



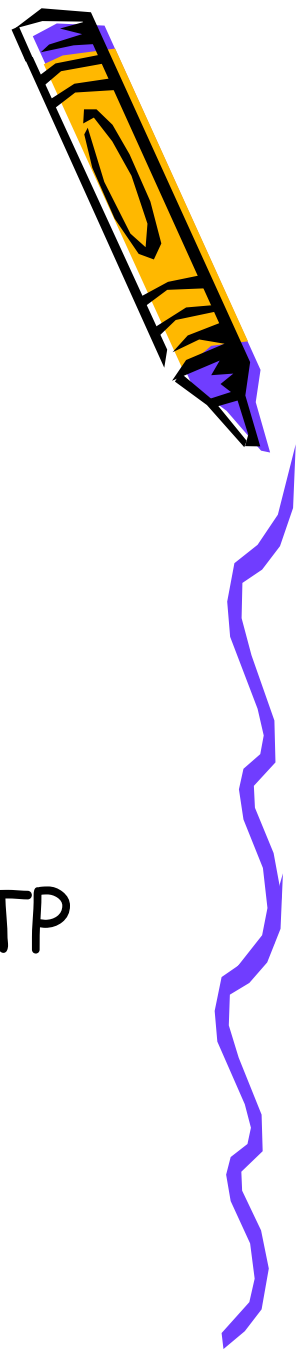
Bodegas de Datos y OLAP

Introducción a la Bodegas de
Datos



Contenido

- SI-Definición y Clasificación
- MIS Vs DSS
- DSS-Definición y Características
- DW-Definición, Elementos, Características, Arquitectura, OLTP Vs DW, Data Marts



SISTEMAS DE INFORMACION-Definición



- Conjunto formal de procesos que operan sobre una colección de datos estructurada según las necesidades de la empresa, ***recopilan, elaboran y distribuyen la información*** necesaria para las operaciones y actividades de dirección y control, para desempeñar las actividades de acuerdo a la estrategia de negocio.



SI-CLASIFICACION



- OLTP: procesan datos (clasificación, cálculo, distribución, resumen y almacenamiento de datos) de las actividades (transacciones) de la empresa.
- **MIS: necesidades de información de los gerentes de una compañía o subunidad.**



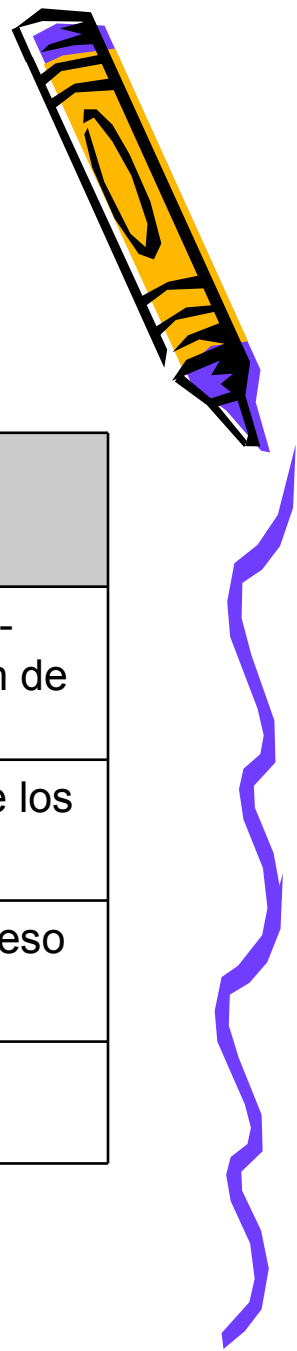
SI-CLASIFICACION



- **DSS:** problemas semiestructurados. Usar la computadora para apoyar "a un *gerente específico* que toma una *decisión específica* para resolver un *problema específico*".
- **SI para oficinas:** elaborar correspondencia, reportes, relaciones y documentos especiales.



MIS Vs DSS



Sistemas de información gerencial (MIS)	Sistemas de soporte a decisiones (DSS)
Se centra en decisiones estructuradas y decisiones rutinarias.	Se centra en las tareas semi-estructuradas que requieren de criterio administrativo.
Se centra en el almacenamiento de información	Se centra en la manipulación de los datos.
Con frecuencia los administradores tienen acceso a los datos.	Los administradores tienen acceso directo a los datos.
Se apoya en expertos en computación.	Se apoya en el juicio del administrador.



DSS-Definición (1)



- C. J. Date, ayudan al análisis de información de negocios. Propósito: Ayudar a la administración a que "Marque tendencias, señale problemas y tome.... decisiones inteligentes".



DSS-Definición (2)



- W. H. Inmon, soportan decisiones de gerencia. Usualmente involucran el análisis de muchas unidades de datos en una forma heurística. Los procesos DSS no involucran la modificación de datos.



