

# TEMA 1. FUNDAMENTOS DEL MODELO DE DATOS MULTIDIMENSIONAL

Natalia Padilla-Zea, Eladio Garv, Jos Samos

# 1.0 Introducción

2

- ¿Qué vamos a hacer en esta asignatura?
  - ▣ Sistemas para tomar decisiones
  
- ¿Cómo se toman las decisiones?
  - ▣ Informes
  
- Informe ← Fuentes + procesamiento
  - ▣ El usuario solicita un informe... ¿quién lo hace? SQL, DI

# 1.0 Introducción

3

- Ejemplo: Biblioteca universitaria
  - ▣ BD de catálogos y préstamos
  - ▣ Director/a: libros que tiene que comprar, organizarlos
  - ▣ Vicerrector/a: asignación de recursos a los centros
  - ▣ Rector/a: presupuesto global, visibilidad de publicaciones

# 1.0 Introducción

4

- Problemas que nos encontramos:
  - ▣ Un nombre que representa cosas distintas en distintas fuentes: ¿qué tiene? ¿a qué se refiere?
  - ▣ Dificultad de reutilización: el tratamiento de datos pierde el significado original
  - ▣ Implica alto coste y retraso temporal
  - ▣ Falta de confianza: errores en la interpretación de los datos (DI)
  - ▣ Necesidad de personal “del contexto” que revise los informes

# 1.0 Introducción

5

- Solución propuesta:
  - ▣ Tener BD orientadas a los informes
  - ▣ Diseño para el tipo de consultas concreto
  - ▣ Integrando distintas fuentes
  - ▣ Disminuye la complejidad, la demora y el coste de la generación de informes

# 1.0 Introducción

6

- Enfoques a considerar:
  - ▣ Sistema transaccional: modelo de datos E-R, OO...
  - ▣ Sistema multidimensional: modelo de datos multidimensional → datos, consultas, informes

# 1.1 Principios de los sistemas transaccionales

7

- Sistema transaccional: orientado a transacciones
  - ▣ Las BD que conocemos: similar frecuencia de consulta y actualización
- Modelo inicial a utilizar :
  - ▣ Modelo de datos E-R, OO...

# 1.1 Principios de los sistemas transaccionales

8

- Caso hipotético:
  - ▣ Hacer el modelo E-R para la biblioteca de la UGR, de tal forma que podamos almacenar:
    - Las distintas bibliotecas que hay en la UGR: campus Aynadamar, campus Ceuta, campus Melilla...
    - El personal que trabaja en la biblioteca
    - Los usuarios de la biblioteca
    - Los libros de que se dispone
    - Los préstamos de libros a los distintos usuarios
    - Los tipos de usuarios



# 1.1 Principios de los sistemas transaccionales

9

- ¿Qué consultas se pueden hacer sobre el modelo anterior?
  - ▣ Libros prestados más de 3 veces al mes
  - ▣ N° de préstamos en una franja horaria
  - ▣ Comparación entre meses
  - ▣ Comparación entre tipos de usuarios...

# 1.1 Principios de los sistemas transaccionales

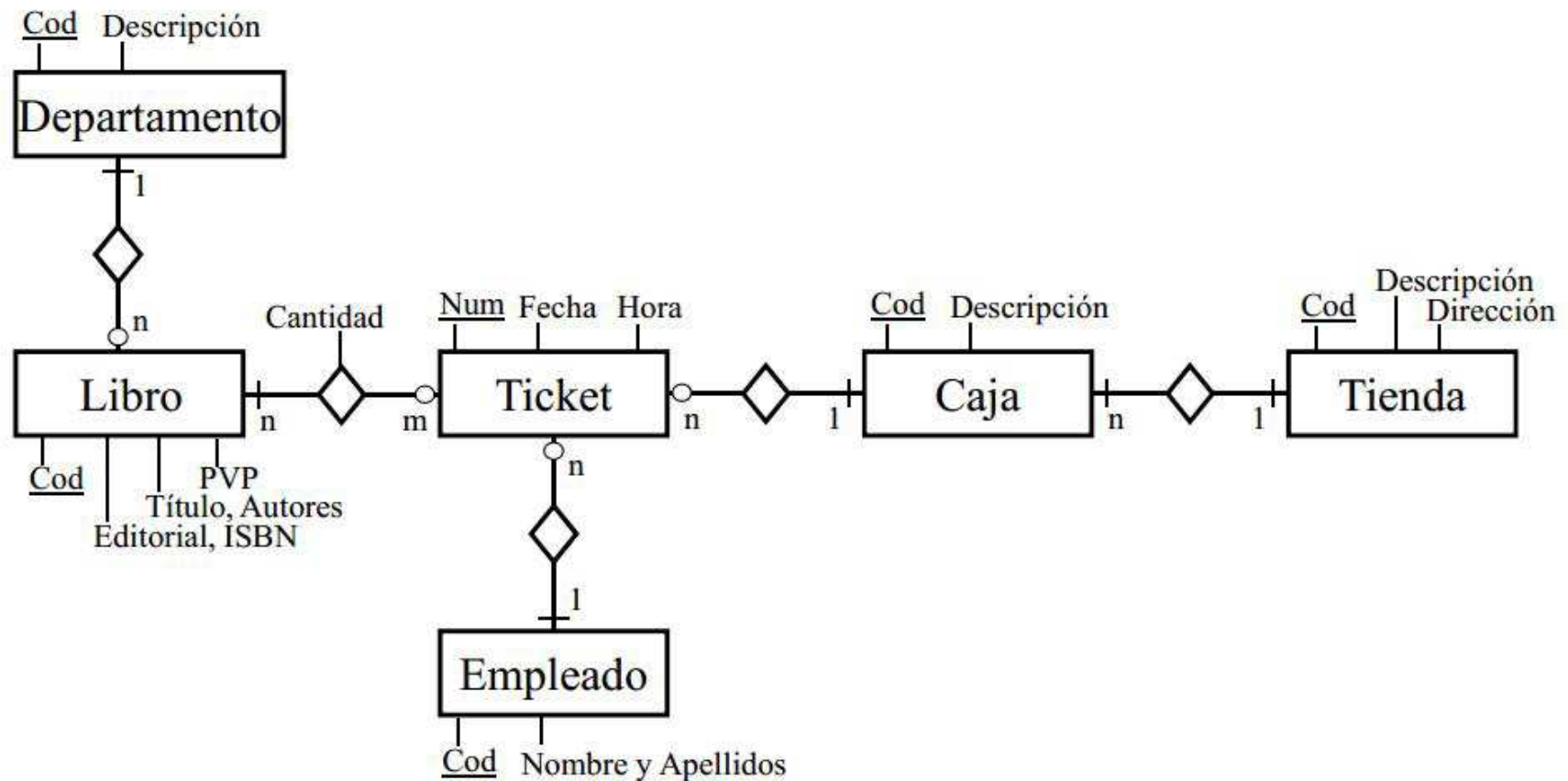
10

- ¿Y si se plantean modificaciones?
  - ▣ La biblioteca de la UGR quiere poder modificar el tiempo que le presta los libros a sus usuarios.
  - ▣ Posibles soluciones:
    - En la tabla de usuarios, añadir el tiempo de préstamo
    - Crear una tabla nueva que, según el tipo de usuario, indique el tiempo y el número de libros que se prestan

# 1.1 Principios de los sistemas transaccionales

11

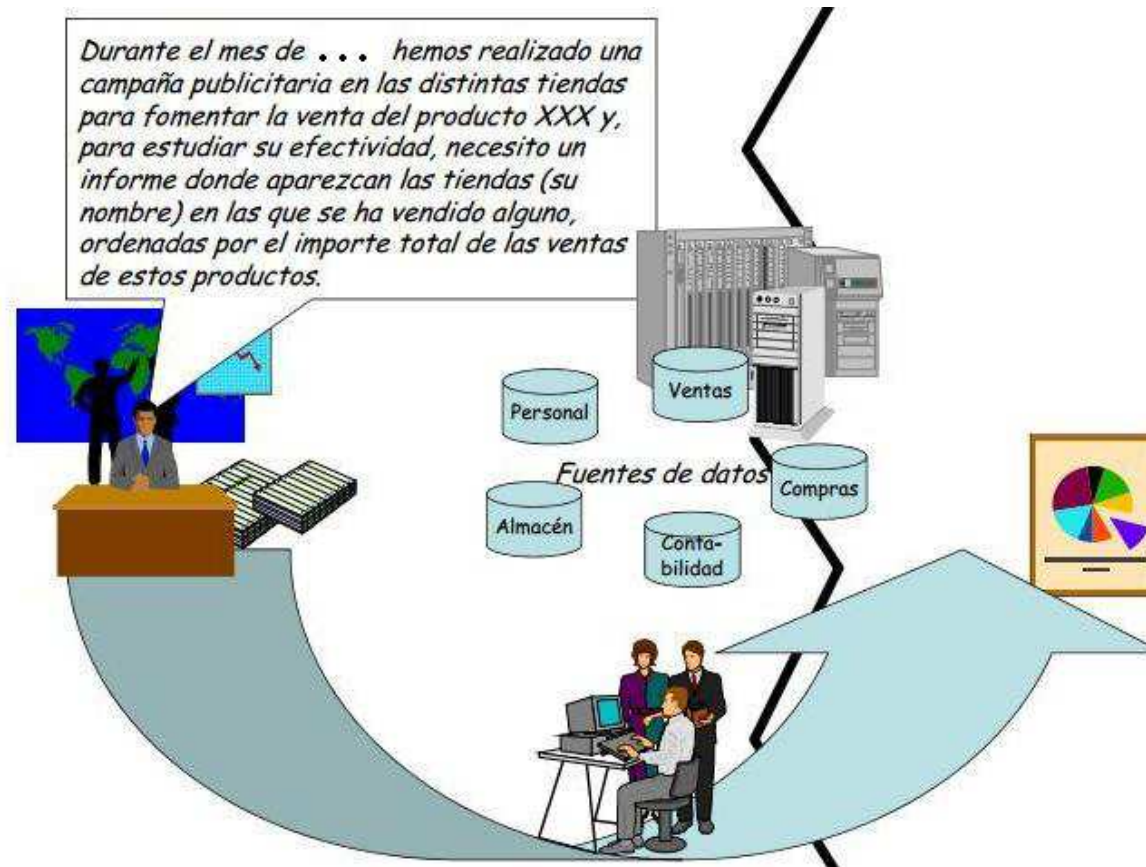
- Caso de referencia: sistema de gestión de ventas en librerías (modelo conceptual)



