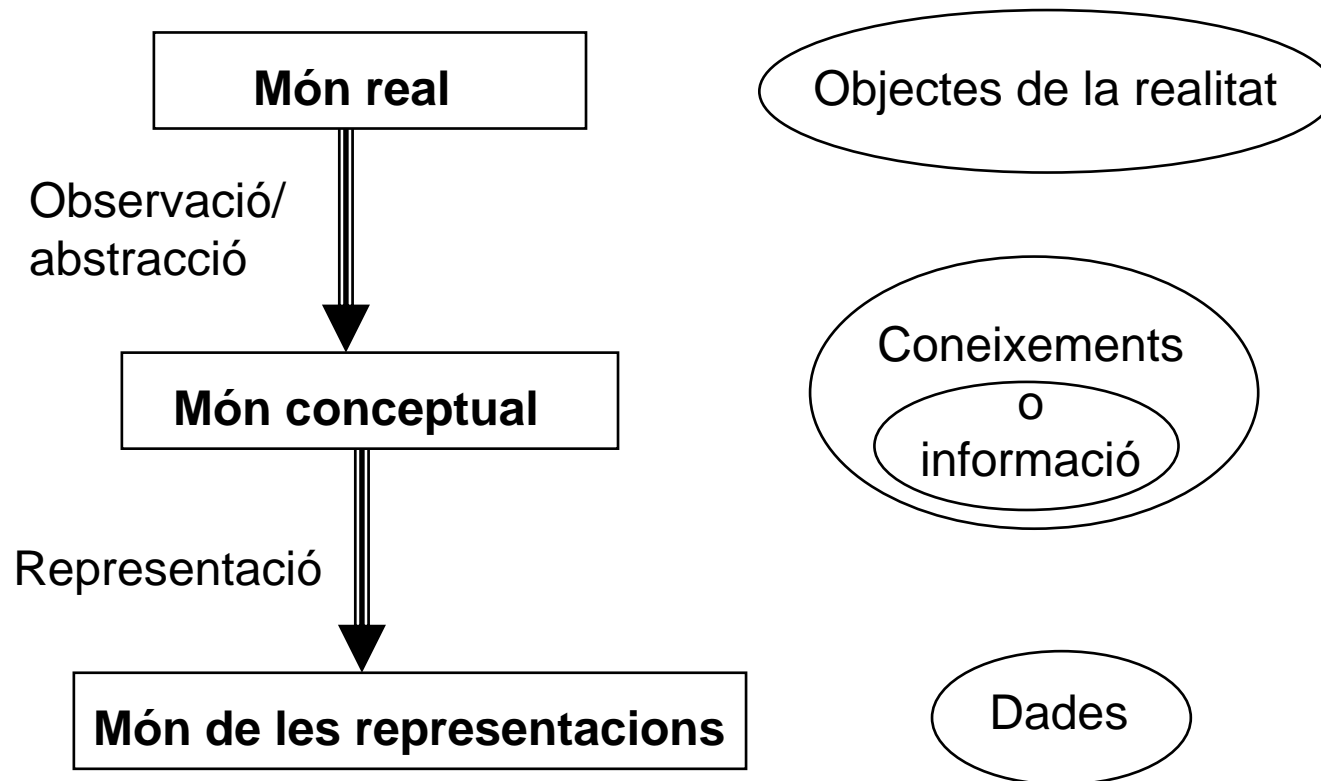

Introducció

- Els tres mons
 - Els tres mons i el disseny de BD
 - Concepte de BD
 - Els SGBD
 - Objectius dels SGBD
 - Arquitectura d'un SGBD
 - Models de BD
 - Usuaris de BD
 - Llenguatges de BD
 - Administració de BD
-

Els tres mons

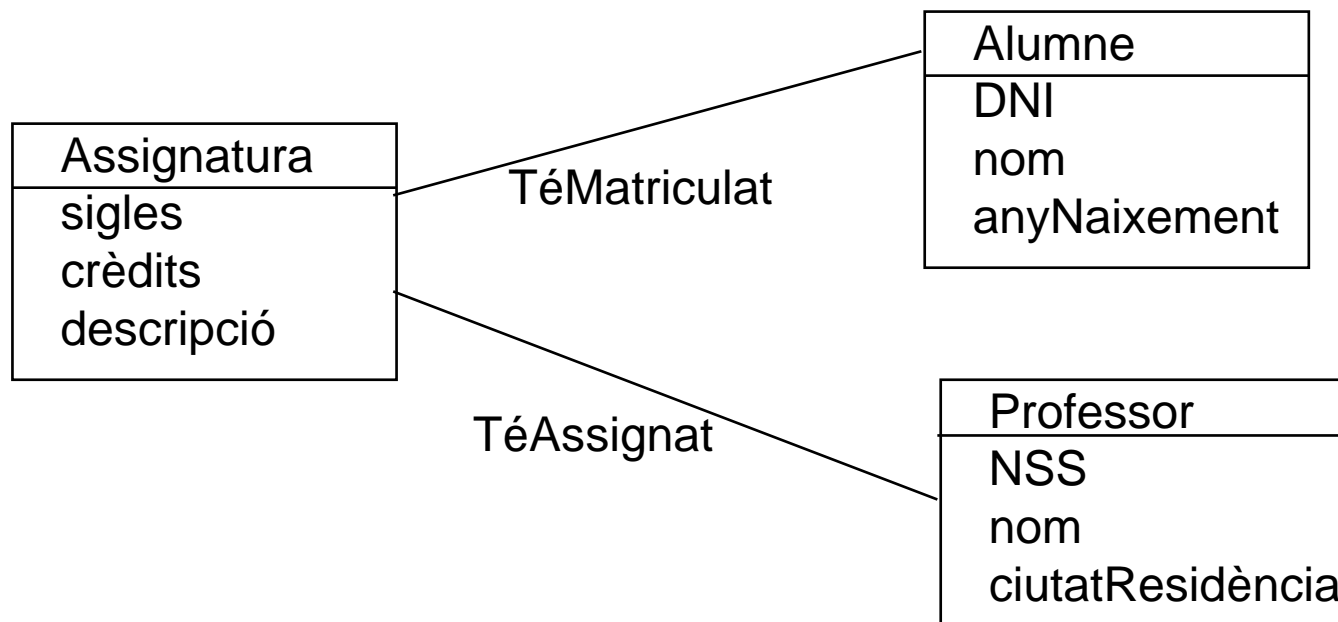


Els tres mons

- **El món real**, la part de realitat que ens interessa, és el que percebem amb els nostres sentits i és compost per objectes concrets, físics o no.
 - El conjunt de **coneixements o informacions** obtinguts observant la part de realitat que ens interessa forma part del que anomenem **món conceptual**.
 - Per poder treballar amb els coneixements i comunicar-los, necessitem representar-los en el **món de les representacions**.
 - Les **dades** són les representacions informàtiques dels coneixements que tenim dels objectes del món real.
-

