

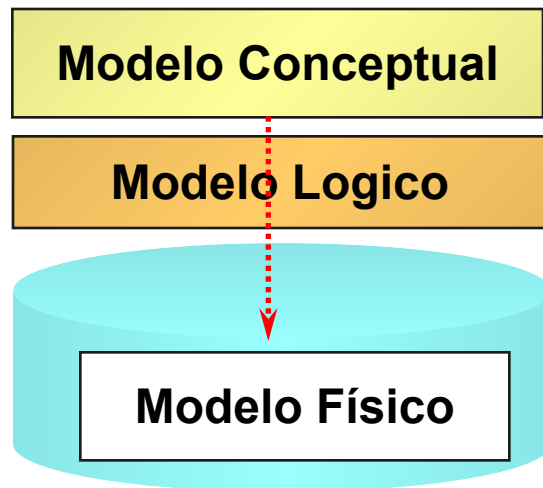


Modelado multidimensional. Parte I : Diseño conceptual

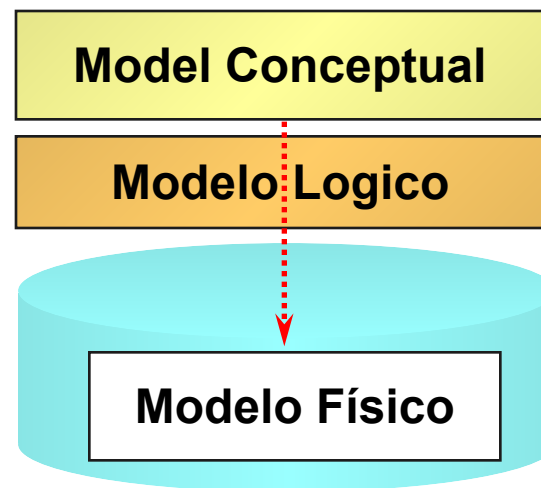
Tema 3

Modelos utilizados en BD

- Modelo Conceptual
- Modelo Logico
- Modelo Físico



**Sistemas
Operacionales**

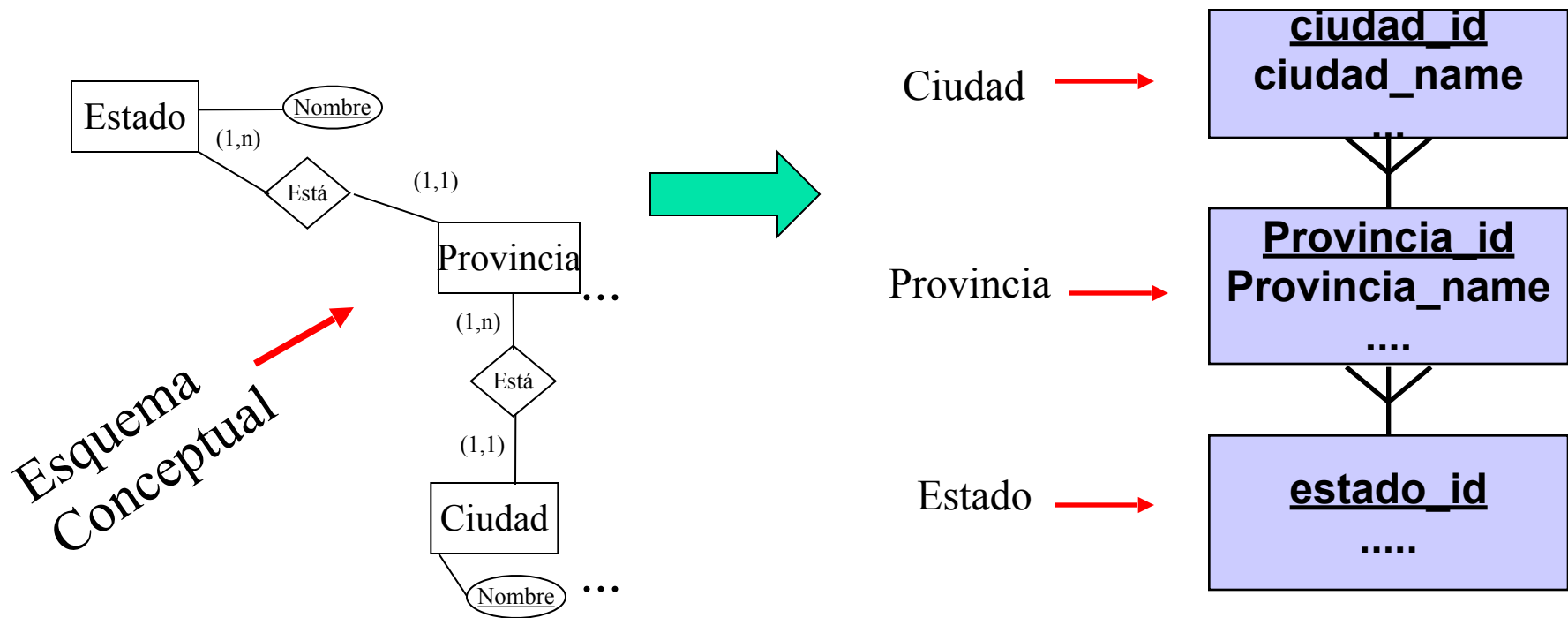


Almacenes de Datos

Modelos utilizados en BD

Sistemas operacionales

- Normalización
 - Sist. Transaccionales





Diseño conceptual de DW

- **Los modelos conceptuales**

- Permiten una mejor comunicación entre diseñadores y usuarios para comprender los requisitos de aplicación
- No hay un modelo conceptual bien establecido para datos multidimensionales
- Varias propuestas basadas en UML, en el modelo ER, o en el uso de notaciones específicas
- Problemas:
 - Enfocados a modelar sistemas transaccionales
 - No reflejan las propiedades multidimensionales (hechos, dimensiones, etc.)
 - La falta de un mapeo a la plataforma de implementación



Diseño conceptual de DW

- Modelado multidimensional (MD)
 - Parte estructural
 - Parte dinámica
- Parte estructural
 - Hechos y dimensiones
- Parte dinámica
 - Operaciones de consulta al modelo MD



Modelado multidimensional

- Tal y como el usuario percibe el mundo real objeto de estudio
 - Perspectiva estructural
 - Modelado o Modelo Multidimensional (MD)
 - Hechos y Dimensiones
 - Perspectiva dinámica
 - Definición de requerimientos iniciales sobre el modelo MD
 - Operaciones de consulta avanzada



Modelado multidimensional

Parte estructural

- Perspectiva estructural → modelo MD
 - Hechos
 - Objeto de análisis
 - Ej. Ventas de productos, compras, alquileres, transportes
 - Dimensiones
 - Diferentes perspectivas para analizar los hechos
 - Ej. Productos, almacenes, tiempo, vehículos, etc.



