



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

**FACULTAD DE ESTUDIOS SUPERIORES
ACATLÁN**



**ADMINISTRACIÓN DE LA INFORMACIÓN
UTILIZANDO DATA WAREHOUSE:
ARQUITECTURA Y DISEÑO**

T E S I N A
QUE PARA OBTENER EL TÍTULO DE
**LICENCIADO EN MATEMÁTICAS
APLICADAS Y COMPUTACIÓN**
P R E S E N T A
ÁNGEL GARDUÑO PÉREZ



ASESOR: LIC. ARISTEO BENJAMÍN GÓMEZ SÁINZ

AGOSTO -2004



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

FIRMA:

FECHA:

NOMBRE:

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.

ESTA TESIS NO SALE
DE LA BIBLIOTECA

Dedicatoria

La ilusión de concluir esta etapa se la debo principalmente y muy en especial a mi hermano, Felipe Viquez Chávez, ya que él me dio la oportunidad de tener una carrera profesional, así como me enseñó y me orientó en otros aspectos que me han sido de mucha utilidad a lo largo de mi camino por esta gran aventura que es la vida.

Vaya pues la dedicatoria de este evento, como resultado de nuestro esfuerzo conjunto, al gran personaje que ha sido y seguirá siendo para mí, mi hermano Felipe. Con profunda admiración, inmenso cariño, gran respeto y el dolor de no tenerlo a mi lado ... De siempre y para siempre,

Angel.

Cita para Felipe:

"Que el deporte debe ser,
como lo dice su nombre,
en el interior del ser,
lo máspreciado del hombre"

Agradecimientos

Hay tanto que agradecer y muchas personas a quienes hacerlo, que bien podría escribir una tesis acerca de ello, pero creo que con esta tesina es suficiente por ahora ...

Quiero agradecer, en lo que vale y que es bastante, a mi familia. Por guiarme, apoyarme, aguantarme y darme todo lo que tuvieron a su alcance con tal de verme como alguien de bien y con una expectativa amplia de un futuro halagüeño.

Infinitas gracias a mi mamá, Guadalupe, que me dio lo mejor de sí, su tiempo, cariño y cuidado incondicionales ...

A mis hermanos: Vicos, Raúl y Tomás, por estar ahí, cerca, en las duras y en las maduras.

A la hermana que nunca tuve, Lupe, por ser tan afable y condescendiente conmigo. Al "pequeño" Milton Carlos que me dio la oportunidad de ser niño nuevamente.

A mis cuñadas, Elvira, Chayo e Irani; y a mis sobrinos Carlos, Ale, Hugo, Stephanie, Cristian, Diana, Cesarín, "la güera" Karla y a Arantxa; a todos y cada uno por existir, por ser parte de esta gran familia y por quererme tanto como yo a ustedes.

Agradezco eterna e infinitamente, a la persona y a la mujer que decidió compartir su vida conmigo, a mi esposa Alma Delia por apoyarme en todo momento, por permitirme ser auténtico ... TAMEE!

No puedo, ni quiero dejar de agradecer a mis amigos y hermanos elegidos, a "La banda", con quienes he compartido la mejor de las etapas como estudiantes y como amigos, con quienes he pasado momentos buenos y otros no tanto, pero que en suma nos han unido de tal manera que sé, a ciencia cierta, que pueden contar conmigo, tanto como yo cuento con ustedes. Muchas gracias Rafa, Toño, Felipe, Daniel, Norma, Elena, César, Maru, Raúl, Sofía, Gustavo, Benji (amigo y asesor), Edith, Franco, Martha, Fernando, Gaby, Chipui, Armando, Laura, Cynthia, Ernesto, Omar, Octavio ... a ustedes y a sus lindas familias.

Ofrezco mi agradecimiento a las personas que han sido parte importante en mi desarrollo, que han sido buenos amigos, compañeros y que han dejado huella en mí, tanto que siempre los recuerdo con agrado, como por ejemplo al DWH Team y a la gente que hizo posible el éxito de mi aventura por España ya que se portaron súper bien conmigo, a todos un abrazo afectuoso.

Sinceramente,

Angel.

ÍNDICE

Introducción	2
Capítulo 1. Definición y principios del Data Warehouse	4
1.1. ¿Qué es un Data Warehouse?	5
1.2. Atributos del Data Warehouse	5
1.3. Escenarios de la información para la dirección	8
1.4. Acceso a la información – Tres ambientes básicos	13
1.5. ¿Qué es la Arquitectura?	17
1.6. Componentes lógicos	19
1.7. Acciones relevantes en una implementación de Data Warehouse	21
Capítulo 2. Componentes dentro de la arquitectura y diseño del Data Warehouse	22
2.1. Topologías de Data Warehouse	23
2.2. Modelado de datos	26
2.3. Administración de la Metadata	28
2.4. Extracción de datos	30
2.5. Administración de copias	38
2.6. Servidor del Data Warehouse	40
2.7. Base de Datos del Data Warehouse	45
2.8. Acceso y Reporte de datos	53
2.9. Procesamiento Analítico en Línea (OLAP)	57
Capítulo 3. Plan de Implementación del DWH	61
3.1. Esquemas de implementación	62
3.2. El Piloto del Data Warehouse	64
3.3. Cuestiones de dirección	65
3.4. Organización del Data Warehouse	67
Capítulo 4. Conclusiones	69
4.1. Conclusiones	70
ANEXO Modelado de datos en el Data Warehouse	75
Glosario	90
Bibliografía	93

Introducción

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

Ambos métodos han evolucionado a través del tiempo y ahora las organizaciones manejan una base de datos no limpia e inconsistente, sobre la cual, la mayoría de las veces, se toman decisiones importantes.

La tarea administrativa reconoce que una manera de elevar su eficiencia está en hacer el mejor uso de los recursos de información que ya existen dentro de la organización. Sin embargo, a pesar de que esto se viene intentando desde hace muchos años, no se tiene todavía un uso efectivo de los mismos.

La razón principal es la manera en que han evolucionado las computadoras, basadas en las tecnologías de información y sistemas. La mayoría de las organizaciones hacen lo posible por conseguir buena información, pero el logro de ese objetivo depende fundamentalmente de su arquitectura actual, tanto de hardware como de software.

Al respecto, es importante saber que un Data Warehouse es una colección de datos en la cual se encuentra integrada la información de la Organización y que se usa como soporte para el proceso de toma de decisiones gerenciales. Aunque diversas organizaciones e individuos logran comprender el enfoque de un Warehouse, la experiencia ha demostrado que existen muchas dificultades potenciales.

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

Un Data Warehouse se crea al extraer datos de una o más bases de datos de aplicaciones operacionales y al alimentarse de otras fuentes alternas e importantes de información. Los datos extraídos son transformados para eliminar inconsistencias y resumir (tablas sumariadas) si es necesario y luego, cargados en el Data Warehouse. El proceso de transformar, crear el detalle de tiempo variante (representa un período de tiempo: fecha de creación, etc.), resumir y combinar los extractos de datos, ayudan a crear el ambiente para el acceso a la información Organizacional. Este nuevo enfoque ayuda a los individuos, en todos los niveles de la empresa, a efectuar su toma de decisiones con mayor responsabilidad.

La innovación de la tecnología de información dentro de un ambiente Data Warehouse, puede permitir a cualquier organización optimizar el uso de los datos, como un ingrediente clave para un proceso de toma de decisiones más efectivo. Las organizaciones tienen que aprovechar sus recursos de datos para crear la información de la operación del negocio, pero deben considerarse las estrategias tecnológicas necesarias para la implementación de una arquitectura completa de Data Warehouse.

El Data Warehouse ha sido el centro de la arquitectura para los sistemas de información en la década de los '90, en la actualidad y es una tendencia futura, al menos para las empresas que quieran tener alguna ventaja competitiva.

Diferentes tipos de información

La tecnología Data Warehouse basa sus conceptos y diferencias en dos tipos fundamentales de sistemas de información en todas las organizaciones: los sistemas técnico-operacionales y los sistemas de soporte a las decisiones.

Hay otras funciones dentro de la empresa que tienen que ver con la planeación, previsión y administración de la organización. Estas funciones son también críticas para la supervivencia de la organización, especialmente en nuestro mundo de rápidos cambios.

Las funciones como "planificación de mercadotecnia", "planeación de ventas" y "análisis financiero", requieren, además, de sistemas de información que los soporten. Pero estas funciones son diferentes de las operacionales, y los tipos de sistemas y la información requerida son también diferentes. Las funciones basadas en el conocimiento son los sistemas de soporte a las decisiones.

Estos sistemas están relacionados con el análisis de los datos y la toma de decisiones, frecuentemente, decisiones importantes sobre cómo operará la empresa, ahora y en el futuro. Estos sistemas no sólo tienen un enfoque diferente al de los operacionales, sino que, por lo general, tienen un alcance diferente.

Mientras las necesidades de los datos operacionales se enfocan normalmente hacia una sola área, los datos para el soporte de decisiones, con frecuencia, son tomados de varias áreas diferentes y necesitan cantidades grandes de datos operacionales relacionados. Son estos sistemas sobre los que se basa la tecnología del Data Warehouse.

A fin de comprender cómo se relacionan todos los componentes involucrados en una estrategia de Data Warehouse, es esencial contar con la arquitectura correspondiente, una Arquitectura de Data Warehouse.

Capítulo 1.

Definición y principios del Data Warehouse

Objetivo:

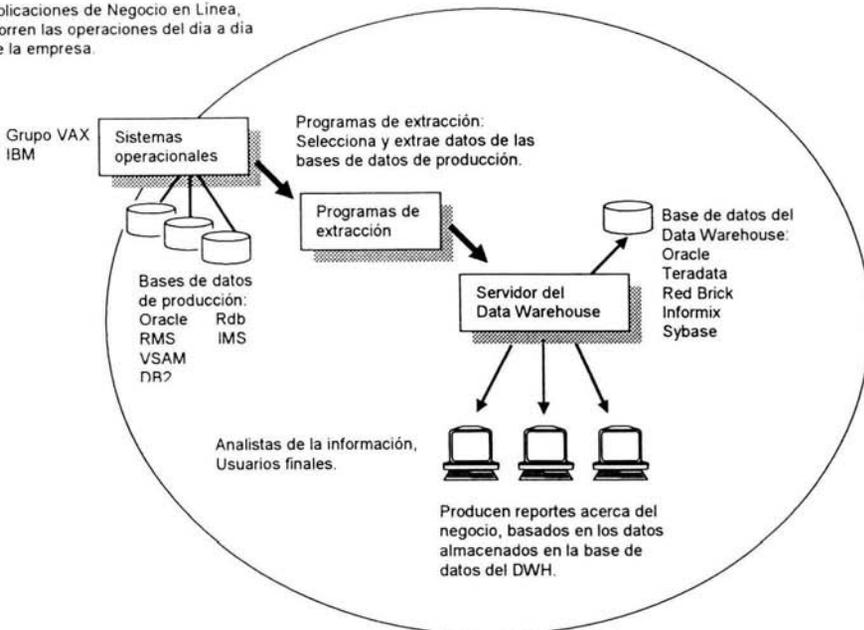
Dar una noción del Data Warehouse (DWH) en el marco de la administración de la información, mostrando los atributos principales de los datos almacenados en el mismo para luego notar la evolución en el acceso a esos datos y a los informes como son vistos por los usuarios finales del DWH.

Mostrar las tres principales configuraciones en la administración de la información con la intención de perfilar una metodología para el desarrollo ordenado de la arquitectura del DWH dividiendo el ambiente en cuatro componentes lógicos e identificando la funcionalidad de cada uno.

1.1. ¿Qué es un Data Warehouse?

Ambiente del Data Warehouse

Aplicaciones de Negocio en Línea,
Corren las operaciones del día a día
de la empresa.



Definición

Un Data Warehouse es un marco de información. Involucra una solución completa la cual consta de una mezcla en armonía de hardware, software, conocimiento del negocio y técnicas de integración de sistemas.

1.2. Atributos del Data Warehouse

Un Data Warehouse es:

- Orientado al negocio
- Integrado
- Variante en el tiempo
- No volátil



una colección de datos *en apoyo a las decisiones de la dirección*¹.

¹ William H. Inmon: *Building the Data Warehouse*, John Wiley, 1996.

Orientado al negocio

Los datos son categorizados y almacenados en el Data Warehouse por tipo más que por aplicación.

Aplicaciones operacionales/ Bases de datos

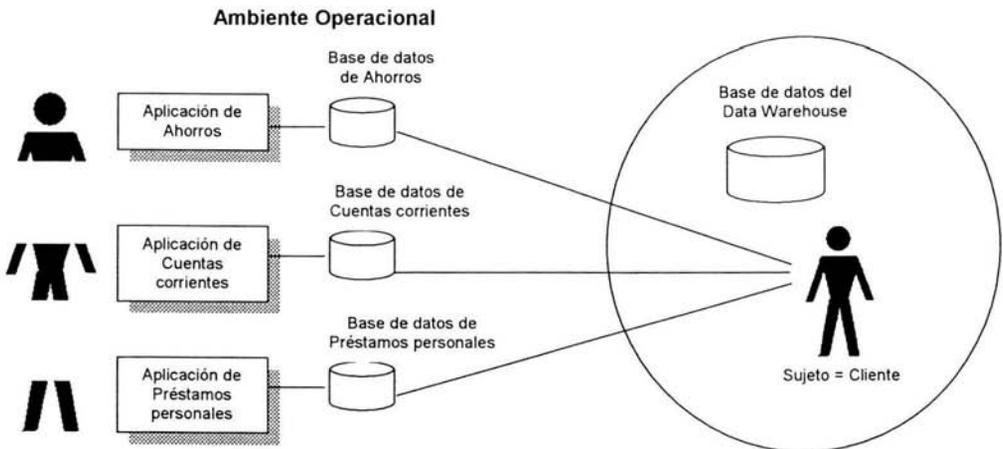
Estados de cuenta
Cuentas por cobrar
Cuentas por pagar
Préstamos
Ahorros
Evaluación de reclamaciones de seguro de vida
Seguro para autos

Contenido del Data Warehouse

Cliente
Reclamaciones
Ventas
Producto

Integrado

Los datos acerca de un sujeto determinado son definidos y almacenados una vez.
Todos los datos corporativos necesarios son reunidos en una sola fuente integrada.

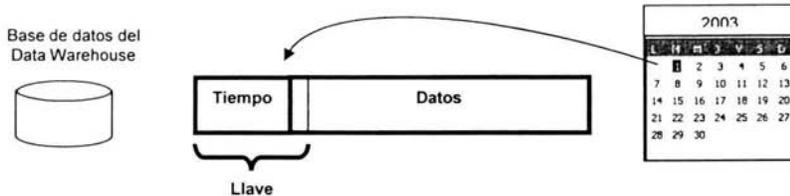


Datos del cliente almacenados en varias bases de datos

Ejemplo: Institución Bancaria

Variante en el tiempo

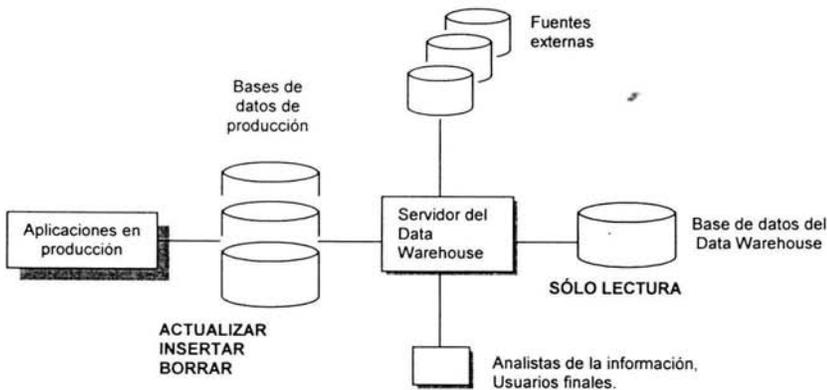
Los datos son almacenados como una serie de snapshots (fotografías instantáneas), cada una representando un periodo de tiempo.



- Los datos son etiquetados con algún elemento de tiempo – fecha de creación, ciclo de negocio, etc.
- Los datos están disponibles por largos periodos de tiempo. Por ejemplo, de cinco a diez años.
 - Nota: Con frecuencia los datos almacenados en los sistemas operacionales se encuentran disponibles solamente por cortos periodos de tiempo. Por ejemplo, de 30 a 60 días y luego son respaldados o borrados.

No volátil

Los datos dentro del Data Warehouse no pueden ser actualizados o borrados.



El Data Warehouse

- Optimiza los datos y su organización para su acceso y análisis.
- Proporciona, a los usuarios finales, un catálogo de datos accesibles.
- Integra los datos de múltiples fuentes.
- Asegura la consistencia de la información utilizada por todos los usuarios.
- Elimina el impacto, en los usuarios finales, del acceso sobre los sistemas operacionales.
- Proporciona datos de toda la empresa a los usuarios finales.

1.3. Escenarios de la información para la dirección

Data Mining (Minería de datos)

Servidores OLAP

Data Marts (Mercados de datos)

Almacenes de Datos Operacionales (ODS)

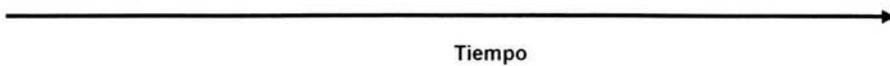
El Data Warehouse

Mismos datos, resultados diferentes

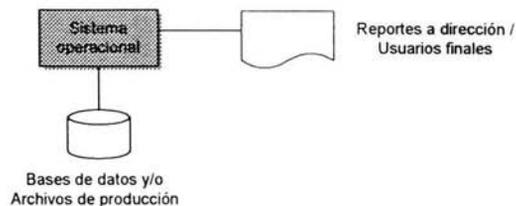
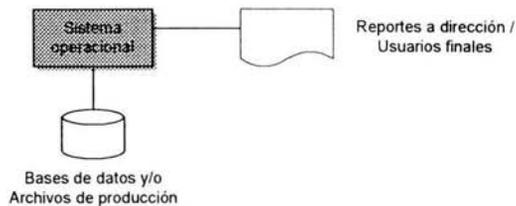
Acceso de los usuarios finales a los datos extraídos

Los Sistemas de Soporte a las Decisiones (DSS) sobrecargan el ambiente de producción

Reportes operacionales

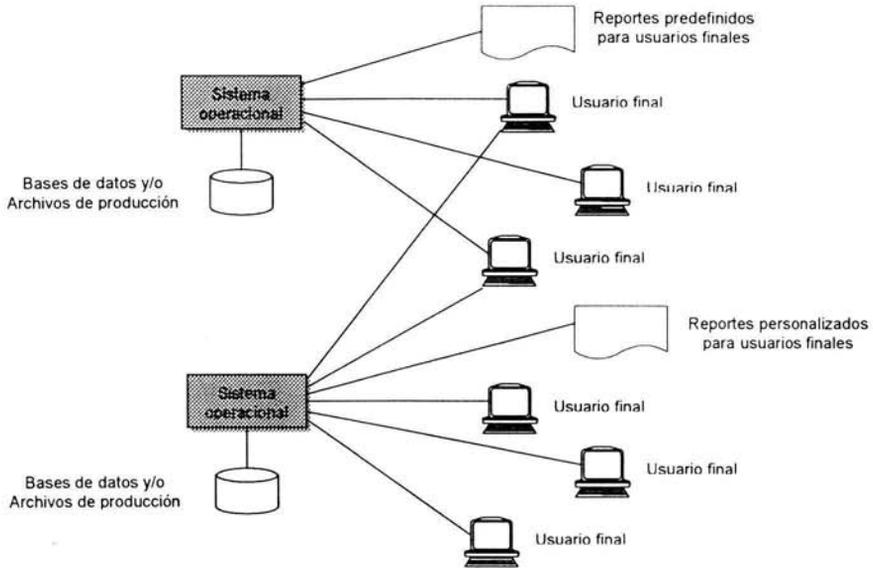


En el inicio: Reportes operacionales



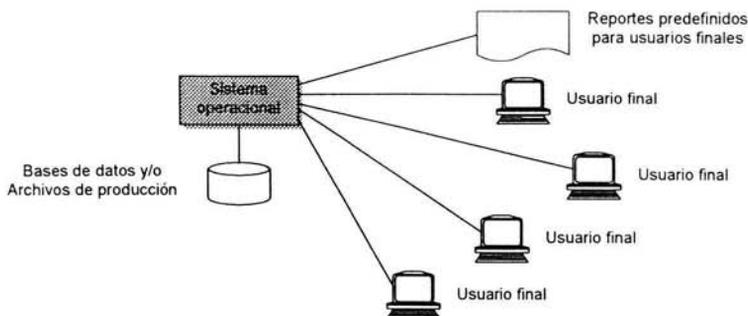
- Todos los reportes son producidos por el sistema de información de la organización.

Acceso de los usuarios finales a los sistemas de producción



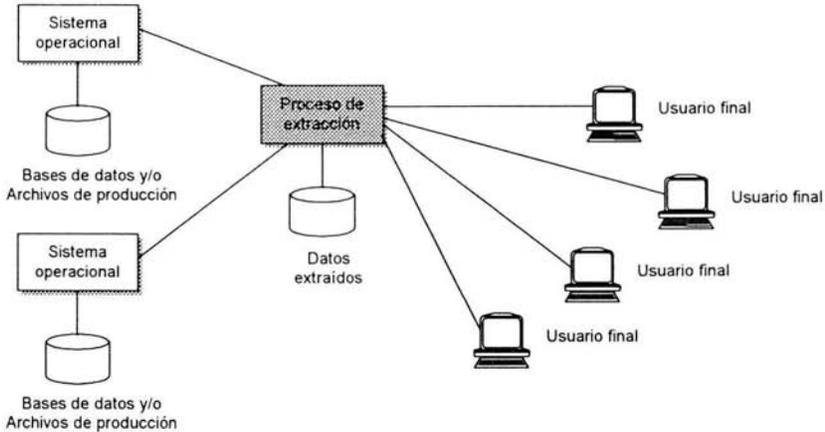
- El legado de las estructuras de datos no está diseñado para ser entendido por usuarios que no sean parte del área de sistemas de información.

Los Sistemas de Soporte a las Decisiones (DSS) sobrecargan el ambiente de producción



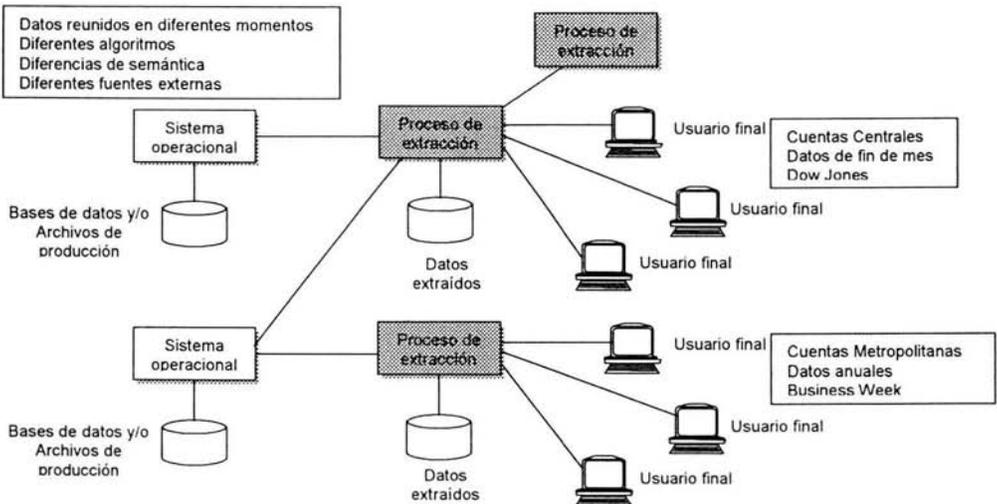
- El acceso de los usuarios finales tiene un impacto negativo en el rendimiento de los sistemas transaccionales en línea, OLTP ya que saturan el acceso al sistema.
- Los usuarios finales deben acceder a datos localizados en múltiples sistemas y bases de datos porque existen diversos sistemas y cada uno tiene su propia base de datos.
- El acceso de los usuarios finales debe ser fuera del proceso de carga porque la información no está disponible mientras se ejecuta este proceso.

