

# Tema 3

Modelos de datos.  
El modelo de datos relacional.

# Contenidos

- Definición de modelo de Datos
- El modelo de datos relacional
  - Estructura de datos
  - Restricciones de integridad
- Otros modelos de datos

# Contenidos

- Definición de modelo de Datos
- El modelo de datos relacional
  - Estructura de datos
  - Restricciones de integridad
- Otros modelos de datos

# Definición de modelo de datos

- Proceso de transformación

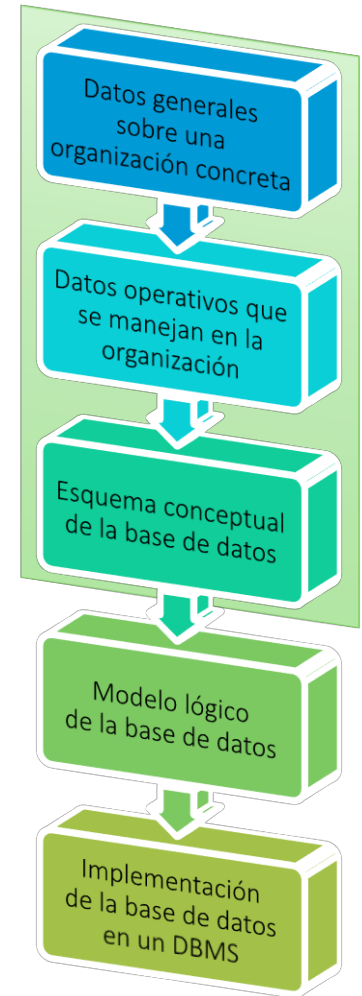
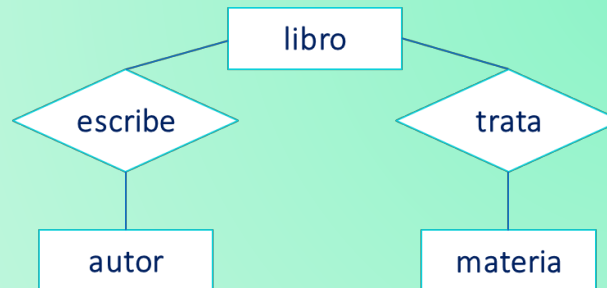
- a) Mundo real
- Delimitación objetivos
  - Selección de datos
  - Hipótesis semánticas

- Biblioteca

- b) Datos operativos
- Atributos
  - Conexiones
  - Restricciones

- Libro: título, isbn, editorial....
- Autor: nombre, nacionalidad,....
- Materia: código, descripción....

- c) Esquema conceptual



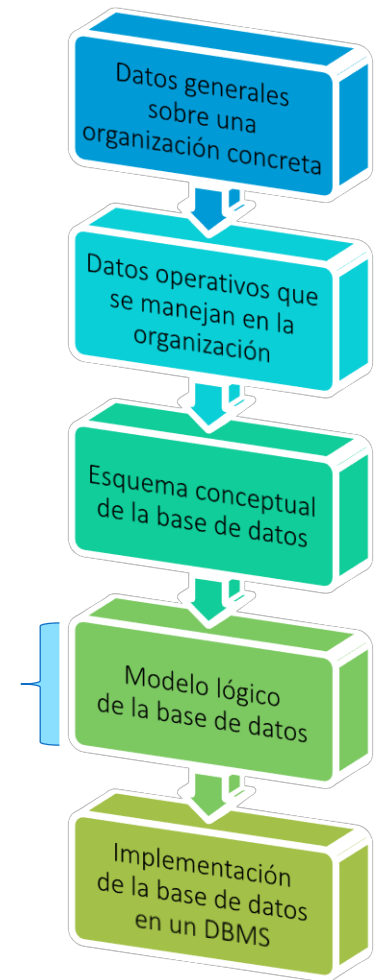
# Definición de modelo de datos

Modelado lógico:

- Trasladamos a un esquema lógico en función de una estructura implementable.

Tabla LIBROS

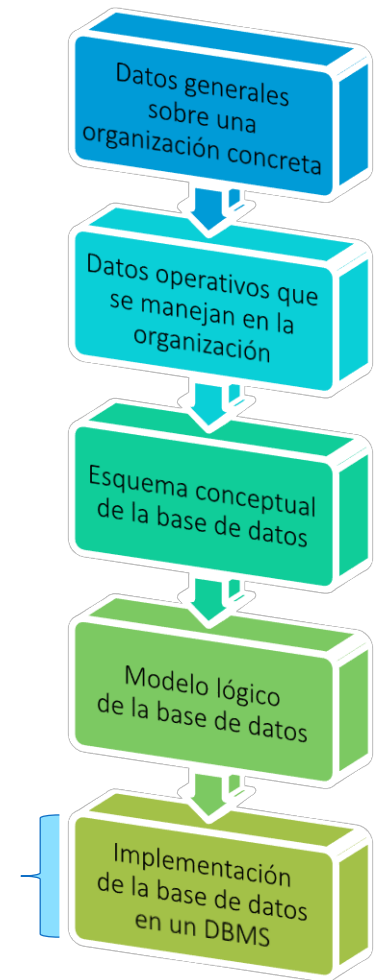
Título	ISBN	Editorial	...
Introducción a las BD	1234-1234	Thomson	...
Cálculo para todos	4321-4321	Delta	...
...	...	...	...



# Definición de modelo de datos

- Lo implementamos en un SGBD:

```
CREATE TABLE LIBROS (  
    titulo char(45) NOT NULL,  
    ISBN char(10) PRIMARY KEY,  
    editorial char(30) REFERENCES ...  
);
```



# Definición de modelo de datos

- Modelo de datos
  - Mecanismo formal para representar y manipular información de manera general y sistemática.
  - Debe constar de:
    - 1.- Notación para describir datos.
    - 2.- Notación para describir operaciones.
    - 3.- Notación para describir restricciones de integridad.
- Historia
  - 1<sup>er</sup> Modelo: Relacional (Codd).
  - Se recuperan los modelos basados en grafos (1974).
  - El modelo E/R (Chen, 1975), otros modelos semánticos.
  - Modelos orientados a objetos (1983, 1986,...).
  - Modelos lógicos (1986,...).

# Definición de modelo de datos

- Necesidad de modelos de datos:
  - Cada esquema se describe utilizando un lenguaje de definición de datos.
  - Este lenguaje es de muy bajo nivel, está muy ligado al SGBD.
  - Hacen falta otros mecanismos de más alto nivel que permitan describir datos de una forma no ambigua y entendible por los usuarios implicados en cada paso del proceso de implantación.



# Definición de modelo de datos

- Objetivo:
  - Describir modelos que representen los datos y los describan de una forma entendible y manipulable.
  - En relación con la Arquitectura ANSI/SPARC:
    - Nivel Externo:
      - Modelo de datos externo.
    - Nivel Conceptual:
      - Modelo de datos conceptual.
    - Nivel Interno:
      - Modelo de datos interno.