El món conceptual

- **Classe d'objectes:** Descriu un conjunt d'objectes que comparteixen propietats, associacions amb altres objectes i tenen una semàntica comuna.
- **Atribut:** Propietat compartida pels objectes d'una classe.
- **Associació:** Interrelació entre classes d'objectes.
- Exemple:



El món de les representacions

- Precursors de les BD: **els fitxers**.

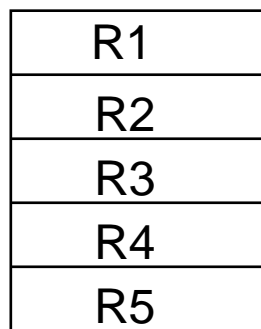
- **Exemple:** fitxer d'alumnes.

DNI	Nom	AnyNaix
12.121.212	Maria Pi	1984
34.343.434	Cristina Pujol	1982
56.565.656	Jordi Martínez	1986
78.787.878	Marc Font	1984

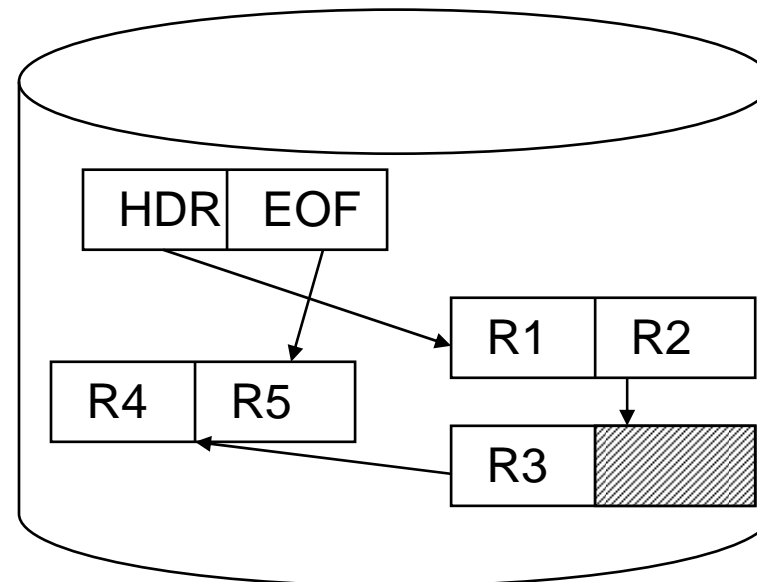
- Un fitxer permet representar una classe d'objectes i els seus atributs univaluats:
 - Cada **camp** representa un atribut univaluat de la classe d'objectes.
 - Cada **registre** correspon a un objecte concret de la classe.
- **Bases de dades:** Faciliten la representació de diverses classes d'objectes, els seus atributs i també les **associacions** entre classes.

El món de les representacions: enregistrament físic

- Les **dades** s'han d'emmagatzemar de manera persistent, no volàtil.
- Els suports de **memòria externa** proporcionen persistència: discos magnètics, cintes, CD-ROM, DVD, etc.
- Els **fitxers** i **bases de dades** s'han d'enregistrar físicament en suports de memòria externa.

