

Unidad 10

Almacenes de Datos

Introducción

Definición

- Los Almacenes de Datos (data warehouse) son colecciones de datos orientadas a la toma de decisiones

Almacenes de Datos y Bases de Datos

- La principal diferencia en cuanto a concepto de una Base de Datos es que no se utilizan con fines operativos
- Los datos almacenados no se consideran actuales y se utilizan para fines de toma de decisiones (muchas veces con Minería de Datos)

OLAP y OLTP

- OLAP
- Procesamiento Analítico en Línea
- (*On Line Analytical Processing*)
- OLTP
- Procesamiento de Transacción en Línea
- (*On Line Transaction Processing*)
- Las Bases de Datos operativas usan OLTP
- Los Almacenes de Datos utilizan OLAP

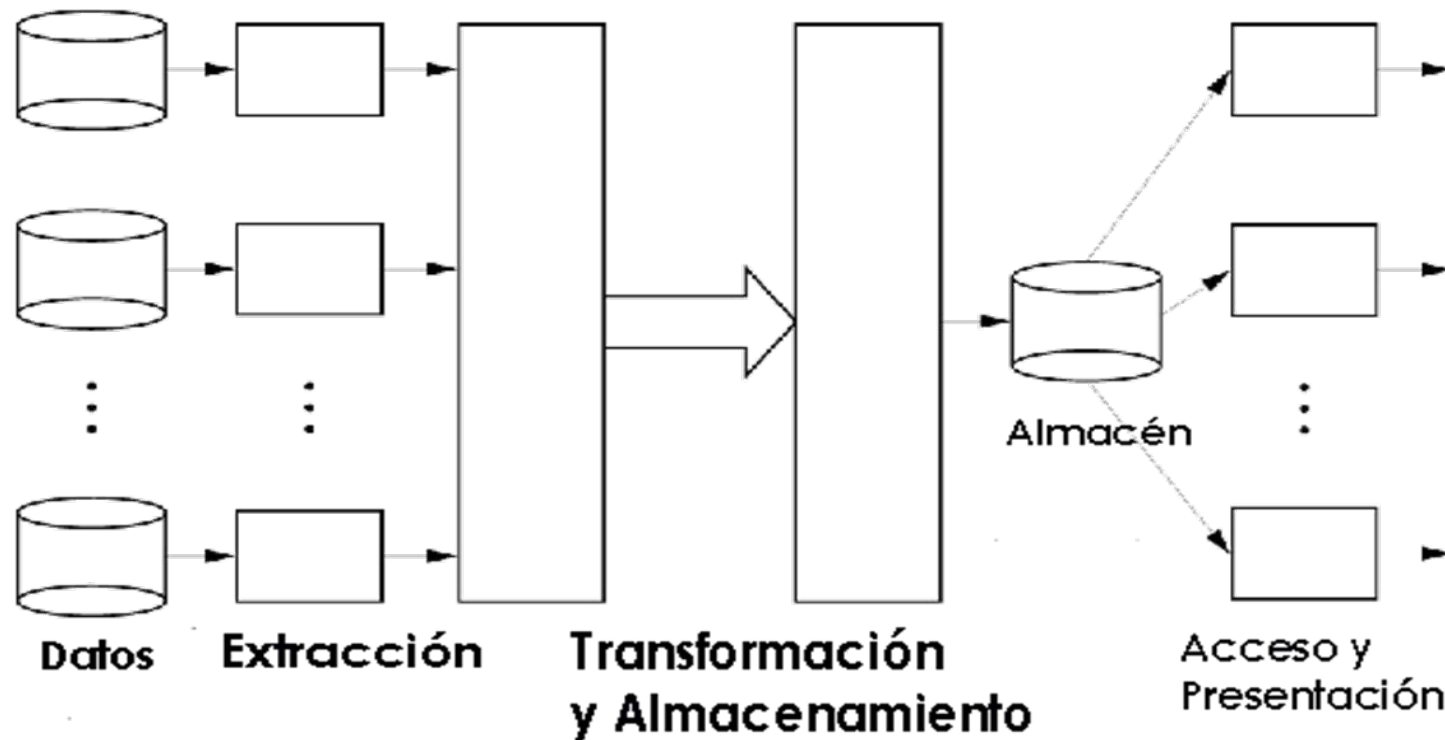
Otras Características

- Un Almacén de Datos tiene datos útiles para una organización (repositorio)
- Contienen datos históricos
- Entregará información adecuada en el momento que se requiera y en el formato más conveniente
- Se utiliza con Sistemas de Soporte de Decisiones
- Son datos de “solo lectura”
- Sus consultas no afectarán la operación de un sistema

Arquitectura de un Almacén de Datos

Arquitectura

- La Arquitectura de un Almacén de Datos está dada por sus componentes y la manera en que se construye



