

Diseño de Bases de Datos

Introducción a la Administración
de BD

Tipos de usuarios

- Distintos tipos de usuarios / responsabilidades / competencias
 - Usuario de BD
 - interactúan con la BD a través de aplicaciones
 - Desarrollador (de aplicaciones)
 - diseño y desarrollo de la aplicación, incluyendo la *estructura* de la BD
 - estimación de requisitos de almacenamiento, eficiencia, seguridad, etc.
 - ¿quién *cuida* del DBMS que permita hacer todo esto de forma adecuada?

...

– Administrador de la BD (DBA)

- instala y actualiza el DBMS y las herramientas asociadas (¿decidir el DBMS a utilizar?)
- establece y reserva el sistema de almacenamiento
- crea los objetos primarios (tablas, vistas, índices), una vez que los desarrolladores han realizado el diseño
- modifica la estructura de la base de datos, integrando la información de distintas aplicaciones
- gestiona el acceso de usuarios y sus privilegios
- establece políticas de seguridad y las monitoriza
- monitoriza y optimiza el rendimiento de la BD
- establece procedimientos de copias / recuperaciones
- gestiona el soporte técnico con la distribuidora del DBMS

Tareas típicas

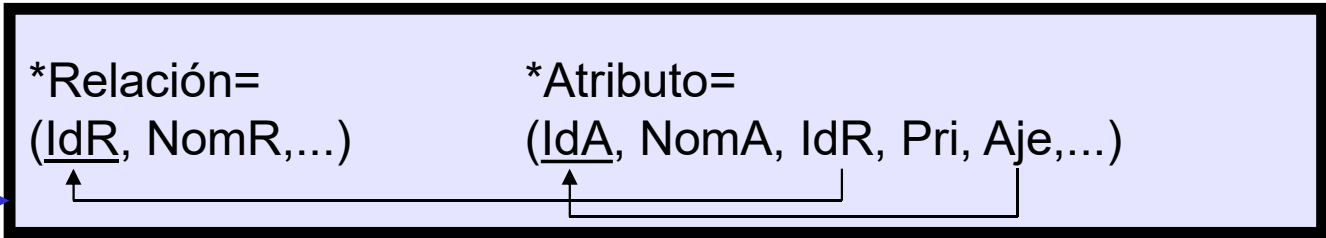
- Instalar el DBMS (inc. evaluación del hw)
 - estructura de almacenamiento (Oracle): *tablespaces*, archivos de datos, discos de almacenamiento
 - efecto en el rendimiento general del sistema
- Creación, arranque y parada de BD
 - *layout* de ficheros, inicialización, tamaño de bloque, ficheros de log, ...
- Gestión de usuarios
 - establecimiento de políticas de seguridad, gestión de usuarios y recursos

...