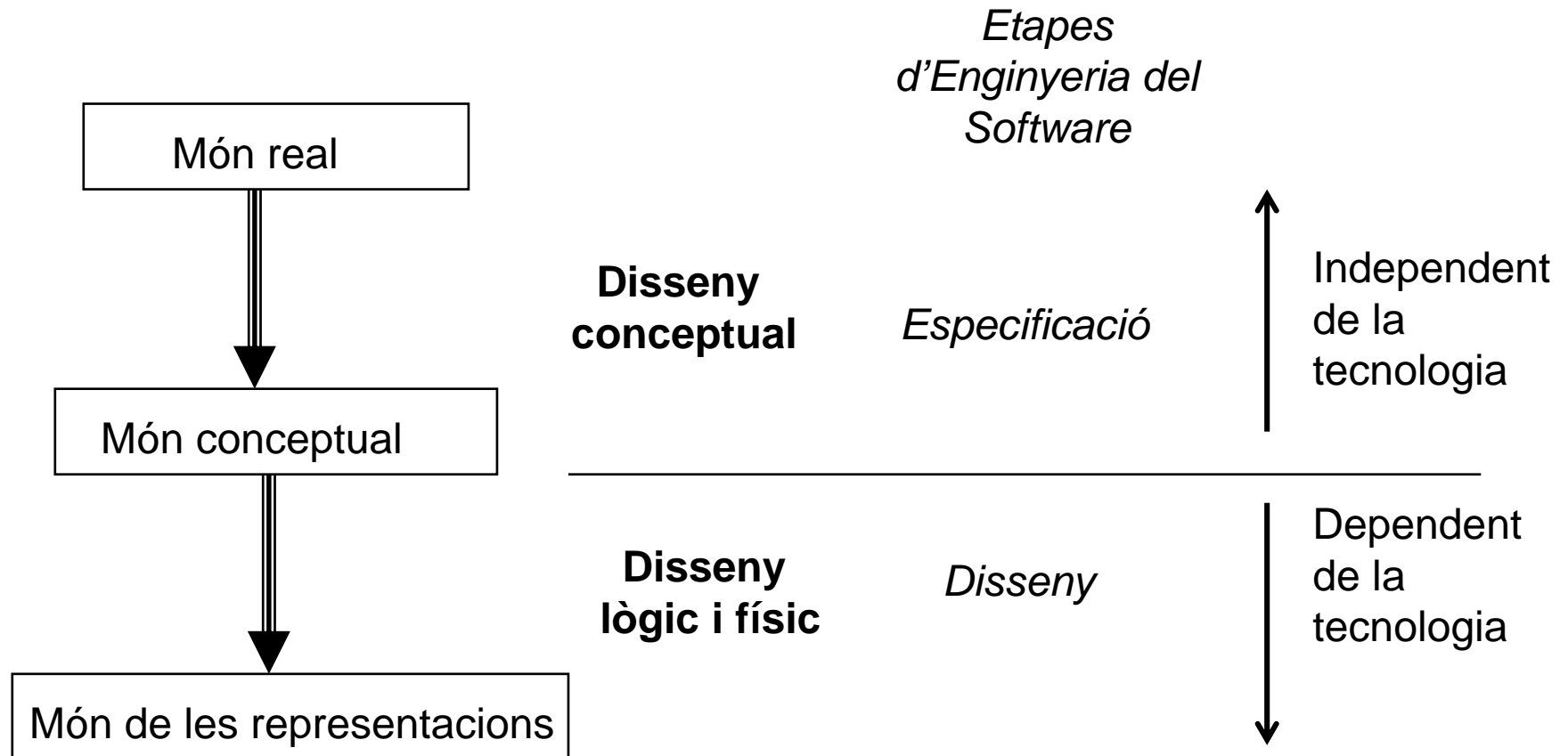


Fitxer



Enregistrament físic

Els tres mons i el disseny de BD



Concepte de BD (1)

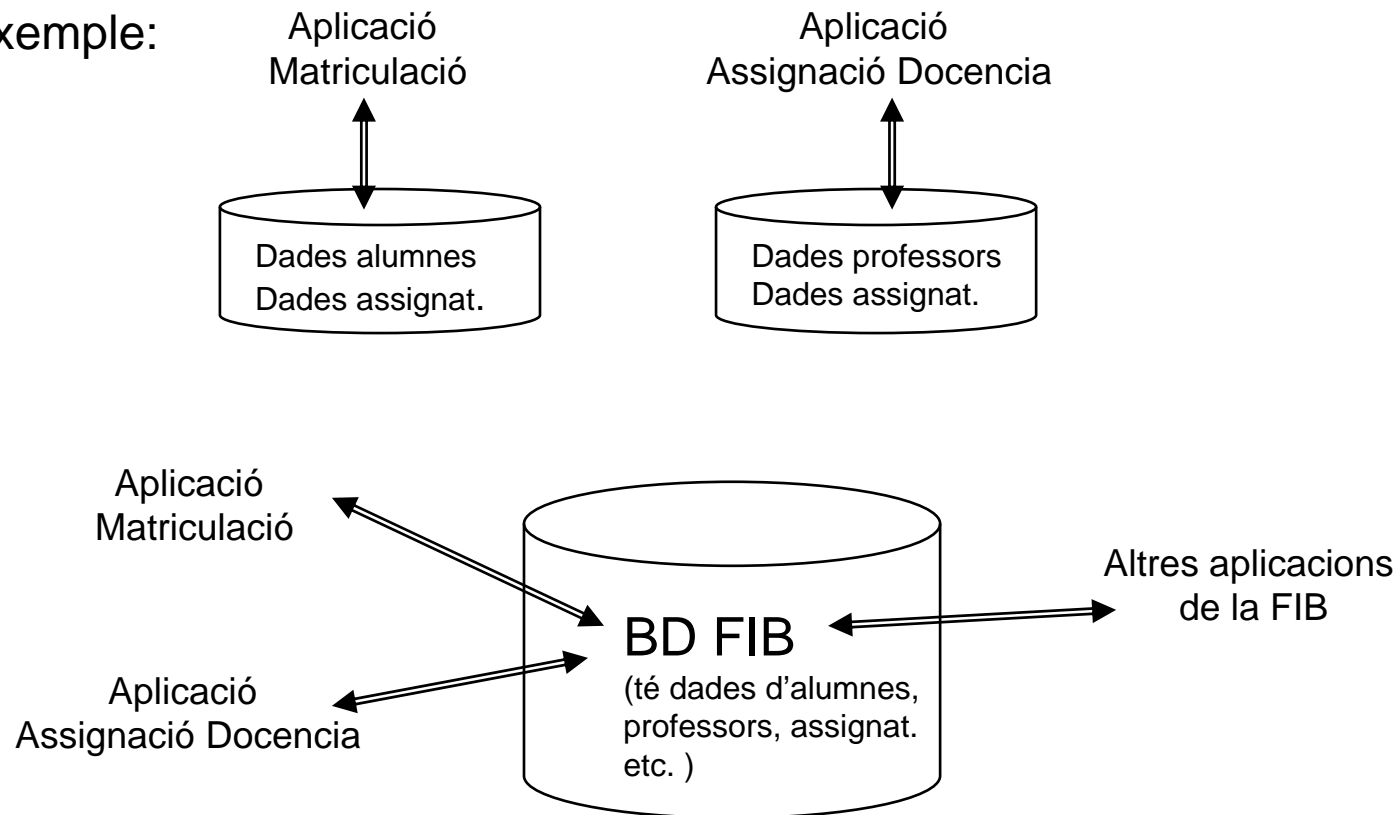
- **Representació de classes d'objectes i de les seves associacions**
 - Les bases de dades proporcionen **estructures de dades** que faciliten aquesta representació.
 - Podem **classificar** les bases de dades segons l'estructura de dades utilitzada per representar classes i associacions.
 - **BD jeràrquiques**: col.lecció d'arbres.
 - **BD en xarxa**: xarxa.
 - **BD relacional**: relacions.
-

Concepte de BD (2)

- Integració de les dades

- Les bases de dades faciliten la **integració de dades** de diverses aplicacions (permet evitar la repetició de dades)

- Exemple:



Concepte de BD (3)