Modelado multidimensional

Parte estructural

- Hechos representan normalmente relaciones muchos a uno con cada dimensión en particular
 - Ej. Ventas de productos (H) por producto (D), almacenes (D) y tiempo (D)
 - Un producto (D) \rightarrow varias ventas (H)
 - Una venta (H) \rightarrow un solo producto (D) y almacén (D)



Modelado multidimensional

Parte estructural

- Sin embargo, a veces hechos son muchos a muchos con dimensiones en particular
 - Ej. Tickets emitidos (H) por
 - Un ticket (H) puede contener muchos productos
- Hechos y Dimensiones se caracterizan por atributos
 - **Hechos** → atributos de hecho o **medidas**
 - **Dimensiones** → atributos de dimensión
 - Definidos por la **jerarquía**



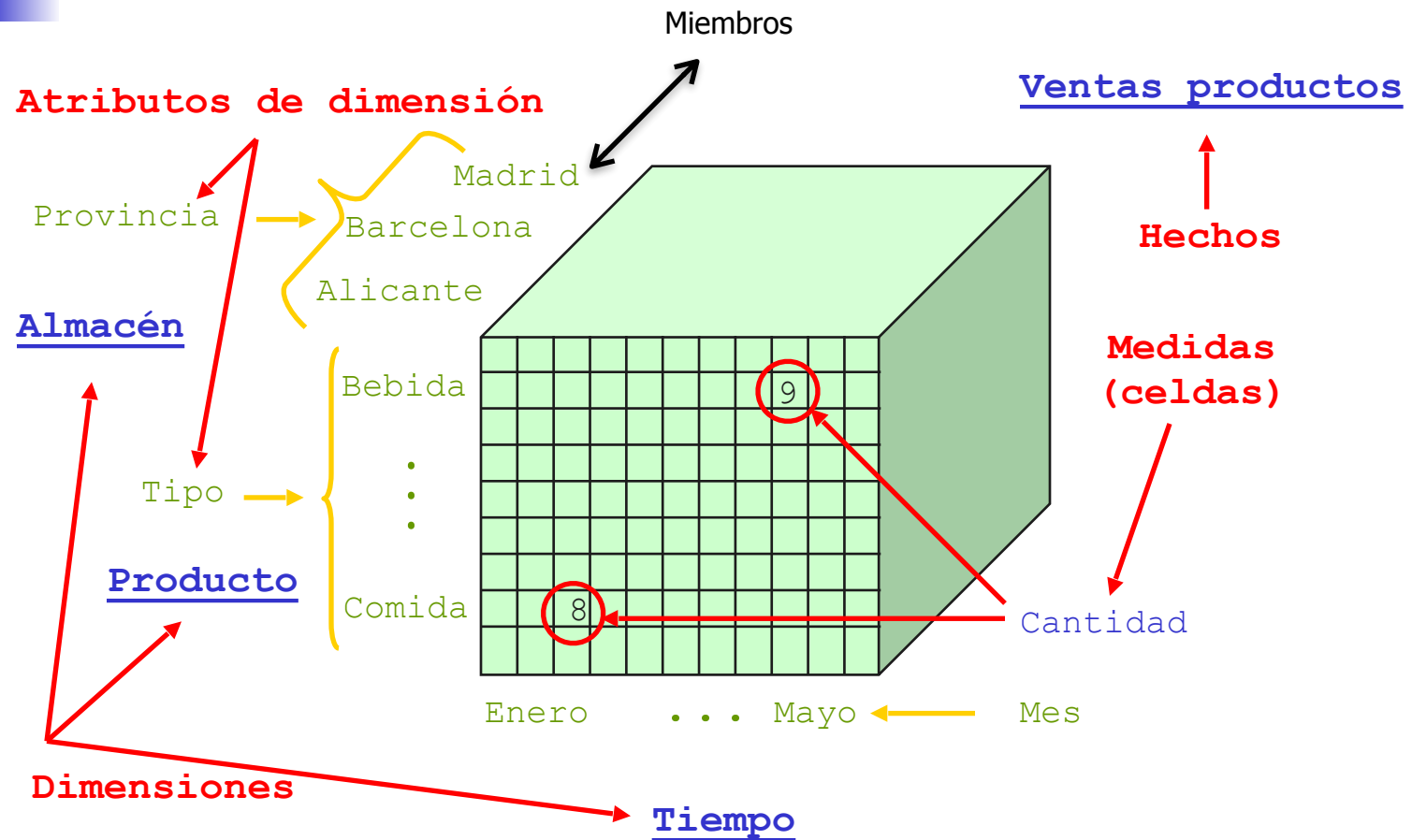
Modelado multidimensional

Parte estructural

- ¿Cómo se representa el modelo MD intuitivamente?
 - Cubos
 - Hipercubos (Cubos sobre cubos)
 - Tablas multidimensionales

Modelado multidimensional

Parte estructural. Cubo



Modelado multidimensional

Parte estructural. Tablas MD.

- Tablas multidimensionales

Ventas			Producto.Grupo = “Supermercado”			
			Comida		Bebida	
			Cong	Fresco	Refresco	Alcohol
Almacén. comunidad = “Comunidad Valenciana”	Alicante	Albatera	100	200	300	400
		Elche	500	600	700	800
	Valencia	Burjasot	900	1000	1100	1200
		Cullera	1300	1400	1500	1600



Modelado multidimensional

Parte estructural. Dimensiones

- Atributos dimensión → jerarquías clasificación
 - Los niveles de jerarquía serán usados para la agregación de las medidas
 - Ej. Ciudad, comunidad, tipos de productos, etc.

Modelado multidimensional

Parte estructural. Dimensiones. Jerarquías de clasificación

- Caminos
 - Simples → Representación mediante árbol

Simétricas

Todos los niveles de la jerarquía son obligatorios y deben estar instanciados.

Asimétricas

Los niveles inferiores de la jerarquía no son obligatorios y pueden no estar instanciados.

Generalizadas

Camino de navegación establecido en función de una categorización de instancias.

- Múltiples → Representación mediante grafo

Modelado multidimensional

Parte estructural. Dimensiones. Jerarquías de clasificación simples.

Todas

Todas instancias

Arbol

Comunidad

C. Valenciana

...

Castilla La Mancha

Provincia

Alicante

...

Valencia

Ciudad Real

...

Albacete

Ciudad

Alicante

...

Villarrobledo

...

Tomelloso

Almacén

Pepito S.L.

