

# Diseño lógico de almacenes de datos

## Tema 3

### Profesores:

Juan C. Trujillo  
Alejandro Reina Reina  
LUCENTIA Research Group



Universitat d'Alacant  
Universidad de Alicante



Departamento de  
Lenguajes y Sistemas  
Informáticos

# Diseño lógico de DW

2

- Modelado lógico multidimensional (MD)
- Modelado multidimensional relacional
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ Resumen/comparativa final
  - ▣ La dimensión tiempo
  - ▣ Otras consideraciones de diseño
- El proceso de diseño de un DW

# Diseño lógico de DW

3

- Modelado lógico multidimensional (MD)
- Modelado multidimensional relacional
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ Resumen/comparativa final
  - ▣ La dimensión tiempo
  - ▣ Otras consideraciones de diseño
- El proceso de diseño de un DW

# Modelado multidimensional

4

- El modelado lógico multidimensional depende del tipo de servidor OLAP
  - ▣ MOLAP (Multidimensional OLAP)
    - Vectores o matrices multidimensionales
  - ▣ ROLAP (Relational OLAP)
    - Esquema estrella de R. Kimball
      - Tablas relacionales para representar hechos y dimensiones
      - Variantes esquema estrella
        - Constelaciones de hechos
        - Copos de nieve

# Diseño lógico de DW

5

- Modelado lógico multidimensional (MD)
- Modelado multidimensional relacional
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ Resumen/comparativa final
  - ▣ La dimensión tiempo
  - ▣ Otras consideraciones de diseño
- El proceso de diseño de un DW

# Diseño lógico de DW

6

- Modelado lógico multidimensional (MD)
- Modelado multidimensional relacional
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ Resumen/comparativa final
  - ▣ La dimensión tiempo
  - ▣ Otras consideraciones de diseño
- El proceso de diseño de un DW

# Modelado multidimensional

## Esquema estrella

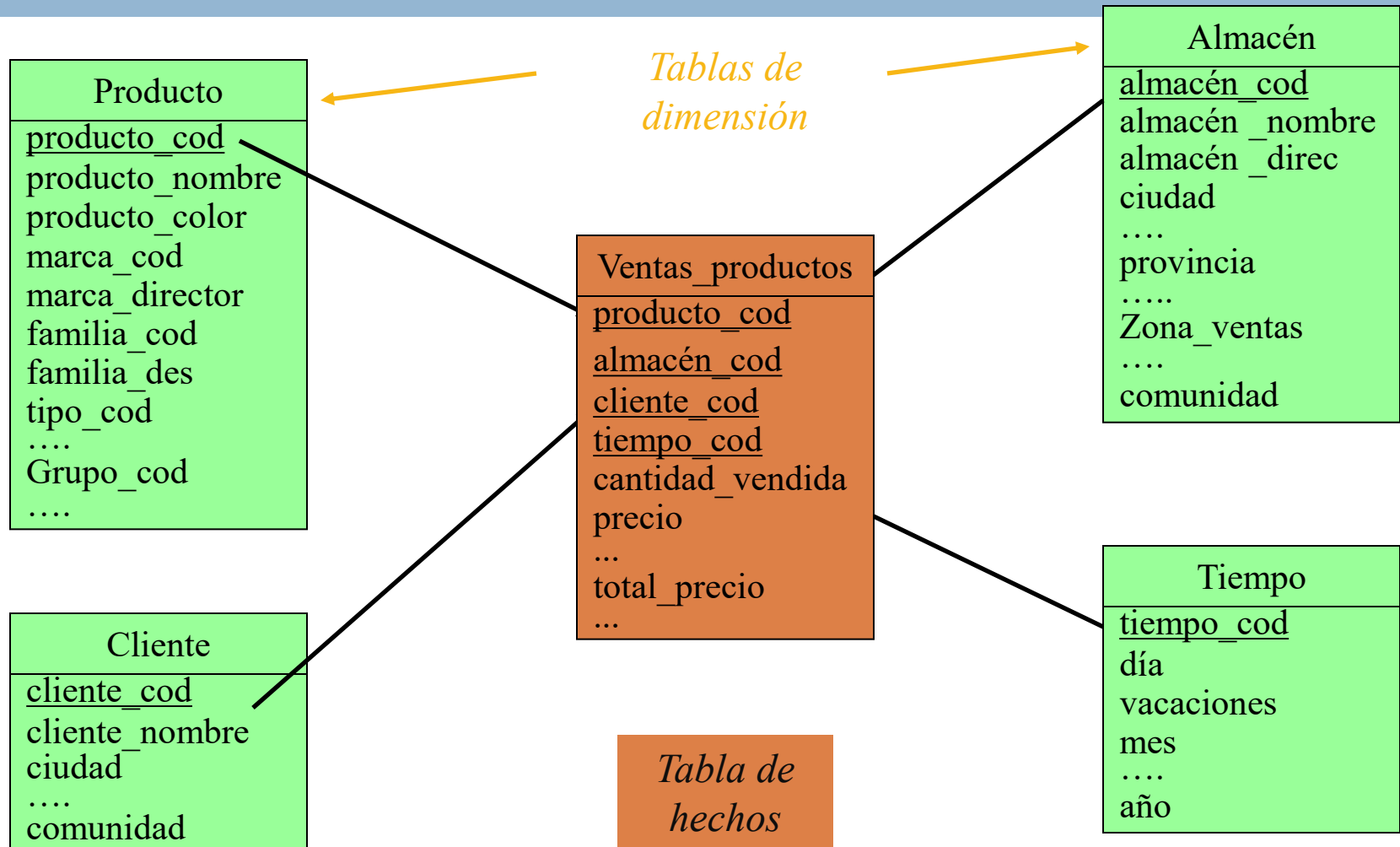
7

- Hechos → tablas de hechos
- Dimensiones → tablas de dimensiones
- Tabla de hechos es relación muchos a muchos con las tablas de dimensiones
  - ▣ Técnica utilizada → desnormalización
  - ▣ Clave primaria tabla de hechos compuesta por claves ajenas de tablas de dimensiones

# Modelado multidimensional

## Esquema estrella

8





# Modelado multidimensional

Esquema estrella. Granularidad.

9

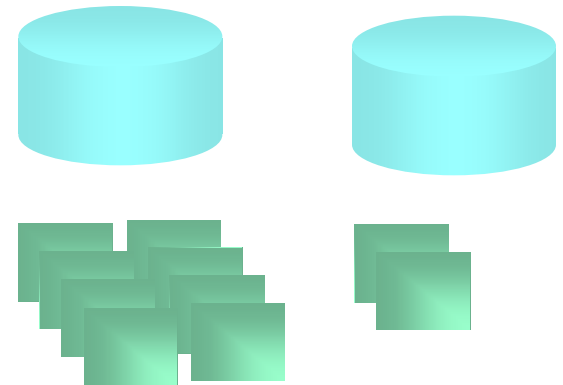
## □ Granularidad

### ▣ Afecta al Almacén de datos

- Tamaño del repositorio
- Grado de análisis
- Flexibilidad

### ▣ Nivel de detalle de los datos ...

- Transacciones individuales
- Resúmenes diarios
- Resúmenes mensuales
- Resúmenes anuales
- Cualquier otro período de tiempo



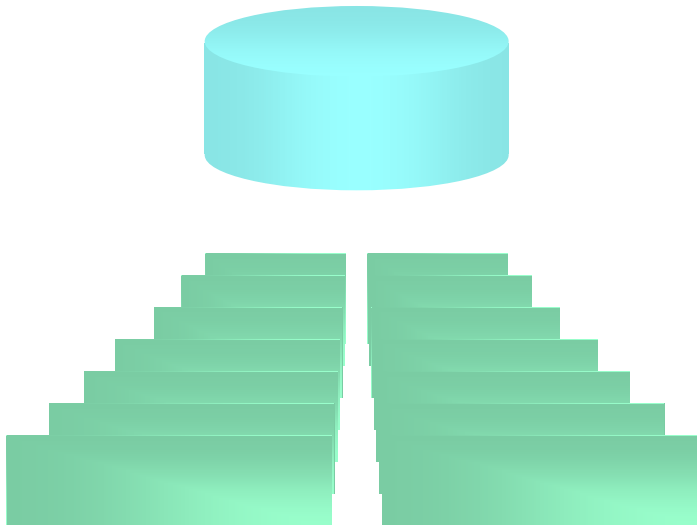
# Modelado multidimensional

Esquema estrella. Granularidad.

10

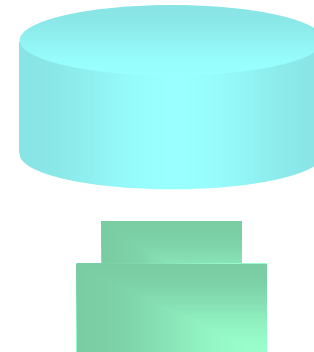
## ■ Nivel alto (fine grain)

**Detalles de las transacciones  
de los clientes del banco por mes**



## ■ Bajo nivel (coarse grain)

**Resumen de las transacciones  
de los clientes del banco por mes**



**Define el nivel en función de  
las necesidades del negocio**

# Modelado multidimensional

## Esquema estrella

11

### □ Tablas de dimensiones

- Describen el contexto para analizar los hechos
  - Datos “textuales” (Alfanuméricos)
  - Datos desnormalizados → redundancia
- Cada fila contiene su clave primaria y los atributos descriptores de todos los niveles de jerarquía
- Tablas más pequeñas que las tablas de hechos

# Modelado multidimensional

## Esquema estrella

12

### □ Tabla de dimensión producto

| <u>Producto_cod</u> | Producto nombre .... | Family_Desc         |
|---------------------|----------------------|---------------------|
| 1                   | Puleva Milk ½ L.     | Productos lácteos   |
| 2                   | Puleva Milk 1L .     | Productos lácteos   |
| 3                   | Yogourt Pascual      | Productos lácteos   |
| 4                   | Mr. Proper           | Productos limpieza  |
| 5                   | Ajax                 | Productos limpieza. |

| Producto            |
|---------------------|
| <u>producto_cod</u> |
| producto_nombre     |
| producto_color      |
| marca_cod           |
| marca_director      |
| familia_cod         |
| familia_des         |
| tipo_cod            |
| ....                |
| Grupo_cod           |
| ....                |

# Modelado multidimensional

## Esquema estrella

13

SQL\*Plus Worksheet

Archivo Editar Hoja de trabajo Ayuda

ORACLE

```
--Connect jtrujillo/****@ibm
select * from almacen;

--set lin 800;

--desc almacen;
```

| REG_ID | REG_NOMBRE   | ALM_DIRECTOR | ALM_TIPO_EDIFICIO | CIT |
|--------|--------------|--------------|-------------------|-----|
| 1      | Eastern      | Jones        | Modern            |     |
| 2      | Mid West     | Smith        | Original          |     |
| 3      | South East   | Davis        | Compact           |     |
| 4      | Pacific      | Johnson      | Modern            |     |
| 4      | Pacific      | Green        | Original          |     |
| 1      | Eastern      | Brown        | Compact           |     |
| 1      | Eastern      | White        | Modern            |     |
| 5      | South West   | Williams     | Original          |     |
| 4      | Pacific      | Stuber       | Compact           |     |
| 5      | South West   | Merz         | Modern            |     |
| 2      | Mid West     | Erickson     | Original          |     |
| 2      | Mid West     | Kalman       | Compact           |     |
| 2      | Mid West     | Inmon        | Modern            |     |
| 6      | Mountain     | Strehlo      | Original          |     |
| 6      | Mountain     | Ollom        | Compact           |     |
| 2      | Mid West     | Mantle       | Modern            |     |
| 7      | Mid Atlantic | Mays         | Original          |     |
| 3      | South East   | Maris        | Compact           |     |
| 1      | Eastern      | Ruth         | Modern            |     |
| 2      | Mid West     | Cobb         | Original          |     |
| 1      | Eastern      |              |                   |     |

# Modelado multidimensional

## Esquema estrella

14

### □ Tablas de hechos

- Actividades básicas de empresa

- Cada fila se compone de:

  - Clave primaria (compuesta por claves ajenas de las dimensiones)

  - Medidas → Datos numéricos

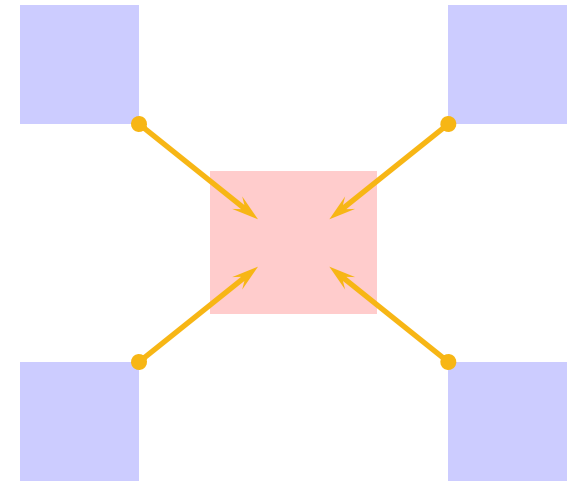
- Generalmente relación m-n con dimensiones y, m-1 en particular con cada dimensión