# 1.1 Principios de los sistemas transaccionales

12

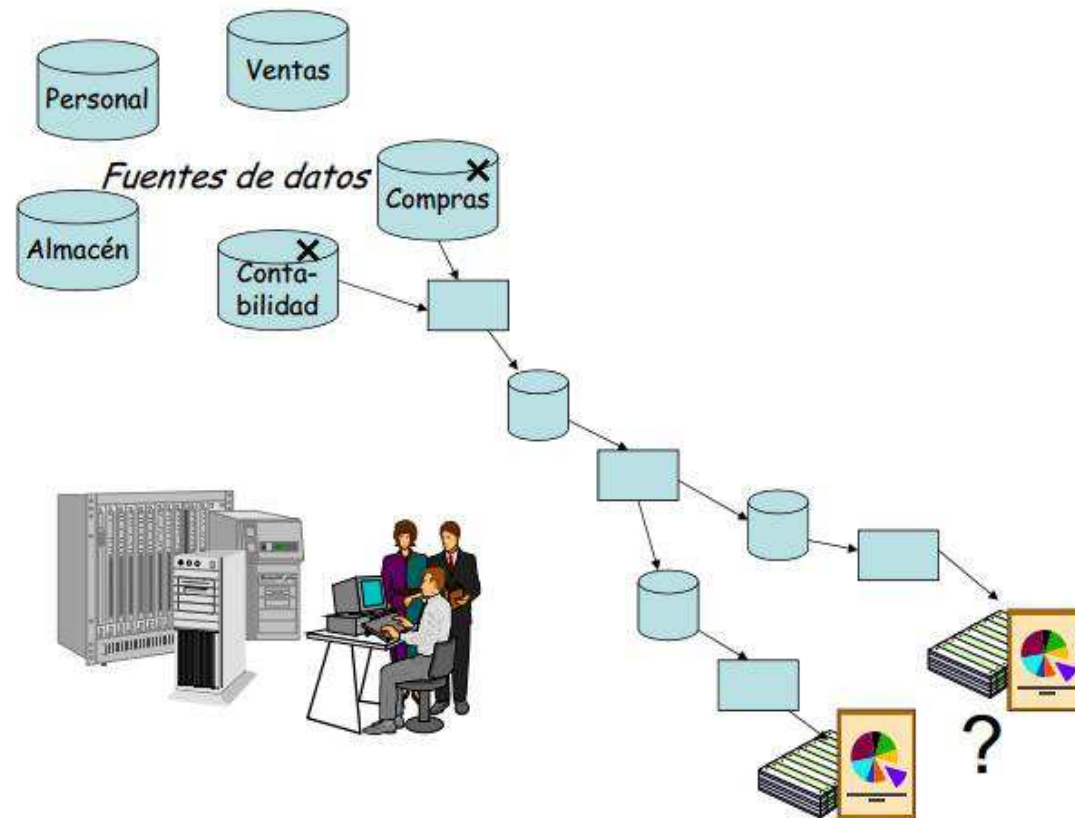
- Caso de referencia: sistema de gestión de ventas en librerías (petición de un nuevo informe)



# 1.1 Principios de los sistemas transaccionales

13

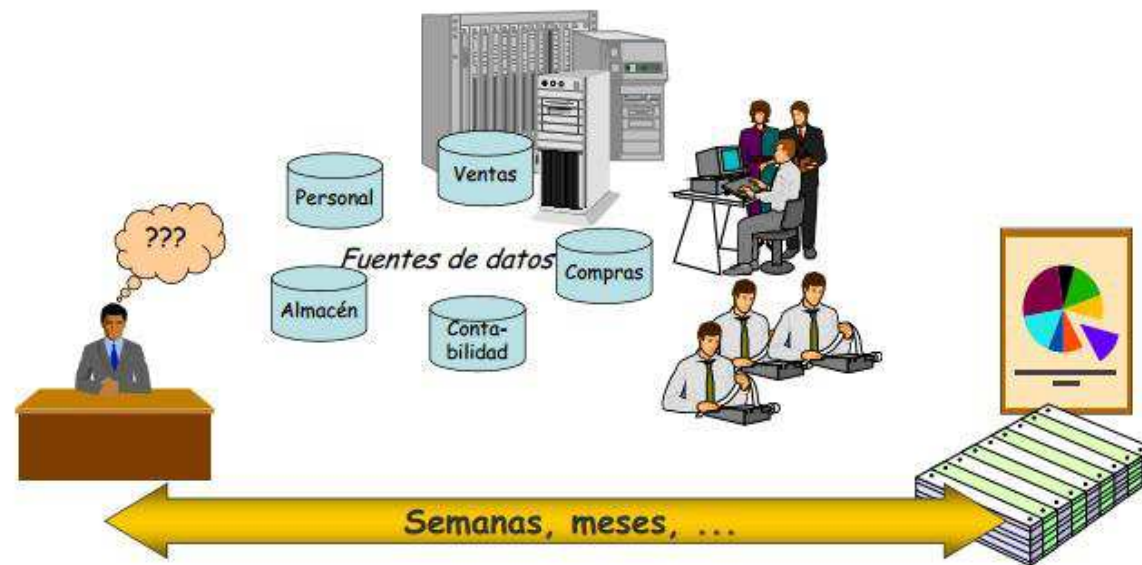
- Caso de referencia: sistema de gestión de ventas en librerías (pasos para obtener el nuevo informe)



# 1.1 Principios de los sistemas transaccionales

14

- Caso de referencia: sistema de gestión de ventas en librerías (dificultades en la obtención del informe)



# 1.1 Principios de los sistemas transaccionales

15

- Al pasar a tablas hay que normalizar:
  - ▣ Que cada dato esté en la BD una sola vez (integridad)
    - Evita errores en las actualizaciones
  - ▣ Eliminar redundancias
  - ▣ Hacer las modificaciones más simples y rápidas
  - ▣ Facilitan los JOIN para hacer consultas

# 1.1 Principios de los sistemas transaccionales

16

- Objetivos de los sistemas transaccionales:
  - ▣ No tener duplicados
  - ▣ Evitar anomalías en las modificaciones
  - ▣ Optimizar las modificaciones de los datos
  - ▣ Registrar transacciones



# 1.1 Principios de los sistemas transaccionales

17

- Tipos de informes:
  - ▣ Fijos: se resuelven con informes predeterminados o parametrizables
  - ▣ A medida:
    - ¿Con informes predeterminados?
    - ¿Pidiéndolos al DI? El DI deberá realizar los pasos de:
      - Localización de fuentes
      - Extracción de datos: consultas SQL, programas al efecto...
      - Transformación e integración de datos: más programas  
Ejemplo → Hombre, Mujer, H, M, 0, 1
      - Carga de datos para los usuarios: más programas
      - Presentación de datos de forma inteligible: más programas
  - Requiere tiempo y revisiones

# 1.1 Principios de los sistemas transaccionales

18

- Pasos en la construcción de los sistemas transaccionales:
  - ▣ Modelado conceptual: E/R, OO...
  - ▣ Modelado lógico: relacional, objeto-relacional...
  - ▣ Modelado físico: índices, archivos...
  - ▣ Implementación: Oracle, Access, MySQL...

# 1.2 Principios de los sistemas multidimensionales

19

- Buscamos alternativas a las dificultades que presentan los sistemas transaccionales
- Un sistema multidimensional es un modelo heterogéneo: las tablas tienen distinta importancia
- El objetivo ahora es hacer consultas e informes:
  - ▣ Queremos que sirvan para tomar decisiones
  - ▣ No consideramos actualizaciones

# 1.2 Principios de los sistemas multidimensionales

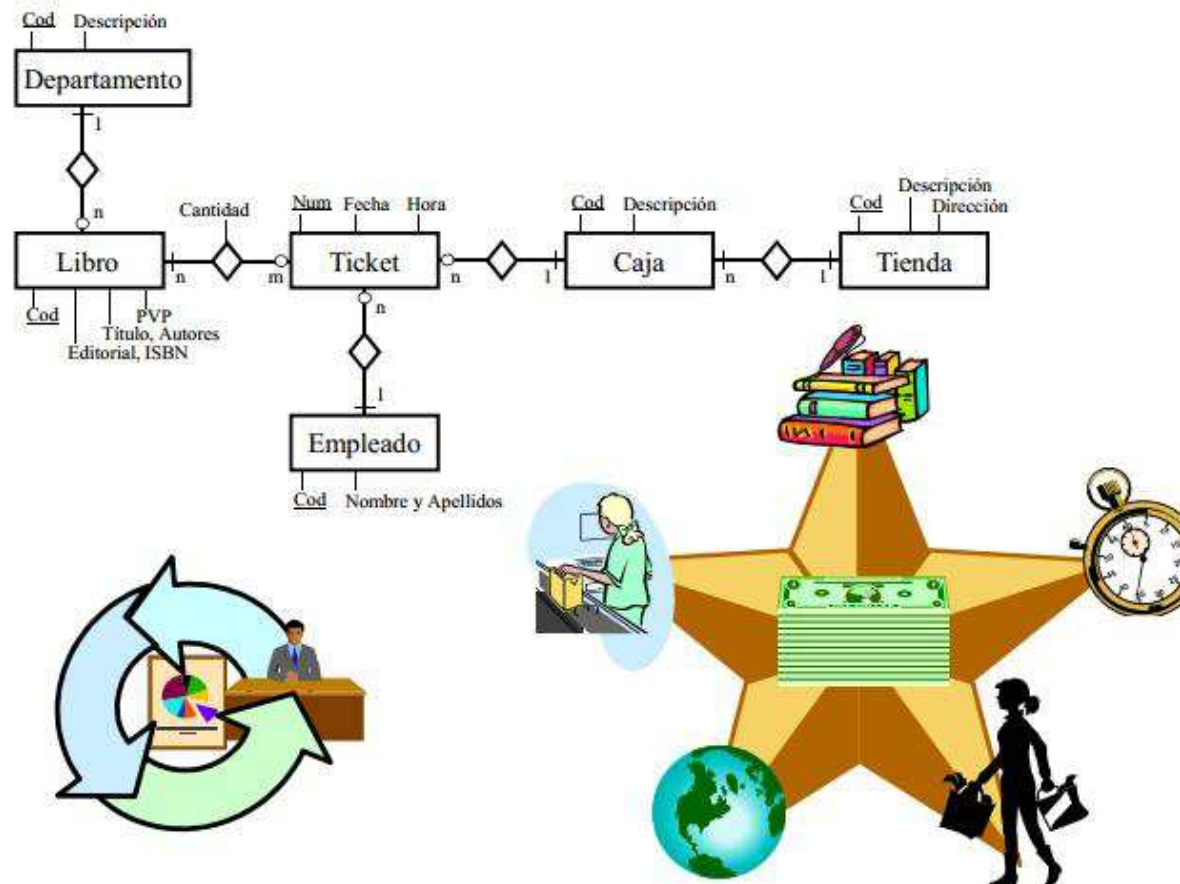
20

- Sistemas Multidimensionales:
  - ▣ Seleccionamos información que pasa a niveles superiores de la organización o empresa
  - ▣ Partimos de un conjunto de datos elementales: sistemas transaccionales
  - ▣ Construiremos sistemas de inferencia con distintas dimensiones desde las que tomar decisiones
- Modelo de partida: modelo de datos multidimensional → datos, consultas, informes