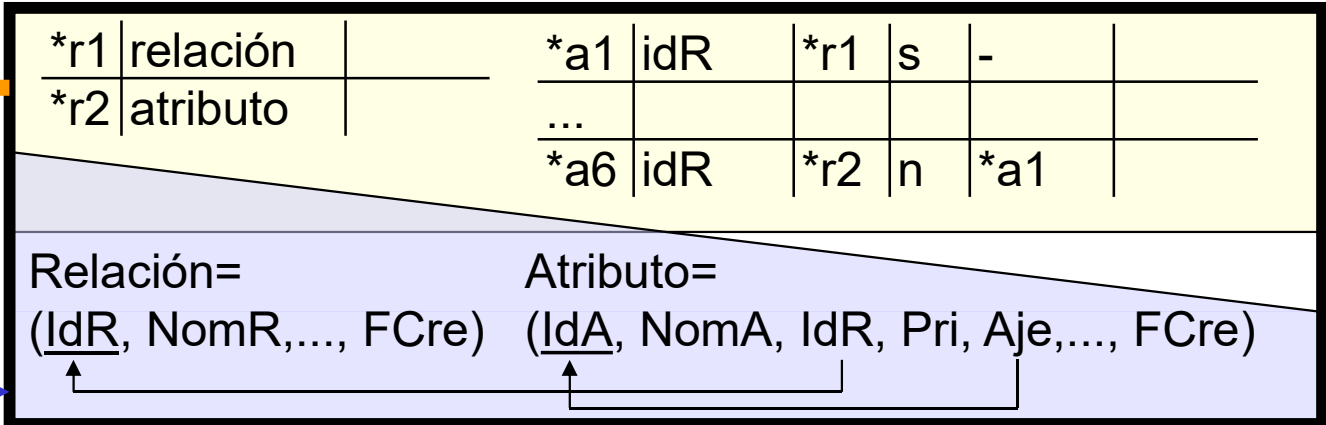
- Implementación del diseño (lógico)
 - objetos del esquema: tablas, campos, claves, vistas, ...
 - estructuras por defecto, índices primarios, ... (d. físico)
- Optimización del rendimiento / ajuste / *tuning*
 - optimización de consultas: árboles relacionales, rutas de acceso y planes de ejecución
 - optimización sintáctica, semántica y estadística; reglas, costes y *pistas*
 - optimización de las estructuras de datos del diseño físico: índices y su influencia en las consultas

Metadatos

- Datos
- Metadatos: definición, uso
 - » ¿son diferentes a los datos?
 - » ¿se almacenan por separado?
 - » ¿se describen con distintos modelos de datos?
 - » ¿existe un esquema para los metadatos —meta-esquema—?
 - » ¿existen meta-esquemas internos y externos?
 - » ¿se gestionan igual que los datos?
 - » ¿se pueden cambiar los esquemas dinámicamente?
 - » ¿los cambios son reflejados automáticamente en los datos?