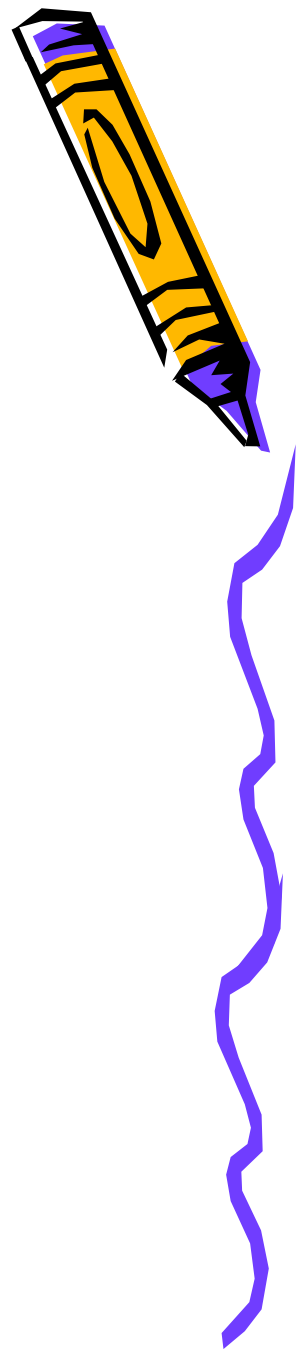
DSS y DW



- El DW es la base del procesamiento DSS.
- En el DW hay una fuente de datos integrados y los datos son asequibles.



DSS- Características



- Interactividad.
- Tipo de decisiones.
- Frecuencia de uso.
- Variedad de usuarios.
- Flexibilidad.
- Desarrollo.
- Interacción ambiental.
- Comunicación íter organizacional.
- Acceso a base de datos.
- Simplicidad.



DW-Definición

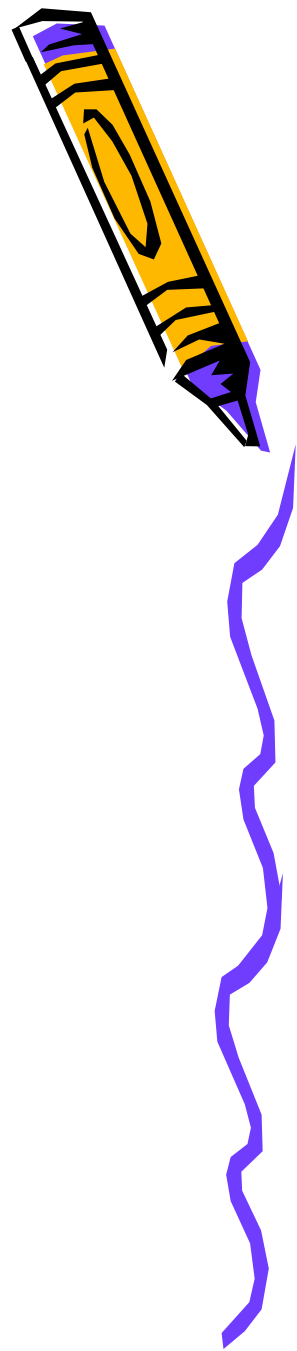


- W. H. Inmon, colección de datos integrados, orientados a temas, que dan soporte a las funcionalidades del DSS, donde cada unidad de dato es relevante en en el tiempo.
- Según R. Kimball, copia de datos transaccionales específicamente estructurados para consultas y análisis. Es la unión de todos los Datamarts.

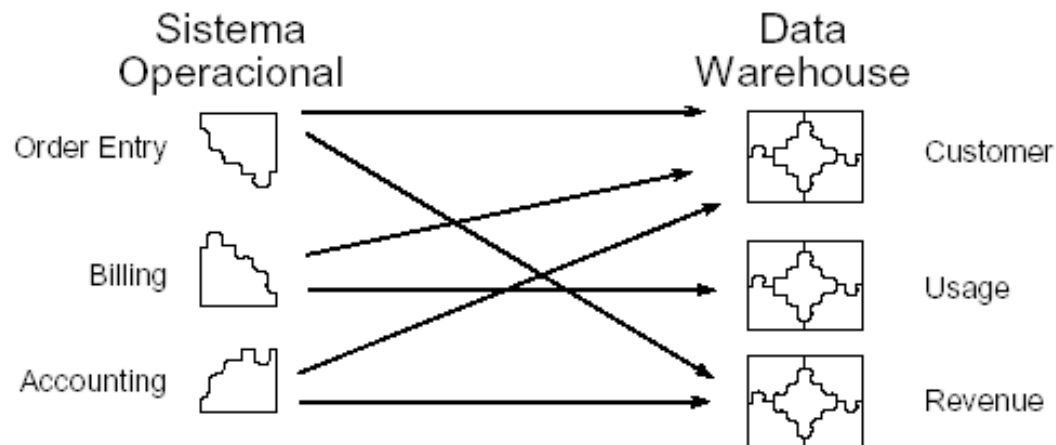


DW-Características

- Orientado a temas
- Integrado
- No volátil
- De tiempo variante



DW-Características (Orientado a Temas)