# Definición de modelo de datos

- Clasificación:
  - Basados en registros.
  - Basados en objetos.
  - Físicos.
- Utilización:
  - Los dos primeros:
    - Nivel externo y conceptual.
  - Físicos:
    - Nivel interno.

# Definición de modelo de datos

- Modelos de datos basados en registros:
  - Modelo de datos jerárquico.
  - Modelo de datos en red.
  - Modelo de datos relacional (Codd, 1969).

# Contenidos

- Definición de modelo de Datos
- El modelo de datos relacional
  - Estructura de datos
  - Restricciones de integridad
- Otros modelos de datos

# El modelo de datos relacional

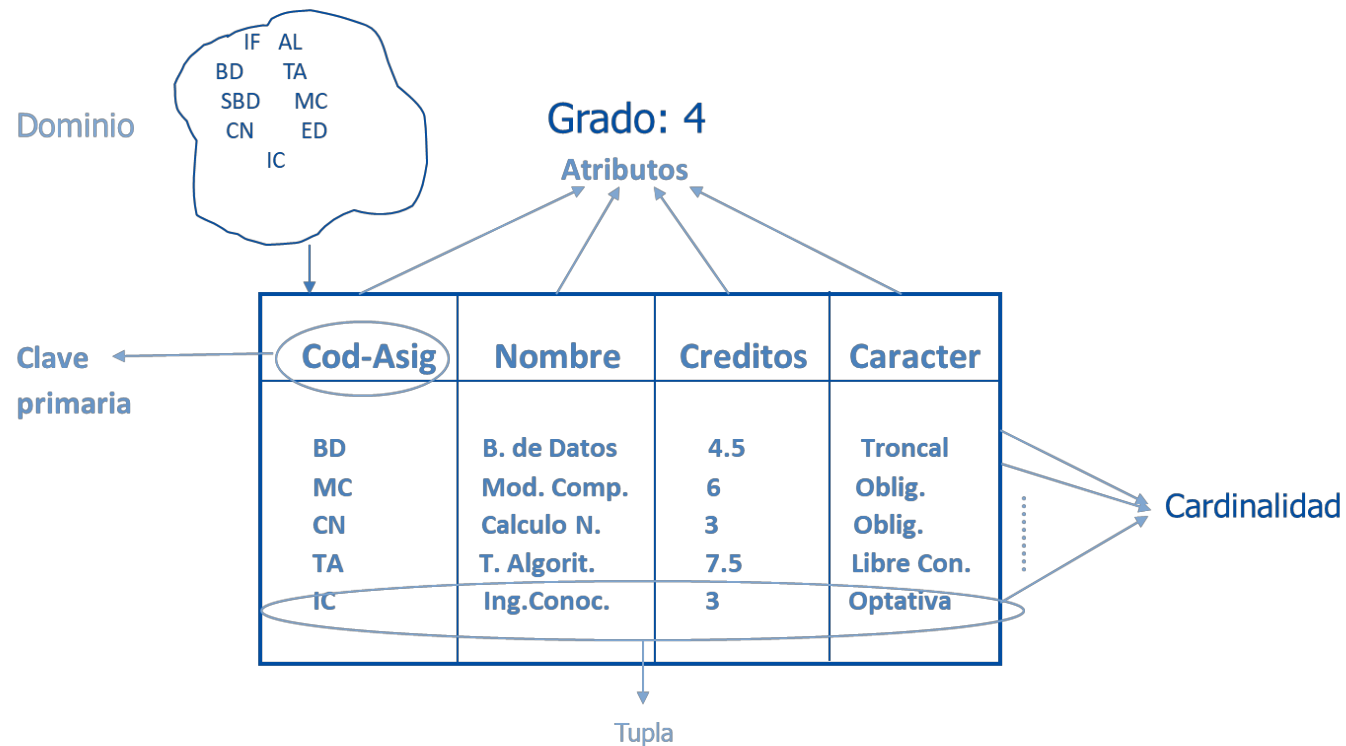
## Estructura de datos

- El modelo de datos relacional organiza y representa los datos en forma de tablas o relaciones.
- Base de datos relacional: colección de tablas cada una de las cuales tiene un nombre único.

# El modelo de datos relacional

## Estructura de datos

### Ejemplo: Asignaturas de una titulación



# El modelo de datos relacional

## Estructura de datos

- Propuesto por E.F. Codd en 1970.
- El modelo relacional abarca tres ámbitos distintos de los datos:
  - Las estructuras para almacenarlos: El usuario percibe la información de la base de datos estructurada en tablas.
  - La integridad: Las tablas deben satisfacer ciertas condiciones que preservan la integridad y la coherencia de la información que contienen.
  - Consulta y manipulación: Los operadores empleados por el modelo se aplican sobre tablas y devuelven tablas.
- La tabla es la estructura lógica de un sistema relacional.
  - A nivel interno/físico, el sistema es libre de almacenar los datos en el formato más adecuado (archivo secuencial, archivo indexado, listas con apuntadores,...).

# El modelo de datos relacional

## Estructura de datos

- Algunos conceptos:
  - Atributo:
    - Cualquier elemento de información susceptible de tomar valores. Notación:  $A_i$ ,  $i=1,2,...$
  - Dominio:
    - Rango de valores donde toma sus datos un atributo. Se considera finito. Notación:  $D_i$ ,  $i=1,2,...$
  - Relación:
    - Dados los atributos  $A_i$  ( $i=1..n$ ) con dominios  $D_i$  no necesariamente distintos, definimos relación asociada a  $A_1,...,A_n$ , y lo notaremos por  $R(A_1,...,A_n)$ , a cualquier subconjunto del producto cartesiano  $D_1 \times D_2 \times .. \times D_n$ .
  - Tupla
    - Cada una de las filas de una relación.



# El modelo de datos relacional

## Estructura de datos

- Cardinalidad de una relación
  - Número de tuplas que contiene.
  - Es variable en el tiempo.
- Esquema de una relación R
  - Atributos de la relación junto con su dominio.
  - $A_1:D_1, \dots, A_n:D_n$
- Grado de una relación
  - Número de atributos de su esquema.
  - Invariable en el tiempo.
- Instancia de una relación
  - Conjunto de tuplas  $\{(x_1, x_2, \dots, x_n)\} \subseteq D_1 \times D_2 \times \dots \times D_n$  que la componen en cada momento.

# El modelo de datos relacional

## Estructura de datos

- Esquema de una base de datos relacional:
  - Colección de esquemas de relaciones junto con restricciones de integridad.
- Instancia o estado de una base de datos:
  - Colección de instancias de relaciones que verifican las restricciones de integridad.
- Base de datos relacional:
  - Instancia de una base de datos junto con su esquema.

# El modelo de datos relacional

## Estructura de datos

- Propiedades

- Condición de normalización

- Todos los valores de los atributos de una relación son atómicos.
  - Un valor atómico es un valor no estructurado.
- Cuando una relación cumple la primera condición de normalización se dice que está en Primera Forma Normal.

- Consecuencias

- No hay valores tipo conjunto.
- No hay valores tipo registro.
- No hay valores tipo tablas.