- **Compartició**

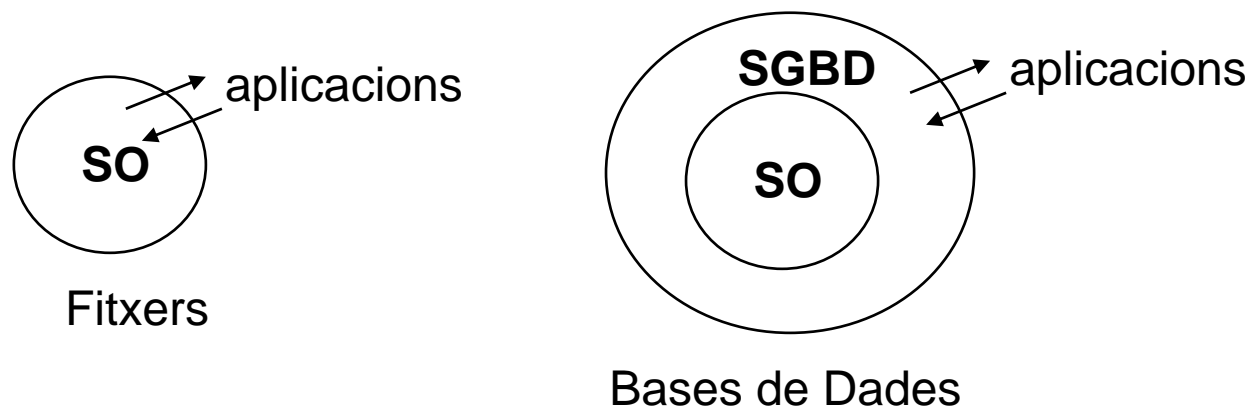
- Les bases de dades faciliten l'**accés compartit** a les dades per molts usuaris de tipus diversos.

- **Persistència**

- Les dades de les bases de dades s'emmagatzemen a memòria externa i són **persistents** (no volàtils).
-

Els SGBD

- **SGBD**: Sistema de Gestió de Bases de Dades
En anglès: Data Base Management System (DBMS)
- Un **SGBD** és un software dissenyat per facilitar diverses tasques de gestió de dades als programes d'aplicació
 - Representació de classes i associacions
 - Integració de dades
 - Compartició
 - etc

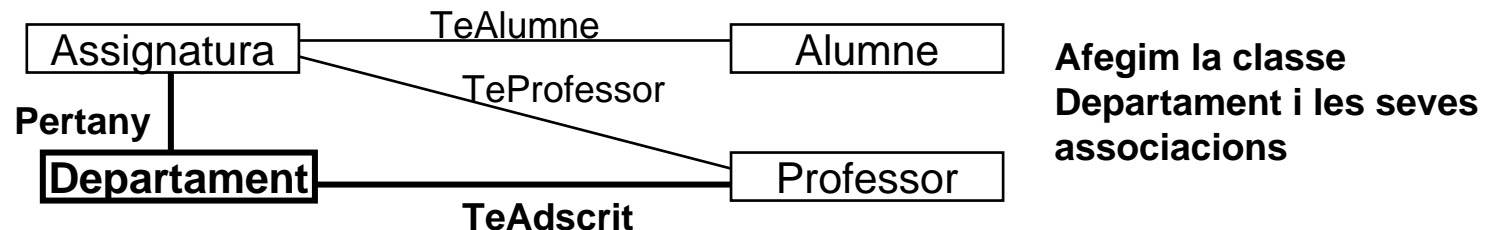


Objectius d'un SGBD

- Flexibilitat i independència
 - Facilitar el control de la redundància de les dades
 - Integritat de les dades
 - Concurrència d'usuaris
 - Recuperació
 - Seguretat
 - Consultes no predefinides i complexes
 - Distribució de les dades
-

Flexibilitat i independència

- **Flexibilitat als canvis:** Ha de ser fàcil fer canvis.
- **Canvis possibles:**
 - Canvis conceptuals: afegir/ suprimir atributs, classes d'objectes o associacions.



- Canvis a l'enregistrament físic de les dades.
- **Independència física de les dades.**
- **Independència lògica de les dades.**
- L'**arquitectura ANSI/SPARC** està orientada a l'obtenció de la independència física i lògica de les dades.

Facilitar el control de la redundància de les dades

- Les **redundàncies** de les dades venen motivades per:
 - Dades **repetides**.
 - Dades **calculables**.
 - **Problemes de les redundàncies:**
 - Ocupació innecessària d'espai (encara que habitualment no és un problema greu).
 - Cost d'actualització.
 - Possibilitat d'inconsistència de les dades.
 - **Redundàncies controlades:**
 - Pot interessar tenir una dada calculable redundant si es vol **millorar el rendiment de la seva consulta**.
 - **Caldrà refer el càlcul** de la dada redundant cada vegada que hi ha una modificació a les dades que serveixen per fer el càlcul. És aconsellable que el mateix SGBD refaci el càlcul o que controli que el càlcul es refaci adequadament per evitar inconsistències.
 - **Cal documentar** les redundàncies controlades.
-

Integritat de les dades

- Les dades d'una base de dades han de complir:
 - **Regles d'integritat del model:** Condicions que les dades han de complir i que són inherents al model de dades que utilitza l'SGBD.
 - **Restriccions d'integritat dels usuaris:** Condicions que les dades han de complir i que venen motivades per requisits dels usuaris de la BD.
 - **Tipus de restriccions d'integritat dels usuaris:**
 - Restriccions d'integritat **estàtiques:** La seva avaluació depèn d'un sol estat de la BD.
 - Restriccions d'integritat **dinàmiques:** La seva avaluació depèn de diversos estats de la BD.
 - L'SGBD ha de **proporcionar mecanismes** per:
 - La **definició** de les restriccions d'integritat dels usuaris.
 - El **manteniment** de les regles d'integritat del model i de les restriccions d'integritat dels usuaris.
-