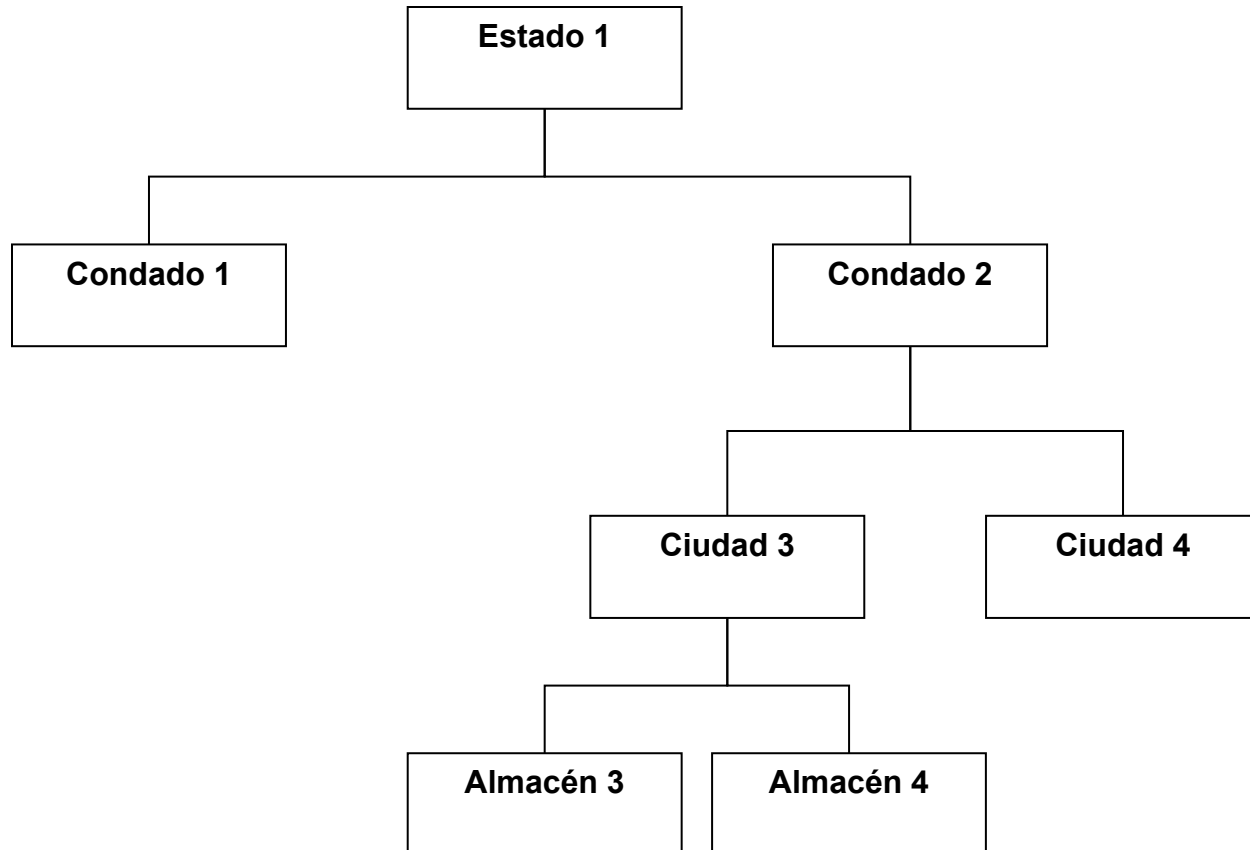
...

Manolo S.L.

Modelado multidimensional

Parte estructural. Dimensiones. Jerarquías de clasificación simples.

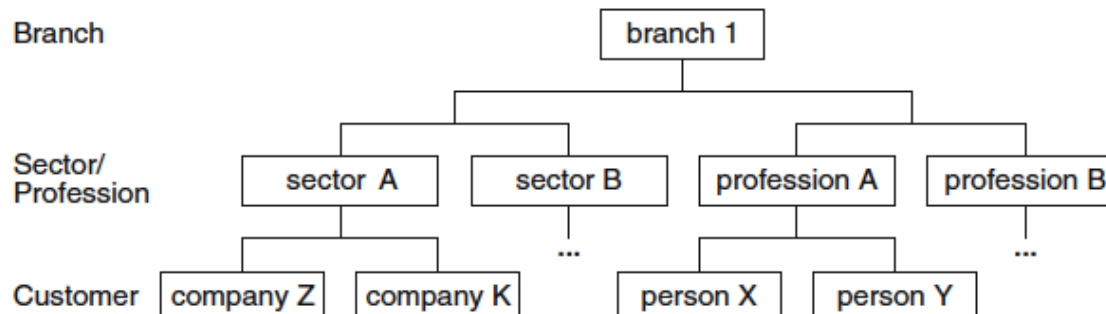
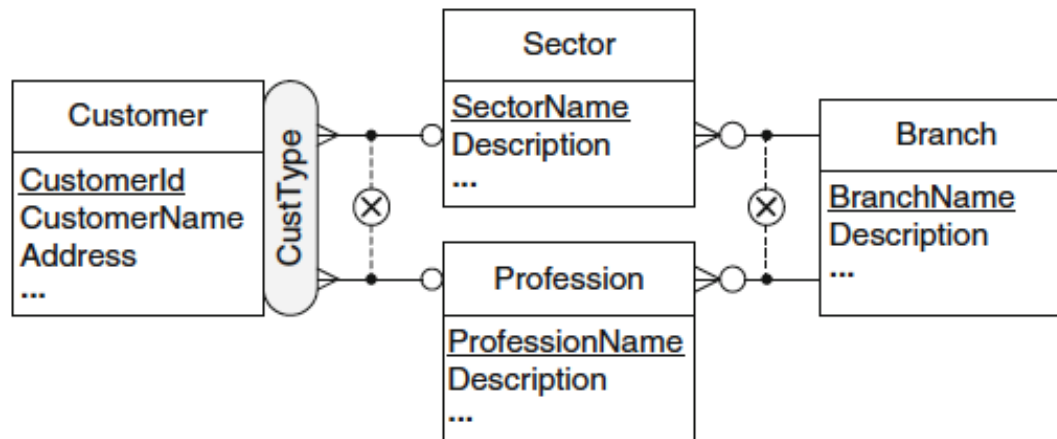
- Ejemplo de Asimétrica



Modelado multidimensional

Parte estructural. Dimensiones. Jerarquías de clasificación simples.