Elementos Básicos en la Construcción

- Integración
- Integración de los datos que vienen de Bases de Datos Distribuidas
- Se pueden tener estructuras de tablas muy distintas entre sí
- Separación
- Los datos de las Bases de Datos Distribuidas y las de Almacenes de Datos deberán ser independientes y no estar almacenados en la misma base de datos

Proceso de Construcción

- Los Datos se extraen de diversas fuentes externas (BD, archivos, etc)
- Se realiza un proceso de limpieza
- Se realiza un proceso de formateo
- Proceso de Carga de información

Mercados de Datos

- También conocidos como Data marts
- Son un sub conjunto de la información que se almacena en un Almacén de Datos
- Por ejemplo información sobre una sola área de una empresa
- Su objetivo también es proporcionar conocimiento para la toma de decisiones
- Pueden procesarse para generar más información para un Almacén

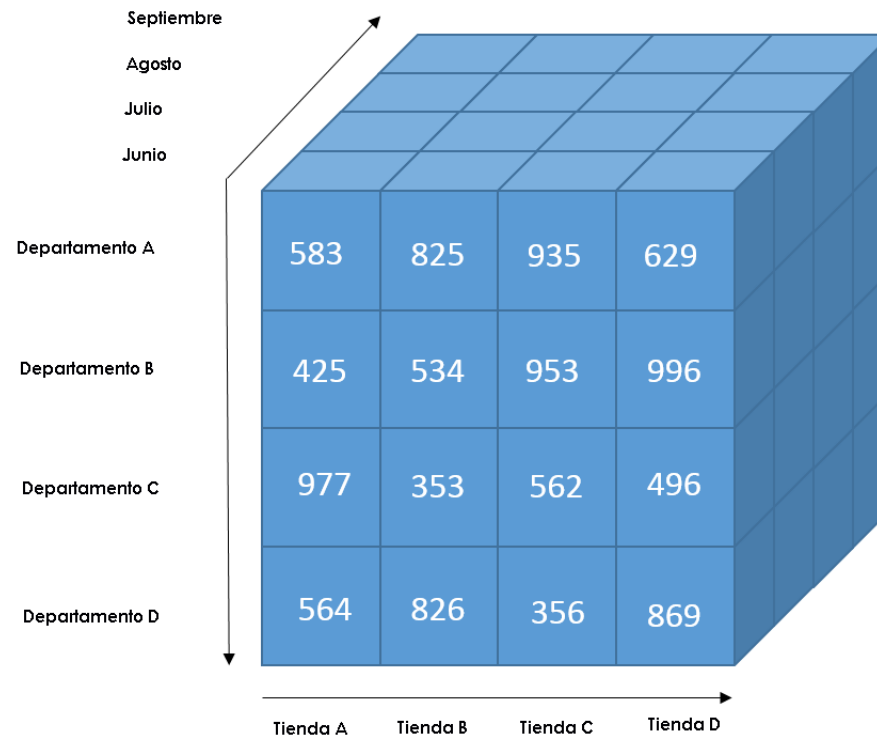
Modelos de Datos

El Cubo OLAP

- Los almacenes de datos utilizan un modelo multidimensional conocido como Cubo de datos
- Cada dimensión se representa mediante su propia tabla
- Se pueden tener más de 3 dimensiones pero no es posible representarlas “físicamente”
- La cara frontal es la información que se está consultando

