



grupo evaluación nuevas tecnologías

Herramientas Informáticas para la Gestión de Datos. Introducción Teórica

Madrid, 27 de mayo de 2009

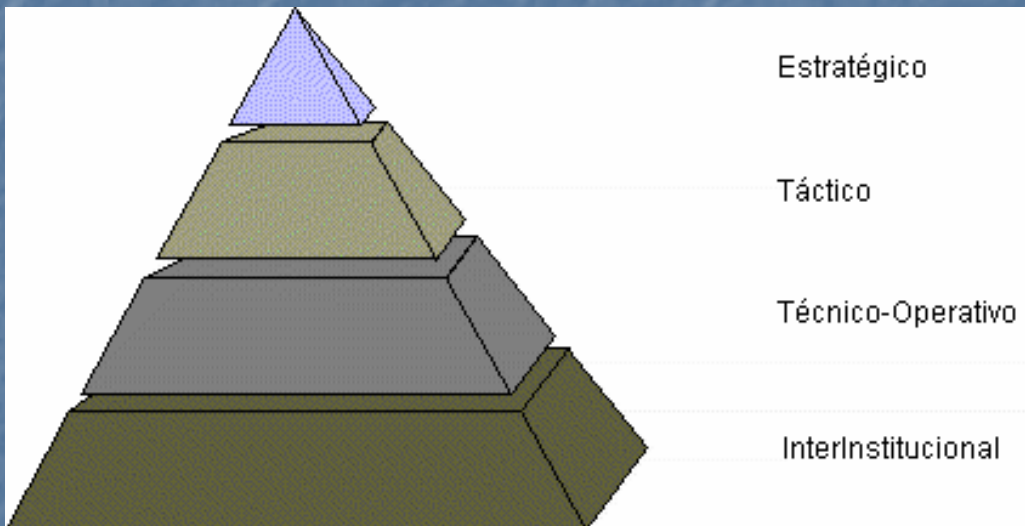
José Antonio Rodríguez Maniega

Ingeniero Informático

Introducción

- Los sistemas de Información para la gestión de datos
- Conceptos básicos
- Tipos de sistemas
- Fases
- Conceptos

Tipos de Sistemas de Información



DW, EIS, Otros

Workflow, Fax, email

HIS, Farmacia

Intranet, extranet,

EIS/DDS

- EIS: Sistema de Información Ejecutiva
- DSS: Sistema de ayuda a la toma de decisión

¿Por qué un sistema específico?

Sistema Transaccional

Datos Operacionales
Orientado a la aplicación
Actual
Detallada
Cambia continuamente

Data Warehouse

Datos del negocio para Información
Orientado al sujeto
Actual + histórico
Detallada + más resumida
Estable

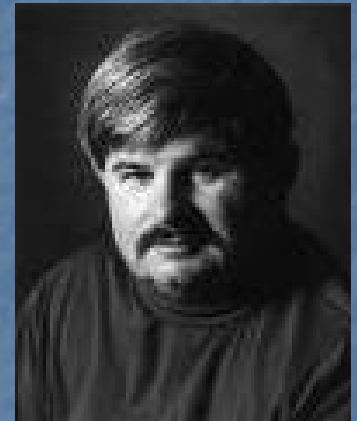
- El ingreso de datos en el data warehouse viene desde el ambiente operacional en casi todos los casos. El data warehouse es siempre un almacén de datos transformados y separados físicamente de la aplicación donde se encontraron los datos en el ambiente operacional

Definición Data Warehouse

- Data Warehouse: Almacén de datos.
 - Ahora se refiere más al sistema de información que almacena datos para ser explotados. El objetivo es el dato, no el proceso.

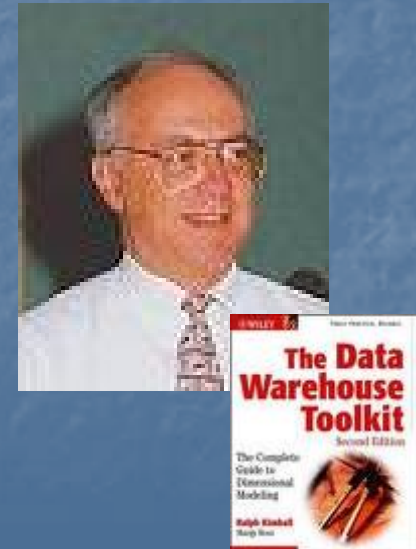
"Padre" y Paradigma Bill Inmon

- **Orientado a temas.**- Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.
- **Variante en el tiempo.**- Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.
- **No volátil.**- La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de *sólo lectura*, y se mantiene para futuras consultas.
- **Integrado.**- La base de datos contiene los datos de todos los sistemas operacionales de la organización, y dichos datos deben ser consistentes.
- **Inmon** defiende una metodología descendente (**top-down**) a la hora de diseñar un almacén de datos,
- RECIENTE- 90's



