

BASE DE DATOS III
LIC. ROMINA ROJAS

Unidad 5: Modelo multidimensional

Contenido

2

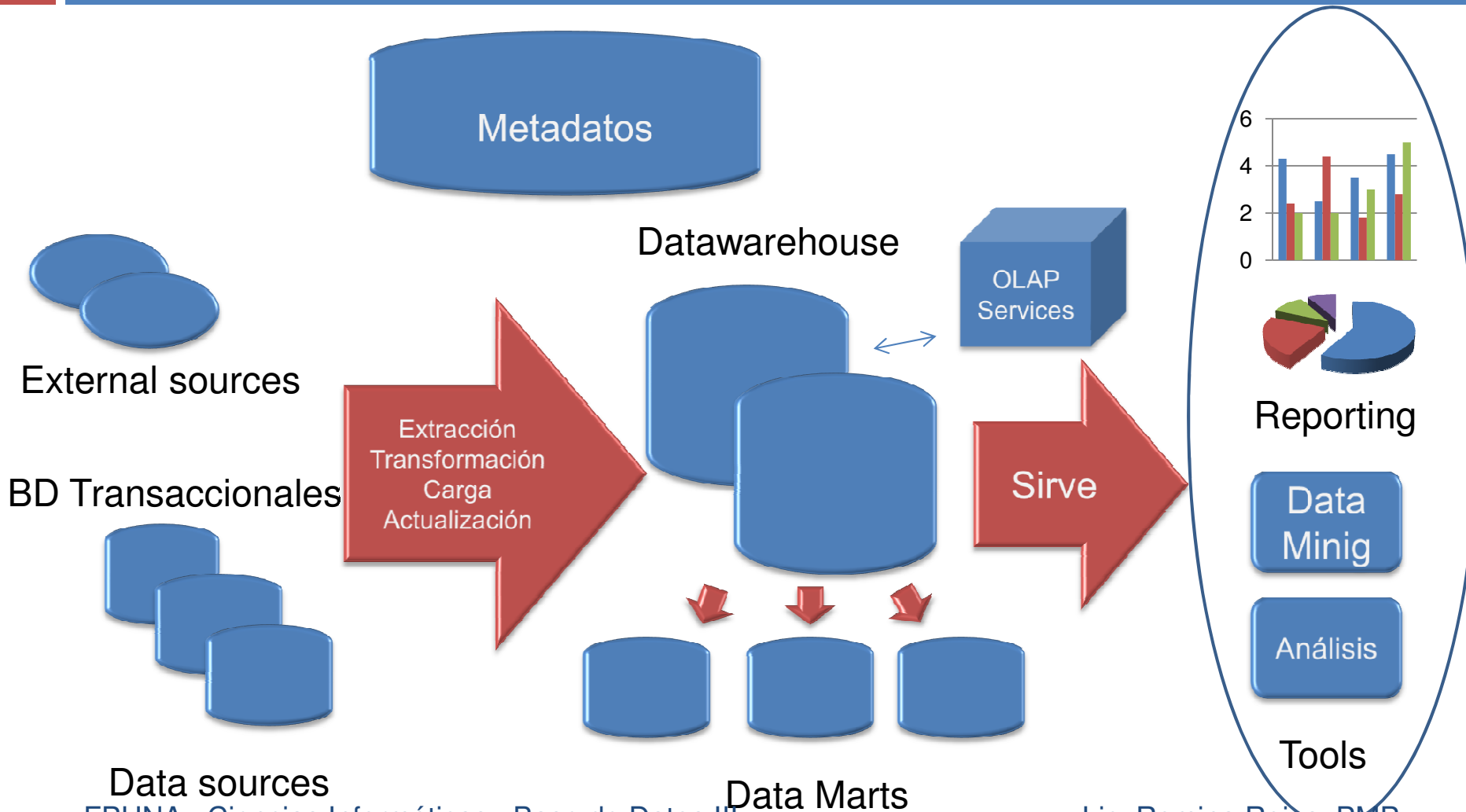
- Introducción
- Procesamiento analítico en línea OLAP.
- Modelos multidimensionales.
 - ▣ Esquemas relacionales.
 - ▣ Esquemas de representación.
- Implementación
- Servidores OLAP.
- Modelamiento multidimensional. Pasos a seguir.

Introducción

- La tecnología Datawarehousing debido a su orientación analítica, impone un procesamiento y pensamiento distinto, la cual se sustenta por un modelamiento de Bases de Datos propio, conocido como Modelamiento Multidimensional, el cual busca ofrecer al usuario su visión respecto de la operación del negocio.

