

DISEÑO CONCEPTUAL DE ALMACENES DE DATOS

Tema 2

Profesores:

Juan C. Trujillo
Alejandro Reina Reina
LUCENTIA Research Group



Universitat d'Alacant
Universidad de Alicante

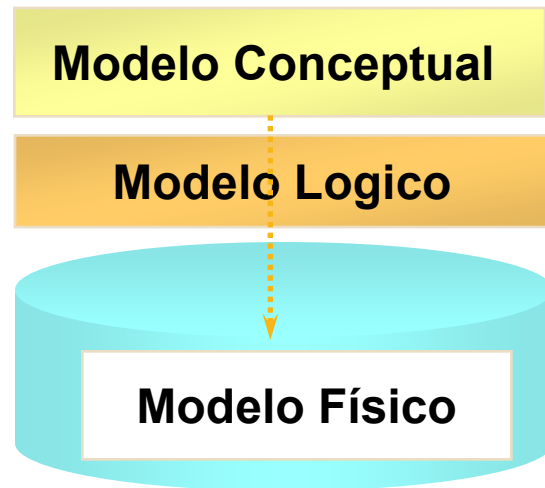


Departamento de
Lenguajes y Sistemas
Informáticos

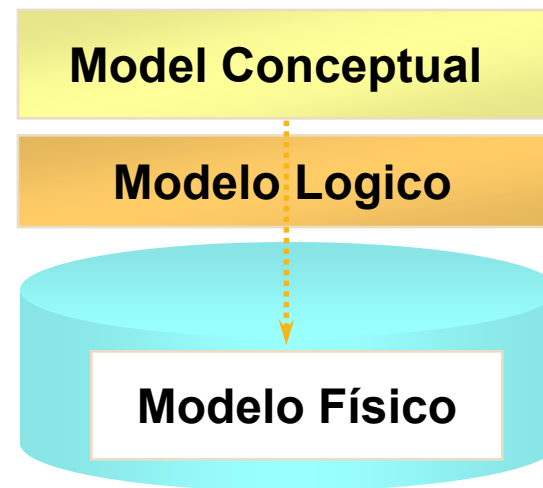
Modelos utilizados en BD

2

- ❑ Modelo Conceptual
- ❑ Modelo Logico
- ❑ Modelo Físico



**Sistemas
Operacionales**



Almacenes de Datos

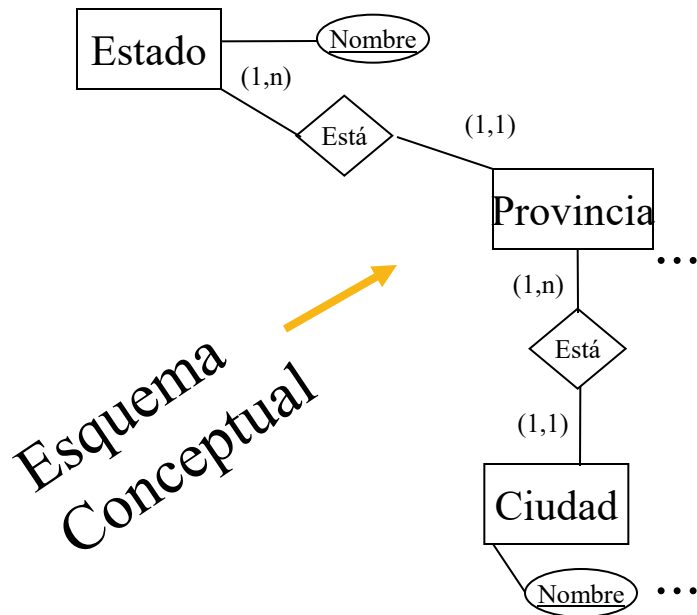
Modelos utilizados en BD

Sistemas operacionales

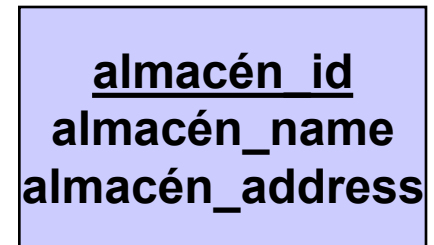
3

□ Normalización

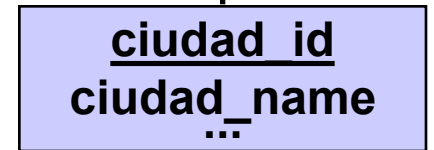
▣ Sist. Transaccionales



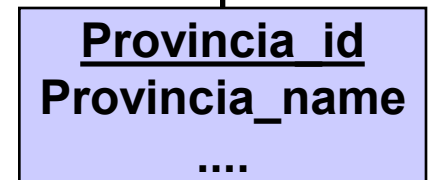
Almacén



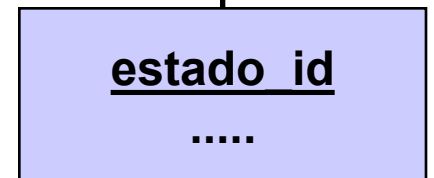
Ciudad



Provincia



Estado



Diseño conceptual de DW

4

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

Modelado multidimensional

5

- 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

6

□ 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

7

- Hechos representan normalmente relaciones *muchos a muchos* con todas las dimensiones y, *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

