (cientos de miles de registros). Ambos paradigmas de acceso a la información, son contrapuestos, pero complementarios. OLTP se encargará de la recogida de la información en un modelo normalizado de Base de Datos, optimizado para procesar las transacciones en menos de un segundo, y la OLAP se dedicará al análisis de esa información, en un modelo multidimensional que facilitará la exploración y acceso a la información, invirtiendo minutos en obtener dichas vistas multidimensionales de los indicadores de rendimiento de la compañía.

### 3.2.5 ¿Cuál Es La Diferencia Entre OLAP y Minería De Datos?

Las herramientas OLAP proporcionan facilidades para “manejar” y “transformar” los datos.

- Producen otros “datos” (más agregados, combinados).
- Ayudan a analizar los datos porque producen diferentes vistas de los mismos.

Las herramientas de Minería de Datos:

- Son muy variadas: permiten “extraer” patrones, modelos, descubrir relaciones, regularidades, tendencias, etc.
- Producen “reglas” o “patrones” (“conocimiento”).

## 3.3 OLTP - “On-line Transaction Processing”.

### 3.3.1 Diferencia Entre Data Warehouse y Bases De Datos Operacionales (OLTP).

Los sistemas tradicionales de transacciones y las aplicaciones de Data Warehousing son polos opuestos en cuanto a sus requerimientos de diseño y sus

características de operación. Es de suma importancia comprender perfectamente estas diferencias para evitar caer en el diseño de un Data Warehouse como si fuera una aplicación de Transacciones en Línea (OLTP).

*Las aplicaciones de OLTP están organizadas para ejecutar las transacciones para los cuales fueron hechos. Por otro lado, un Data Warehouse está organizado en base a conceptos, como por ejemplo: clientes, facturas, productos, etc.*

Otra diferencia radica en el número de usuarios. Normalmente, el número de usuarios de un Data Warehouse es menor al de un OLTP. Es común encontrar que los sistemas transaccionales son accedidos por cientos de usuarios simultáneamente, mientras que los Data Warehouse sólo por decenas. *Los sistemas de OLTP realizan cientos de transacciones por segundo mientras que una sola consulta de un Data Warehouse puede tomar minutos.* Otro factor es que, frecuentemente, los sistemas transaccionales son menores en tamaño a los Data Warehouses, esto es debido a que un Data Warehouse puede estar formado por información de varios OLTP's.

Existen también diferencia en el diseño. El OLTP normalmente está formado por un número mayor de tablas, cada una con pocas columnas, mientras que en un Data Warehouse 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 con día, mientras que los Data Warehouse son actualizados en batch de manera periódica.

Las estructuras de los OLTP son muy estables, rara vez cambian, mientras las de los Data Warehouses sufren cambios constantes derivados de su evolución.

Un Data Warehouse es diferente de las Bases de Datos Operacionales que soportan las aplicaciones de un Procesamiento de Transacción en Línea OLTP, On-Line Transaction Processing. El Data Warehouse es lo siguiente:

- Está orientado a una materia.
- Administra grandes cantidades de información.
- Guarda información en diversos medios de almacenamiento.
- Comprende múltiples versiones de un esquema de base de datos.

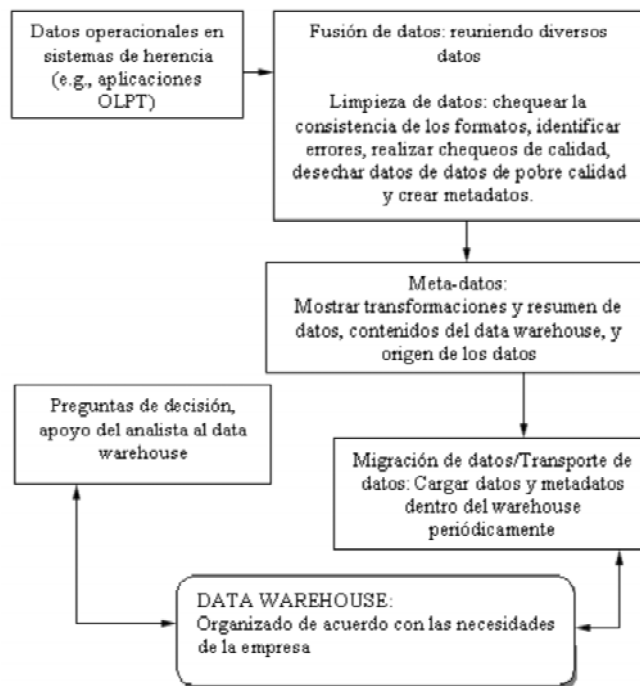


Figura 3.2: Data Warehouse y OLTP.

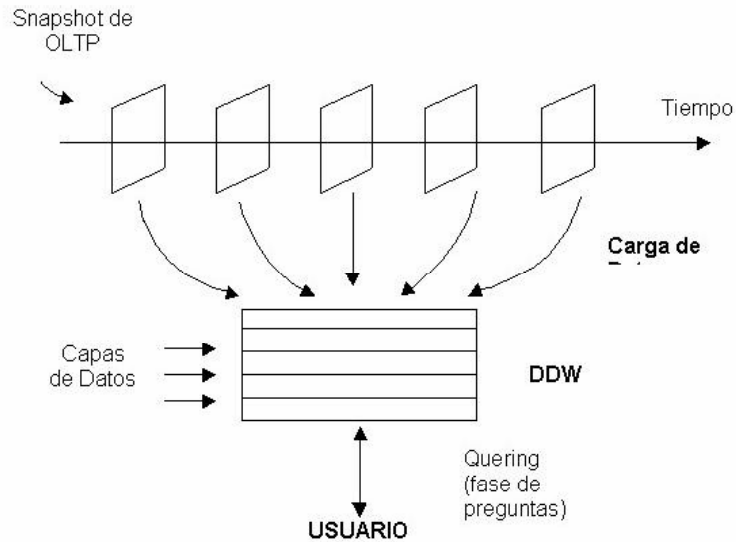


Figura 3.3: Esquema de Relación entre OLTP Y DDW.