- Problema

- Todas las representaciones son extensivas: no se puede representar directamente información del tipo “el valor del atributo asignaturas de un alumno es : (FBD, ALG, LD)”.

Handwritten example illustrating the problem of representing sets of values in a relational database:

	CONY 1	CONY 2	HIJOS	} Así no
	JUAN	MARÍA	AUCIA, ANTONIO	
↓	JUAN	MARÍA	AUCIA	} Así sí
	JUAN	MARÍA	ANTONIO	

# El modelo de datos relacional

## Estructura de datos

- Consecuencias de la definición:
  - No hay tuplas duplicadas.
    - Por la definición conjuntista de relación.
  - No hay orden en las filas ni en los atributos.
    - Al no estar ordenados ni los atributos ni las filas (conjuntos) el acceso es por Nombre de Atributo y Valor.

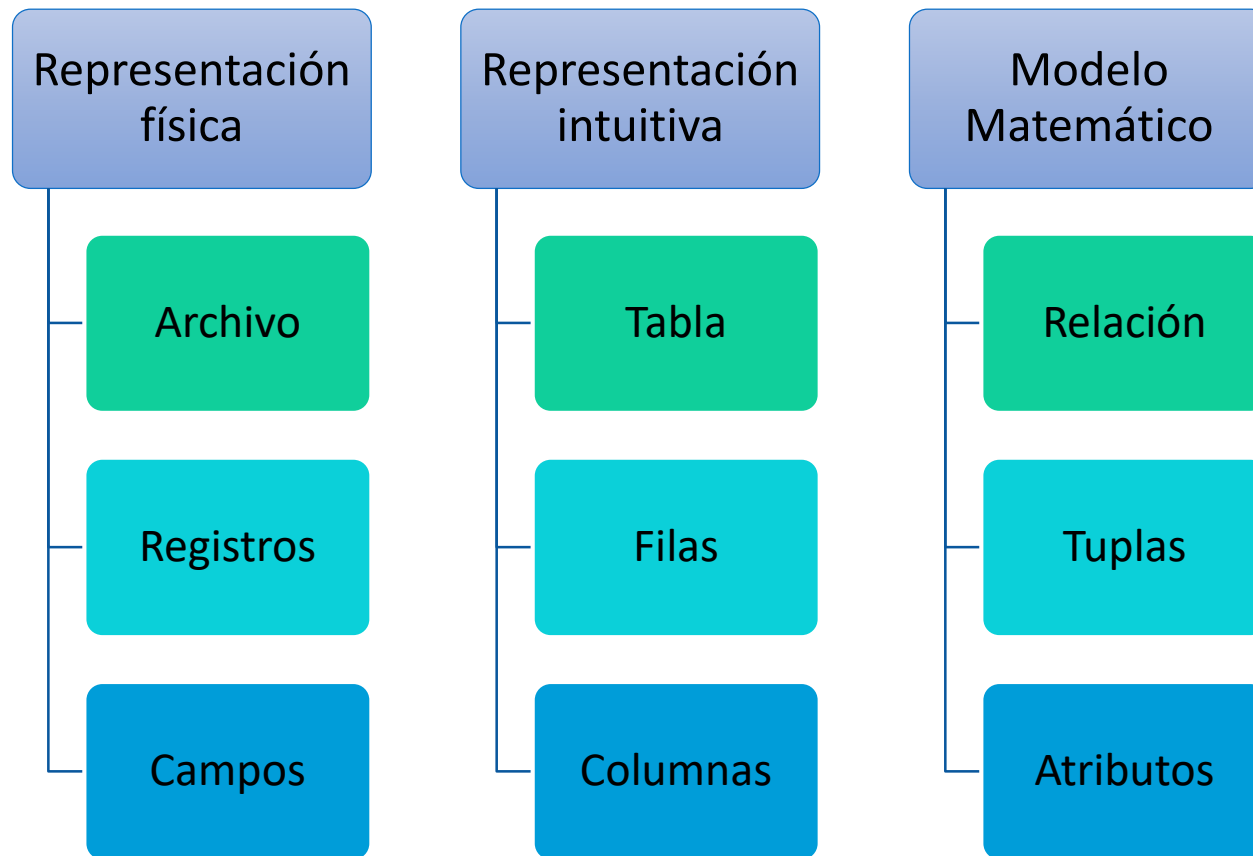
*Nunca vamos a buscar por posiciones, sino por el atributo de una entidad concreta  
(Aquí se refleja la independencia física)*

A	B	C	D	E
a1	b1	c1	d1	e1
a1	b2	c2	d2	e1
a2	b1	c3	d3	e1
a2	b1	c4	d3	e1
a3	b2	c5	d1	e1

A	B	C	D	E
a2	b1	c4	d3	e1
a2	b1	c3	d3	e1
a1	b2	c2	d2	e1
a3	b2	c5	d1	e1
a1	b1	c1	d1	e1

# El modelo de datos relacional

## Estructura de datos



# El modelo de datos relacional

## Estructura de datos

- Notación a utilizar

- Relación:  $R, S, T, \dots$
- Atributos:  $A, B, \dots$
- Esquema de relación:  $R(A_1, A_2, \dots, A_n)$
- Instancia de relación  $R$ :  $r, \dots$
- Tuplas de una instancia:  $t_1, t_2, \dots \in r$
- Valor de un atributo  $A_i$  en una tupla  $t_j$ :  $t_j[A_i]$  ó  $A_{ij}$

# El modelo de datos relacional

## Estructura de datos

- Algunas veces no se conoce el valor de un atributo para una determinada tupla. En esos casos a ese atributo de esa tupla se le asigna un valor nulo.
  - Un valor nulo puede ser un valor desconocido.
  - Un valor nulo puede ser un atributo no aplicable.
  - En cualquier caso, ese valor es un valor más de todos los dominios de la base de datos.

# El modelo de datos relacional

## Restricciones de integridad

- Las restricciones o reglas de integridad son condiciones para preservar la corrección semántica de una base de datos.
- Específicas del problema:
  - $0 \leq \text{edad} \leq 100$
  - $\text{créditos} > 0$
  - $\text{carácter} \in (\text{'troncal'}, \text{'obligatoria'}, \text{'optativa'}, \dots)$
- Propias del papel de los atributos en el esquema:
  - $\text{imparte.NRP} \in \text{profesor.NRP}$   
(un profesor inexistente no puede impartir una asignatura)
  - $\text{cod\_asig} \neq \text{nulo}$   
(siempre debe conocerse el código de una asignatura)



# El modelo de datos relacional

## Restricciones de integridad

- Superclaves y claves (candidatas y primarias)
  - Superclave: Cualquier conjunto de atributos que identifica unívocamente a cada tupla de una relación.
- Clave: superclave minimal.
  - Ejemplo:
    - Por ejemplo, en la relación Asignaturas, el conjunto de atributos {Cod\_Asig, Nombre} identifica unívocamente cada tupla. Sin embargo, no es minimal y no puede considerarse como una clave. Cod\_Asig por sí solo, es una clave.
- En una relación puede que más de un conjunto de atributos cumplan las condiciones para ser clave. Estos conjuntos de atributos se llaman claves candidatas.
- De entre las claves candidatas, hay que elegir una como principal. Esta clave se denomina Clave Primaria. → *Cuanto más corta mejor*
  - Criterio de selección: Tamaño, significado, capacidad para recordarla, fusión con otras tablas,....