Acceso de los usuarios finales a los datos extraídos



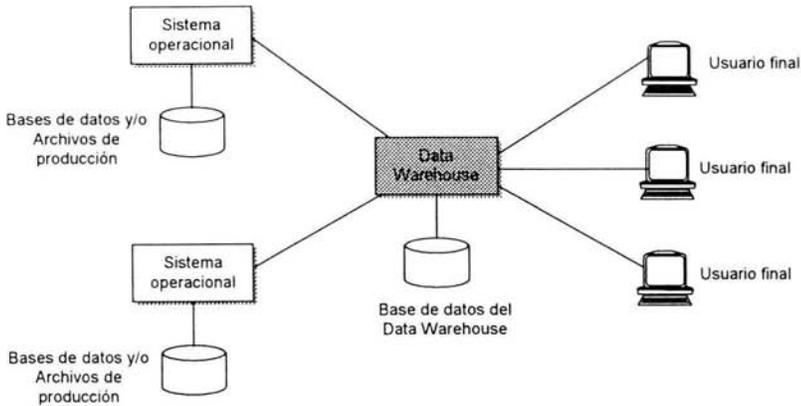
- El acceso de los usuarios finales está fuera del proceso de carga del ambiente operacional.
- El ambiente de los DSS (Decision Support Systems) es sobrecargado.

Mismos datos, resultados diferentes



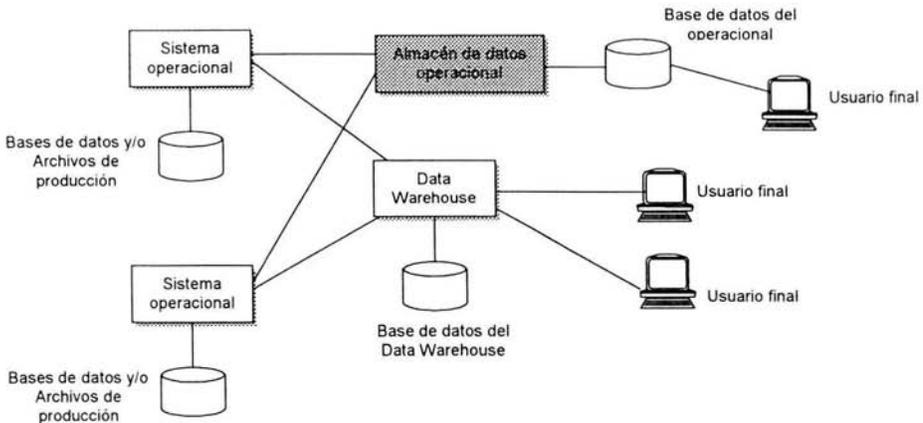
- Diferentes grupos obtienen diferentes resultados. La credibilidad es un riesgo.

El Data Warehouse



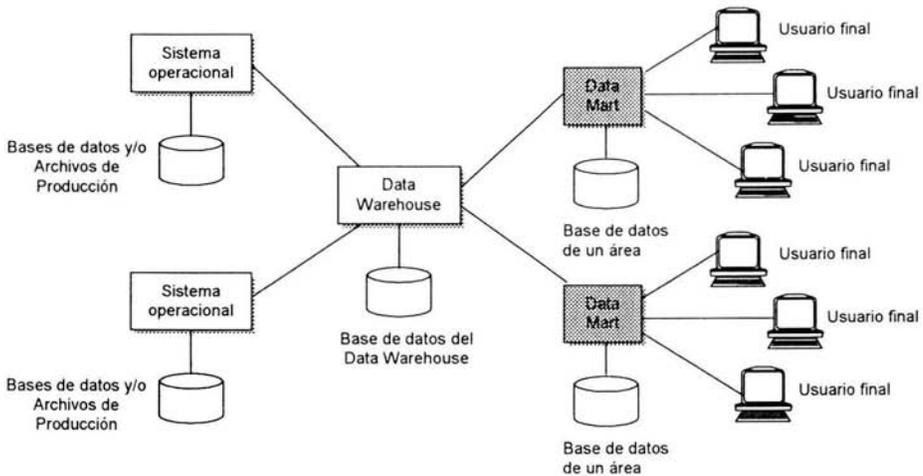
- El Data Warehouse es el centro de distribución para los datos de producción.

Almacenes de datos operacionales



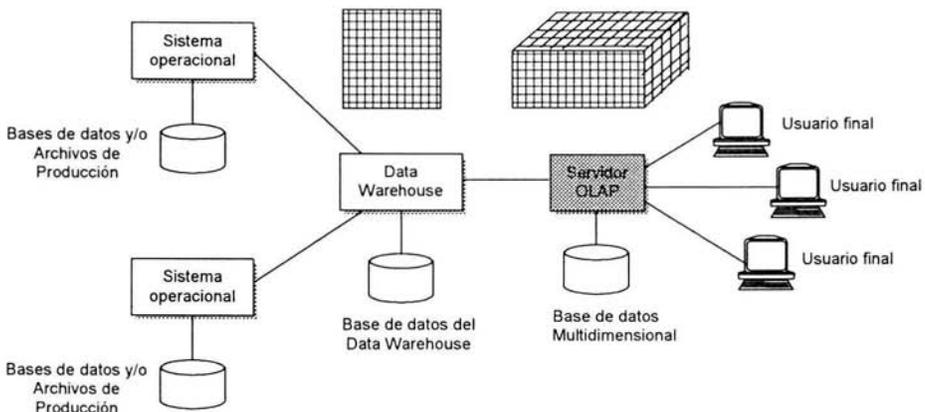
- El almacén de datos operacional es el centro de distribución para los valores actuales de los datos.
- A partir de éste se toman decisiones día con día.

Data Marts (Mercados de Datos)



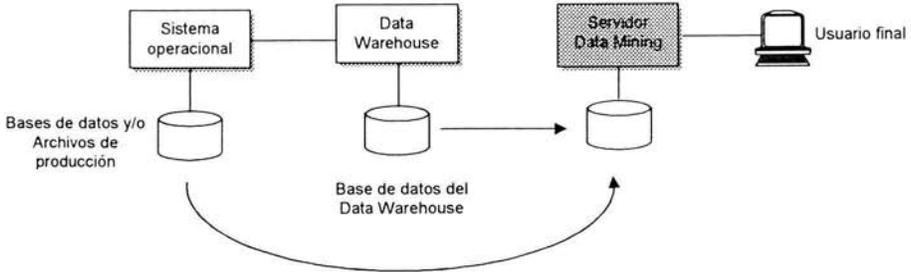
- Los Data Marts extraen y adaptan porciones del Data Warehouse para satisfacer requerimientos únicos de un área o departamento.

Servidores OLAP (On-Line Analytical Processing)



- Porciones del Data Warehouse son movidas para un propósito especial, "base de datos multidimensional", para optimizar el análisis de los datos.

Data Mining (Minería de Datos)

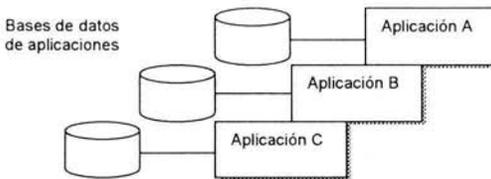


- Examina patrones de datos, asociaciones, etc., sobre los datos recibidos del ambiente operacional o del Data Warehouse.

1.4. Acceso a la información – Tres ambientes básicos.

Mixto; Producción y DSS

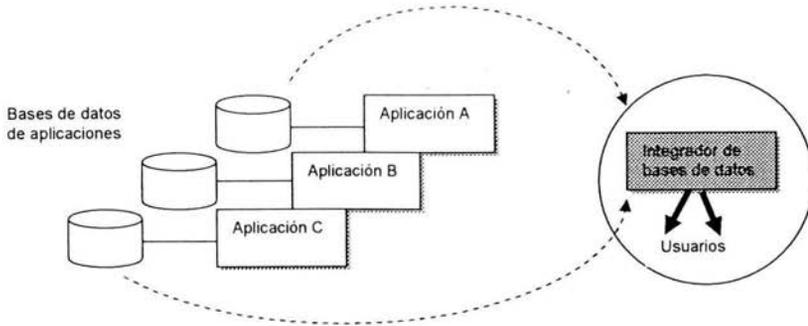
Ambiente operacional



Prioridades	Buen rendimiento, disponibilidad y recuperación del ambiente de misión crítica.
Utilización del procesador	Altamente previsible.
Tiempo de respuesta	Fraciones de segundo a pocos segundos.
Base de datos	Jerárquica, Red, Relacional, Archivos planos. Organizada por aplicación.
Contenido de datos	Datos necesarios, en una aplicación específica, para procesar todas las transacciones del negocio.
Naturaleza de los datos	Dinámicos, cambiando constantemente.
Procesamiento de aplicaciones	Altamente estructurado, procesamiento repetitivo. Operaciones de inserción, actualización y borrado.
Usuarios finales	Empleados.

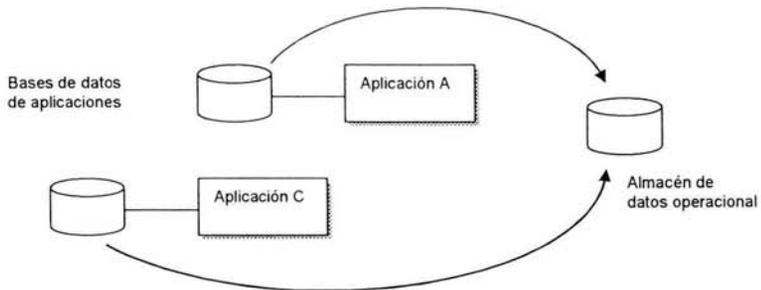
- Propósito: Ejecutar las operaciones del día a día de la empresa.

Accesando datos desde los sistemas operacionales



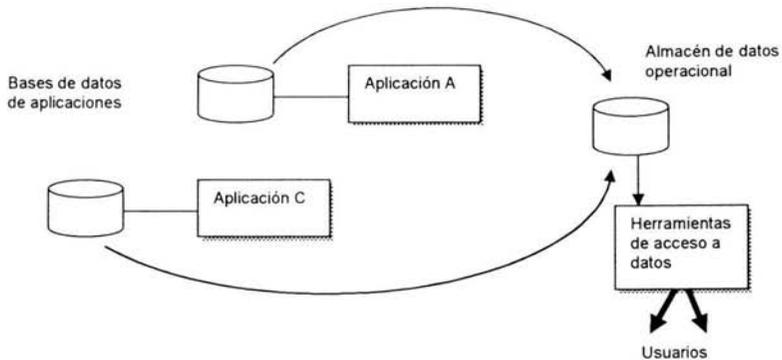
- **Ventajas**
Acceso sencillo a múltiples bases de datos.
Puede acceder múltiples bases de datos para satisfacer una consulta simple del usuario.
- **Desventajas**
Puede impactar severamente el rendimiento de las aplicaciones de producción.
No hay datos históricos.
Necesidad de un experto en consultas (SQL) por parte de los usuarios.

Almacén de Datos Operacional (ODS, por sus siglas en inglés)



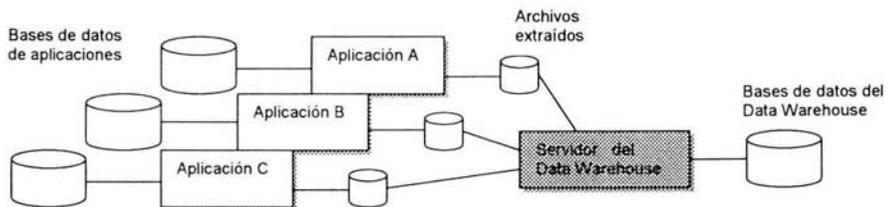
Prioridades	Disponibilidad.
Utilización del procesador	Previsible – la mayor parte del tiempo.
Tiempo de respuesta	Segundos a minutos.
Base de datos	Relacional, Otros. Organizado por sujetos, ejemplo: Ventas, Cliente. Datos actuales.
Contenido de datos	Todos los datos necesarios, integrados para procesar las consultas operacionales.
Naturaleza de los datos	Dinámicos, cambiando constantemente.
Procesamiento de aplicaciones	Algo estructurado, algo de procesamiento analítico.
Usuarios finales	Usuarios de aplicaciones.

- **Propósito:** Tomar las decisiones del día a día.



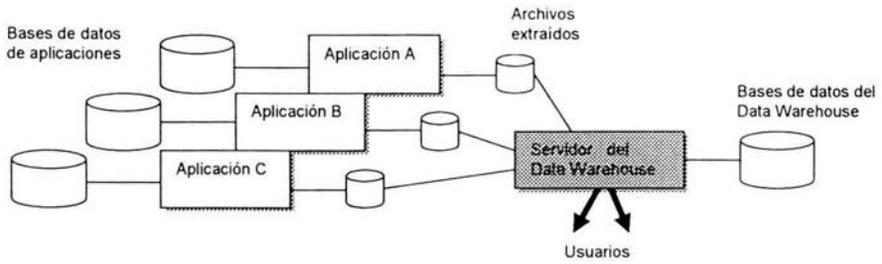
- **Ventajas**
Datos operacionales integrados, disponibles para los usuarios.
- **Desventajas**
Podría impactar el rendimiento de las aplicaciones de producción.
No hay datos históricos.

Data Warehouse



Prioridades	Facilidad de uso, Flexibilidad en el acceso a los datos.
Utilización del procesador	Altamente imprevisible.
Tiempo de respuesta	Segundos a varios minutos.
Base de datos	Relacional, Multidimensional. Organizado por sujetos, ejemplo: Ventas, Cliente. Snapshot (Foto instantánea) de los datos.
Contenido de datos	Todos los datos integrados, necesarios por la dirección.
Naturaleza de los datos	Estáticos.
Procesamiento de aplicaciones	No estructurado, procesamiento heurístico y analítico.
Usuarios finales	Usuarios de aplicaciones, Directivos.

- **Propósito: Administrar la información de la empresa.**

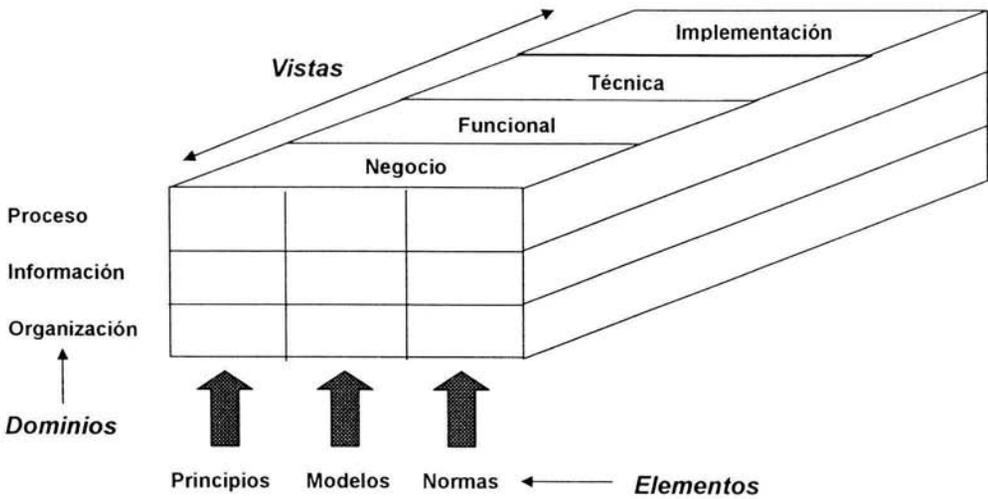
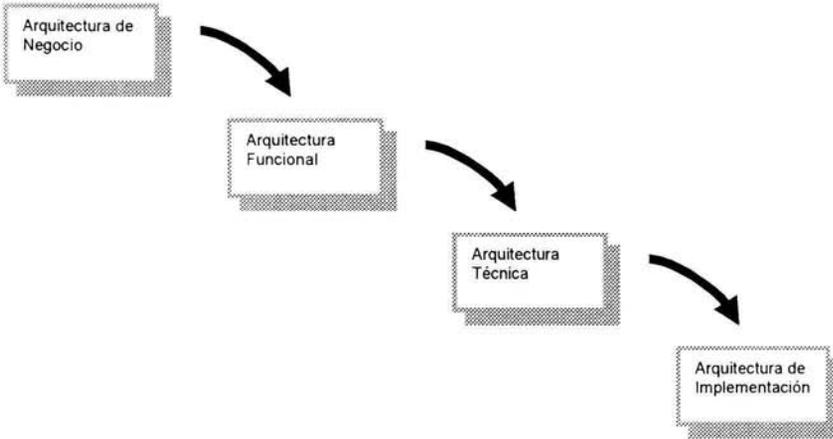


- **Ventajas**
Comunicación, en términos semejantes a los de negocio, entre los usuarios y el Data Warehouse.
Sistemas de producción aislados desde el análisis de la carga.
Disponibilidad de datos históricos.
Fuentes de datos comunes para el análisis del negocio; los resultados pueden ser replicados.
- **Desventajas**
Nuevo ambiente de acceso a la información. Capacitación necesaria para aprender el concepto.

1.5. ¿Qué es la Arquitectura?

Un marco de referencia para tomar decisiones tecnológicas en la fase de cambio.

Tipos de Arquitectura



Las **Vistas** de la arquitectura (también pueden ser llamadas "Capas") son:

- **Negocio** – porque éste es necesario
- **Funcional** – que podría parecer como lo ve el usuario
- **Técnico** – como puede ser implementado utilizando Tecnología de Información
- **Implementación** – el despliegue y organización del significado (gente, así como productos)

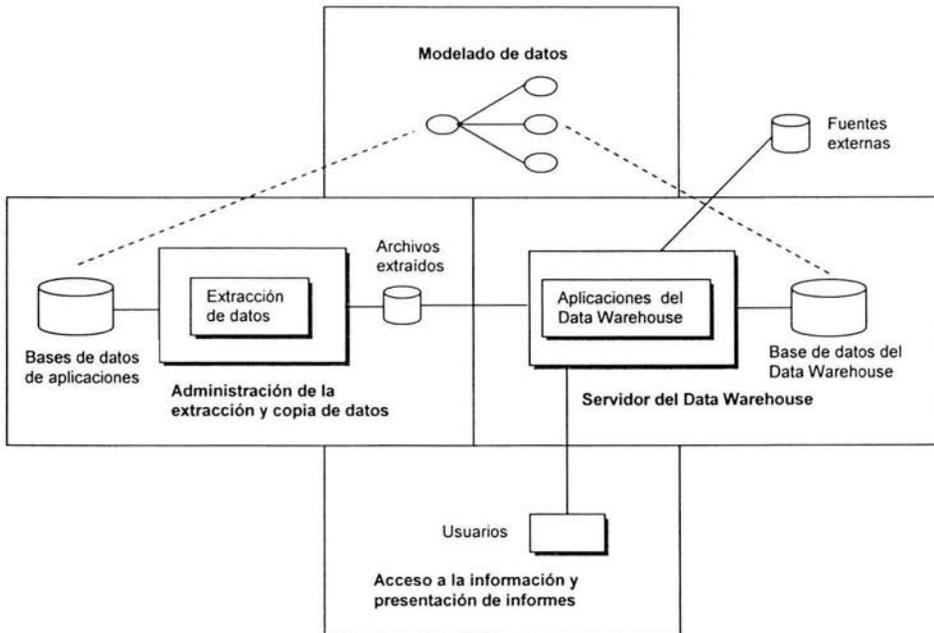
Las vistas de la arquitectura son descritas con los siguientes elementos:

- **Dirigentes del negocio** – comunicados de metas de negocio, direcciones, acciones y prioridades las cuales dirige la Tecnología de Información.
- **Principios** – comunicados que fundamentalmente dan forma y conectan las vistas.
- **Modelos** – descripciones ilustradas de las vistas (modelos de procesos, modelos de aplicaciones, modelos de datos, etc.)
- **Normas** – comunicados los cuales condicionan las vistas.

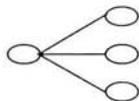
Visión general de las vistas de la Arquitectura

Vista	Objetivo	Participantes	Salidas típicas
Negocio	Establecer un entendimiento claro de la naturaleza de la empresa, sus principales objetivos, actividades y procesos.	Empresarios y Directores de TI	Normas. Principios de negocio. Modelos de proceso de negocio. Modelos organizacionales. Métricas de negocio. Datos de la empresa. Escenarios de topografía y distribución. Factores críticos de éxito.
Funcional	Proporcionar el anteproyecto para construir los sistemas de información, componentes e interfases requeridos para soportar el negocio.	Empresarios y Directores de TI, Usuarios, Líderes de proyectos, Desarrolladores, Personal de soporte técnico	Principios de sistemas. Modelos de procesos. Recomendaciones de inversión. Normas de sistemas. Anteproyecto de alto nivel (modelos funcionales) Estrategia de implementación. Acciones. Modelos de datos.
Técnica	Definir el enfoque y estilo para la infraestructura técnica y el criterio de selección para los componentes de hardware y software.	Directores de TI, Administradores de proyectos, Arquitectos de soluciones, Analistas, Desarrolladores, Personal de soporte técnico	Principios de tecnología. Modelos de infraestructura. Normas y métricas técnicas. Estrategia de implementación. Oportunidades técnicas.
Implementación	Definir el enfoque y estilo para implementar la arquitectura.	Directores del negocio y de TI, Desarrolladores, Personal de soporte técnico	Recomendaciones de inversión. Plan de implementación. Componentes de hardware y software. Pautas organizacionales.

1.6. Componentes lógicos



Modelado de datos



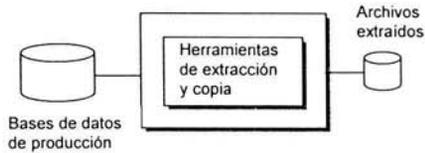
Metodología para mapear los datos de los sistemas operacionales en información orientada a sujetos, tales como: Ventas, Clientes, etc.

- Identificar los sujetos del negocio.
- Definir relaciones entre los objetos del negocio.
- Nombrar los atributos para cada sujeto del negocio.

Desarrollar un modelo completo de datos, corporativo, puede ser complejo y llevar mucho tiempo.

El mejor enfoque es comenzar con uno o dos sujetos del negocio, de interés inmediato para la empresa.

Administración de la extracción y copia de datos

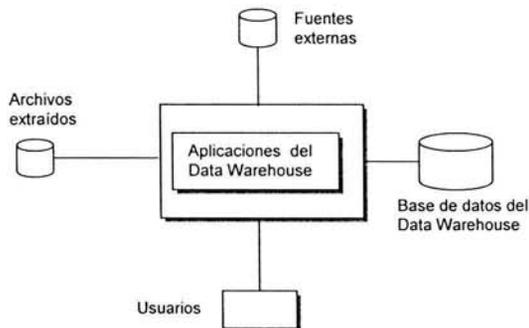


Involucrado con la selección y movimiento de los datos operacionales al Data Warehouse.

- Los datos operacionales pueden sufrir varias transformaciones para aumentar su utilidad, tales como agregaciones, agrupamiento, extrapolación, convertir estructuras y formatos, y expansión de campos clave.

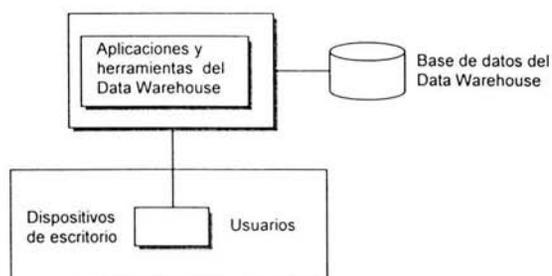
Tal vez, el área más compleja del Data Warehouse.