- Condensa y agrega información.
- Integra y asocia información de muchas fuentes de información.

### 3.3.2 Relación Entre Ambos Esquemas.

Se definen dos fases en el Data Warehouse Dimensional (DDW): carga y consultas. En la primera se carga la *snapshot* para un tiempo dado, y en la segunda se pueden hacer consultas en la base de datos sin que haya cambios en ella durante el proceso.

Para visualizar fácilmente la relación entre ambos esquemas (OLTP y DDW), se muestra en la figura 3.3 de la pagina 49

### 3.3.3 ¿Una Máquina o Dos?

Es mucho más habitual encontrar separado el DW del OLTP, debido a factores bien específicos y de considerable relevancia para su desempeño. Los

argumentos que favorecen el tener en máquinas separadas el OLTP del DW son:

- El DW tiene una significativa y altamente variable demanda de recursos, por lo tanto puede entorpecer considerablemente el desempeño del OLTP.
- Los sistemas en cuestión son configurados muy diferentemente.
- A veces los datos del DW son integrados de múltiples sistemas OLTP remotos, y por lo tanto el DW puede verse como un conjunto de recursos centralizados. Es obvio entonces que estén físicamente separados ambos sistemas.

La razón para tenerlos en la misma máquina está en el hecho de que al ser la estructura básica del DW distinta a la del OLTP, el dato tiene que ser copiado y reestructurado por el DW. Para ahorrar envíos de datos entre máquinas, es mejor realizar este proceso dentro de una sola.

## Capítulo 4

# CUBE VIEWS

### 4.1 DB2 Business Intelligence - DB2 Cube Views

### 4.2 ¿Que Es DB2 Cube Views?

- Acelerador de Herramientas *OLAP*.
  - Acelera el rendimiento de las preguntas y aplicaciones *OLAP*.
- Genera un eficaz y sofisticado *DB2 MQT*.
- Herramienta de Productividad *OLAP*.
  - Interfase de facil uso (GUI).
- Para crear metadatos y un rápido despliegue de *OLAP*.
  - Compartir los *metadatos OLAP* con las *herramientas analíticas de BI*.
- Define *metadatos OLAP* solo una vez.
  - Simplifique la explotación de las características avanzadas de DB2 BI.
- Resume las tablas de dirección.

- Funciones Analíticas.
- Almacén Centralizado de *metadatos OLAP*.
  - Guardar y Mantner los metadatos *OLAP* dentro del Data warehouse.
- *API* abierto para los *metadatos OLAP*.
  - Cualquier herramienta Analítica puede llevar a cabo el acceso a los *metadatos Cube Views*.
- Proveedor de la Infraestructura *OLAP*.
- Analítica y de reportes para *DB2 Alphablox*.

Integre las aplicaciones de la inteligencia de negocios (comercial o BI) con el almacén de los datos fácilmente, compartiendo el metadata entre la base de datos correlativa y las aplicaciones de inteligencias de negocio en lugar de manejar cada aplicación individualmente, usted puede modelar los datos una vez en el almacén, y desplegar ese modelo con cada aplicación. [38] [5] [39] [40]

### 4.3 ¿Que Es DB2 MQT?

DB2 MQT (Materialized Query Table) es un tipo especial de tablas:

- Pre-agregado de los datos correlativos. [16]
- Mejora dramáticamente el interrogante de la ejecución para las herramientas y aplicaciones de OLAP.

#### 4.3.1 Compartiendo Los Metadatos OLAP Con El DB2 Cube Views.

Ver figura 4.1 de la pagina 53.

A través de los metadatos *API* del *Cube Views* y la funcionalidad de Importar/Exportar, el *Cube Views* habilita a compartir bi-direccionalmente

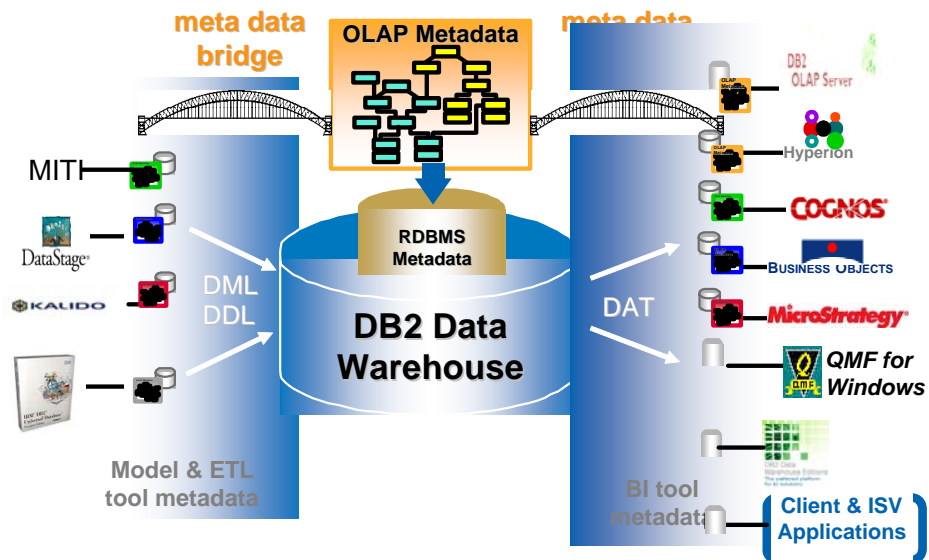


Figura 4.1: Metadatos Olap y DB2 Cube Views

los catálogos de los metadatos *OLAP* con cualquier número de herramientas analíticas, incluyendo: [40] [15]

- OLAP server tools.
- OLAP reporting tools.
- ETL tools.
- Metadata Management tools.
- Custom applications.