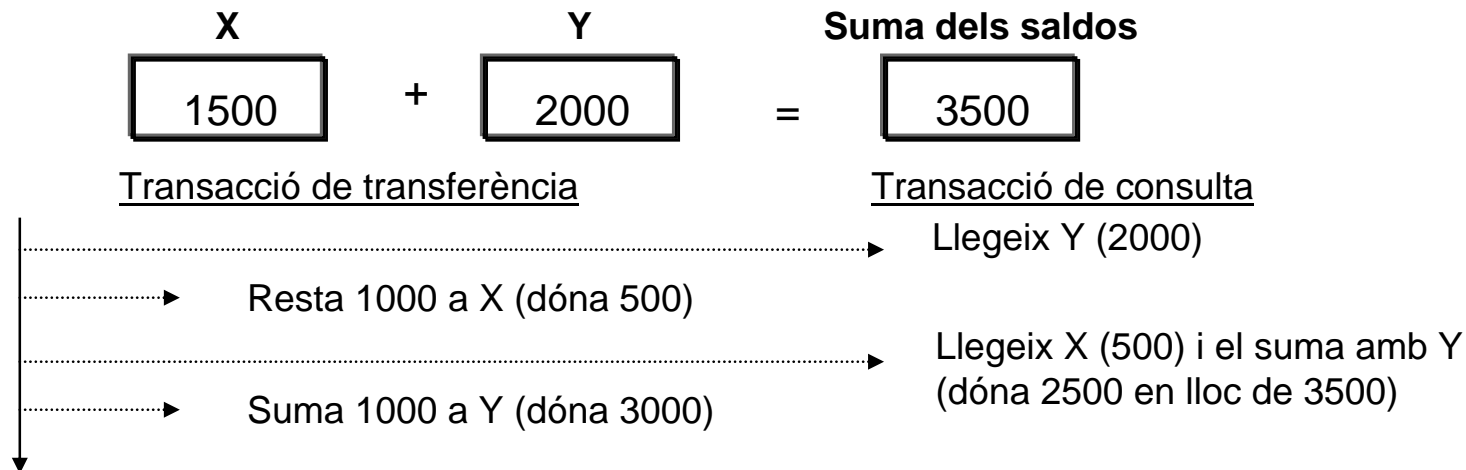
Concurrencia d'usuaris

- Un objectiu fonamental dels SGBD és permetre que diversos usuaris puguin **accedir concurrentment** a la mateixa BD.
 - Per a tractar els accessos concurrents els SGBD fan servir el concepte de **transacció**:
 - **Transacció**: conjunt d'operacions simples que s'executen com una unitat.
 - Ex: Transferència de diners d'un compte X a un compte Y
 - Càrrec a X de la quantitat q de diners
 - Abonament a Y de la mateixa quantitat q de diners
 - Una transacció pot acabar de dues maneres:
 - COMMIT: Transacció confirmada.
 - ROLLBACK: L'SGBD ha de desfer la transacció.
 - Poden donar-se **problemes de concurrència** entre transaccions.
-

Problemes de concurrència

- **Exemple** de problema de concurrència entre transaccions:

Transferència de 1000 euros del compte X al compte Y concurrentment amb una consulta de la suma dels saldos d'X i Y.



- Per aconseguir que les transaccions s'executin com si estiguéssin aïllades, els SGBD usen diverses tècniques, la més coneguda de les quals és el **bloqueig** (lock)
- El bloqueig d'unes dades per una transacció consisteix a posar limitacions als accessos que les altres transaccions podran fer a aquestes dades.

Recuperació

- L'SGBD ha de garantir:
 - **Integritat de les dades.**
 - Per les **transaccions**:
 - No s'han de perdre els efectes d'una transacció confirmada.
 - Els canvis produïts per una transacció fallida s'han de desfer.
 - **No pèrdua de dades.**
 - **Causes** externes a les transaccions que poden interferir en els objectius anteriors:
 - **Fallades del sistema.**
 - **Fallades del sistema d'emmagatzematge.**
 - Per fer front a aquests casos, és necessari que l'SGBD pugui fer una **recuperació** de les dades.
 - Els mecanismes de recuperació dels SGBD es basen en:
 - Obtenció de **còpies de seguretat** periòdiques (backup).
 - Manteniment continu d'un **diari** (log) on l'SGBD va enregistrant totes les operacions d'actualització que fan les transaccions.
-

Seguretat

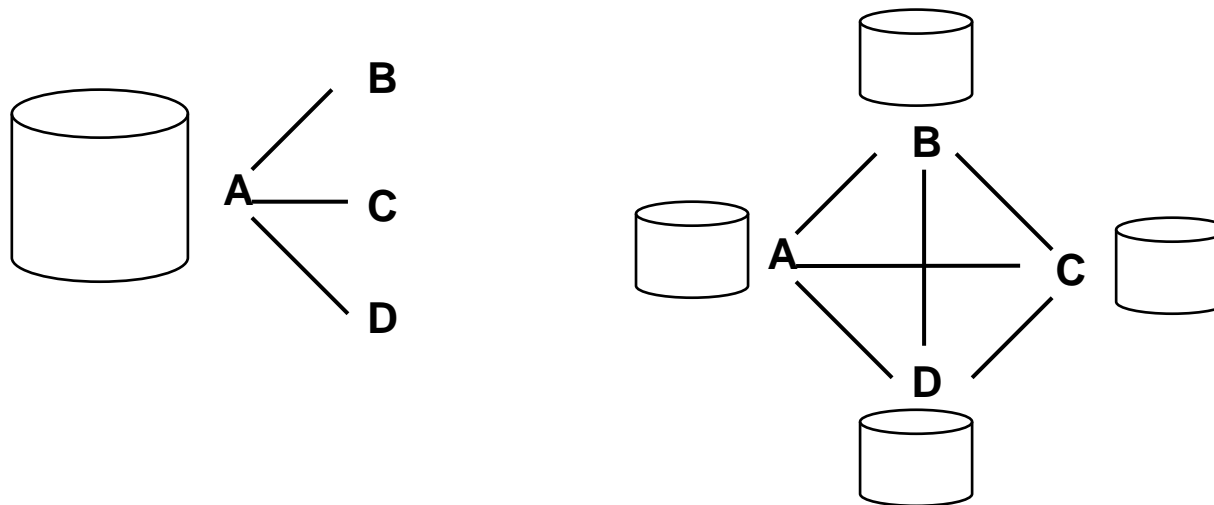
- El terme **seguretat** se sol utilitzar per fer referència als temes relatius a la confidencialitat, les autoritzacions, els drets d'accés, etc.
- Un objectiu dels SGBD és el de garantir que les dades emmagatzemades només poden ser accedides per les persones autoritzades i de la forma autoritzada.
- Per assolir aquest objectiu cal:
 - **Identificació i autenticació dels usuaris**: mitjançant una paraula clau, targeta, tècniques biomètriques, ...
 - **Possibilitat de definir autoritzacions o drets d'accés**: Una autorització dóna dret a un usuari o grup d'usuaris a accedir a unes determinades dades de la BD per fer unes determinades operacions.
 - La definició de les **dades** que s'autoritzen s'ha de poder fer amb diferents nivells de granularitat: al nivell global de tota la BD, al nivell de classe, al nivell d'atribut.
 - S'han de poder autoritzar **operacions** concretes sobre les dades: consulta, inserció, esborrat, modificació.