# 1.2 Principios de los sistemas multidimensionales

21

- Caso de referencia: sistema de gestión de ventas en librerías (planteamiento general)



## 1.2 Principios de los sistemas multidimensionales

22

- Caso de referencia: sistema de gestión de ventas en librerías (ejemplo de informe)

for Mes de Julio, Las aventuras de Ulises		
	A	B
1	Tienda	Sum of Importe Sorted Down
2	Todo Libro Vélez-Málaga	€1.049,95
3	Todo Libro VÍcar	€827,95
4	Todo Libro Jerez	€661,85
5	Todo Libro Almería	€651,70
6	Todo Libro Berja	€552,55
7	Todo Libro Huetor Tájar	€438,15
8	Todo Libro Aljaraque	€433,05
9	Todo Libro Estepona	€410,15
10	Todo Libro Orio	€357,60
11	Todo Libro Pegalajar	€348,30
12	Todo Libro Donostia	€344,05
13	Todo Libro Granada 2	€341,50
14	Todo Libro Medina Sidonia	€338,95
15	Todo Libro Villacarriedo	€317,80

# 1.2 Principios de los sistemas multidimensionales

23

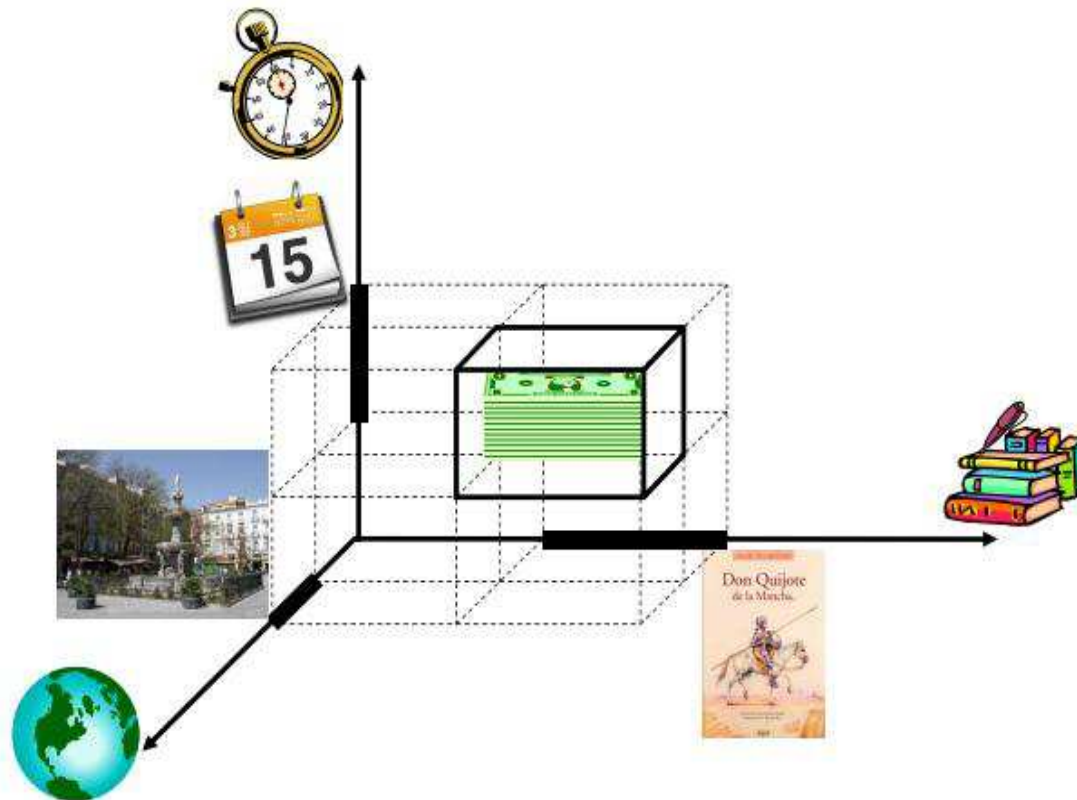
- Caso de referencia: sistema de gestión de ventas en librerías (elementos para la obtención del informe)



## 1.2 Principios de los sistemas multidimensionales

24

- Caso de referencia: sistema de gestión de ventas de librerías (representación espacial –multidimensional– de los elementos del informe)





# 1.2 Principios de los sistemas multidimensionales

25

- Definimos:
  - ▣ Foco de atención: lo que queremos estudiar
    - Son los “*hechos*”
    - Buscamos lo que se puede medir
  - ▣ Dimensiones: características de los hechos
    - Responden a preguntas tipo
- Haremos consultas sobre el foco de atención restringidas a las dimensiones deseadas

# 1.2 Principios de los sistemas multidimensionales

26

## □ Ejemplo: Tienda

▣ ¿Cuál es el foco de atención? La venta

### ■ Mediciones

- ¿De qué tipo son los datos?
- ¿A qué pregunta responden?
- ¿Qué relación tienen con los hechos?

# 1.2 Principios de los sistemas multidimensionales

27

## □ Ejemplo: Tienda

### ▣ ¿Cuáles son las dimensiones?

Cliente	Fecha	Lugar	Pago	Artículo
Emilio Roldán	13/02/2015	Ceuta	Tarjeta	Libro: BDD
M <sup>a</sup> Clara Pérez	14/02/2015	Granada	Efectivo	Bolígrafo azul

- Emilio Roldán, M<sup>a</sup> Clara Pérez → ¿Quién?
- 13/02/2015, 14/02/2015 → ¿Cuándo?
- Ceuta, Granada → ¿Dónde?
- Tarjeta, Efectivo → ¿Cómo?
- Libro, Bolígrafo → ¿Qué?

# 1.2 Principios de los sistemas multidimensionales

28

- Podemos asumir que, en general, los informes tienen:
  - ▣ Hechos que estudiamos: cuánto
  - ▣ Dimensiones desde las que los miramos: quién, cuándo, dónde, cómo, qué.
- Solución:
  - ▣ Usar un modelo de datos que contemple esta estructura y sea más cercano al decisor
  - ▣ Representación gráfica del modelo conceptual

# 1.2 Principios de los sistemas multidimensionales

29

- El modelo multidimensional permite que los informes los genere un “no informático”, es decir, no hay que acudir al DI
- ¿Cómo lo vamos a hacer?
  - ▣ Con Star Tracker
  - ▣ Con Excel
- Dificultad: modelo de datos diferente, que hay que implementar para dárselo al decisor → Alguien lo tiene que construir: el DI

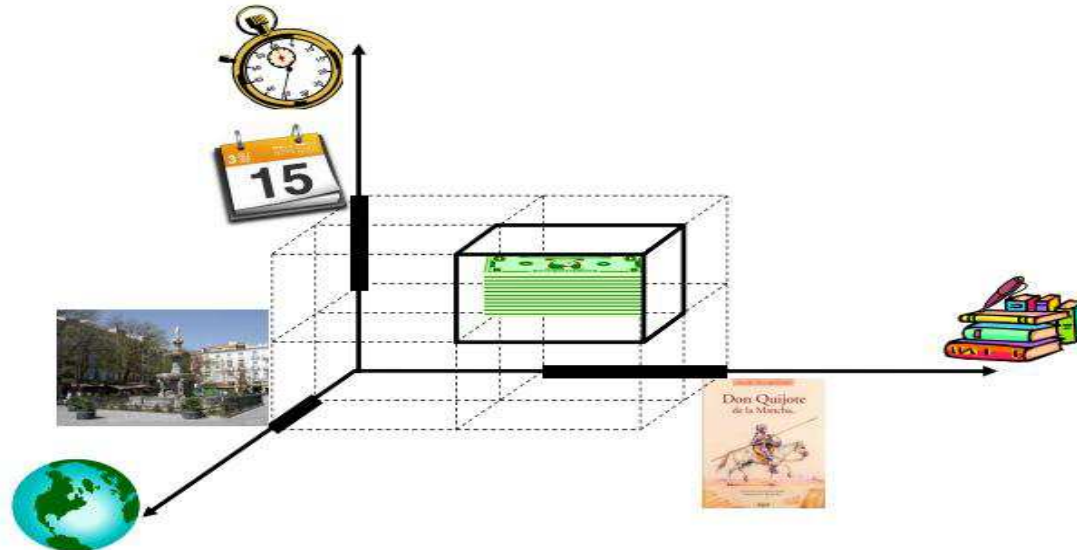
# 1.2 Principios de los sistemas multidimensionales

30

## □ Funcionamiento del modelo multidimensional

### ▣ Representación espacial:

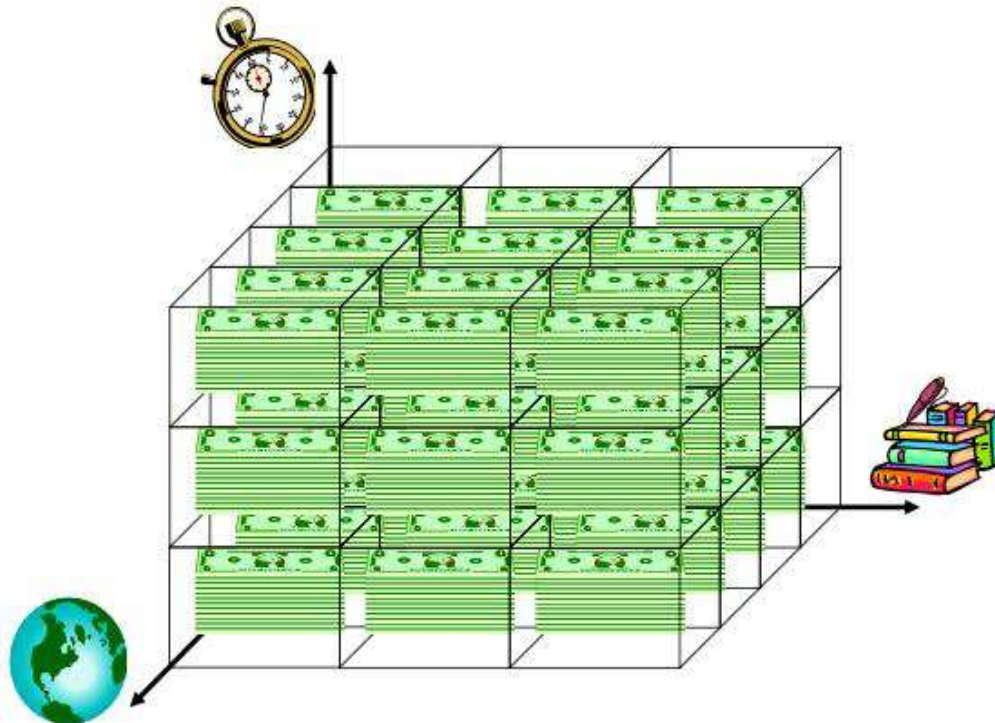
- Las dimensiones están en los ejes
- Un hecho es un punto en el espacio n-dimensional
- La colección de hechos genera un hipercubo  
(n-dimensional) → cubo