#### 4.4 ¿Por Qué Cube Views?

Las principales razones para utilizar *Cube Views* son las siguientes:

- Metadatos *OLAP* centrados en DB2.

- Los metadatos *OLAP* están dentro de la Base De Datos.
- Comparte metadatos con las *herramientas analíticas de BI*.
- El acceso de los datos es transparente.
  - Las *herramientas Analíticas BI* no necesitarán modificar su generación de preguntas para acceder a DB2.
- Integración con *DB2 Alphablox*.
- Vendedores de BI que adoptan *Cube Views*:
  - Brio, Business Objects, Cognos, Crystal, Hyperion, Kalido, MicroStrategy, Ascential.
- Herramientas de Productividad DBA.
  - Modelo basado en la creación de *MQT*.
  - Quick Started Wizard.
  - Interfaz fácil de usar.
- Creación de *MQT* perfeccionados.
  - Cube Views Asesor de Optimización.
- Generación de super agregados *MQT*'s con Data Sampling, Statistics, Query Type, Disk Space, OLAP Model.
- Rollup, Grouping Sets.
- Validación de las reglas *MQT*:
  - Primary Keys, Foreign Keys, Null columns.
- Creación de índices de *MQT*.

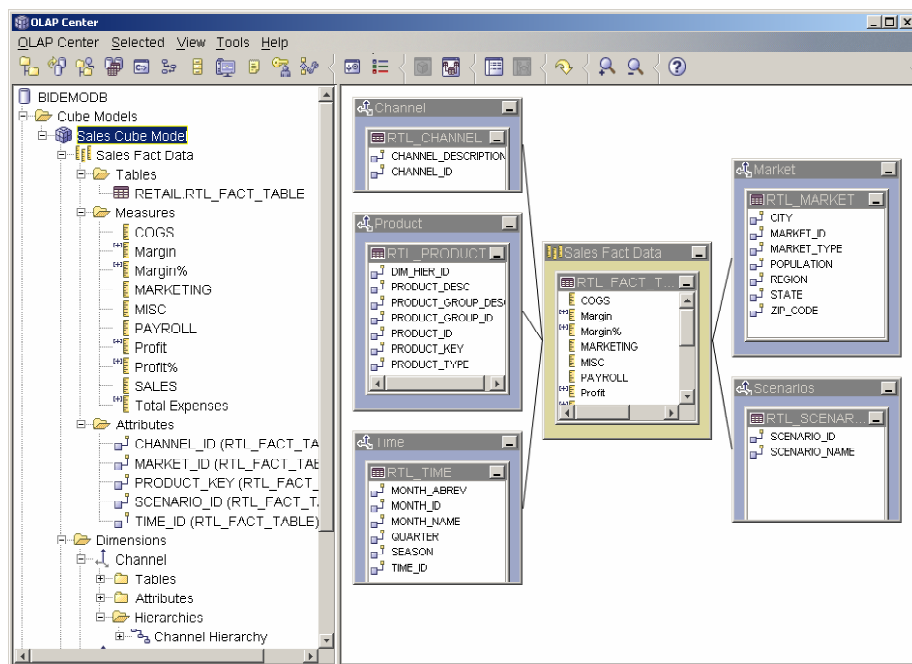


Figura 4.2: Modelo de cubo

### 4.4.1 Modelo Del Cubo

El *Cube Model* del *DB2 Cube Views* es una representación lógica de un esquema de estrella o un esquema de copo de nieve (ver gráfico 4.2 de la pág. 55). El *Cube Model* es una agrupación de objetos de dimensión pertinente, alrededor de un objeto Facts central. Cada dimensión puede tener múltiples jerarquías.

La información estructural sobre cómo unir las tablas que usan Facts y los objetos de dimensión, son referenciados por el Cube Model. El Cube Model, también tiene la información suficiente para recuperar los datos OLAP. Otros reportes y herramientas de OLAP que entienden al Cube Model y pueden desplegar múltiples vistas de una dimensión específica, pueden beneficiar a éste con su uso.

Los *Cube Model* definen un juego complejo de relaciones y pueden usarse para exponer selectivamente hechos pertinentes y dimensiones a una aplicación. [15]

Cada objeto Join que conecta una dimensión a un objeto Facts Central es guardado con la dimensión correspondiente como un conjunto.

Los subconjuntos de los componentes del *Cube Model* pueden ser usados por muchos cubos para diferentes propósitos de análisis.

Se puede crear un *Cube Model* vacío en el OLAP Center usando el *Cube Model Wizard*. Un *Cube Model* vacío no tiene un objeto Facts o cualquier dimensión. Con los *Cube Model Wizard* en el OLAP Center, se puede completar el *Cube Model* creando los objetos Facts y una o más dimensiones.

También puede crear a un *Cube Model* completo usando el *Quick Start Wizard*. *DB2 Cube Views* validará su *Cube Model*, asegurando que tiene los siguientes componentes obligatorios:

- Un Objeto Facts.
- Por lo menos una dimensión.
- Por lo menos una jerarquía correspondiente para cada dimensión.
- Joins (puentes-uniones) entre el objeto Facts existente y las dimensiones.
- Los atributos (cualquiera) que referencien tablas válidas.