Elementos del Datawarehouse

4



Procesamiento Análítico en Línea

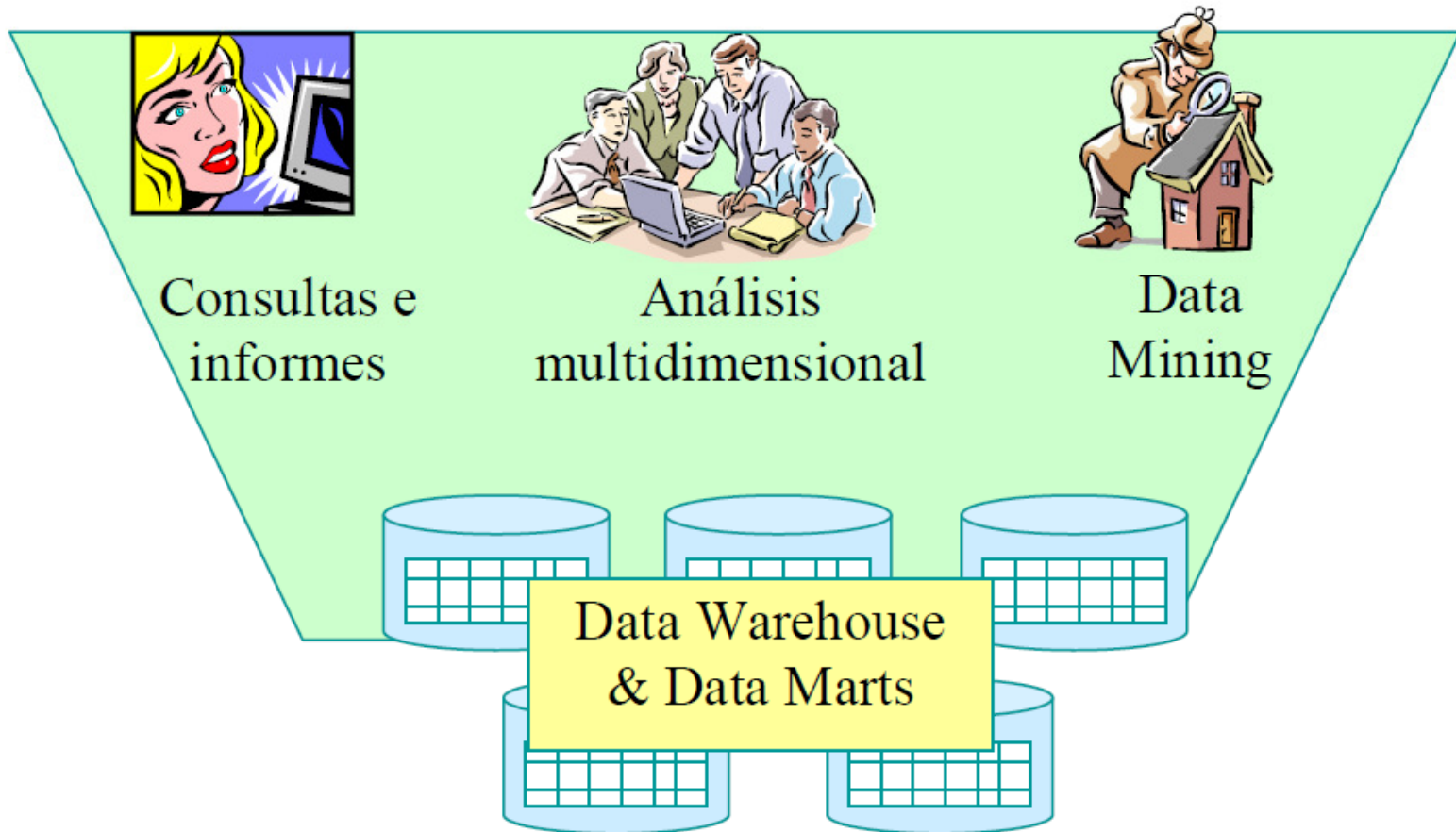
OLAP

5

- El procesamiento analítico en línea es una tecnología que permite a los analistas y administradores visualizar y navegar los datos accediendo a una amplia variedad de vistas posibles de la información de manera interactiva, rápida y eficiente.
- Como parte de los servicios de explotación, los datos tienen que ser modelados multidimensionalmente para satisfacer los requerimientos de desempeño de este tipo de consultas en línea.