# 1.2 Principios de los sistemas multidimensionales

31

- Funcionamiento del modelo multidimensional
  - ▣ ¿Cómo son los cubos?
    - Tienen huecos, están dispersos
      - Cada producto no se vende en todas las tiendas todos los días



# 1.2 Principios de los sistemas multidimensionales

32

## □ Ejemplo: venta de bebidas → Dimensión QUÉ

### ■ Un producto:

- Cerveza Alhambra Especial
- Cerveza Alhambra 1925
- Cerveza Mahou Sin
- Cerveza Mahou 1965

### ■ Una marca:

- Alhambra
- Mahou

### ■ Tipo de bebida:

- Cerveza
- Refresco

### ■ TODO

+ genérico





# 1.2 Principios de los sistemas multidimensionales

33

## □ Ejemplo: venta de bebidas → Dimensión DÓNDE

### ■ Un pueblo o ciudad:

- La Zubia
- Maracena

### ■ Una provincia:

- Granada
- Ceuta

### ■ Una región

- Andalucía

### ■ Un país

- España
- Marruecos

### ■ TODO

+ genérico



# 1.2 Principios de los sistemas multidimensionales

34

## □ Ejemplo: venta de bebidas → Dimensión CUÁNDO

### ■ Un instante:

■ 13/02/2016, a las 15:30 h

### ■ Una franja:

■ 13/02/2016, por la tarde

### ■ Un día:

■ 13/02/2016

### ■ Un mes:

■ Febrero de 2016 (¿y si fuese “Febrero”?)

### ■ Un año:

■ 2016

### ■ TODO

+ genérico



# 1.2 Principios de los sistemas multidimensionales

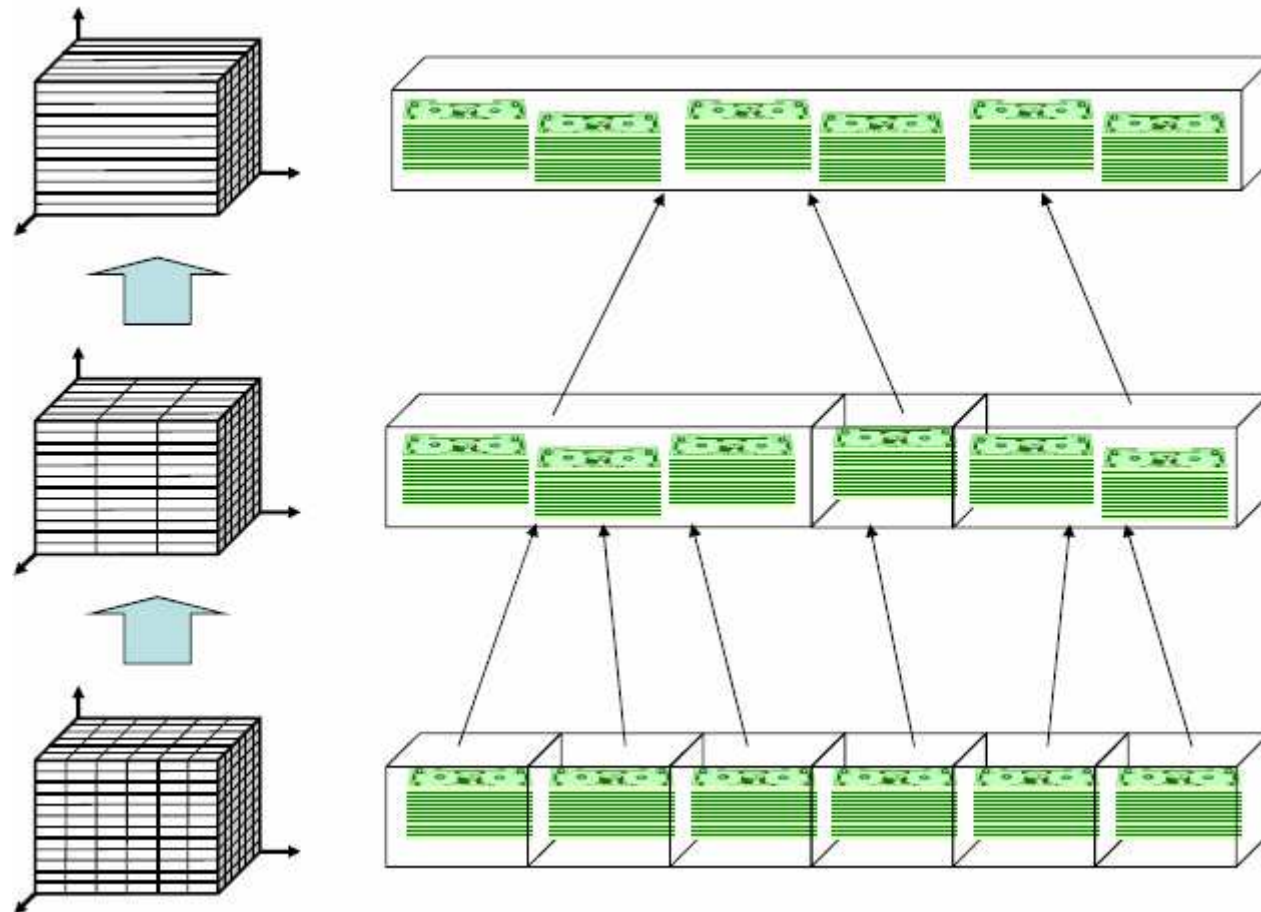
35

- Ejemplo: venta de bebidas → Modelo conceptual
  - ▣ Otras preguntas que se podrían querer contestar:
    - ¿Qué marca...?
    - ¿En qué localidad...?
    - ¿Qué tipo de producto...?
  - ▣ ¿Resulta fácil?
    - Tenemos herramientas que permiten hacer estas operaciones de forma automática
      - Star Tracker, que usaremos en prácticas
    - Las mediciones se pueden agregar (en general) porque son datos numéricos
      - Agregar significa “sumar”

# 1.2 Principios de los sistemas multidimensionales

36

## □ Proceso de agregación



# 1.2 Principios de los sistemas multidimensionales

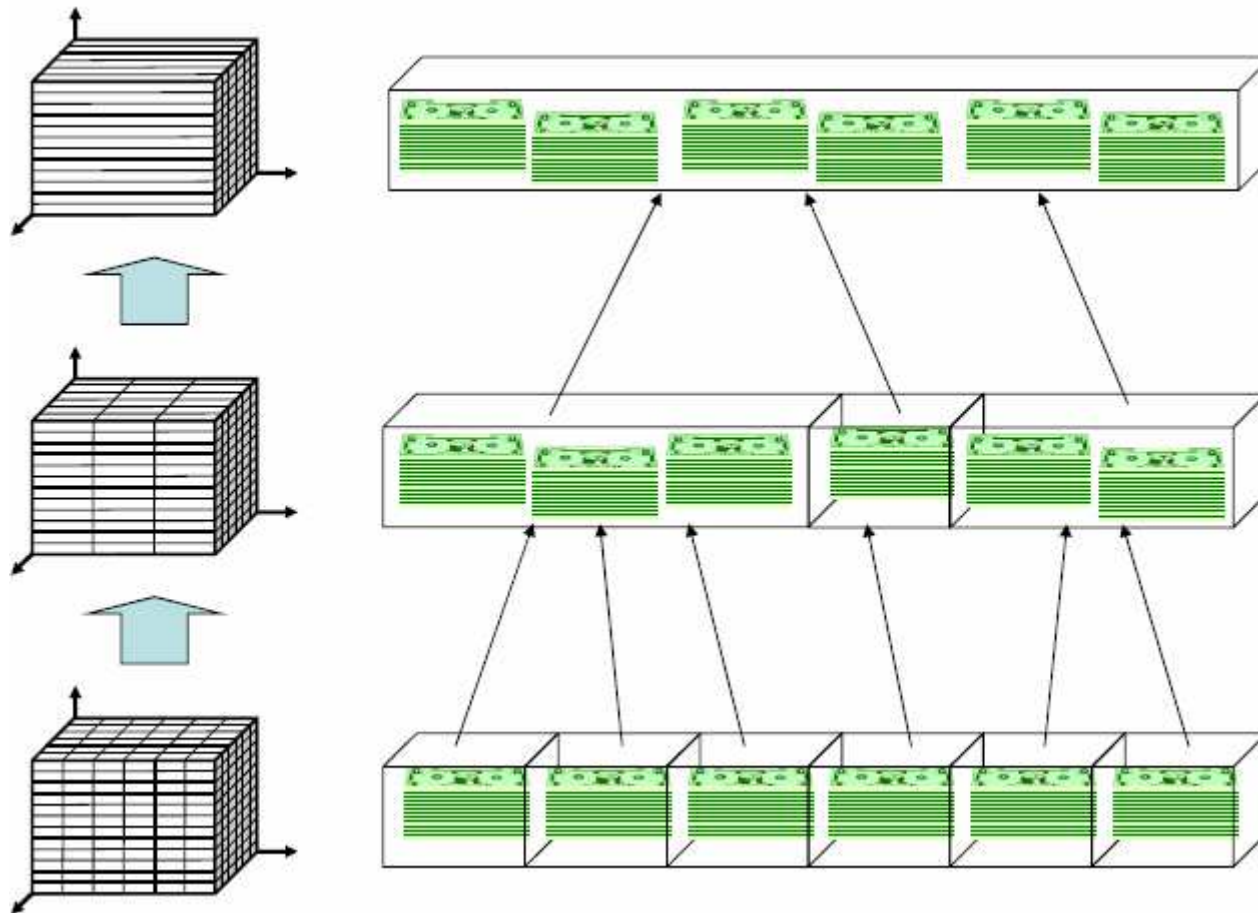
37

- Ejemplo: dinero que llevan las personas
- Hay que considerar tres cosas:
  - ▣ Las consultas que serían interesantes para nuestro decisor
    - Y sus agrupaciones
    - Se hacen fácilmente gracias a la definición de niveles
  - ▣ Si las agrupaciones tienen sentido
    - Son datos numéricos...
  - ▣ Tipos de operaciones que podemos realizar en las agregaciones
    - ROLL-UP, DRILL-DOWN, SLICE & DICE