Servidor del Data Warehouse



- Es el centro de información para la empresa.
- Asegura que los datos del negocio sean correctos y estén completos para entregarlos y que sean autorizados por los usuarios.
- Maneja los requerimientos de interoperabilidad con los sistemas operacionales, las fuentes de datos externas y el ambiente de escritorio.

Acceso a la información y presentación de informes



- Recupera datos del Data Warehouse y los transforma en información para el negocio.

1.7. Acciones relevantes en una implementación de Data Warehouse

Para el ambiente actual de Acceso de datos y Reporte de información, prepare lo siguiente:

- Topología de hardware
- Inventario de software
- Herramientas DSS (Sistemas de Soporte a Decisiones, tales como: sistemas expertos, inteligencia artificial, redes neuronales) y EIS (Sistemas de Información Ejecutiva, como por ejemplo InSight que permite a los usuarios acceder a la información requerida y relevante directamente, sin necesidad de conocimientos de programación)

Para cada componente del ambiente del Data Warehouse prepare la Revisión # 0.1 de la arquitectura del almacén de datos:

- Principios del Data Warehouse
- Modelos del Data Warehouse
- Normas del Data Warehouse

Capítulo 2.

Componentes dentro de la arquitectura y diseño del Data Warehouse

Objetivo:

Revisar formas potenciales de configurar el ambiente del Data Warehouse (DWH), presentando las funciones, diseño y cuestiones de implementación asociadas con cada componente principal del DWH.

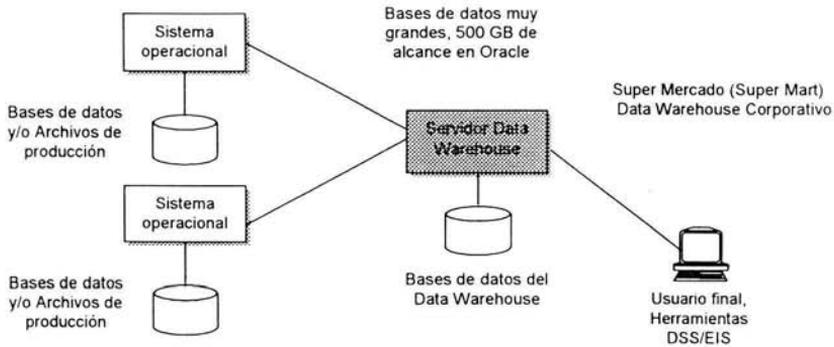
Listar algunas herramientas de software disponibles para implementar cada componente del DWH, tratando de proporcionar algunas recomendaciones para seleccionar las más adecuadas y que mejor se ajusten a las necesidades.

2.1. Topologías de Data Warehouse

Tipos de Data Warehouse

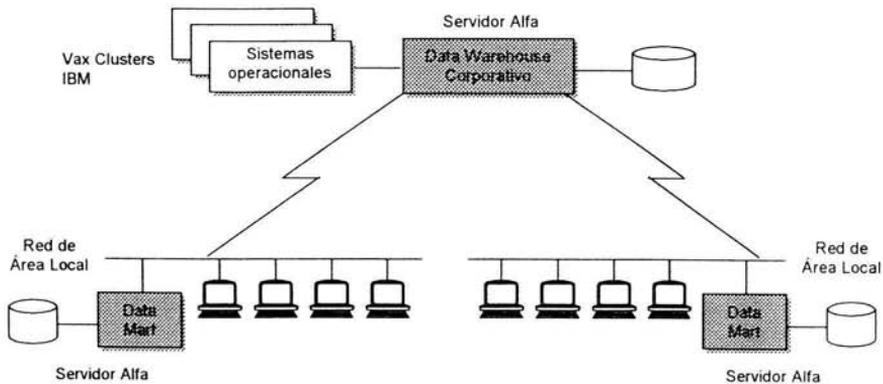
- Almacén Corporativo
- Almacén Geográfico
- Almacenes funcionales
 - Ventas
 - Recursos Humanos
 - Producto
- Almacenes Organizacionales

Data Warehouse Centralizado, ambiente Cliente/Servidor



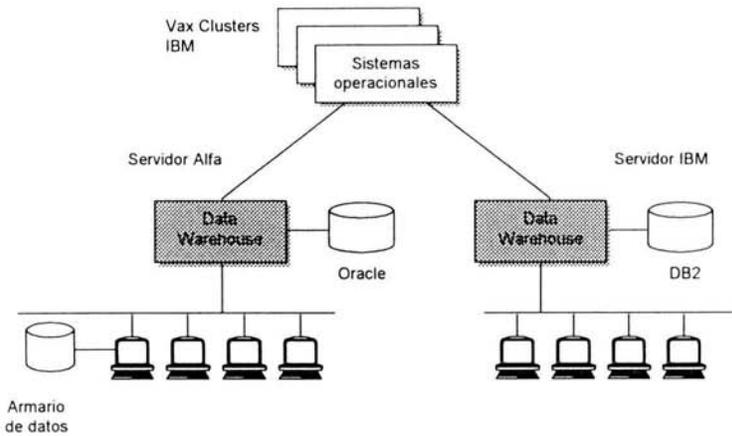
- Punto central de distribución.
- Única fuente autorizada de datos.

Data Warehouse Distribuido, Ejemplo 1



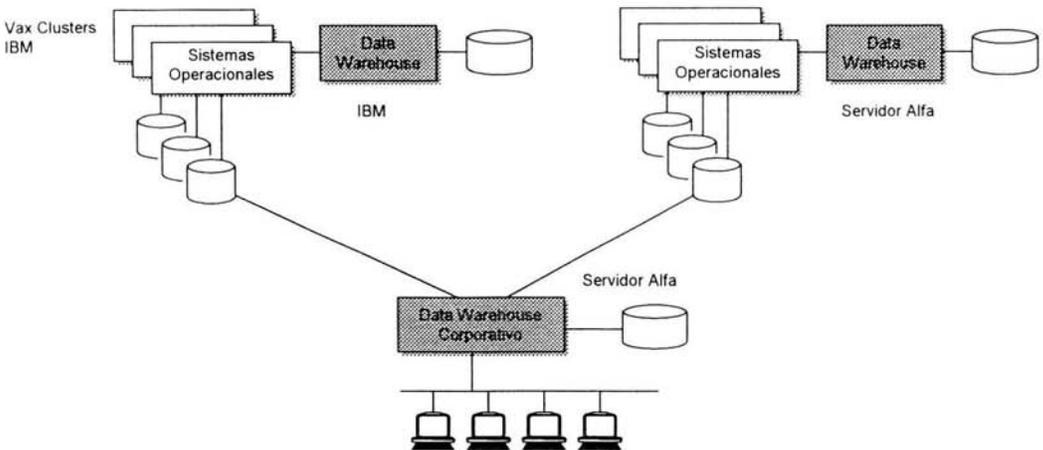
- Los datos del Data Warehouse son distribuidos hacia los Data Marts funcionales/departamentales.

Data Warehouse Distribuido, Ejemplo 2



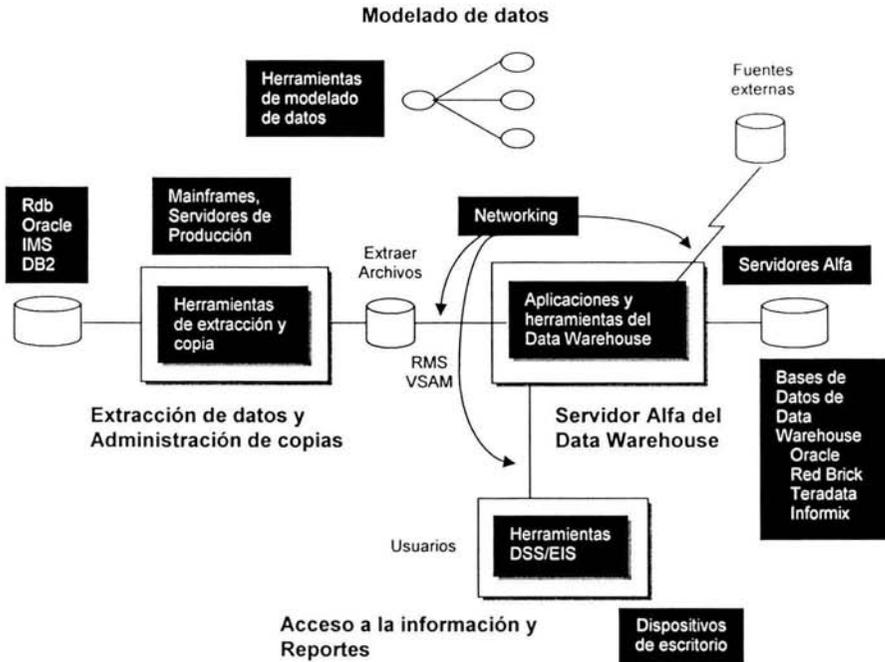
- Data Warehouse Múltiple.

Data Warehouse Distribuido, Ejemplo 3



- Data Warehouse Corporativo y Departamental.

Ambiente del Data Warehouse, Planificación, herramientas e integración de sistemas



Acciones relevantes en una implementación de Data Warehouse:

- Topología del ambiente del Data Warehouse actual y planeado.
- Componentes principales de Hardware y Software.
- Ambiente y protocolos de comunicaciones.

2.2. Modelado de Datos

Modelos de datos de negocio

Modelos funcionales

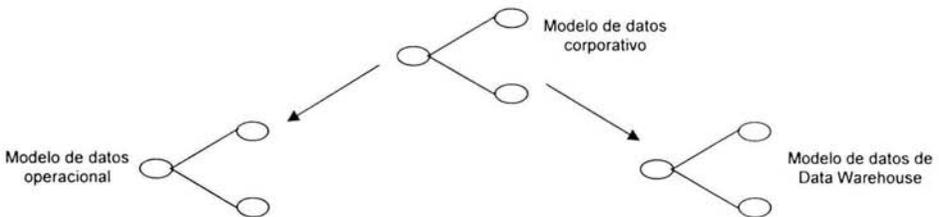
- ¿Qué hacen los negocios?
- Funciones, procesos y actividades.
- Plantean los requerimientos de información.

Modelos de datos

- Definen las entidades de negocio.
- Establecen relaciones entre las entidades de negocio.
- Identifican reglas de negocio y políticas.
- Sirven como vehículo de comunicación para los desarrolladores de sistemas.



Modelos de datos operacionales vs. Modelos de datos de Data Warehouse



Modelos de datos OLTP	Modelos de datos de Data Warehouse
Orientación de la línea de negocios.	Orientación a sujetos.
Consideraciones de rendimiento en todas partes.	Agrega elementos de tiempo.
Datos normalizados. Redundancia de datos mantenida en un mínimo.	Datos desnormalizados.
Activa la integridad referencial.	Trata la integridad referencial como objetos.

Modelos de datos de Data Warehouse

Desarrollados internamente

- Consume mucho tiempo.
- Requiere personas y recursos especializados.

Desarrollados con ayuda externa

- Puede ayudar a acelerar el desarrollo.
- Puede contribuir en la corrección de los modelos de datos.

Modelos de datos de datos genéricos – Ayuda externa

Funciones típicas

Contabilidad/Finanzas
Recursos humanos
Mercadotecnia/Ventas
Contratos
Impuestos corporativos

Industrias típicas

Líneas aéreas
Manufactura
Telefonía
Banca/Financiera
Distribución
Distribución indirecta
Exploración de gas y petróleo
Empresas públicas
Seguros

- Es bueno revisar el modelo contra otros que ya han sido hechos.
- Hay que considerar la utilización de un Modelo de Datos Genérico como punto de partida.

Consideraciones en la implementación

- Limitar el alcance del Modelo de Datos. Para el Data Warehouse es de suma importancia, por ejemplo, el número de temas a implementar.
- Definir las entidades y sus relaciones, en una empresa grande, lleva un tiempo considerable.
- Los Modelos de Datos pueden llevar a un óptimo diseño teórico.
- Estar enterados de los constraints impuestos por los modelos de datos operacionales existentes.
- Crear un modelo de datos completo para una empresa grande que dure para siempre.

Herramientas de Modelado de Datos

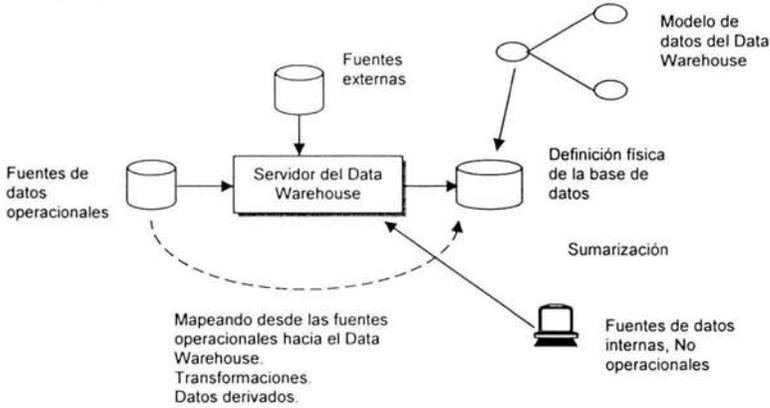
Compañía	Producto
Data Stage	Data Models
Oracle	Designer
Texas Instruments	ADW Encyclopedia IEF CASE MSP Reltech Brownstone CA-IDD Bachman
Erwin	Erwin

Acciones relevantes en una implementación de Data Warehouse:

- Temas (Sujetos) del Data Warehouse: actuales y posteriores.
- Herramienta de Modelado de Datos.
- ¿Quién es el responsable del Modelado de Datos?

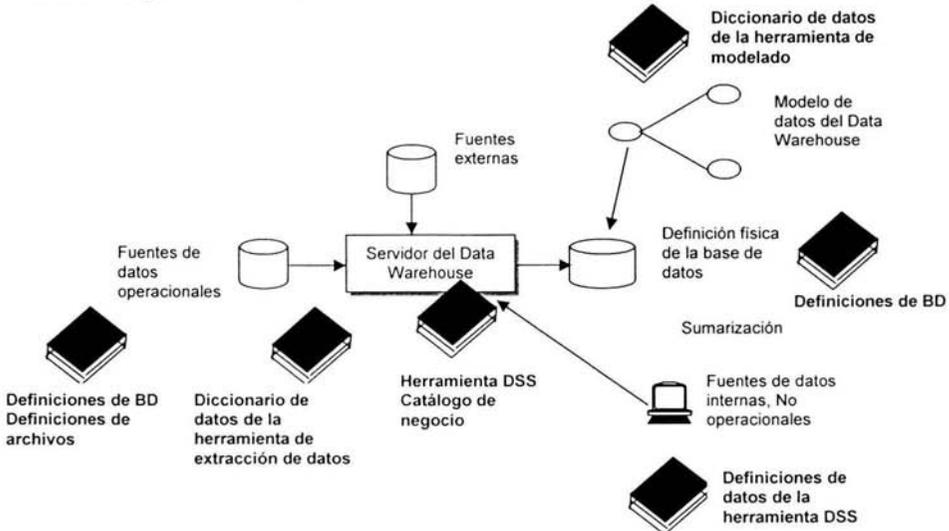
2.3. Administración de la Metadata

Metadata del Data Warehouse



- **Metadata: Información acerca de los datos.**

Metadata y Diccionarios de datos



Metadata del Data Warehouse

Puede involucrar metadatos de múltiples fuentes:

- Modelado de datos
- Código fuente
- Bases de datos
- Sistemas de archivos
- Fuentes externas
- Fuentes internas no-operacionales
- Herramientas DSS/EIS
- Herramientas de extracción de datos

Las versiones de la Metadata pueden requerir el soporte de la historia de los cambios a través del tiempo

- Definiciones de campos estándar
- Rutas fuente
- Tablas destino
- Mapeo de datos
- Criterio de selección
- Transformaciones de datos
- Cálculos de datos derivados

Consideraciones para la implementación

- Principal cuestión en la implementación: Sincronización de la Metadata.
 - Existe una carencia de Metadatos estándar para el Data Warehouse.
 - La actual tendencia es tener proveedores de herramientas DSS/EIS, olvidando las relaciones con los proveedores de soluciones de Data Warehouse (por ejemplo: Business Objects/Data Stage, MicroStrategy/Informática, Esperant/Carleton)
- ⇒ Considerar las herramientas DSS/EIS que permitan a los usuarios finales la facilidad de ver los tipos y características de los datos almacenados en el Data Warehouse.

Herramientas de Administración de la Metadata

Metadata para Administradores de Datos

Compañía	Producto
Informática	Informática Manager
Data Stage	Data Stage Directory Manager
Carleton	Metadata Dictionary
Data Integration	DD/D Builder
Evolutionary Technologies, Inc.	ETI

Metadata para Usuarios Finales

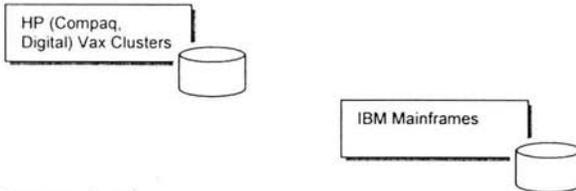
Compañía	Producto
Business Objects	Business Objects
MicroStrategy	MicroStrategy
Cognos	Cognos
Open Data	Find Out!
Trinzic	InfoPump

Acciones relevantes en una implementación de Data Warehouse:

- Fuentes de Metadatos.
- Diccionarios y repositorios utilizados y planeados.
- Requerimientos de Metadatos vs. Funcionalidad disponible en las herramientas existentes.

2.4. Extracción de Datos

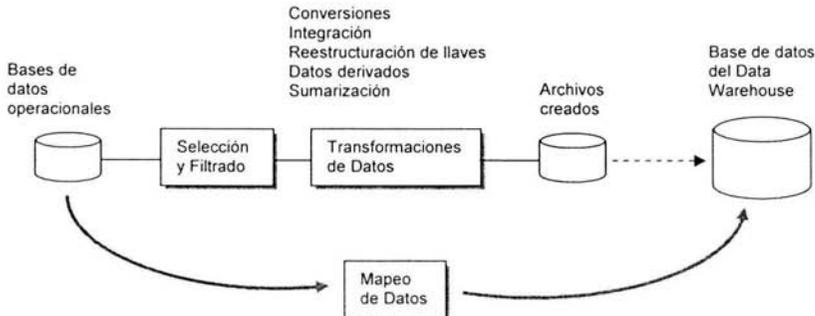
Fuentes internas de datos



El ambiente operacional:

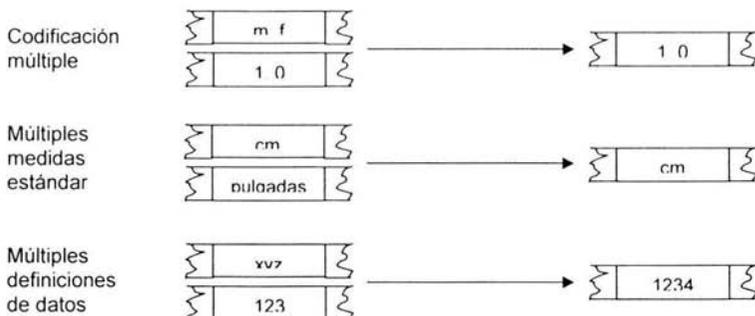
- Múltiples plataformas de Sistemas Operativos.
- Múltiples plataformas de Hardware.
- Múltiples organizaciones de Bases de Datos.
- Múltiples organizaciones de Archivos.

Proceso de Extracción de Datos

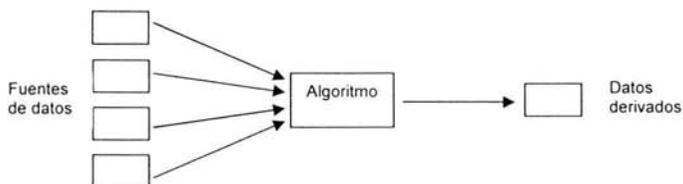


⇒ La extracción y la transformación de datos son una parte muy importante en el esfuerzo requerido para implementar el Data Warehouse.

Conversiones de Datos



Datos derivados



Clasificación de clientes

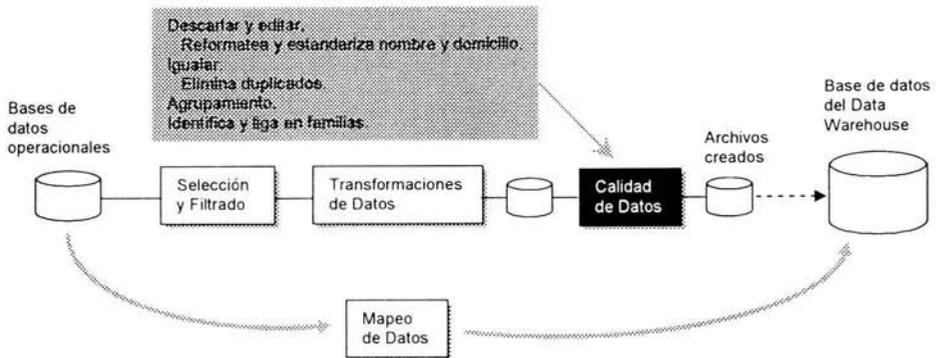
- Ventas totales, descuento total, intereses perdidos debido a pagos tardíos, costo de moneda, etc.

Cálculos anuales

- Total de salario, total de impuestos, total de seguro social, etc.

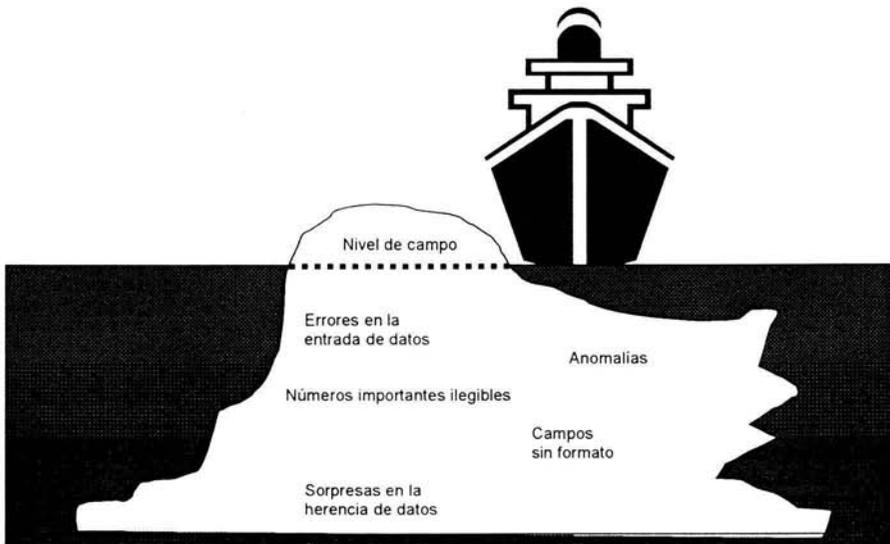
Calculándolo una vez, está disponible para siempre.

Credibilidad del Data Warehouse – Calidad de datos



Los proveedores de herramientas de Calidad de Datos (Trillium, Vality, Innovative Systems), usualmente se asocian con los proveedores de Extracción de Datos (Informática, Data Stage, Carleton, Data Integration, ETI), para introducir calidad de datos al Data Warehouse.

Anomalías en los datos operacionales



La calidad de datos requiere revisar los datos que existen, no solamente las definiciones de los mismos.