# Modelado multidimensional

## Esquema estrella

15

□ Tabla de hecho “ventas”



| <u>CliKey</u> | <u>ProductoKey</u> | <u>AlmacénKey</u> | <u>Tiempo key</u> | Sale Amount |
|---------------|--------------------|-------------------|-------------------|-------------|
| 1             | 1                  | 1                 | 1                 | 100€        |
| 1             | 2                  | 1                 | 2                 | 120€        |
| 1             | 3                  | 1                 | 3                 | 200€        |
| .             | .                  | .                 | .                 |             |

# Modelado multidimensional

## Esquema estrella

16

SQL\*Plus Worksheet

Archivo Editar Hoja de trabajo Ayuda

ORACLE

```
--Connect jtrujillo/****@ibm
--set lin 2000;
select * from ventas;
```

60 filas seleccionadas.

| ALM_ID | COSTES | DA_SMART_KEY | NUMERO_CLIENTES | PRM_ID | PRO_ID | UNIDADES_VENDIDAS | VENTAS |
|--------|--------|--------------|-----------------|--------|--------|-------------------|--------|
| 10     | 13,65  | 89           | 14              | 11     | 18     | 23                | 16,05  |
| 7      | 23,45  | 90           | 36              | 11     | 18     | 40                | 27,66  |
| 4      | 50,39  | 91           | 64              | 11     | 18     | 87                | 60,08  |
| 16     | 1,48   | 92           | 2               | 11     | 18     | 2                 | 1,66   |
| 10     | 11,45  | 93           | 13              | 11     | 18     | 20                | 13,97  |
| 17     | 35,76  | 94           | 37              | 11     | 18     | 63                | 43,24  |
| 3      | 34,22  | 95           | 52              | 11     | 18     | 60                | 41,25  |
| 7      | 19,28  | 96           | 24              | 11     | 18     | 33                | 23,08  |
| 12     | 30,82  | 97           | 53              | 11     | 18     | 54                | 37,2   |
| 4      | 46,43  | 98           | 65              | 11     | 18     | 76                | 52,42  |
| 19     | 51,59  | 99           | 67              | 11     | 18     | 93                | 64,05  |
| 8      | 21,88  | 100          | 30              | 11     | 18     | 39                | 26,9   |
| 5      | 32,85  | 101          | 30              | 11     | 18     | 53                | 36,59  |
| 4      | 14,12  | 102          | 13              | 11     | 18     | 23                | 16,07  |
| 7      | 41,08  | 103          | 41              | 11     | 18     | 73                | 50,66  |
| 4      | 35,36  | 104          | 41              | 11     | 18     | 58                | 39,68  |
| 1      | 43,38  | 105          | 74              | 11     | 18     | 79                | 54,17  |

2025



# Modelado multidimensional

## Esquema estrella

17

SQL\*Plus Worksheet

Archivo Editar Hoja de trabajo Ayuda

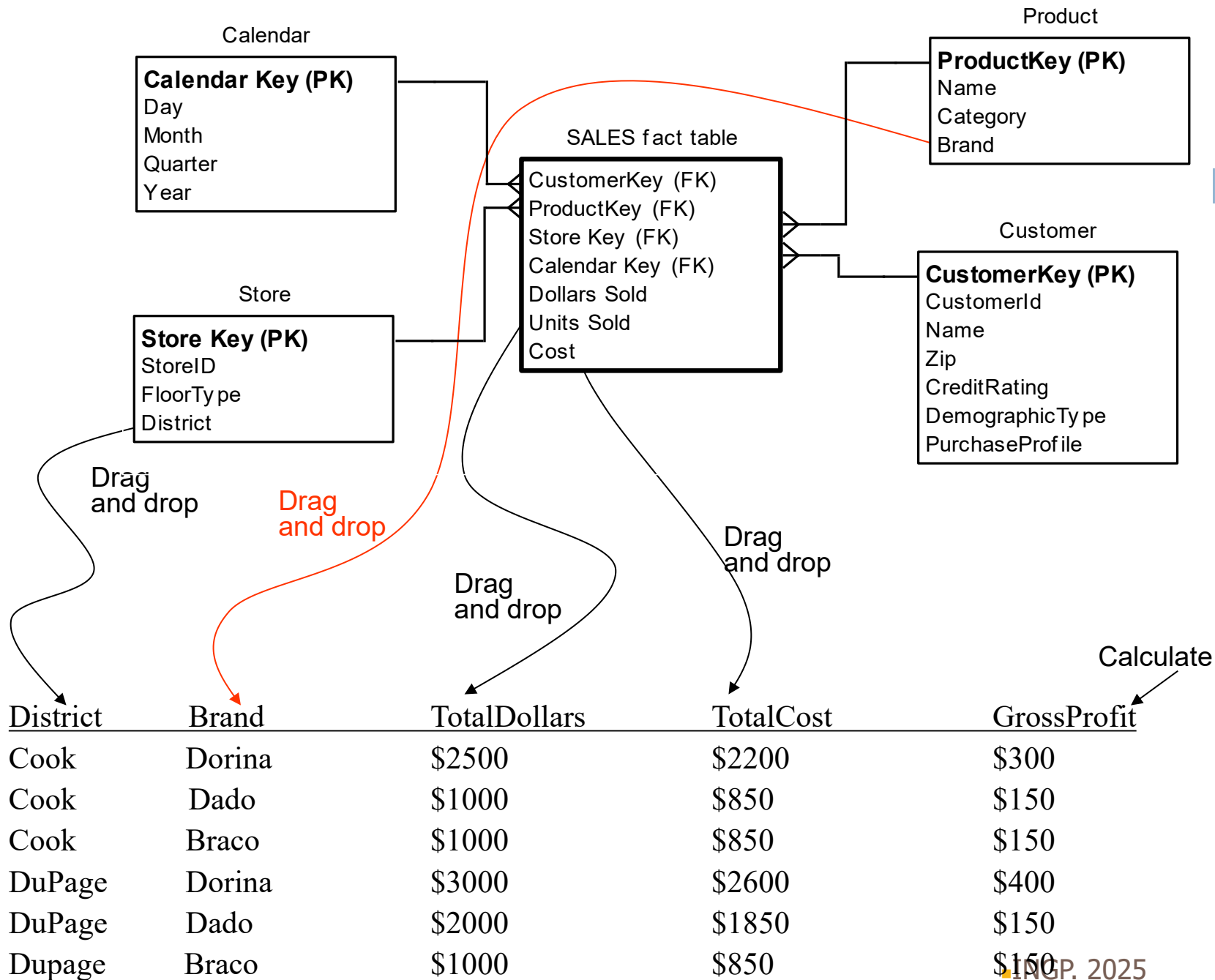
ORACLE

```
--Connect jtrujillo/****@ibm
--set lin 2000;
select * from ventas;
```

|    |        |     |    |    |    |    |        |
|----|--------|-----|----|----|----|----|--------|
| 4  | 58,68  | 128 | 24 | 11 | 56 | 44 | 68,84  |
| 6  | 118,14 | 129 | 84 | 11 | 56 | 88 | 136,87 |
| 20 | 94,09  | 130 | 40 | 11 | 56 | 72 | 111,26 |
| 3  | 16,52  | 131 | 9  | 11 | 56 | 13 | 19,94  |
| 8  | 52,61  | 132 | 27 | 11 | 56 | 41 | 63,84  |
| 15 | 34,28  | 133 | 18 | 11 | 56 | 26 | 40,21  |
| 6  | 86,07  | 134 | 62 | 11 | 56 | 68 | 106,25 |
| 11 | 83,55  | 135 | 36 | 11 | 56 | 65 | 101,23 |
| 10 | 38,73  | 136 | 23 | 11 | 56 | 30 | 47     |
| 12 | 98,24  | 137 | 58 | 11 | 56 | 72 | 111,05 |
| 6  | 113,28 | 138 | 66 | 11 | 56 | 83 | 128,84 |
| 16 | 89,66  | 139 | 43 | 11 | 56 | 70 | 109,33 |
| 13 | 19,92  | 140 | 8  | 11 | 56 | 15 | 22,78  |
| 11 | 117,51 | 141 | 72 | 11 | 56 | 86 | 133,07 |
| 6  | 11,23  | 142 | 4  | 11 | 56 | 8  | 13,14  |
| 16 | ,43    | 143 | 0  | 11 | 56 | 0  | ,51    |
| 12 | 26,15  | 144 | 17 | 11 | 56 | 20 | 31,55  |
| 20 | 6,51   | 145 | 4  | 11 | 56 | 5  | 7,71   |

11044 filas seleccionadas.

Inicio Microsoft Pow... d:\orant9\bin... SQL\*Plus Wo... 19:05



# Modelado multidimensional

Esquema estrella. Vista relacional.

19

## SQL Statement:

```

select      District, Brand
            sum(DollarsSold) TotalDollars,
            sum(Cost) TotalCost,
            sum(DollarsSold) - sum(Cost) GrossProfit,
from        Sales, Store, product
where       Sales.store_key = Store.store_key and
            Sales.product_key = Product.product_key
group by    District, Brand
  
```

## RESULT

| District | Brand  | TotalDollars | TotalCost | GrossProfit |
|----------|--------|--------------|-----------|-------------|
| Cook     | Dorina | \$2500       | \$2200    | \$300       |
| Cook     | Dado   | \$1000       | \$850     | \$150       |
| Cook     | Braco  | \$1000       | \$850     | \$150       |
| DuPage   | Dorina | \$3000       | \$2600    | \$400       |
| DuPage   | Dado   | \$2000       | \$1850    | \$150       |
| Dupage   | Braco  | \$1000       | \$850     | \$150       |

# Modelado multidimensional

Esquema estrella. Vista relacional.

20

**Productos**

(dimensión en el esquema estrella)

**Almacenes (loc. geog)**

(dimensión)

**Cantidad**

| Producto | Ciudad   | Día | Mes     | Año  | Unidades |
|----------|----------|-----|---------|------|----------|
| Milk     | Toronto  | 15  | January | 1997 | 1200     |
| IntBread | Kingston | 22  | January | 1997 | 1050     |
| Juice    | Ottawa   | 25  | January | 1997 | 1350     |
| IntBread | Ottawa   | 28  | January | 1997 | 1500     |

# Modelado multidimensional

Esquema estrella. Vista Multidimensional.

21

Corte en la tabla de dimensión: **Enero** Datos derivados

**Almacén (dimensión)**

**Lugar**

| Producto  | Toronto | Ottawa | Kingsto | Lake Reg. | Other Regs. | Total |
|-----------|---------|--------|---------|-----------|-------------|-------|
| Juice     | 0       | 1350   | 0       | 0         | 1350        | 1350  |
| Milk      | 1200    | 0      | 0       | 1200      | 0           | 1200  |
| IntBread  | 0       | 1500   | 1050    | 1050      | 1500        | 2550  |
| Cheese    | 1600    | 0      | 0       | 1600      | 0           | 1600  |
| Dairy     | 2800    | 0      | 0       | 2800      | 0           | 2800  |
| Breads    | 0       | 1500   | 1050    | 1050      | 1500        | 2550  |
| Beverages | 0       | 1350   | 0       | 0         | 1350        | 1350  |

**Producto (dimensión)**

**Cantidad y datos derivados**

**Corte en el Cubo Multidimensional**

# Modelado multidimensional

Esquema estrella. Particularidades.

22

## □ Dimensiones degeneradas

- ▣ Las tablas de hechos contienen información (atributo) sobre una dimensión que no existe físicamente
  - Este atributo puede/suele formar parte de la clave primaria de la tabla de hechos
    - Suelen referirse a dimensiones físicas en mundo real
      - Ej. Número de ticket, cod\_factura, cod\_albarán

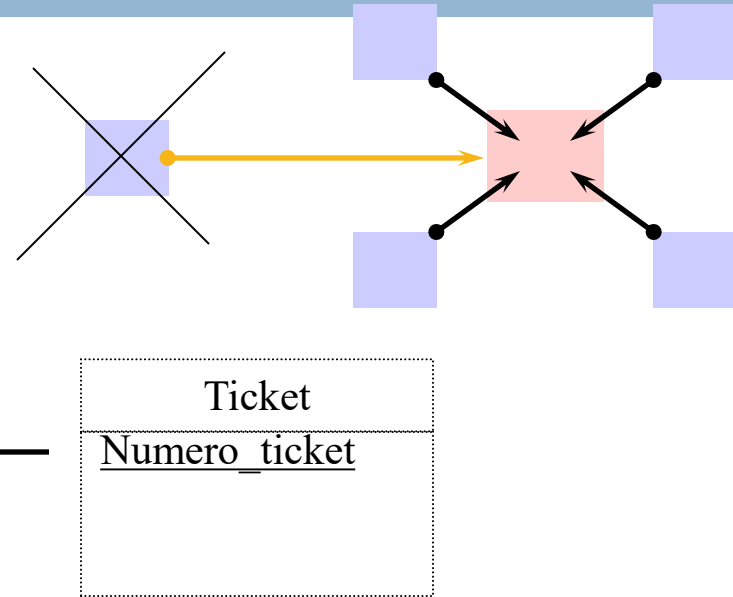
# Modelado multidimensional

## Esquema estrella

23

### ■ Dimensiones degeneradas

| Ventas_productos     |  |
|----------------------|--|
| <u>producto_cod</u>  |  |
| <u>almacén_cod</u>   |  |
| <u>cliente_cod</u>   |  |
| <u>tiempo_cod</u>    |  |
| <u>Número_ticket</u> |  |
| cantidad_vendida     |  |
| precio               |  |
| ...                  |  |
| total_precio         |  |
| ...                  |  |



- Normalmente indica que hay relación muchos a muchos en particular con una dimensión
- A veces a estos esquemas se les denomina Multi Star Schema

# Modelado multidimensional

Esquema estrella. Particularidades.