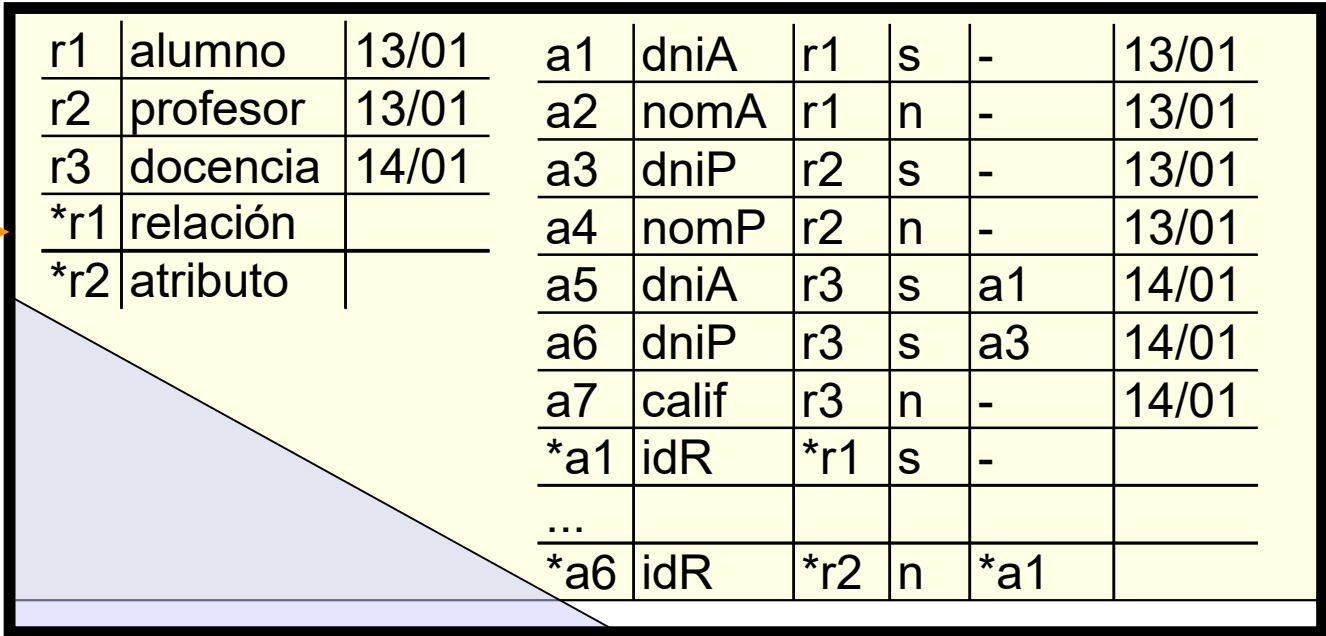


③
ESQUEMA
DICCIONARIO



DATOS
MODELO

②
ESQUEMA
APLICACIÓN

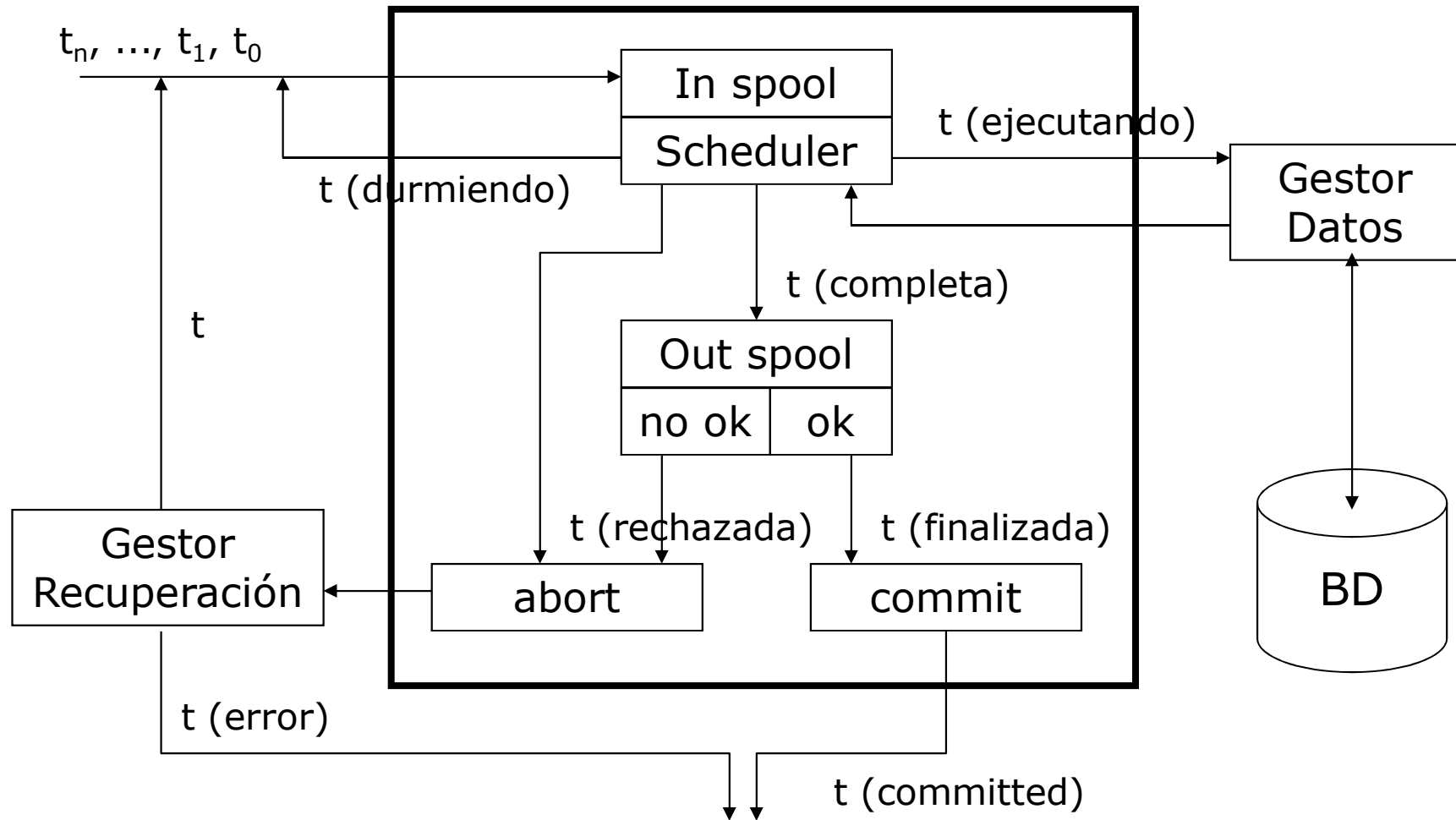


DATOS
DICCIONARIO

Concurrencia / transacciones

- Concurrencia
 - Acciones sobre las mismas tablas al mismo tiempo
 - Problemas: update *perdido*, lectura *sucia*, problema *fantasma*