Procesamiento analítico en línea OLAP

6



Modelamiento multidimensional

7

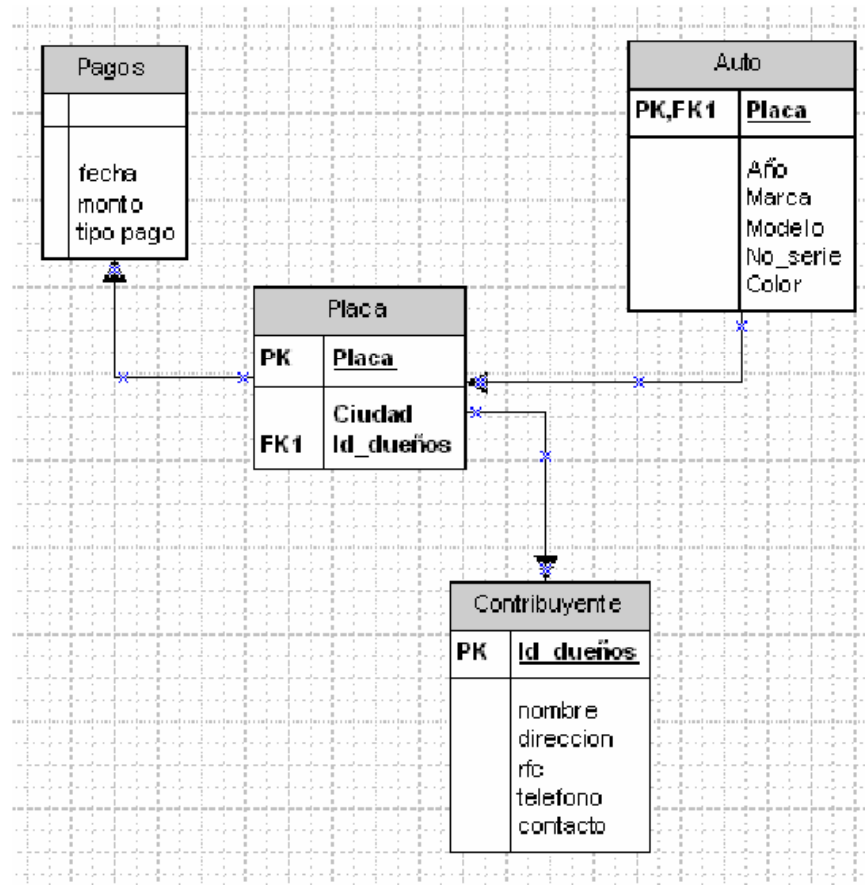
- Es una técnica para modelar bases de datos simples y entendibles al usuario final. La idea fundamental es que el usuario visualice fácilmente la relación que existe entre los distintos componentes del modelo.

Modelos multidimensionales

- Un modelo de datos es la representación de los datos y sus relaciones con otros datos que se utiliza para conocer como se organizarán los datos en bases de datos u otro medio de almacenaje y administración de datos.

Ejemplo Modelo ER

9



Esquemas relacionales

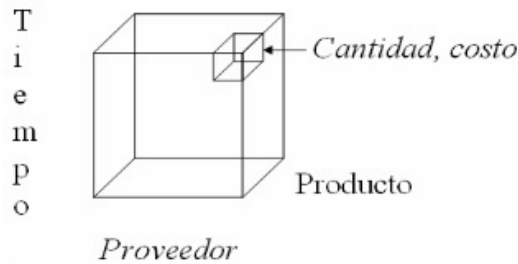
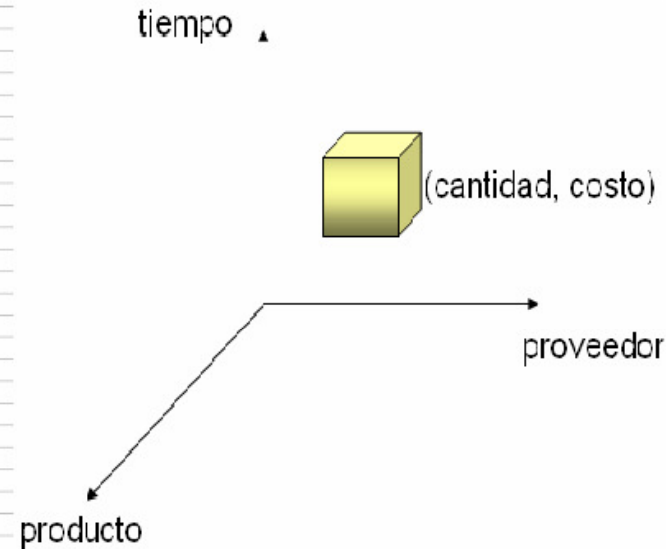
10

- Un modelo multidimensional o cubo es una colección de medidas las cuales dependen de un conjunto de dimensiones, es una representación de los datos que permite organizarlos en la forma de hechos, dimensiones y agregados.
 - Hechos: Contienen medidas, es decir, la información a nivel transaccional que vamos a analizar.
 - Dimensiones: Contiene información descriptiva de esas transacciones. Ej.: Fecha, Cliente, Producto, etc.
- Un modelo multidimensional se utiliza para el análisis de información.