# 1.2 Principios de los sistemas multidimensionales

38

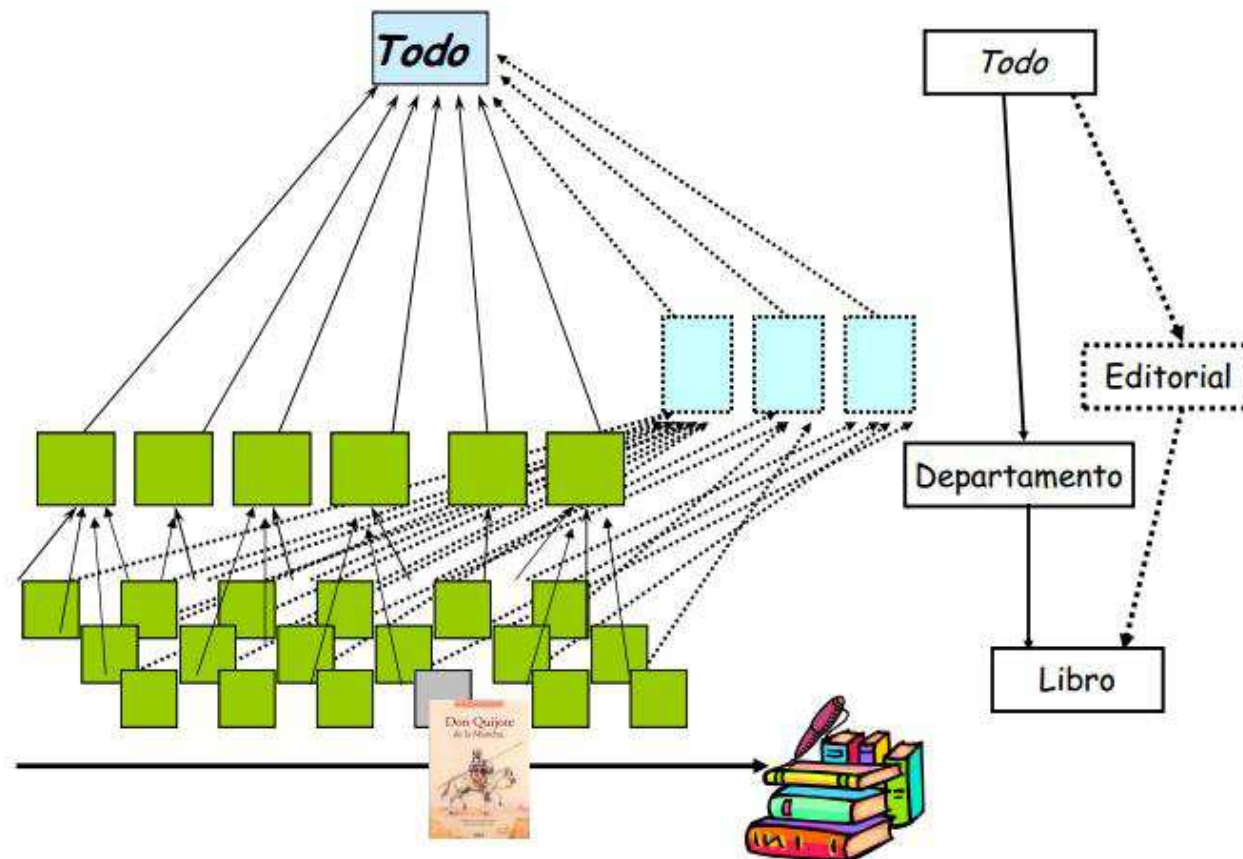
- Roll-up: menor nivel de detalle (agregar)



# 1.2 Principios de los sistemas multidimensionales

39

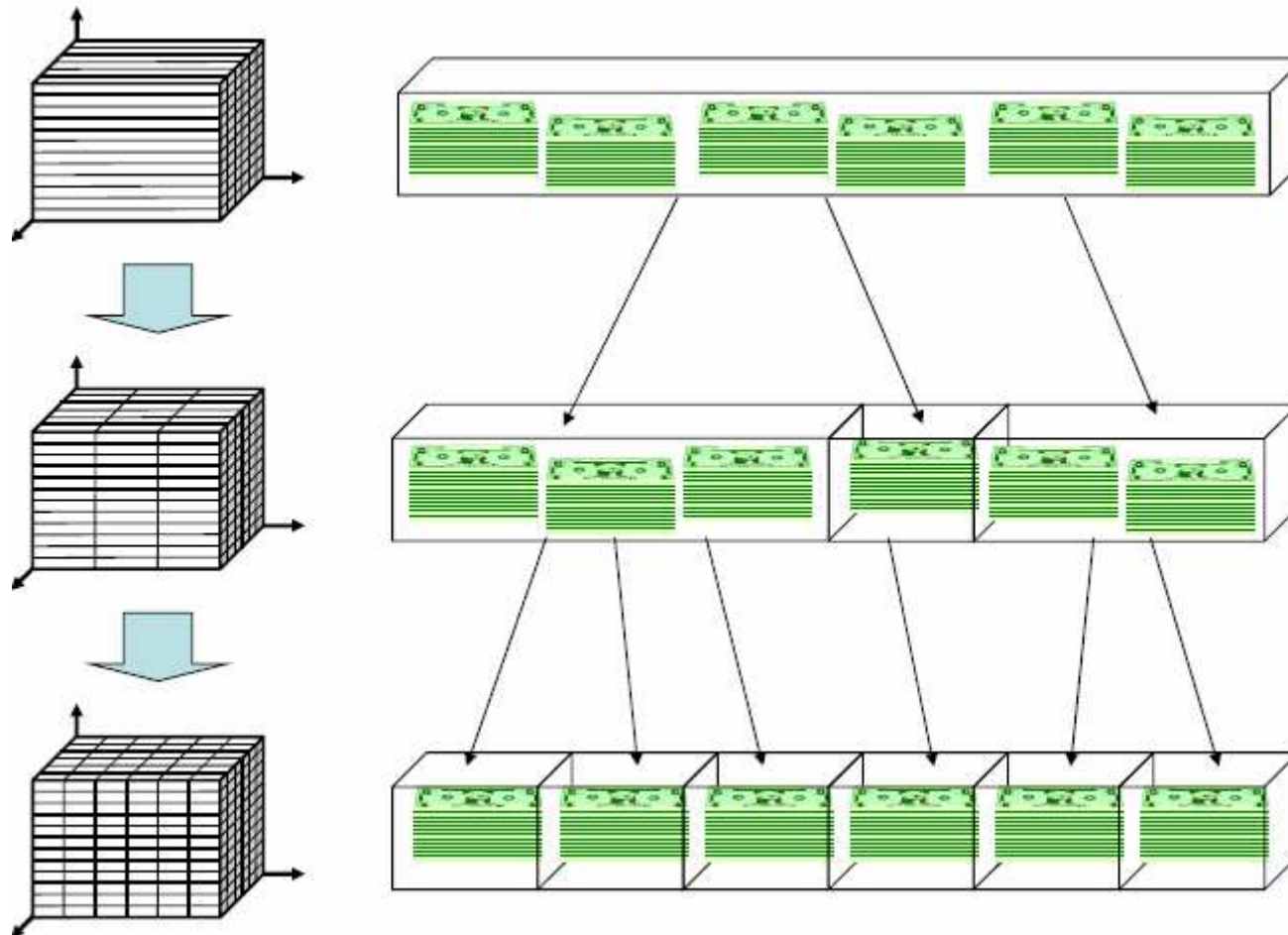
## □ Ejemplo de Roll-up



# 1.2 Principios de los sistemas multidimensionales

40

- Drill-down: mayor nivel de detalle (desagregar)

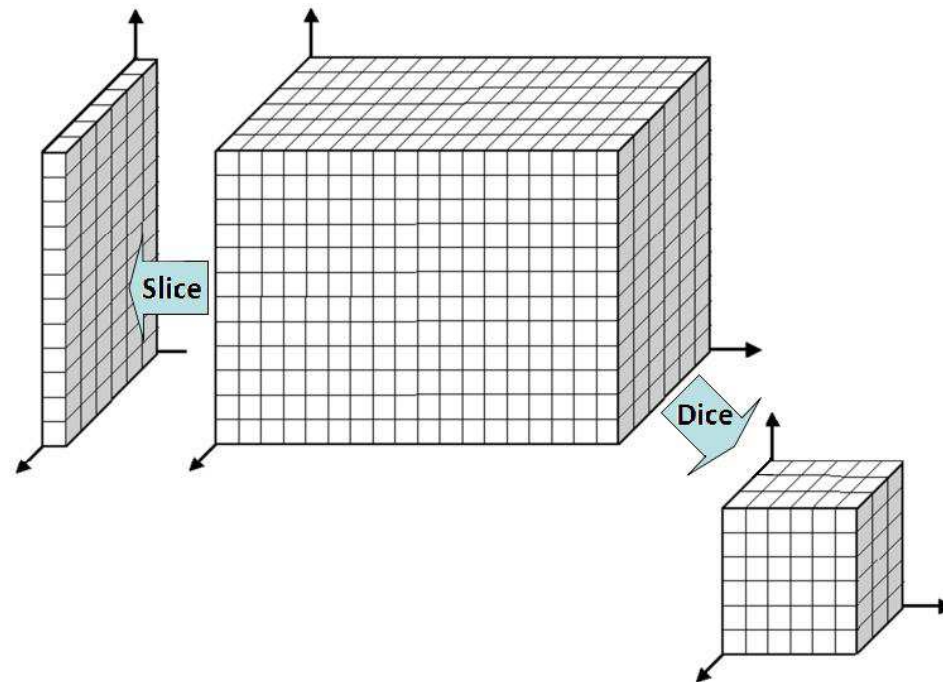




# 1.2 Principios de los sistemas multidimensionales

41

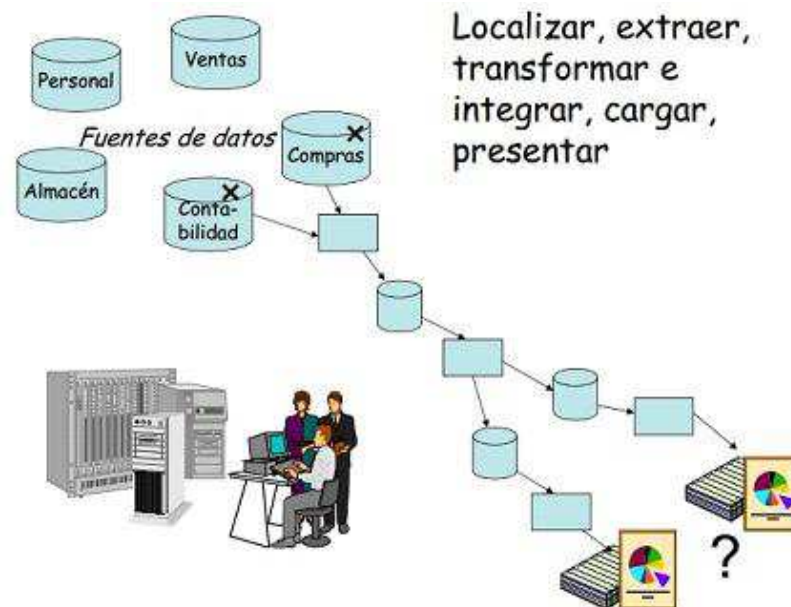
- Slice & Dice: no cambia el nivel de detalle (seleccionar)
  - ▣ Slice: una dimensión
  - ▣ Dice: varias dimensiones