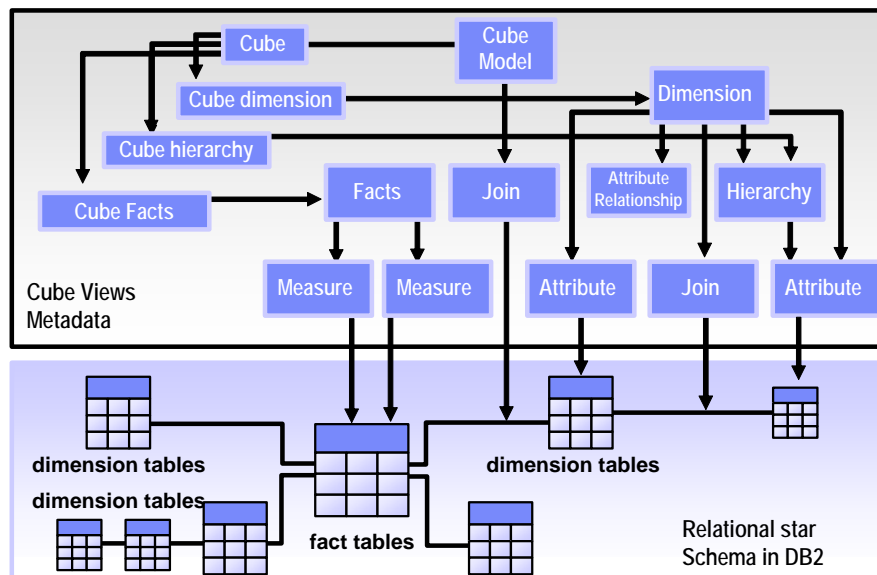


Figura 4.3: The OLAP-aware Database (OLAP)

#### 4.4.2 DB2 Cube Views - The OLAP-aware Database (OLAP)

Los aspectos más significativos se muestran en la figura 4.3 de la pág. 57.

Una jerarquía de cubo es una versión del alcance de una jerarquía que se utiliza en un cubo. Una jerarquía del cubo menciona la jerarquía de la cual se deriva y puede tener un subconjunto de los atributos de la jerarquía del padre. Adicionalmente, un objeto de la jerarquía de cubo menciona las relaciones del atributo que aplican en el cubo. Sólo una jerarquía de cubo se puede definir para una dimensión de un cubo.

Una jerarquía define las relaciones entre un conjunto de uno o más atributos, dentro de una dimensión de un Cube Model. Definir estas relaciones proporciona medios de navegación y cálculos para atravesar una dimensión dada. Se pueden definir múltiples jerarquías para una dimensión de un Cube Model. [16]

La dimensión del objeto del cubo menciona la dimensión de la cual se deriva y la jerarquía pertinente del cubo dado.

La dimensión del objeto proporciona una manera de clasificar un conjunto

de atributos relacionados que describen un aspecto de una medida. Las dimensiones se utilizan en el Cube Model para organizar los datos calculados en los objetos Facts según las categorías lógicas como la Región, Producto, o Tiempo; los atributos relacionados y los joins, para agrupar estos atributos, necesitan ser definidos en las propiedades específicas de la dimensión del objeto. Las dimensiones referencian a una o más jerarquías.

La medida de un objeto define la medida de una entidad y se usa en los objetos Facts. Las medidas se vuelven significantes dentro del contexto de una dimensión.

El objeto “*Join*” une tablas relacionales que son mencionadas por dos objetos metadata. Dos objetos *metadata* pueden unir una o más partes de los atributos de los objetos formando las columnas de las tablas correlativas.

Un cubo Facts tiene un subconjunto de medidas en una lista ordenada desde un objeto Facts específico. Un objeto *Facts* da la flexibilidad a un cubo para alcanzar un modelo de cubo *Facts*. Los grupos de objetos Facts relacionan las medidas que interesan a una aplicación dada. Pueden unirse tablas Facts correlativas múltiples con los atributos específicos para trazar las medidas relacionadas adicionales. El objeto Facts almacena información acerca de los atributos utilizados por la dimensión de los Joins, los atributos y los Joins son usados para trazar las medidas adicionales a través de múltiples tablas de la base de datos.

Un atributo representa la abstracción básica de las columnas de las tablas de base de datos; puede ser más de una columna de datos. Un atributo se define por una expresión SQL que puede ser una asignación simple a una columna de la tabla, puede involucrar columnas múltiples y otros atributos, y puede involucrar todas las funcionalidades de la base de datos subyacente, como el usuario definió las funciones.

### 4.4.3 DB2 Cube Views.

Los principales aspectos se detallan en la figura 4.4 de la pág. 59.

Un esquema estrella es un diseño de la base de datos que consiste en una *tabla Facts* grande y varias tablas de dimensión más pequeña. Generalmente, la tabla Facts tiene millones de registros (filas) de datos mientras que las tablas de dimensiones tienen una magnitud menos -quizás 100 o 10.000 registros-.

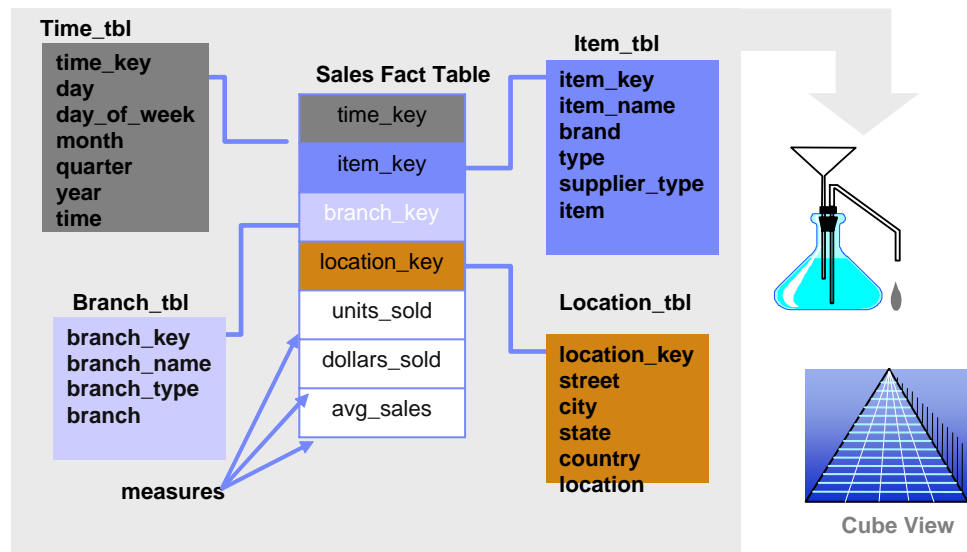


Figura 4.4: DB2 Cube Views

Cuándo se habla acerca de cómo se quiere mirar los datos, se quiere generalmente ver alguna clase de datos agregados. Estos datos se llaman *Measures*. Estas Measures son los valores numéricas que son mensurables y aditivos.