Falta de estándares

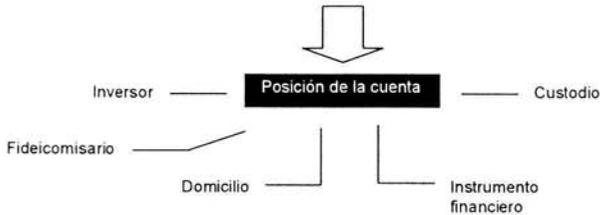
	Nombre	Localización
Base de Datos 1	Roberto González	MA93
	José G. García	CT15
	Carlos y Lety Víquez	IL21
Base de Datos 2	González, Roberto	MA93
	García, José	CT15
	Víquez, Carlos y Lety	IL21
Base de Datos 3	Roberto González Beas	MA93
	Sr. García José	CT15
	Sr. y Sra. Carlos Víquez	IL21

¿Cómo estandarizar nombres y otros atributos importantes?

Legado de información sepultado en campos sin formato

¿Quién es el cliente?

Nombre-Línea 1	Roberto González Ing. Roberto González Jr.
Nombre-Línea 2	Banco Nacional de México
Nombre-Línea 3	IPN Torres y Marco Ramírez UNAM
Domicilio-Línea 1	Ahuehuetes 25 Los Morales 3-30-8959
Domicilio-Línea 2	Esq. Alcanfores
Domicilio-Línea 3	Cuautilán, Méx. 54800



La pesadilla de las anomalías

NumCte	Nombre	Domicilio	Tipo
90328574	Hewlett Packard	Periférico 1520 Blvd. San Miguel CP 06720	OC
90328575	HP	Perif. 1520 Boulevard San Miguel CP 06720	OC
90328575	Hewlett P	Periferico 1520 B Sn Miguel CP 06720	\$#%
90233479	Hewlett Packard Corp	Periferico 1520 Boul. Sn. Miguel CP 06720	Comp
90233489	HP Consulting	Reforma 275 Cuauhtémoc CP 02200	Consult
90345672	H Packard Services	Centro Santa Fe México. D.F/ CP 06531	Sys Int

No es llave única

Anomalías

Ortografía incorrecta

No estandarizado

Basura en campos opcionales

¿Cómo identificar y consolidar correctamente las anomalías de millones de registros?

Sorpresas en valores de datos

Nombre	NSS	Teléfono
Roberto González Ing.	228-03-1975	6172380900
Josee García	99999999	2354516
Carlos y Lety Viquez	025-34-9987	
Ahuehuetes 25 Los Morales	34-2345678	415-234-200
Banco de México	10101010	508-276-400
Jorge Rasso, Inspector	18-2345678	Fax 5287776

- Metadata ... limita la vista a descripciones de datos solamente.
- Prácticas de negocio ... cambian con el tiempo.
- Ingreso de datos ... burla las reglas de negocio.

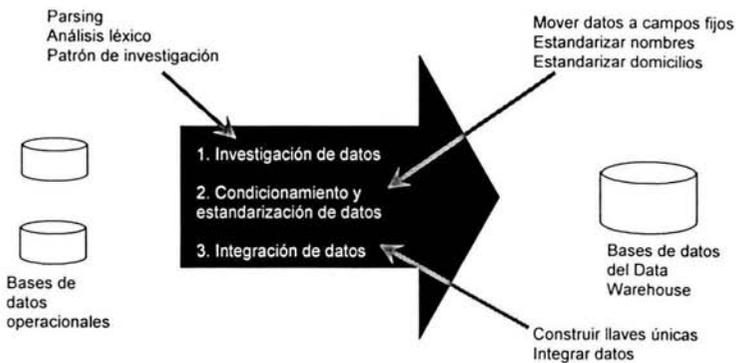
Cuestiones en la migración de datos

- No existe llave única.
- Muchos nombres en una línea.
- Un nombre en dos líneas.
- Nombre y domicilio en la misma línea.
- Nombre y razón social mezclados.
- Diferentes domicilios para el mismo miembro.
- Diferentes nombres y ortografía para el mismo cliente.

Respuestas a la migración de datos

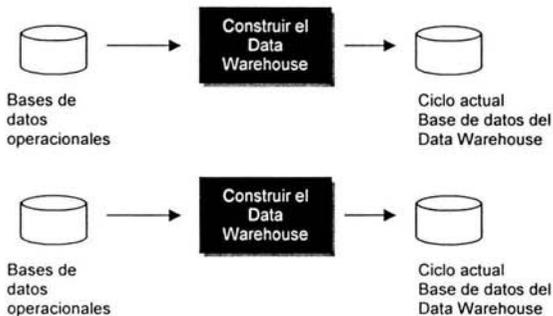
- Cobol y 4GL convencional
 - No pueden predecir y escribir si-entonces-reglas para todas las posibles variantes.
 - Dificultad para empatar y consolidar.
- Herramientas de propagación de la Metadata
 - Las descripciones de los datos no dicen la historia completa acerca de ellos.
- Herramientas de conversión de cajas negras
 - Aplicaciones genéricas pueden no solucionar sus problemas.
- Conversión personalizada Procesos/Herramientas.

Conversión personalizada Procesos/Herramientas



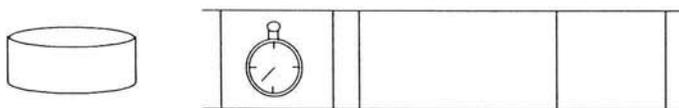
Poblando el Data Warehouse

Reemplazo total



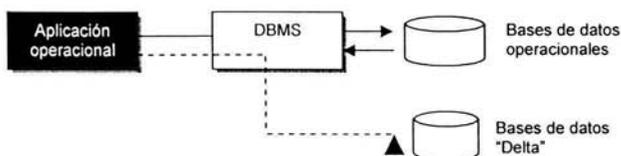
- **El Data Warehouse se construye desde cero en cada ciclo de negocio**
 - Costoso.
 - Los datos históricos están limitados a los datos almacenados en los sistemas operacionales.

Datos operacionales Time-stamping (marcas de tiempo)



- Las marcas de tiempo (Time stamps) identifican los datos que han cambiado desde la última extracción de datos.
 - El tiempo es parte de la llave de búsqueda.
 - No pueden reflejar los datos borrados.

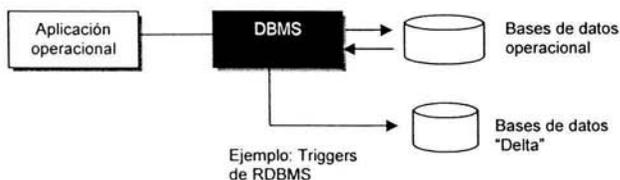
Base de datos "Delta"



- La base de datos "Delta" contiene los datos modificados.
 - Ocurre muy frecuentemente.
 - El mantenimiento de la aplicación es requerido ya que nuevos atributos/temas(sujetos) son agregados al Data Warehouse.

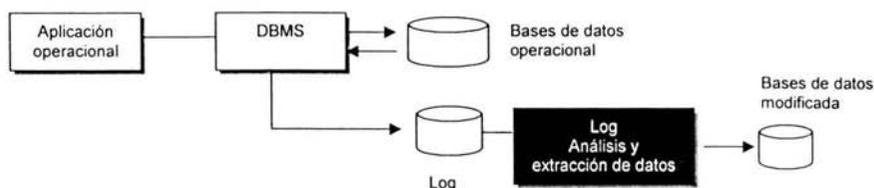
Triggers DBMS

Un trigger de BD es una unidad de programa en PL/Sql, almacenado en la BD, asociado con una tabla específica de la BD.



- Los datos modificados son intersectados al nivel del DBMS.
 - Los triggers pueden ser empleados para verificar la modificación de datos en una tabla, para derivación de valores en columnas de manera automática, para mantenimiento general de las tablas, etc.
 - Dependen del DBMS.
 - Requerimiento de I/O extra mientras el sistema está en línea.

Log de base de datos



- La cinta del Log contiene el historial de los cambios efectuado en la BD.
- Se pueden determinar las causas de corrupción y pérdida de datos, permite monitorear los cambios en los datos y es útil en la recuperación de datos.
- Puede ser difícil interpretar el formato del Log.
- Debe esperar por el sistema de control.

Diferencia entre instancias de la base de datos



- Los datos modificados resultan de comparar dos instancias de la misma base de datos.
- Tiempo muy costoso de CPU.
- Podría resultar muy complejo.

Consideraciones para la implementación

- La calidad de datos es el atributo más importante del Data Warehouse. Un proceso de extracción de datos bien definido es la clave para hacer que esto ocurra.
- Debido a la complejidad de los datos almacenados heredados de las bases de datos, con frecuencia las herramientas de extracción/movimiento de datos no alcanzan a cubrir las expectativas.
- A pesar de la funcionalidad proporcionada por las herramientas de extracción de datos, los clientes, quizá, dejan todavía la tarea de escribir un importante número de datos extraídos a programas de limpieza y migración.
- No se debe desestimar la complejidad de recursos necesarios para ejecutar la extracción de datos.

Herramientas de extracción de datos

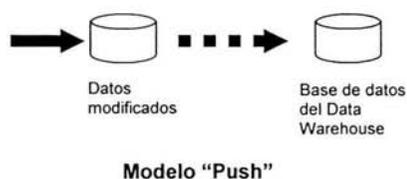
Compañía	Producto
Data Stage	Data Stage
Informática	Informática
Carleton	Carleton Passport Workbench
Evolutionary Technologies, Inc.	Extract Tool Suite
Vality	Vality Integrity Data Re-engineering Tool, Data Cleansing/Data Quality
Innovative Systems, Inc.	ISI Tool Set, Data Cleansing/Data Quality

Acciones relevantes en una implementación de Data Warehouse:

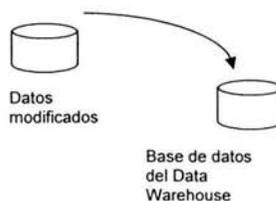
- Identificar el sistema de registro.
- Procesos de transformación de datos/limpieza de datos.
- Herramienta de extracción de datos.
- Herramienta de calidad de datos.
- Técnicas de modificación de datos en uso/planeadas.
- Frecuencia de refresco (reparación)

2.5. Administración de copias

Modelos de copias



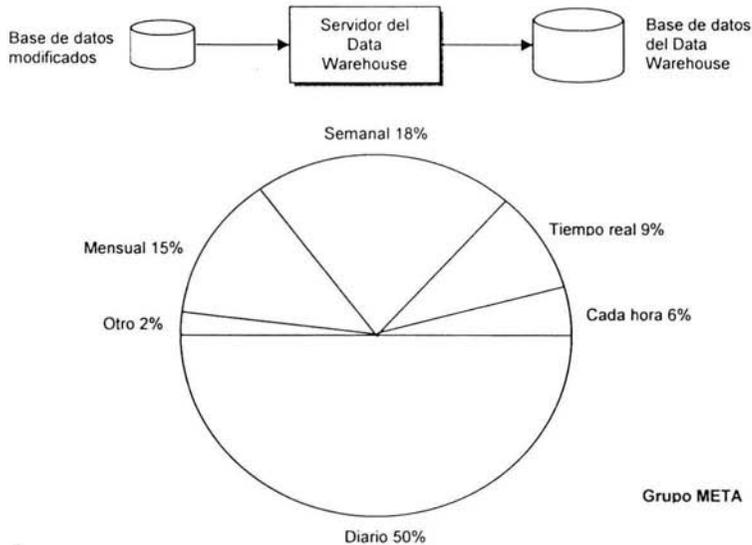
Modelo "Push"



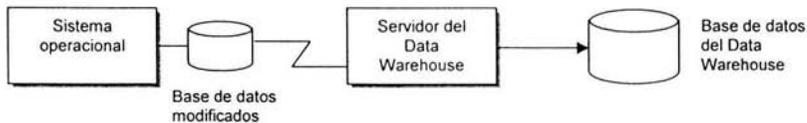
Modelo "Pull"

- **"Modelo Push"**: El ambiente operacional programa e inicia el proceso de movimiento de datos.
- **"Modelo Pull"**: El ambiente del Data Warehouse programa e inicia el proceso de movimiento de datos.

Refresco de datos



Cálculo de transferencia de datos



Supuestos:

- 300 Gb para la base de datos del Data Warehouse.
- 30 Gb de datos modificados.

Tasa de transferencia

- 1 Mb por segundo
- 10 Mb por segundo
- 100 Mb por segundo

Tiempo de transferencia

- 8.3 hrs.
- 50 mins.
- 5 mins.

Consideraciones para la implementación

- Volúmen de cálculo de datos.
- Ancho de banda de red – Tiempo de transferencia de datos.
- Frecuencia de movimiento de datos.
- Responsabilidad para disponibilidad de datos.
- Recreación de informe de los mismos datos debido a errores.
- Modelo Push/Pull.

Herramientas de extracción de datos

Compañía	Producto
Data Stage	Data Stage
Informática	Informática
Carleton	Carleton Passport
Evolutionary Technologies, Inc.	ETI Extract Tool Suite
Trinzic	Trinzic InfoPump
Sybase	Sybase Replication Server
Oracle	Oracle 7 Distributed Option
Legent	Legent Data Mover
Platinum	Platinum Integrator

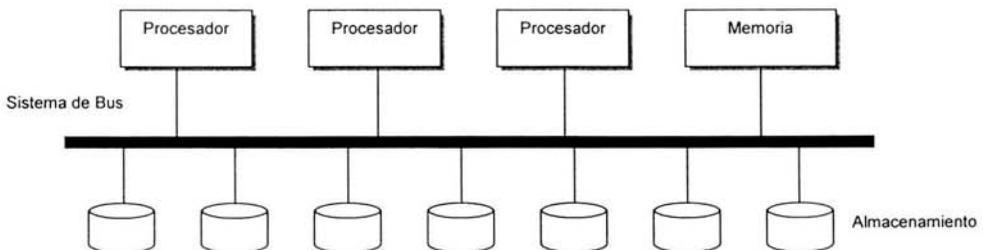
Acciones relevantes en una implementación de Data Warehouse:

- Volúmen de datos modificados.
- Cálculos de transferencia de datos.
- Responsabilidad de disponibilidad de datos.

2.6. Servidor del Data Warehouse

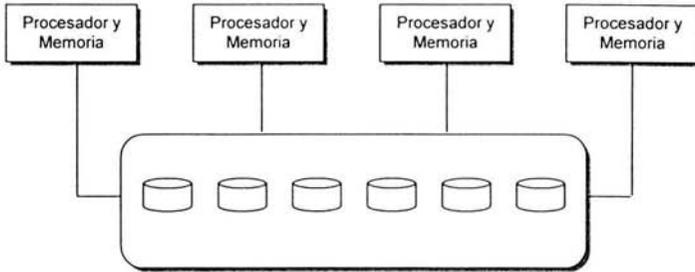
Arquitecturas de computadoras

Procesamiento de memoria compartida



- Servidores:
 - Rendimiento
 - Escalabilidad

Arquitectura agrupada

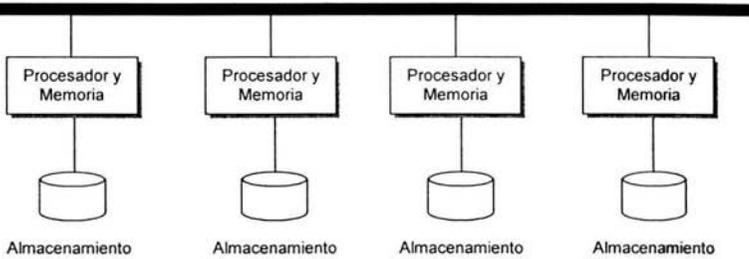


Memoria dedicada, periféricos compartidos

- Clusters:
 - Escalabilidad
 - Rendimiento
 - Disponibilidad 7x24 (días/horas)

Procesamiento masivo en paralelo

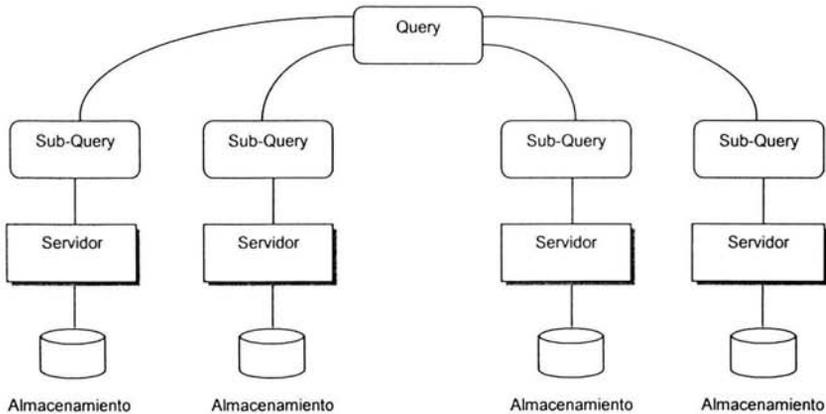
Interconexión escalable



Memoria dedicada, almacenamiento o rutas de almacenamiento dedicadas

- Teradata, Tandem Himalaya

Data Warehouse y Arquitecturas paralelas

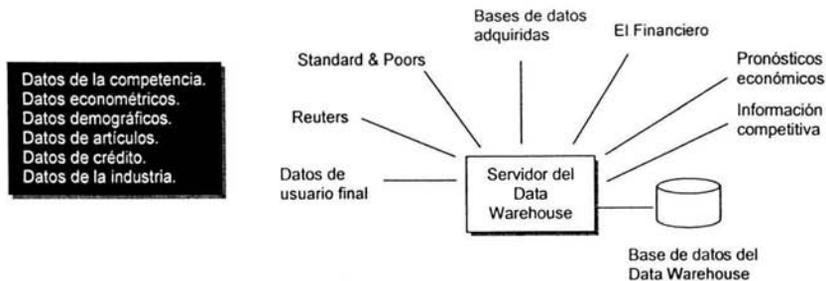


- Cargando en paralelo.
- Creación de índices en paralelo.
- Joins en paralelo.
- Exploración multi-usuario en paralelo.

Funcionalidad del servidor del Data Warehouse

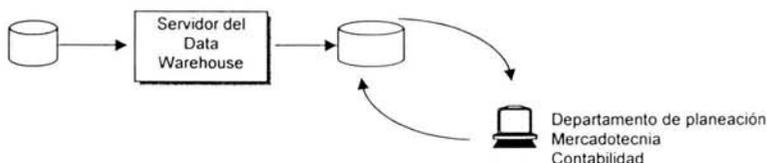
- Proporciona una fuente de datos sencilla, amplia y autorizada.
- Es responsable de la integración de los datos de múltiples sistemas operacionales.
- Contiene un catálogo del negocio.
- Administración Ad-Hoc y suscripción de ordenamiento de datos.
- Administración de la seguridad para los consumidores(usuarios), proveedores y autorizadores(quienes aprueban la información obtenida)

Datos externos (No operacionales)



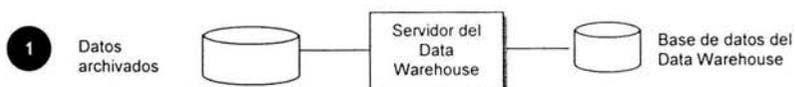
- Complementan los datos internos.
- Los DSS/EIS sólo pueden utilizar porciones seleccionadas de estos datos.
- Deben permitir a los usuarios finales seguir la pista de sus fuentes, por ejemplo: publicación, fecha, autor.
- Los usuarios finales deben ser capaces de saber el tipo y la disponibilidad de esos datos.

Fuentes internas de datos no operacionales

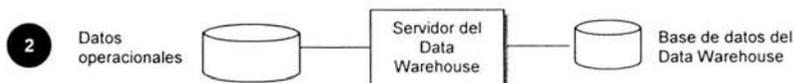


- Datos generados por un grupo o individuo para usarse por toda la organización.
 - Presupuestos
 - Pronósticos de productos
 - Objetivos de Ventas/Gastos/Ganancias
- La metadata incluye fuente y tiempo.

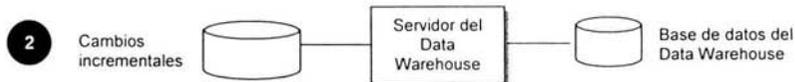
Cargando el Data Warehouse



Con datos archivados. Actividad de una sola vez. Rara vez hecha.

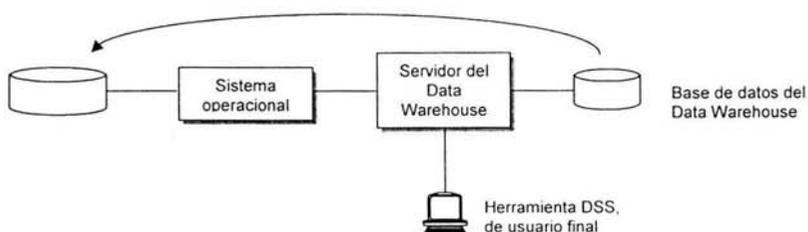


Con datos operacionales del día a día. Actividad de una sola vez.



Con datos modificados. Actividad actual.

Retroalimentación del Data Warehouse



- Los usuarios finales ejecutan la herramienta DSS. Los resultados son enviados al ambiente operacional.
- Paquete promocional dirigido a incrementar las ventas. Dos por uno, tarifas rebajadas, etc.

Consideraciones para la implementación:

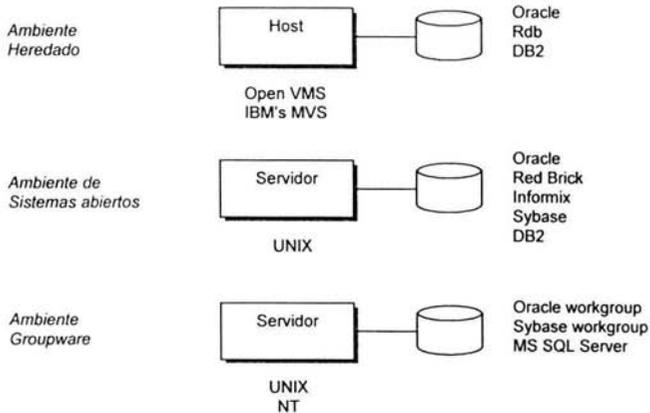
- Controles de seguridad y acceso para usuarios finales y administradores de datos.
- Escalabilidad del servidor. Tendencia del Data Warehouse a crecer muy rápido.
- Como mínimo, conduzca "detrás de los cálculos del sobre" para todas las actividades del establecimiento de una red. Evitar sorpresas.
- Sincronización de datos, por ejemplo: para un ciclo dado, todas las fuentes relevantes deben estar disponibles antes de que los datos sean enviados a los usuarios finales.
- Contabilizar la carga hacia atrás.
- Control de fugas de consultas.
- Puede ser necesario escribir programas tradicionales para refrescar el Data Warehouse.
- Ambiente/Estandares de sistema abierto.
- Conectividad de red. Apilados múltiples de la comunicación.

Acciones relevantes en una implementación de Data Warehouse:

- Configuración del hardware.
- Configuración del software.
- Fuentes externas.
- Fuentes internas (no operacionales)
- Mecanismos de retroalimentación del Data Warehouse.

2.7. Base de datos del Data Warehouse

Ambientes de base de datos



- Casi de manera invariable la base de datos del Data Warehouse es una base de datos relacional

OLTP vs. Bases de datos del Data Warehouse

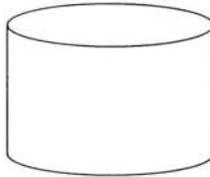
Base de datos OLTP	Base de Datos Data Warehouse
Altamente normalizado – Mejorar el rendimiento y proteger la integridad de datos	Altamente desnormalizado
Implementa integridad referencial en la base de datos (declarativa o disparadora)	Sin integridad referencial
Utiliza índices en poca cantidad	Altamente indexado
Utiliza candados extensivamente	Optimizado para lectura

Tamaño de la base de datos

- Base de datos multiusuario Low-end
 - Ambiente de trabajo en grupo
 - 500 Mb a 2 Gb
- Base de datos empresarial
 - 100 Gb a 500 Gb
 - Sistemas abiertos/heredados
- Base de datos muy grande
 - 2 Tb+
 - 1995: Oracle, Informix

Las bases de datos del Data Warehouse escalan a menudo/ rápidamente por encima de una orden de magnitud.