- Sistema transaccional
 - Ejecución todo-o-nada, sin interferencias entre programas; (cuasi-)paralelos
- Protocolos de bloqueos
 - Hace que los accesos a los datos por distintas transacciones sean mutuamente excluyentes

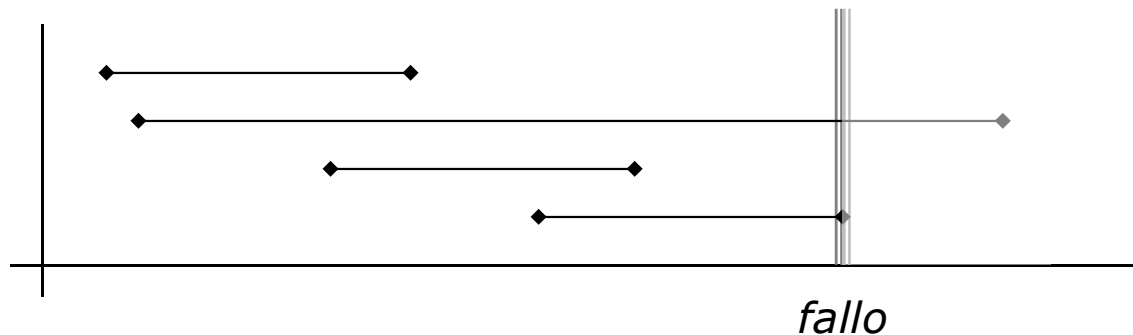
...



Recuperación

» x/0, deadlock, ...