# 1.2 Principios de los sistemas multidimensionales

42

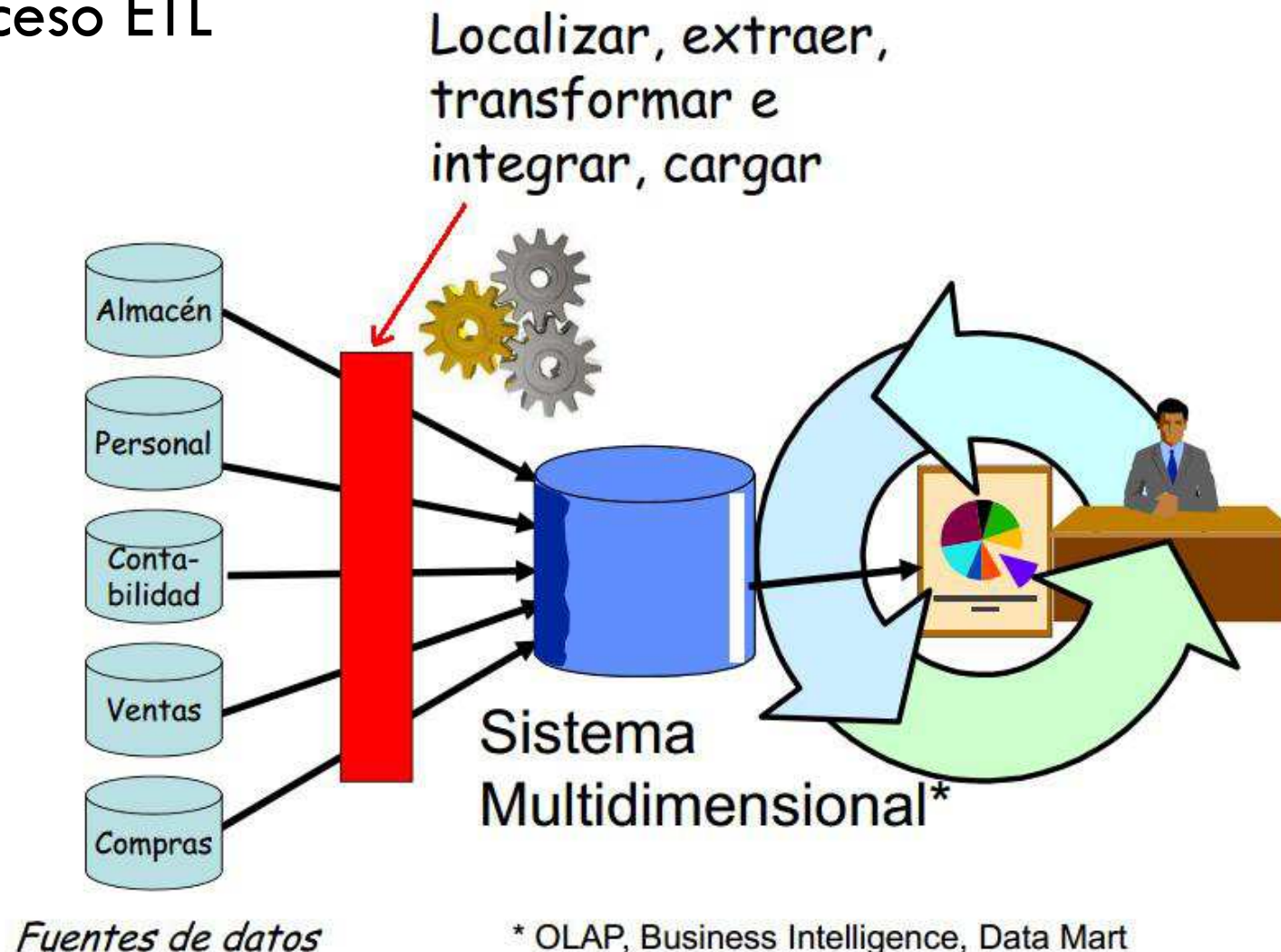
- Componente ETL
  - ▣ Objetivo del Sistema Multidimensional: base de datos centrada en un foco de atención
  - ▣ Necesidad de construir una nueva BD para el Sistema Multidimensional
  - ▣ Se modifica el proceso seguido en los sistemas transaccionales  
(ver páginas 13 y 17)



# 1.2 Principios de los sistemas multidimensionales

43

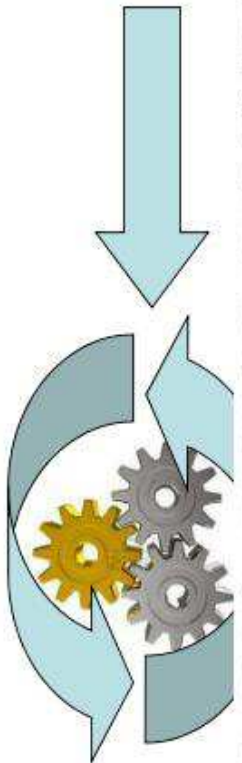
## □ Proceso ETL



# 1.2 Principios de los sistemas multidimensionales

44

## □ Pasos del proceso ETL



The diagram illustrates the ETL process flow. A large light blue arrow points downwards from the top, indicating the initial direction of the process. Below this, a circular flow of three arrows (two light blue, one grey) surrounds a central gear mechanism consisting of three interlocking gears (two grey, one yellow). This visualizes the iterative and interconnected nature of the ETL steps.

Function	Process step	Frequency
<i>Administration</i>	1. Identify the source data.	Once
<i>Administration</i>	2. Identify or define the target data.	Once
<i>Administration</i>	3. Create the mapping between source and target.	Once
<i>Administration</i>	4. Define the replication mode.	Once
<i>Administration</i>	5. Schedule the process of replication.	Once
<i>Capture</i>	6. Capture the required data from the source.	Frequently
<i>Data transfer</i>	7. Transfer the captured data between source and target.	Frequently
<i>Transformation</i>	8. Transform the captured data based on the defined mapping.	Frequently
<i>Apply</i>	9. Apply the captured data to the target.	Frequently
<i>Process management</i>	10. Confirm the success or failure of the replication.	Frequently
<i>Process management</i>	11. Document the outcome of the replication in the metadata.	Frequently
<i>Administration</i>	12. Maintain the definitions of source, target, and mapping.	As needed

# 1.3 El modelo de datos MD

45

- Para definir un modelo de datos se requiere:
  - ▣ Definir los elementos del modelo
  - ▣ Definir las relaciones entre elementos
  - ▣ Definir las operaciones del modelo
  
- Elementos del Modelo de Datos Multidimensional (MDMD)
  - ▣ ¿Cuáles son?
    - Hechos
    - Dimensiones

# 1.3 El modelo de datos MD

46

## □ Elementos del MDMD

### ▣ Dimensiones:

- Niveles → Relación parte-todo exclusiva
- Jerarquías
- Atributos (se definen en los niveles)

### ■ Ejemplo: Dimensión “producto”

- ¿Cuáles son sus atributos?
- ¿La marca sería un atributo?
- ¿Un producto puede estar en varias marcas?
- ¿Un producto puede no tener marca?

# 1.3 El modelo de datos MD

47

## □ Elementos del MDMD

### ▣ Dimensiones:

- Ejemplo: Dimensión “producto”
  - Otras formas de agrupar los datos: tipo y marca
- La estructura de una dimensión es un retículo o rejilla
- Nivel: se define con 1 o más atributos que lo identifican
- Descriptor: el resto de atributos asociados a un nivel
  - ¡No sirven para el Roll-Up! Permiten Slice & Dice
  - Ejemplo en dimensión producto
- Ejercicios: diseñar las dimensiones “dónde” y “cuándo”

# 1.3 El modelo de datos MD

48

## □ Elementos del MDMD

### ▣ Hechos:

- ¿Qué son?
  - ¿A qué pregunta responden?
  - ¿Qué tienen los hechos?
- 
- Ejercicio: Definir el foco de atención de “Ventas”
    - ¡Cuidado! En el modelo, el foco de atención se nombra en singular



# 1.3 El modelo de datos MD

49

## □ Relaciones entre los elementos del MDMD: bases y aditividad

### ▣ Bases

- Necesidad de identificar un hecho concreto
- Son conjuntos de dimensiones que identifican un hecho
- Ejemplo: si todos comprásemos una botella de agua de 750 ml en la cafetería de nuestro centro, ¿cuántas instancias habría en nuestro sistema?

# 1.3 El modelo de datos MD

50

- Relaciones entre los elementos del MDMD: bases y aditividad

*(Jensen, Pedersen y Thomsen 2010)*

A **cube** is a truly multidimensional data structure for capturing and analyzing data.

A cube consists of uniquely identifiable **cells** at each of the dimensions' intersections.

A nonempty cell is called a **fact**.

A fact has a number of **measures** associated with it. These are numerical values that "live" within the cells.

# 1.3 El modelo de datos MD

51

## □ Relaciones entre los elementos del MDMD: bases y aditividad

### □ Bases

- Entonces, ¿qué representan las bases? ¿qué significan?
  - Indican el nivel de detalle de los hechos
  - Identifican a cada hecho particular → Permiten relacionar las dimensiones con los hechos
- ¿Cuántas bases puede haber?
  - Varias
  - Tienen que identificar a los hechos unívocamente

# 1.3 El modelo de datos MD

52

## □ Relaciones entre los elementos del MDMD: bases y aditividad

### ▣ Aditividad

- Define cómo se agrega una medición a lo largo de las dimensiones
- ¿Cómo se detecta? → ¿Tiene sentido sumar esta medición?
  - Sí implica que la agregación mediante la suma conserva el sentido de la medición
- Ejemplo: saldo de una cuenta del banco
  - ¿Tiene sentido sumar el saldo?

# 1.3 El modelo de datos MD

53

- Relaciones entre los elementos del MDMD: bases y aditividad
  - ▣ Aditividad: debe comprobarse en todas las dimensiones
    - Aditivo: tiene sentido sumar en todas sus dimensiones
    - No aditivo: no tiene sentido sumar en ninguna dimensión
    - Semi-aditivo: sólo tiene sentido sumar en algunas dimensiones
    - En los ejemplos vistos, ¿son aditivas las mediciones?
      - ¿Importe?
      - ¿Cantidad?
      - ¿PVP?
      - ¿Saldo?