24

- Tablas de hechos que no son hechos (*Factless fact tables*)
  - No contienen medidas o atributos de análisis
  - Recogen la ocurrencia de un evento

| Impartir_clase        |
|-----------------------|
| <u>Profesor_dni</u>   |
| <u>Asignatura_cod</u> |
| <u>Alumno_dni</u>     |
| ...                   |



# Modelado multidimensional

Esquema estrella. Particularidades

25

- Claves primarias de las tablas de dimensiones
  - ▣ (i) Autogeneradas
    - Rendimiento
    - Incluso en la dimensión Tiempo
    - Más fáciles de manejar para procesos ETL
  - ▣ (ii) Con significado semántico

# Modelado multidimensional

Esquema estrella. Particularidades

26

## □ Otra vez !!!! → Razones para desnormalizar

### □ Rendimiento

#### ■ Eliminar el número de uniones entre tablas

- Elevado número de filas

- Unión es la operación más cara

### □ Intuitivo

- Más fácil para consultar por parte del usuario no experto

# Modelado multidimensional

Esquema estrella. Particularidades

27

- Dimensiones que cambian lentamente
  - ▣ Las instancias de dimensiones se pueden considerar fijas a lo largo del tiempo
  - ▣ Sin embargo, a veces deseamos registrar cambios que puedan suceder en algunas dimensiones para seguir analizando datos históricos
    - Normalmente cambios en descripción de productos o de clientes
  - ▣ *Kimball* propone tres soluciones básicas y dos híbridas

# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

28

- Sol. 1: Sobrecribir el valor antiguo en la fila correspondiente de la dimensión
  - La fila registra el último valor válido
    - Perdemos la habilidad de analizar “la historia”
  - Se suele utilizar cuando el valor antiguo deja de tener significado o cuando se corrigen errores
  - Aproximación simple
  - Ejemplos: descripción productos, nombres nuevos, etc.

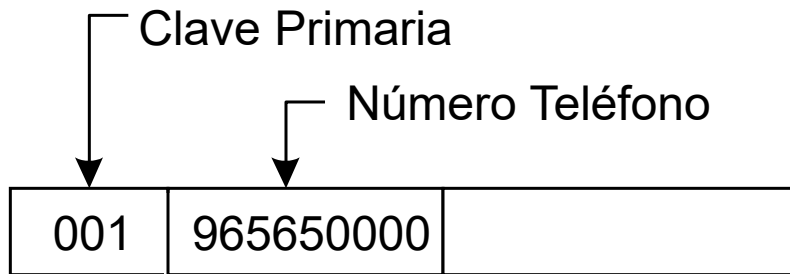
# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

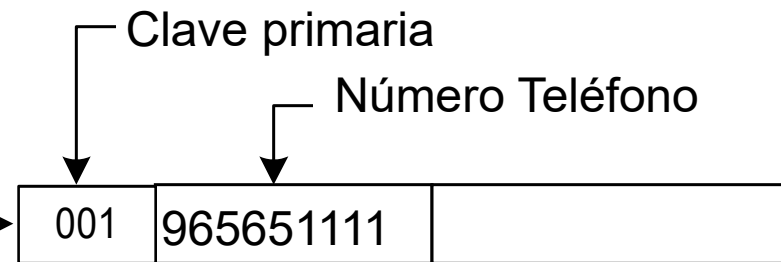
29

- Sol. 1: Sobrecribir el valor antiguo en la fila correspondiente de la dimensión

*Estado actual*



*Después de cambios*



# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

30

- Sol. 2: Añadir una fila nueva utilizando un nuevo valor de clave autogenerada
  - ▣ Permite registrar el cambio del valor de un atributo en una dimensión
  - ▣ Representa un cambio físico real en la dimensión
  - ▣ Imprescindible utilizar claves autogeneradas
  - ▣ Añadir un campo que identifique al valor actual
    - Añadir un campo que sea un “flag” al *valor actual*

# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

31

- Sol. 2: Añadir una fila nueva utilizando un nuevo valor de clave autogenerada...
  - ▣ Se puede utilizar dos atributos añadidos para registrar la fecha de comienzo y fin de validez de cada instancia
  - ▣ Inconvenientes
    - Necesita almacenamiento extra para los atributos extras
    - Realizar test para excluir las versiones antiguas
  - ▣ Ej. Dirección, Región...

# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

32

- Sol. 2: Añadir una fila nueva utilizando un nuevo valor de clave autogenerada...
  - ▣ Segmenta la historia
    - Particiona la tabla de hechos
      - Dos “productos” distintos con claves generadas distintas estarán registrados en las ventas
  - ▣ Queries...
    - Definir restricciones sobre atributo que ha cambiado el valor
    - Si queremos mismo producto, actuaremos sobre clave primaria original



# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

33

- Sol. 2: Añadir una fila nueva utilizando un nuevo valor de clave autogenerada...
  - ▣ Suele ser la más utilizada para las verdaderamente *Dimensiones que cambian lentamente*

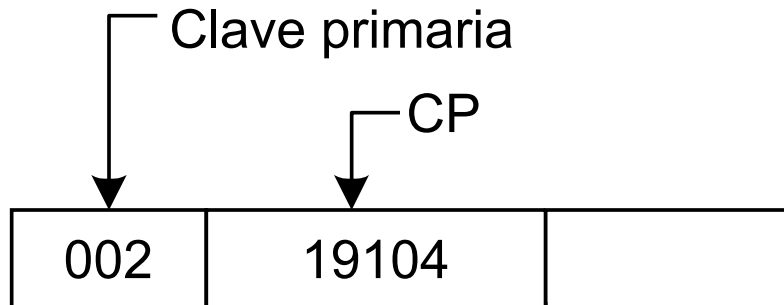
# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

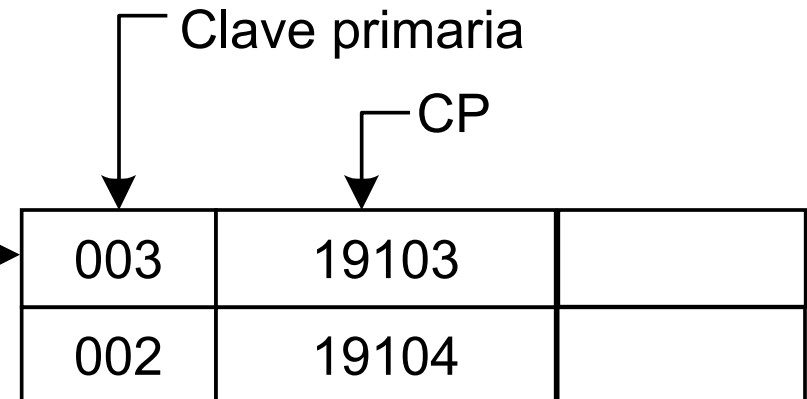
34

- Sol. 2: Añadir una fila nueva utilizando un nuevo valor de clave autogenerada...

*Estado actual*



*Después de cambios*



# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

35

- Sol. 3: Añadir atributos para registrar el valor nuevo y el inmediatamente anterior
  - ▣ Suele utilizarse en cambios mínimos como por ejemplo redefinición de *descripción de producto*
  - ▣ Utilizado cuando deseamos seguir la historia de tanto los valores antiguos como nuevos
  - ▣ Se puede / suele añadir un atributo de *fecha\_efectiva*
  - ▣ Poco recomendada / utilizada

# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

36

- Sol. 3: Añadir atributos para registrar el valor nuevo y el inmediatamente anterior
  - Inconveniente:
    - Sólo permite analizar los datos de la tabla de hechos o bien por el valor antiguo o bien por el nuevo
      - Ya que se utiliza la misma clave autogenerada

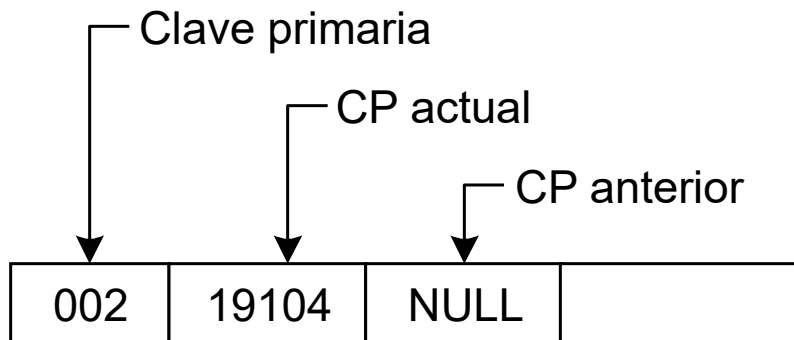
# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

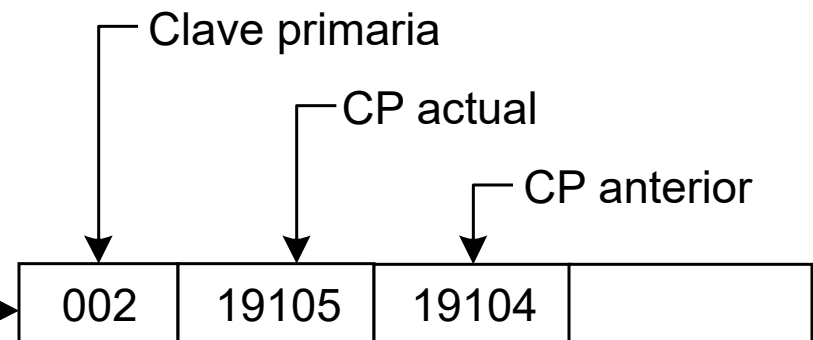
37

- Sol. 3: Añadir atributos para registrar el valor nuevo y el inmediatamente anterior

*Estado actual*



*Después de cambios*



# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

38

- Sol híbridas son más flexibles aunque proporcionan mayor complejidad de manejo
  - ▣ Muy seguro de ser adecuado para el negocio y los posibles requerimientos que se pueden lanzar
- Sol. Híbrida 1: cambios predecibles con versiones múltiples que se superponen
  - ▣ Existen cambios predecibles y regulares
  - ▣ Usar la aproximación de tipo 3 generalizada

# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

39

- Sol. Híbrida 1: cambios predecibles con versiones múltiples que se superponen

| Sales Rep Dimension  |
|----------------------|
| <u>Sales rep key</u> |
| Sales rep name       |
| Sales Rep address    |
| Current district     |
| District 2001        |
| District 2000        |
| District 1999        |
| ...                  |

- Conocen los años de análisis de ventas
- Se puede consultar el estado del distrito en cada año

# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

40

- Sol. Híbrida 2: cambios no predecibles con versiones simples que se superponen
  - ▣ Existen cambios irregulares e impredecibles
  - ▣ Necesidad de preservar los valores históricos
  
- ▣ Ninguna de las soluciones estándares anteriores contempla estas dos características



# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

41

- Sol. Híbrida 2: cambios no predecibles con versiones simples que se superponen
  - ▣ Utilizar los tipo 1 + tipo 2 + tipo 3
    - Añadir una nueva fila para capturar los cambios
    - Añadir una nueva columna para controlar los valores actuales
    - Utilizar la aproximación de tipo 1 para cambios de valores muy recientes donde la historia no interesa
    - Más información: cap. 4 (*R. Kimball*), pag. 104