# El modelo de datos relacional

## Restricciones de integridad

- Clave candidata y primaria (definición formal).
  - Sea  $R(A_1, A_2, \dots, A_n)$ ;  $CC \subseteq \{A_1, A_2, \dots, A_n\}$  se denomina clave candidata si y solo si:
    - Unicidad:  $\forall r$  instancia de  $R$  y  $\forall t_1, t_2 \in r, t_1 \neq t_2 \Rightarrow t_1[CC] \neq t_2[CC]$
    - Minimalidad: No existe  $CC' \subset CC$  que verifique la unicidad.
  - Es decir, una clave candidata es un atributo o conjunto de atributos que identifica a cada tupla en la relación, para el que, además, no existe un subconjunto propio del mismo que también identifique a cada tupla de la relación.
  - La clave primaria es la clave candidata elegida por el diseñador para desempeñar el papel de identificar.
  - Si  $CC$  verifica la unicidad y no la minimalidad, entonces solo es superclave.

# El modelo de datos relacional

## Restricciones de integridad

- Completamos la notación para describir una relación, subrayando los atributos que forman su clave primaria, si existen otras claves candidatas también se subrayan etiquetando el subrayado con CC:

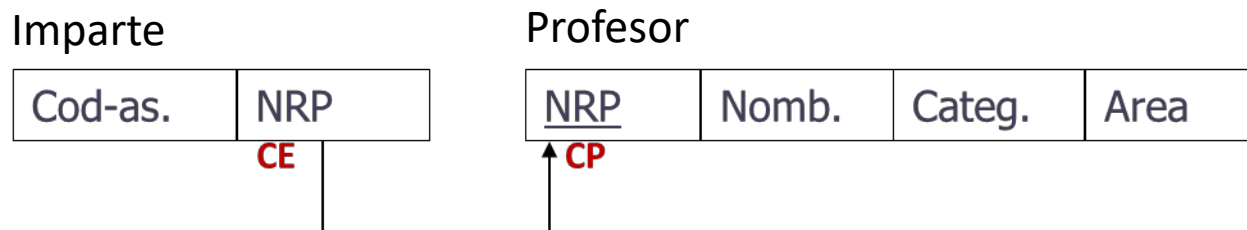
Socio (#Socio, DNI, Nombre, Dirección)

CC

# El modelo de datos relacional

## Restricciones de integridad

- Clave externa (ajena):
  - Consideramos una relación R (relación que referencia) y un subconjunto de atributos de su esquema CE (clave externa); y una relación S (relación referenciada) cuya clave primaria CP coincide con CE.
  - Si CE es una clave externa de R que referencia a CP en S, entonces:
    - $\forall t \in r, \exists t' \in s$  tal que  $t[CE] = s[CP]$ , donde r y s son las instancias de R y S en la base de datos.
  - Eventualmente, R y S pueden ser la misma relación.



# El modelo de datos relacional

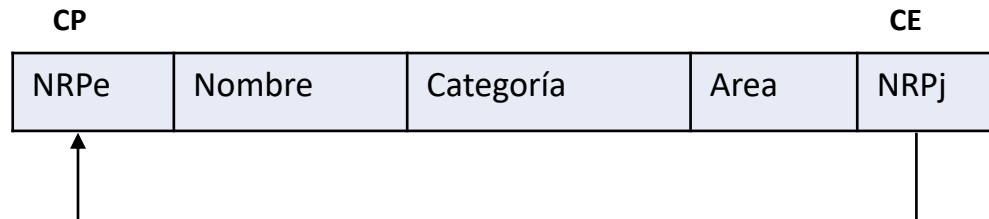
## Restricciones de integridad

- Dominio activo:
  - Subconjunto de valores del dominio de un atributo de una relación que está presente en la instancia de la relación.
    - Dada R una relación, su instancia r y un atributo A de R, el dominio de A es el siguiente conjunto:
$$\{t[A] | t \in r\}$$
  - Podemos extender el concepto de dominio activo a un conjunto de atributos.
    - Dada R una relación, su instancia r y un subconjunto de atributos CA de R, el dominio de CA es el siguiente conjunto:
$$\{t[CA] | t \in r\}$$
- Las claves externas establecen relaciones de inclusión entre los dominios activos de la clave externa y la clave referenciada.

# El modelo de datos relacional

## Restricciones de integridad

- Claves externas
  - Puede haber más de una clave externa en una relación.
  - Puede haber una clave externa a la clave primaria de la propia relación.
- Ejemplo



# El modelo de datos relacional

## Restricciones de integridad

- Conceptos generales
  - Condiciones de integridad:
    - Normas que mantienen la corrección semántica de una base de datos.
  - Integridad Genérica: depende del papel que juega un atributo en el diseño de la tabla.
    - Son metarreglas (generan reglas de integridad aplicadas a una base de datos concreta).
    - Hay dos:
      - Integridad de entidad.
      - Integridad referencial.

# El modelo de datos relacional

## Restricciones de integridad

- Integridad de entidad
  - No se debe permitir que una entidad sea representada en la base de datos si no se tiene una información completa de los atributos que son clave primaria de la entidad:
    - La clave primaria, o una parte de la misma, no puede ser un valor nulo.

*Un atributo que forma parte de la clave primaria de una tupla en una relación no puede tener un valor nulo.*



# El modelo de datos relacional

## Restricciones de integridad

- Integridad referencial:
  - Una base de datos en la que todos los valores no nulos de una clave externa referencian valores reales de la clave referenciada en la otra relación cumple la regla de integridad referencial.

*Si una relación incluye una clave externa conectada a una clave primaria, el valor de la clave externa debe ser, bien igual a un valor ya existente en el dominio activo de la clave primaria, o bien completamente nulo (si la semántica lo permite).*

- La integridad referencial mantiene las conexiones en las bases de datos relacionales.

# El modelo de datos relacional

## Restricciones de integridad

- EL SGBD debe encargarse de mantener las siguientes restricciones:
  - La unicidad de la clave primaria y de las claves candidatas.
    - Frente a operaciones de Inserción y Actualización, el SGBD debe rechazar los valores introducidos que sean iguales a los presentes en la BD para los atributos que el diseñador ha definido como clave primaria y como claves candidatas.
  - La integridad de identidad.
    - Frente a operaciones de Inserción y Actualización, el SGBD debe rechazar las modificaciones que asignen un valor NULO a algún atributo de la clave primaria.