Rastrear las medidas simplemente no es suficiente. Se necesita mirar las medidas que se usan “para las condiciones”, a las cuales las llaman las dimensiones.

Hay casi siempre una dimensión de *tiempo* en algo que se pide. Se puede querer saber las ventas por categoría o por producto.

Por consiguiente, diseñando un esquema estrella, el primer orden de negocio es normalmente determinar lo que se quiere ver (las medidas) y cómo se quiere ver (las dimensiones).

*DB2 Cube Views* opera principalmente desde los sistemas de esquema estrella, pero pueden usarse con los esquemas de copo de nieve. Los esquemas de copo de nieve tienen múltiples tablas Fact y tablas de dimensión vinculadas a cada uno.

Generalmente, varias dimensiones de tablas son compartidas a través de las diferentes tablas *Facts*.

## Performance Advisor

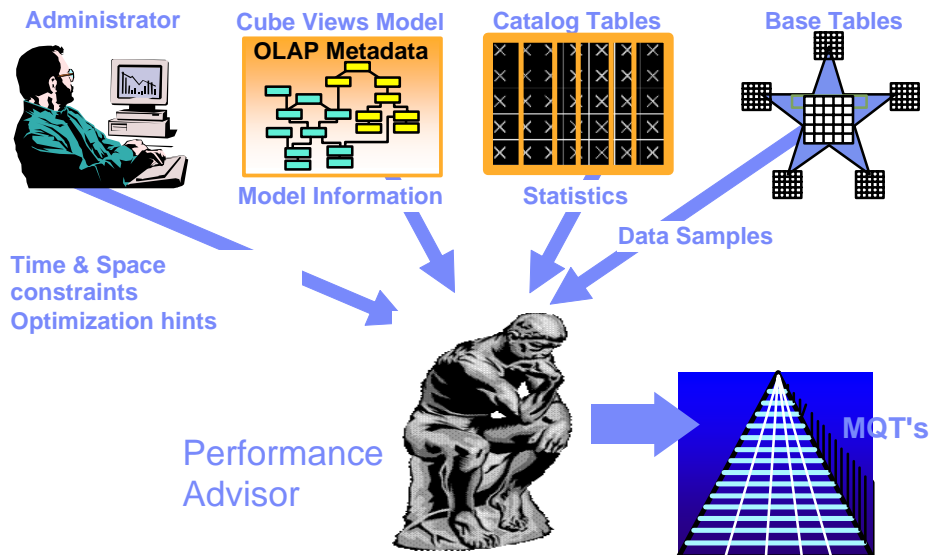


Figura 4.5: Performance Advisor

Lógicamente tiene sentido, pero hay muchas maneras en que esto puede suceder. La integridad referencial es crucial para Cube Views. DB2 permite encender un RI (integridad referencial) que supervisa la opción tal que nunca se puede insertar una fila en la tabla Facts con una “llave extranjera” (foránea) que no se une a una fila real en la tabla de dimensiones.

Existe un optimizador del asistente de diseño en el Cube Views que considera muchos factores formando antes el diseño de los super MQT's (ver figura 4.5 de la pág. 60).

La mayoría de los valores del DB2 Cube Views están en el optimizador. Considera los datos diseñados en el esquema estrella, las estadísticas de las filas y las columnas, el plan multidimensional y la cantidad de espacio en el disco que el DBA desea usar para este MQT.

Mientras es posible que un DBA haga esto en forma manual, es extremadamente complejo. Esto involucra software de optimización que equilibra docenas de variables para un mejor resultado.

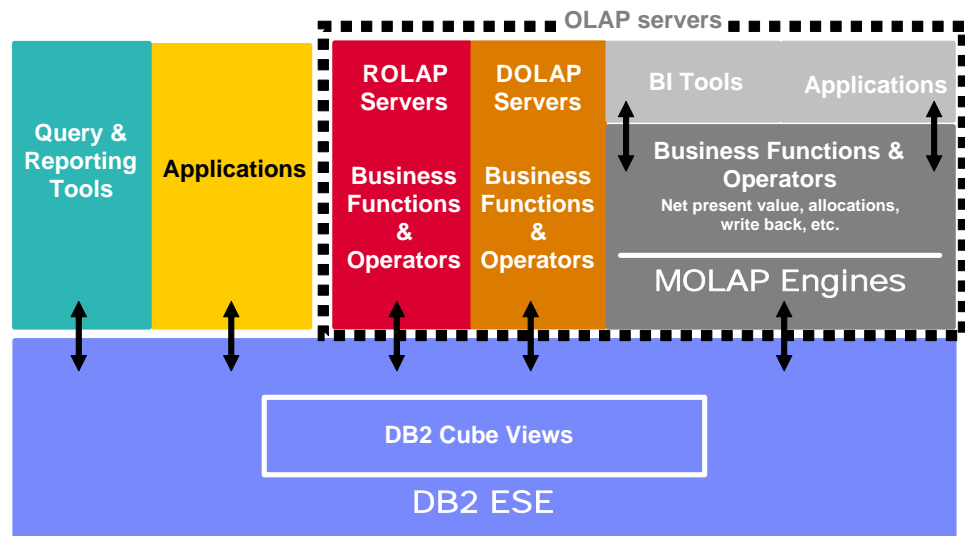


Figura 4.6: DB2 and OLAP ACCELERATOR.

