

Tecnología de Gestión de Datos

Tema 1: Evolución histórica

MU Ingeniería y Tecnología de Sistemas Software
Profesor: Juan Carlos Casamayor

Tecnología de Gestión de Datos:

Tecnología (software) para la gestión de grandes volúmenes de datos en un computador.

- modelo de estructuración de los datos
- lenguaje de manipulación
- implementación de la tecnología

| Etapa | Década | Sistemas | |
|-------|--------|----------------------------|--|
| 1ª | 1950 | Sistemas de ficheros | Sistemas de ficheros en cinta magnética |
| 2ª | 1960 | | Sistemas de ficheros en disco |
| 3ª | 1970 | Sistemas de bases de datos | SGBD prerrelacionales (jerárquico, red) |
| 4ª | 1980 | | SGBD relacionales |
| 5ª | 1990 | | SGBD postrelacionales (objeto-relacionales, orientados a objetos). |
| 6ª | 2000 | Big Data | SGBD NoSQL |

Sistemas de ficheros: conceptos previos:

Fichero:

- estructura de datos en memoria secundaria
- secuencia de registros.
- organizaciones de ficheros: ordenado, desordenado, con direccionamiento calculado (Hash).

Índice:

- estructura de datos en memoria secundaria
- permite el acceso directo (y secuencial) a los registros de un fichero por el valor de un campo(s).
- organizaciones de índices: árbol B, árbol B+, de mapa de bits, de Join, ...

Sistemas de ficheros.

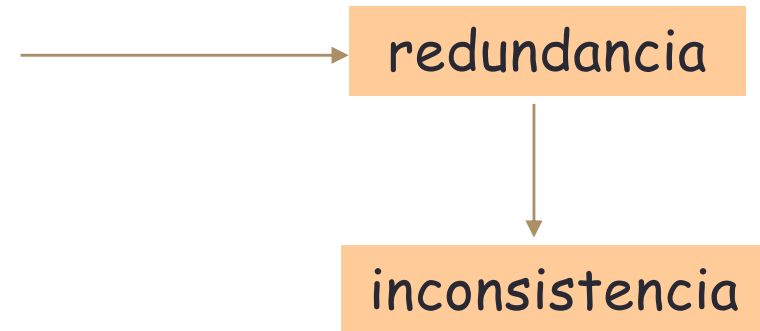
| Sistemas de ficheros | Etapa | Década | Sistemas | Eventos |
|----------------------|-------|--------|---|--|
| | 1ª | 50 | Sistemas de ficheros en cinta magnética | 1945: aparición de la cinta magnética. 1965: aparición del disco magnético 1965: sistemas generalizados de acceso a ficheros (métodos) 1965: índices en árbol AVL 1972: índices en árbol B |
| | 2ª | 60 | Sistemas de ficheros en disco | |

Sistemas de ficheros.

| Sistemas de ficheros | Etapa | Década | Sistemas | Características |
|----------------------|-------|--------|---|--|
| | 1ª | 50 | Sistemas de ficheros en cinta magnética | <ul style="list-style-type: none">- dispositivo de acceso secuencial- acceso secuencial a los registros: ficheros ordenados y desordenados. |
| | 2ª | 60 | Sistemas de ficheros en disco | <ul style="list-style-type: none">- dispositivo de acceso directo- acceso directo a los registros: ficheros hash, ficheros indexados- múltiples caminos de acceso (índices). |

Limitaciones de los sistemas de ficheros.

- desarrollo de aplicaciones independientes para los distintos departamentos de la organización.



- dependencia de los programas respecto a los datos: la descripción del fichero y el acceso formaba parte del código de las aplicaciones.

- limitaciones en el acceso concurrente, y en el control de la seguridad y la consistencia.

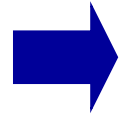
| Etapa | Década | Sistemas | |
|-------|--------|----------------------------|--|
| 1ª | 1950 | Sistemas de ficheros | Sistemas de ficheros en cinta magnética |
| 2ª | 1960 | | Sistemas de ficheros en disco |
| 3ª | 1970 | Sistemas de bases de datos | SGBD prerrelacionales (jerárquico, red) |
| 4ª | 1980 | | SGBD relacionales |
| 5ª | 1990 | | SGBD postrelacionales (objeto-relacionales, orientados a objetos). |
| 6ª | 2000 | Big Data | SGBD NoSQL |

Sistemas de bases de datos

| Sistemas de bases de datos | Etapa | Década | Sistemas | Eventos |
|----------------------------|-------|--------|---|---|
| | 3ª | 70 | SGBD prerrelacionales (jerárquico, red) | <p>1969: IMS (IBM). Modelo jerárquico.</p> <p>1969: primer informe CODASYL (DDL, DML). Modelo en red</p> <p>1970: Codd propone el Modelo relacional.</p> <p>1971: nuevo informe CODASYL (esquema, subesquema).</p> <p>1973: Inicio del proyecto INGRES (Berkeley).</p> <p>1974: Inicio del proyecto System R (IBM)</p> |
| | 4ª | 80 | SGBD relacionales | <p>1975: arquitectura ANSI/SPARC.</p> <p>1978: nuevo informe CODASYL (esquema, subesquema, esquema interno).</p> <p>1986: estándar SQL1 de ANSI.</p> |
| | 5ª | 90 | SGBD postrelacionales | <p>1990: "The objet database system manifesto".</p> <p>1992: nuevo estándar SQL2.</p> <p>1995: "The third manifesto".</p> <p>1996: "Active database system manifesto".</p> <p>1999: nuevo estándar SQL3.</p> <p>2003, 2005, 2008: extensiones del SQL3</p> |

Sistemas de bases de datos: conceptos previos

Base de
datos:




colección estructurada de
datos en memoria secundaria

Los mecanismos de estructuración de datos (estructuras de datos) que se pueden utilizar dependen del sistema informático (SGBD) con el que se vaya a crear y manipular la base de datos



Sistema de gestión de bases de datos (SGBD)

SGBD:  herramienta (software) para la gestión (creación y manipulación) de bases de datos.