# El modelo de datos relacional

## Restricciones de integridad

- La integridad referencial.
  - En inserción:
    - Rechazar la tupla insertada si el valor de la clave externa no concuerda en la relación referenciada para alguna tupla en el valor su clave primaria.
    - Si el valor para la clave externa es NULO y el diseño no lo permite, habrá de rechazar también esa inserción.
  - En actualización:
    - Si se actualiza la clave externa: rechazar la modificación si se produce alguna de las circunstancias descritas en punto anterior.
    - Si se actualiza la clave primaria de la relación referenciada: Actualizar en cadena las claves externa que la referencien (o impedir la actualización mientras existan referencias a valor anterior).
  - En borrado:
    - Si se borra la clave primaria en la relación referenciada: borrado en cadena de todas las tuplas que la referencian o poner valor nulo en la clave externa de todas esas tuplas.

# Contenidos

- Definición de modelo de Datos
- El modelo de datos relacional
  - Estructura de datos
  - Restricciones de integridad
- Otros modelos de datos

# Otros modelos de datos

## Modelo Jerárquico

- Fue el primero en implementarse físicamente
  - Nivel externo: aplicaciones Cobol.
  - No había interactividad:
    - Carecía de un lenguaje de consulta.
- Estructura de datos básica:
  - Árbol
    - Registro maestro.
    - Registros secundarios.
- La BD es una colección de instancias de árboles.

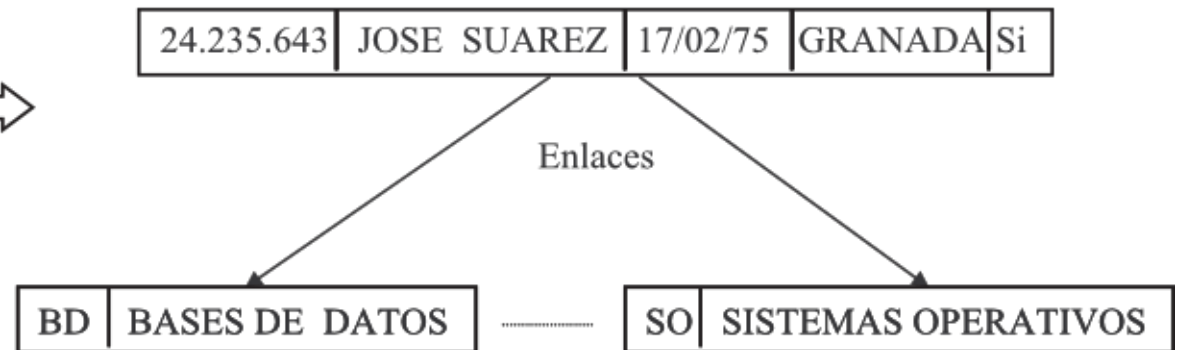
# Otros modelos de datos

## Modelo Jerárquico

### Esquemas



### Instancias



# Otros modelos de datos

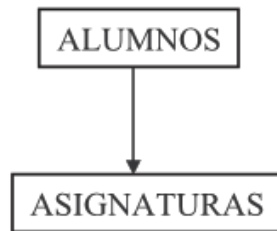
## Modelo Jerárquico

- Esta estructura plasma de forma muy directa:
  - Relaciones muchos a uno.
  - Relaciones uno a uno.
- Relaciones muchos a muchos:
  - Hay que duplicar toda la información sobre las entidades involucradas.

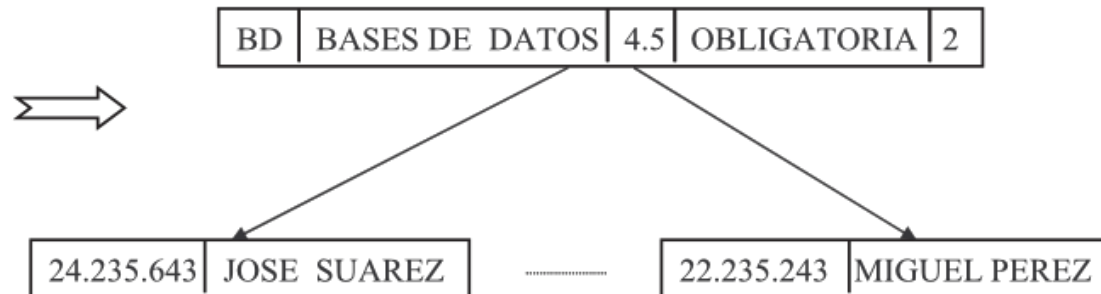
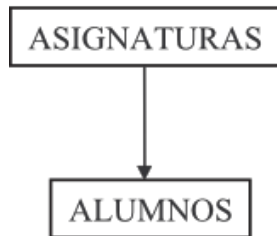
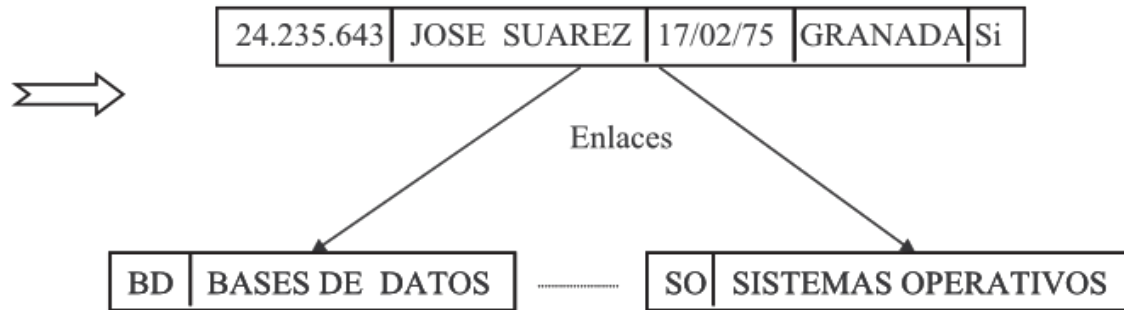
# Otros modelos de datos

## Modelo Jerárquico

### Esquemas



### Instancias





# Otros modelos de datos

## Modelo Jerárquico

- Inconvenientes:
  - Almacenar árboles en ficheros es complejo:
    - Varios tipos de registros.
    - Punteros que hay que mantener.
  - DML difícil de:
    - Implementar.
    - Usar.
  - Dependencia existencial obligatoria de los registros de tipo secundario con respecto a los de tipo raíz:
    - No se podrá insertar un registro de tipo secundario mientras no exista uno de tipo raíz con el que “enlazarlo”.
  - Redundancia necesaria para plasmar relaciones muchos a muchos.
    - La integridad de los datos es costosa de mantener.