8

- 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

Modelado multidimensional

Parte estructural

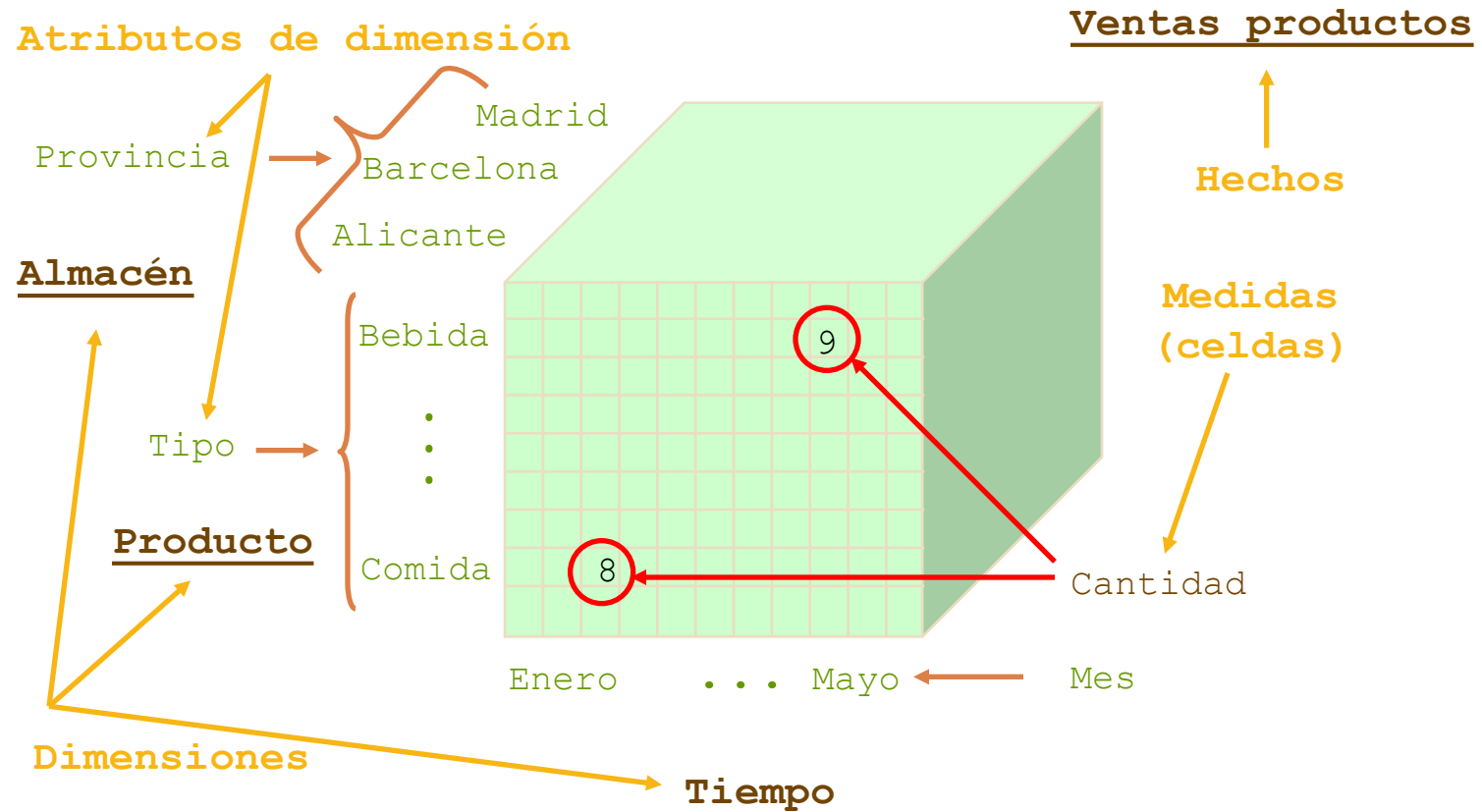
9

- ¿Cómo se representa el modelo MD intuitivamente?
 - ▣ Cubos
 - ▣ Hipercubos (Cubos sobre cubos)
 - ▣ Tablas multidimensionales tipo hoja de cálculo, etc.

Modelado multidimensional

Parte estructural. Cubo

10



Modelado multidimensional

Parte estructural. Tablas MD.

11

■ Tablas multidimensionales

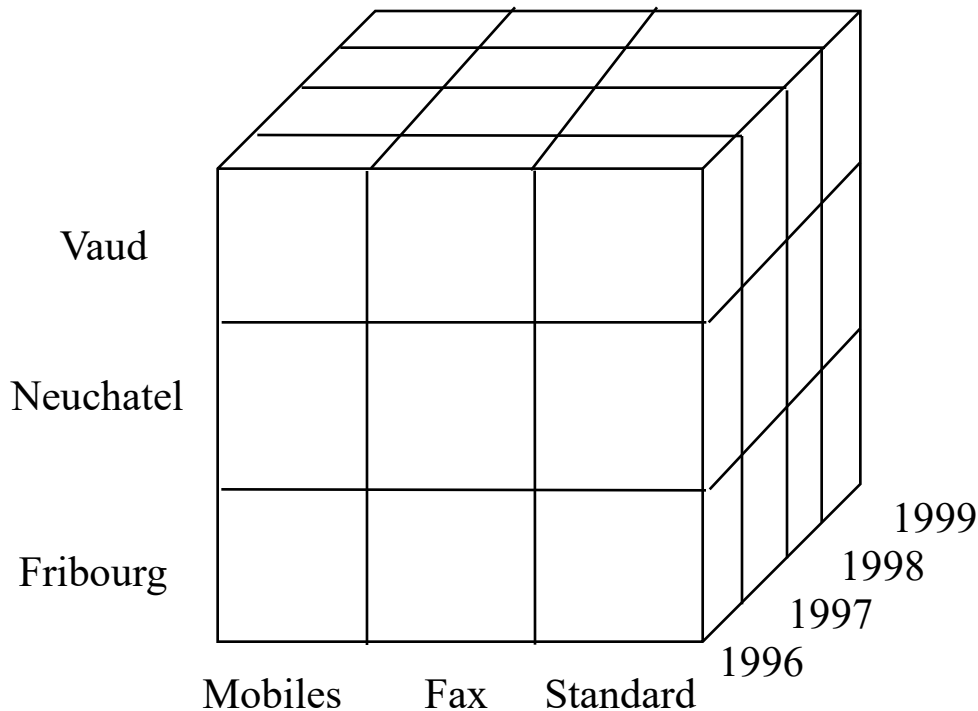
Ventas			Producto.Grupo = “Supermercado”			
			Comida		Bebida	
			Cong	Fresco	Refresco	Alcohol
Almacén. comunidad = “Comunidad Valenciana”	Alicante	Albatera	100	200	300	400
		Elche	500	600	700	800
	Valencia	Burjasot	900	1000	1100	1200
		Cullera	1300	1400	1500	1600

Modelado multidimensional

Parte estructural. Cubo vs. Tabla relacional.

12

- Para recuperar los datos necesarios:
 - ▣ En una SGBDR: $4 \times 3 \times 3 = 36$ tuplas (filas)
 - ▣ En una BD Multidimensional: $4 + 3 + 3 = 10$ valores en los ejes



Region	Producto	Año	Llam	tupla
Vaud	Fax	1997	12	1
Vaud	Mobiles	1998	23	2
Vaud	Standard	1999	22	3
Fribourg	Fax	1997	34	4
Fribourg	Mobiles	1998	45	5
Fribourg	Standard	1999	48	6
Neuchatel	Fax	1986	55	7
Neuchatel	Mobiles	1987	66	8
...
Neuchatel	Standard	1998	55	35
Neuchatel	Mobiles	1999	66	36

Modelado multidimensional

Parte estructural. Dimensiones

13

- Puede haber alto grado de categorización
 - ▣ Atributos en función de instancias
 - Ej. Volumen y porcentaje de alcohol sólo para bebidas
 - Ej. Tiempo y modo preparación sólo para comidas
- 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

14

- ▣ Las instancias de niveles \equiv miembros

▣ Clasificación

▣ Cardinalidad

- Por defecto \rightarrow estrictas ($1-m$)