# Modelado multidimensional

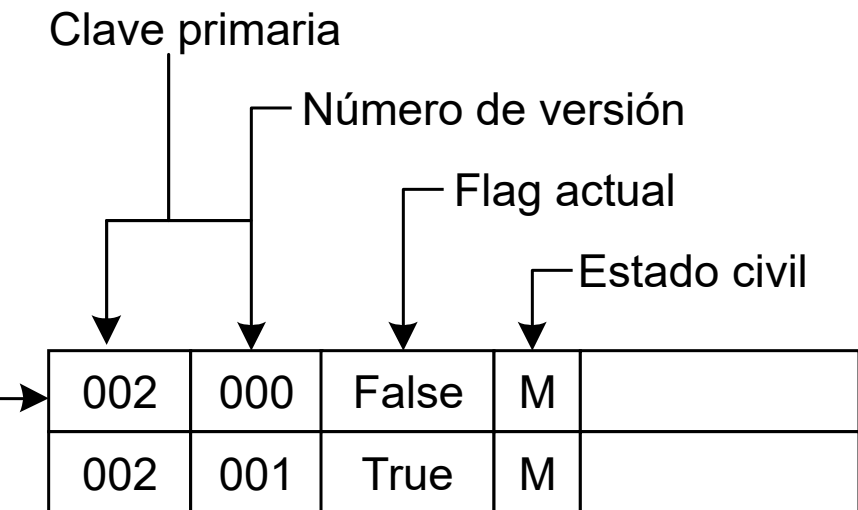
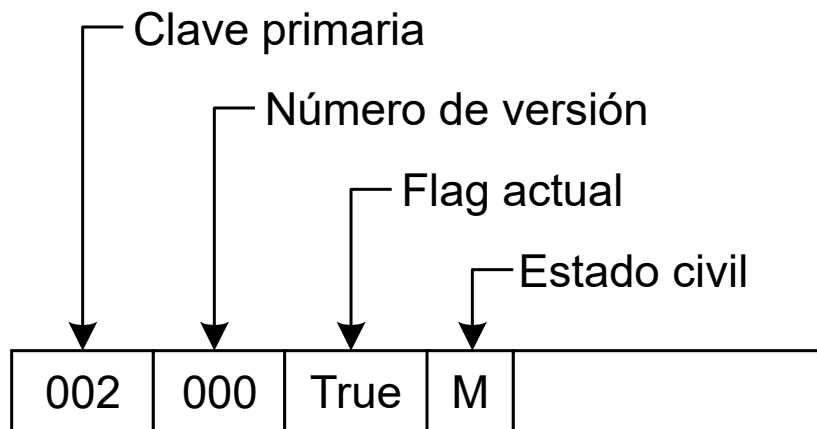
Esquema estrella. Dimensiones que cambian lentamente

42

- Otros métodos
- Método 4: Registrando versiones con *flag's*

*Estado actual*

*Después de cambios*



# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

43

- Método 5: Utilizando la clave primaria de los sistemas OLTP con flags

*Estado actual*

*Después de cambios*

| Clave primaria | Cliente ID | CP    | Actual | Fecha     |
|----------------|------------|-------|--------|-----------|
| 002            | 111        | 19104 | True   | 5/31/2000 |

| Clave primaria | Cliente ID | CP    | Actual Flag | Fecha     |
|----------------|------------|-------|-------------|-----------|
| 202            | 111        | 19107 | True        | 5/31/2001 |
| 002            | 111        | 19104 | False       | 5/31/2000 |

# Modelado multidimensional

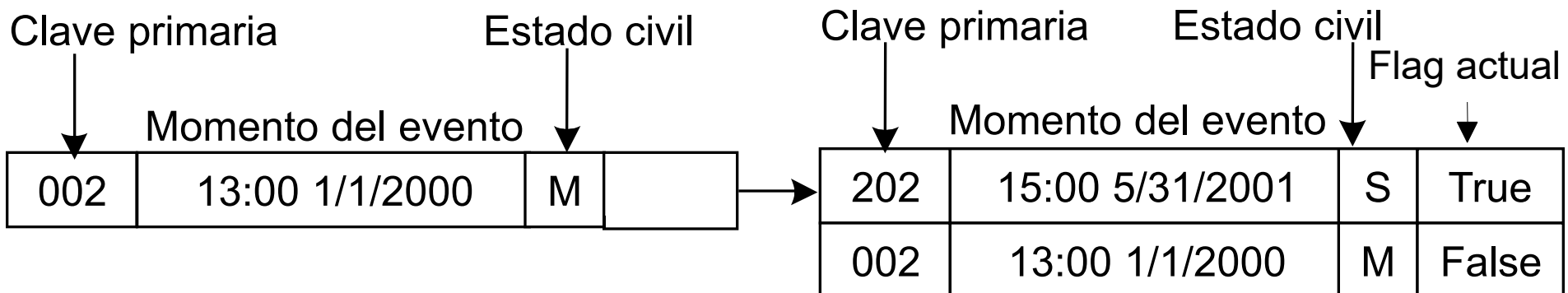
Esquema estrella. Dimensiones que cambian lentamente

44

- Método 6: Modelar eventos con flag's
  - ▣ Utiliza claves de OLTP. Prácticamente tipo 2

*Estado actual*

*Después de cambios*



# Modelado multidimensional

Esquema estrella. Dimensiones que cambian lentamente

45

## □ Método 7: Modelado del estado con clave OLTP y flag

*Estado actual*

| Clave primaria | Fecha comienzo | Fecha final | Estado civil |
|----------------|----------------|-------------|--------------|
| 002            | 09:00 1/1/1998 | NULL        | S            |

*Después de cambios*

| Clave primaria | Cliente ID | Fecha comienzo | Fecha final | Estado civil | Flag actual |
|----------------|------------|----------------|-------------|--------------|-------------|
| 202            | 111        | 1/1/2000       | NULL        | M            | True        |
| 002            | 111        | 1/1/1998       | 1/1/2000    | S            | False       |

# Modelado multidimensional

## Esquema estrella

46

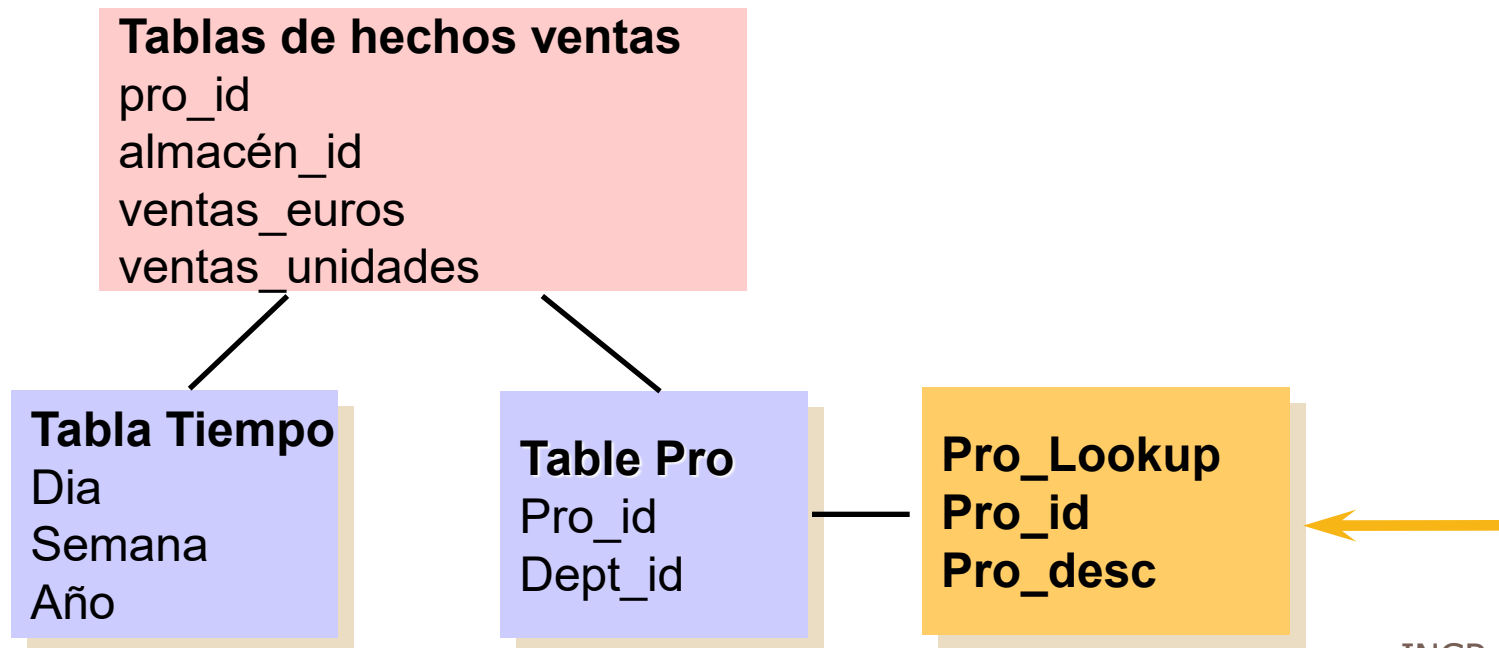
- Qué información podemos saber a partir del esquema estrella ....
  - ▣ ¿Cuál es la jerarquía de clasificación en las dimensiones?
  - ▣ ¿Es la jerarquía entre almacén y zona\_Ventas estricta?
  - ▣ Hay demasiada información explícita → necesidad de documentar el diseño

# Modelado multidimensional

Esquema estrella. Tablas y datos de referencia.

47

- ▣ Proporciona apoyo para la gestión de las dimensiones
- ▣ Se reduce el volumen del almacén de datos



# Diseño lógico de DW

48

- Modelado lógico multidimensional (MD)
- **Modelado multidimensional relacional**
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ Resumen/comparativa final
  - ▣ La dimensión tiempo
  - ▣ Otras consideraciones de diseño
- El proceso de diseño de un DW



# Diseño lógico de DW

49

- Modelado lógico multidimensional (MD)
- Modelado multidimensional relacional
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ Resumen/comparativa final
  - ▣ La dimensión tiempo
  - ▣ Otras consideraciones de diseño
- El proceso de diseño de un DW

# Modelado multidimensional

## Esquema constelaciones de hechos

50

- Esquema de constelaciones de hechos (fact constellations)
  - ▣ Además de la tabla de hechos base, se define una tabla de hechos para cada nivel de agregado
  - ▣ A veces se les denomina *Multi Fact Tables* por contener más de una tabla de hechos
  - ▣ Acelera la consulta de datos agregados

# Modelado multidimensional

## Esquema constelaciones de hechos

51

### □ Esquema de constelaciones de hechos (fact constellations)

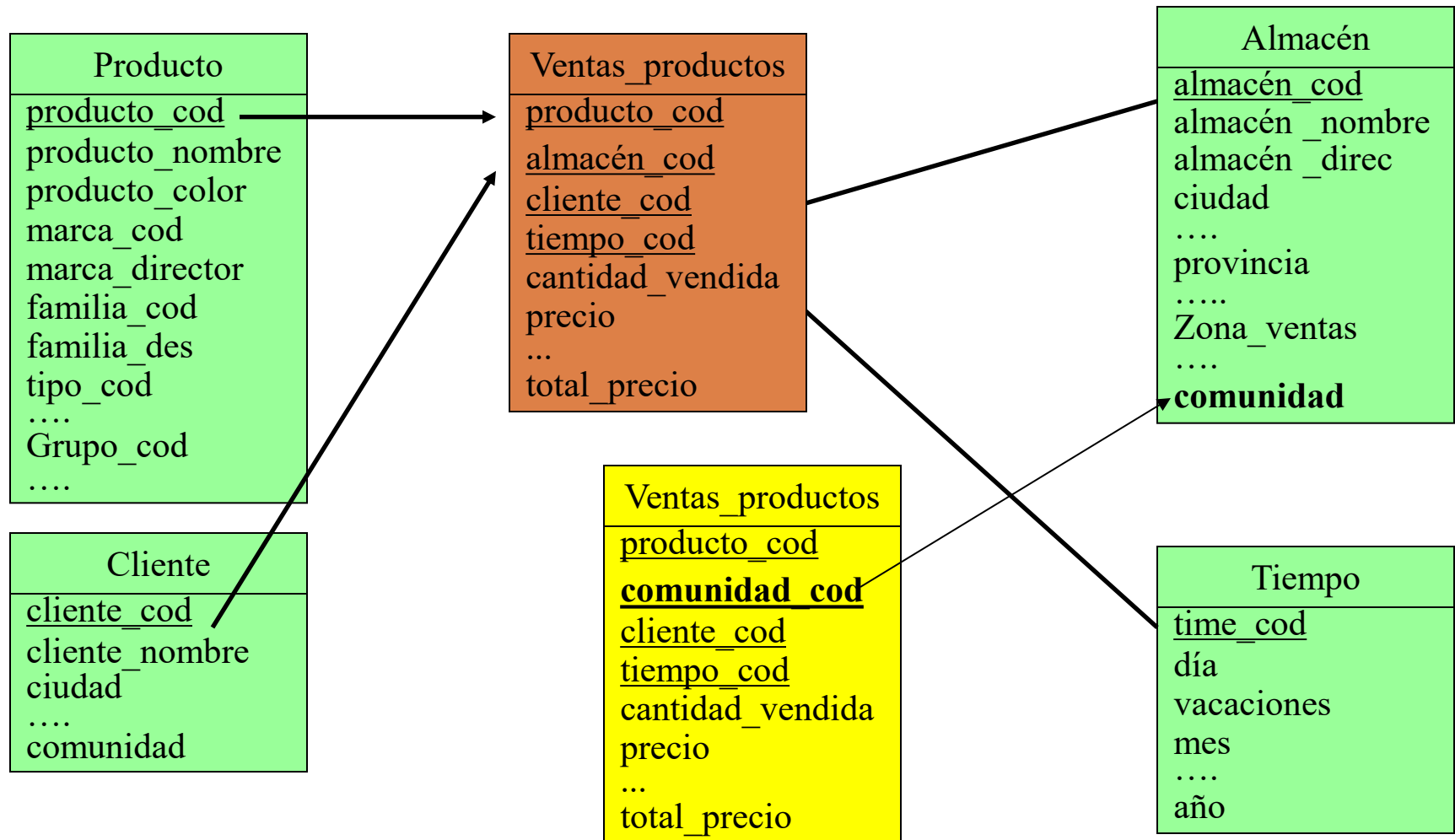
#### □ Inconvenientes

- Cada tabla de agregados se usa para calcular su nivel
  - Navegar por jerarquías requiere escanear (browse) distintas tablas
- Aumenta el tamaño de los metadatos
- Dificulta su gestión y mantenimiento ya que para cada carga nueva de datos se han de recalculan todas las tablas de hechos
- Puede haber requerimientos que necesiten varias tablas