SGBD



modelo de datos



- ✓ tipos de estructuras de datos
- ✓ operadores asociados (lenguajes)

Características de la tecnología de bases de datos: superación de las limitaciones de los sistemas de ficheros.

soporte del SI
de la
organización



- ✓ integración de la información de la organización
- ✓ persistencia de datos


abstracción
de datos



- ✓ descripción unificada de los datos e independiente de las aplicaciones: esquema de base de datos
- ✓ independencia de las aplicaciones respecto a la representación física (implementación) de los datos: arquitectura ANSI/SPARC

Características de la tecnología de bases de datos

servicio a
distintos usuarios



- ✓ definición de vistas parciales de los datos para distintos usuarios
- ✓ acceso simultáneo para distintos usuarios (conurrencia)

consistencia
de los datos



- ✓ ejecución consistente y recuperación de transacciones: principio ACID


seguridad de
los datos



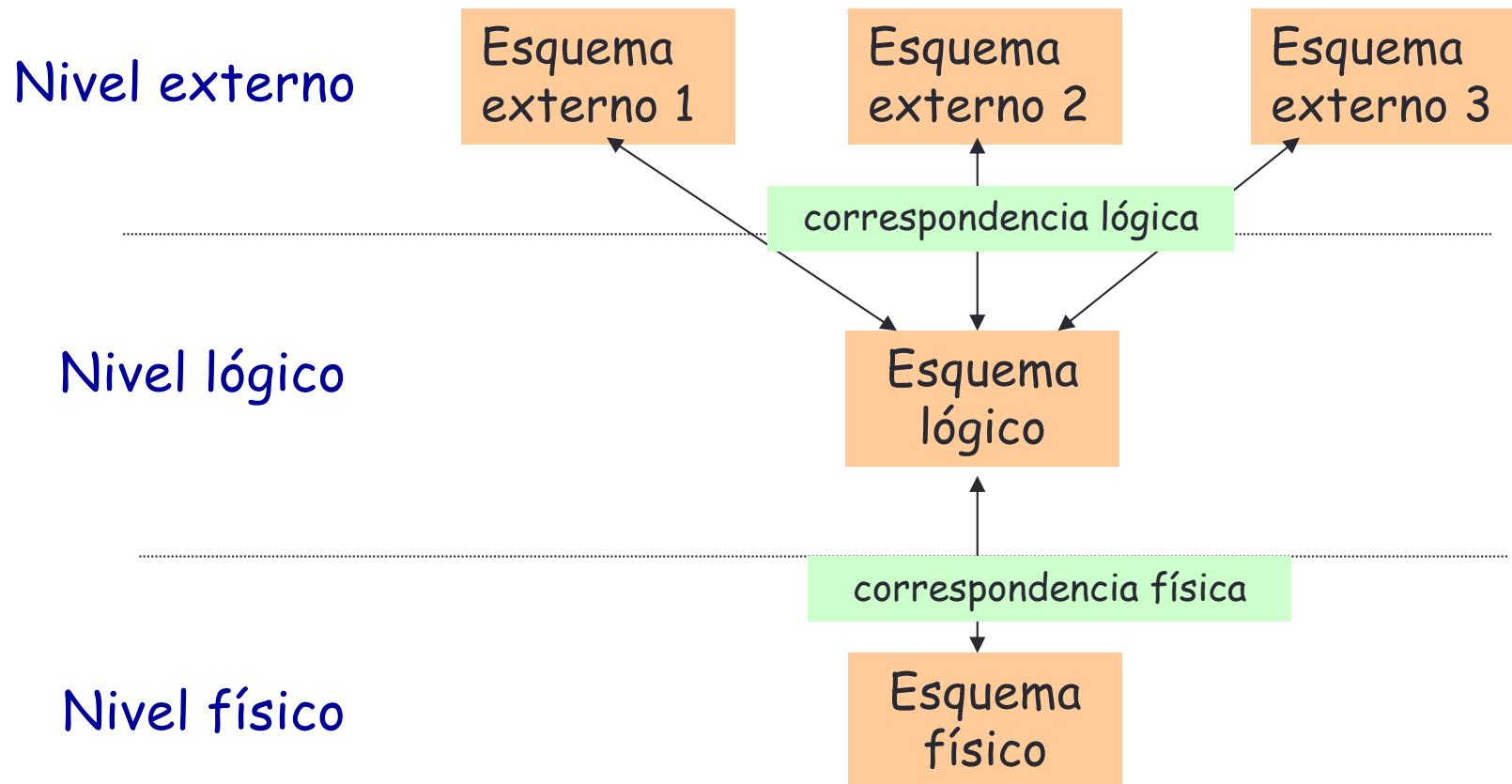
- ✓ privacidad de los datos.

Evolución de los modelos de datos.

| SGBD | modelo | estructuras de datos |
|--------------|------------|------------------------------------|
| jerárquicos | jerárquico | registro, árbol |
| en red | red | registro, lista (set) |
| relacionales | relacional | registro (tupla), tabla (relación) |

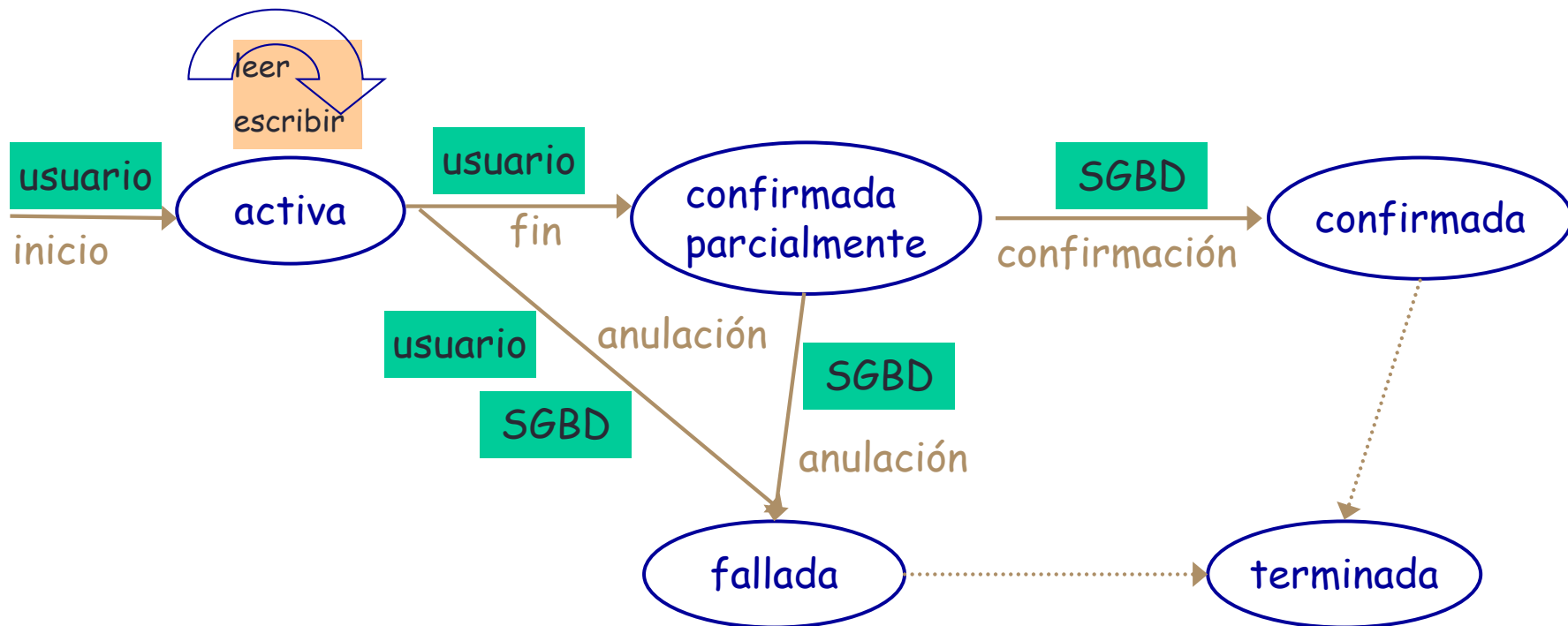


Arquitectura ANSI/SPARC.