- Base de datos ultra grande
 - 2 Tb++
 - 1995: Tandem, Teradata
 - 1996/98: Los comerciantes de DBMS harán obsoletas las máquinas de bases de datos.



Espacio del sistema
Datos
Índices

Datos: Espacio del sistema + Índices
1: 4

- Regla de Thumb: 20 Mb de datos utilizan 100 Mb de almacenamiento.

Atributos de base de datos

- Orientado a Sujetos.
 - Organizado por sujetos más que por aplicación.
- Alto Nivel de Integración de Datos.
 - Datos codificados, definiciones de datos estándar, etc.
- Volatilidad de Actualización.
 - Cargas por volumen.
- Longevidad de Datos.
 - 1-5 años.

Requerimientos de tecnología

- Consulta de datos en paralelo.
- Indexación de datos: eficiente, en paralelo.
- Carga de datos: eficiente, en paralelo.
- Escalabilidad de Base de Datos.
- Alta disponibilidad. Un ambiente similar al de producción.
- Amplia administración de sistemas.
- Estándares de sistemas abiertos.

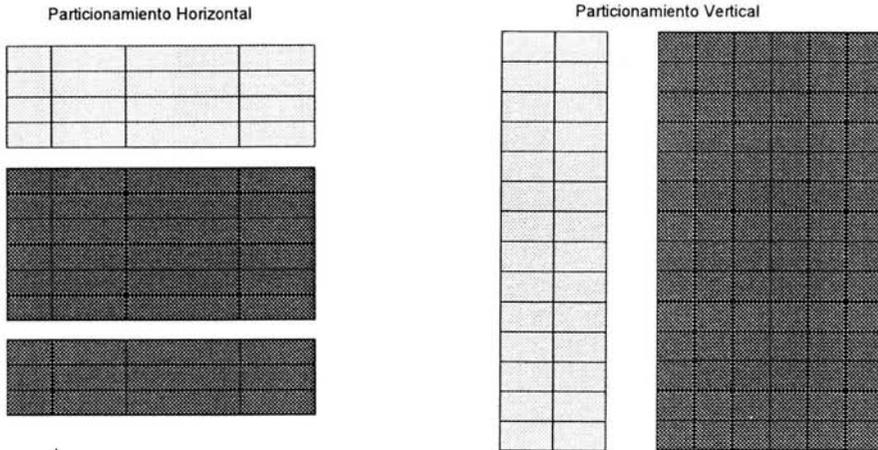
Diseño de base de datos lógico/físico

- No hay una buena razón para que falle el proyecto de Data Warehouse.
- Existen herramientas de base de datos que son suficientes para Data Warehouse basados en RDBMS.
- Probadas metodologías de diseño fácilmente disponibles.
- Estudiar la conducta de capacidad de planeación.
- Ejecuta herramientas para monitorear y afinar el rendimiento.
- Una pronta inversión en diseño de base de datos es la llave.

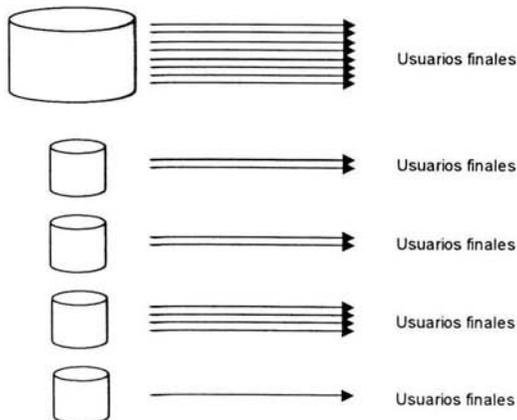
Dos temas del Data Warehouse

1. Particionamiento

"En el ambiente del Data Warehouse, la cuestión no es si los datos serán particionados, sino cómo serán particionados"²

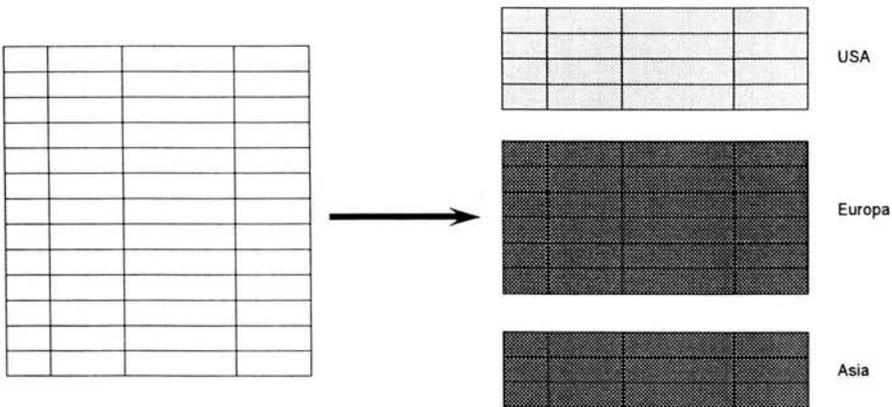


Particionamiento a nivel de aplicación



² William H. Inmon; *Building the Data Warehouse*, John Wiley, 1996

- Criterios para Particionamiento.
 - Basado en el tiempo: Año, Mes.
 - Geografía: Norte, Sur.
 - Frecuencia de acceso: Bajo y alto volúmen de consultas.
 - Granularidad: Bajo y alto nivel de detalle.
 - Funcional: Mercadotecnia, Manufactura.
 - Producto: Dispositivos domésticos, bienes industriales.

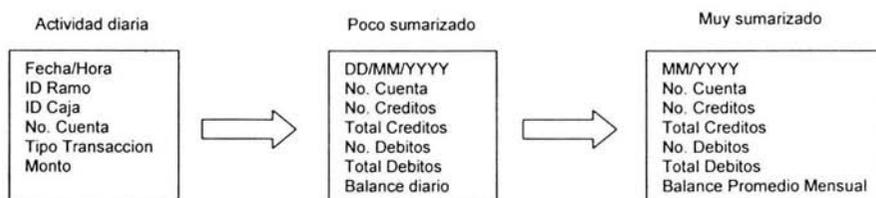


- Fomenta el paralelismo – Consultas, cargas, indexación. Cada partición contiene su propio conjunto de índices.
- Podría dividirse para separar el procesador, por ejemplo: configuración de grupo.
- Pero ... múltiples tablas que administrar.

Ventajas del particionamiento de la base de datos del Data Warehouse

- Mejorando el tiempo de respuesta para las consultas.
- Distribuye I/O a través de los discos para búsquedas paralelas.
 - Distribución de datos estática, física.
 - Por valor o rango, expresión, round-robin.
- Las particiones no afectadas son saltadas durante el proceso de consulta.
- Transparente para aplicaciones y usuarios.

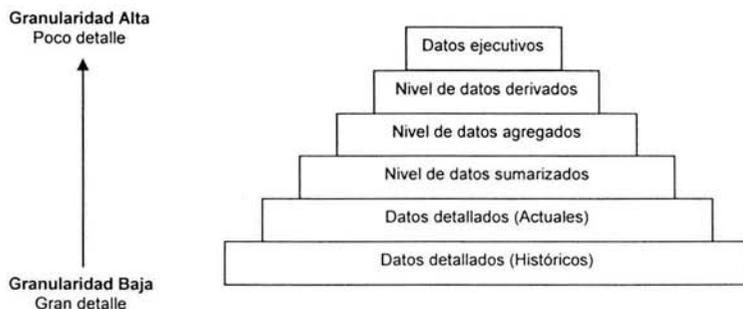
2. Granularidad



Ejemplo de las Actividades Bancarias

- ¿Qué datos se deben mover al Data Warehouse y cuáles son las compensaciones?
 - ¿En qué ramo hace mi cliente la mayoría de sus negocios?
 - ¿Cuál es el período donde ocurre la mayor parte de su actividad?

Niveles de granularidad



Ejemplo de Tarjeta de Crédito:



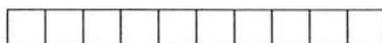
Detalle actual

ESTE MES
Compras detalladas
Anticipos detallados
Cargos financieros
Ratino



Actividad mensual – Poco sumariado

ESTE MES
Compras totales
Anticipos totales
Cargos financieros
Ratino



Actividad anual – Muy sumariado

ESTE MES
Compras anuales
Anticipos anuales
Cargos financieros anuales
Ratino suietado

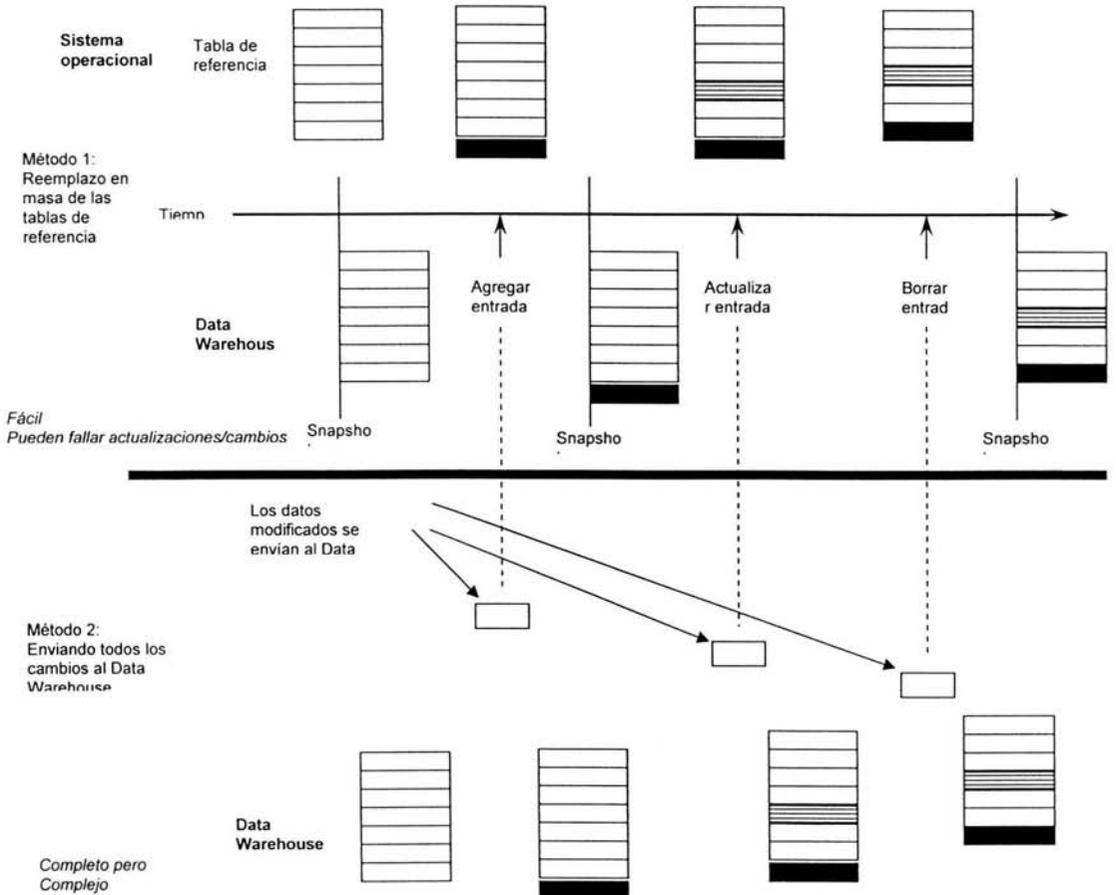
- Mayor nivel de detalle = Menor nivel de granularidad

Múltiples niveles de granularidad

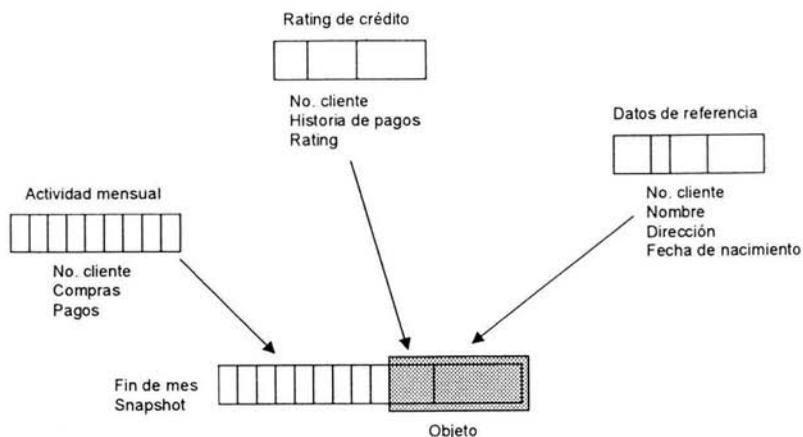
Ventajas:

- Servicio mejorado para usuarios finales
 - Mejor tiempo de respuesta.
- Mejor utilización de recursos
 - Almacenamiento y CPU.
- Realiza el proceso de análisis
 - Permite hacer drill-down (exploración) hacia datos de un nivel mayor de detalle.
- Permite múltiples tipos de datos sumariados
 - Acumulativo simple, acumulativo multi-nivel, resúmenes de balance.

Tablas de referencia



Integridad referencial – Objetos del Data Warehouse



Consideraciones para la Implementación

- Enfocarse primero en la flexibilidad.
 - Guardar el dato de la transacción básica en el menor nivel de granularidad que sea posible.
- Ajustar el diseño para cumplir con los constraints físicos.
 - Desnormalizar.
 - Sumarizar
- Ajustar el diseño para cumplir con los requerimientos del usuario final.
 - Tiempo de respuesta.
 - Requerimientos de exploración (drill-down)

El diseño real de la Base de Datos es siempre una cuestión de compensaciones.

Sistemas de Administración de Base de Datos Relacional (RDBMS)

Compañía	Producto
Oracle Corp.	Oracle 11i
	Rdb
Red Brick Systems	Red Brick Warehouse VPT
Microsoft	MS SQL Server
Informix Software, Inc.	Informix – OnLine Dynamic Server for Unix
	CA - Ingress
Sybase	Sybase
	Allbase

Pasaporte de datos (Gateway)

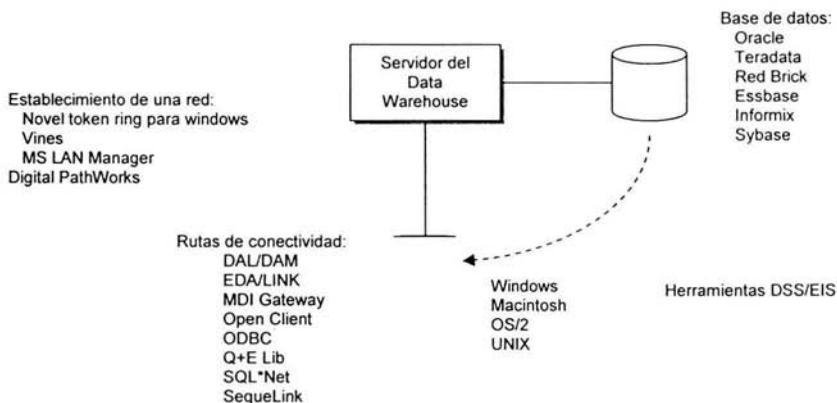
Compañía	Producto
Sybase	Enterprise Connect
Trinzic	InfoHub
Information Builders, Inc.	EDA/SQL
TechGnosis	PC SequeLink
Gupta	DB2 Gateway CA-Ingress Gateways
Oracle Corp.	Open Connect
Intersolv	Q&E
Netwise	Trans-Access

Acciones relevantes en una implementación de Data Warehouse:

- Software de base de datos.
- Gateways de acceso a datos.
- Administrador de datos.
- Cálculos del tamaño de la base de datos.
- Particionamiento.
- Granularidad.
- Programas de restauración de la base de datos.

2.8. Acceso y reporte de datos

Ambiente de acceso y divulgación de los datos



¿Quiénes son los usuarios finales?

- Altos Ejecutivos
- Gerentes
- Analistas
- Planificadores
- Desarrolladores
- Usuarios

Tipos de usuarios finales

- Usuarios que manejan la información.
 - Planificadores
 - Analistas
 - Gerentes
 - Desarrolladores



- Ejecutivos de alto nivel
 - Director financiero
 - Vicepresidente de mercadotecnia
 - Estrategas corporativos



¿Qué es lo que hacen los usuarios finales?

- Usuarios que utilizan la información (quienes conocen el negocio)
 - Ante todo – una persona del negocio. En segundo lugar – un técnico.
 - Ejecutan herramientas DSS.
 - Entregan reportes analíticos a la gerencia.
- Altos ejecutivos. Los últimos usuarios a los que llega la información del Data Warehouse.
 - Dirigen el negocio.
 - Ponen énfasis en la competitividad y rentabilidad.

Mentalidad del usuario final

“Deme lo que digo que deseo, entonces puedo decirle lo que realmente deseo”

Para tener éxito, el usuario final debe ser capaz de explorar las posibilidades.

Perfiles de los usuarios finales

- **Gerentes (Encargados)**
 - Buscan el fondo del asunto.
 - El paradigma “presionar un botón”. La complejidad de las consultas es ocultada debajo de las herramientas EIS.
- **Analistas.**
 - Juegan en aguas conocidas (Aprovechan el conocimiento del medio)
 - Consultas “paramétricas” – de acuerdo con el punto de partida con alguna variación en la ejecución.
- **Usuarios.**
 - Obtienen victorias y derrotas.
 - Completamente ad-hoc, buscando relaciones no anticipadas.

Diseñe el Data Warehouse para los Analistas. Más de un conjunto de herramientas se puede requerir para satisfacer a los Encargados, Analistas y Usuarios.

Características de los usuarios finales

Gerentes (Encargados)

- Respuestas en formatos específicos.
- Resultados repetitivos, presentación constante.
- Herramientas "invisibles".
- Ningún requisito para aprender o recordar.

Analistas

- Una cierta flexibilidad – permite variables en tiempo de ejecución.
- Adaptable para ciertas funciones.
- Resultado final en la herramienta de acceso.
- Análisis de escritorio ligado.

Usuarios

- Cualquier pregunta, cualquier dato, en cualquier momento.
- Crea los resultados solicitados cada vez.
- Soporta muchas iteraciones.
- Utilizan el gran poder, de la mejor manera posible, de las herramientas de escritorio.

Consideraciones para la implementación

- Elegir herramientas de DSS/EIS: Funcionalidad de DBMS + Establecimiento de una red, soportadas.
- Herramientas de DSS/EIS ya utilizadas pueden limitar las opciones.
- Soporte a usuarios finales: Help Desk, capacitación.
- Entregar mecanismos a los usuarios finales (solicitados vs. suscripciones)
- Administración de seguridad.
- Un experto en desarrollo de aplicaciones puede ser requerido para modificar y proporcionar la funcionalidad necesaria por los usuarios finales.

Herramientas de Usuario final

Permiten a los usuarios acceder datos del Data Warehouse. Proporcionan funciones tales como consultas, reportes, gráficas, estadísticas, modelado, razones de negocio y/o herramientas de análisis.

- Herramientas de soporte a las decisiones (DSS)
- Herramientas de soporte a ejecutivos (EIS)
- Herramientas de consultas y reportes.

Herramientas DSS/EIS

Compañía	Producto
Software Interfaces	SQLASSIST
Business Objects Inc.	Business Objects
MicroStrategy Inc.	DSS Agent
Software AG	Esperant, Gentium GQL
Open Data Corp.	FindOut!
IQ Software Corp.	Intelligent Query
Andyne Computing Ltd.	GQL and Pablo
Brio Technology	BrioQuery Designer, BrioQuery Navigator, BrioQuery Explorer
Cognos	PowerPlay, Impromptu
Holistic Systems	Holos
Methaphor	DIS
Information Advantage	IA Suite
Comshare	Commander EIS
Centura	Centura SQL Windows
IntelligenceWare Inc.	Iconic Query
Microsoft Corp.	MS Access
Knowledgeware	Clear Access
Trinzic	Forest & Trees
SAS Institute	SAS Tools
Pilot (D&B)	LightShip

Herramientas de Desarrollo de Aplicaciones

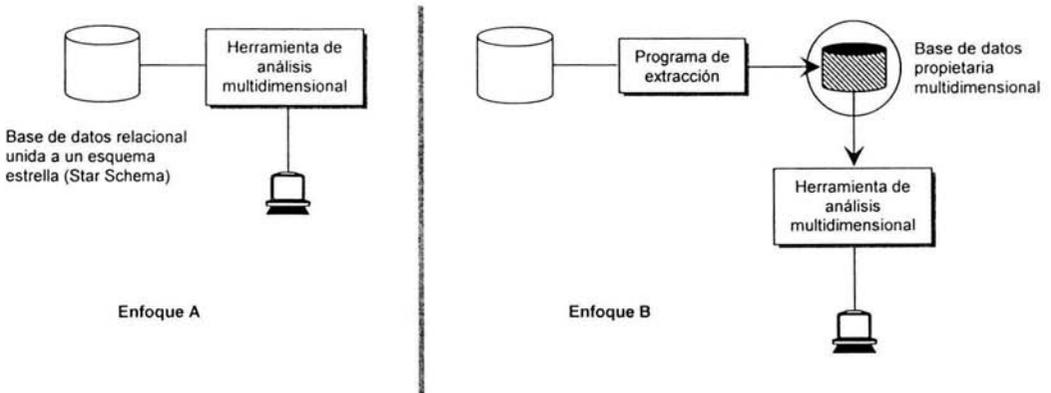
Compañía	Producto
Microsoft Corp.	Visual Basic
Information Advantage Inc.	AXSYS
Information Builders Inc.	Focus/RIS for Windows
Planning Sciences Inc.	Gentium
Powersoft Corp.	Power Builder
Easel Corp.	Object Studio
Business Objects Inc.	Business Objects
Intersolv Inc.	Q+E
Holistic Systems Inc.	Holos

Acciones relevantes en una implementación de Data Warehouse:

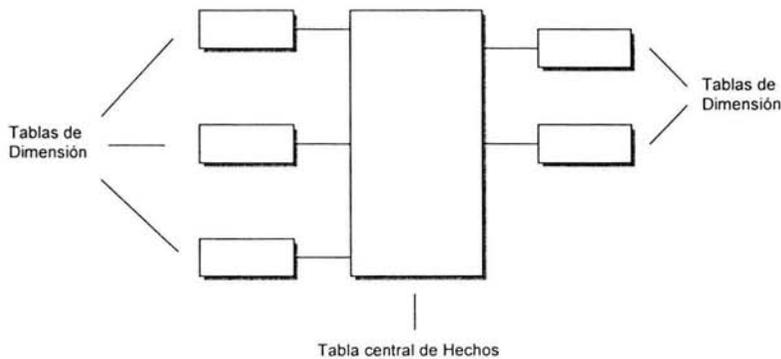
- Herramientas DSS/EIS.
- Herramientas de desarrollo de aplicaciones.
- Número de usuarios finales.
- Tipo de consultas permitidas.

2.9. Procesamiento Analítico en Línea (OLAP)

Bases de datos OLAP – El debate, bases de datos multidimensionales vs. relacionales



Diseño lógico – Modelado dimensional³



- Tabla de Hechos.
 - Tabla central en la base de datos.
 - Tabla grande, muy grande.
 - Contiene números variables en una transacción (por ejemplo: artículos de línea en una factura)
- Tabla de Dimensión.
 - Contienen constraints, usados al formar subconjuntos de los datos.
 - Tablas pequeñas.

³ Ralph Kimball: *Datawarehouse (Proceso de datos automático)*, 1 de Junio de 1994.

Diseño lógico – Modelado multidimensional⁴

Definir las respuestas deseadas con un diseño dimensional es un proceso intuitivo que implica constraints de los hechos que se desean ver con las cualidades encontradas en las tablas de dimensión. Constraints típicos son:

- Dimensión Artículo
 - Artículos de los cuales un encargado específico es responsable.
 - Artículos usados por la industria de la impresión.
 - Todos los artículos excepto aquellos que pesan más de 500 libras.
 - Todos los artículos también vendidos por sus competidores de la costa este.