- Una instancia sólo se relaciona con una instancia del nivel superior de jerarquía

- Ej. Un almacén está ubicado en una sola ciudad

- Sin embargo algunas pueden ser **no estrictas** ($m-m$)

- Ej. Un almacén pertenece a más de una zona de ventas

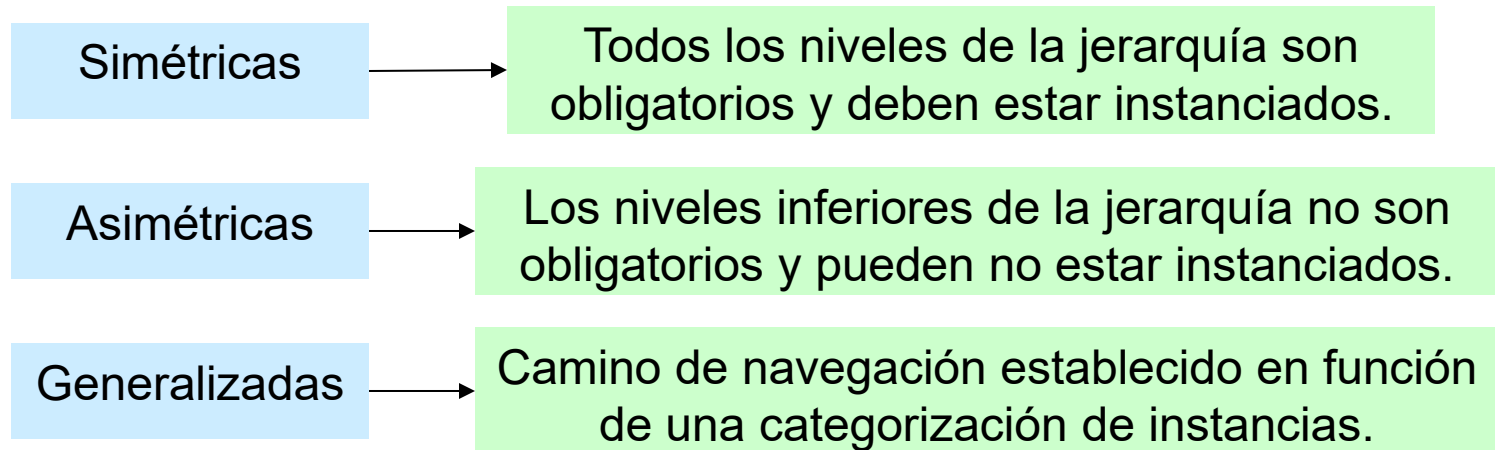
Modelado multidimensional

Parte estructural. Dimensiones. Jerarquías de clasificación

15

□ Caminos

■ Simples → Representación mediante árbol

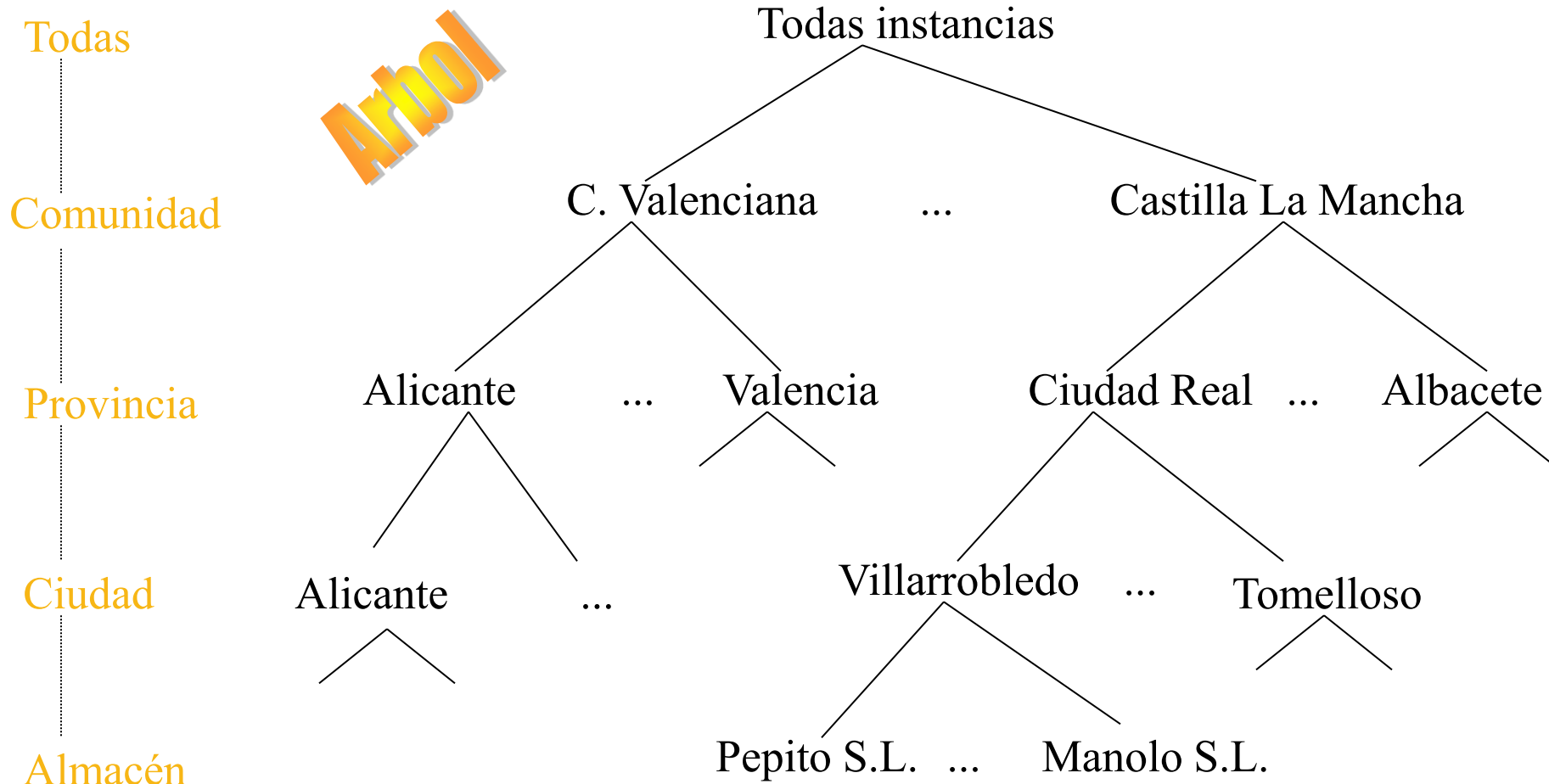


■ Múltiples → Representación mediante grafo

Modelado multidimensional

Parte estructural. Dimensiones. Jerarquías de clasificación simples.