- Los datos operacionales son organizados por procesos específicos y se mantienen en sistemas separados
- Los datos del DW son organizados por área y poblados desde muchos sistemas operacionales

DW-Características (Integrado)

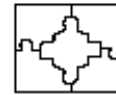
Sistema
operacional



Aplicaciones específicas

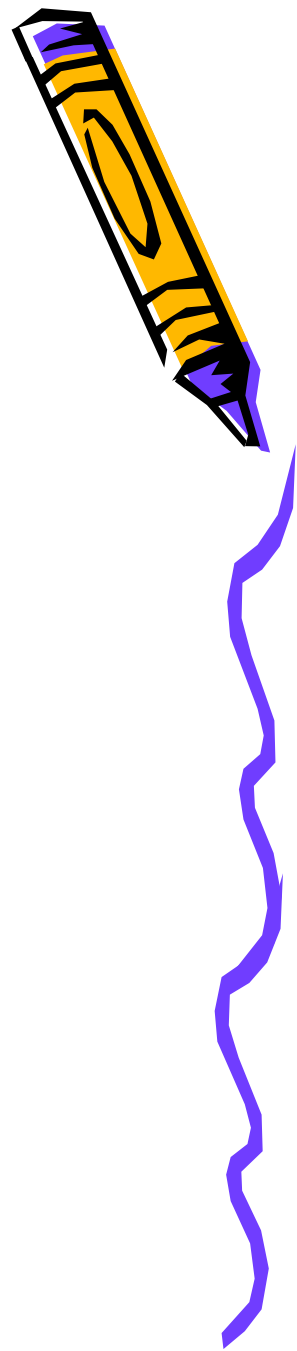
- Las aplicaciones y sus bases de datos son diseñadas y construidas por separado
- Cubre largos períodos de tiempo

Data
Warehouse

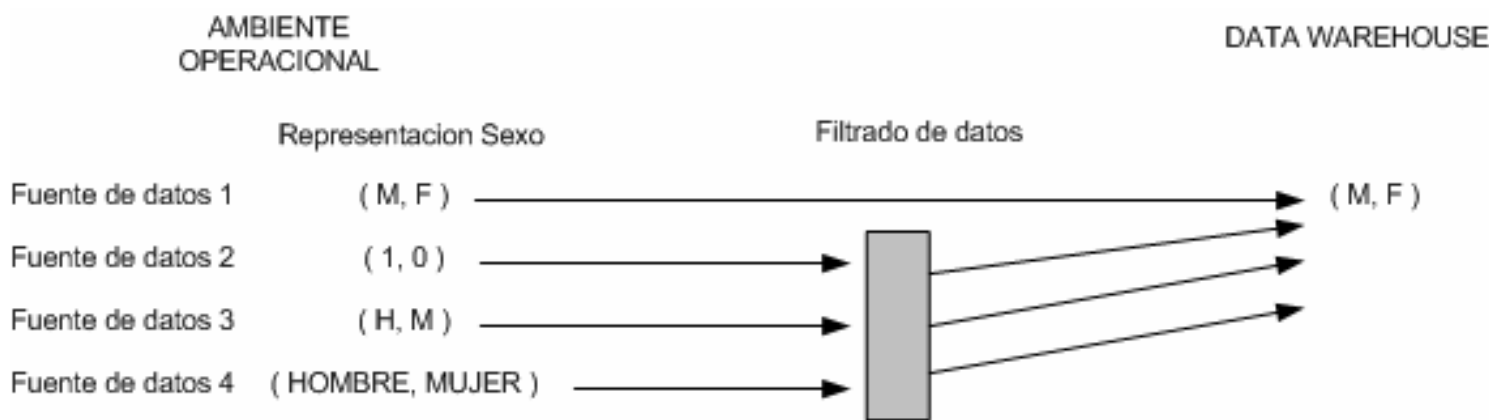


Integradas

- Integradas desde el comienzo
- Diseño realizado una sola vez, implementación iterativa sobre períodos cortos de tiempo



DW-Características (Integrado)

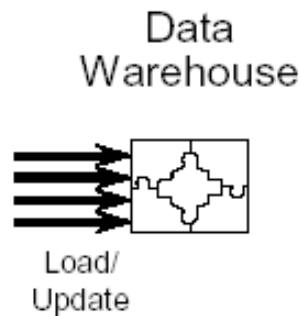


DW-Características (No volátil)