Correspondencia: definición de cada *elemento* de un esquema en términos de *elementos* del esquema inmediatamente inferior

Procesamiento de transacciones



Transacción: unidad lógica de ejecución

Principio **ACID** para el procesamiento de transacciones:

Atomicidad

Persistencia

Consistencia

Aislamiento

Características de la tecnología de bases de datos.

Ideas básicas:

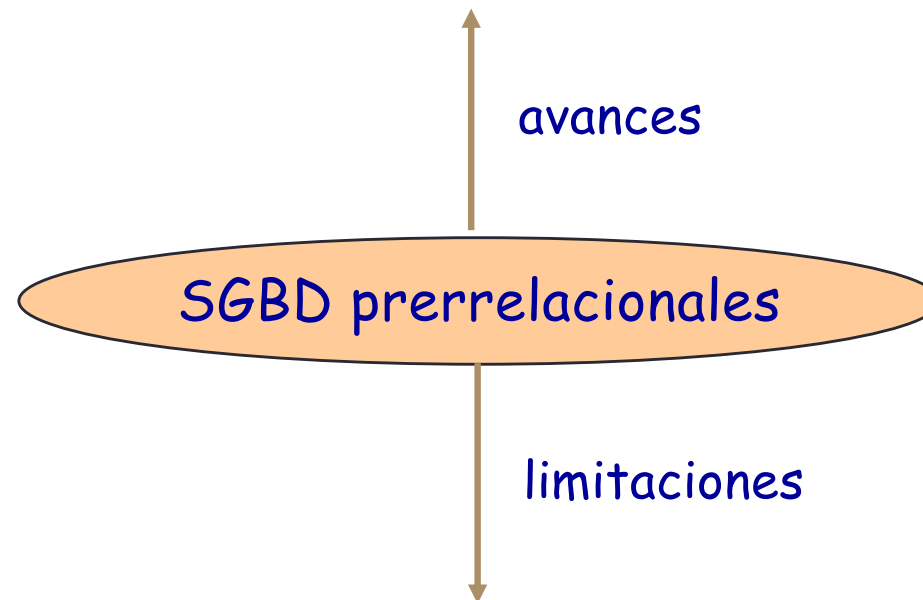
- Datos estructurados: distintos modelos de datos
- Concepto de esquema de base de datos: definición unificada de los datos
- Arquitectura ANSI/SPARC: distintos esquemas de la base de datos (independencia de datos)
- Principio ACID: ejecución consistente de transacciones en un entorno concurrente

Sistemas de bases de datos prerrelacionales.

| <i>SGBD</i> | <i>modelo</i> | <i>estructuras</i> | <i>SGBD</i> |
|--|---------------|-----------------------|---------------------------------|
| jerárquicos (1969) | jerárquico | registro, árbol | IMS (IBM) |
| en red (Codasyl) (1969, 1971, 1978) | red | registro, lista (set) | IDMS (Bull) DMS1100 (Univac) |

Sistemas de bases de datos prerrelacionales. (ver Anexo I)

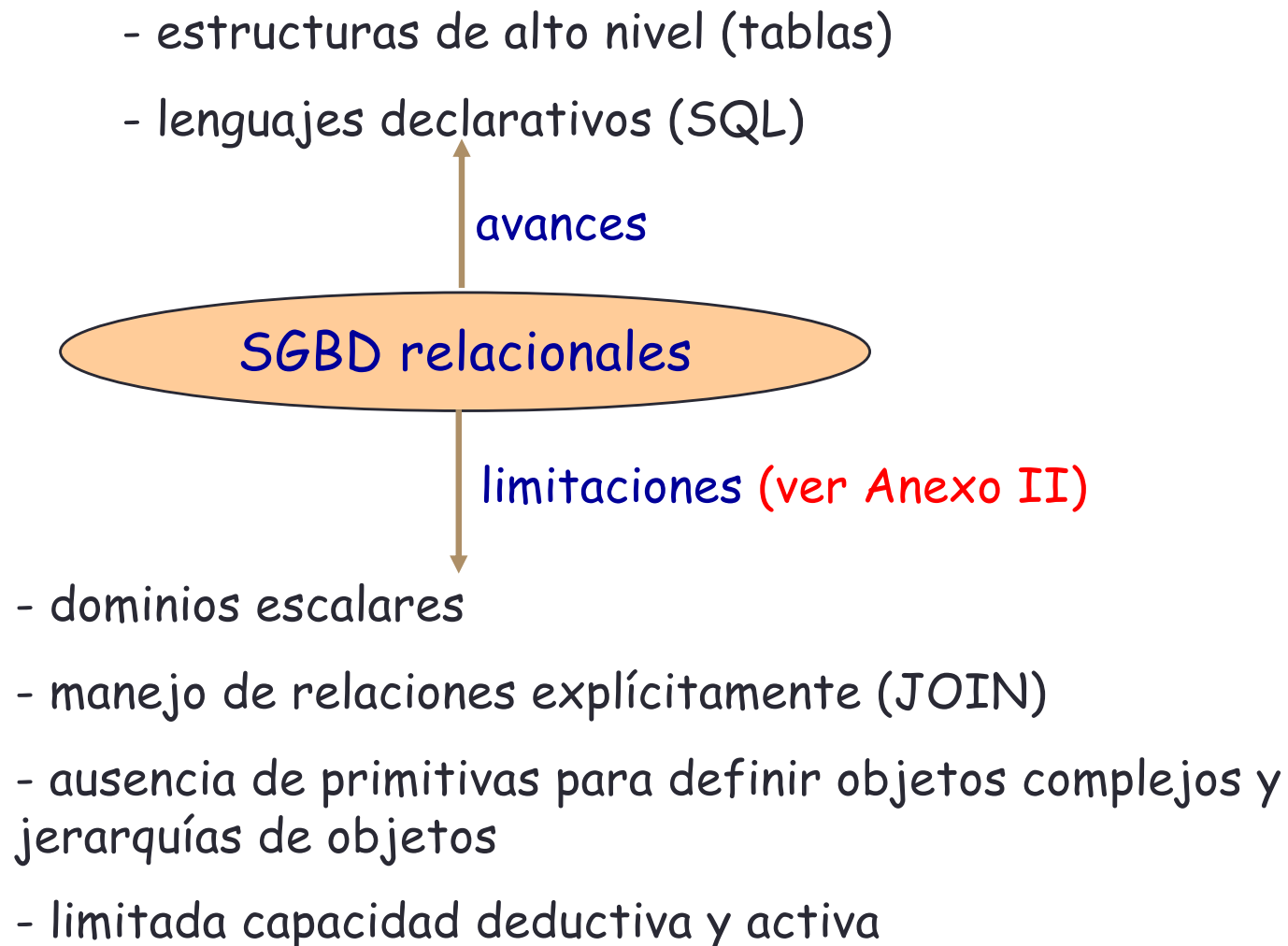
- integración de los datos: descripción unificada
- independencia de datos: esquemas de la base de datos