Además, DB2 Cube Views mejorará año tras año en su habilidad de perfeccionar, basado en los compromisos del cliente.

#### 4.4.4 DB2 Cube Views = An OLAP Accelerator

Los principales aspectos a considerar en la comparación se muestran en la figura 4.6 de la pág. 61.

El DB2 Cube Views es un acelerador de OLAP, no un servidor de OLAP. No puede hacer OLAP procesando con DB2 Cube Views. En cambio, se puede acelerar la actuación de varios OLAP frente a herramientas MOLAP, DOLAP, y artefactos de ROLAP.

Esto es porque necesita guardar el software de OLAP y porque las DB2 Cube Views apoyan todos los OLAP fabricados en lugar de competir con ellos.

#### 4.4.5 Cube Views Performance

Se consideran los datos de la figura 4.7 de la pág. 62.

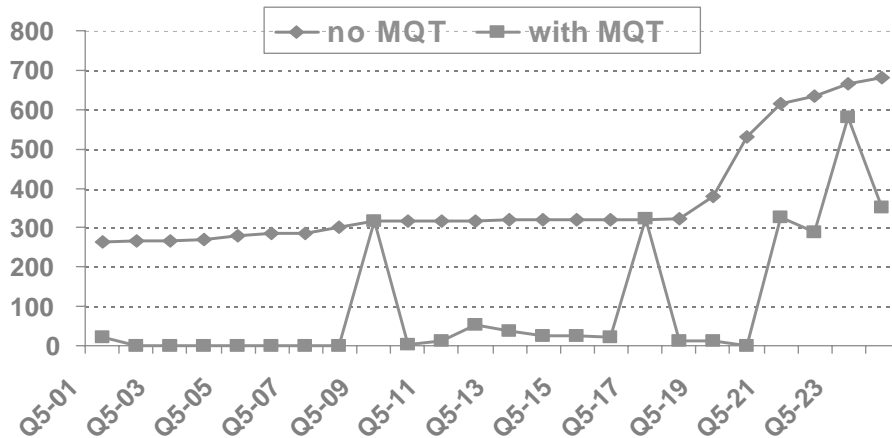


Figura 4.7: Cube Views Performance

No todas las preguntas redireccionan a los MQTs ni deben hacerlo. DB2 Cube Views acelera una multitud de preguntas dentro de los límites del espacio del disco asignado [40].

La base de datos del 5<sup>o</sup> mes era más representativa del trabajo del cliente que la base de datos del 1<sup>er</sup> mes.

Como la base de datos del 1<sup>er</sup> mes, una mayoría de preguntas era capaz de redireccionar a los DB2 Cube Views para generar MQT, mientras producía un 90%-99% de reducciones en tiempo pasado.

DB2 usa las técnicas tradicionales simplemente para lograr una buena relación con data warehousing query performance.

Si un DBA quisiera acelerar todas las preguntas con DB2 Cube Views, con unas visitas al OLAP Center Optimization Advisor y mucho más espacio del disco lo haría. En estos casos, los límites de espacio del disco permitieron MQTs, así como el levantamiento de medidas y exigencias eliminó la generación de MQTs, a la profundidad que habría acelerado todas las preguntas.

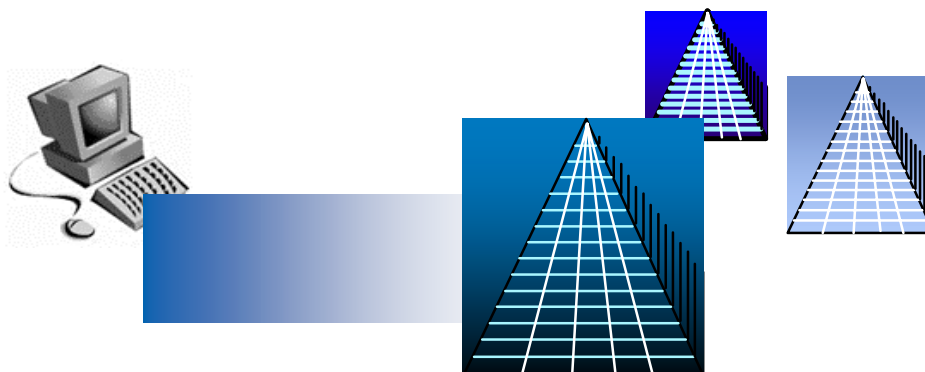


Figura 4.8: DB2 Cube Views

#### 4.4.6 DB2 Cube Views – Beneficios

Se considera la figura 4.8 de la pág. 63.

Los principales beneficios son los siguientes:

- Rápidos informes de resumen.
- OLAP mejora la performance para OLAP Híbrido:
  - MOLAP, DOLAP, y herramientas de ROLAP.
- Se prefiere que las herramientas LOB trabajen con DB2 Cube Views.
- OLAP está disponible para cualquier Aplicación:
  - BI, web, batch applications entregan el análisis multidimensional de DB2.
  - Soporta BI en tiempo real para las decisiones inteligentes en línea.
  - Nuevas aplicaciones habilitadas por los usuarios.

El valor primario de DB2 Cube Views es la velocidad. Los informes multidimensionales proporcionados por las herramientas de usuario final y artefactos de OLAP logran buena performance (dependiendo de muchas variables).

Esto ayuda a preservar y extender el valor de los servidores de OLAP existentes. Por ejemplo, se entregarán ahora más rápidamente los informes de Microstrategy con menos uso de recursos en el servidor de almacén de datos.

Siendo 100% basado en DB2 SQL, no hay necesidad de comprar o aprender las nuevas herramientas. Esto reduce costos y complejidad para la mayoría de las organizaciones.