Esquemas relacionales

11

1	FECHA	COMPRA-\$	CANTIDAD	PROVEEDOR	PRODUCTO
2	Ene-04	1302	2	AMAZON	LIBRO
3	Ene-04	232	34	GANDHI	REVISTA
4	Ene-04	234	2	BARNES	REVISTA
5	Feb-04	2	45	CRYSTAL	LIBRO
6	Feb-04	488	3	GANDHI	LIBRO
7	Mar-04	745	2	CRYSTAL	LIBRO
8	Mar-04	224	6	AMAZON	REVISTA
9	Abr-04	1248	5	CRYSTAL	LIBRO
10	Abr-04	121287	32	GANDHI	REVISTA
11	May-04	41212	6	CRYSTAL	LIBRO
12	May-04	122	21	CRYSTAL	LIBRO
13	Jun-04	1545	1	AMAZON	LIBRO
14	Jun-04	4545	3	SOTANO	REVISTA
15	Jul-04	4545	45	SOTANO	LIBRO
16	Jul-04	1212	21	SOTANO	REVISTA
17	Jul-04	4548	1	AMAZON	LIBRO
18	Ago-04	121285	3	SOTANO	LIBRO
19	Ago-04	42	6	GANDHI	REVISTA
20	Sep-04	1212	2	AMAZON	LIBRO
21	Oct-04	21245	1	GANDHI	REVISTA
22	Oct-04	1515	5	SOTANO	REVISTA
23	Nov-04	87854	2	SOTANO	LIBRO
24	Nov-04	6568	1	SOTANO	LIBRO
25	Nov-04	266	4	AMAZON	REVISTA
26	Dic-04	56497	1	AMAZON	LIBRO
27	Dic-04	266	5	BARNES	LIBRO



Esquemas relacionales

12

- **Dimensiones:** Perspectivas o entidades respecto a las cuales una organización quiere mantener sus datos organizados (p.ej. tiempo, localización, clientes, proveedores...)
 - Miembros: Nombres o identificadores que marcan una posición dentro de la dimensión. Ejemplos:
 - Meses, trimestres y años son miembros de la dimensión tiempo.
 - Ciudades, regiones y países son miembros de la dimensión localización.
 - Jerarquías: Los miembros de las dimensiones se suelen organizar en forma de jerarquías.

Esquemas relacionales

13

- **Hechos:** Colecciones de datos relacionados compuestas por medidas y un contexto.
 - ▣ Las dimensiones determinan el contexto de los hechos.
 - ▣ Cada hecho particular está asociado a un miembro de cada dimensión.
- **Medidas:** Atributos numéricos asociados a los hechos (lo que realmente se mide).
 - ▣ Ejemplos: Volumen de las ventas, coste asociado a un producto, número de transacciones efectuadas, porcentaje de beneficios.

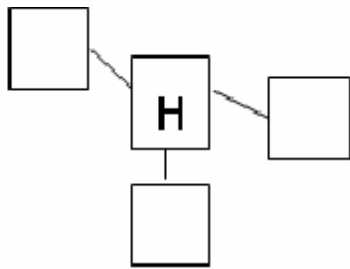
Esquemas relacionales

14

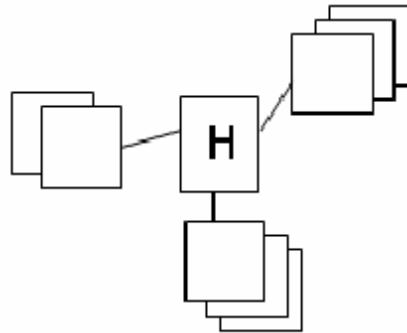
- En el modelo multidimensional cada eje corresponde a una dimensión particular .
- La dimensionalidad de nuestra base estará dada por la cantidad de ejes (o dimensiones) que le asociemos.
- Cuando una base puede ser visualizada como un cubo de tres o más dimensiones, es más fácil para el usuario organizar la información e imaginarse en ella cortando y rebanando el cubo a través de cada una de sus dimensiones, para buscar la información deseada.
- La descripción de una organización típica es: “Nosotros vendemos productos en varios mercados, y medimos nuestro desempeño en el tiempo”: Un diseñador dimensional lo verá como: “Nosotros vendemos productos en varios mercados, y medimos nuestro desempeño en el tiempo.”