- estructuras de datos de bajo nivel (árbol, lista)
- lenguajes navegacionales (orientados al registro)

| Sistemas de bases de datos | Etapa | Década | Sistemas | Características |
|----------------------------|-------|--------|---|--|
| | 3ª | 70 | SGBD prerrelacionales (jerárquico, red) | <ul style="list-style-type: none"> - integración de los datos - descripción unificada de los datos - independencia lógica y física (débil) - estructuras de datos de bajo nivel (árboles, listas) - lenguajes navegacionales (orientados al registro) |
| | 4ª | 80 | SGBD relacionales | |
| | 5ª | 90 | SGBD postrelacionales | |

Sistemas de bases de datos relacionales.



Sistemas de bases de datos

| Etapa | Década | Sistemas | Características |
|-------|--------|---|---|
| 3ª | 70 | SGBD prerrelacionales (jerárquico, red) | <ul style="list-style-type: none"> - integración de los datos - descripción centralizada de los datos - independencia lógica y física (débil) - estructuras de datos de bajo nivel (árboles, listas) - lenguajes navegacionales (orientados al registro) |
| 4ª | 80 | SGBD relacionales | <ul style="list-style-type: none"> - independencia lógica y física (fuerte) - estructuras de datos de alto nivel (tablas) - lenguajes declarativos (orientados al conjunto) |
| 5ª | 90 | SGBD postrelacionales | |

Limitaciones del modelo relacional

- ✓ atributos simples sobre dominios escalares.
 - ✓ identificación de las entidades por valor (contenido)
 - ✓ manipulación de las relaciones explícitamente
 - ✓ ausencia de primitivas para representar entidades complejas (agregación)
 - ✓ ausencia de primitivas para representar jerarquías de entidades (generalización)
 - ✓ lenguaje de manipulación (SQL3) computacionalmente incompleto
 - ✓ operadores predefinidos y genéricos (INSERT, DELETE, UPDATE)
-
- ✓ limitaciones en la definición de información implícita: vistas recursivas
-
- ✓ limitaciones en la definición de comportamiento activo: disparadores

BD objeto-
relacionales

BD
deductivas

BD
activas

Sistemas de bases de datos

| Etapa | Década | Sistemas | Características |
|-------|--------|---|---|
| 3ª | 70 | SGBD prerrelacionales (jerárquico, red) | <ul style="list-style-type: none"> - integración de los datos - descripción centralizada de los datos - independencia lógica y física (débil) - estructuras de datos de bajo nivel (árboles, listas) - lenguajes navegacionales (orientados al registro) |
| 4ª | 80 | SGBD relacionales | <ul style="list-style-type: none"> - independencia lógica y física (fuerte) - estructuras de datos de alto nivel (relaciones) - lenguajes declarativos (orientados al conjunto) |
| 5ª | 90 | SGBD postrelacionales | <ul style="list-style-type: none"> - SQL3 (objeto-relacional, deductivo, activo) - SGBD orientados a objetos - Integración de BD: SGBD federadas, SGBD distribuidas - Nuevas aplicaciones: Datawarehouse, Multimedia, GIS (nuevos desarrollos tecnológicos) |

| Etapa | Década | Sistemas | |
|-------|--------|----------------------------|--|
| 1ª | 1950 | Sistemas de ficheros | Sistemas de ficheros en cinta magnética |
| 2ª | 1960 | | Sistemas de ficheros en disco |
| 3ª | 1970 | Sistemas de bases de datos | SGBD prerrelacionales (jerárquico, red) |
| 4ª | 1980 | | SGBD relacionales |
| 5ª | 1990 | | SGBD postrelacionales (objeto-relacionales, orientados a objetos). |
| 6ª | 2000 | Big Data | SGBD NoSQL |

Big Data

Nuevas aplicaciones de la tecnología de gestión de datos:

aplicaciones web

sistemas de información masivos

...



SGBD clásicos

limitaciones

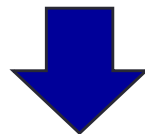
- datos muy estructurados
- poca escalabilidad
- rigidez de procesamiento: principio ACID

Big Data

Características:

- Datos semiestructurados o no-estructurados
- Ausencia de esquema de la base de datos
- Alta distribución de los datos
- Alta escalabilidad (horizontal)
- Procesamiento eventualmente consistente de transacciones.

...