Algunos de los principales beneficios son los siguientes:

- La facilidad de uso para codificar (SQL).
- Facilidad para utilizar esquemas estrella existentes o esquemas de copo de nieve.
- Facilidades de los MQTs para usar el paralelismo de DB2 rápidamente.
- Reducción del trabajo de los DBA:
  - La exportación e importación de metadata evita rekeying de planes.
- Compartición de las definiciones de OLAP vía metadata bridge:
  - Incremento de las velocidades de las herramientas de BI.
- Reducción de repetición y errores en múltiples almacenes de metadata.

Los informes rápidos mejorarán la satisfacción y el interés del usuario final de usar los datos con warehouse.

Además, las Vistas del Cubo reducirán el uso de recursos de disco de E/S y CPU.

Desde que las *DB2 Cube Views* están 100 % basadas en SQL, no hay ninguna nueva herramienta o habilidad para ser desarrollada. Esto ahorra los costos del diseñador y asegura que la mayoría del personal de programación puede usar las *DB2 Cube Views* inmediatamente.

DB2 Cube Views trabajan mejor con los esquemas de estrella y esquemas de copo de nieve, que son común en el almacenaje de los datos.

La habilidad de importar o exportar el diseño del metadata puede reducir significativamente las tareas de DBA.

## Cube Views Model

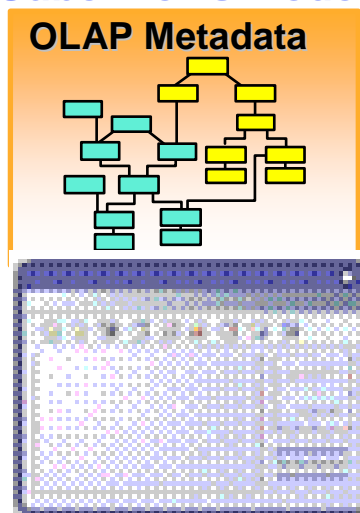


Figura 4.9: Modelo Cube Views



# Bibliografía

- [1] Job Vigía Abukari Kobana. *Business Intelligence in Action*. <http://www.netlibraryebooktoolkit.com/content/library/patron/support/3>, 2003.
- [2] Jhon Wiley Alan Simon and Sons. *Data Warehouse, Data Mining and OLAP*. USA, 1997.
- [3] Mc Graw Hill Alex Berson, Stephen J. Smith. *Data Warehouse, Data Mining and OLAP*. USA, 1997.
- [4] María Mercedes Marqués Andrés. *Clasificación de los Sistemas de Gestión de Bases de Datos*. <http://www3.uji.es/mmarmarques/f47/apun/node31.html>, 2005.
- [5] Mariano Berruezo. *DB2 V8 Ú SQL V: Reglas y Restricciones*. IBM-Bs.As.-Argetina, 2005.
- [6] James A. O´ Brien. *Sistemas de Información Gerencial*. Editorial Nomos S.A., Argentina, 2003.
- [7] Casares C. *Data Warehousing*. <http://www.programacion.com/bbdd/tutorial/warehouse/15/>, 2005.
- [8] Goodwin Candice. *Technology. Business Intelligence - Assault on the Data Mountain*. <http://www.netlibraryebooktoolkit.com/content/library/patron/support/3>, 2003.
- [9] Don Chamberlin. *IBM´s Object Relational Database System*. [www.software.ibm.com/data/db2/udb](http://www.software.ibm.com/data/db2/udb), 2005.
- [10] Don Chamberlin. *Soluciones para un Mundo Pequeño*. [www.software.ibm.com/data/db2/udb](http://www.software.ibm.com/data/db2/udb), 2005.
- [11] Don Chamberlin. *Using the New DB2*. [www.software.ibm.com/data/db2/udb](http://www.software.ibm.com/data/db2/udb), 2005.

- [12] A. González del Alba Baraja; V. Yague Galaup; L. Joyanes Aguilar. *Impacto de las Tecnologías en la Gestión de los Sistemas de Información en II Congreso Internacional de Sociedad de la Información y del Conocimiento*. McGraw Hill, Madrid-España, 2003.
- [13] Barry Devlin. *Information Integration-Extending the Data Warehouse*. IBM Press, USA, 2003.
- [14] Jhon Wileyand Sons Eric Thompson. *OLAP Solutios: Building Multidimensional Information Sysmens, Segunda Edición*. USA, 1997.
- [15] Daniel Graham Glen Sheffield. *Exploring Cube Views Metadata-Querying data via Cube Views Metadata-Cube Views Metadata Bridges*. IBM Press, IBM-Bs.As.-Argetina, 2005.
- [16] Daniel Graham Glen Sheffield. *Optimizing OLAP Models-DB2 Cube Views Optimization Advisor-DB2 Materialized Query Tables (MQTs)*. IBM Press, IBM-Bs.As.-Argetina, 2005.
- [17] R. Groth. *A Hands-on Approach for Business Professionals*. Prentice Hall, NJ-USA, 1998.
- [18] IBM Software Group. *Enterprise Data Warehousing whit DB2: The 10 Terabyte TPC-H Benchmark*. IBM Press, USA, 2003.
- [19] IBM. *DB2 Universal Database (DB2 UDB)*. <http://www.ibm.com>, 2005.
- [20] María José Ramírez Quintana José Hernández Orallo. *Extracción Automática de Conocimiento en Bases de Datos e Ingeniería del Software*. España, 2003.
- [21] Manuel Palomar Juan C. Trujilla. *Diseño de Almacenes de Datos*. España, 2002.
- [22] María Cristina López Lina Marcela Cepeda Díaz, Grace Lennis Benavides. *Inteligencia Artificial*. <http://www.monografias.com/trabajos16/la-inteligencia-artificial/>, 2005.
- [23] Torres H.; Visitación M.; Grau A.; Mar M.; Barranco H.; Soldado M. *Bases de datos y Data Warehouse: Herramientas Estratégicas para la Eficacia Comercial*. <http://www-lsi.ugr.es/rosana/investigacion/bd-efsi04.pdf>, 2005.