16

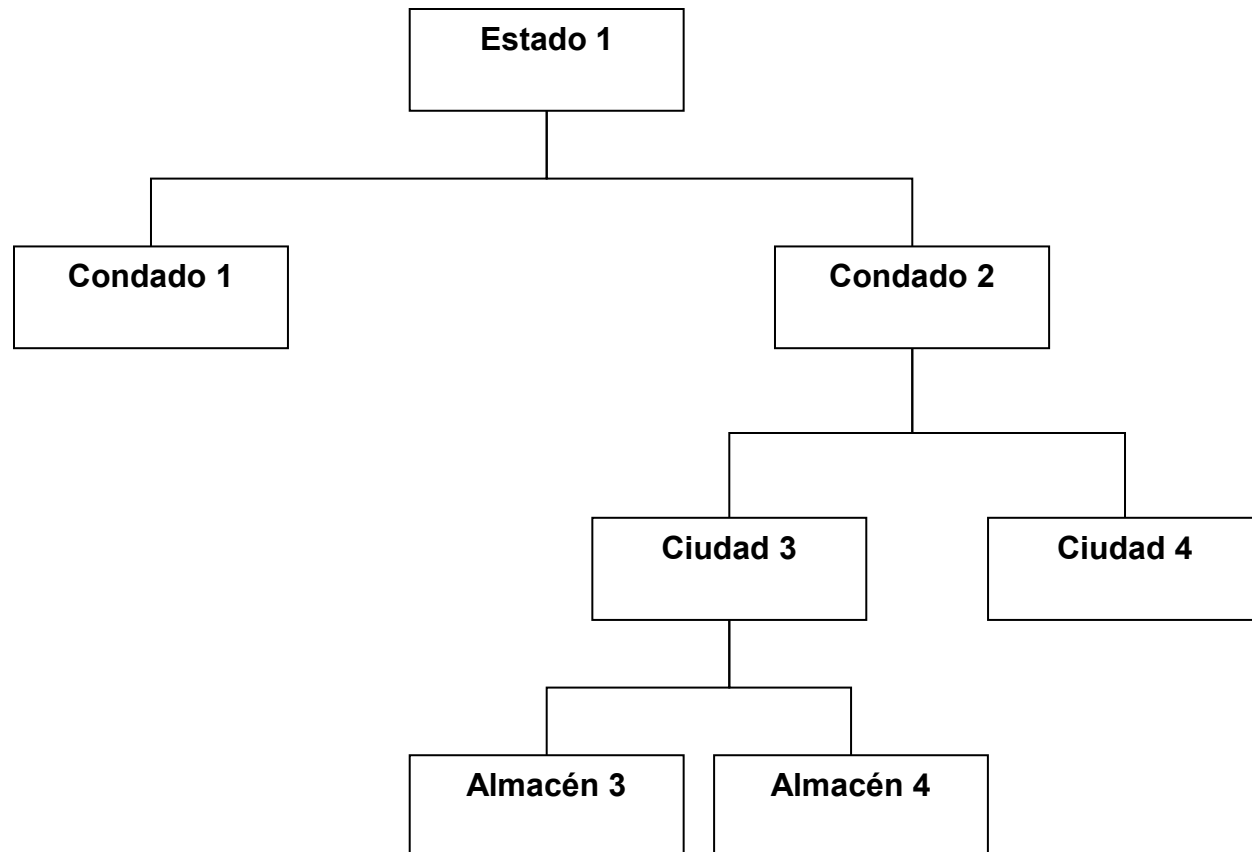


Modelado multidimensional

Parte estructural. Dimensiones. Jerarquías de clasificación simples.

17

■ Ejemplo de Asimétrica



Modelado multidimensional

Parte estructural. Dimensiones. Jerarquías de clasificación

18

▣ Jerarquías múltiples y de camino alternativo

- Rep. → D.A.G.

- 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

- Independiente → Las distintas jerarquías no comparten niveles

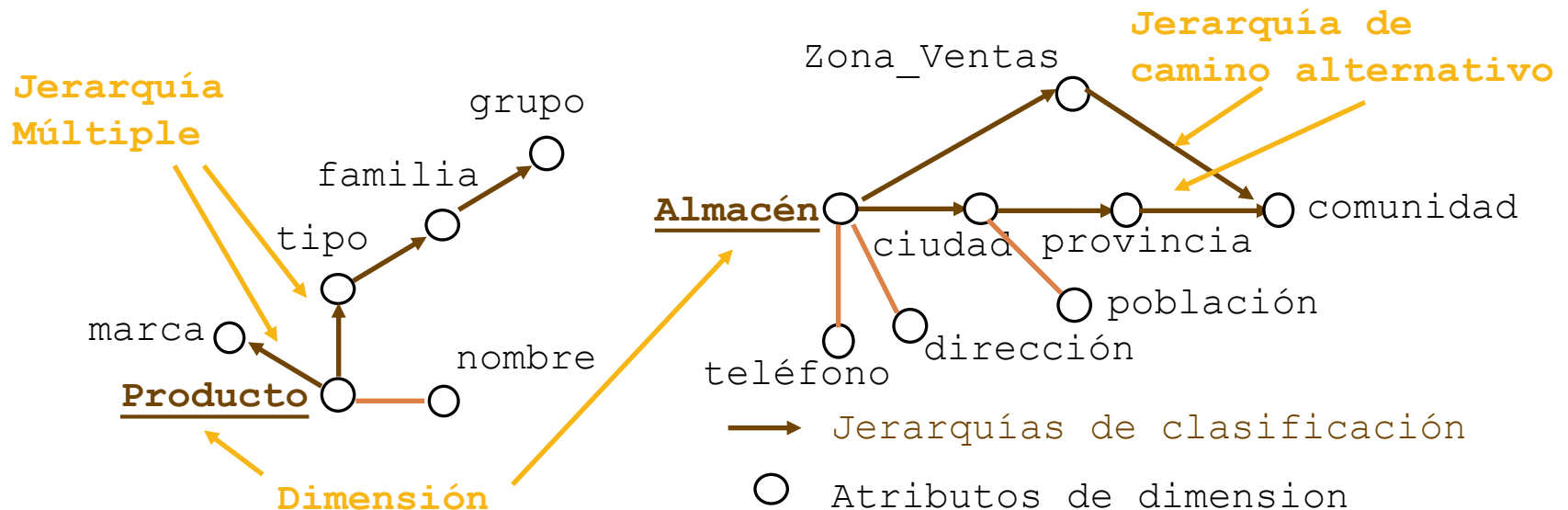
- Dependientes → Las distintas jerarquías comparten algún nivel