Representación de Cubo

- La cara frontal es el mes de Junio, por lo que se está consultando la cantidad de ventas por tienda y departamento en el mes de Junio



Pivoteo

- Se refiere a que los usuarios pueden ver los datos por cualquier dimensión que se desee

Modificaciones

- Roll-Up
 - Se refiere a tener una granularidad más gruesa sobre alguna de las dimensiones
- Drill-down
 - Se proporciona información sobre cierta dimensión usando una granularidad más fina

Roll-Up

- Información acerca de una Tienda, haciendo Roll-up sobre los departamentos

	Junio	Julio	Agosto	Septiembre
Tienda A	2549	2750	2700	2987
Tienda B	2538	2740	2678	2488
Tienda C	2806	3000	2995	2788
Tienda D	2990	3200	3100	2899

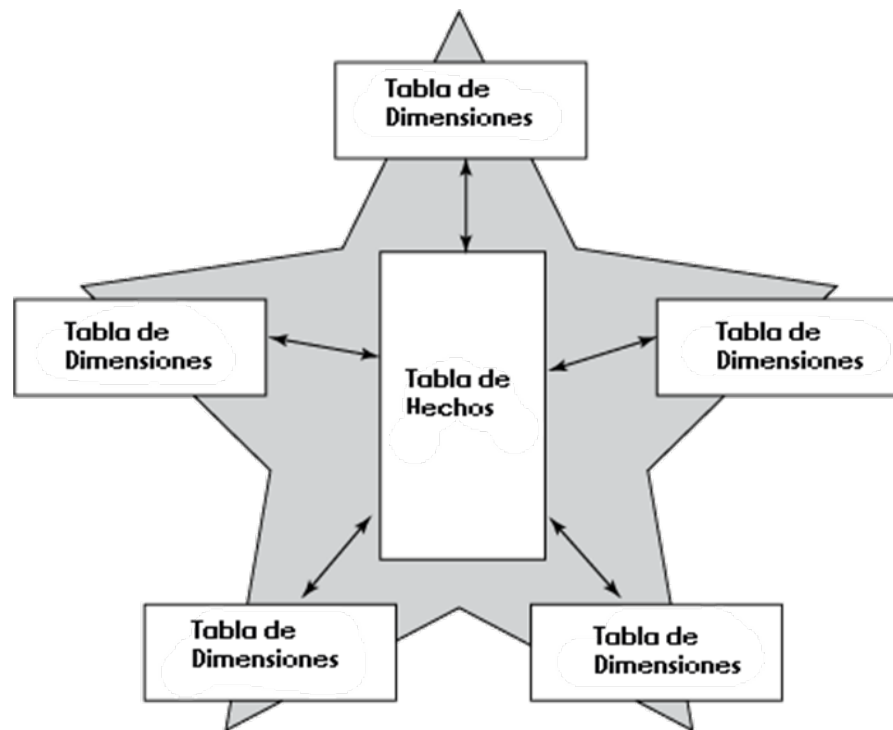
Drill Down

- Información específica de un Mes haciendo Drill-Down sobre los Departamentos

	Tienda A	Tienda B	Tienda C	Tienda D
Departamento A				
Producto 1	125	145	167	155
Producto 2	145	156	177	144
Departamento B				
Producto 1	176	144	155	167
Producto 2	189	199	201	176
Departamento C				
Producto 1	165	187	187	198
Producto 2	148	176	154	165
Departamento D				
Producto 1	167	165	178	163

Representación de Estrella

- Existe una tabla central conocida como Tabla de Hechos
- La rodean (como los picos de una estrella) las Tablas Dimensión



Representación de Copo de Nieve

- Las Tablas de Dimensiones tienen dimensiones causadas por alguna normalización