“Gurú” y Paradigma Kimbal

- Ralph Kimbal
- Ha creado el “modelo dimensional” estándar actual
- Comprensible por encima de todo



José A. R. Maniega
Data Warehouse

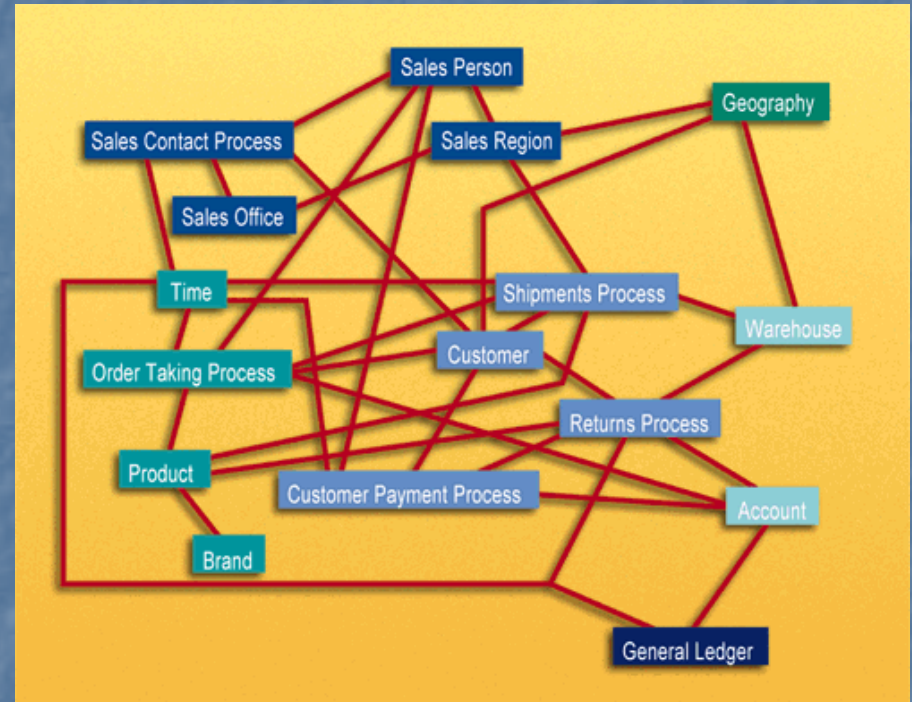
Modelo Inicial

- Comienzo: del papel al ordenador

proveedor	artículo	cantidad	precio	total	Fecha
Bayer	Adiro	30	2	60	12/2/09

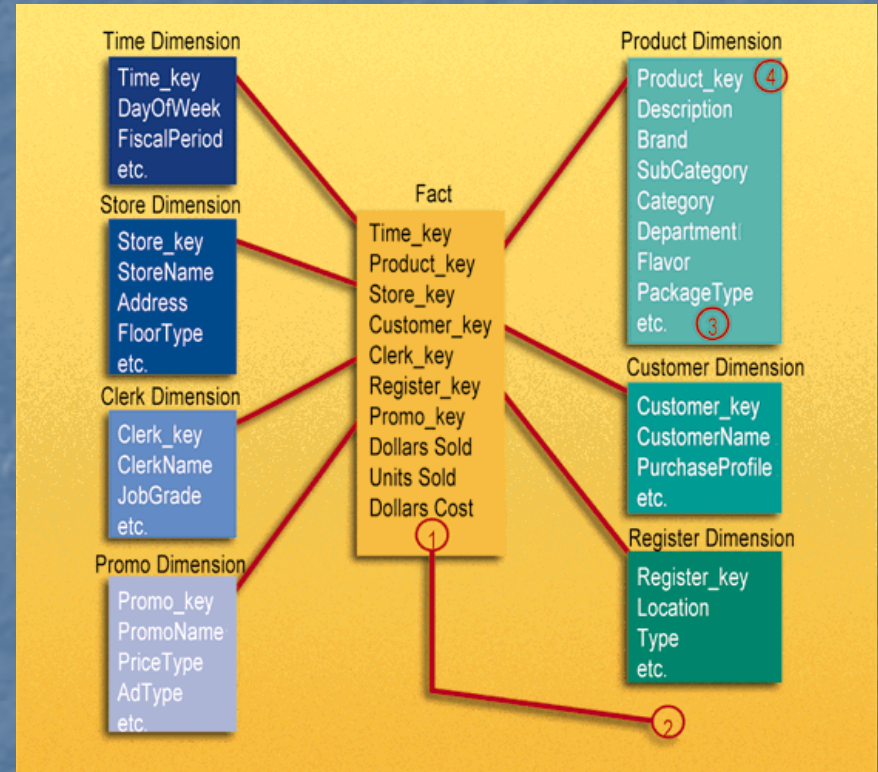
Modelo Entidad Relación

- Modelo Entidad-Relación. Para eliminar inconsistencias por duplicidad. Coste: complejidad en la transacción, bloqueos...



Modelo Dimensional

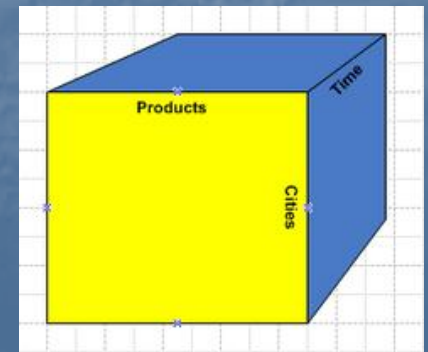
- Técnica de diseño lógico que persigue plantear los datos en una estructura estándar e intuitiva que permita el acceso a los datos a gran velocidad