Modelado multidimensional

Parte estructural. Dimensiones. Jerarquías de clasificación

19

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



Modelado multidimensional

Parte estructural. Dimensiones. Jerarquías de clasificación

20

▣ Jerarquías completas

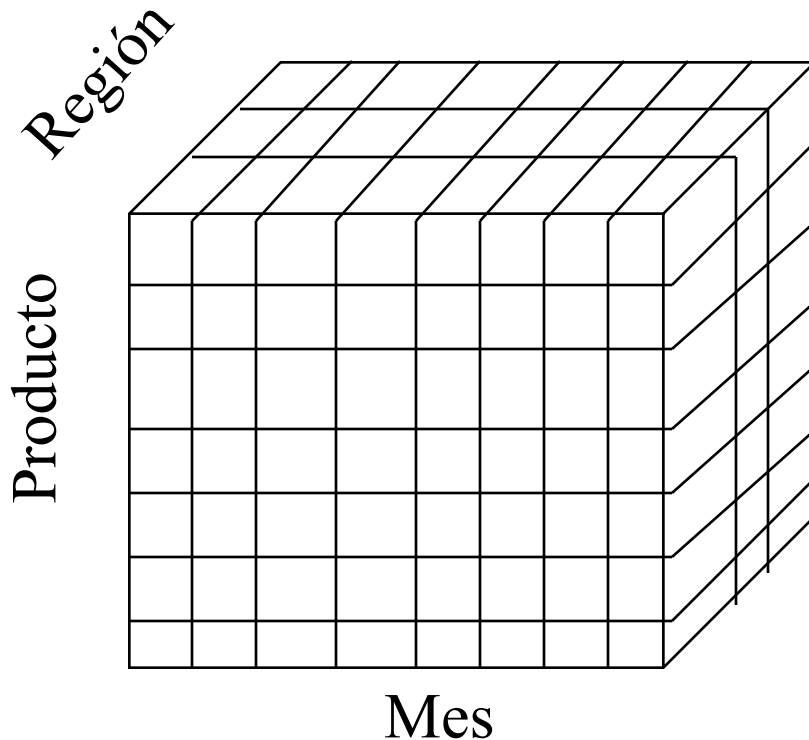
- Además de estrictas, un miembro o instancia de un nivel superior está compuesto únicamente por los del nivel inferior
→ relación fija entre instancias

Modelado multidimensional

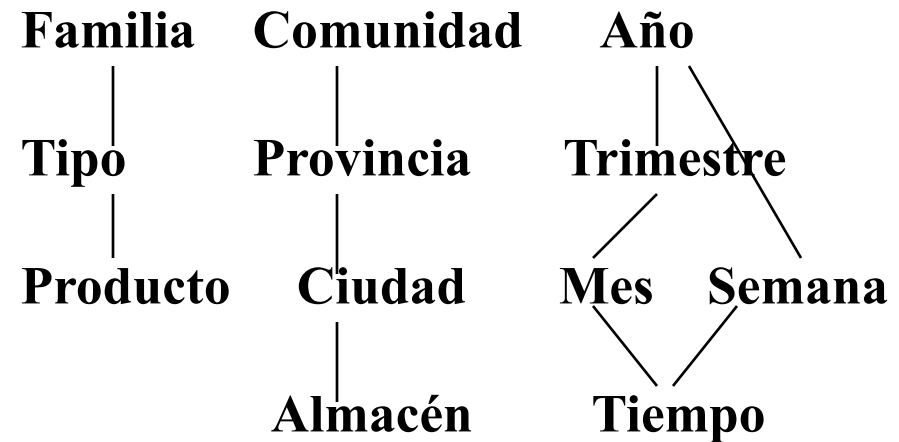
Parte estructural. Dimensiones. Jerarquías de clasificación y cubos

21

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



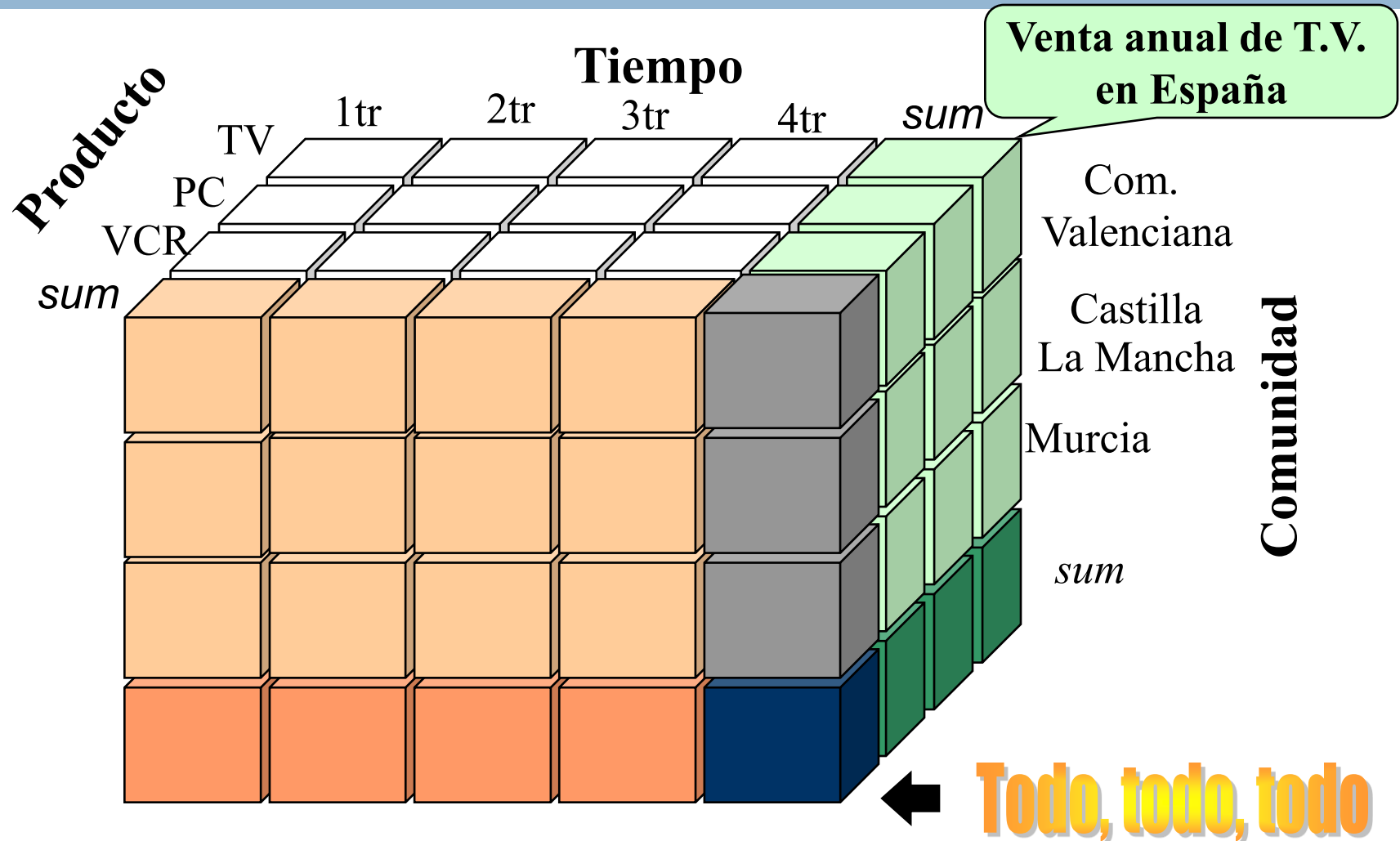
Dimensiones: Producto, Almacén, Tiempo
Caminos de jerarquía por los que agregar