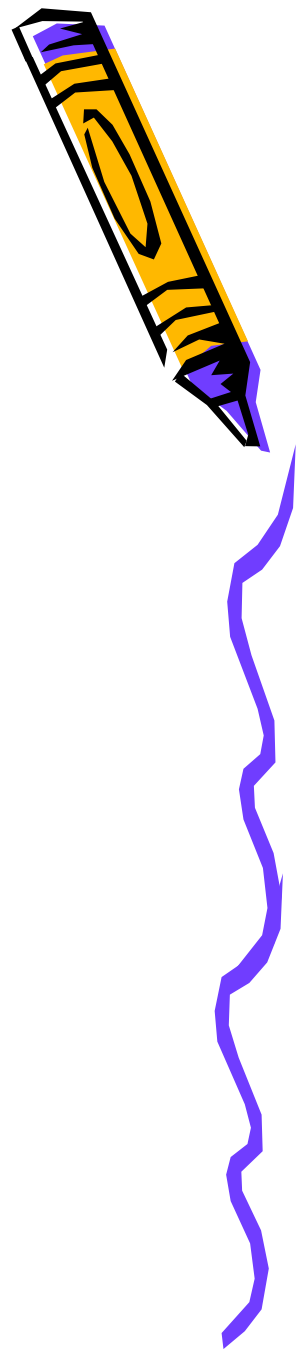
Cambios constantes

- Modificación constante
- Cambio en los datos de acuerdo a la necesidad



Consistentes en un punto del tiempo

- Se agregan datos regularmente, pero rara vez se realizan cambios directamente
- No significa que los DW nunca sufran cambios!!



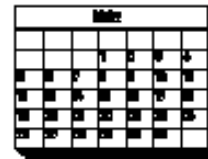
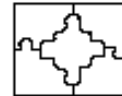
DW-Características (De tiempo variante)

Sistema
Operacional

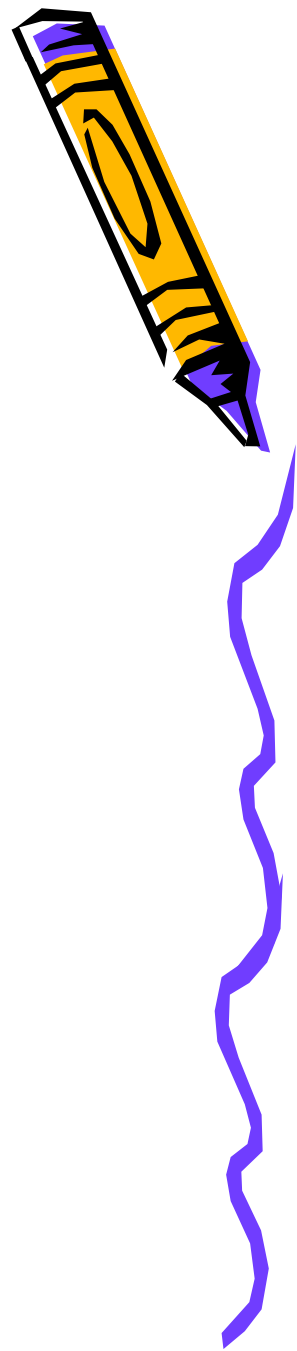


- Principalmente, datos de hoy!