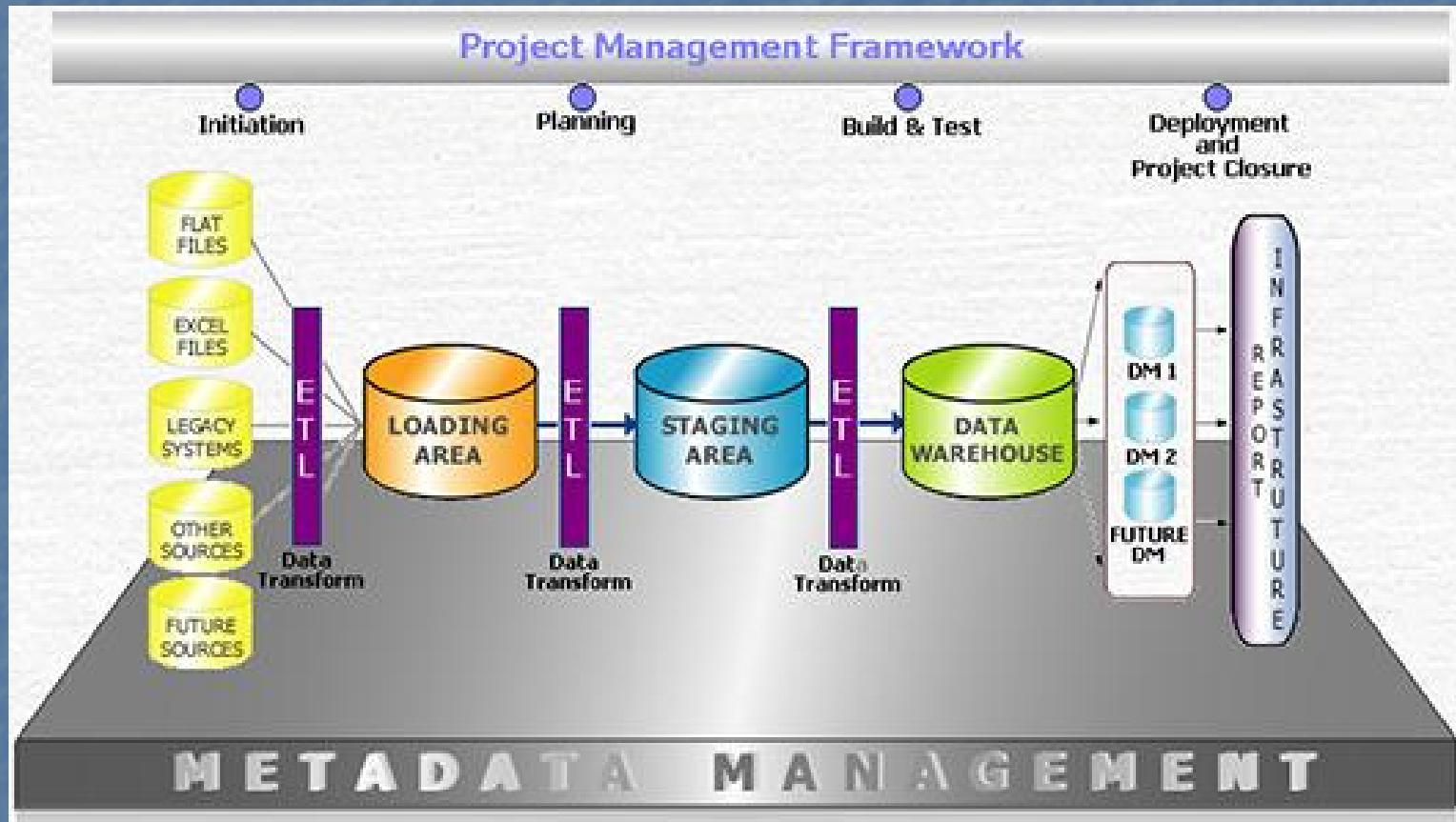
<http://www.dbmsmag.com/9708d15.html>

Cubo OLAP

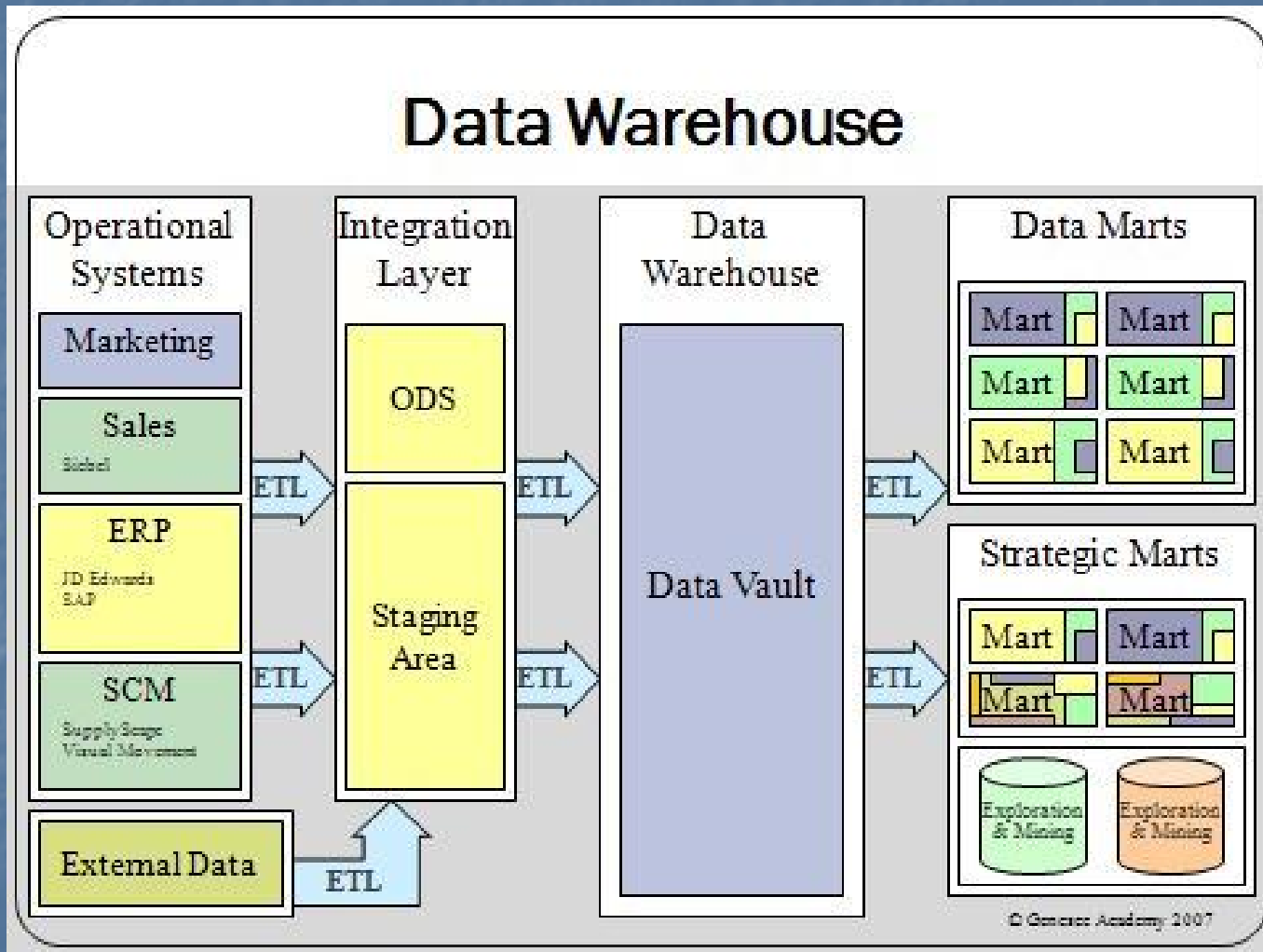
- Nos referimos a cubos OLAP cuando hablamos de bases de datos multidimensionales, en las cuales el almacenamiento físico de los datos se realiza en vectores multidimensionales.
- (Los cubos OLAP se pueden considerar como una ampliación de las dos dimensiones de una hoja de cálculo)
- Por ejemplo, una empresa podría analizar algunos datos financieros por producto, por período de tiempo, por ciudad, por tipo de ingresos y de gastos, y mediante la comparación de los datos reales con un presupuesto. Estos parámetros en función de los cuales se analizan los datos se conocen como dimensiones.
- NO admite cambios en la estructura