Consultes no predefinides i complexes

- Els usuaris han de poder fer consultes de qualsevol tipus i **complexitat** directament a l'SGBD.
 - Aquestes consultes poden **no** ser **preestablertes** i l'SGBD ha de respondre immediatament, és a dir, sense que s'hagi d'escriure, compilar i executar un programa específic per a cada consulta.
 - L'usuari ha de poder formular la consulta en un **llenguatge senzill**:
 - Que es quedi a nivell lògic i no requereixi coneixements de la representació física per a la formulació de la consulta.
 - El sistema l'ha d'interpretar directament.
 - El llenguatge estàndard relacional, **SQL**, facilita l'assoliment d'aquest objectiu.
 - Els SGBD inclouen la funció del **processament de consultes** que fa la transformació de la consulta a passos de baix nivell que la implementen i alhora l'optimitzen.
-

Distribució de dades

- Segons hi hagi o no distribució de les dades, les BD poden ser **centralitzades** o **distribuïdes**.
- **BD centralitzades**: totes les dades són a una mateixa localització.
- **BD distribuïdes**: les dades estan repartides a localitzacions diferents.



- En les BD distribuïdes les dades locals estan més disponibles i es poden processar més eficientment.
- En les BD distribuïdes augmenta el cost de la gestió de les transaccions, manteniment d'integritat, recuperacions, etc.

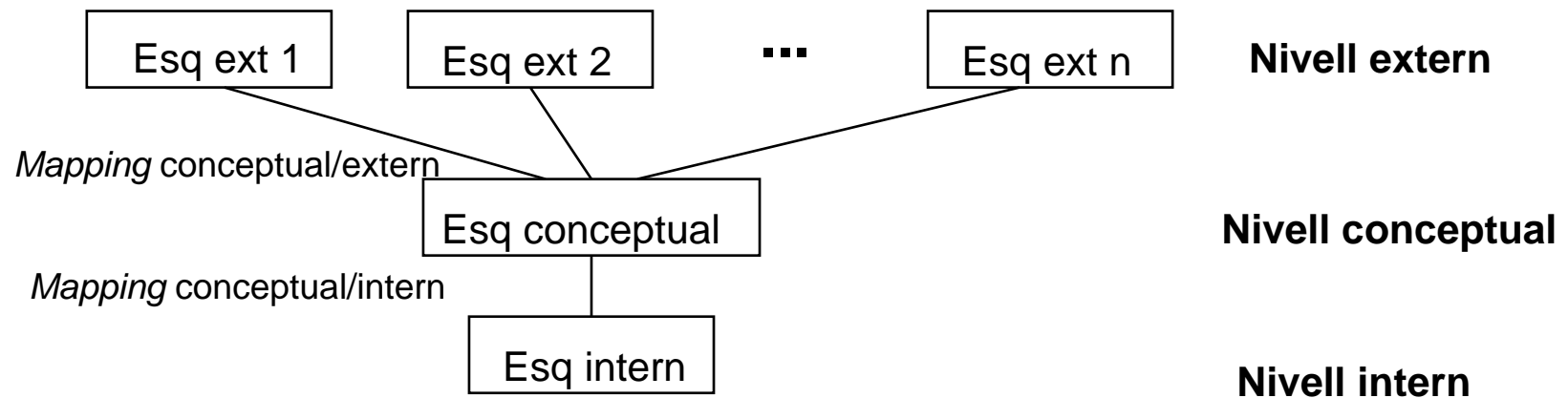
Arquitectura d'un SGBD

- Arquitectura ANSI/SPARC
 - Nivells i esquemes
 - Independència de les dades
 - Flux de dades i control
-

Arquitectura ANSI/SPARC

- El comitè ANSI/SPARC va recomanar (1975-82) que l'arquitectura dels SGBD preveïés tres nivells d'esquemes que corresponen a tres nivells d'abstracció diferents.