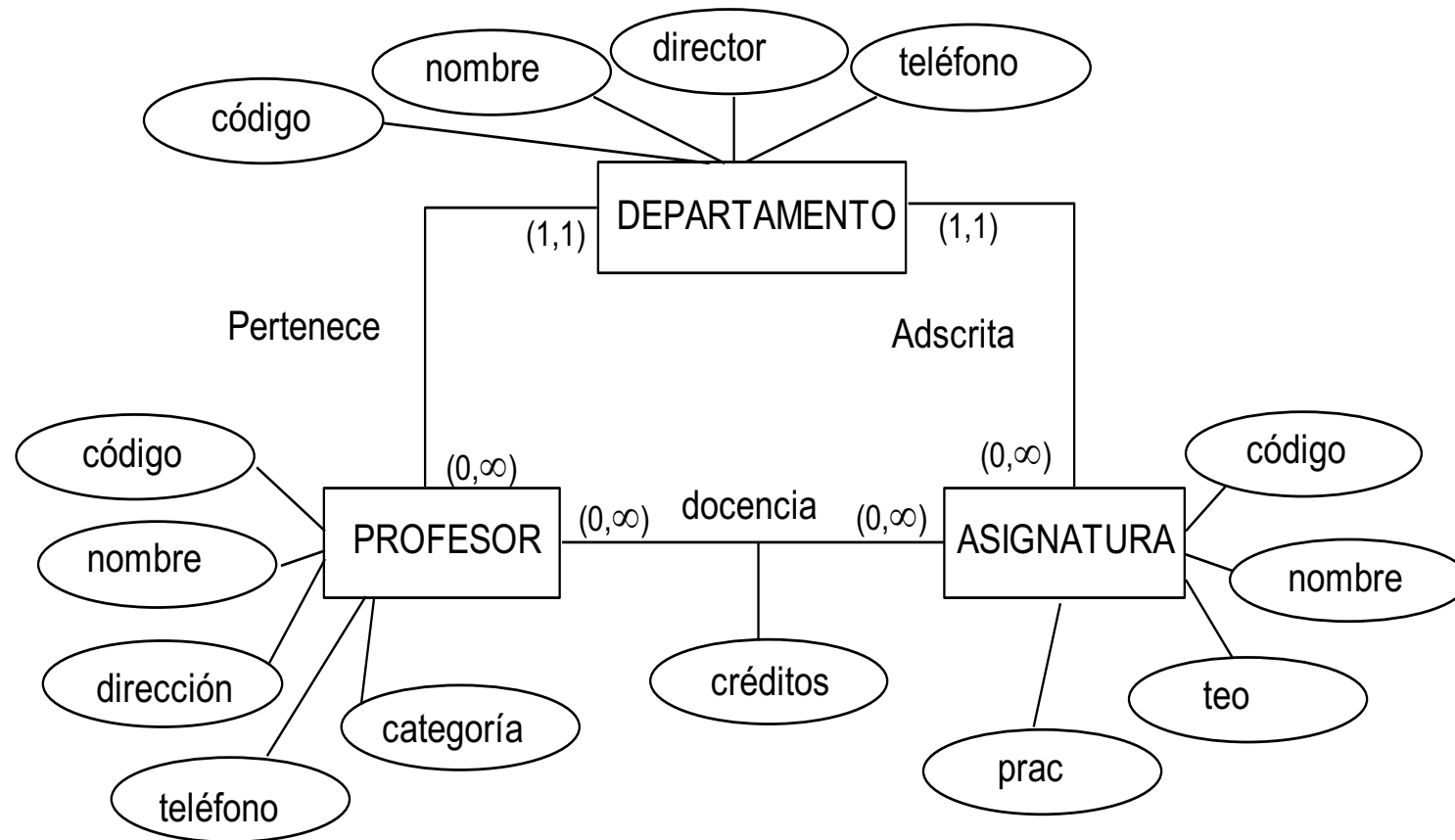


Sistemas NoSQL ("not only SQL")

Anexo I

Sistemas de bases de datos prerrelacionales

Sistemas de bases de datos prerrelacionales.



Esquema conceptual

Hay dos departamentos con la siguiente información:

- código = DSIC, nombre = Sistemas Informáticos y Computación, director = Juan García, teléfono = 3570
- código = DFA, nombre = Física, director = José Ruíz, teléfono = 3540.

Hay tres profesores:

- código = JCP, nombre = Juan Cerdá Pérez, dirección = Olta 23, teléfono = 3222, categoría = TEU
- código = LBP, nombre = Luis Bos Pérez, dirección = Jesús 91, teléfono = 3545, categoría = TEU
- código = PMG, nombre = Pedro Martí García, dirección = Cuenca 12, teléfono = 3412, categoría = TEU

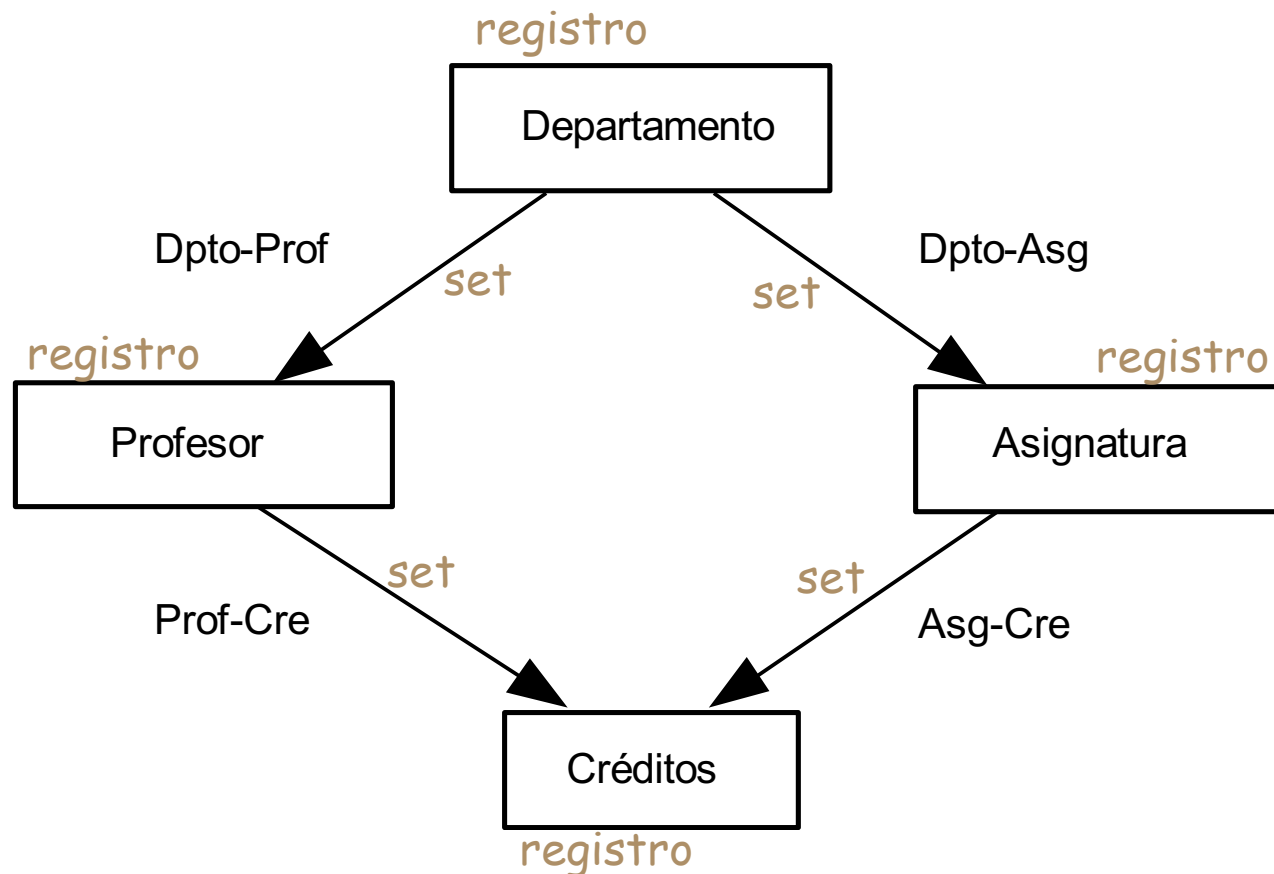
Hay tres asignaturas:

- código = AD1, nombre = Algoritmos y estructuras de datos 1, teo = 3, prac = 3
- código = IP, nombre = Introducción a la programación, teo = 1.5, prac = 1.5
- código = AD2, nombre = Algoritmos y estructuras de datos 2, teo = 3, prac = 3

Respecto a las relaciones, la información es la siguiente:

- los tres profesores pertenecen al DSIC.
- las tres asignaturas están adscritas al DSIC.
- Juan Cerdá Pérez imparte 9 créditos de AD1 y 9 de IP, Luis Bos Pérez imparte 9 créditos de IP y Pedro Martí García imparte 9 créditos de AD1.