Flujo



Mapa Lógico



Minería de Datos/ Data Mining

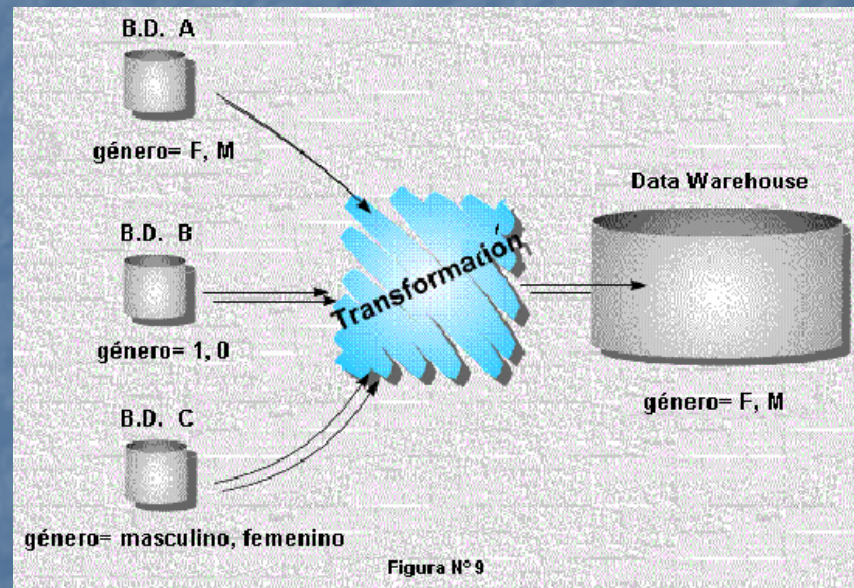
- **Extracción de información previamente oculta**
- Un proyecto de minería de datos tiene varias fases necesarias :
 - **Comprensión** del negocio y del problema que se quiere resolver.
 - Determinación, obtención y limpieza de los datos necesarios.
 - Creación de modelos matemáticos.
 - Validación, comunicación, etc. de los resultados obtenidos.
 - Integración, si procede, de los resultados en un sistema transaccional o similar.
- En realidad es un proceso muy complejo.
- Ejemplo: Comportamiento de diferentes servicios sobre un mismo fármaco

Data Mart

- Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades. (Parte del DW)
- Según Inmon un data mart dependiente es un subconjunto lógico (vista) o un subconjunto físico (extracto) de un almacén de datos más grande, que se ha aislado

ETL

■ Extract, Treat, Load



Uso de los DW

Uso de Base de Datos
Operacionales

Muchos usuarios
concurrentes

Consultas predefinidas y
actualizables

Cantidades pequeñas de
datos detallados

Requerimientos de
respuesta inmediata

Uso de Data Warehouse

Pocos usuarios concurrentes

Consultas complejas, poco
anticipadas.

Cantidades grandes de
datos detallados

Requerimientos
de respuesta no críticos

Farmacia

- ¿Es necesario tener información exhaustiva?
- ¿Hacen falta modelos predictivos ?
- ¿Existen varias fuentes de datos a integrar ?
- Hacen falta decisiones ejecutivas basadas en información

¿Puede disminuir el gasto?

- ¿Hay información “oculta” ?
- Es necesario “prueba”
- Información para la crisis..



José A. R. Maniega
Data Warehouse

Muchas Gracias

**Una colección de conocimientos debe ser una farmacia
donde se encuentra remedio a todos los males.
Voltaire (1694-1778)**

jose_antonio.rodriguez@siemens.com