Data
Warehouse

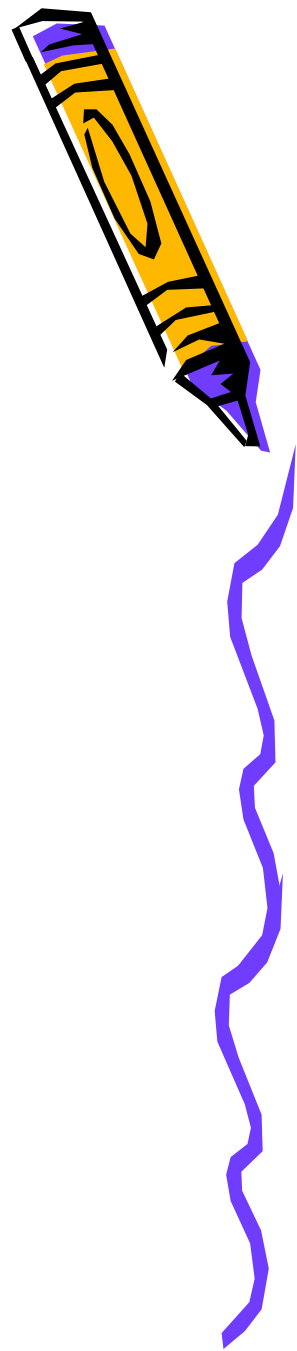


- Generalmente, mantiene datos históricos



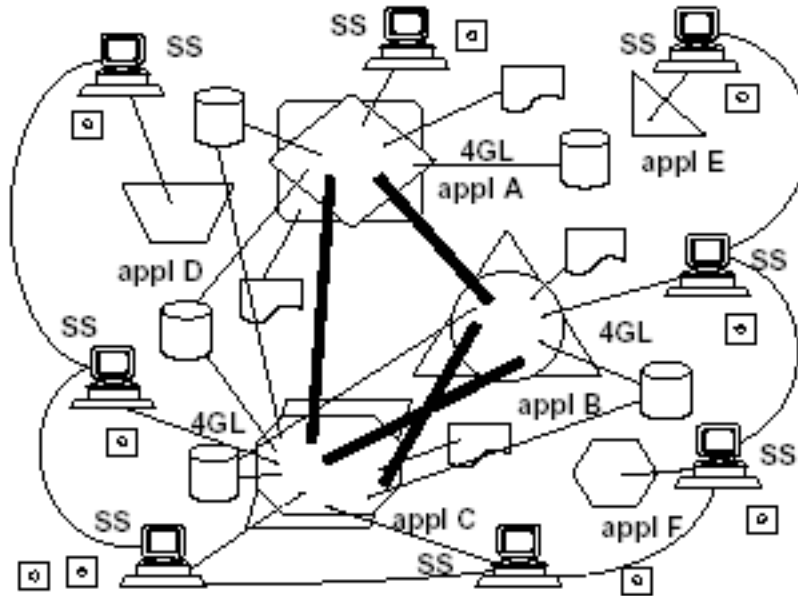
OLTP Vs DW (1)

Sistema Operacional (OLTP)	Data Warehouse
Almacena datos actuales	Almacena datos históricos
Almacena datos de detalle	Almacena datos de detalle y datos agregados a distintos niveles
Los datos son dinámicos (actualizables)	Los datos son estáticos
Los procesos (transacciones) son repetitivos	Los procesos no son previsibles
El número de transacciones es elevado	El número de transacciones es menor o medio
Dedicado al procesamiento de transacciones	Dedicado al análisis de datos
Orientado a los procesos de la organización	Orientado a la información relevante
Soporta decisiones diarias	Soporta decisiones estratégicas
Sirve a muchos usuarios (administrativos)	Sirve a técnicos de dirección

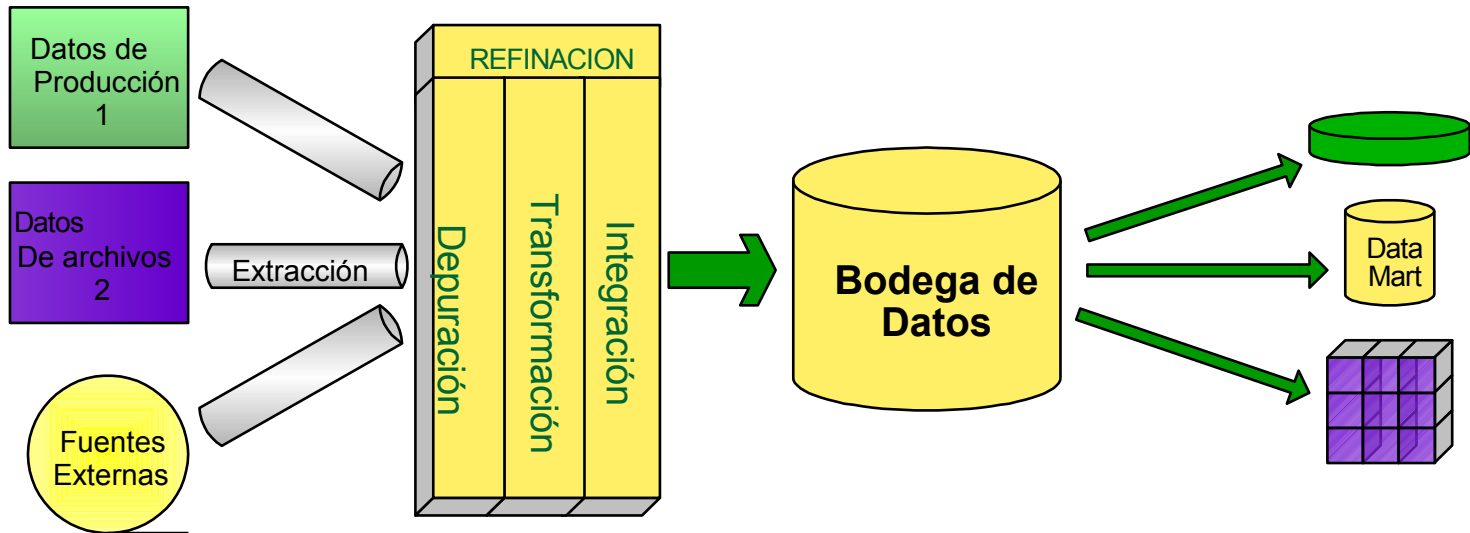


OLTP Vs DW (2)

La tradicional red de sistemas solitarios, contiene los datos que queremos. El problema es que no fue pensada para proveer de información, sino mantener datos.