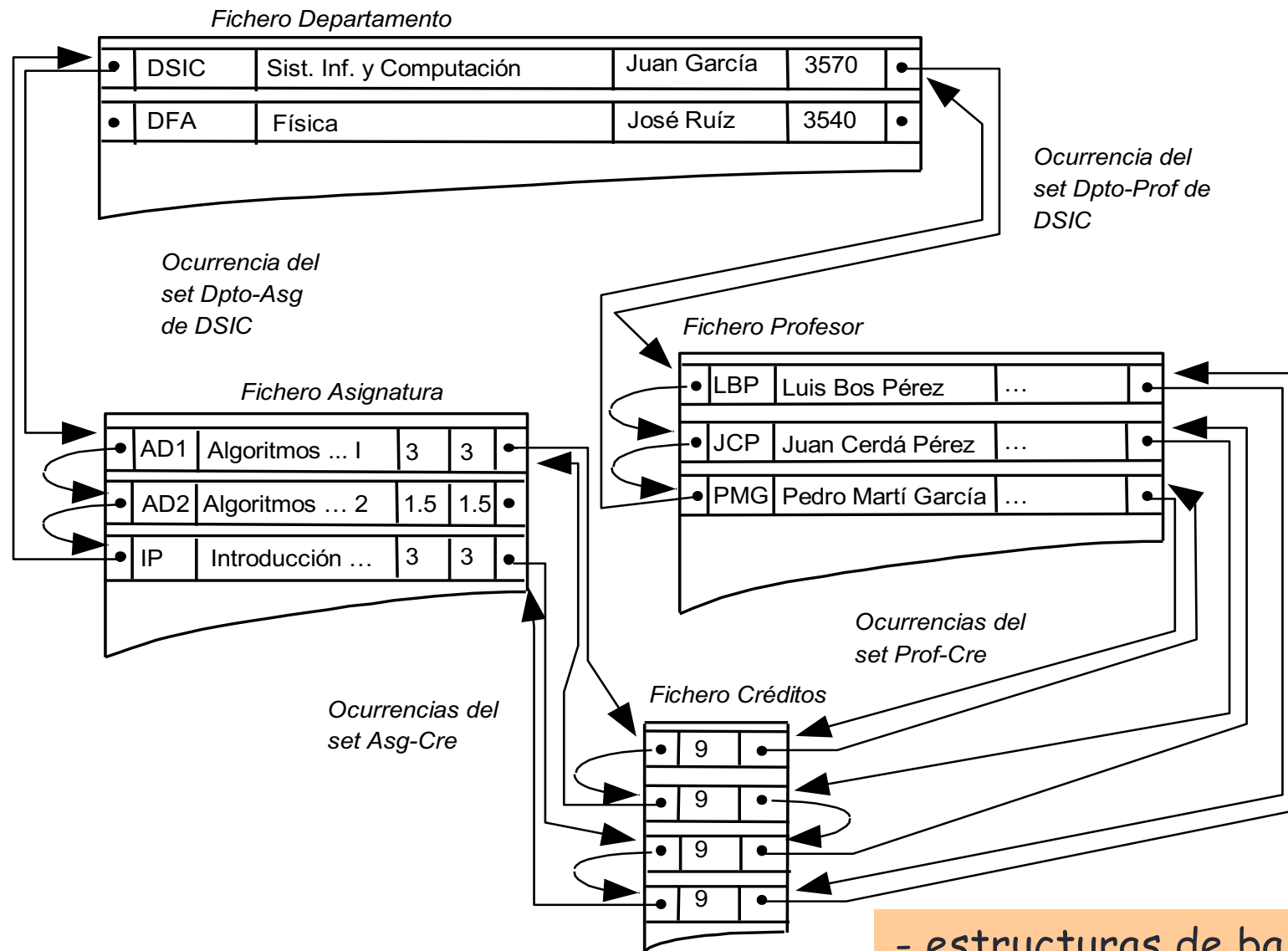
Modelo en red de CODASYL.



Esquema de la base de datos (CODASYL):

- Ficheros de registros
- Listas circulares de registros (SET)

Modelo en red de CODASYL.



- estructuras de bajo nivel

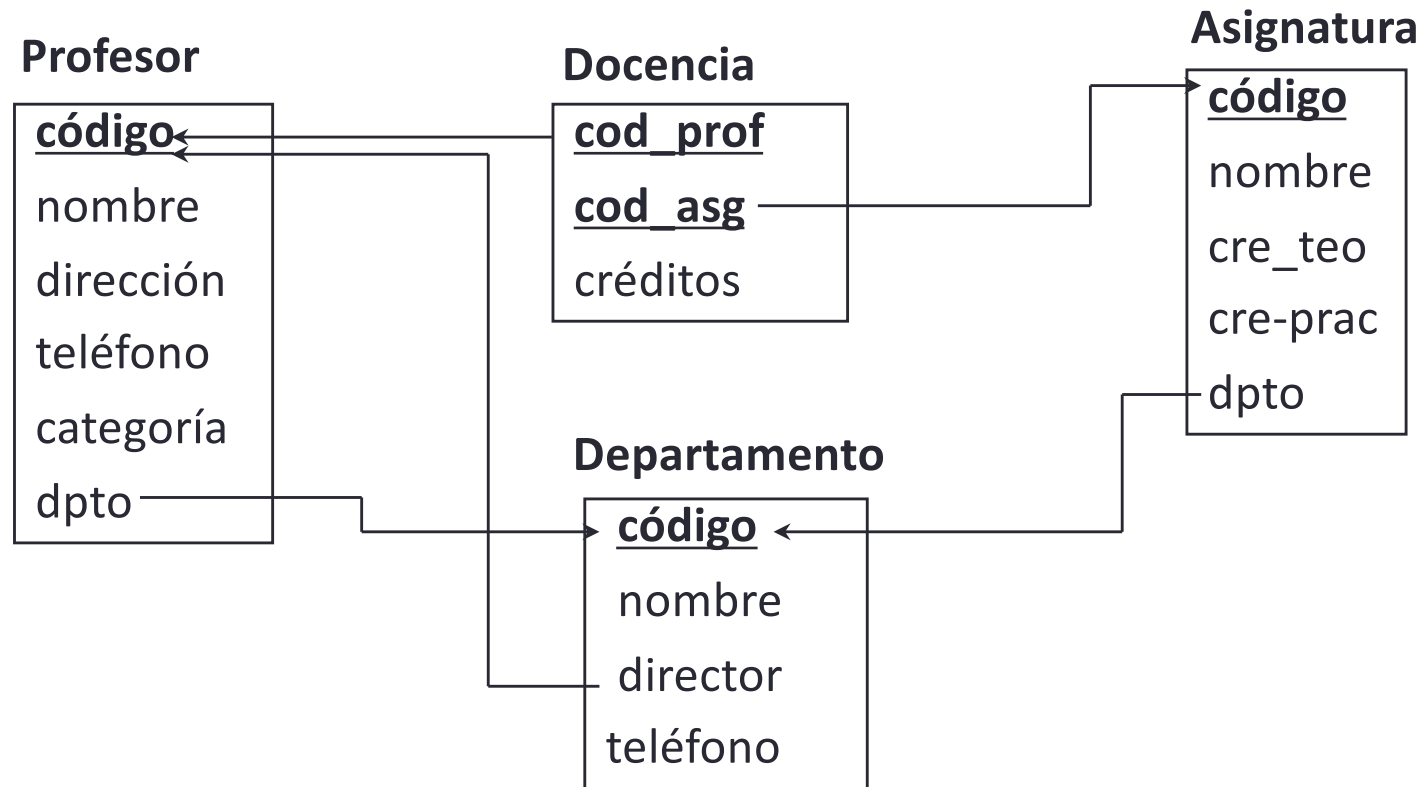
Modelo en red de CODASYL.