Esquemas de Representación

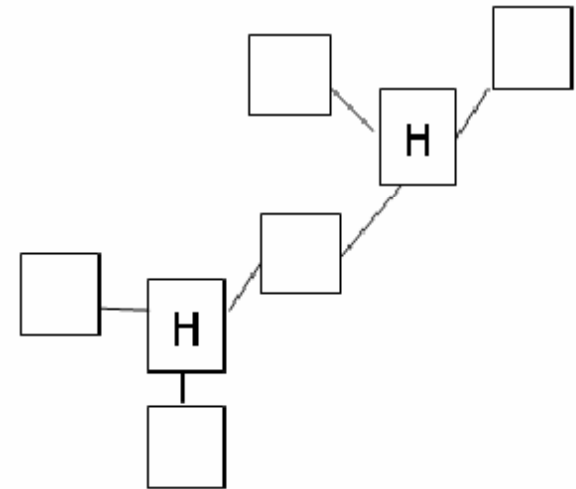
15



Estrella



Copo de nieve



Constelación

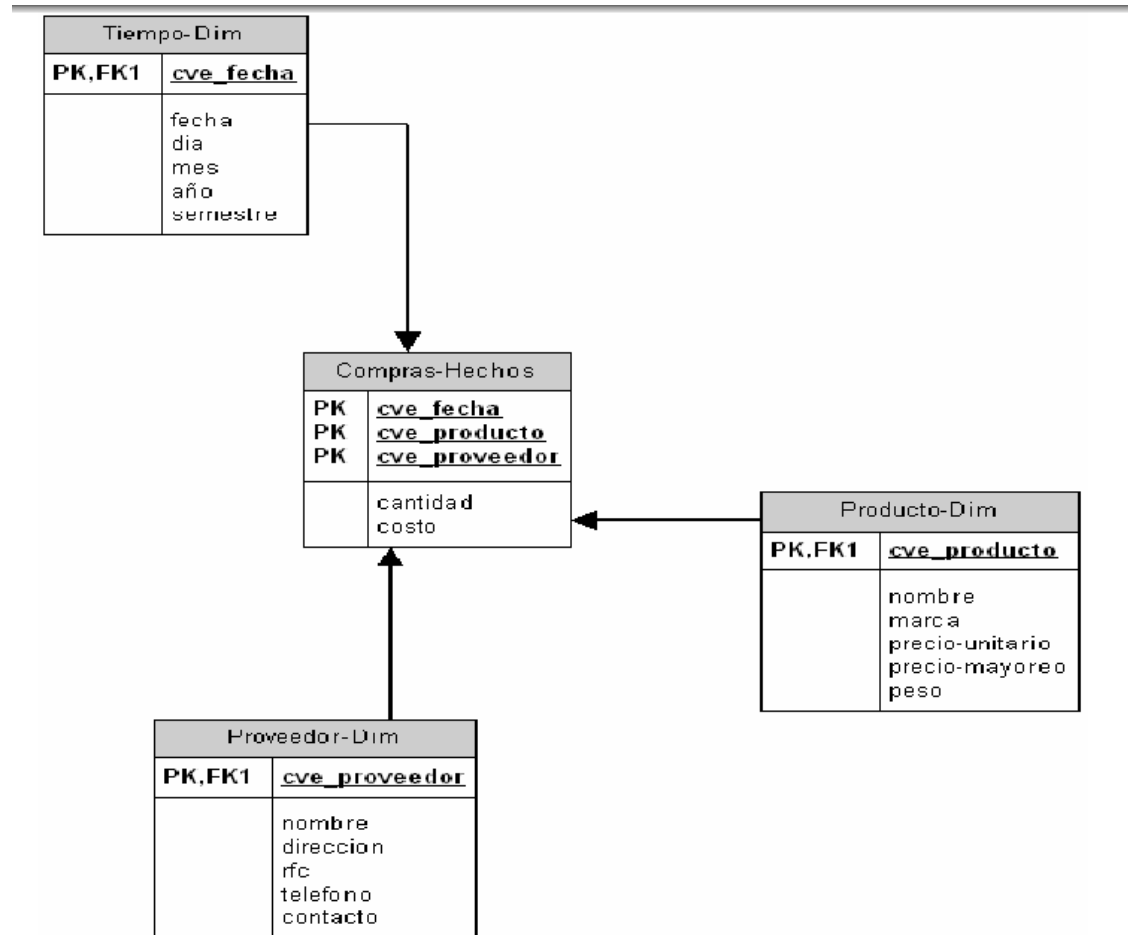
Modelo o esquema de Estrella

16

- Técnica de modelado de datos usado para hacer corresponder un Modelo multidimensional a una BD relacional.
- Tiene 4 componentes: hechos, dimensiones, atributos y jerarquía de atributos.
- Los hechos y dimensiones son representados por tablas físicas en el Dw.
- La tabla de hechos está relacionada a cada dimensión en una relación uno a muchos, a través de claves foráneas.