- Dimensión Cliente
 - Clientes, no de la cadena, en estados del sur.
 - Clientes que llevan solamente su producto.
 - Nuevos clientes en la región de ventas del este.

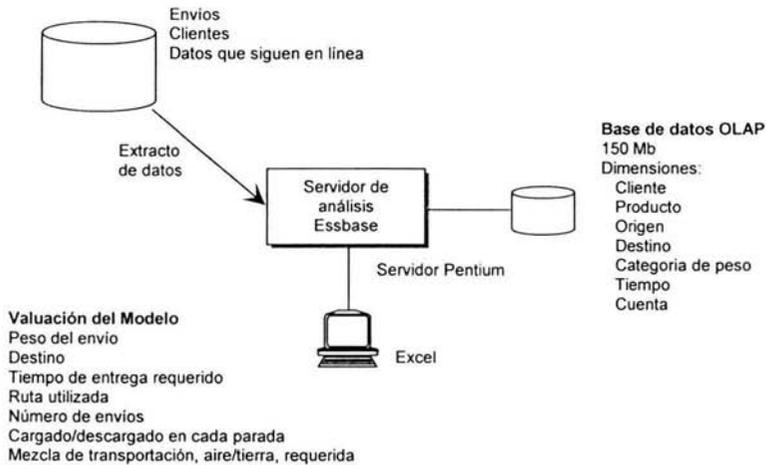
- Dimensión Tiempo
 - Primer trimestre fiscal de 1994.
 - Primer trimestre fiscal de 1995.
 - Este mes en el periodo de ventas del día del padre.

Bases de Datos Multidimensionales vs. Relacionales

Multidimensional	Relacional
<p><u>Ventajas</u></p> <ul style="list-style-type: none"> • Optimizado para consultas de DSS (analítico) • Representa el panorama del negocio • Soporta cruces de cálculos en la dimensión 	<p><u>Ventajas</u></p> <ul style="list-style-type: none"> • Optimizado para consultar/actualizar registro por registro • Acceso ilimitado a cualquier combinación de atributos • Menos capacitación de producto a producto
<p><u>Desventajas</u></p> <ul style="list-style-type: none"> • El rendimiento se degrada mientras la base de datos aumenta • Limitaciones de la base de datos (2 a 20 Gb) • Carece de funcionamiento de afinación, herramientas de monitoreo • Respaldo/almacenamiento/refresco incremental/seguridad, no son sofisticados 	<p><u>Desventajas</u></p> <ul style="list-style-type: none"> • Debe estar de-normalizada • Tiene que almacenar agregaciones para funcionar eficientemente

⁴ Ralph Kimball: *Datamation (Proceso de datos automático)*, 1 de Junio de 1994.

Aplicaciones OLAP



Checklist de funcionalidad de herramientas OLAP

- Visión multidimensional.
- Pivote/Rotación.
- Exploración (Drill Down/Up)
- Cruce de cálculos en Dimensiones.
- Conjuntos dinámicos.
- Filtros.
- Capacidades de decisión Groupware.
- Browsing plegable.
- Definiciones flexibles del periodo.
- Acceso directo a Bases de Datos Relacionales.
- Metadata (Información acerca de los datos)
- Ayuda escasa del esquema de matriz.
- Lectura y escritura.
- Generación de consultas.
- Credibilidad.

Consideraciones para la implementación

- Los servidores son típicamente bloqueados por LAN.
- Rara vez pueden procesar conjuntos de datos arriba de 10 Gb.
- RDBMS comerciales que se mueven en el mercado de los MDBMS.

Herramientas OLAP

Compañía	Producto
Pilot	Pilot Lightship
Arbor Software Inc.	Essbase Analysis Server
IRI	Express
Kenan	Accumate
Dimensional Insight Inc.	Cross Target
Information Advantage Inc.	IA Decision Support Suite

Acciones relevantes en una implementación de Data Warehouse:

- Herramientas OLAP.
- Tamaño del MDBMS (Sistema Manejador de Base de Datos Multidimensional)
- Número de usuarios.
- Topología.

Capítulo 3.

Plan de Implementación del DWH

Objetivo:

Discutir acerca de una plantilla de implementación para el proyecto de Data Warehouse perfilando las cuestiones críticas más importantes en la realización de un piloto exitoso de Data Warehouse.
Discutir cuestiones acerca de un gran proyecto de Data Warehouse.

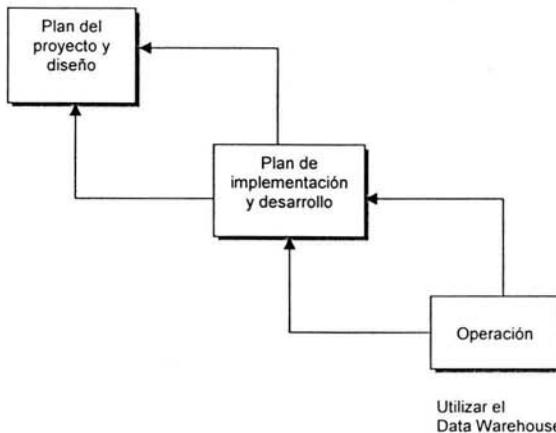
3.1. Esquemas de implementación

Enfoque evolutivo (“Iterativo”)

- Inicio pequeño
- Probar la utilidad
- Construir en éxitos

Evite una gran explosión del enfoque

Plantilla del enfoque evolutivo (“Iterativo”)



El Data Warehouse se construye de una manera ordenada, iterativa y un paso a la vez.

Paso 1: El Piloto del Data Warehouse.

Plan del proyecto

- Actividades
 - Seleccione un área funcional.
 - Identificar la forma de dar los reportes los directivos de alto nivel.
 - Definir las tareas del proyecto y los entregables.
 - Definir los participantes.
 - Determinar la agenda y el presupuesto del proyecto.
 - Obtener recursos.
 - Capacitar al equipo del Data Warehouse.
- Entregables
 - Plan de trabajo del proyecto, agenda y presupuesto.
 - Participantes capacitados.
 - El equipo del Data Warehouse de acuerdo con los objetivos del proyecto.

Actividades de diseño

- Actividades
 - Identificar requerimientos.
 - Revisar el modelo de datos.
 - Desarrollar la arquitectura del Data Warehouse.
 - Definir el ambiente óptimo de tecnología.
 - Definir los incrementos del Data Warehouse.
 - Desarrollar el plan de implementación.
- Entregables
 - Arquitectura del Data Warehouse.
 - Plan de Implementación del Data Warehouse.
 - Plan de soporte al Data Warehouse.

Plan de desarrollo

- Actividades
 - Definir el modelo lógico de datos del Data Warehouse.
 - Diseño del modelo físico del Data Warehouse.
 - Determinar las fuentes de datos.
 - Desarrollar especificaciones de conversión/traducción.
- Entregables
 - Modelos de datos, lógico y físico del Data Warehouse.
 - Mapa de datos fuente y destino.
 - Especificaciones de migración y transformación.

Desarrollo

- Actividades
 - Construir el ambiente físico de base de datos.
 - Desarrollar programas de migración y transformación.
 - Ejecutar programas de extracción y carga de datos hacia el Data Warehouse.
 - Realizar pruebas de validación y verificación.
- Entregables
 - El Data Warehouse trabajando.

Operación

- Actividades
 - Capacitar y dar soporte a los usuarios.
 - Realizar extracciones y cargas periódicas.
 - Optimizar el rendimiento de la base de datos.
- Entregables
 - Un Data Warehouse mejorado.
 - Datos para análisis y toma de decisiones.

3.2. El Piloto del Data Warehouse

Criterio para la selección del piloto

- **Importancia** – La información entregada por el piloto debe ser relevante para el negocio. Ser específico. Por ejemplo: reducir el problema de préstamos.
- Elegir el Project Manager adecuado.
- Riesgo técnico bajo: realizable. **Inicio pequeño.**
- Necesidad clara de la integración del nivel de detalle de los datos de la cuenta.
- Necesidad de datos que cubran múltiple períodos de tiempo, cálculos y sumalizaciones.
- Uso significativo anticipado.

Template de Implementación

- Construir o conseguir un modelo de datos de alto nivel.
- Seleccionar una parte del modelo de datos para implementarla.
- Construir un modelo de datos para el piloto.
 - Un tema.
 - Un subconjunto de un tema.
 - Un subconjunto de múltiples temas.
- Determinar la configuración del Data Warehouse.
- Crear las tablas de la base de datos.
- Utilizar herramientas o construir programas de extracción de datos.
- Cargar y refrescar la información del Data Warehouse.
- Seleccionar herramientas de acceso a datos/reporteo.
- Emplear un piloto.

Reunir requerimientos

- Revisar los objetivos de negocio de la empresa.
- Observar el trabajo hecho por compañías que han tratado áreas similares.
- Dirigir sesiones de descubrimiento con la gente interesada en el éxito del Data Warehouse.
- Revisión de planes de negocio estratégicos.
- Familiarizarse con los actuales reportes de DSS.
- Revisar cómo se calculan los indicadores críticos del negocio.
- Atender los seminarios de la industria.
- Obtener los templates específicos de los modelos de datos de la industria.
- Entrevistar a la gente correcta: Ejecutivos, Jefes, Usuarios finales.
- Afinar preguntas. Por ejemplo: ¿Qué información necesita para que funcione el negocio?, en comparación con,
- ¿Qué información necesita en el Data Warehouse?

3.3. Cuestiones de dirección

Comité de dirección

- Considerar un comité de dirección compuesto por ejecutivos del negocio y usuarios finales para ...
 - Proporcionar la dirección del negocio.
 - Tomar decisiones respecto a cuestiones de implementación en grandes empresas.
 - Establecer prioridades.
 - Ayudar con la asignación de recursos.
- ⇒ No sustituye al Project Manager del Data Warehouse.



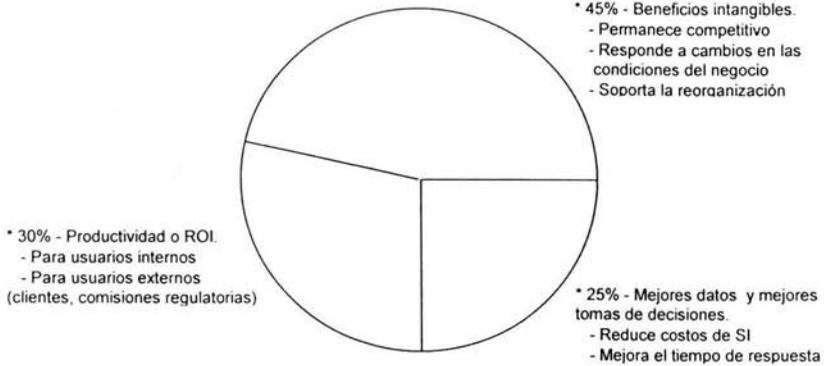
Defensor del Data Warehouse

- Llevar el Data Warehouse a la vida.
- Promoverlo activa y entusiastamente.
- Respaldar la construcción.
- Vencer la resistencia.
- Articular la visión del Data Warehouse.

Factores críticos de éxito

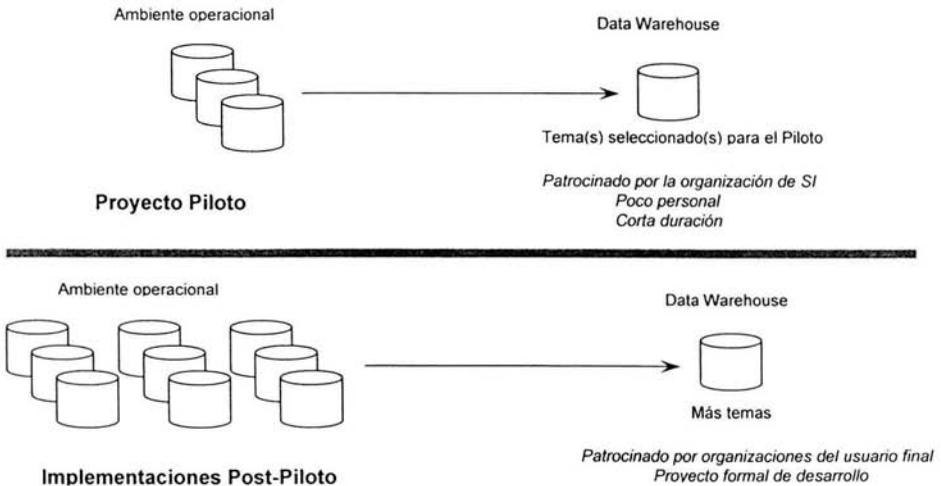
- El Data Warehouse será exitoso si ...
 - 3 Agrega valor al proceso de toma de decisiones.
 - 3 Inicia con pocos temas de negocio
 - 3 Almacena una cantidad de datos manejable.
 - 3 Puede ser actualizado de una manera oportuna.
 - 3 Soporta el acceso a modelos solicitados por los usuarios de la información.
 - 3 Puede ser entendido por la comunidad del negocio.
 - 3 Da integridad a los datos.
- ⇒ Tiene un sólido compromiso de la dirección.

Justificación financiera



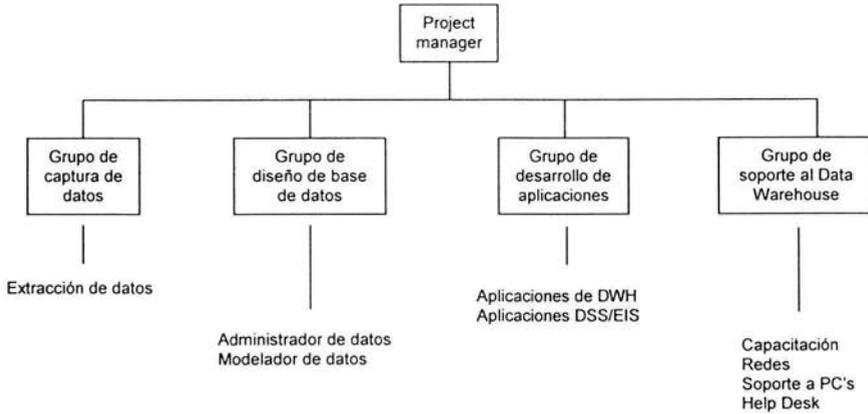
De Asociados en Bases de Datos, "Data Warehouse en la Práctica", Junio de 1993.

Patrocinador del Data Warehouse



De Bill Inmon - Orlando, Marzo de 1995.

3.4. Organización del Data Warehouse



Capacitación a usuarios finales

- Capacitación acerca del negocio
 - Cómo utilizar los datos.
 - Cómo utilizar el Data Warehouse para responder a cuestiones del negocio.
 - Técnicas de análisis del negocio.
 - Consultas y reportes predefinidos del negocio.
 - Cómo y cuándo llamar a help desk.
- Capacitación acerca de la tecnología
 - Cómo utilizar las herramientas.
 - Herramientas DSS/EIS.
 - Texto sobre tecnología relacional (tablas, renglones, columnas, relaciones)
 - Elementos de datos almacenados en el Data Warehouse.
 - Guías y tutoriales de capacitación.
 - Desarrollo de consultas ad-hoc.
- Capacitación acerca del Data Warehouse
 - Cómo tomar ventaja del ambiente del Data Warehouse.
 - Consultas a la metadata (información acerca de los datos)

Cuestiones de implementación (Trampas de la puesta en práctica)

- Utilizar la misma base de datos para OLTP y DSS.
- Tomar demasiada información.
- No mostrar resultados suficientes con prontitud.
- Presupuesto insuficiente.
- Paralización de la evaluación.
- Pobre calidad y disponibilidad de datos.
- Respaldo insuficiente de los ejecutivos.
- Fijar expectativas excesivas.

Línea final

- Personal del equipo de implementación con gente de experiencia.
 - Con reputación.
 - Con historial.
- Seleccionar los mejores servidores de SMP (Multi-Procesamiento Simétrico)
- Utilizar tecnología RDBMS con capacidad de procesamiento de consultas en paralelo.
- Afinar el rendimiento de consultas.
- Diseñar los procesos de extracción y carga será la tarea más difícil.

Capítulo 4.

Conclusiones

Objetivo:

Destacar los puntos principales en la generación de un Data Warehouse, por medio de la teoría descrita y la experiencia adquirida en este concepto de la administración de sistemas.

4.1. Conclusiones

Esta tecnología (Data Warehouse) permite que la organización disponga, en forma integrada y estandarizada, de la información correspondiente a la operación de la empresa, así como, proporciona a los usuarios, que tienen a su cargo la toma de decisiones, las herramientas adecuadas, para que a través de consultas rápidas, ellos mismos accedan a la información requerida.

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

El ingreso de datos en el Data Warehouse viene desde el ambiente operacional en casi todos los casos. El Data Warehouse es siempre un almacén de datos transformados y separados físicamente de la aplicación donde se encontraron los datos en el ambiente operacional.

Una de las razones por las que el desarrollo de un Data Warehouse crece rápidamente, es que realmente es una tecnología muy entendible. De hecho, el Data Warehouse puede representar mejor la estructura amplia de una empresa para administrar los datos informativos dentro de la organización. Soporta el procesamiento informático al proveer una plataforma sólida, a partir de los datos históricos para hacer el análisis. Facilita la integración de sistemas de aplicación no integrados. Organiza y almacena los datos que se necesitan para el procesamiento analítico-informático sobre una amplia perspectiva de tiempo.

El Data Warehouse es, actualmente, el centro de atención de las grandes organizaciones, porque provee un ambiente para que hagan un mejor uso de la información que está siendo administrada por diversas aplicaciones operacionales.

Una Arquitectura Data Warehouse es una forma de representar la estructura total de datos, comunicación, procesamiento y presentación, que existe para los usuarios finales que disponen de una computadora dentro de la empresa.

Hay algunas consideraciones adicionales que deben tenerse en cuenta al construir y administrar el Data Warehouse:

La primera consideración es respecto a los *índices*. La información de los niveles de esquematización más altos pueden ser libremente indexados, mientras que las de los niveles más bajos de detalle, por ser tan voluminosa, pueden ser indexados moderadamente.

Por lo mismo, los datos en los niveles más altos de detalle pueden ser reestructurados fácilmente, mientras que el volumen de datos en los niveles más inferiores es tan grande, que los datos no pueden ser fácilmente reestructurados.

Por consiguiente, el modelo de datos y el diseño clásico fundamentan que el Data Warehouse se aplique casi exclusivamente al nivel actual de detalle. En otras palabras, las actividades de modelado de datos no se aplican a los niveles de esquematización, en casi todos los casos.

Otra consideración estructural es la *partición de la información* en el Data Warehouse. El nivel de detalle actual es casi siempre particionado.

La partición puede hacerse de dos maneras: al nivel de DBMS y al nivel de la aplicación. En la partición DBMS, se conocen las particiones y se administra por consiguiente. En el caso de la partición de las aplicaciones, sólo los programadores de las mismas conocen las particiones y la responsabilidad de su administración es asignada a ellos.

Al interior de las particiones DBMS, muchos de los trabajos de infraestructura se hacen automáticamente. Pero existe un elevado grado de rigidez asociada con la administración automática de las particiones. En el caso de las particiones de las aplicaciones del Data Warehouse, la mayor parte del trabajo recae sobre el programador, pero el resultado final es que la administración de datos es más flexible.

Hay factores decisivos para determinar el desarrollo de un Data Warehouse:

Los datos "sucios" son un serio peligro para el éxito de un proyecto de Data Warehouse. Dependiendo del alcance del problema, simplemente podría no ser posible dirigirlo rápidamente y abaratarlo.

Los principales factores son:

- El tiempo que toma la programación interna
- El costo de las herramientas

Los gerentes de proyectos de Data Warehouse necesitan evaluar el problema con realismo, los recursos internos disponibles para distribuirlos y seleccionar la solución que se adapte al equipo y presupuesto del proyecto, o modificar el equipo y el presupuesto para solucionar el problema.

Fase de implementación

En esta fase, el proyecto de Data Warehouse debe tener asignado el liderazgo adecuado, así como los recursos humanos, recursos tecnológicos y el presupuesto apropiado.

Sin embargo, deben evaluarse otros aspectos, como desarrollar un proyecto en su totalidad o por fases y además, diferenciar el tipo de proyecto a realizar.

Se tienen que considerar los siguientes elementos en la implementación:

- Proyecto total o proyecto en fases

Es más viable el desarrollo de un proyecto en fases que produzcan resultados a corto plazo que el desarrollo de un proyecto que entregue resultados al término de varios años. Por ello, el proyecto debe estar centrado en un área o un proceso.

- Modelo lógico de datos

El modelo lógico de datos debe tener un alcance más alto y cubrir todas las áreas de interés, así como los procesos más estratégicos de cada una de ellas.

Ejemplo: Puede cubrir las áreas de mercadeo, crédito y comercialización y los procesos de segmentación, scoring para retención, scoring para crédito y administración de clientes, productos y canales de ventas.

- Proyecto especializado o proyecto base

Decidir sobre qué tipo de proyecto, es algo complicado. Un proyecto especializado soporta directamente un proceso específico, por ejemplo: retención de clientes.

Un proyecto base entrega capacidad genérica de análisis a todos los usuarios que tengan acceso al Data Warehouse, pero no tiene, entre sus funcionalidades, la solución de un problema específico o el soporte especializado de un proceso específico.

Un proyecto base es más económico y fácil de acabar que uno especializado, más costoso y difícil de terminar.

En cuanto a estrategias para el proceso de implementación, deben definirse las siguientes:

- Identificar el problema en el cual el uso estratégico de la información detallada, permita conseguir una solución para generar una ventaja competitiva o un ahorro de costos.

Ejemplo: Un problema puede ser la ausencia de un modelo para estudios de retención de clientes.

- Definir el modelo lógico de datos a implementar para resolver el problema planteado.

Ejemplo: Se puede dar un modelo lógico cuando se presenta al usuario la información en términos de dimensiones (clientes, productos, canales de ventas, promociones, adquirientes, etc.) básicas del modelo de datos y hechos(Facts) que se registrarán para estas dimensiones (medidas de ventas, de costos, de producción, de facturación, de cartera, de calidad, de servicio, etc.)

- Reunir los datos para poblar ese modelo lógico de datos.
- Tomar iniciativas de complementación de información para asegurar la calidad de los datos requeridos para poblar el modelo de datos.

Estas definiciones deben estar acompañadas de un servidor apropiado para el Data Warehouse, así como elementos de comunicaciones, nodos cliente, el manejador de la base de datos del Data Warehouse y demás hardware y software requeridos para la implementación del proyecto.

Y deben plantearse las siguientes estrategias en la implementación:

- Definir el mejor diseño físico para el modelo de datos. El diseño físico debe estar orientado a generar buen rendimiento en el procesamiento de consultas, a diferencia del modelo lógico que está orientado al usuario y a la facilidad de consulta.
- Definir los procesos de extracción, filtro, transformación de información y carga de datos que se deben implementar para poblar ese modelo de datos.
- Definir los procesos de administración de la información que permanece en el Data Warehouse.
- Definir las formas de consultas a la información del Data Warehouse que se le proporcionará al usuario. Para esto, debe considerarse la necesidad de resolver un problema y la potencia de consulta.
- Completar el modelo de consulta base, relativo al área seleccionada.
- Implementar los procesos estratégicos del área de trabajo, es decir, implementar herramientas especializadas de scoring, herramientas especializadas para inducción de conocimiento (Data Mining), etc.
- Completar las áreas de interés, en forma similar a lo descrito anteriormente.

Evaluación de rendimiento de la inversión.

Cuando se evalúan los costos, el usuario del Data Warehouse puede no tener el contenido de los costos en mente, pero las preguntas mínimas que puede comenzar a hacerse son las siguientes:

- ¿Qué clases de costos excedieron el presupuesto en más del 10% en cada uno de los 12 meses pasados?
- ¿Se aumentaron los presupuestos en más de 5% para cualquier área dentro de los últimos 18 meses?