- Ejemplo de Generalizada





Modelado multidimensional

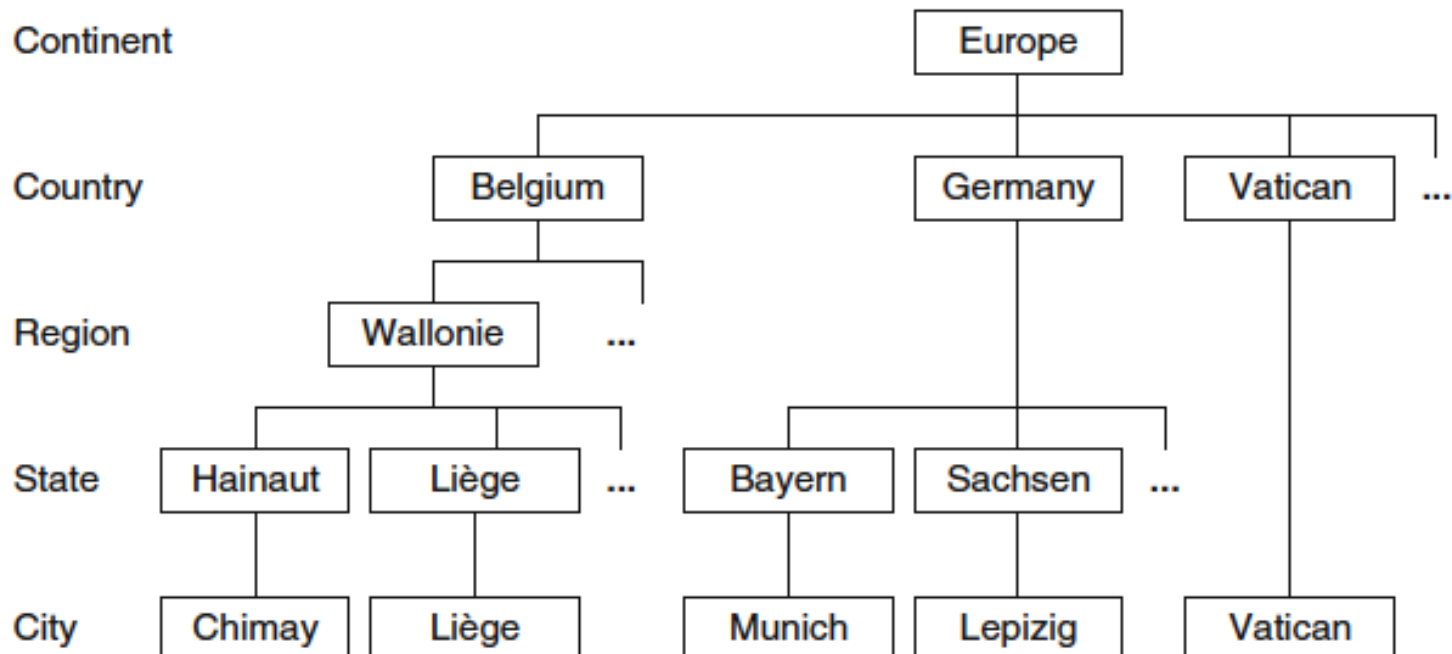
Parte estructural. Dimensiones. Jerarquías de clasificación simples.

- Un caso especial dentro de las jerarquías generalizadas es la jerarquía **desigual**
- Las rutas alternativas se obtienen **saltándose** uno o varios niveles intermedios
- En el nivel de instancia, cada miembro hijo tiene solo un miembro padre, aunque la longitud del camino desde las hojas hasta ese nivel padre puede ser **diferente** para diferentes miembros

Modelado multidimensional

Parte estructural. Dimensiones. Jerarquías de clasificación simples.

- Ejemplo de jerarquía desigual





Modelado multidimensional

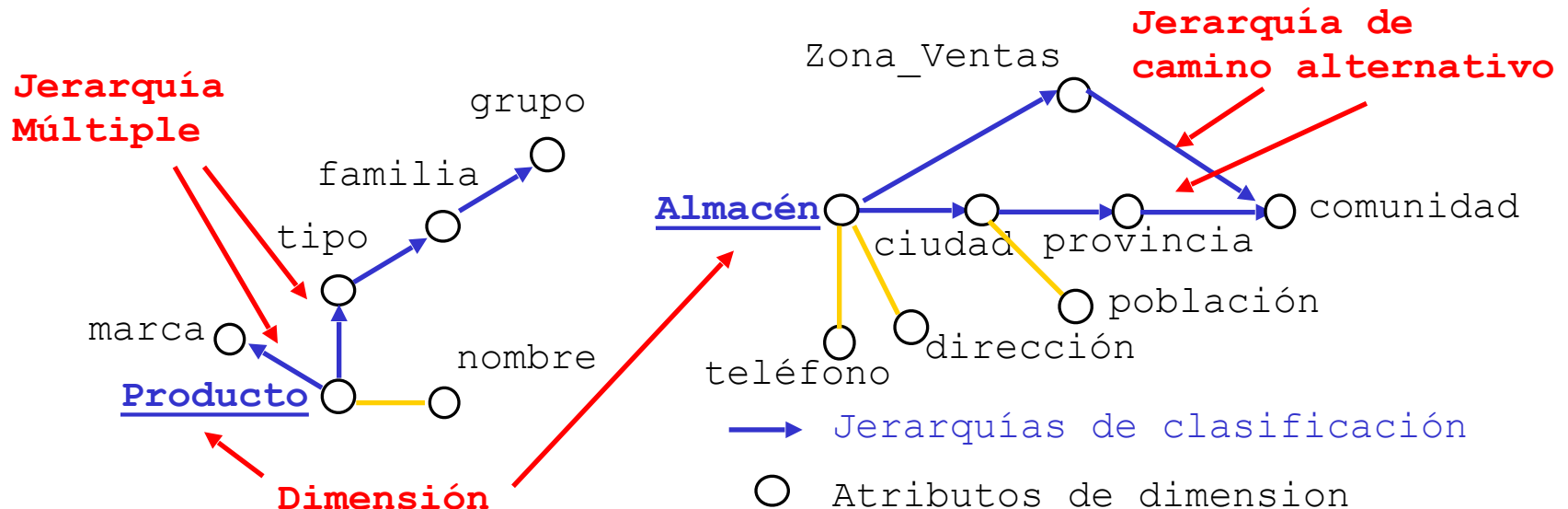
Parte estructural. Dimensiones. Jerarquías de clasificación

- Jerarquías múltiples y de camino alternativo
 - Ej. Ciudad se puede clasificar en comunidad y,
 - Ciudad también se puede clasificar en zona ventas
- Jerarquías paralelas → Más de una jerarquía definida para la misma dimensión