Modelo o esquema de Estrella

17



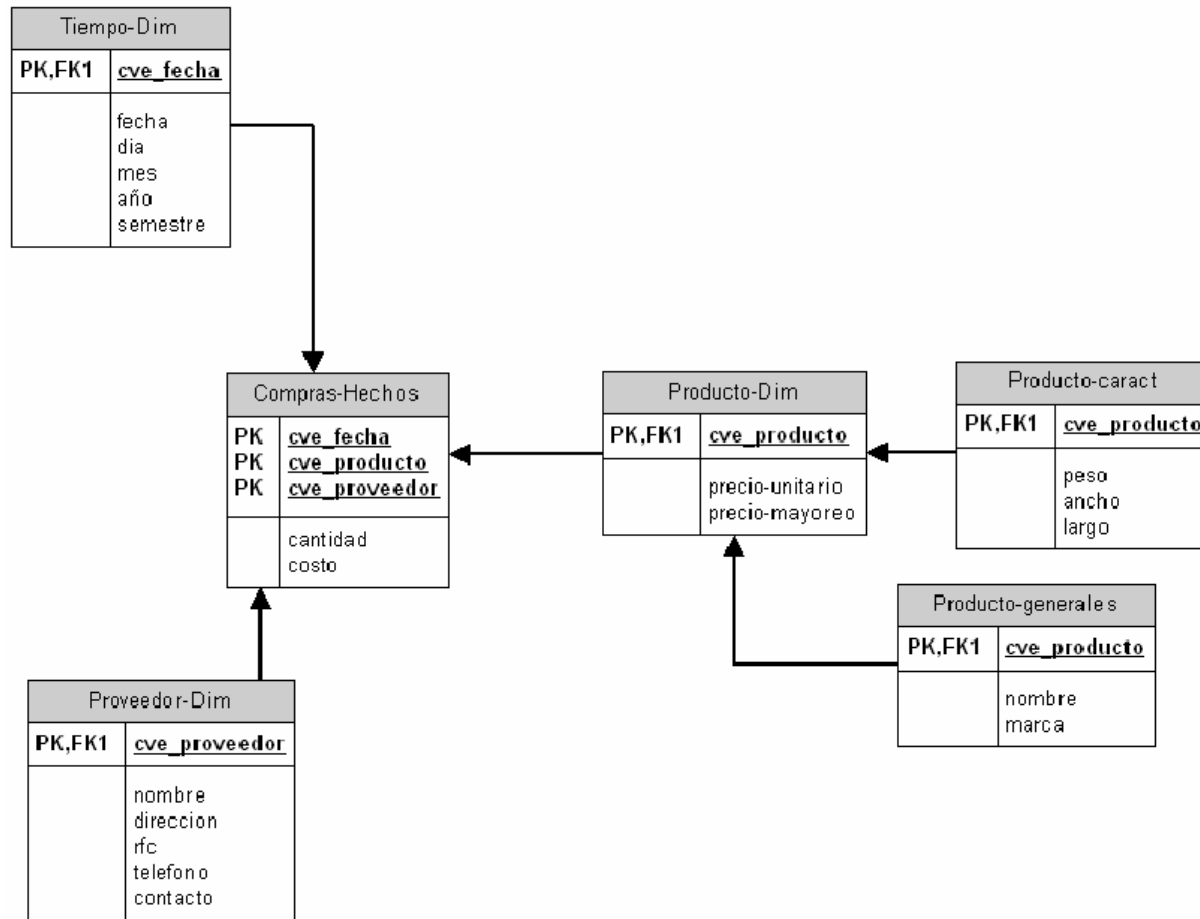
Modelo de copo de nieve

18

- Es una variación de la estrella tradicional.
- En cada dimensión se almacenan jerarquías de atributos o bien simplemente se separan atributos en otra entidad por razones de desempeño y mejor utilización del espacio.