Modelado multidimensional

Parte estructural. Un ejemplo de cubo de datos

22



Modelado multidimensional

Parte estructural. Hechos

23

□ 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

24

□ Aditividad

- Conjunto de operadores de agregación (SUM, AVG, etc.) que se pueden aplicar para agregar los valores de medidas a lo largo de las jerarquías de clasificación (Kimball, 1996)
- Es aditiva → SUM sobre todas las dimensiones
- Semi-aditiva → SUM sólo sobre algunas dimensiones
- No aditiva → SUM sobre ninguna dimensión

Modelado multidimensional

Parte estructural. Hechos

25

- Si no aditiva → otros operadores pueden aplicarse (ej. AVG, MIN, etc.)
 - ▣ Ej. Atributos que miden niveles (ej. Inventarios) no son aditivos sobre la dimensión tiempo
 - Es aditivo sobre la dimensión producto
 - ▣ Las medidas de temperatura no son aditivas
 - ▣ Algunas son semánticamente incorrectas
 - Ej. Atributo número de clientes que cuenta el número de tickets emitidos no es aditiva sobre la dimensión producto

Modelado multidimensional

26

- 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

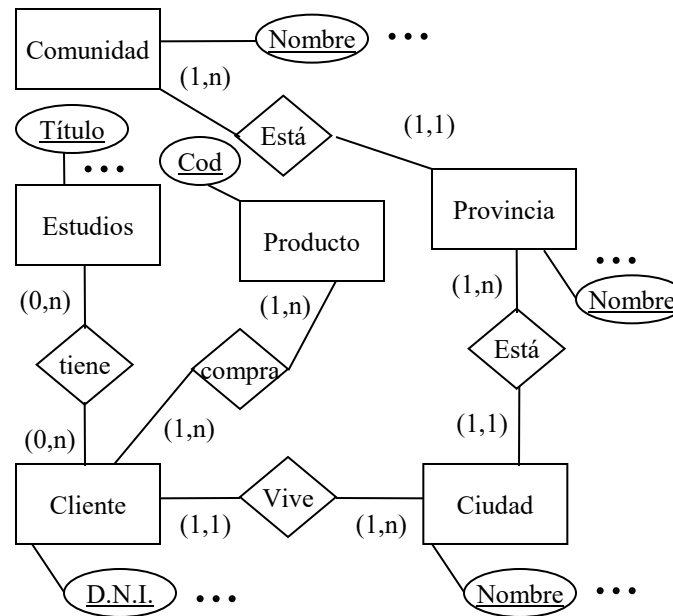
27

□ ¿Hechos?

□ ¿Dimensiones?

▣ ¿Jerarquías?

Transformar



■ ¿Podría un analista interrogarlo?

Modelado multidimensional

Parte dinámica.

28

- BD Multidimensionales parecen más naturales y
- ⇒ Queries son también más naturales
- Ejemplo: llamadas de tfno. por producto y región
 - En un cubo: *print total.(calls keep product, region)*

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

Modelado multidimensional

Parte dinámica.

29

- Definición de requerimientos iniciales de usuario
 - ▣ Están basados en jerarquías definidas en Dimensiones

Cantidad vendida de productos comestibles agrupados por su familia y tipo, vendidos en la comunidad valenciana y, agrupados por la provincia y ciudad donde se vendieron

Modelado multidimensional

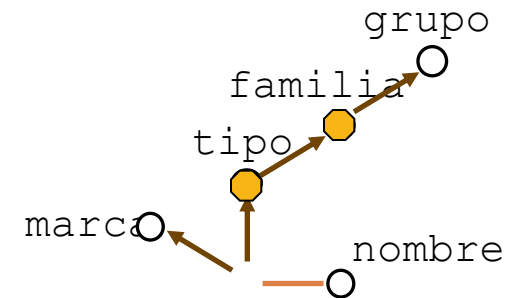
Parte dinámica.

30

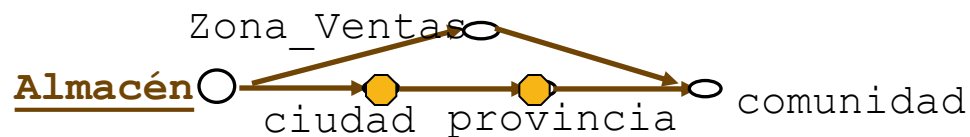
Operaciones de consulta (OLAP)

Roll-up

- Agregar valores de medidas a lo largo de jerarquías de clasificación



Ventas'		Producto.Grupo = "alimentación"	
		Comida	Bebida
Almacén. comunidad = "Comunidad Valenciana"	Alicante	1400	2200
	Valencia	4600	5400



Modelado multidimensional

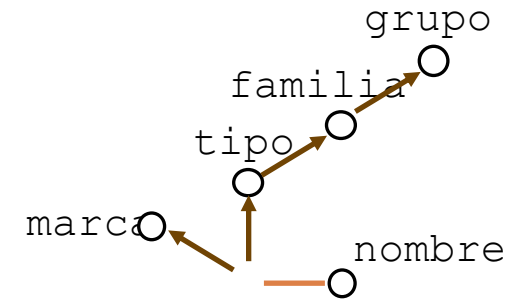
Parte dinámica.

31

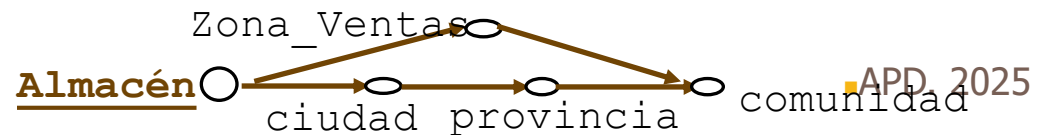
Operaciones de consulta (OLAP)

Drill-down

- Desagregar valores de medidas a lo largo de jerarquías de clasificación



Ventas			Producto.Grupo = "alimentación"			
			Comida		Bebida	
			Cong.	Fresca	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



APD, 2025

Modelado multidimensional

Parte dinámica.