- ¿Cómo especificar las clases de gasto entre diferentes departamentos? ¿Entre divisiones? ¿A través de las regiones geográficas?
- ¿Cómo tener márgenes de operación sobre los dos últimos años en cada área de negocio? Donde han disminuido los márgenes, ¿se han incrementado los costos?

Con frecuencia, los aspectos realmente importantes identificados por una administración mayor, tienen un valor agregado, en el que ellos saben si tuvieron la información que estaban buscando, lo que significaría una mejora de (por ejemplo) las ventas en 0.5% a 1% - que, si su operación estuvo por los billones de dólares en un año, puede resultar en cientos de millones de dólares. En algunos casos, el costo del depósito inicial se ha recobrado en un periodo de 6 a 8 meses.

Al hacerse preguntas de este tipo, los usuarios comienzan a identificar las áreas en las que los costos han aumentado o disminuido significativamente y pueden evaluar cada una de estas áreas con más detalle.

Costos y beneficios

Se han identificado diversos costos y beneficios en la elaboración de un proyecto de construcción de un Data Warehouse, tales como:

- Costos
 - *Costos preliminares*
 - Planificación
 - Diseño
 - Modelado/Ingeniería de Información
 - *Costos iniciales*
 - Plataforma de hardware
 - Software de base de datos
- Herramientas de extracción, transformación y carga, de limpieza de datos y de front end
 - *Costos en procesamiento*
 - Mantenimiento de datos
 - Desarrollo de aplicaciones
 - Capacitación y soporte
- Beneficios
 - *Beneficios Tácticos*
 - Impresión y emisión de reporte reducido
 - Demanda reducida para consultas de clientes
- Entrega más rápida de información a los usuarios
 - *Beneficios Estratégicos (Potencialidad)*
 - Aplicaciones y herramientas de acceso para los usuarios finales
 - Decisiones con mayor información
 - Toma de decisiones más rápida
 - Capacidad de soporte a la información organizacional

Beneficios a obtener

- Para la empresa

El Data Warehouse hace lo posible por aprovechar el valor potencial enorme de los recursos de información de la empresa y volver ese valor potencial en valor verdadero.

o Para los usuarios

El Data Warehouse extiende el alcance de la información para que puedan acceder directamente en línea, lo que a la vez contribuye en su capacidad para operar con mayor efectividad las tareas, rutinarias o no.

Los usuarios del Data Warehouse pueden acceder a una riqueza de información multidimensional, presentado coherentemente como una fuente única, confiable y disponible a ellos por medio de sus terminales de trabajo.

Los usuarios pueden usar sus herramientas familiares, hojas de cálculo, procesadores de textos y software de análisis de datos y análisis estadístico para manipular y evaluar la información obtenida desde el Data Warehouse.

o Para la organización de tecnologías de información

El Data Warehouse enriquece las capacidades del usuario autosuficiente y hace lo factible para ofrecer nuevos servicios a los usuarios, sin interferir con las aplicaciones cotidianas de producción.

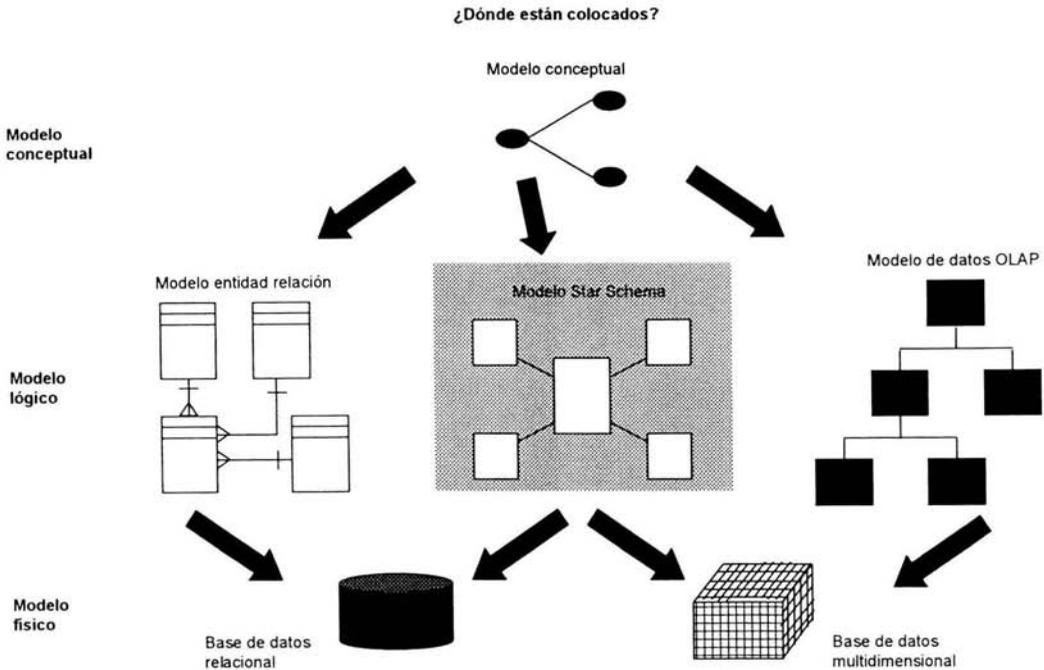
La pugna constante por resolver las necesidades de usuarios que piden acceso a los datos operacionales, finaliza con la implementación de un Data Warehouse. La mayoría de los usuarios no necesita acceder más a los datos actuales, porque ellos tienen información más útil disponible desde el Data Warehouse.

Un Data Warehouse aumenta el valor de las inversiones en tecnologías de información, en aplicaciones y bases de datos operacionales. Como estas bases de datos alimentan información, al evolucionar el Data Warehouse, llegan a ser imprescindibles no solamente para las operaciones diarias, sino además como la fuente de información del negocio de amplio rango.

ANEXO: Modelado de datos en el Data Warehouse

Introducción al modelado de datos

Tipos de modelos de datos



Modelos de datos OLTP

Orientados hacia la operación del negocio
 Afinado para el procesamiento transaccional
 Datos normalizados
 Integridad referencial activa

Modelos de datos de Data Warehouse

Orientados hacia el análisis del negocio
 Afinado para el acceso a los datos
 Datos desnormalizados
 La integridad referencial es considerada como un artefacto

El proceso de modelado de datos del Data Warehouse

- Selecciona y analiza el área del tema (desde el punto de vista de requerimientos de negocio)
- Desarrolla el modelo lógico de datos.
- Mapea las entidades, atributos y relaciones a partir del OLTP o utiliza modelos de datos genéricos.
- Hace reingeniería (si es posible) de los sistemas fuente.
- Aplica el criterio de diseño de Data Warehouse.

Resultado: Modelo Lógico de Datos

Modelado dimensional

Diseño principal,

No impresiona a los usuarios finales con la complejidad.

Organización de la base de datos

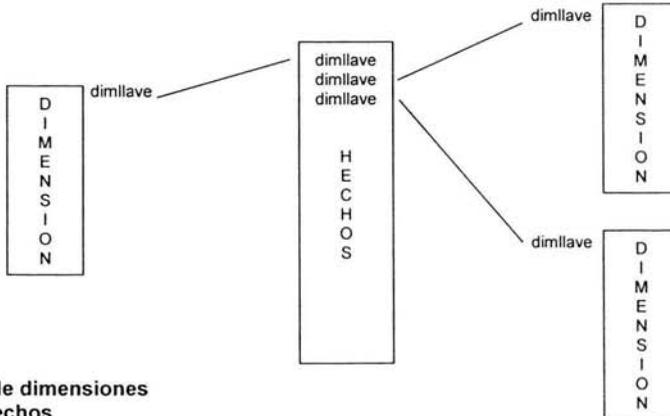
- Debe parecerse al negocio.
- Reconocida por los usuarios finales.
- Accesible para los usuarios finales.
- Relevante para los usuarios finales.

Se buscan algunas buenas tablas ...

Star Schema y Snowflakes: Los modelos dimensionales

Star Schema

- Diferente enfoque para examinar los datos.
- Tomar en consideración las necesidades del usuario del negocio:
 - Generalmente no desea saber sobre base de datos.
 - Generalmente no desea ir a usted para el consejo sobre cómo conseguir datos.
 - Los datos prueban cómo piensa el usuario del negocio.



n número de dimensiones
Tabla de hechos

Dimensiones

- Las Dimensiones son típicamente:
 - Tiempo
 - Mercado/Geografía
 - Cliente/Demografía
 - Producto
 - Estructura de la organización
- Las Dimensiones son cómo el negocio ve los datos:
 - En sus términos
 - Contienen descripciones así como códigos
 - Manejan pocos registros (generalmente < 1 millón)
- Las Dimensiones son tablas pequeñas:
 - Definidas en terminología del usuario del negocio
 - Registros amplios (en cantidad de columnas) de texto altamente descriptivo
 - Usualmente menos de 1 millón de registros.
- Tablas de Dimensión:
 - Representan típicamente el 90% de los elementos de datos
 - Ocurre comúnmente en constraints y en cláusulas Group by
 - Muy indexadas

Ejemplos de Dimensiones

<p>Dimension Tiempo</p> <p>Llave Tiempo</p> <p>Año Semestre Trimestre Mes Semana Dia</p>	<p>Dimension Almacen</p> <p>Llave Almacen</p> <p>Numero almacen Numero territorio Desc territorio Numero region Desc region Numero division Desc division</p>	<p>Dimension Producto</p> <p>Llave Producto</p> <p>Codigo producto Desc producto Codigo categoria Desc categoria Codigo subcategoria Desc subcategoria Codigo color Desc color Codigo tamaño Desc tamaño</p>	<p>Dimension Orden</p> <p>Llave Orden</p> <p>Numero orden Numero linea orden Id persona ventas Nombre persona ventas Tipo pago Desc pago</p>
--	---	--	--

MiniDimensiones

Estas deben ser utilizadas para romper una dimensión grande por razones de rendimiento. Por ejemplo:

<p>Llave Cliente</p> <p>Nombre Apellido paterno Apellido materno Direccion Ciudad Estado Codigo postal Edad Ingreso Sexo Estado civil Conducta de compra Otros demograficos</p>
--

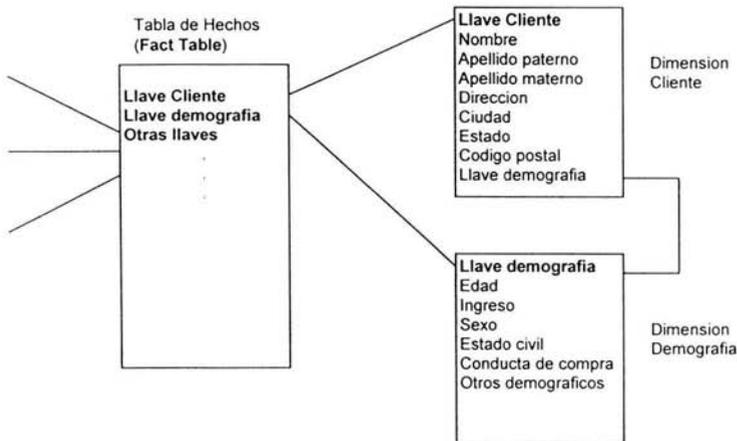
Esta podría ser bastante grande en términos del tamaño de la tabla

Se podría romper ésta, de entrada, en 2 MiniDimensiones

Llave Cliente
Nombre
Apellido paterno
Apellido materno
Direccion
Ciudad
Estado
Codigo postal
Llave demografia

Llave demografia
Edad
Ingreso
Sexo
Estado civil
Conducta de compra
Otros demograficos

y podría quedar como sigue



Dimensiones típicas

- Periodos de tiempo
- Región geográfica (mercados, ciudades)
- Productos (código de producto, números de parte)
- Promociones (descuentos)
- Clientes (número de cuenta)
- Ventas (representante, comprador)
- Nivel de servicio, procedimiento, operación
- Tipo de sitio, servicio, clasificación, asiento
- Vendedor, distribuidor, almacén

Hechos (The Facts)

- Generalmente datos numéricos sin procesar, medidas relevantes solamente.
- Renglones "delgados", algunas columnas en la mayoría.
- De 10 millones hasta mil millones de registros.
- Los artículos son típicamente aditivos.
- Se accesan vía dimensiones.

Definición de Fact Table

- Llave primaria compuesta.
- Las columnas con frecuencia producen funciones
 - sum
 - max
 - min
 - avg
 - rank (si está disponible)
 - otras (desviación estándar, media, ...)

Las columnas aditivas para la Fact Table consisten de:

- Tipos de datos numéricos tales como:
 - Unidades vendidas
 - Dólares vendidos vs. dólares por unidad
 - Montos reclamados
 - Dólares rebajados
 - Ganancia antes de impuesto
 - Cargos por servicio
 - Número de llamadas

Algunas Fact Tables típicas

- Ventas y compras
- Ventas diarias, semanales, mensuales, trimestrales
- Políticas de venta, reclamaciones de venta
- Ordenes, envíos
- Pronósticos de presupuesto, presupuesto real
- Demográficas
- Promociones, publicidad

Tabla de Hechos (Fact Table)	Hechos Ventas
	Llave Almacen Llave Tiempo Llave Producto Llave Orden Unidades vendidas Costo de unidad Costo de venta Monto de impuesto Ventas totales

ESTA TESIS NO SALE
DE LA BIBLIOTECA

Factless Fact Table

Se puede crear un Data Warehouse para proporcionar información sin hechos medidos

- ¿Cuáles son las instalaciones más utilizadas?
- ¿Qué instalaciones de qué departamentos son utilizadas por más estudiantes de otros departamentos?
- ¿Qué cursos son más solicitados?
- ¿Qué cursos se saturan en un cierto plazo?
- ¿Qué estudiantes asistieron a qué cursos fuera de su especialidad?



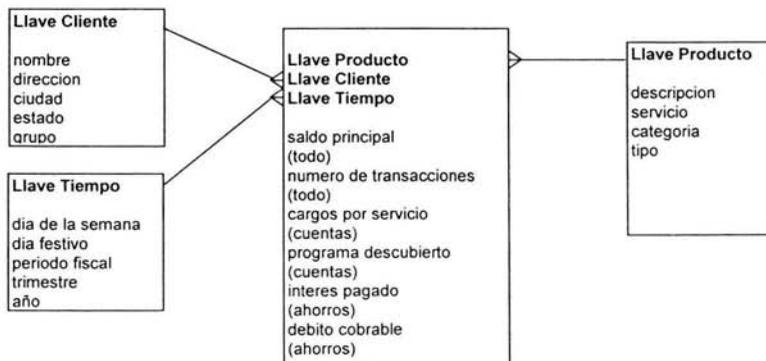
- La existencia se coloca siempre en 1
- Esto hace más fácil el SQL

Star Schema

La estrella ubicua

- Una estrella para cada industria
 - Telecomunicaciones
 - Transportación
 - Seguros
 - Salud
 - Fabricantes
 - Banca
 - Gobierno

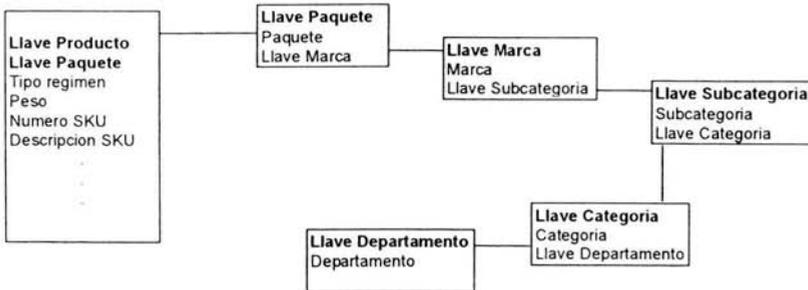
Ejemplo: **Banco,**



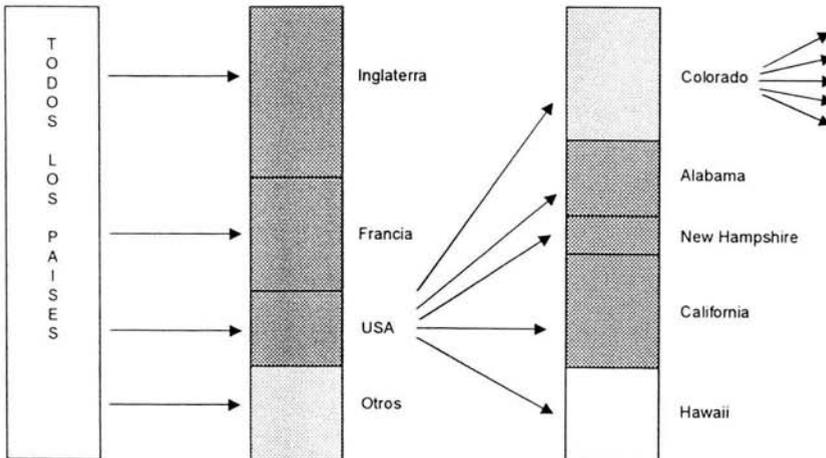
Esquema Snowflake

- Similar al Estrella, excepto porque:
 - Tablas de dimensión más normalizadas
 - Hace que los modeladores de datos se sientan mejor
 - No es tan intuitivo como la Estrella para el usuario de negocio

Dimensión Producto



Esquema Snowflake



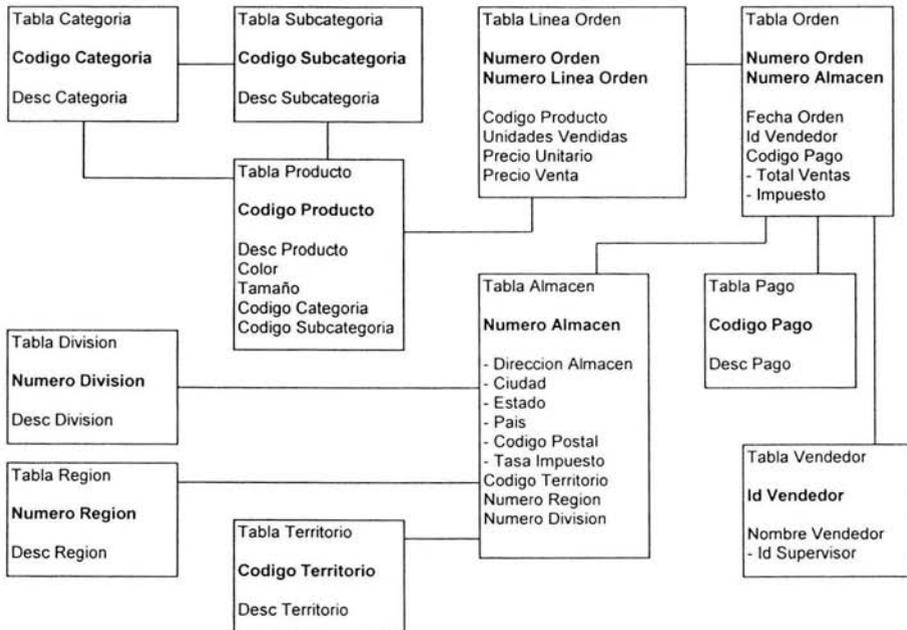
Parece un Drill Down (exploración a detalle)

El Proceso del modelo dimensional

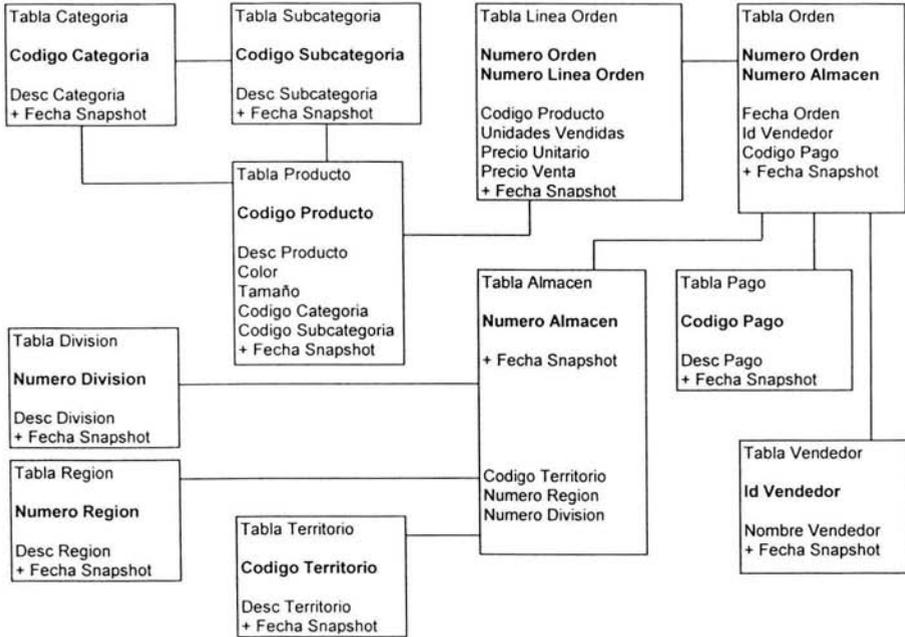
o cómo crear un modelo de datos para el Data Warehouse Star/Snowflake utilizando los datos del modelo OLTP como entrada?