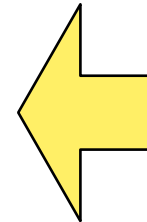


DW-Arquitectura (1)



METADATOS

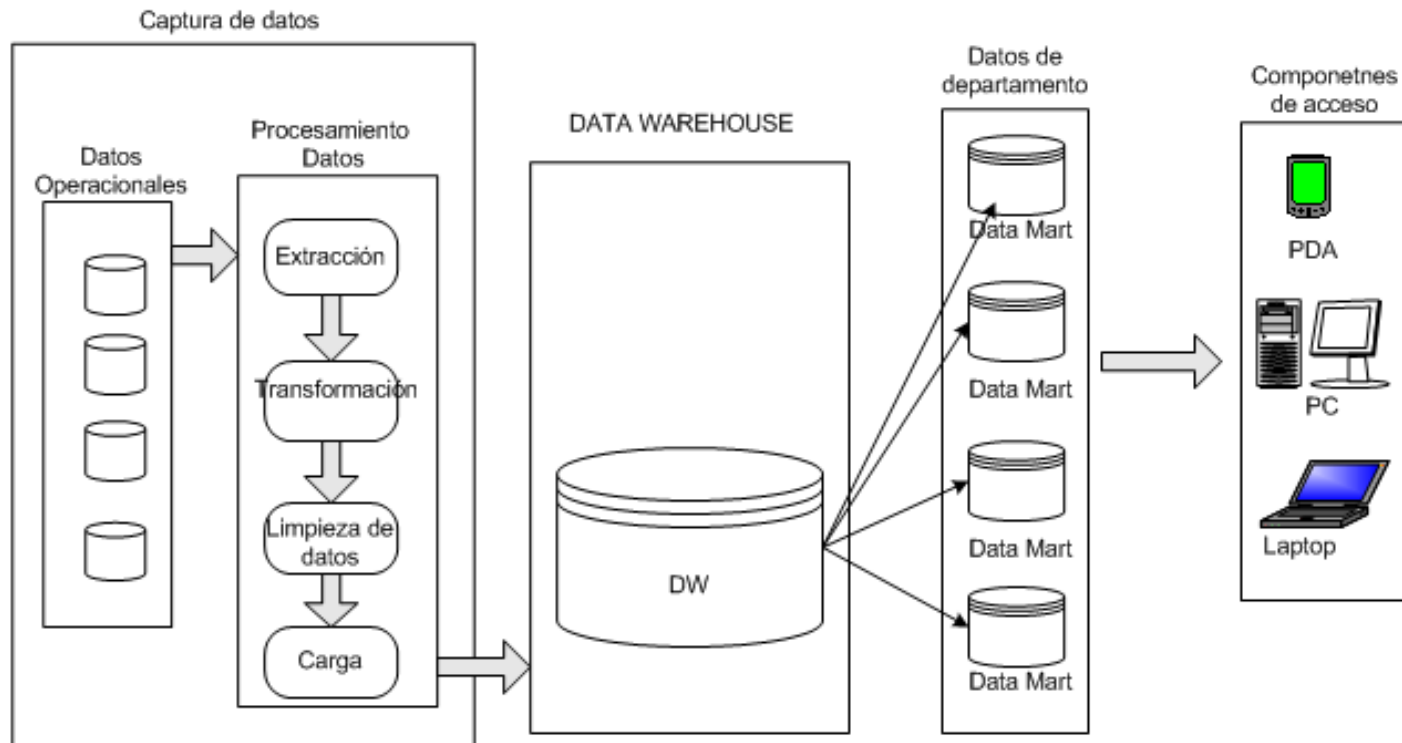
Requerimientos del
Negocio



Usuario
Final



DW-Arquitectura (2)