32

□ Operaciones de consulta (OLAP)

▣ Drill-across

■ Consultar medidas de varios hechos en el mismo cubo

- Ej. Que en la tabla MD analizáramos el ratio de ventas respecto de compras.

- $1000 / 400$

Modelado multidimensional

Parte dinámica.

33

□ Operaciones de consulta (OLAP)

▣ Slice-dice

- Definir restricciones sobre niveles de jerarquías

- Ej. Analizar datos donde el año sea 1999

Ventas'			Producto.Grupo = "Alimentación"	
			Comida	
			Congelada	Fresca
Almacén. Comunidad = "Comunidad Valenciana"	Alicante	Albatera	100	200
		Elche	500	600

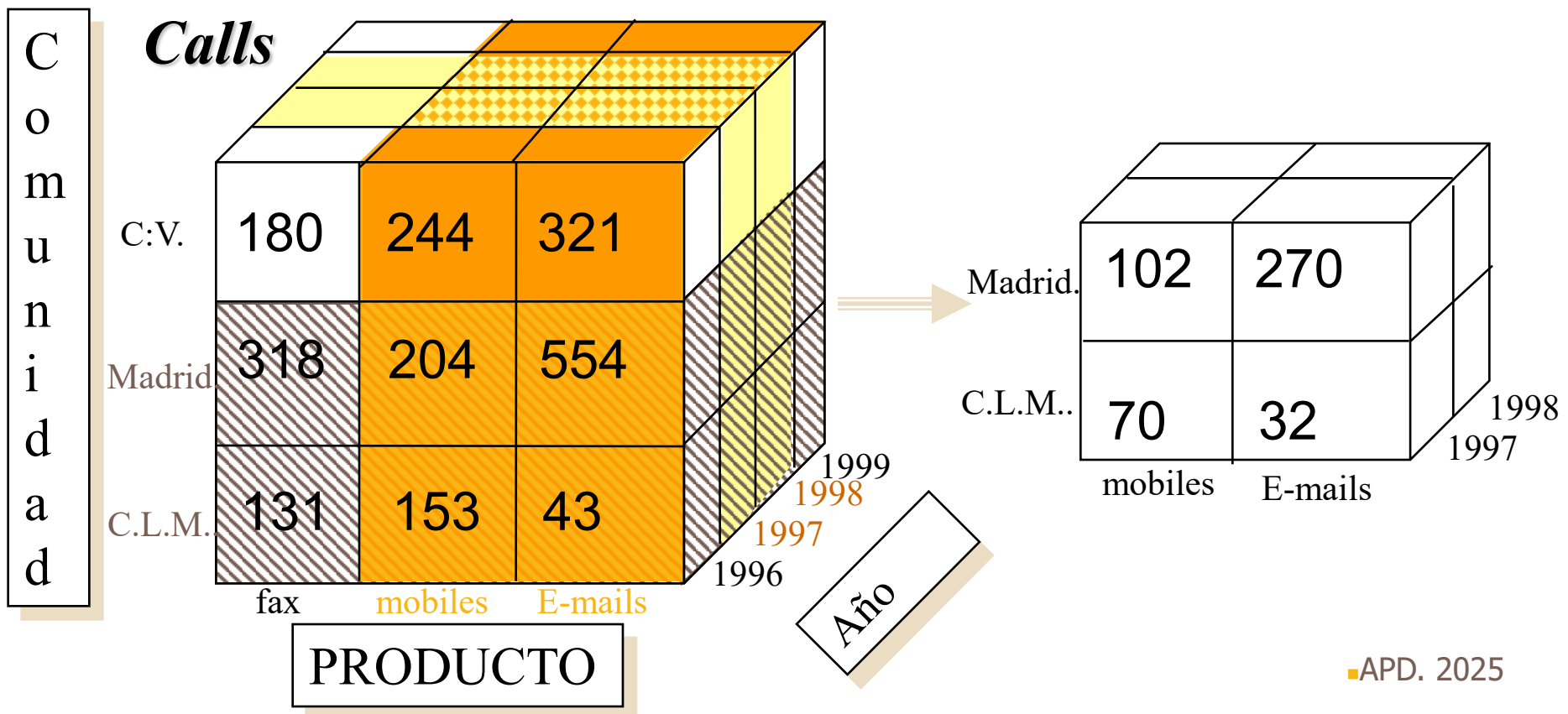
Modelado multidimensional

Parte dinámica.

34

■ Slice-dice (cont.)

■ Otro ejemplo



Modelado multidimensional

Parte dinámica.

35

□ Operaciones de consulta (OLAP)