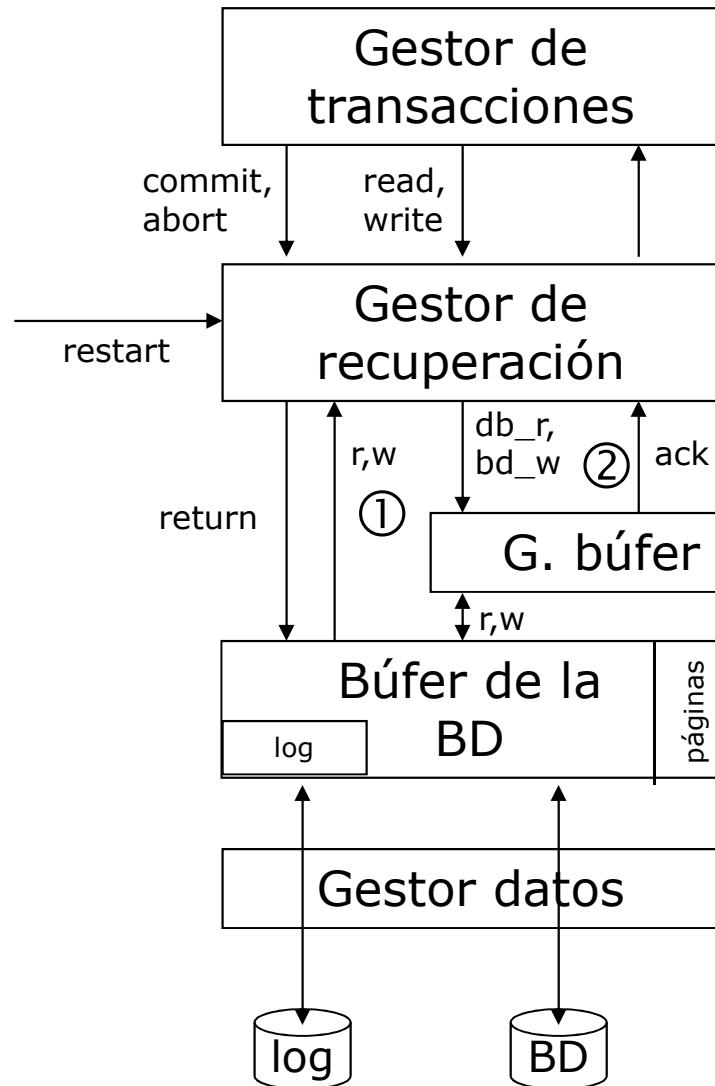
■ Fallo de una transacción; pérdida datos



1. transacciones terminadas (committed) pero no persistentes (en la bd) \Rightarrow REDO
2. transacciones no terminadas (quizás algunas escrituras hechas en la bd) \Rightarrow UNDO

■ Copia que permita restaurar un estado consistente

...



- ① lecturas/escrituras en el búfer si el objeto está disponible
—valor de retorno—
 - ② si el objeto no está disponible, pide al G.búfer que lo haga disponible
—ack—
- Log: información sobre qué acciones de qué transacciones ya han afectado al estado actual de la bd
<transacción, objeto, acción>
 - imagen_antes
 - imagen_después

Optimización

```
select distinct P.nombre
from prof P, calif C
where P.dni=C.dni_prof
      and C.nia_alumno='----';
```

#(prof) = 100
 #(calif) = 10.000

1	sin optimización		10^6
i.	prof x calif	leer 10^4t , 100 veces escribir 10^6 (no m.p.)	10^6 10^6
ii.	restricción	leer $106t$, 50 m.p.	10^6
iii.	proyección	-	-
2	optimización A: cambiar orden de operaciones		10^4
i.	restric. calif	leer 10^4t , 50 a m.p.	10^4
ii.	x prof	leer 100	10^2
iii.	proyección	-	-
3	optimización B: índice sobre nia_alumno		10^2
i.	restric. calif	leer índice, 50 a m.p.	10^2
ii.	x prof	leer 100	10^2
iii.	proyección	-	-

Diseño físico

- Representación física
 - Atributo \leftrightarrow *campo*
 - Tupla \leftrightarrow *registro*
 - Colección de tuplas \leftrightarrow *bloque*
 - Relación \leftrightarrow *fichero*
- Estructuras de datos
 - Secuenciales
 - Índices
 - Árboles
 - Dispersión

Secuenciales

» n registros almacenados uno_detrás_de_otro
en bloques con b registros

- **Búsqueda**

Todo el fichero: $O(n/b)$

- **Inserción**

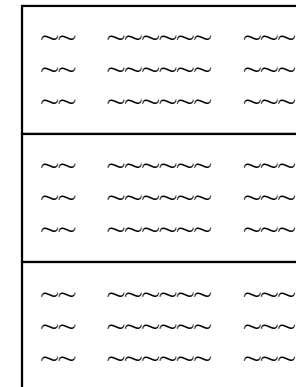
En el último bloque si hay espacio libre

- **Borrado**

Deja un espacio libre en el bloque

⇒ reorganización (*recolección de basura*)