# Modelado multidimensional

## Esquema constelaciones de hechos

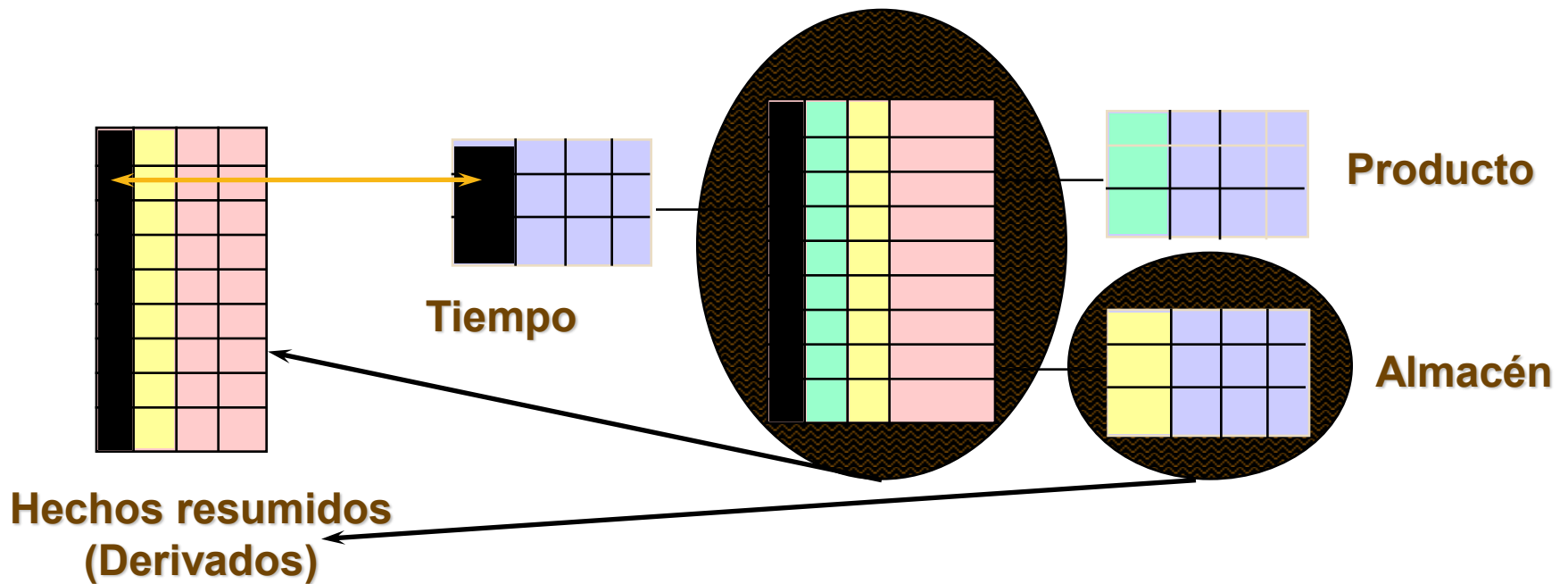
52



# Modelado multidimensional

Esquema constelaciones de hechos

53



# Modelado multidimensional

Esquema constelaciones de hechos

54

- ▣ Acceso rápido a datos precalculados
- ▣ Reduce el uso de
  - I/O
  - CPU
  - Memoria
- ▣ Diferencia con fuentes...
  - Sistemas fuente de datos – poco resumidos
  - Resúmenes precalculados – muy resumidos
- ▣ Determinar los requerimientos pronto

# Diseño lógico de DW

55

- Modelado lógico multidimensional (MD)
- **Modelado multidimensional relacional**
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ Resumen/comparativa final
  - ▣ La dimensión tiempo
  - ▣ Otras consideraciones de diseño
- El proceso de diseño de un DW

# Diseño lógico de DW

56

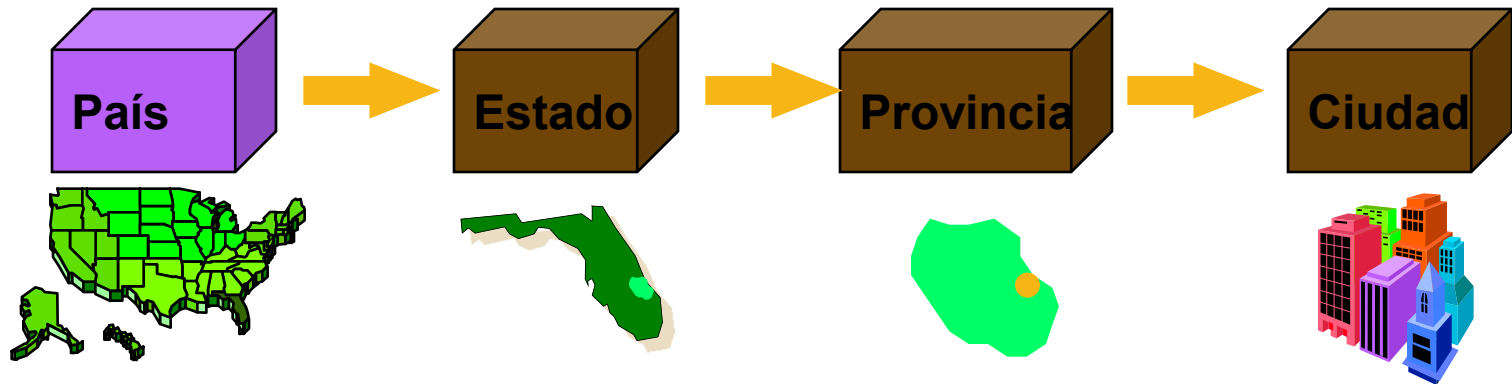
- Modelado lógico multidimensional (MD)
- Modelado multidimensional relacional
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ Resumen/comparativa final
  - ▣ La dimensión tiempo
  - ▣ Otras consideraciones de diseño
- El proceso de diseño de un DW



# Modelado multidimensional

## Esquema de copos de nieve

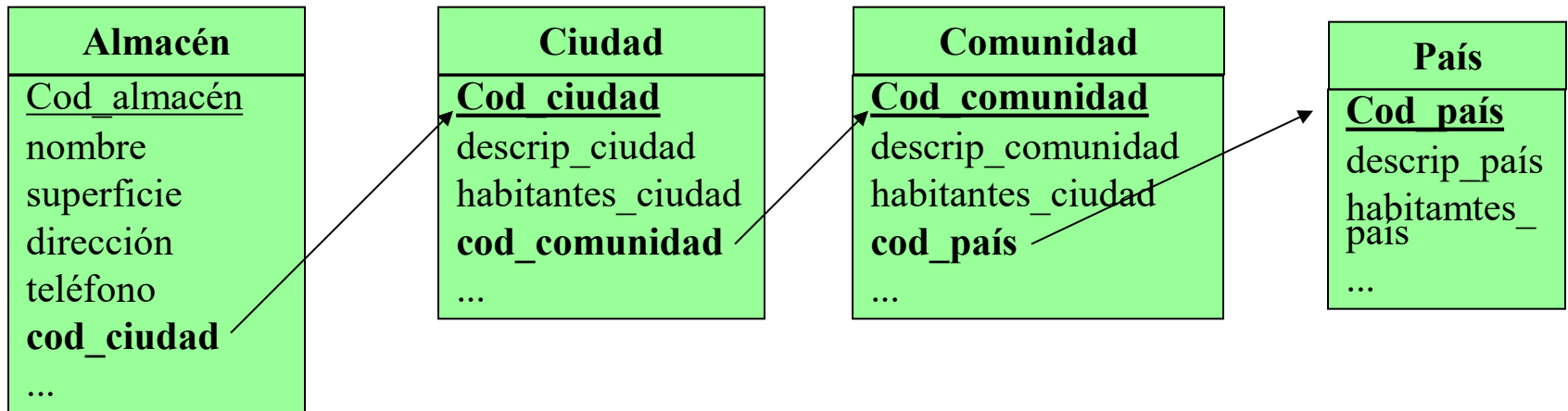
- Las dimensiones se normalizan con los niveles de jerarquía de dimensiones
  - Tabla dimensión → valores del mínimo nivel de jerarquía



# Modelado multidimensional

## Esquema de copos de nieve

58



# Modelado multidimensional

## Esquema de copos de nieve

59

### □ Ventajas

- ▣ Podría salvar espacio en disco, pero no demasiado
- ▣ Consultas (browsing) de atributos simples es rápido
- ▣ Mejora el rendimiento cuando mayoría de requisitos solicitan niveles de agregación superiores
  - Disminuye el tamaño de tablas a escanear

# Modelado multidimensional

Esquema de copos de nieve

60

## □ Inconvenientes

- ▣ Aumenta el número de tablas → aumenta el número de uniones (join)
  - Algunos requisitos pueden demorarse en exceso
- ▣ Consultar (browsing) atributos de más de una tabla es más lento
- ▣ Demasiado complejo para que los usuarios finales definan sus propias consultas ad-hoc a partir de ellos

# Modelado multidimensional

Esquema de copos de nieve

61

## □ Inconvenientes

- ▣ Requiere una clave primaria más por cada nivel de jerarquía normalizado
- ▣ Aumenta la complejidad de diseño y mantenimiento
- ▣ No soportado por todas las herramientas del mercado

# Modelado multidimensional

Esquema de copos de nieve

62

## □ Recomendaciones

### ▣ Normalmente no se recomienda

- Realmente cuando el espacio en disco es un problema

### ▣ Normalmente se recomienda normalizar una o dos de las dimensiones más grandes

- Existen un gran número de filas

### ▣ Suele aplicarse cuando muchos atributos caracterizan a los niveles más altos de las jerarquías

# Modelado multidimensional

## Esquema de copos de nieve

63

### □ Recomendaciones

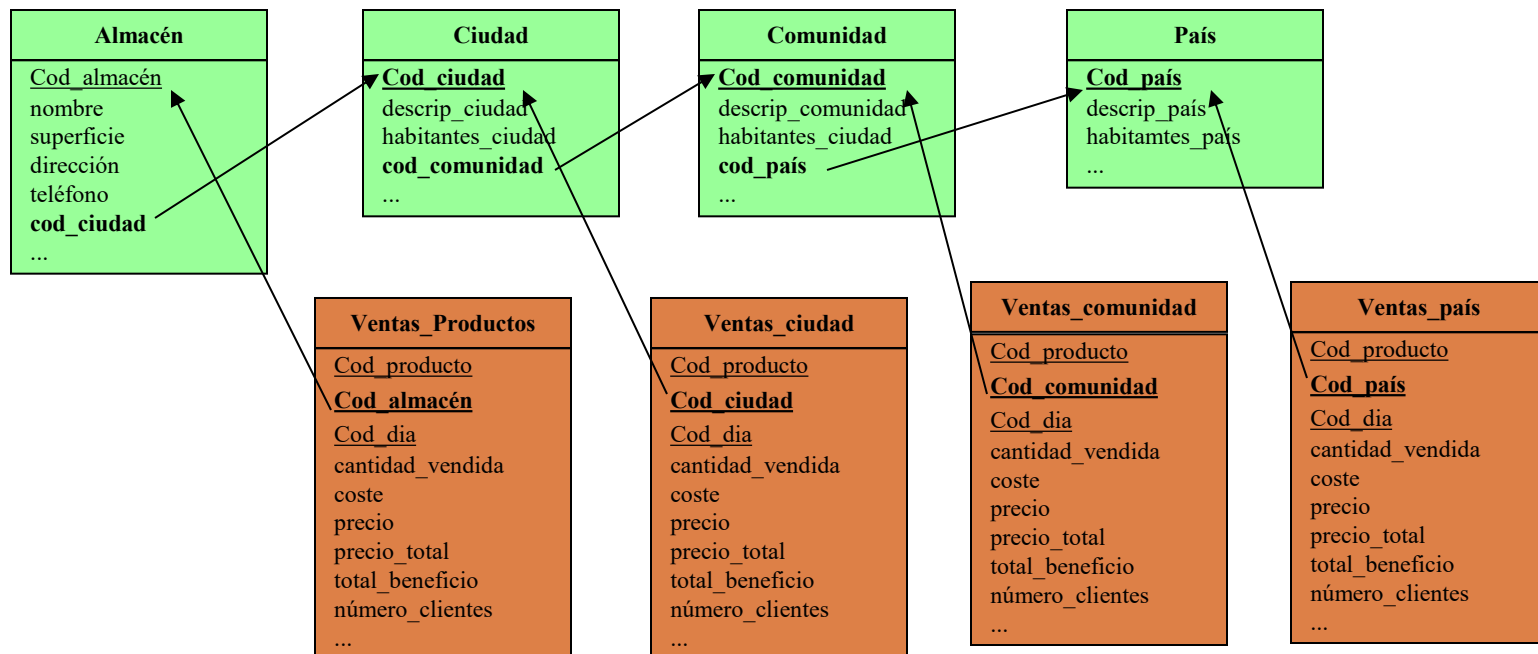
- ▣ Cuando se pueden utilizar índices bitmap para los atributos de granularidad mínima
- ▣ Utilizar sólo cuando las ventajas son muy explícitas:
  - Ahorro en disco significativo
  - Muchos atributos en los niveles más altos de jerarquías
- ▣ Estadísticamente, el espacio en disco ahorrado utilizando *snowflake schemas* es del 1% del espacio total en disco