- [24] José Antonio Rodero Rodero Mario Piattini Velthuis. *Auditoría de los Data Warehouse*. España, 1998.
- [25] Frank Mcguff. *Designing The Perfect Datawarehouse*. <http://www.techguide.com/>, 2005.
- [26] Frank Mcguff. *Putting Metadata To Work in the Warehouse*. <http://www.techguide.com/>, 2005.
- [27] Frank Mcguff. *Building A Decision Support Architecture For Datawarehousing*. <http://www.techguide.com/>, 2006.
- [28] Lic. Ricardo Sánchez Montoya. *Business Intelligence. To BI or not to*. <http://www.monografias.com/trabajos14/bi/bi.shtml>QUEES, 2005.
- [29] José Hernández Orallo. *Almacenes de Datos-Parte I-II-III*. Departamento de Sistemas Informáticos Y computación-Universidad Politécnica de Valencia-España, España.
- [30] Universidad Latina De Panamá. *Sistemas de Bases de Datos*. <http://pdf.rincondelvago.com/trabajos-global/informatica/2>, 2005.
- [31] IBM Press. *IBM DB2 Intelligent Miner for Data: Utilización del Visualizador de Asociaciones*. IBM Press, USA, 1999.
- [32] IBM Press. *IBM DB2 Warehouse Manager Guía de Instalación Version 8*. IBM Press, USA, 2001.
- [33] IBM Press. *IBM DB2 Intelligent Miner for Data: Utilización de Intelligent Miner for Data*. IBM Press, USA, 2002.
- [34] IBM Press. *IBM DB2 Intelligent Miner Visualization: Using the Intelligent Miner Visualizers*. IBM Press, USA, 2002.
- [35] IBM Press. *IBM DB2 Universal Database Versión 8.1 Visión General Rápida: Iniciador*. IBM Press, USA, 2002.
- [36] IBM Press. *DB2 UDB/WebSphere Actuación que Pone a punto la Guía*. IBM Press, USA, 2005.
- [37] Cynthia Roberts. *Sistemas de Bases de Datos*. <http://www3.uji.es/mmarques/f47/apun/node31.html>.

- [38] Eduardo Fontana; Paola Rodríguez. *DB2: Construido para Business Intelligence - DB2 Data Warehouse Center, DB2 Warehouse Manager y Information Catalog Manager*. IBM-Bs.As.-Argetina, 2005.
- [39] Paola Rodríguez. *DB2 Business Intelligence-Business Intelligence Technical Specialist*. IBM Press, IBM-Bs.As.-Argetina, 2005.
- [40] Paola Rodríguez. *DB2 Cube Views Getting Started with MicroStrategy*. IBM Press, IBM-Bs.As.-Argetina, 2005.
- [41] Julio Tejada. *Bases de Datos*. <http://apuntes.rincondelvago.com/trabajos-global/informatica/2/>, 2005.
- [42] Jhon Wiley W. H. Inmon and Sons. *Building the Data Warehouse*. USA, 1996.
- [43] Colin J. White. *IBM Enterprise Analytics for the Intelligent e-Business*. IBM Press, USA, 2001.

# Índice de Materias

- Arquitectura de BI, 13
- Base De Datos
  - Ventajas, 2
- Bases De Datos
  - Arquitectura, 5
  - Estructura, 3
  - Objetivos, 2
- bases de datos, 1
- BI, 10, 11
  - Componentes de BI, 12
  - inteligencia de negocios, 1
  - Que es Bi, 11
  - Que puedo hacer con BI, 12
- BI para DB2 UDB, 10
- Cube Model, 56
- Cube Views, 51, 53
  - Definicion, 51
- Data Marts, 9, 35
  - Administracion, 38
  - Arquitectura, 37
  - Definicion, 36
- Data Warehouse (DW), 19
- Datawarehouse, 26
  - Almacenes De Datos, 21
  - Analisis De Datos Para La Toma De Desiciones, 21
  - Arquitectura, 28
  - Beneficios, 26
  - Componentes y Funciones, 28
  - Costos y Valor De DW, 33
  - Definicion, 24
  - Estructura Básica De La Arquitectura, 30
  - Impactos, 31
  - Problemas, 22
  - Requerimientos, 27
- DB2, 6
  - Característica y Funciones, 6
  - Common Server, 7
  - Funciones Complementarias, 9
  - Introducción al BI, 15
  - Parallel Edition, 7
  - valor estratégico, 7
- DB2 Cube Views, 51
  - Beneficios, 63
  - Esquema Estrella, 58
  - Modelo De Cubo, 56
- DB2 MQT, 52
- DB2 Para BI
  - Arquitectura De ETL, 18
  - Conceptos Generales, 15
  - Datawarehouse, 19
  - Estructura de BI, 15
  - Funcines de BI, 15
  - Que Es ETL, 17
- db2 para bi
  - ETL, 15
- DBMS, 3
  - funciones, 3
- ETML, 10

- ETT, 32
  - Extraccion-Transformacion y Transporte, 31
- Herramientas
  - Para La Toma De Decisiones, 43
- IMS
  - Information Management System, 1
- J2EE, 6
- measures, 59
- Microsoft
  - .NET, 6
- OLAP, 8, 43
- Olap
  - Definicion, 43
  - Herramientas, 44
  - Olap vs Oltp, 45
- OLTP, 7, 8, 35
- oltp
  - relacion entre oltp y datawarehouse, 49
- OLTP-(Procesameinto Transaccional en Linea), 46
- open source, 6
- RDBMS, 1, 7
- SMP, 7
- tabla
  - facts, 58
- TI, 10
- XML, 7