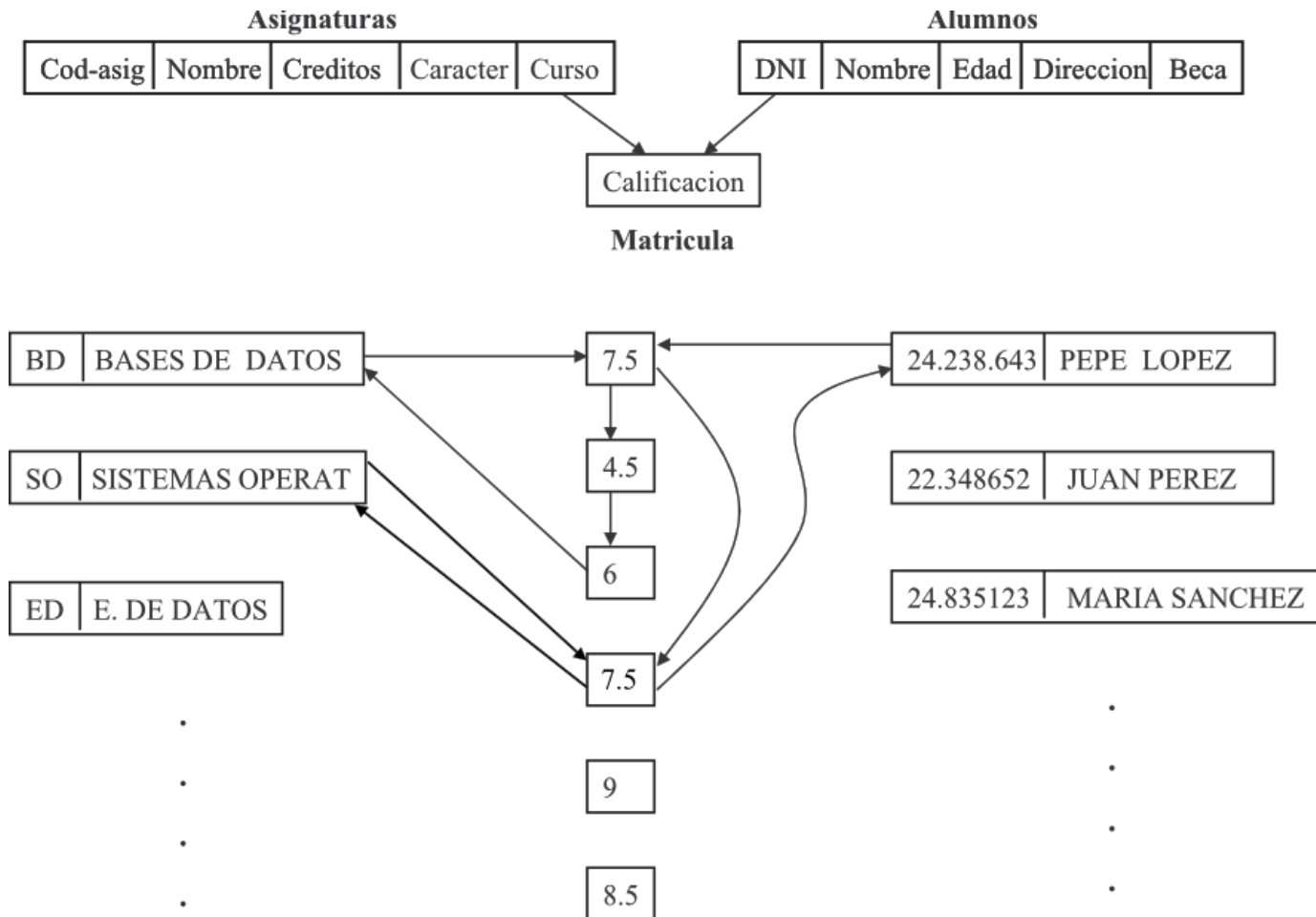
# Otros modelos de datos

## Modelo en Red

- Estructura de datos:
  - Grafos cuya topología depende de las conexiones existentes entre las entidades
    - Nodos: registros.
    - Arcos: enlaces entre registros (punteros).
    - Relaciones entre conjuntos de entidades:
      - Conectores: registros especiales (atributos propios de la relación).
      - Cada ocurrencia de un conector representa una asociación distinta.
  - Cualquier registro puede relacionarse con cualquier registro.
- Base de datos:
  - Colección de instancias de grafos.
- La estructura es muy genérica:
  - Permite plasmar todo tipo de relaciones.
  - Implementa directamente las relaciones muchos a muchos.

# Otros modelos de datos

## Modelo en Red



# Otros modelos de datos

## Modelo en Red

- Ventajas:
  - Estructura algo más homogénea.
  - Permite insertar nuevas entidades en un conjunto de forma independiente.
- Problemas:
  - La existencia de enlaces entre los registros hace que las operaciones del DDL y el DML sigan siendo complejas de implementar y utilizar.

# Otros modelos de datos

## Comparativa

- Con respecto a la representación

- Relacional

- Un solo elemento para la representación (esencialidad).
    - Conexiones lógicas.
    - Representación relaciones n:m simétrica.
    - Identidad por valor.

- Basado en grafos

- Dos elementos para la representación.
    - Conexiones en el modelo físico subyacente.
    - Representación conexiones n:m: imposible en modelos jerárquicos, difícil en modelos en red.
    - Identidad por posición.

- Con respecto a la consulta

- Relacional

- Obtención de la consulta como resultado global.



- Lenguajes declarativos (y también procedimentales).

- Basado en grafos

- Mecanismo de navegación por punteros.



- Lenguajes procedimentales.

# Otros modelos de datos

## Orientación a objetos

- Debilidades de los SGBD relacionales:
  - Pobre representación de las entidades del mundo real.
  - Sobrecarga semántica de la estructura básica: la relación.
  - La estructura relacional es muy estricta:
    - Todas las tuplas (filas) han de tener los mismo atributos.
    - Los valores de un atributo (columna) pertenecen al mismo dominio.
    - Atributos han de tener un valor atómico.
  - SQL puramente relacional permite un conjunto de operaciones limitado. No permite modelar el comportamiento de muchos objetos del mundo real.
  - Object-relational impedance mismatch.