# Modelado multidimensional

Esquema de copos de nieve y constelaciones de hechos

64

- Utilizar copos de nieve con constelaciones de hechos (agregados)





# Diseño lógico de DW

65

- Modelado lógico multidimensional (MD)
- **Modelado multidimensional relacional**
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ Resumen/comparativa final
  - ▣ La dimensión tiempo
  - ▣ Otras consideraciones de diseño
- El proceso de diseño de un DW

# Diseño lógico de DW

66

- Modelado lógico multidimensional (MD)
- **Modelado multidimensional relacional**
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ **Resumen/comparativa final**
  - ▣ La dimensión tiempo
  - ▣ Otras consideraciones de diseño
- El proceso de diseño de un DW

# Modelado multidimensional

Resumen/comparativa final del esquema estrella y sus variantes

67

## □ Esquema estrella

### ▣ Ventajas

- Fácil de entender por los usuarios
- Reduce número de uniones físicas
  - Respuestas rápidas para la mayoría de las consultas
  - Fácil de entender por usuarios finales
- Metadatos sencillos
- Soportado por la inmensa mayoría de aplicaciones

# Modelado multidimensional

Resumen/comparativa final del esquema estrella y sus variantes

68

## □ Esquema estrella

### ▣ Inconvenientes

- El aumento del tamaño de la tabla de hechos con datos agregados puede empeorar el rendimiento general
  - Por ello se recomienda tablas de hechos agregados al margen
- Las dimensiones tienen un tamaño enorme
  - Alrededor de 50 atributos (*Kimball*)
- La necesidad de manejar el atributo de nivel para datos agregados en algunos sistemas
- Es poco robusto o susceptible a cambios
- Más lento de construir

# Modelado multidimensional

Resumen/comparativa final del esquema estrella y sus variantes

69

## □ Esquema constelaciones de hechos

### □ Ventajas

- No se necesita el atributo de nivel en las tablas de dimensiones que lo utilizan explícitamente
- Los datos agregados no se almacenan con los del nivel de detalle más bajo en las tablas de hechos
  - Aumenta el rendimiento al hacer browse de datos almacenados precalculados (sumados)

# Modelado multidimensional

Resumen/comparativa final del esquema estrella y sus variantes

70

## □ Esquema constelaciones de hechos

### □ Inconvenientes

- Un gran número de tablas de agregados
  - Complica el mantenimiento de los metadatos
  - Necesidad del usuario de conocer la existencia de dichas tablas
- Algunos requisitos pueden necesitar consultar datos de varias tablas y mermar el rendimiento general del sistema

# Modelado multidimensional

Resumen/comparativa final del esquema estrella y sus variantes

71

## □ Esquema copos de nieve

### ▣ Ventajas

- No se necesita el atributo de nivel en los sistemas que lo gestionan explícitamente
- Más flexible y adecuado para requerimientos
- Carga de datos (ETL) más rápida y sencilla

# Modelado multidimensional

Resumen/comparativa final del esquema estrella y sus variantes

72

## □ Esquema copos de nieve

### □ Ventajas

- Directamente implementados por algunas jerarquías
  - No Todas
- Mejora considerablemente el rendimiento cuando un gran número de requisitos solicita datos agregados o de niveles superiores de jerarquías
  - Los requerimientos escanean un reducido número de filas



# Modelado multidimensional

Resumen/comparativa final del esquema estrella y sus variantes

73

## □ Esquema copos de nieve

### ▣ Inconvenientes

- Aumenta la complejidad de mantener los metadatos
  - Debido al aumento del número de tablas
- Si no se dispone de la suficiente cantidad de tablas de agregados, el rendimiento general podría disminuir

# Diseño lógico de DW

74

- Modelado lógico multidimensional (MD)
- **Modelado multidimensional relacional**
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ Resumen/comparativa final
  - ▣ La dimensión tiempo
  - ▣ Otras consideraciones de diseño
- El proceso de diseño de un DW

# Diseño lógico de DW

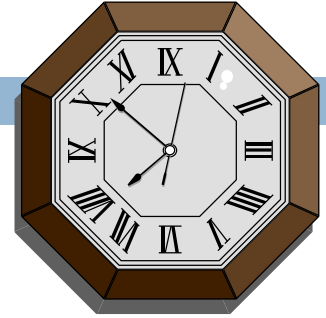
75

- Modelado lógico multidimensional (MD)
- **Modelado multidimensional relacional**
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ Resumen/comparativa final
  - ▣ **La dimensión tiempo**
  - ▣ Otras consideraciones de diseño
- El proceso de diseño de un DW

# Modelado multidimensional

## La dimensión tiempo

76

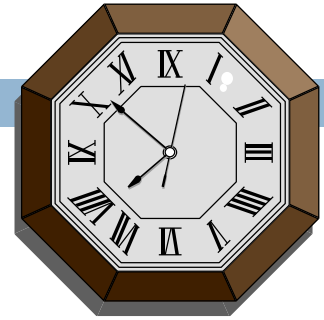


- La dimensión tiempo
  - Es obligatoria
  - Es única
  - Es potente
  - Necesita ser diseñada cuidadosamente
  
- Incluye períodos del tiempo de negocio
  
- Incluye fechas especiales

# Modelado multidimensional

## La dimensión tiempo

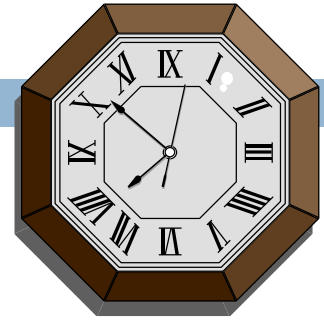
77



- ▣ Analizar y planificarla
- ▣ Considerar rango de fechas relevante
- ▣ Identificar la granularidad
- ▣ Utilizar las características de forma uniforme para navegar por la dimensión
- ▣ Modelar de forma flexible

# Modelado multidimensional

## La dimensión tiempo

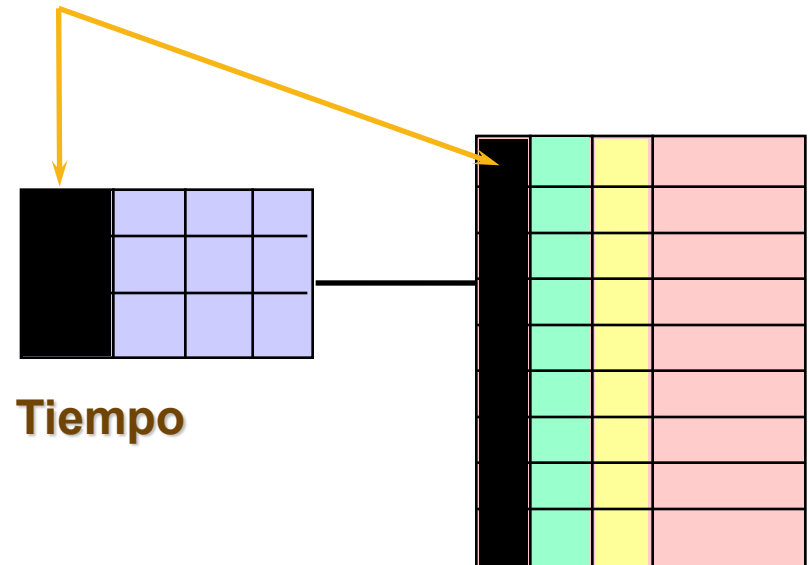


### ■ Tabla de dimensión tiempo

- Día laborable
- Período fiscal
- Principal evento
- Mes
- Vacaciones

**Clave de la Tabla tiempo**

**Una columna simple: generl. auto generada**



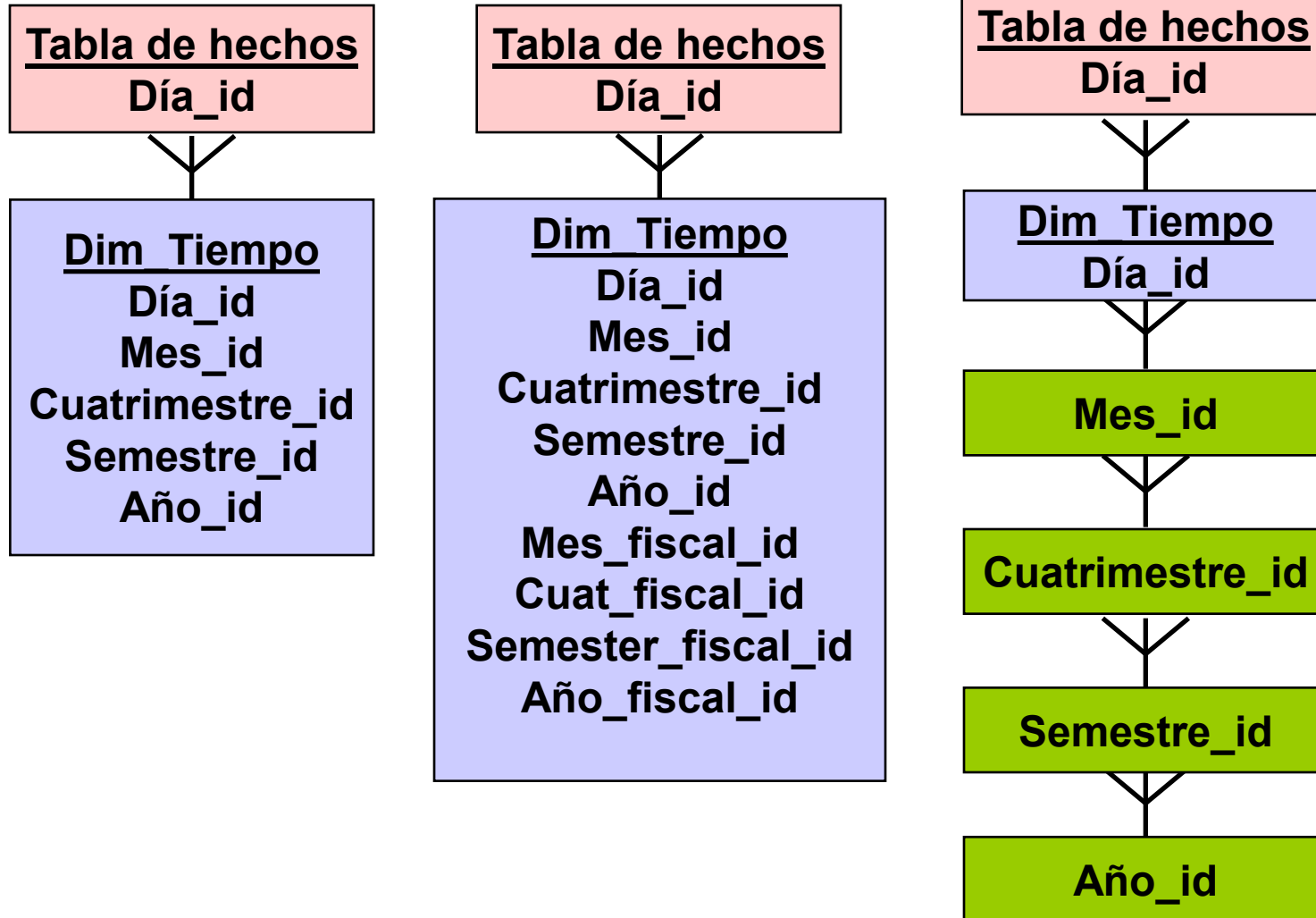
- Permite un análisis más flexible

# Modelado multidimensional

La dimensión tiempo

■ Normalizada

79



# Modelado multidimensional

## La dimensión tiempo

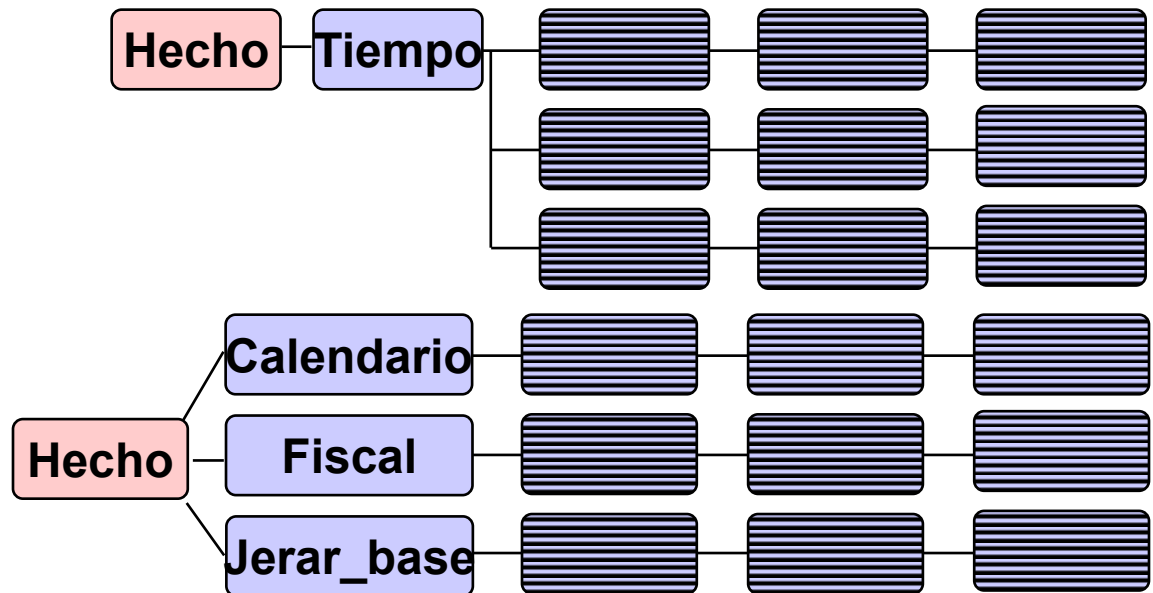
80

- Identificar requerimientos fecha

- Normales
- Fiscales
- Jerarquías

- Crear ...