Consulta: Obtener las asignaturas impartidas por el profesor de código LBP.

```
Profesor.código ← "LBP"  
FIND Profesor USING clave_prof  
FIND FIRST Créditos RECORD WITHIN Prof_Cre SET  
mientras no_error  
    FIND OWNER RECORD WITHIN Asg_Cre SET  
    escribir (datos_asignatura)  
    FIND NEXT Créditos RECORD WITHIN Prof_Cre SET  
fin_mientras
```

- lenguajes navegacionales

Sistemas de bases de datos relacionales.



Esquema de la base de datos (relacional):

- tablas de tuplas (registros)

Relación Departamento

| Código | Nombre | Director | Teléfono |
|---------------|-------------------------------------|-----------------|-----------------|
| DSIC | Sistemas Informáticos y Computación | Juan García | 3570 |
| DFA | Física | José Ruíz | 3540 |

Relación Profesor

| Código | Nombre | Dirección | Categoría | Teléfono | Dpto |
|---------------|--------------------|------------------|------------------|-----------------|-------------|
| JCP | Juan Cerdá Pérez | Olta 23 | TEU | 3222 | DSIC |
| LBP | Luis Bos Pérez | Jesús 91 | TEU | 3545 | DSIC |
| PMG | Pedro Martí García | Cuenca 12 | TEU | 3412 | DSIC |

Relación Asignatura

| Código | Nombre | Teo | Prac | Dpto |
|---------------|--------------------------------------|------------|-------------|-------------|
| AD1 | Algoritmos y estructuras de datos I | 3 | 3 | DSIC |
| IP | Introducción a la programación | 1.5 | 1.5 | DSIC |
| AD2 | Algoritmos y estructuras de datos II | 3 | 3 | DSIC |

Relación Docencia

| cod_prof | cod_asg | créditos |
|-----------------|----------------|-----------------|
| JCP | AD1 | 9 |
| JCP | IP | 9 |
| LBP | IP | 9 |
| PMG | AD1 | 9 |

- estructuras de alto nivel

Sistemas de bases de datos relacionales.

Consulta: Obtener las asignaturas impartidas por el profesor de código LBP.

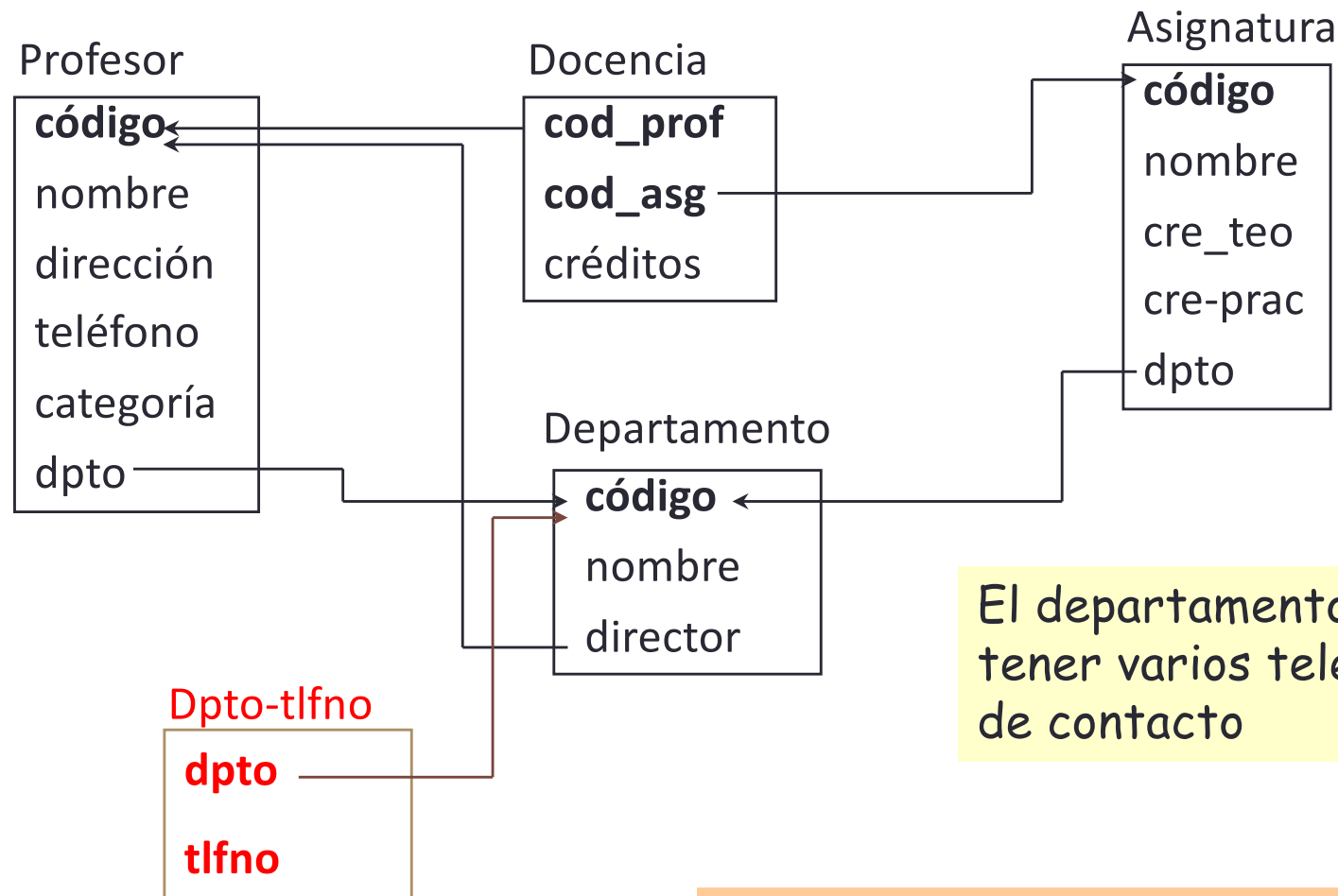
```
SELECT A.nombre  
FROM Asignatura A, Docencia D  
WHERE A.codigo=D.cod_asg AND D.cod_prof='LBP'
```