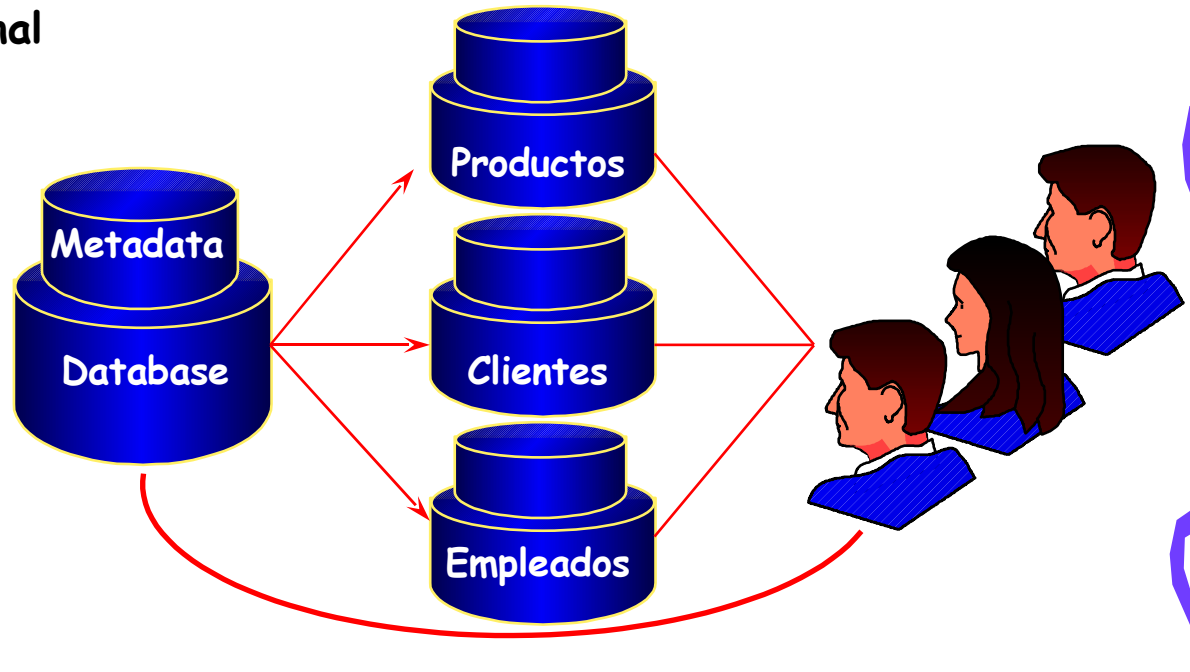


DW-Data Marts

Datos Operación Transaccional

Datos Operación Batch

Datos externos



DW-Caso de Estudio



- ¿En qué se parecen los pañales a la cerveza?
 - Si hacemos un análisis clásico de comportamiento del consumidor, aparentemente nada.
 - Pero si se establecen correlaciones entre los datos, "el bosque" se despeja.
 - Situación: "A las 19:00 horas de los días viernes hay un aumento notorio en las ventas de pañales y cervezas"
 - ¿Qué sucede?
 - Decisión táctica: Colocar la góndola de las cervezas cerca de los pañales.



DW-Caso de Estudio



- Productos Correlacionados
 - La idea es buscar que combinaciones de productos producen relaciones asociativas: pan + mantequilla ==> Alta probabilidad de comprar jamón.
 - A partir de preguntas que surgen de los datos, se toman acciones.
 - Al analizar el comportamiento de compra, se puede inferir que producto es el que induce la compra de otro.

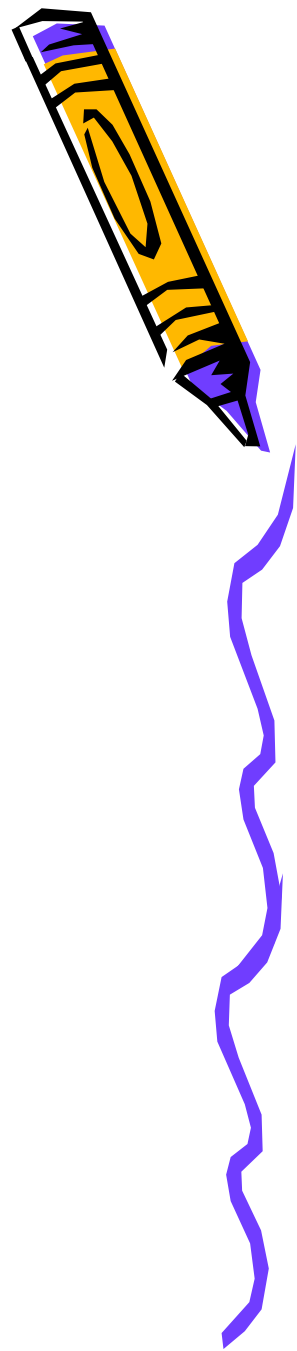


DW-Caso de Estudio

Escenarios posibles

- Se acerca el fin de semana.
 - Hay un bebé en casa.
 - Hay que comprar pañales.
 - El papá compra los pañales camino a casa.
 - Se acuerda que hay un partido de fútbol.
 - Desea comprar cerveza
- Se acerca el fin de semana.
 - Hay que comprar cerveza, vienen los amigos a ver el partido.
 - Pero, el bebé esta en casa, también la señora!!
 - Hay que comprar pañales primero.