- Jerarquías múltiples
- Dimensiones independientes
- Vistas





# Modelado multidimensional

## La dimensión tiempo

81

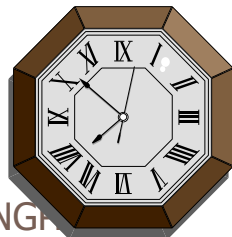
### □ Tipo de consultas sobre dimensión tiempo

#### ■ Absolutas

- Ej. Ventas desde julio 1999 hasta julio 2000

#### ■ Relativas

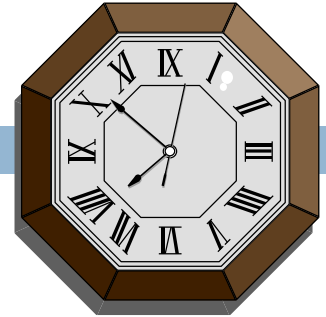
- Ej. Obtener ventas para el mes actual del año pasado



# Modelado multidimensional

## La dimensión tiempo

82

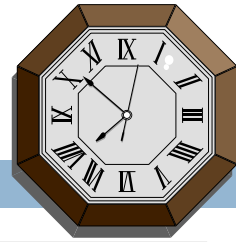


### ■ Algunos sistemas de forma interna

- Añaden un atributo para indicar período actual y el previo
  - Suele hacerse en el mínimo nivel de granularidad
  - Aunque hay que definir muy bien los periodos de comparación
- Proporcionar códigos de tiempo secuencial que identifiquen a los elementos de las jerarquías

# Modelado multidimensional

## La dimensión tiempo



83

| Cod Dia | Día | Cod Mes | Mes   | Cod Semestre | Semestre | Cod Año | Año  | Periodo actual | Agg level |
|---------|-----|---------|-------|--------------|----------|---------|------|----------------|-----------|
| 1300    | 2   | 500     | Sept  | 200          | S3       | 50      | 1999 |                | 1         |
| 1301    | 3   | 500     | Sept  | 200          | S3       | 50      | 1999 | <b>P</b>       | 1         |
| ...     | ... | ...     | ...   | ...          | ...      | ...     | ...  | ...            | ...       |
| ...     | ... | ...     | ...   | ...          | ...      | ...     | ...  | ...            | ...       |
| 1420    | 1   | 504     | Enero | 202          | S1       | 51      | 2000 |                | 1         |
| 1421    | 2   | 504     | Enero | 202          | S1       | 51      | 2000 |                | 1         |
| ...     | ... | ...     | ...   | ...          | ...      | ...     | ...  | ...            | ...       |
| 1665    | 2   | 512     | Sept  | 204          | S3       | 51      | 2000 |                | 1         |
| 1666    | 3   | 512     | Sept  | 204          | S3       | 51      | 2000 | <b>S</b>       | 1         |

# Diseño lógico de DW

84

- Modelado lógico multidimensional (MD)
- **Modelado multidimensional relacional**
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ Resumen/comparativa final
  - ▣ La dimensión tiempo
  - ▣ Otras consideraciones de diseño
- El proceso de diseño de un DW

# Diseño lógico de DW

85

- Modelado lógico multidimensional (MD)
- **Modelado multidimensional relacional**
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ Resumen/comparativa final
  - ▣ La dimensión tiempo
  - ▣ **Otras consideraciones de diseño**
- El proceso de diseño de un DW

# Modelado multidimensional

## Otras consideraciones de diseño

86

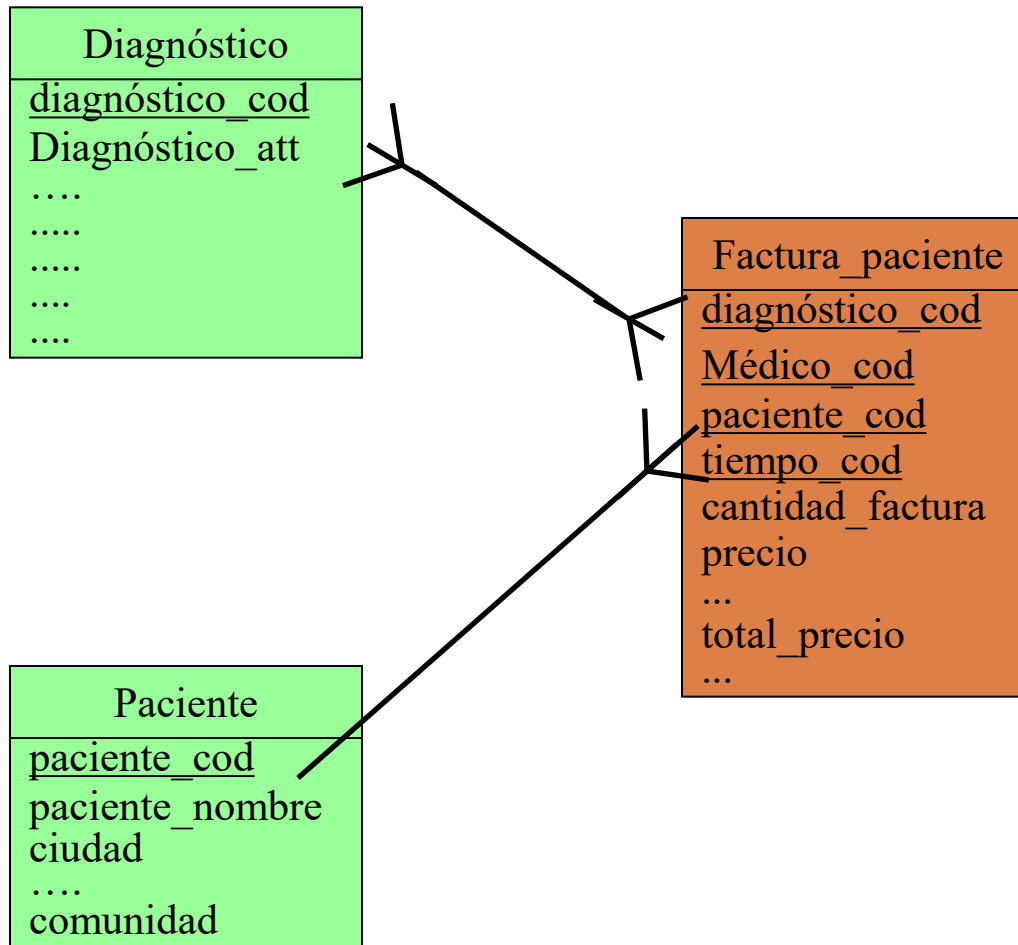
### □ *Many-to-many dimensions*

- ▣ Cuando entre la tabla de hechos y una tabla de dimensión existe una relación de muchos a muchos (m-n)
- ▣ Ejemplo: Health Care System
  - Un paciente puede tener varios diagnósticos
  - ¿Qué hacer cuando un paciente tiene varios diagnósticos?

# Modelado multidimensional

## Otras consideraciones de diseño

87



# Modelado multidimensional

## Otras consideraciones de diseño

88

### □ *Many-to-many dimensions*

#### ▣ Problemas

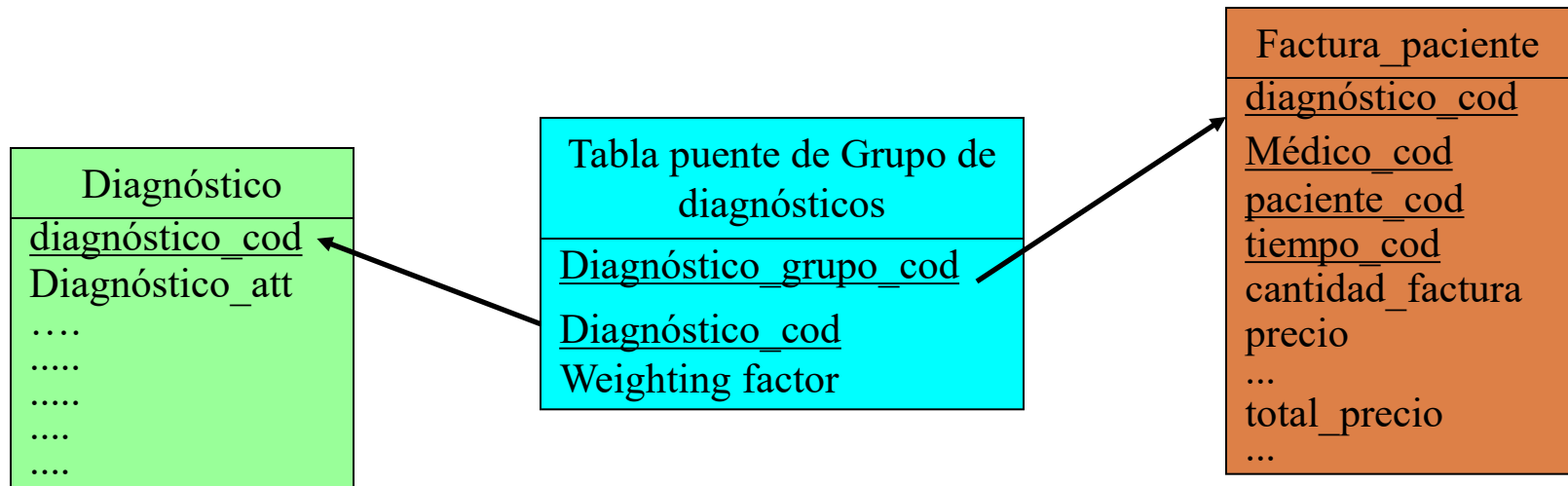
- Sólo se podría visualizar uno de forma sencilla
- Para visualizar todos, la generación del informe es más compleja
  - Necesidad de más uniones entre tablas
- Se suele utilizar los denominados *weighting factors*
- Kimball propone el uso de tablas puente (*bridge tables*)



# Modelado multidimensional

## Otras consideraciones de diseño

89



# Modelado multidimensional

## Otras consideraciones de diseño

90

- *The Data Warehouse Bus Architecture*
  - ▣ Arquitectura detallada de cómo construir todo el DW corporativo
  - ▣ Cada Data Mart debería estar basado en la mínima granularidad del proceso de negocio que contempla
  - ▣ Dos aproximaciones:
    - Crear una arquitectura que defina todo el DW
    - Supervisar la construcción de cada Data Mart en particular
  - ▣ Utilizar las dimensiones y hechos comunes (*conformed*)

# Modelado multidimensional

## Otras consideraciones de diseño

91

- Dimensiones comunes (*Conformed dimensions*)
  - ▣ Es una misma dimensión que existe en todos los DM
    - Ejemplo. Tabla maestro de clientes o productos
      - Tablas que se mantienen de forma independiente a los DM
  - ▣ Cuando se definen dimensiones comunes
    - No obviarlas ni menospreciarlas
    - Definir el mínimo nivel de granularidad posible
    - Utilizar claves auto generadas (*surrogate keys*)
  - ▣ Este tipo de dimensiones puede ocupar el 80% del esfuerzo del desarrollo total

# Modelado multidimensional

## Otras consideraciones de diseño

92

### □ Dimensiones comunes (*Conformed dimensions*)

#### ▣ Ventajas

- Una misma dimensión se puede utilizar contra varios hechos
- Interfaces de usuario y datos son consistentes
- Permite navegar entre Data Marts (DM)

### □ Hechos comunes (*Conformed facts*)

#### ▣ Medidas utilizadas en más de un DM

- Ejemplos: beneficio, coste, precio etc.