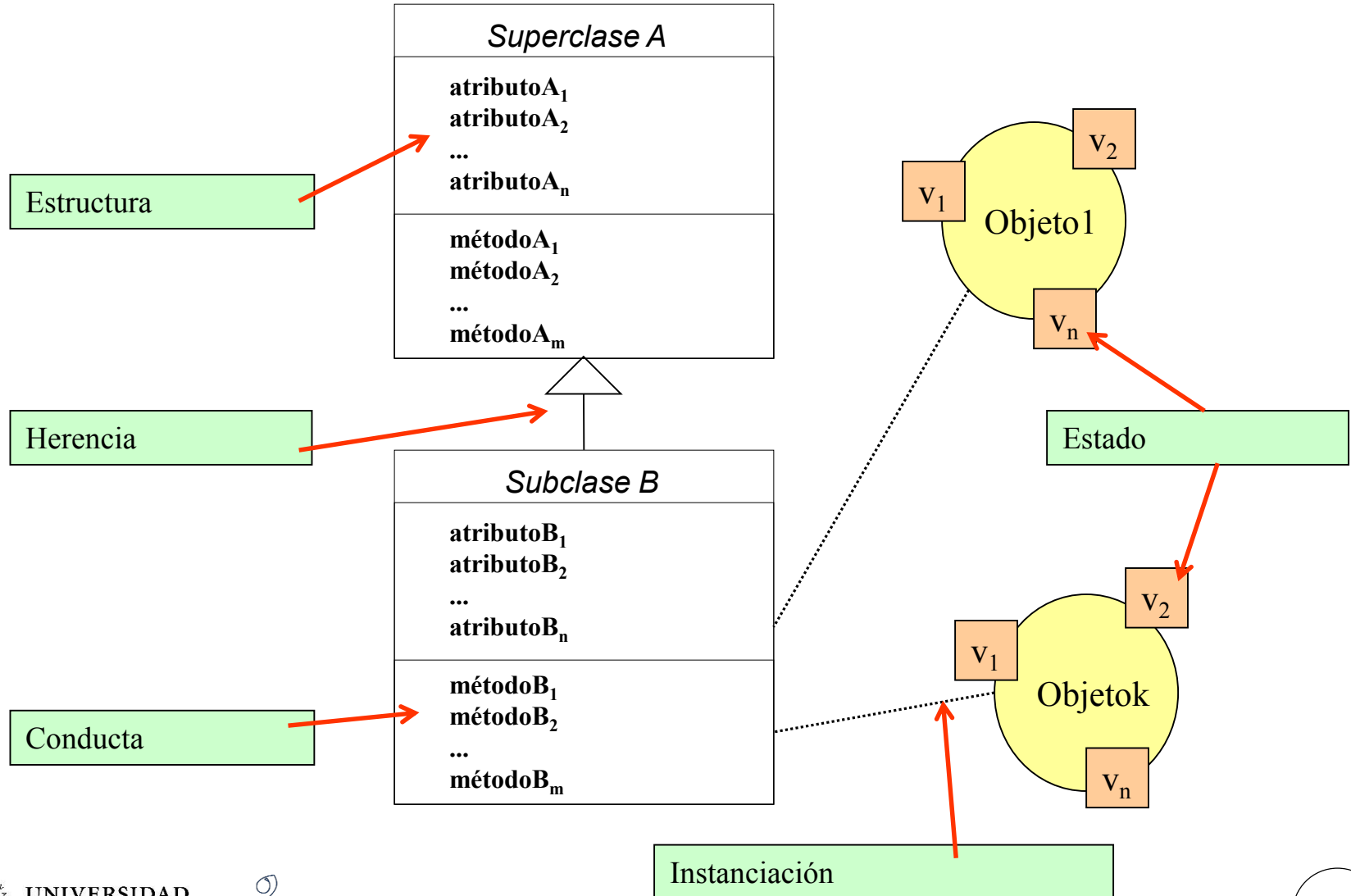
# Otros modelos de datos

## Orientación a objetos

- Filosofía del modelado orientado a objetos:
  - Abstractar un modelo de la realidad en forma de conjunto de objetos que interaccionan entre sí por medio de mensajes.
- Conceptos:
  - Estado / comportamiento.
  - Propiedades / métodos.
  - Encapsulamiento.
  - Herencia.
  - Polimorfismo.

# Otros modelos de datos

## Orientación a objetos





# Otros modelos de datos

## Orientación a objetos

- En el mundo de las Bases de Datos:
  - Primero: Orientación a objetos
    - Gran capacidad de modelado.
    - Dificultades de implementación del SGBD.
  - Después: Objeto-relacional
    - Solidez de SGBD relacionales.
    - Capacidad de modelado de la orientación a objetos.

# Otros modelos de datos

## Cubos de datos – sistemas multidimensionales

- Dos modelos de BD en la empresa:
  - Bases de datos operacionales:
    - Modelos relacional o de objetos.
    - OLTP.
  - Bases de datos analíticas:
    - Modelo multidimensional.
    - OLAP.

# Otros modelos de datos

## Cubos de datos – sistemas multidimensionales

- OLTP: On-line Transaction Processing
  - Aplicaciones OLTP
    - Soporte a operaciones diarias: estructuradas, repetitivas.
  - Requieren datos detallados y al día.
  - Las operaciones afectan fundamentalmente a *pocos* registros a los cuales se accede principalmente a través de la clave primaria.
  - Es de vital importancia:
    - Consistencia.
    - Fiabilidad.
  - Criterio esencial de rendimiento:
    - Optimizar la gestión de transacciones.

# Otros modelos de datos

## Cubos de datos – sistemas multidimensionales

- OLAP: On-line Analytical Processing
  - Aplicaciones OLAP
    - Orientadas al soporte de decisiones.
    - No son tan previsibles.
  - Datos importantes
    - Consolidados.
    - Resumidos.
    - De tipo histórico.
  - Estos sistemas necesitan resolver consultas:
    - Complejas.
    - Fundamentalmente “ad hoc”, aunque también predefinidas.
    - Pueden involucrar a millones de registros.
  - Necesita estructuras de datos diferentes.
  - El procesamiento de consultas y el tiempo de respuesta son más importantes que el control de transacciones.

# Otros modelos de datos

## Cubos de datos – sistemas multidimensionales

BD Operacional		Orientada a ...	DBAs, Programadores de aplicaciones, Operadores ...	Función	Soporte al funcionamiento diario de la organización.	Procesamiento	De transacciones.
DW			Ejecutivos.		Soporte a la toma de decisiones.		Analítico.
Datos	Actuales, aislados, relacionales...	Uso	Repetitivo.	Acceso	Lectura/escritura. Transacciones simples.	Crucial	Gestión de transacciones. Consistencia de los datos.
	Históricos, resumidos, integrados, multidimensionales...		"Ad hoc".		Lectura. Consultas complejas.		Gestión de consultas. Oportunidad de los datos.