Modelado multidimensional

Parte estructural. Dimensiones. Jerarquías de clasificación

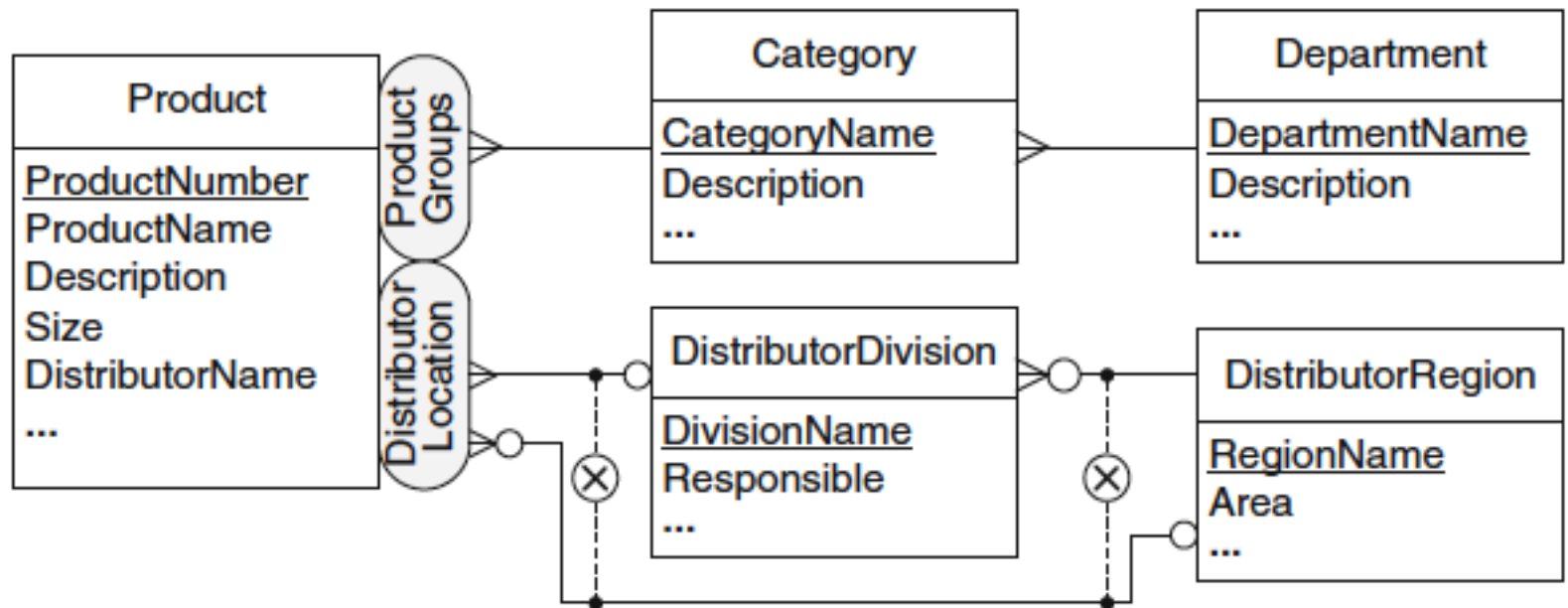
- Multiples → Normalmente se representan mediante Grafos Acíclicos dirigidos (G.A.D.)



Modelado multidimensional

Parte estructural. Dimensiones. Jerarquías de clasificación

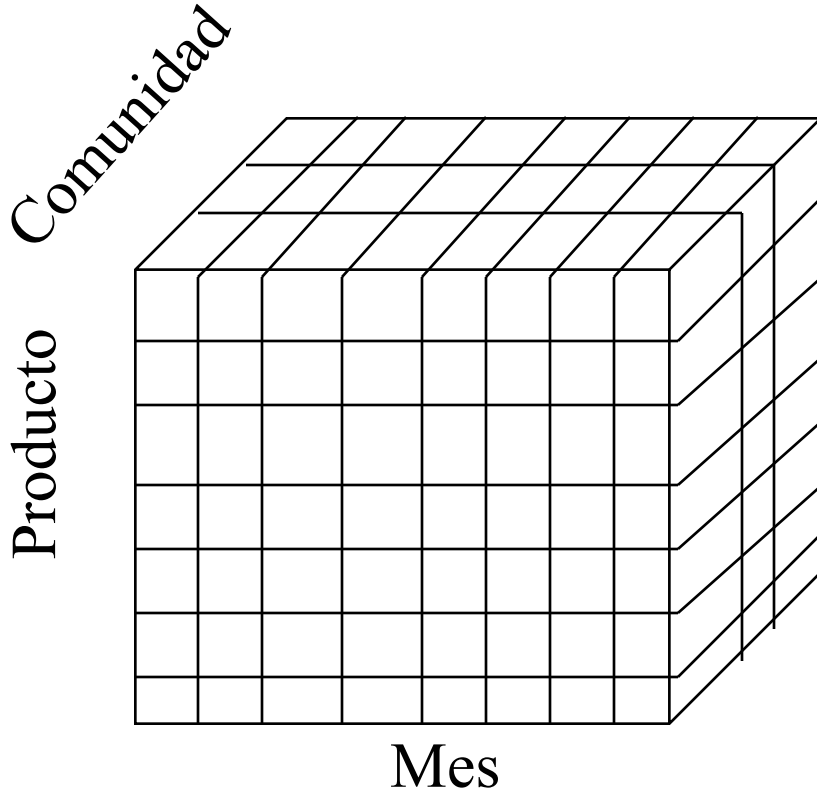
- Jerarquías paralelas independientes



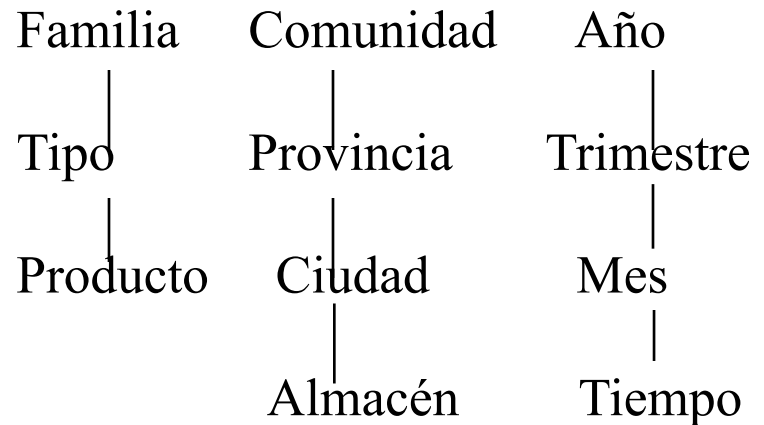
Modelado multidimensional

Parte estructural. Dimensiones. Jerarquías de clasificación y cubos

- Si utilizamos un cierto nivel de agregación
 - Ventas es una función del producto, mes y región