# 1.3 El modelo de datos MD

54

- Operaciones del modelo
  - ▣ Ya conocemos las siguientes:
    - Roll-up
    - Drill-down
    - Slice & dice
  - ▣ ¿Ejemplos?
    - Usad como referencia alguno de los que ya tenemos de clase

# 1.3 El modelo de datos MD

55

- Operaciones del modelo
  - ▣ Podríamos tener varios sistemas (focos de atención) en una misma empresa
    - Almacén, personal, venta...
  - ▣ Queremos comprobar lo que pasó en cada uno de los focos de atención en las mismas circunstancias
    - Ejemplo: ventas en Granada, en julio de 2014, de productos de limpieza
      - ¿Qué paso en el Almacén de Granada, en julio de 2014, con los productos de limpieza?

# 1.3 El modelo de datos MD

56

- Operaciones del modelo
  - ▣ Drill-Across: llevar las condiciones definidas sobre un cubo con un foco de atención a otro cubo con otro foco de atención distinto.
  - ▣ Requisito:
    - Que haya dimensiones en común entre los cubos
      - Al menos, niveles en común
      - Ejemplo:
        - Cubo 1 → Venta: tienda, fecha, producto
        - Cubo 2 → Almacén: producto, provincia, fecha
        - Cubo 3 → Personal: persona, tienda, fecha



# 1.3 El modelo de datos MD

57

## □ Operaciones del modelo: Drill-Across

▣ Uso: realizar informes combinados de distintos focos de atención

■ Ejemplo: Informe de Ventas

Producto	Venta
Producto 1	0
Producto 2	0
...	...

Producto	Almacén
Producto 1	0
Producto 2	N
...	...

Producto	Venta	Almacén
Producto 1	0	0
Producto 2	0	N (≠0)
...	...	...

Esto se añade con  
**Drill-Across**

En vez de crear un cubo nuevo, se combinan los focos de atención

# 1.4 Niveles de abstracción

58

## □ OLTP: On-line Transactional Processing

### ▣ Nivel conceptual: ¿Qué?

- Modelo Entidad Relación (E/R)
- Modelo Orientado a Objetos (O-O)

### ▣ Nivel lógico: ¿Cómo?

- Modelo Relacional

### ▣ Transformaciones entre niveles:

Conceptual

Lógico

Físico

E/R



Mod. Relacional



SGBD Relacional

O-O



Object Relational Mapping

SGBD Relacional u  
Objeto-Relacional

# 1.4 Niveles de abstracción

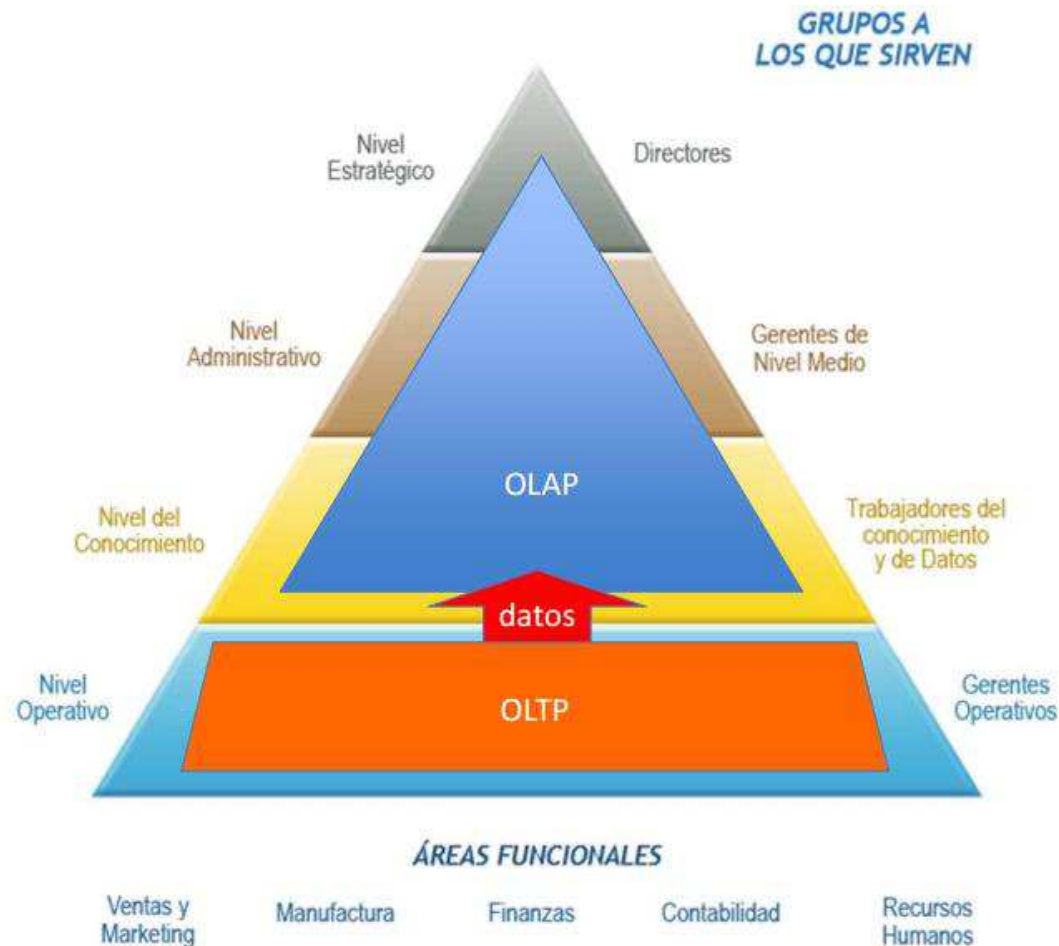
59

- OLAP: On-line Analytical Processing
  - ▣ Funcionalidad de análisis en tiempo real
  - ▣ Nivel conceptual: ¿Qué?
    - Modelo de Datos Multidimensional
  - ▣ Nivel lógico: ¿Cómo?
    - Con una base de datos
    - Veremos cómo se define más adelante

# 1.4 Niveles de abstracción

60

- OLTP y OLAP: ¿tienen alguna relación?



# 1.4 Niveles de abstracción

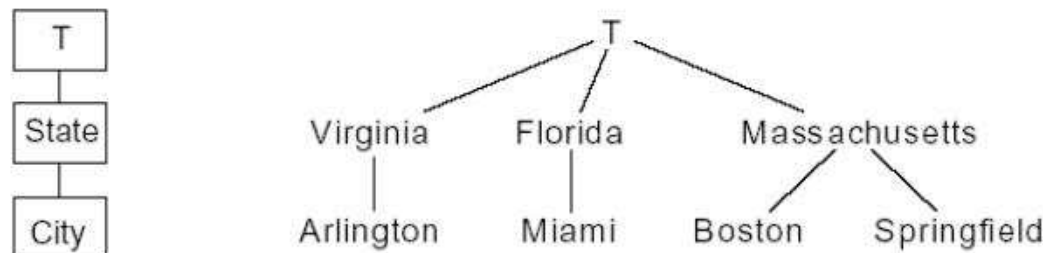
61

## □ Visiones del nivel conceptual

*(Jensen, Pedersen y Thomsen 2010)*

A **dimension** is organized into a containment-like hierarchy composed of a number of **levels**, each of which represents a level of detail that is of interest to analyses to be performed.

The instances of the dimension are typically called **dimension values** or **dimension members**. Each such value belongs to a particular level.



The T (“top”) level represents **all** of the dimension.

# 1.4 Niveles de abstracción

62

## □ Visiones del nivel conceptual

*(Jensen, Pedersen y Thomsen 2010)*

Facts are the objects that represent the **subjects** of the desired analyses, i.e., the interesting “things,” or events or processes, that are to be analyzed to better understand their behavior.

A fact then has a certain **granularity**, which is determined by the levels from which its dimension values are drawn.

A **measure** has two components: a **numerical property** of a fact, e.g., the sales price or profit, and a **formula** (most often a simple aggregation function such as SUM) that can be used to combine several measure values into one.

It is important to distinguish among three classes of measures, namely *additive*, *semi-additive*, and *non-additive* measures, as these behave quite differently in computations.

**Additive** measure values can be combined mediante la suma meaningfully along any dimension.

**Semi-additive** measure values cannot be combined along one or more of the dimensions, most often the Time dimension.

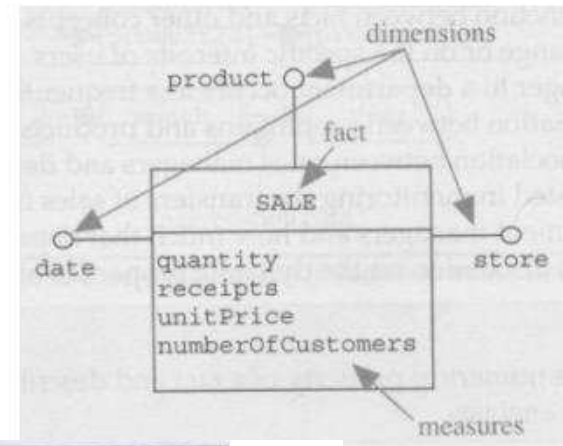
**Non-additive** measure values cannot be combined along any dimension, usually because of the chosen formula. For example, this occurs when averages for lower-level values cannot be combined into averages for higher-level values

# 1.4 Niveles de abstracción

63

## □ Visiones del nivel conceptual

*(Golfarelli y Rizzi, 2009)*



### **Fact**

A *fact* is a concept relevant to decision-making processes. It typically models a set of events taking place within a company.

### **Measure**

A *measure* is a numerical property of a fact and describes a quantitative fact aspect that is relevant to analysis.

### **Dimension**

A *dimension* is a fact property with a finite domain and describes an analysis coordinate of the fact.

### **Primary Event**

A *primary event* is a particular occurrence of a fact, identified by one n-ple made up of a value for each dimension. A value for each measure is associated with each primary event.

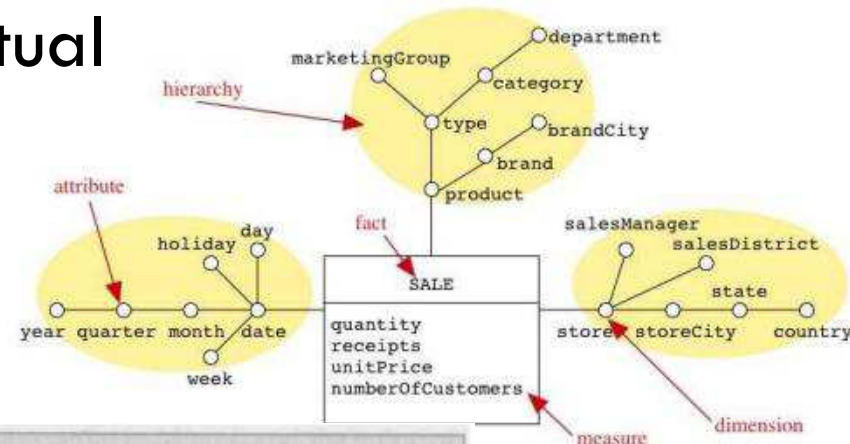


# 1.4 Niveles de abstracción

64

## □ Visiones del nivel conceptual

(Golfarelli y Rizzi, 2009)



### Hierarchy

A *hierarchy* is a directed tree<sup>1</sup> whose nodes are dimensional attributes and whose arcs model many-to-one associations between dimensional attribute pairs. It includes a dimension, positioned at the tree's root, and all of the dimensional attributes that describe it.