# Otros modelos de datos

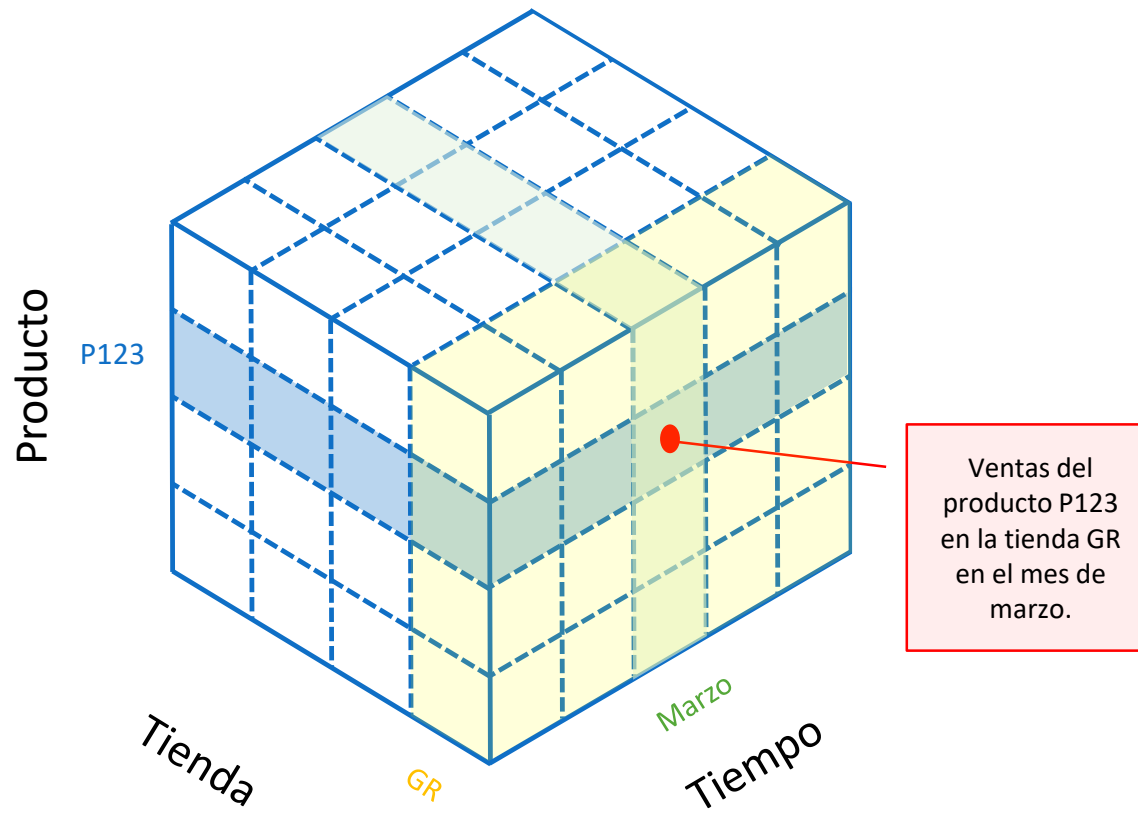
## Cubos de datos – sistemas multidimensionales

- Modelado Dimensional:

- Técnica de modelado que permite organizar los datos como un conjunto de medidas que están descritas por aspectos comunes del negocio.
- Es de probada utilidad para agregar/desagregar datos y reordenarlos para el análisis.
- Está enfocado para trabajar sobre datos numéricos: valores, conteos, ratios, ....
- Si se compara con otros modelos más propios de sistemas operacionales (modelado E/R, los diagramas de clases, o modelos de datos lógicos como el relacional o el objeto-relacional), puede considerarse más fácil de entender y usar (es más sencillo) y visualmente más atractivo.

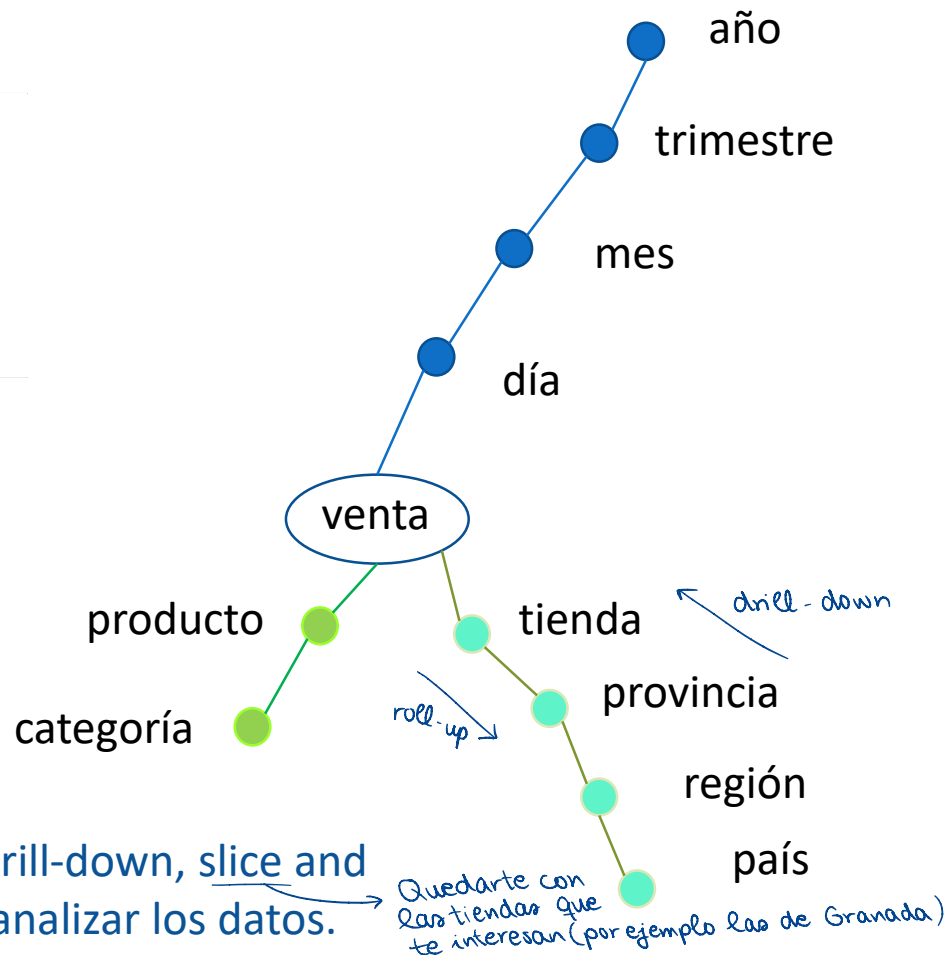
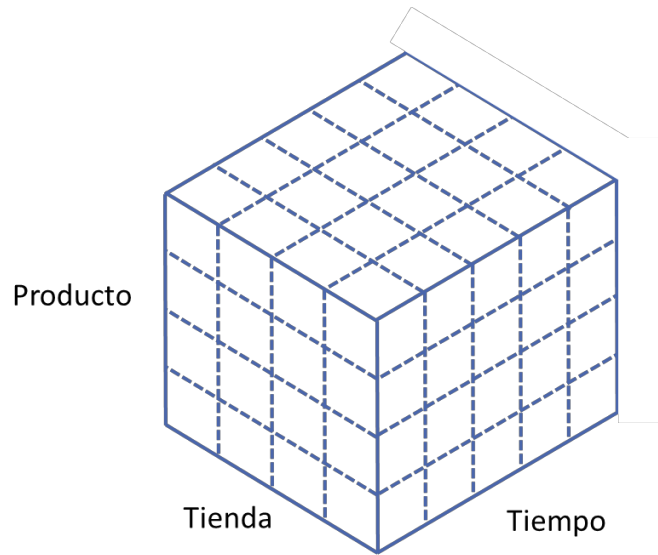
# Otros modelos de datos

## Cubos de datos – sistemas multidimensionales



# Otros modelos de datos

## Cubos de datos – sistemas multidimensionales



Las operaciones OLAP (roll-up, drill-down, slice and dice, pivot) permiten consultar/analizar los datos.



# Contenidos

- Definición de modelo de Datos
- El modelo de datos relacional
  - Estructura de datos
  - Restricciones de integridad
- Otros modelos de datos

# Imágenes

- Imágenes tomadas de [Pixabay](#)
  - Portada
    - Imagen de [Manfred Steger](#)
  - Cabecera
    - Imágenes de [Gerd Altmann](#)