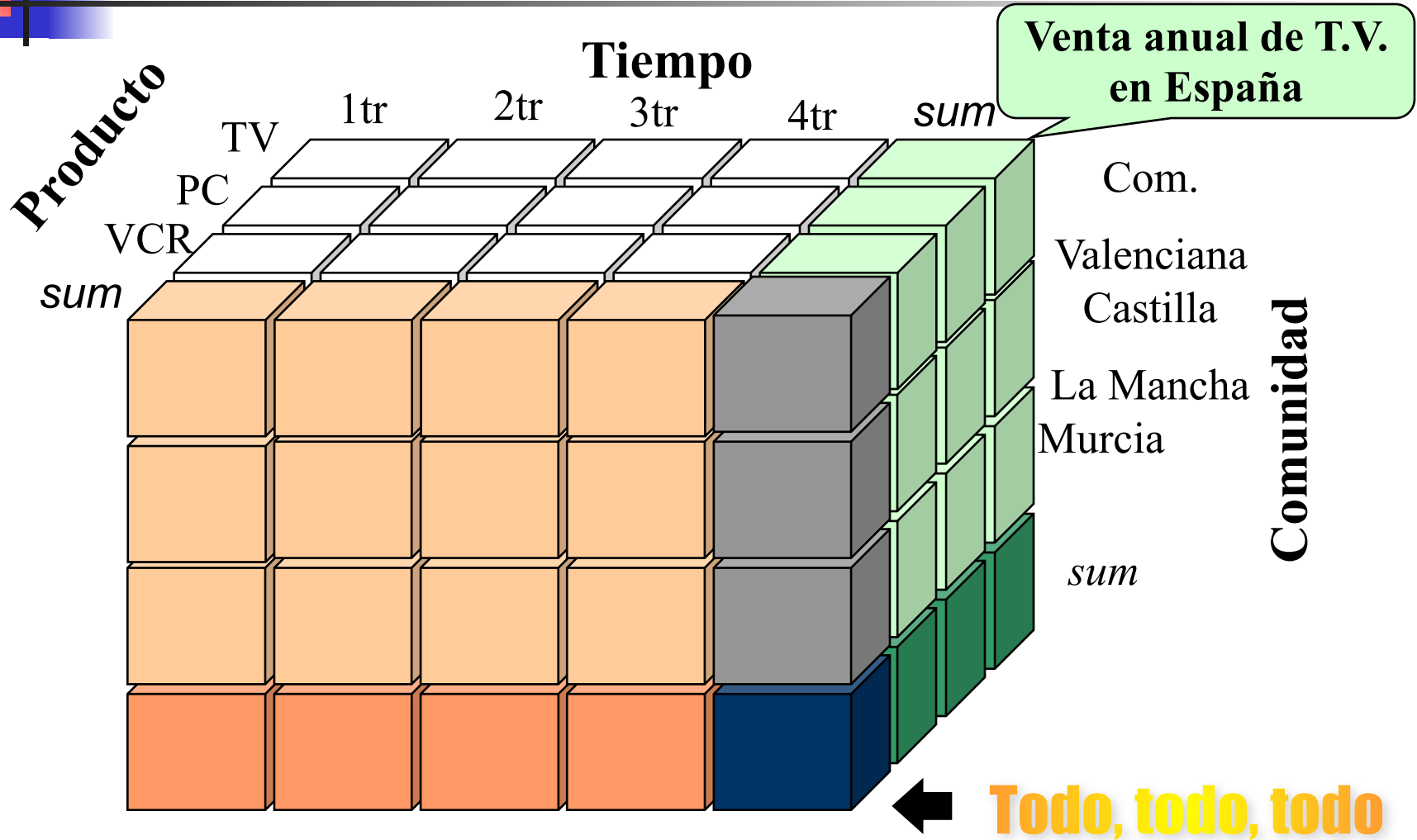


Dimensiones: Producto, Almacén, Tiempo
Caminos de jerarquía por los que agregar



Modelado multidimensional

Parte estructural. Un ejemplo de cubo de datos





Modelado multidimensional

Parte estructural. Hechos

- Atributos de hecho o medidas
 - Atómicos
 - Ej. Cantidad vendida, precio, etc.
 - Derivados
 - Utilizan una fórmula para calcularlos
 - Ej. $\text{Precio_total} = \text{precio} * \text{cantidad_vendida}$



Modelado multidimensional

Parte estructural. Hechos

- **Aditividad**
 - SUM, AVG, ... agregar los valores de medidas a lo largo de las jerarquías de clasificación
 - Es aditiva → SUM sobre todas las dimensiones
 - Semi-aditiva → SUM sólo sobre algunas dimensiones
 - No aditiva → SUM sobre ninguna dimensión
- No aditiva → otros operadores pueden aplicarse (ej. AVG, MIN, etc.)



Modelado multidimensional

- En aplicaciones OLTP...
 - Modelado conceptual → Entidad-Relación (EER)
 - ¿ Podría reflejar la multidimensionalidad de los datos ?
 - Hechos,...
 - Dimensiones,...
 - ¿ Podría ser interrogado por un analista de la información ?

Modelado multidimensional

Parte dinámica.

- BD Multidimensionales parecen más naturales y
- ⇒ Queries son también más naturales
- Ejemplo: llamadas de tfno. por producto y región
 - En un cubo:
 - *¿Cuál sería la consulta que genera el siguiente table/cubo?*

Producto/Comun.	C. Valenciana	Cast. Mancha
Fax	44	28
E-mail	27	51
Mobiles	46	11

Modelado multidimensional

Parte dinámica.