### Dimensional Attribute

The general term *dimensional attributes* stands for the dimensions and other possible attributes, always with discrete values, that describe them.

A *descriptive attribute* is functionally determined by a dimensional attribute of a hierarchy and specifies a property of this attribute. Descriptive attributes often are tied to dimensional attributes by one-to-one associations and do not actually add useful levels of aggregation. Sometimes, they have domains with continuous values, so they cannot be used for aggregation.

### Secondary Events

Given a set of dimensional attributes (generally belonging to separate hierarchies), each n-ple of their values identifies a *secondary event* that aggregates all of the corresponding primary events. Each secondary event is associated with a value for each measure that sums up all the values of the same measure in the corresponding primary events.



# 1.4 Niveles de abstracción

65

- Nivel lógico
  - ▣ Partimos del nivel conceptual: ¿Qué queremos modelar?
    - Utilizamos como base una representación gráfica del modelo conceptual

# 1.4 Niveles de abstracción

66

## □ Nivel lógico

### ▣ Posibilidades de implementación (1ª):

#### ■ Sabemos:

- Estructuras de datos
- Algoritmos
- Grabar y leer en disco

#### ■ Implementación directa con estructuras multidimensionales

#### ■ Se denomina sistema MOLAP

# 1.4 Niveles de abstracción

67

- Nivel lógico
  - ▣ Posibilidades de implementación (2º):
    - Sabemos:
      - BD Relacionales: tablas y relaciones
    - Se denomina sistema ROLAP

# 1.4 Niveles de abstracción

68

- Nivel lógico

- Posibilidades de implementación (3ª):

- Sabemos:

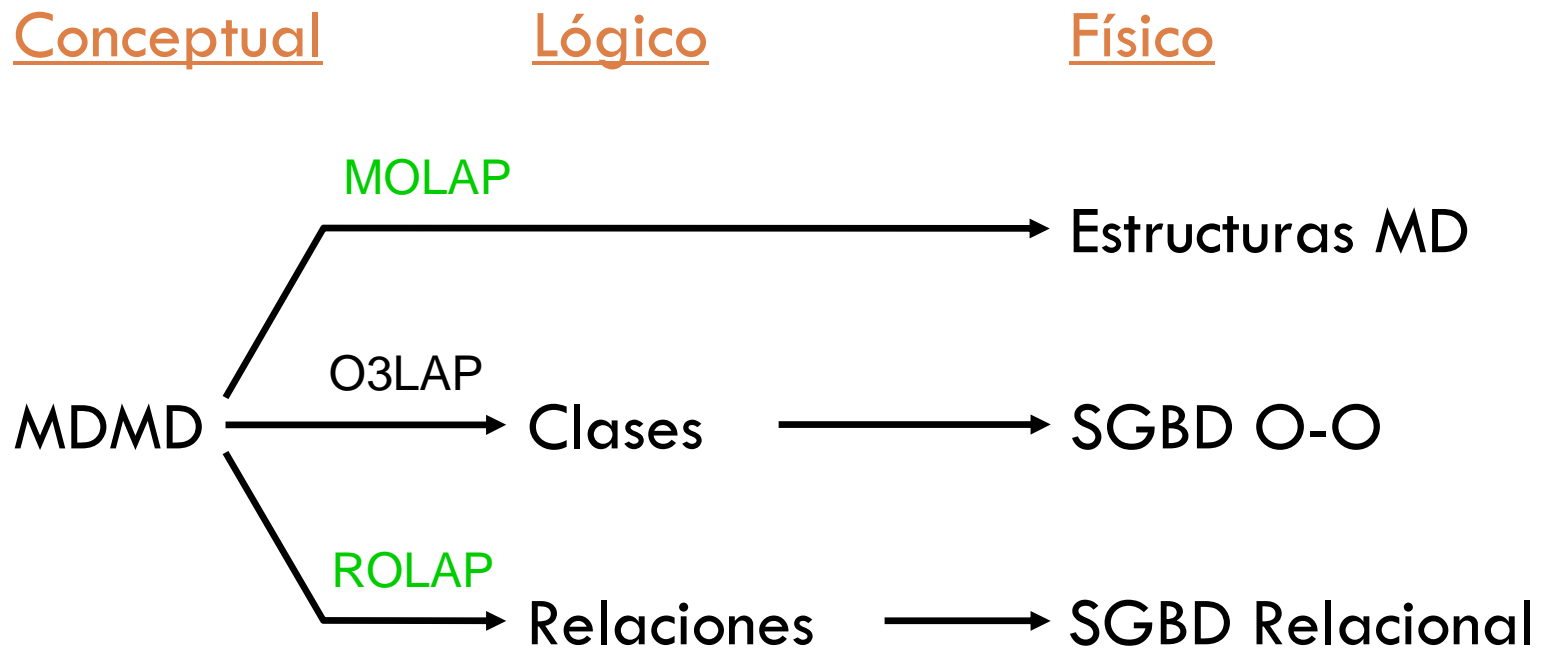
- BD Orientada a Objetos: clases y relaciones

- Se denomina OOOLAP → O3LAP

# 1.4 Niveles de abstracción

69

- OLAP
  - ▣ Transformaciones entre niveles:



# 1.4 Niveles de abstracción

70

- Nivel lógico: ROLAP
  - ▣ La cardinalidad de las relaciones entre los niveles de la jerarquía son 1-N
    - ¿Cómo se implementa a nivel relacional?
      - Usando tablas con llaves primarias y llaves externas
  - ▣ Si se normalizan las tablas:
    - Dimensiones normalizadas
    - Cada nivel requiere una tabla
    - Los hechos están en una tabla
    - Hay un núcleo y ramificaciones
      - Implementación en COPO DE NIEVE

# 1.4 Niveles de abstracción

71

- Nivel lógico: ROLAP
  - ▣ Si no se normalizan las tablas:
    - Una tabla de hechos
    - Una tabla para cada dimensión
    - Tablas no normalizadas
      - Implementación en ESTRELLA
  - ▣ Vamos a utilizar ROLAP-Estrella
  - ▣ La R hace referencia a Relacional
    - Implica que la implementación se hará con una BD Relacional

# 1.4 Niveles de abstracción

72

## □ Nivel lógico

### □ Diferencias entre ROLAP-Estrella y ROLAP-Copo de Nieve

	ROLAP-Estrella	ROLAP-Copo de Nieve
<b>Tabla para cada</b>	Dimensión	Nivel
<b>Hechos</b>	Una tabla	Una tabla
<b>Normalización de tablas</b>	No	Sí

### □ ¿Cuál es mejor?

- Según BD → Copo de nieve
- Normalización de tablas para:
  - Evitar redundancia de datos
  - Disminuir problemas de actualización de datos
  - Proteger la integridad de los datos



# 1.4 Niveles de abstracción

73

- Nivel lógico
  - ▣ Realmente, ¿cuál es mejor?
    - Depende...

	Estrella	Copo de nieve
<b>Complejidad</b>		Más complejo
<b>Tamaño de datos</b>	Ocupa más	
<b>Modificaciones</b>	--	--
<b>Consultas</b>	Un solo JOIN entre los hechos y las dimensiones	Requiere más de un JOIN

# 1.4 Niveles de abstracción

74

## □ Nivel lógico

### ▣ ¿Cuánto espacio adicional se requiere en Estrella?

#### ■ Ejemplo:

##### ■ Dimensiones:

- 8 tiendas
- 5 años
- 80000 productos en el catálogo

##### ■ Hechos:

- Venta: nivel tienda-fecha → dos personas que compran el mismo producto en la misma tienda en la misma fecha genera una única venta – hecho –
- Calcular el número de instancias y el espacio ocupado en cada implementación: estrella y copo de nieve

# 1.4 Niveles de abstracción

75

## □ Nivel lógico

### ▣ ¿Cuánto espacio adicional requiere Estrella?

- Más espacio para las dimensiones
- El tamaño de los hechos no varía respecto a Copo de Nieve
- El espacio no es un problema actualmente

### ▣ Copo de nieve:

- Ventajas: ocupa menos y no hay redundancia de datos (tablas normalizadas)
- Desventajas: requiere más JOIN, peor rendimiento, lentitud de respuesta en los informes

# 1.4 Niveles de abstracción

76

## □ Nivel lógico

### ▣ Reglas generales:

- Un esquema conceptual MDMD pasa a un esquema lógico ROLAP-Estrella
- Cada dimensión será una tabla
- Los hechos serán una tabla
- SIEMPRE emplear llaves generadas

# 1.4 Niveles de abstracción

77

## □ Nivel lógico

### ▣ Reglas generales:

- Las tablas de las dimensiones estarán desnormalizadas
- A las tablas de dimensión se les añade una clave generada que ocupa espacio adicional, pero compensa en cuanto a rendimiento y tamaño de hechos
- Los hechos se relacionan con las dimensiones mediante las claves generadas

# 1.4 Niveles de abstracción

78

## □ Nivel lógico

### ▣ HOLAP: Hybrid On-line Analytical Processing

- Es una implementación híbrida
- Parte de los datos se almacena como ROLAP y otra parte como MOLAP
- Ventajas MOLAP:
  - Ocupa menos que ROLAP
  - Indexación natural de matrices y vectores
  - Pre-estructuración de datos agregados
- Desventajas MOLAP:
  - Dificultades con muchas dimensiones
  - Carga de datos lenta (carga incremental)

# 1.4 Niveles de abstracción

79

## □ Nivel lógico

### ▣ HOLAP: Hybrid On-line Analytical Processing

#### ■ Dos formas de implementar HOLAP:

##### ■ Partición vertical:

- Agregaciones en MOLAP para velocidad de consultas
- Datos detallados en ROLAP para velocidad de procesamiento del cubo

##### ■ Partición horizontal:

- Datos más recientes en MOLAP
- Datos antiguos en ROLAP

# 1.4 Niveles de abstracción

80

- Nivel lógico

- ▣ Herramientas:

- Star Tracker → ROLAP

- Excel → MOLAP



# 1.4 Niveles de abstracción

81

- Nivel físico:
  - ▣ Se verá en el tema 3...
  - ▣ Tratará de índices, fragmentación de tablas y, en general, la implementación física del sistema multidimensional