# Diseño lógico de DW

93

- Modelado lógico multidimensional (MD)
- **Modelado multidimensional relacional**
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ Resumen/comparativa final
  - ▣ La dimensión tiempo
  - ▣ Otras consideraciones de diseño
- El proceso de diseño de un DW

# Diseño lógico de DW

94

- Modelado lógico multidimensional (MD)
- Modelado multidimensional relacional
  - ▣ Esquema estrella
  - ▣ Esquema constelaciones de hecho
  - ▣ Esquema copos de nieve
  - ▣ Resumen/comparativa final
  - ▣ La dimensión tiempo
  - ▣ Otras consideraciones de diseño
- El proceso de diseño de un DW

# Modelado multidimensional

El proceso de modelado de un almacén de datos

95

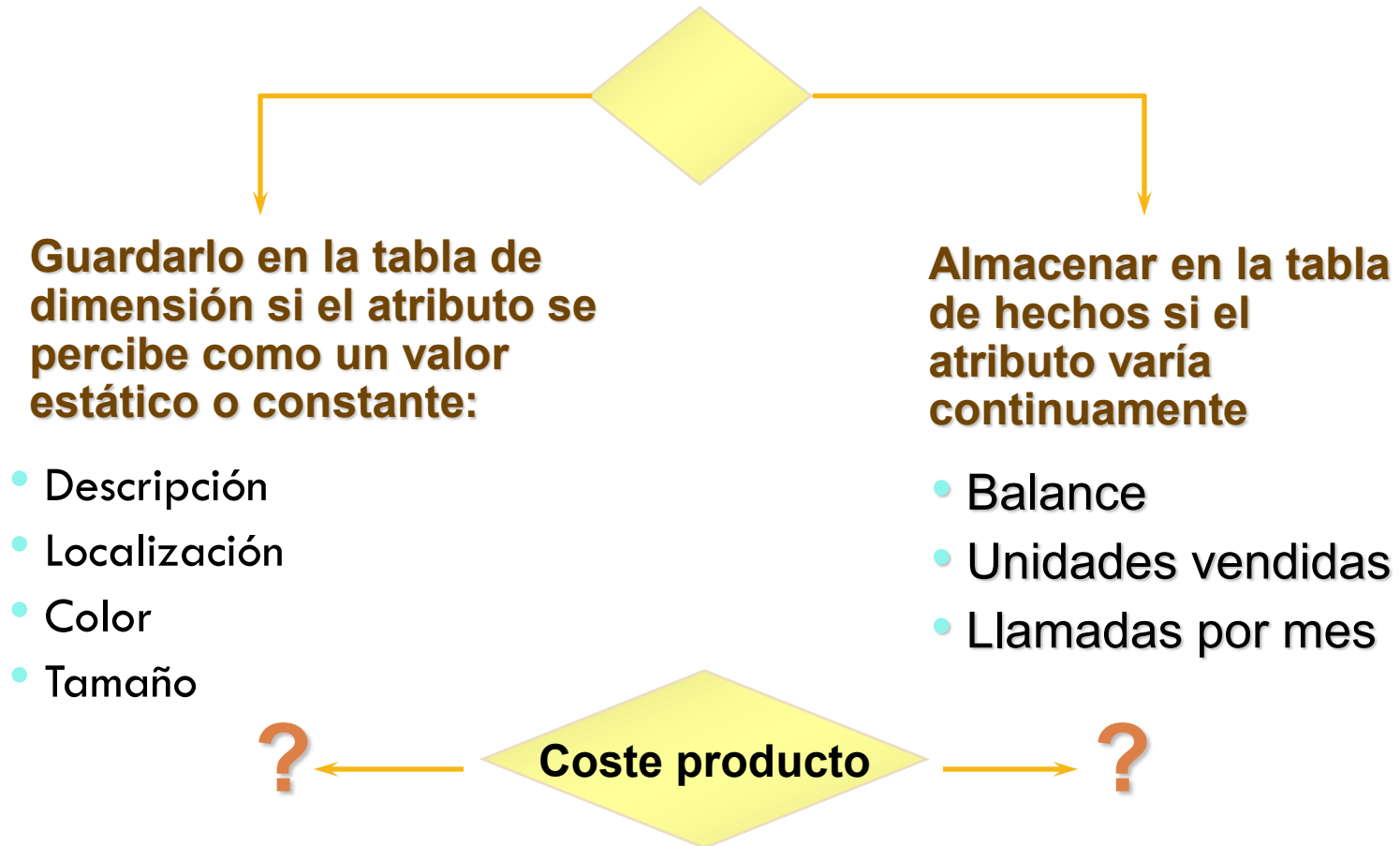
- ▣ A partir de análisis de requisitos
  - Objeto de análisis → hechos
  - Contexto de análisis → dimensiones
    - Jerarquías, etc...
  - A veces no es intuitivo y se adoptarán decisiones sobre si considerar un elemento como dimensión o como hecho

# Modelado multidimensional

El proceso de modelado de un DW: ¿¿¿ Hechos o dimensiones ????

96

## Unidades vendidas – Llamadas en un mes - Color





# Modelado multidimensional

El proceso de modelado de un DW

97

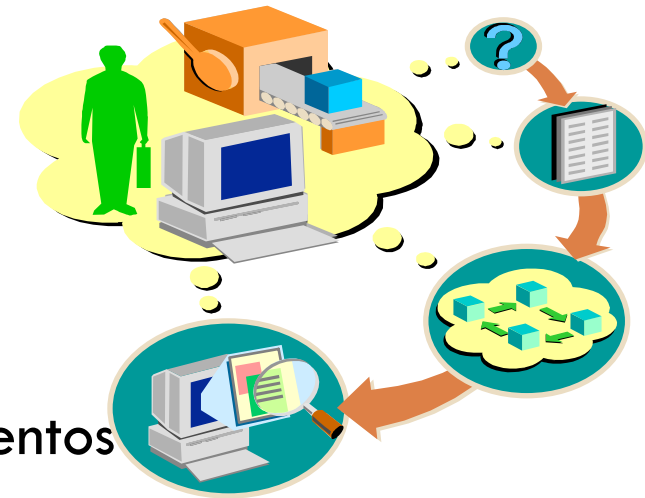
- ¿?? Diseño conceptual ???
  - Algunos diseñadores lo prefieren por la documentación
- Diseño lógico a partir del modelo corporativo/transac.
  - Extraer de la documentación o modelo → esquema estrella
    - Tablas de hechos
    - Tablas de dimensiones
      - Jerarquías de clasificación
      - Atributos
  - En función de los requisitos más comunes decidir
    - Normalizar algunas dimensiones → copos de nieve
    - Estudiar qué agregados definir → constelaciones de hechos

# Modelado multidimensional

El proceso de modelado de un DW

98

- ▣ Determinar los requerimientos de usuario
- ▣ Asistir a los usuarios en la comprensión de la tecnología
- ▣ Casi necesario...
  - ▣ Definir requerimientos analíticos
  - ▣ Objetivos de rendimiento
- ▣ Mantenerse informado con los cambios de requerimientos
- ▣ Proporcionar soporte para más requerimientos que los detectados inicialmente

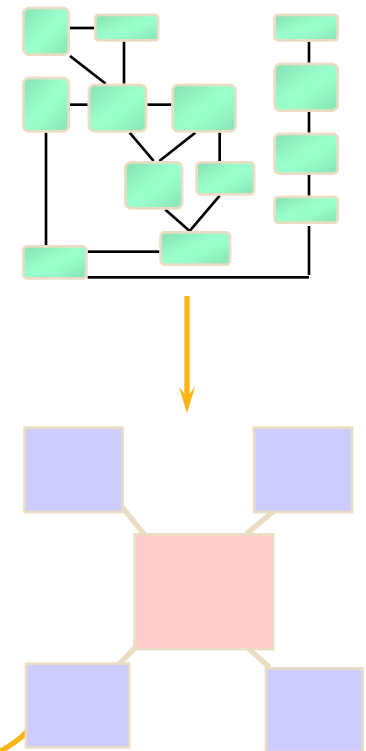
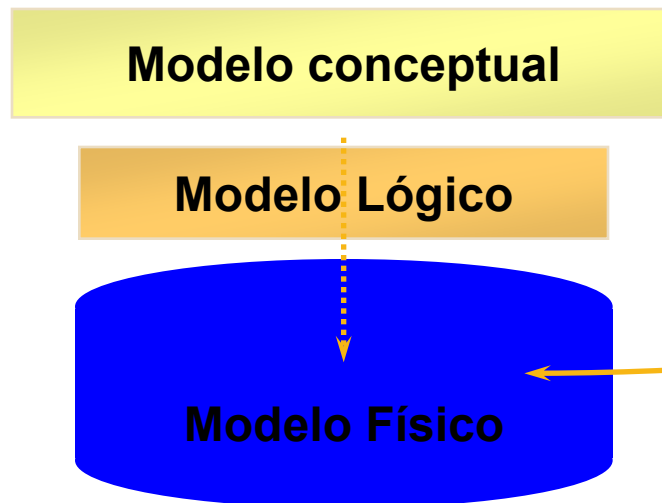


# Modelado multidimensional

El proceso de modelado de un DW

99

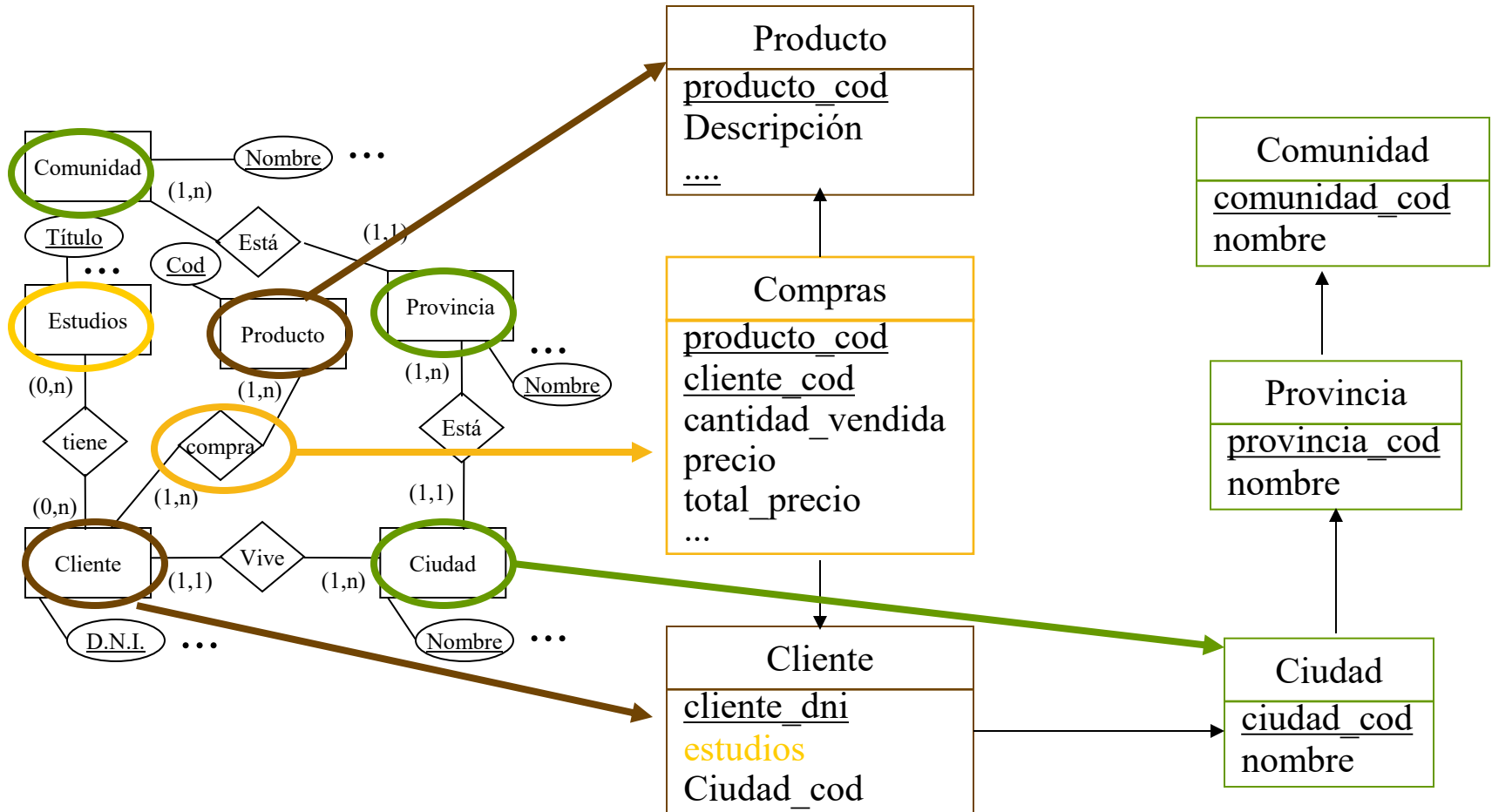
- ▣ Crear un modelo de negocio
  - Opcional
- ▣ Crear un modelo dimensional
- ▣ Crear un modelo físico



# Modelado multidimensional

El proceso de modelado de un DW. Modelo EER → Esq. Estrella

100



## Bibliografía

101

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

# Diseño lógico de almacenes de datos

## Tema 3

### Profesores:

Juan C. Trujillo  
Alejandro Reina Reina  
LUCENTIA Research Group



Universitat d'Alacant  
Universidad de Alicante



Departamento de  
Lenguajes y Sistemas  
Informáticos