- BD Multidimensionales parecen más naturales y
⇒ Queries son también más naturales
- Ejemplo: llamadas de tfno. por producto y región
 - En un cubo:
 - ROLLUP*(calls, producto->producto, localizacion->comunidad, SUM(numLlamadas))

Producto/Comun.	C. Valenciana	Cast. Mancha
Fax	44	28
E-mail	27	51
Mobiles	46	11

- En tablas relacionales:

Select producto, comunidad, sum(llamadas)
from Llamadas, Comunidad
where llamadas.comunidad= Comunidad.comunidad
Group by producto, comunidad
Order by producto, comunidad;

Product	Region	Sum(calls)
Fax	Vaud	44
Fax	Valais	28
Mobile	Vaud	27
Mobile	Valais	51
Standard	Vaud	46
Standard	Valais	11



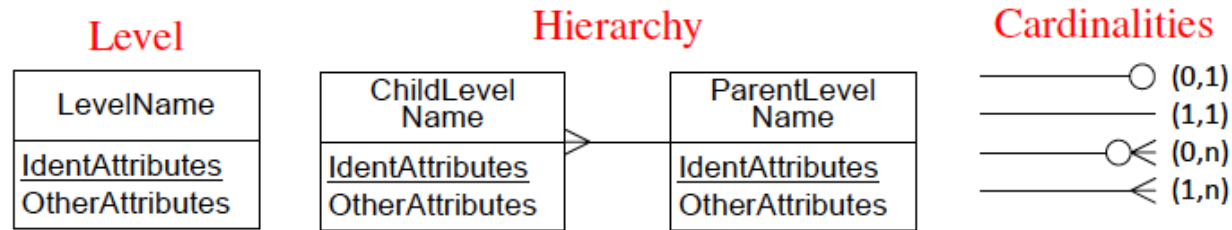
Diseño conceptual de almacenes de datos

- Es el primer paso para el diseño de un almacén de datos
- Se parte de la documentación relacionada con la base de datos integrada (ER, relacional, ...) y consta de:

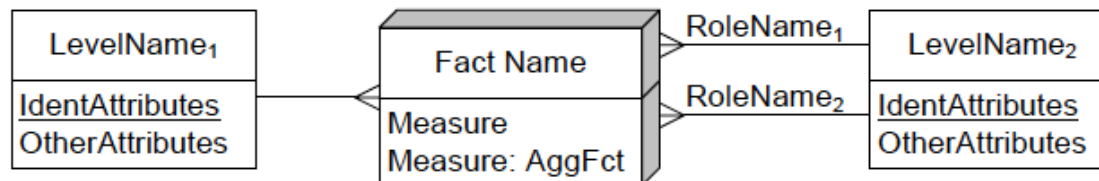
Hecho:

- Dimensiones
- Medidas
- Jerarquías

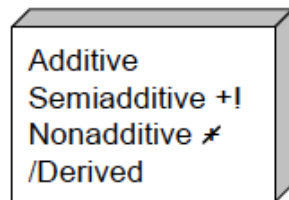
Representación Modelo multidimensional



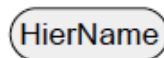
Fact with measures and associated levels



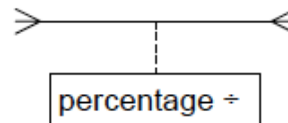
Types of measures



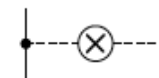
Analysis criterion



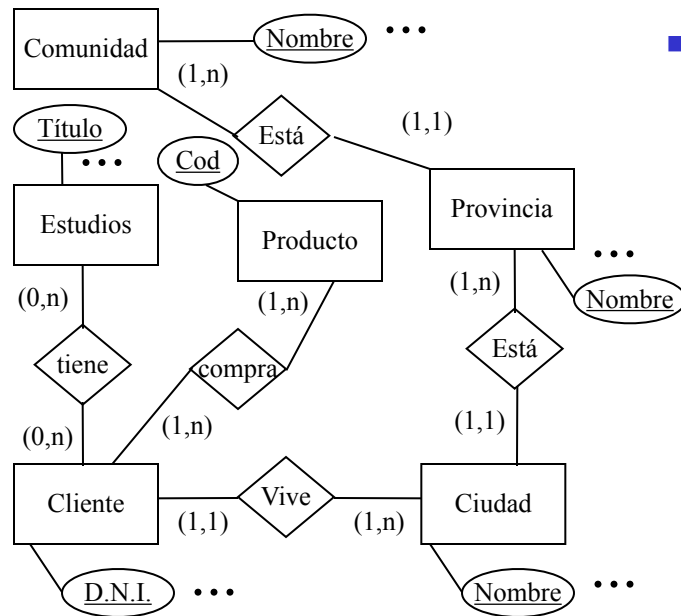
Distributing factor



Exclusive relationships



Modelado multidimensional

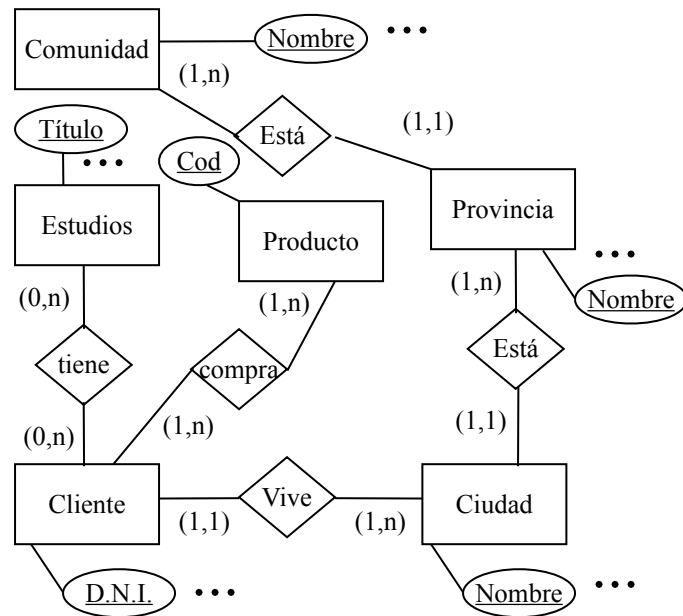


- ¿Hechos?
- ¿Dimensiones?
- ¿Jerarquías?