Pañales → Cerveza



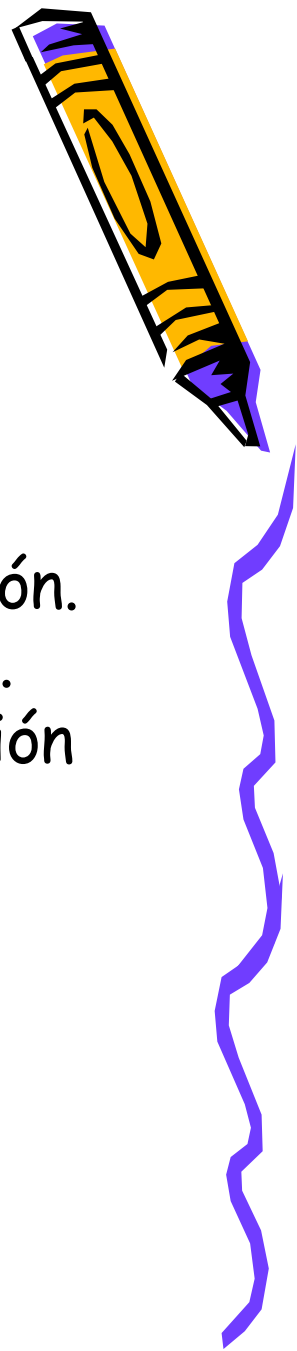
DW-Caso de Estudio



- Reglas Intrínsecas
 - El ejemplo anterior muestra reglas intrínsecas en el comportamiento de compra.
 - El problema es "¿cómo obtenerlas?"
 - Desde el punto de vista estadístico, es posible obtener correlaciones entre los datos. Lo malo es que no siempre se puede inferir una regla.
 - Existen paquetes de data mining, que permiten generar las reglas a partir de los datos.



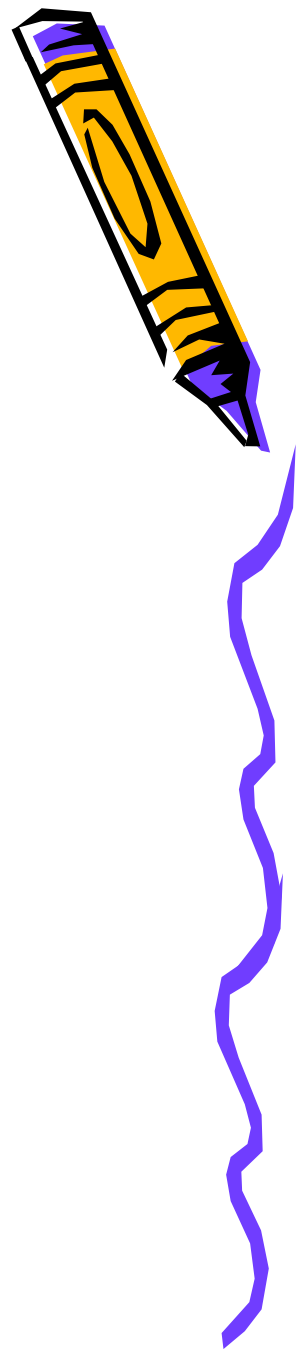
DW-Caso de Estudio



- Problemas a resolver
 - Definición de las fuentes de datos.
 - Consolidación y uniformidad de la información.
 - Definición de un repositorio de información.
 - Automatización de los procesos de extracción de información.
 - Minería de datos.
 - Apoyo de los expertos del negocio.
 - Mantenimiento y administración del conocimiento.



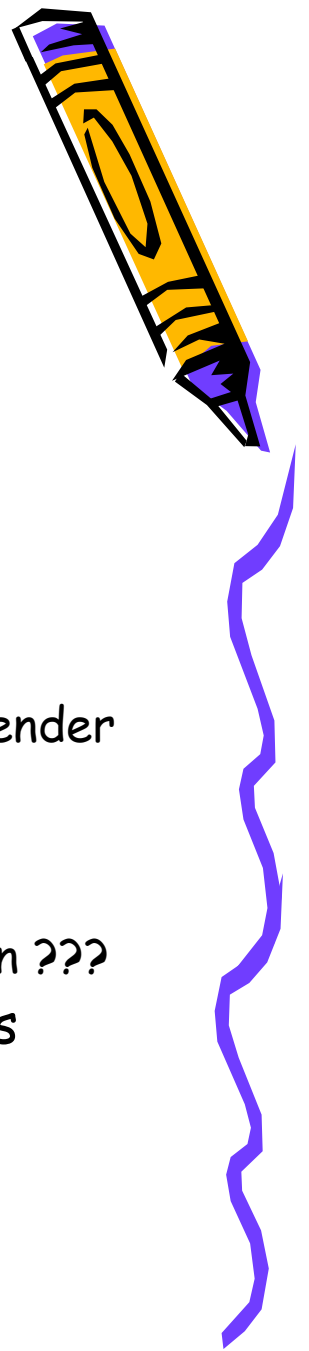
DW-Caso de Estudio



- Preguntas que se pueden resolver
 - Conocimiento: Qué, Quién, Cómo, Cuándo?
 - Predicción: Cómo será mañana?
 - Comprensión: Porqué?



DW-Caso de Estudio



- Qué NO se puede medir con modelos Multidimensionales
 - Preguntas complejas como:
 - Cuál es el siguiente producto (P2) que debo intentar vender al cliente X, si éste ya tiene el producto P1 ????
 - Cuáles son los clientes a los cuales debo enfocar ("targeting") esta campaña de mercadeo ????
 - Por qué razones se retiran los clientes de la institución ???
 - Este tipo de preguntas requiere de Minería de Datos