Modelo de copo de nieve

19



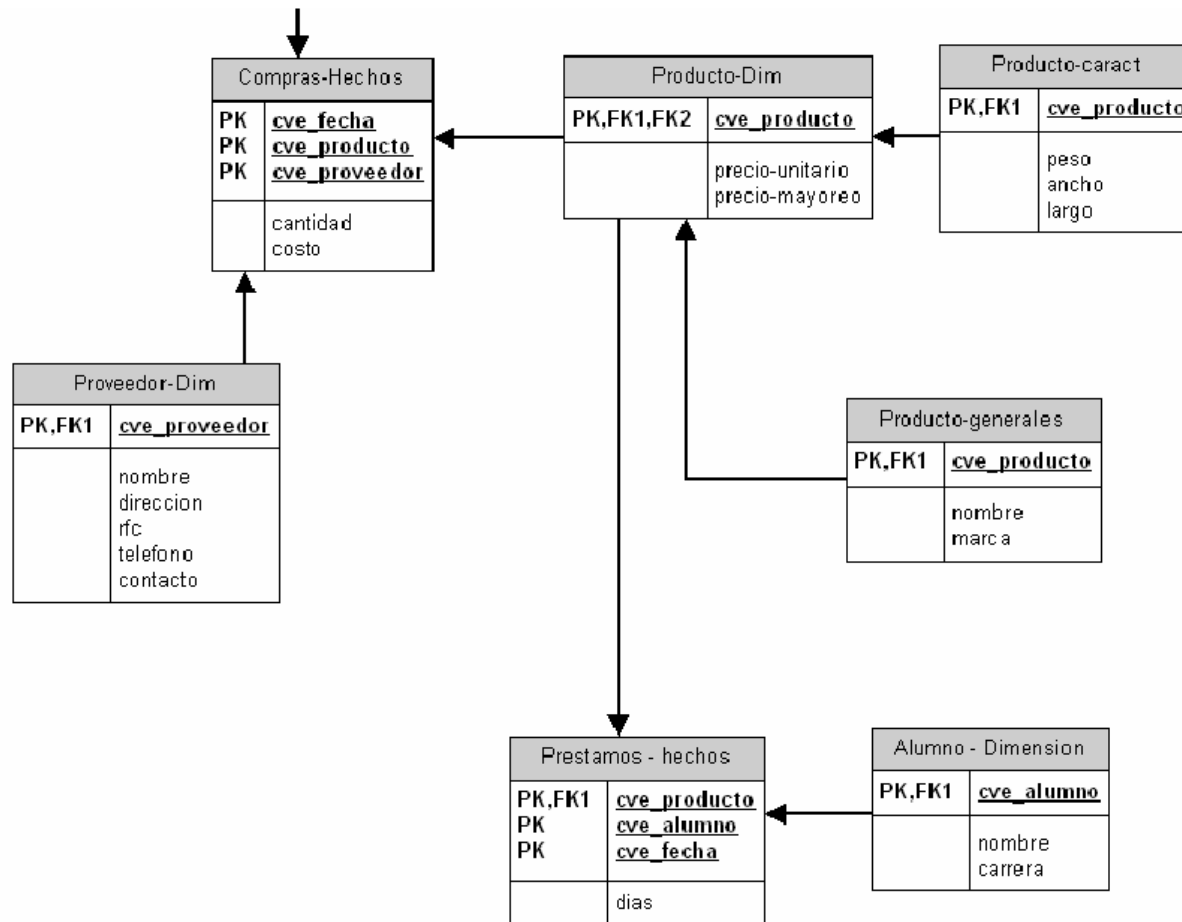
Modelo de constelaciones

20

- Es una variación del esquema estrella tradicional.
- Algunos atributos de las dimensiones se separan formando una nueva entidad que puede ser compartida por otro cubo.
- La utilidad principal de este modelo es que al tener dimensiones que pueden ser compartidas por diferentes cubos se tendrá un mejor uso del espacio de almacenamiento evitando la redundancia.

Modelo de constelaciones

21



Implementación

22

- El mantenimiento de los Data warehouses se realizan en forma separa de las bases de datos operacionales.

- El data warehouse se puede implementar utilizando servidores OLAP:
 - Multidimensionales: MOLAP
 - Relacionales: ROLAP.

Servidores OLAP

23

- Permite definir y navegar un modelo multidimensional.
- Es un administrador de BD especializado en el almacenamiento de cubos y en responder de manera óptima a las operaciones sobre ellos.

Servidor ROLAP

(Relational OLAP)

24

- Permite ser implementado sobre un DBMS relacional.
- Los esquemas más comunes sobre los que se trabaja son estrella o copo de nieve.
- Almacena los cubos en tablas relacionales.
- Las operaciones sobre ellos se realizan con SQL o con SQL extendido para manejar cubos.
- Proporciona algunas funcionalidades para implementar tanto los cubos como las operaciones sobre ellos. Ejemplo: las vistas materializadas.

Servidor MOLAP

(Multidimensional OLAP)

25

- Almacena los datos multidimensionales directamente en estructuras de datos especiales (arreglos) e implementa las operaciones OLAP sobre esas estructuras.
- Para optimizar los tiempos de respuesta, el resumen de la información es usualmente calculado por adelantado.
- Estos valores precalculados o agregaciones son la base de las ganancias de desempeño de este sistema.
- Algunos sistemas utilizan técnicas de compresión de datos para disminuir el espacio de almacenamiento en disco debido a los valores precalculados.