- lenguajes declarativos

Anexo II

Limitaciones del modelo relacional de datos

Sistemas de bases de datos relacionales: limitaciones



El departamento puede tener varios teléfonos de contacto

Limitaciones del modelo relacional (SQL92) en la definición de información compleja

- dominios escalares
- ausencia de primitivas para definir objetos complejos

Sistemas de bases de datos relacionales: limitaciones

Consulta: Obtener las asignaturas impartidas por el profesor de código LBP.

```
SELECT Asignatura.nombre  
FROM Asignatura, Docencia  
WHERE Asignatura.código = Docencia.cod_asg  
and  
Docencia.cod_prof = 'LBP'
```

Consulta de la base de datos

- lenguajes declarativos (SQL)
- manipulación explícita de las relaciones (JOIN) para reconstruir la información de un objeto.

Sistemas de bases de datos relacionales: limitaciones

Se desea registrar en la base de datos información sobre los prerrequisitos (incompatibilidades) entre las asignaturas.

La propiedad de ser prerrequisito entre dos asignaturas es una propiedad que se transmite transitivamente.

A partir de un información básica sobre la relación de prerrequisitos entre las asignaturas a un primer nivel (prerrequisitos directos), se puede derivar información sobre la relación de prerrequisitos a cualquier nivel (prerrequisitos indirectos), es decir a través de otras asignaturas:

“Si la asignatura X tiene como prerrequisito a la asignatura Y e Y tiene como prerrequisito a la asignatura Z, entonces X tiene como prerrequisito a Z”.

Sistemas de bases de datos relacionales: limitaciones

Se desea registrar en la base de datos información sobre los prerrequisitos (incompatibilidades) entre las asignaturas.

Definición declarativa de la propiedad "ser prerrequisito":

Regla 1: $\text{Prerrequisito}(x,y) \leftarrow \text{prer}(x,y)$

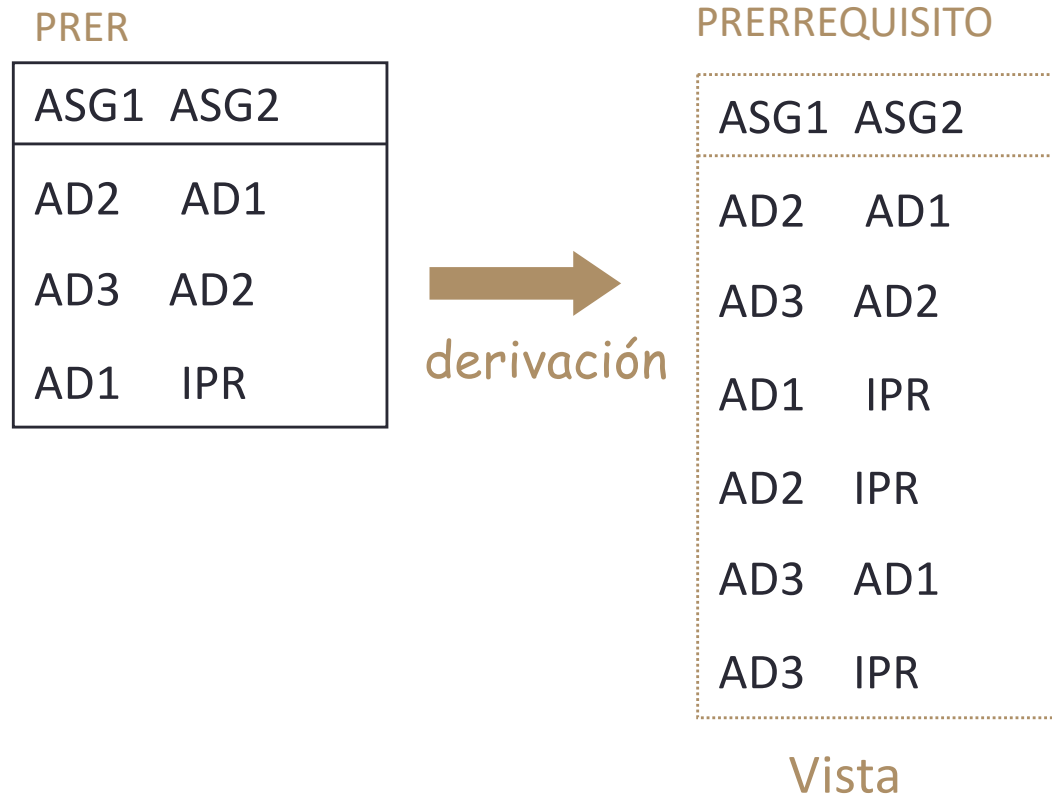
Regla 2: $\text{Prerrequisito}(x,y) \leftarrow \text{prer}(x,z) \wedge \text{prerrequisito}(z,y)$

Prer: representa la propiedad "ser prerrequisito directo"

Prerrequisito: representa la propiedad "ser prerrequisito directa o indirectamente"

Sistemas de bases de datos relacionales: limitaciones

Limitaciones del modelo relacional (SQL92) en la definición de vistas:



Sistemas de bases de datos relacionales: limitaciones

Limitaciones del modelo relacional (SQL92) en la definición de vistas:

Relación derivada PRERREQUISITO  **VISTA**

```
SQL: CREATE VIEW PRERREQUISITO AS
      SELECT asg1, asg2
      FROM PRER
           UNION
      SELECT asg1, asg2
      FROM PRER, PRERREQUISITO
      .....
```

- limitaciones en la definición de información implícita: vistas recursivas.

Sistemas de bases de datos relacionales: limitaciones

Limitaciones del modelo relacional (SQL92) en la definición de comportamiento activo:



- limitaciones en la definición de comportamiento activo: disparadores.