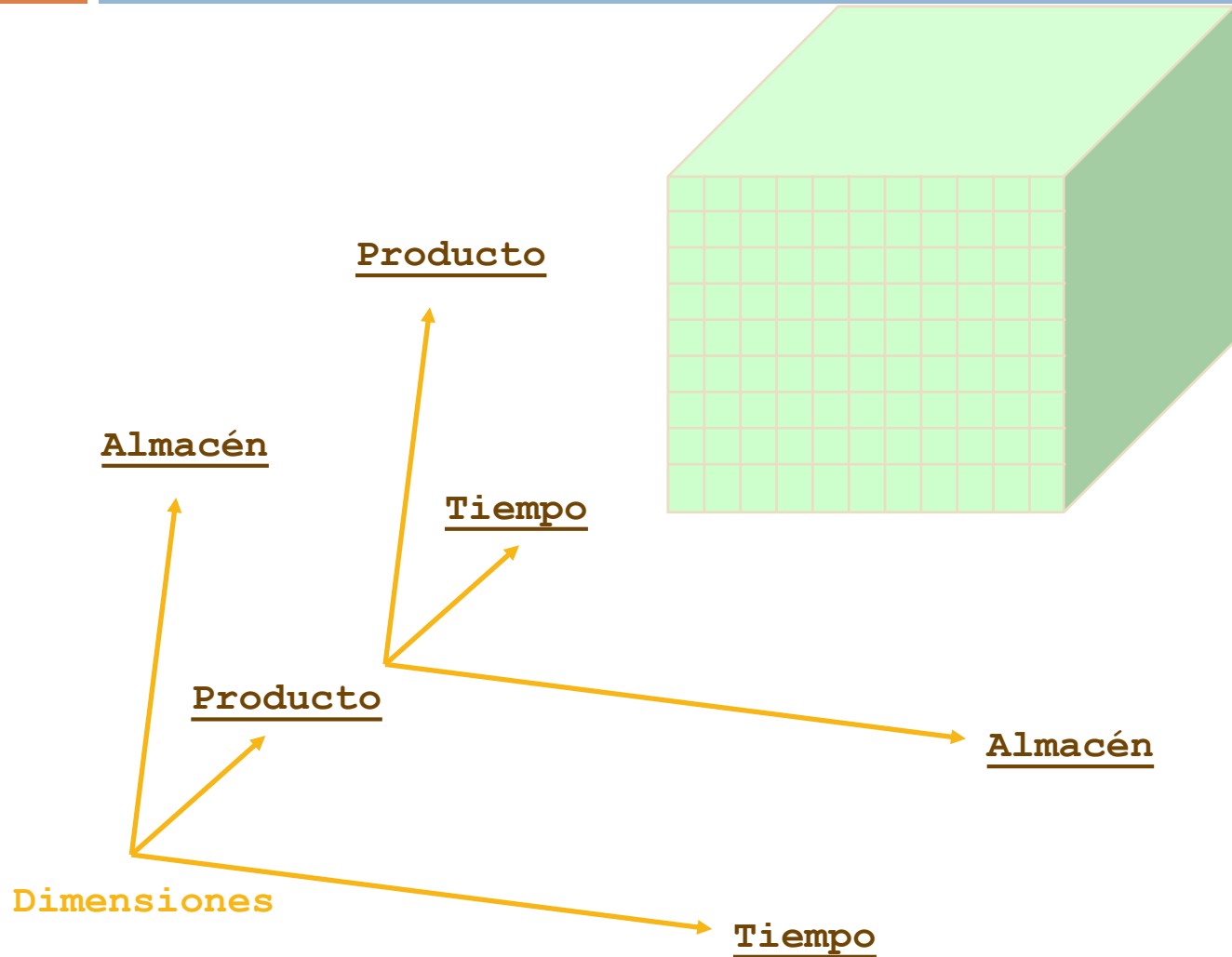
▣ Pivoting

- Reorientar la vista multidimensional de los datos, es decir, cambiar la distribución de filas/columnas
 - Algunos autores consideran también el intercambio de medidas y hechos como pivoting (kimball, 1996) (Inmon, 1996)

Modelado multidimensional

Parte dinámica.

36



Diseño conceptual de almacenes de datos

Bibliografía

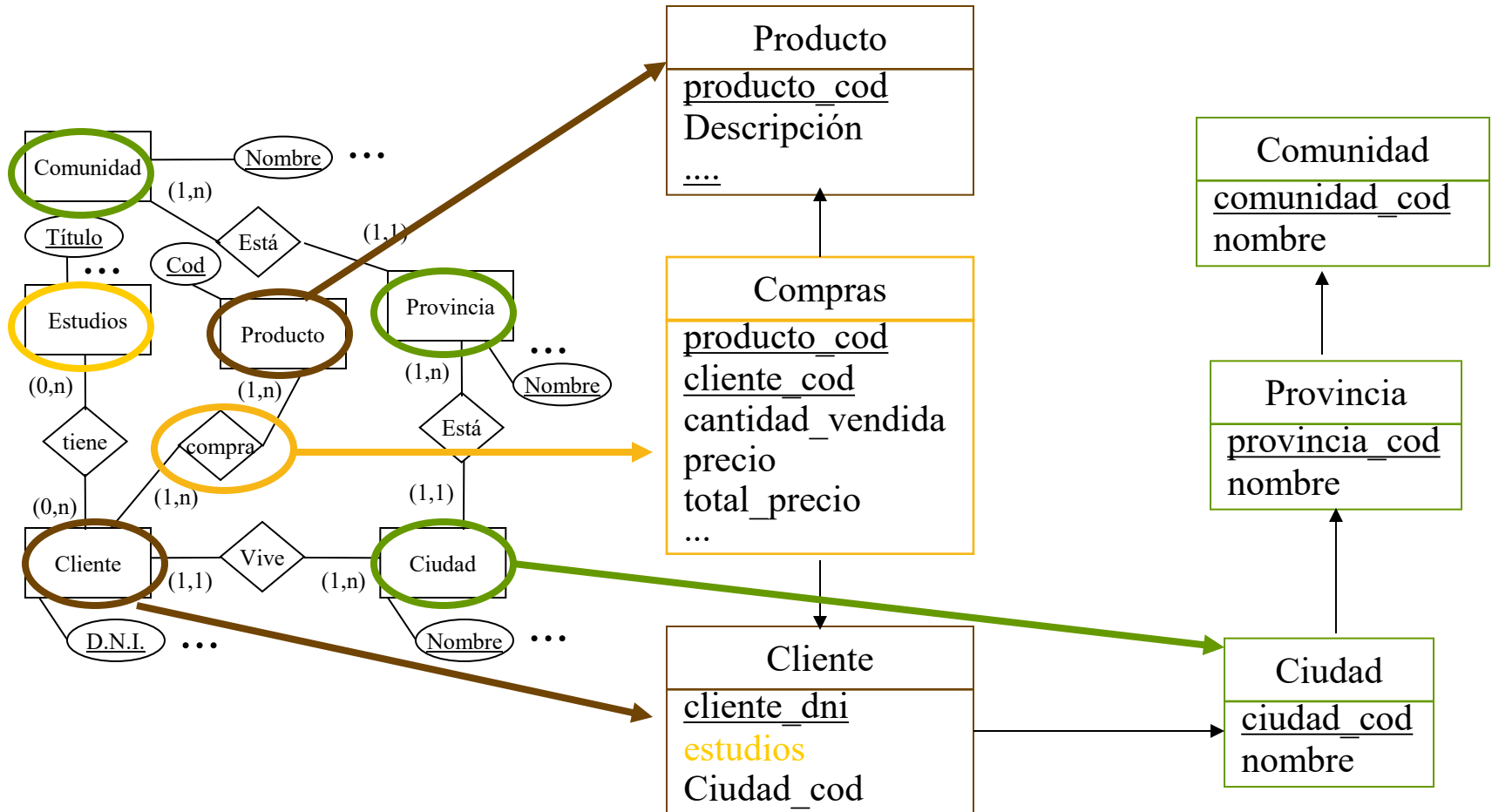
37

- Giovinnazo (2000). Object-Oriented Data Warehouse Design: Building a star schema
- Inmon (2005). Building the Data Warehouse (4^a ed.)
- Kimball & Ross (2013). The Data Warehouse Toolkit (4^a ed.)
- Thomsen (2000). OLAP solutions: Building Multidimensional Information Systems

Diseño conceptual de almacenes de datos

Apéndice: Guías de diseño del Modelo EER → Esq. Estrella

38



DISEÑO CONCEPTUAL DE ALMACENES DE DATOS

Tema 2

Profesores:

Juan C. Trujillo
Alejandro Reina Reina
LUCENTIA Research Group



Universitat d'Alacant
Universidad de Alicante



Departamento de
Lenguajes y Sistemas
Informáticos