Servidor HOLAP

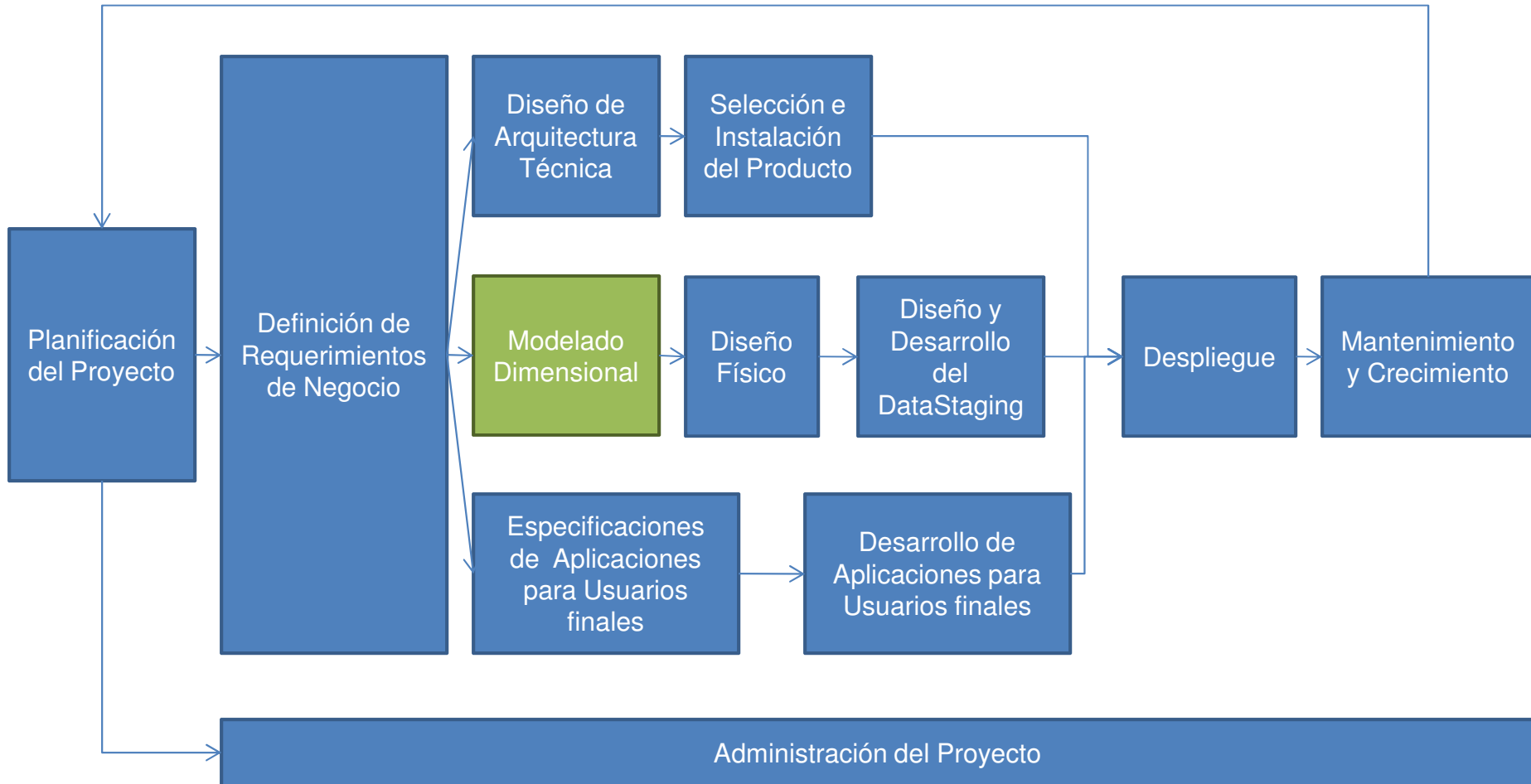
(Hybrid OLAP)

26

- Implementa las ventajas de ROLAP y MOLAP.
- Utiliza cubos en MOLAP y cuando se quiere ir al detalle de la información utiliza la que se encuentra en tablas relaciones.

Modelamiento multidimensional

27



Pasos básicos

28

- **Paso 1:** Decidir cuáles serán los procesos de negocios a modelar, basándose en el conocimiento de éstos y de los datos disponibles. Ejemplo: Gastos realizados por cada mercado para cada ítem a nivel mensual. Productos vendidos por cada mercado según el precio en cada mes.

Pasos básicos

29

- **Paso 2:** Decidir el Granularidad de la tabla de Hechos de cada proceso de negocio. Ejemplo: Producto x mercado x tiempo.
- En este punto se debe tener especial cuidado con la magnitud de la base de datos, con la información que se tiene y con las preguntas que se quiere responder.
- El grano decidirá las dimensiones del DW.
- Cada dimensión debe tener el grano más pequeño que se pueda puesto que las preguntas que se realicen necesitan cortar la base en caminos precisos (aunque las preguntas no lo pidan explícitamente).

Pasos básicos

30

- **Paso 3:** Decidir las dimensiones a través del grano.
- Las dimensiones presentes en la mayoría de los DW son: tiempo, mercado, producto, cliente.
- Un grano bien elegido determina la dimensionalidad primaria de la tabla de hechos.
- Es posible usualmente agregar dimensiones adicionales al grano básico de la tabla de hechos, donde estas dimensiones adicionales toman un solo valor para cada combinación de las dimensiones primarias.
- Si se reconoce que una dimensión adicional deseada viola el grano por causar registros adicionales a los generados, entonces el grano debe ser revisado para acomodar esta dimensión adicional.

Pasos básicos

31

- **Paso 4:** Elegir las mediciones del negocio para la tabla de hechos. Se deben establecer los ítems que quedarán determinados por la clave compuesta de la tabla de hechos. Ejemplos: cantidades, montos.

Claves subrogadas

32

- Una Clave Subrogada es un **campo numérico** de una tabla cuyo único requisito es **almacenar un valor numérico único para cada fila de la tabla**, actuando como una **clave sustituta**, de forma totalmente independiente a los datos de negocio, que habitualmente **no tiene significado por sí misma**. En consecuencia, es posible crear una Clave Subrogada en cualquier tabla (se trate de una Tabla Versionada o no), aunque habitualmente resultan especialmente útiles al trabajar con Tablas de Dimensiones Versionadas.
- Ejemplo: Un cliente puede cambiar de zona, si el usuario desea analizar los datos por zonas, conviene tener claves subrogadas para el cliente, entonces en la Base de datos Operacional existirá un único registro para dicho cliente; sin embargo, en el DW podremos tener varios registros de dicho cliente (Tabla dimensional versionada).

- Preguntas???
- Nos vemos en la siguiente Clase!!!