Nivells i esquemes

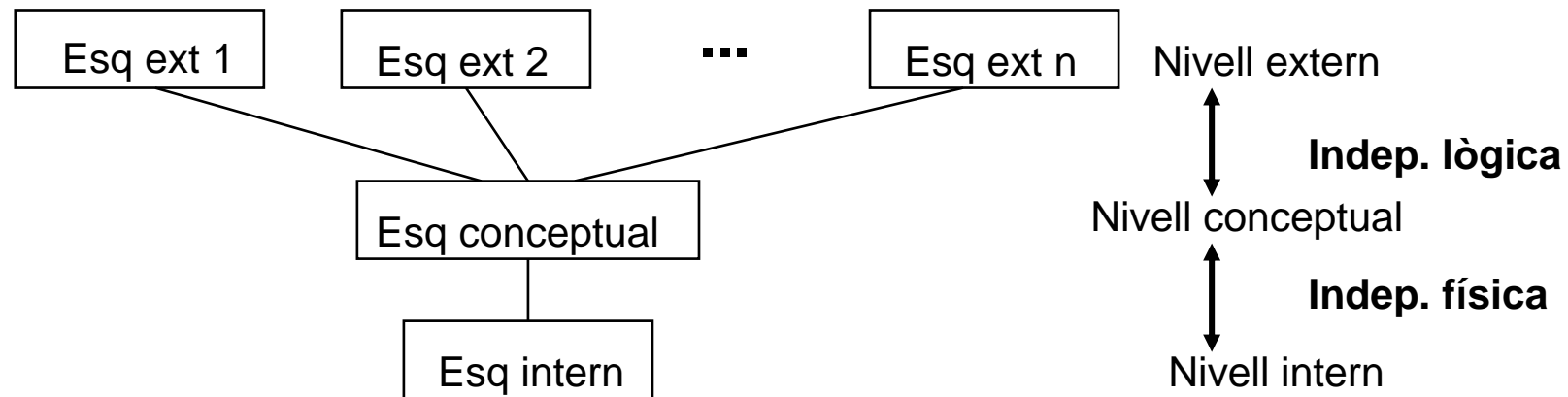


- **Nivell intern:** correspon a l'emmagatzematge de les dades.
- **Nivell conceptual:** correspon a l'estructura de la base de dades per a tota la comunitat d'usuaris.
- **Nivell extern:** correspon a les diferents visions dels diferents tipus d'usuaris o aplicacions que utilitzen la BD. Hi haurà un esquema extern per cadascun.
- **Esquema:** descripció d'elements que pertanyen a un determinat nivell.

Contingut dels esquemes

- **Esquema conceptual:** Conté la descripció de:
 - Classes d'objectes
 - Atributs
 - Associacions
 - Restriccions d'integritaten termes del model de dades de l'SGBD.
 - **Esquema intern:** Conté la descripció de:
 - Organització dels fitxers que emmagatzemen la base de dades
 - Estructures de dades auxiliars que agilitzen l'accés (mètodes d'accés).
 - **Esquema extern:** Conté:
 - La cita d'un subconjunt de dades de l'esquema conceptual
 - La descripció de dades calculables a partir de l'esquema conceptual que interessen a un determinat tipus d'usuari o aplicació.
-

Independència de les dades

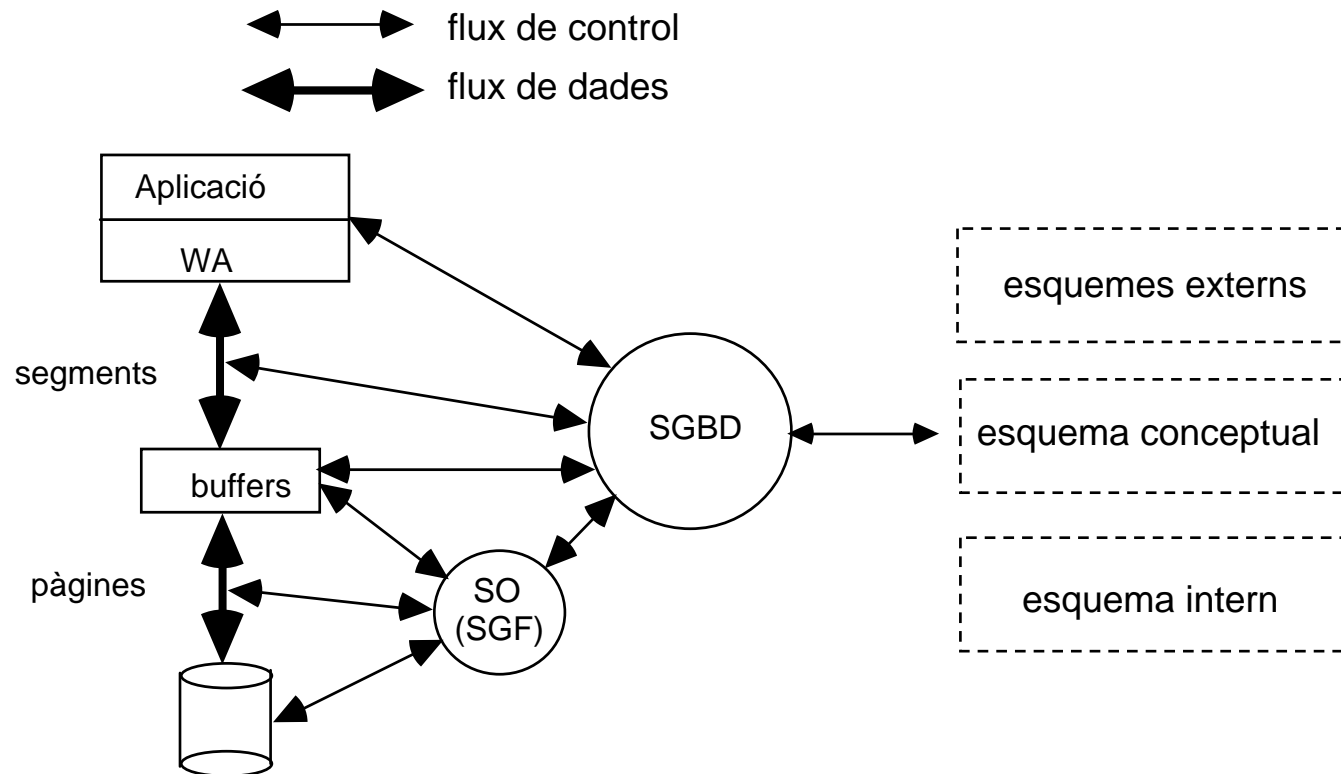


- L'arquitectura de tres nivells ANSI/SPARC ens permet aconseguir:
 - Independència física de les dades.
 - Independència lògica de les dades.