Transformar

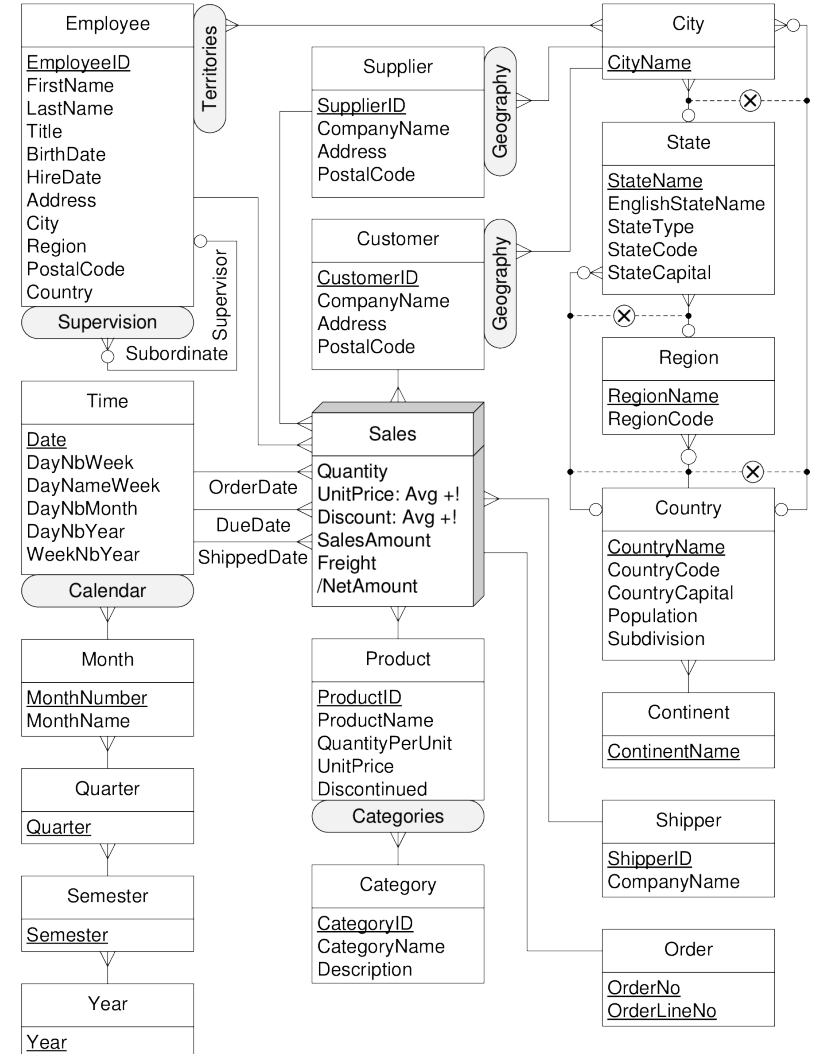
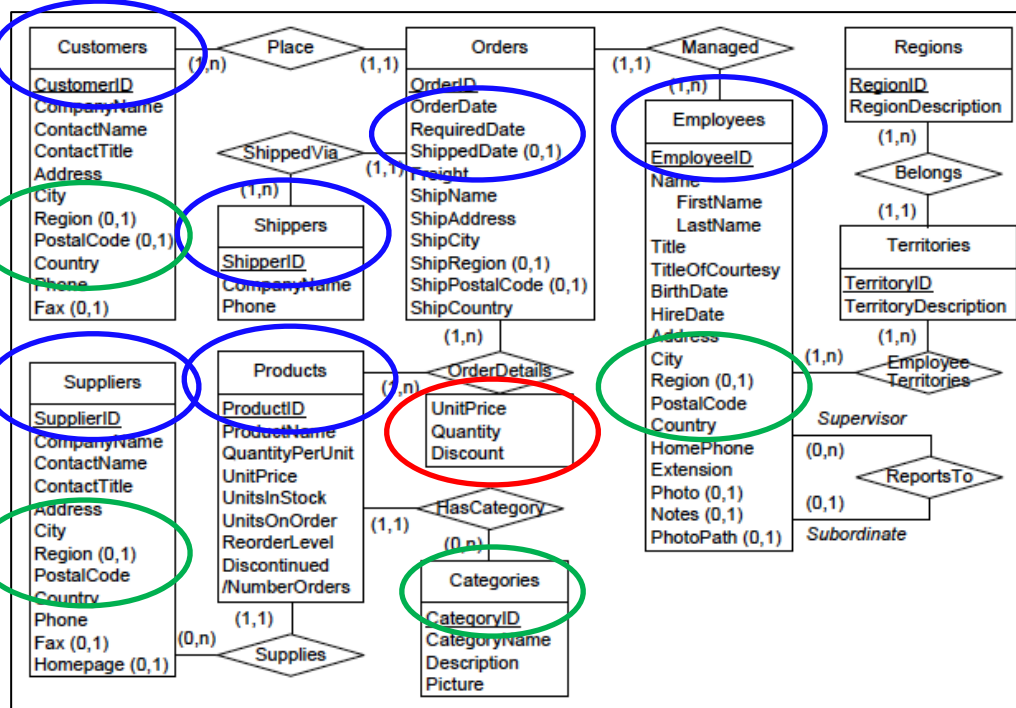
- ¿Podría un analista interrogarlo?

Diseño conceptual de almacenes de datos



Ejemplo 1.
Determina el
diseño conceptual
de un almacén a
partir de
diagrama ER

Ejemplo 2. DB Northwind





Ejercicio 2

- Una compañía de seguros requiere el diseño de almacenes de datos para el análisis de los accidentes de sus clientes. En particular, la empresa requiere evaluar el tipo de accidentes relacionados con los clientes y el tipo de pólizas.
- Objetivo: Evaluar el historial de accidentes y relacionarlas con las pólizas y los clientes
- Diseñar el modelo entidad relación, añadiendo los atributos y relaciones que consideres
- Diseñar el almacén de datos para modelar y hacer preguntas sobre los accidentes, en referencia al número y al gasto que le supone a la compañía. Elije hechos, medidas y dimensiones.



Tarea 3. Diseño conceptual y consultas OLAP

- **PRIMERA PARTE**

En esta tarea vamos a realizar el diseño conceptual del almacén de datos que se indica en el ejercicios 2 de las transparencias de este tema. Se trata de generar un DFM (Dimensional Fact Model) para representar los hechos, dimensiones y jerarquías de cada almacén de datos.

- Los objetivos son:

- Primeramente, hacer una búsqueda de herramientas de modelado conceptual de almacenes de datos. Resumir las que encontréis y comprobar si es posible usar alguna.
- Diseñar el DFM del ejercicio 2. (Puede hacerse a papel o con alguna de las herramientas encontradas)

- **SEGUNDA PARTE**

Realizar al menos tres consultas en OLAP algebraico sobre cada uno de los ejercicios.



Tarea 3. Diseño conceptual y consultas OLAP

- La tarea se realizará por parejas.
- Se pide realizar una memoria que incluirá la información
 - Resumen de las herramientas para la construcción de DFMs encontradas, mostrando algún ejemplo de uso si es posible.
 - El DFM del ejercicio propuestos, con comentarios sobre las consideraciones tomadas en cada uno de ellos.
 - Detallar las consultas realizadas y las operaciones en OLAP algebraico correspondientes.
- La memoria tendrá el nombre de los miembros del grupo (Sólo la sube un miembro)
- Se deberá subir la memoria en pdf.



Diseño conceptual de almacenes de datos

Bibliografía

- Giovinnazo (2000). Object-Oriented Data Warehouse Design: Building a star schema
- Inmon (2002). Building the Data Warehouse (3ª ed.)
- Kimball (2002). The Data Warehouse Toolkit (3ª ed.)
- Thomsen (2000). OLAP solutions: Building Multidimensional Information Systems