- **Modificación**



Indexados

- **Búsqueda**

Denso: valores únicos: binaria + localización en el bloque

no únicos: binaria + localización del 1º + secuencial

No denso: binaria + (si no localizado) secuencial

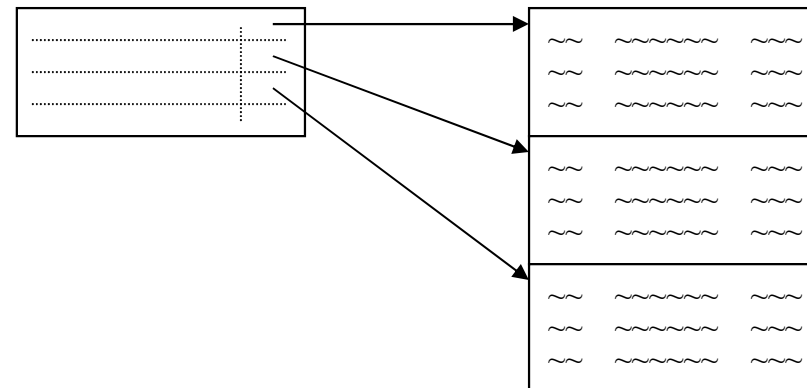
- **Inserción**

Similar a ficheros ordenados, salvo el índice si no denso

- **Borrado**

Similar a ficheros ordenados, salvo el índice si no denso

- **Modificación**



Árboles

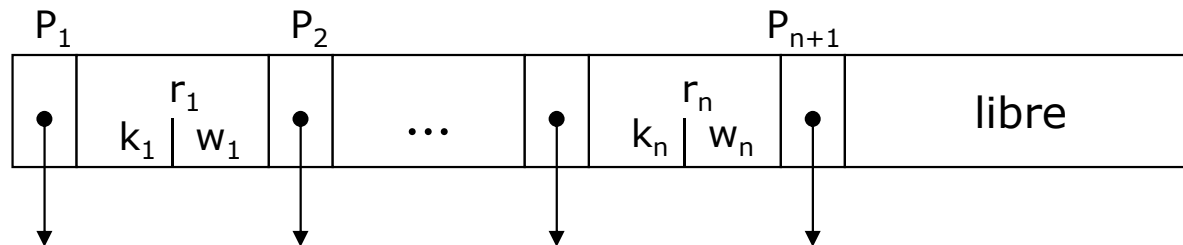
■ Árboles B

» árbol dirigido balanceado

cada nodo (bloque) contiene entre d y $2d$ registros

cada bloque se ordena por valores clave

cada registro tiene una parte clave y una no clave



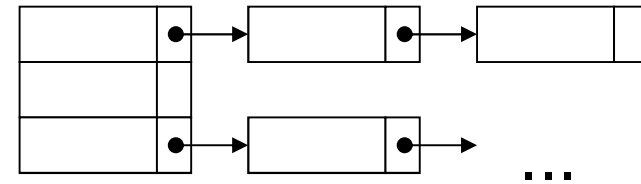
todos los valores contenidos en el nodo referenciado por P_i son menores que k_i (resp. mayores para P_{i+1})

- Búsqueda
- Inserción / Borrado

Dispersión

» no se recorre un índice, sino que la dirección (bloque) es calculada directamente a partir de los valores de la clave

- **Espacio de dispersión**
n direcciones organizadas en bloques



- **Función de dispersión**
método del resto del valor respecto al espacio

- **Colisiones**
cuando coincide la dirección (bloque) para dos valores distintos

a) Overflow (disp. abierta)

b) Redispersión (disp. cerrada)

- **Redistribución (disp. dinámica)**

Índices secundarios

1. Columnas de restricción, no de selección
2. Columnas de join
3. La más restrictiva
Devuelve pocos y éstos se tratan más fácilmente
 - Depende de:
 - veces que se ejecuta la consulta
 - tiempo aceptable de respuesta
 - importancia de la consulta
 - Coste de mantenimiento