Traducción del modelo de datos OLTP

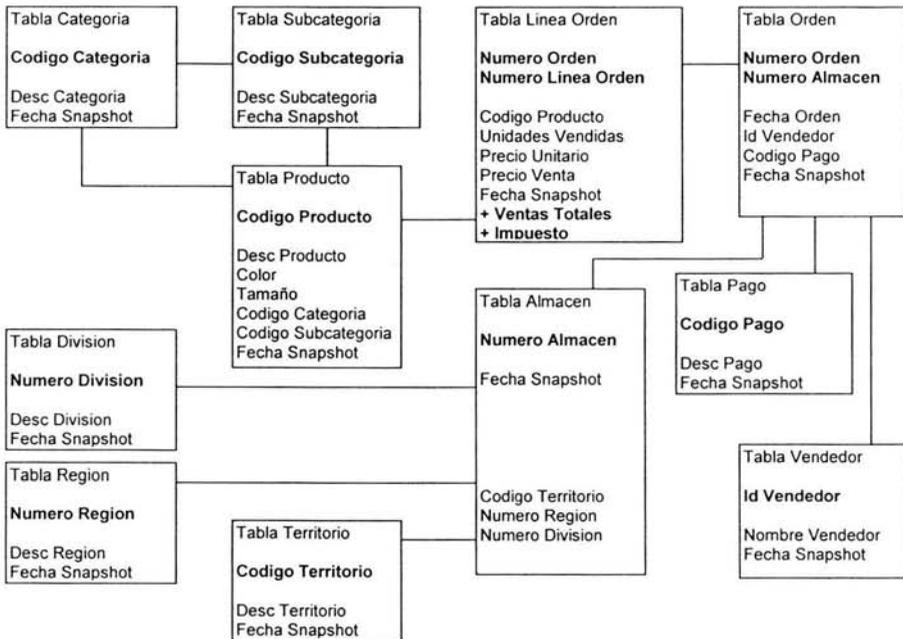
- Primero: Remover datos operacionales innecesarios
 - Datos que son utilizados para transacciones discretas:
 - Banderas y switches (interruptores)
 - Instrucciones (qué hacer, cómo enviar)
 - Crear/Actualizar/Borrar usuarios y fechas
 - Revisar pistas de información



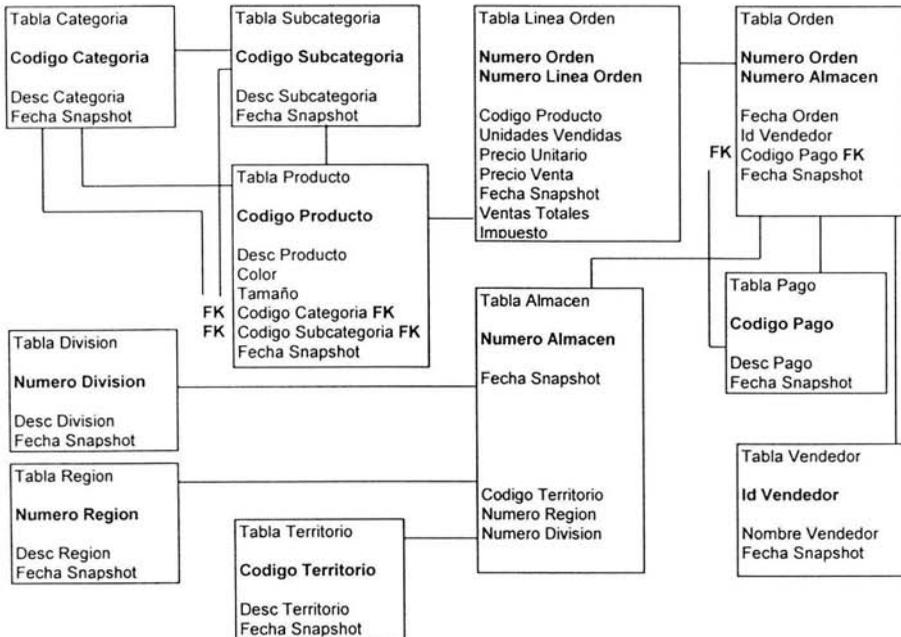
- Segundo: Agregar la dimensión tiempo (después de todo es historia)
 - Cada pieza de datos es asignada en un momento en el tiempo
 - Lapso de tiempo o fechas de snapshot (fotografía instantánea)



- Tercero: Definir datos derivados
 - Cualquier cálculo que sea necesario hacerse
 - ¿Se almacenan promedios o sumas?
 - Asegurarse que los datos pueden ser comparados
 - Longitud promedio del último viaje



- Cuarto: Crear relaciones de objetos
 - Capturar las llaves foráneas para definir las reglas de negocio en uso en aquella época
 - Incluir solamente aquella información de la relación que es útil y obvia



- Quinto: Fusionar datos similares
 - Fusionar datos comunes de las fuentes del modelo de datos
 - Combinar dentro de una vista común

Problema: **Calidad de datos**

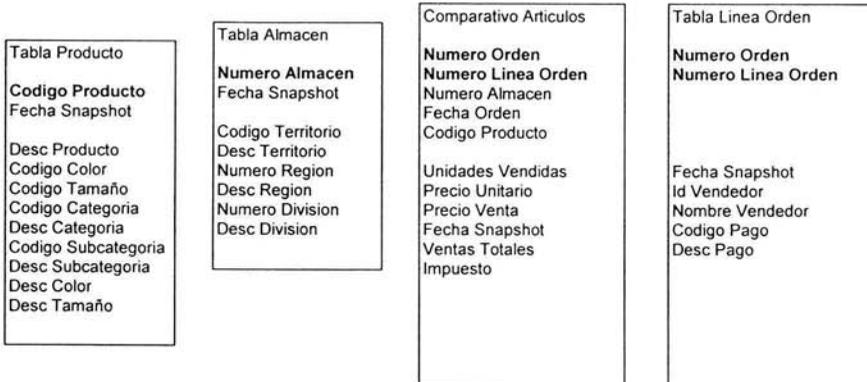
- Este es el paso más difícil y tomará la mayor parte del tiempo para completarla

Tabla Producto
Código Producto
Fecha Snapshot
Desc Producto
Código Color
Código Tamaño
Código Categoría
Desc Categoría
Código Subcategoría
Desc Subcategoría
Desc Color
Desc Tamaño

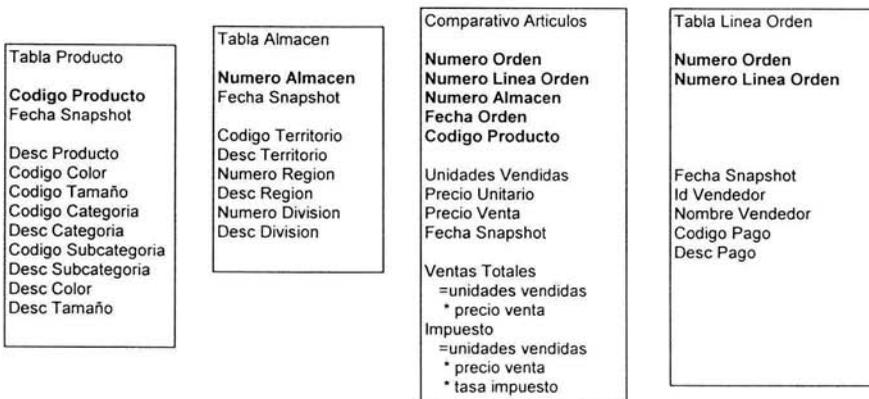
Tabla Almacen
Número Almacen
Fecha Snapshot
Código Territorio
Desc Territorio
Número Región
Desc Región
Número División
Desc División

Tabla Línea Orden
Número Orden
Número Línea Orden
Código Producto
Unidades Vendidas
Precio Unitario
Precio Venta
Fecha Snapshot
Ventas Totales
Impuesto
Número Almacen
Fecha Orden
Id Vendedor
Nombre Vendedor
Código Pago
Desc Pago

- Sexto: Segregar los datos
 - Crear los órdenes si se conocen los incrementos fijos
 - Separar fuera del texto descriptivo



- Séptimo: Aplicar el contexto necesario a los datos
 - Aplicar los criterios de contexto básicos
 - Definir la estructura de cada entidad/atributo
 - Definir la codificación/métrica de los datos
 - Nombramiento de las convenciones/mapeo de datos para el OLTP



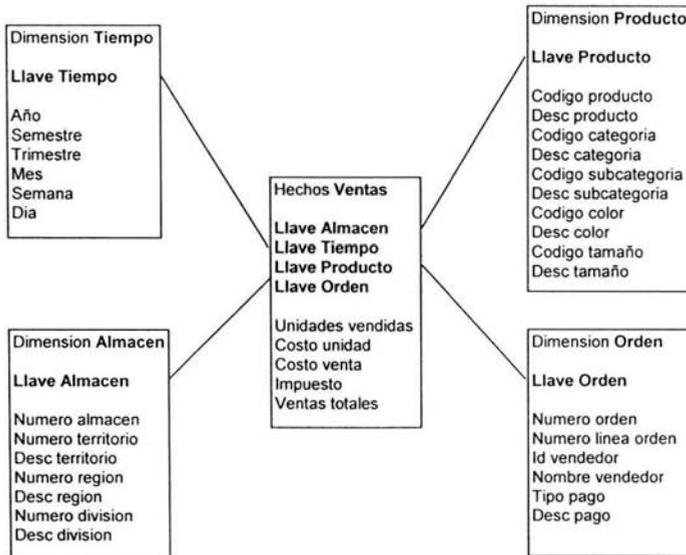
Mapeo lógico a físico

- Determinar llaves
- Crear dimensión de tiempo (si no existe)
- Definir integridad referencial (pk y fk)

Recuerde que la intersección de Dimensiones es la tabla de Hechos (Fact Table)

¿Se terminó con una Estrella (Star) o con un Copo de nieve (Snowflake)?

Traducción del modelo de datos OLTP:

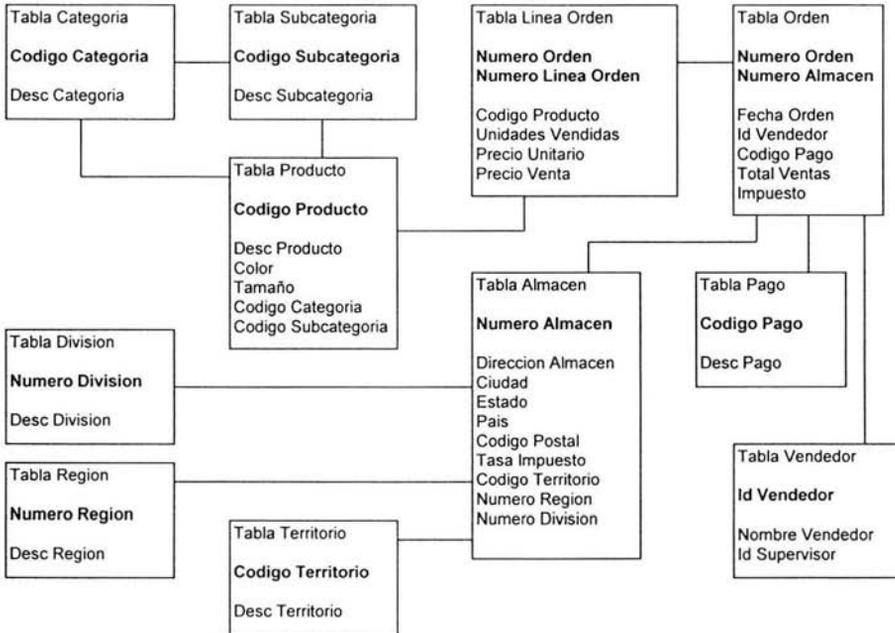


Conclusiones

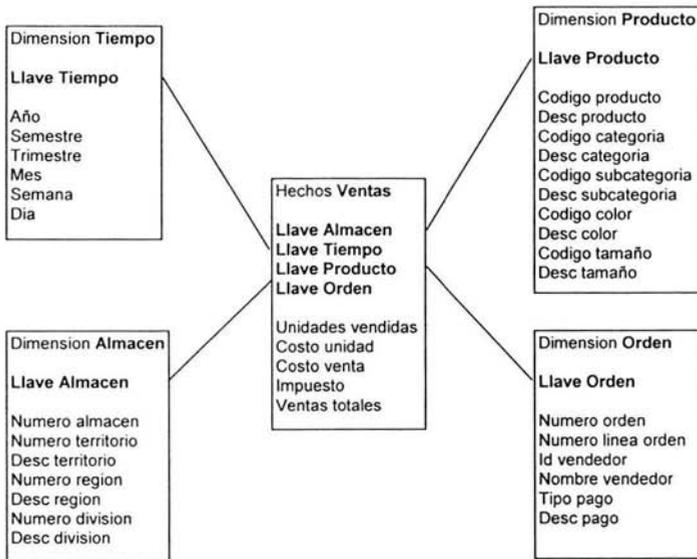
- Limitar el alcance a un tema con el gran proyecto en mente.
- Si es posible, utilizar el modelo OLTP existente más exacto como punto de partida.
- Entender el propósito del esquema estrella (Star Schema)
- Entender dónde el modelo lógico de datos encaja bien dentro del ciclo de desarrollo del Data Warehouse.
- Entender el proceso de traducción.

ESQUEMAS

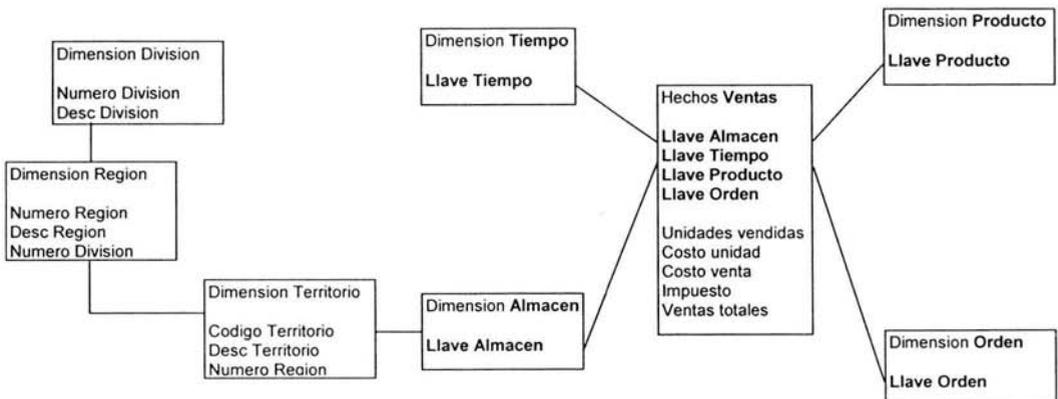
Esquema OLTP



Esquema Estrella (Star Schema)



Esquema Copo de Nieve (Snowflake Schema)



Glosario

Arquitectura de Data Warehouse.- Un conjunto integrado de productos que permiten la extracción y transformación de datos operacionales a ser cargados a una base de datos para hacer análisis y generar reportes por parte de los usuarios finales.

Base de Datos Relacional (RDBMS).- Un tipo de DBMS que almacena datos en forma de tablas relacionadas. Las bases de datos relacionales son poderosas porque requieren pocas suposiciones acerca de cómo están relacionados los datos o cómo pueden ser extraídos de la base de datos. Facilitan el trabajo con registros individuales.

Base de Datos Multidimensional (MDBS y MDBMS).- Una poderosa base de datos que permite a los usuarios analizar grandes cantidades de datos. Una MDBS captura y representa datos como arreglos que pueden ser organizados en múltiples dimensiones.

Cliente/Servidor.- Un enfoque de la tecnología distribuida donde el procesamiento es dividido por función. El servidor ejecuta funciones compartidas, como administración de las comunicaciones, proporciona servicios de base de datos, etc. El cliente ejecuta funciones individuales de usuario, como proporcionar interfaces personalizadas, navegación pantalla a pantalla, ofrece funciones de ayuda, etc.

Constraints.- Son restricciones en alguna columna en la definición de una tabla. Hay varios tipos de restricciones disponibles tanto para las tablas como para las columnas, y se implementan de varias maneras. Algunas restricciones son implementadas como triggers, algunas como índices, y otras son atributos de la definición de una columna.

Una restricción consiste en la definición de una característica adicional que tiene una columna o una combinación de columnas, suelen ser características como valores no nulos (campo requerido), definición de índice sin duplicados, definición de clave principal y definición de clave foránea (clave ajena o externa, campo que sirve para relacionar dos tablas entre sí).

Dato.- Artículos que representan hechos, texto, gráficas, imágenes, sonido, segmentos de video analógico o digital. El dato es la materia prima de un sistema proporcionado por los productores de datos y es utilizado por los consumidores para crear información. El dato es utilizado para consulta, análisis y reporte.

Dato Agregado (Aggregate data).- Dato que es el resultado de aplicar un proceso a elementos de datos combinados. Datos que son tomados colectivamente o sumariados.

Dato del Negocio.- Información acerca de la gente, lugares, cosas, reglas de negocio y eventos, los cuales son utilizados para operar el negocio. Este no es la metadata.

Dimensiones.- Las dimensiones son los conductores de los Data Marts. Las dimensiones son las plataformas para revisar los valores de constraints aceptables y lanzar estos constraints. Las dimensiones son la fuente de los encabezados en los reportes definitivos del usuario, éstos llevan el vocabulario de la empresa a los usuarios.

Drill (Down/Up).- Un método de exploración detallado de datos que son utilizados en la creación de un nivel sumariado de los mismos. Los niveles de exploración (drill) dependen de la granularidad de los datos en el Data Warehouse. Drill Down es el más antiguo y más venerable tipo de exploración en un Data Warehouse. Si Drill Down es agregar grupos de columnas de las tablas de dimensión, entonces Drill Up es sustraer grupos de columnas.

Data Mart.- Genéricamente implementados en departamentos o áreas donde sólo un subconjunto específico de información funcional es necesario para apoyar a los individuos o grupos de trabajo.

Data Warehouse.- Una implementación de una base de datos utilizada para almacenar datos fuente compartidos de bases de datos operacionales. Es típicamente una base de datos de sujetos que permite a los usuarios explotar la información, dentro de un gran almacén de datos operacionales de la compañía, para seguir la pista, responder a tendencias de negocio y facilitar los esfuerzos de pronóstico y planeación.

DSS (Decision Support Systems, Sistemas de Soporte a las Decisiones).- El término DSS es un término amplio que en general cubre a todos aquellos sistemas de cómputo que ayudan a tomar decisiones.

Los sistemas de soporte a las decisiones (DSS) son sistemas de cómputo que combinan modelos e información para tratar de resolver problemas no estructurados con una amplia intervención de usuario. Los beneficios de la integración de un sistema de soporte a las decisiones eran básicamente una alta calidad en la toma de decisiones, mayor comunicación, reducción de costos, mayor productividad, ahorro de tiempo y la satisfacción del cliente y los empleados. El tomar una decisión errónea en una organización se traduce en dinero, esfuerzo y tiempo tres factores importantes que podrían marcar la supervivencia de una organización, es por eso que los sistemas de soporte a las decisiones son una herramienta necesaria en las organizaciones pues ayudan a tomar mejores decisiones, en un menor tiempo y con una mayor satisfacción.

EIS (Executive Information System, Sistemas de Información Ejecutiva).- Herramientas programadas para proporcionar reportes fijos o documentos a los altos ejecutivos. Ofrecen fuertes capacidades de reporte y exploración. Estas herramientas permiten consultas ad-hoc comparadas con una base de datos multidimensional, y la mayoría ofrecen aplicaciones analíticas a lo largo de líneas funcionales tales como ventas y análisis financiero.

Son sistemas de gestión empresarial que apoyan de manera interactiva a las empresas en la toma de decisiones claves. Gracias a ellos se pueden construir indicadores de una empresa, evaluar el desempeño del negocio en el tiempo y compararse con los líderes de su sector.

ETL (Extraction, Transformation and Load).- Herramienta que automatiza el proceso de extracción, transformación y carga de datos, prácticamente de cualquier fuente de datos a cualquier destino de datos.

Fact Table (Tabla de Hechos).- Las Fact Tables son las tablas grandes e importantes "del centro" de un esquema dimensional. Siempre tienen una llave compuesta, en la cual cada componente de la llave relaciona a una única tabla de dimensión.

Factless Fact Table.- Se habla mucho sobre los valores numéricos aditivos en las Fact Tables, que puede resultar sorprendente que dos tipos de Fact Tables muy útiles no tengan ningunos hechos en absoluto. Ellas pueden consistir en nada más que llaves y por esto son llamadas Factless Fact Tables (Tablas de Hechos sin Hechos)

Gateway (Pasaporte de datos).- Software que permite a aplicaciones basadas en SQL acceder fuentes de datos relacionales y no relacionales.

Mapeo de Datos.- El proceso de asignar un elemento de dato fuente a un elemento de dato destino.

Metadata.- La metadata, es decir, datos acerca de los datos, describe los contenidos del Data Warehouse. La metadata consiste de definiciones de los elementos de datos en el Data Warehouse. La metadata es almacenada dentro de un diccionario de datos y un repositorio. Esto aísla al Data Warehouse de los cambios en el esquema operacional. El componente final del Data Warehouse es el de la Metadata.

Minería de Datos (Data Mining).- Es el proceso de examinar cuidadosamente grandes cantidades de datos para producir relaciones del contenido de los datos. Una técnica, que utiliza herramientas de software adaptadas al usuario quien típicamente no sabe exactamente qué es lo que está buscando, pero se está guiando por patrones o tendencias particulares.

Modelo de datos (Data Model).- Un mapa lógico que representa las propiedades inherentes de los datos, independiente del software, hardware o consideraciones de rendimiento de la máquina. El modelo muestra los elementos de datos agrupados en registros, así como la asociación alrededor de esos registros.

Modelado de datos (Data Modeling).- Un método utilizado para definir y analizar los requerimientos de datos necesarios para soportar las funciones de negocio de una empresa. Estos requerimientos de datos son registrados como un modelo de datos conceptual con definiciones de datos asociados. El modelado de datos define las relaciones entre elementos y estructuras de datos.

Motor de Base de Datos.- Bases de Datos Relacionales (RDBMS) y Bases de Datos Multidimensionales (MDBMS). Los motores de bases de datos para el Data Warehouse requieren fuertes capacidades de consulta, mecanismos de carga rápida y grandes requerimientos de almacenamiento.

ODBC.- Open Database Connectivity (Conectividad de Base de Datos Abierta). Un estándar para acceso a bases de datos co-optado por Microsoft del consorcio SQL Access Group.

OLAP.- On-Line Analytical Processing (Procesamiento Analítico en Línea). Se utiliza comúnmente con las bases de datos multidimensionales.

OLTP.- On-Line Transaction Processing (Procesamiento Transaccional en Línea). Describe los requerimientos para un sistema que es utilizado en un ambiente operacional. Generalmente se refiere a las bases de datos relacionales.

Particionamiento.- El proceso de particionar lógica y/o físicamente los datos en segmentos, los cuales son más fáciles de mantener o acceder. Los actuales sistemas RDBMS proporcionan este tipo de funcionalidad de distribución. El particionamiento de datos ayuda en el rendimiento y la utilidad del procesamiento.

Scoring.- Calificación que se le da a un grupo de clientes/productos que mide la propensión a compras, ventas, retiro, llegadas, etc.

Bibliografía

- BARQUIN, Ramon C. Y EDELSTEIN, Herbert A.: *Building, using and managing the Data Warehouse*, Prentice Hall, 1997.
- BERSON, Alex y SMITH, Stephen J.: *Data Warehousing, Data Mining and OLAP*, McGraw Hill, 1997.
- DEVLIN, Barry: *Data Warehouse: from architecture to implementation*, Addison-Wesley, 1997.
- INMON, William H.: *Building the Data Warehouse*, John Wiley, 1996.
- KIMBALL, Ralph: *The Data Warehouse Toolkit*, John Wiley, 1996.
- KIMBALL, Ralph: *DBMS - Data Warehouse Architect: Is ER Modeling Hazardous to DSS*, October, 1995 (17,18)
- *DBMS - Data Warehouse Architect: The Aggregate Navigator*, November, 1995 (18, 20)
 - *DBMS - Data Warehouse Architect: The Problem with Comparisons*, January, 1996, (16, 18)
 - *DBMS - Data Warehouse Architect: SQL Roadblocks and Pitfalls*, February, 1996, (14, 16)
 - *DBMS - Data Warehouse Architect: Drilling Down, Up, and Across*, March, 1996, (14, 16, 18)
 - *DBMS - Data Warehouse Architect: Slowly Changing Dimensions*, April, 1996, (18, 20)
 - *DBMS - Data Warehouse Architect: Monster Dimensions*, May, 1996, (14, 16)
 - *DBMS - Data Warehouse Architect: Mastering Data Extraction*, June, 1996, (14, 16)
 - *DBMS - Data Warehouse Architect: Automating Data Extraction*, July, 1996, (16, 18)
 - *DBMS - Data Warehouse Architect: Aggregate Navigation with (Almost) No Metadata*, August, 1996, (S15, S18, S20, S22)
 - *DBMS - Data Warehouse Architect: Dangerous Preconceptions*, August, 1996, (14, 16)
 - *DBMS - Data Warehouse Architect: Factless Fact Tables*, September, 1996 (16, 18)
 - *DBMS - Data Warehouse Architect: Dealing with Dirty Data*, September, 1996, (55, 56, 58, 60)
 - *DBMS - Data Warehouse Architect: Causal (Not Casual) Dimensions*, November, 1996, (16, 18)
 - *DBMS - Data Warehouse Architect: Letting the Users Sleep – Part 1*, December, 1996, (12, 14)
 - *DBMS - Data Warehouse Architect: Letting the Users Sleep – Part 2*, January, 1997, (16, 18, 30)
 - *DBMS - Data Warehouse Architect: Features for Query Tools*, February, 1997, (12, 14)
 - *DBMS - Data Warehouse Architect: Hackers, Crackers, and Spooks*, April, 1997, (14, 16)
 - *DBMS - Data Warehouse Architect: Traveling through Databases*, May, 1997, (16, 18)
 - *DBMS - Data Warehouse Architect: It's Time for Time*, July, 1997, (16, 18, 99)
 - *DBMS - Data Warehouse Architect: Data Warehouse Role Models*, August, 1997, (14)
- The Data Warehouse Institute: Journal of Data Warehousing*, Spring 1999 (Volume 4 Number 1).
- MARTÍNEZ DE SOUSA, José: *Manual de estilo de la lengua española*, 2.ª ed., Ediciones Trea, S.L., 2001.