Independència de les dades: lògica i física

- **Independència física de les dades:**
 - **Canvis a l'esquema intern:** No afecten a l'esquema conceptual ni als esquemes externs
 - Canvis possibles a l'esquema intern: canvis als mètodes d'accés, mida de les pàgines, etc.
 - Els programes d'aplicació i els usuaris directes no es veuran afectats per aquests canvis. En canvi, sí que caldrà refer les correspondències entre l'esquema conceptual i l'intern i, en la majoria de casos, caldrà refer la BD física.
 - **Independència lògica de les dades:**
 - **Canvis a l'esquema conceptual:** Un canvi d'aquest tipus no afectarà als esquemes externs que no facin referència a les classes, atributs o associacions modificats.
 - **Canvis als esquemes externs:** Un canvi en un esquema extern no afectarà als altres esquemes externs, ni l'esquema conceptual, ni l'esquema intern.
-

Flux de dades i control



Segment: Dades que es transfereixen entre els buffers i l'aplicació

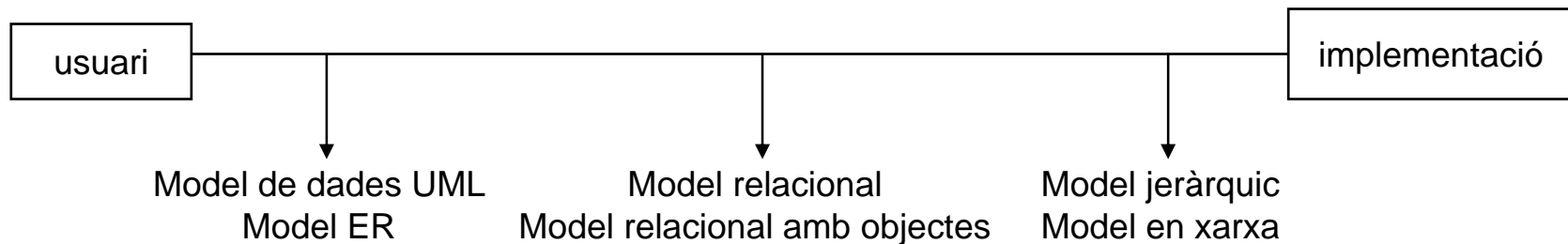
Pàgina: Dades que es transfereixen entre els buffers i la memòria externa

Models de BD

- El conjunt de components o eines conceptuais que proporciona un SGBD per estructurar i manipular les dades rep el nom de **model de BD**. Alguns exemples de models de BD són: jeràrquic, xarxa, relacional, relacional amb objectes.
 - Tot model de BD ens proporciona tres tipus d'eines:
 - **Estructures de dades** amb les quals es pot construir la BD.
 - **Operacions** per manipular i consultar les dades.
 - **Regles d'integritat** que l'SGBD haurà de fer complir a les dades.
 - Els **models de dades** permeten també estructurar les dades però no són necessàriament utilitzats pels SGBD del mercat. Aquests models s'usen en el procés de disseny conceptual de la BD. Exemples: el model de dades de la notació UML, el model ER, etc.
-

Models de dades i de BD en el procés de disseny

- Situació dels models de dades i de BD segons el seu **nivell d'abstracció**: més proper a la implementació (representació física) o més proper a la visió dels usuaris:



- **Procés de disseny de BD**

- **Disseny conceptual:** S'usa un model de dades amb alt nivell d'abstracció (encara que no tingui suport directe en els SGBD del mercat) com ara el model de dades d'UML o el model ER.
- **Disseny lògic:** Consisteix en la traducció del model de dades anterior al model de BD de l'SGBD escollit per a la implementació. Per exemple: traducció del model d'UML al model relacional.
- **Disseny físic:** Consisteix en la transformació de l'estructura obtinguda en el disseny lògic per aconseguir una major eficiència i en la compleció dels aspectes de representació física que depenen de l'SGBD escollit.

Usuaris de BD

- **Usuaris informàtics:** dissenyadors, programadors d'aplicacions, administradors de la BD.
 - **Usuaris no informàtics:**
 - **Usuaris paramètrics:**
 - Usen programes d'aplicacions prèviament dissenyats, implementats i provats pels usuaris informàtics.
 - No han de conèixer la BD ni cap llenguatge de BD.
 - **Usuaris finals:**
 - Necessiten realitzar ocasionalment consultes (no previstes) a la BD i la informació a accedir pot ser diferent en cada ocasió.
 - Usen un llenguatge de consulta d'alt nivell per programar-les ells mateixos.
 - Han de conèixer part de l'estructura de la BD i el llenguatge de consulta.
-

Llenguatges de BD: primera classificació

- **DDL** (*Data Definition Language*)
 - Serveix per definir l'estructura de les dades. Concretament, permet definir l'esquema conceptual, l'esquema intern i els esquemes externs.
 - Destinat a usuaris informàtics.
 - **DML** (*Data Manipulation Language*)
 - Serveix per fer operacions de consulta i actualització sobre la base de dades.
 - Destinat a usuaris informàtics i, en alguns casos, a usuaris finals.
 - El llenguatge **SQL** inclou tant sentències de tipus DDL com de tipus DML i, a més, inclou sentències de control de l'entorn (com ara les de delimitació de transaccions).
-

Llenguatges de BD: segona classificació (pels DML)

- **Declaratius**

- Llenguatges amb els quals s'especifica **què** es vol fer sense explicar com s'ha de fer.
- Ex: SQL
- Destinats a usuaris informàtics i, en alguns casos, a usuaris finals.

- **Procedimentals**

- Llenguatges amb els quals cal detallar **com** s'han de fer les operacions.
 - Ex: àlgebra relacional
 - Destinats a usuaris informàtics.
-

Llenguatges de BD: tercera classificació

- **Autosuficients**

- És el cas de llenguatges de BD que no necessiten de cap altre llenguatge per ser utilitzats.
- Ex: SQL interactiu.
- Destinats a usuaris informàtics i, de vegades, a usuaris finals.

- **No autosuficients**

- És el cas de llenguatges de BD que s'utilitzen conjuntament amb un llenguatge de programació.
 - Destinats a usuaris informàtics.
 - Poden ser:
 - **Hostatjats**
 - **Amb crides a rutines predefinides**
-

Administració de BD

- L'**administració de la BD** agrupa les funcions de gestió destinades a assegurar que l'exploració de la BD és correcta.
 - L'**administrador de la BD** (ABD) és el professional que es responsabilitza de les esmentades funcions.
 - Algunes **tasques de l'ABD** són les següents:
 - 1) Manteniment, administració i control dels esquemes. Comunicació dels canvis als usuaris.
 - 2) Assegurar la màxima disponibilitat de les dades (còpies de seguretat, dietaris, reconstruccions,...).
 - 3) Resolució d'emergències.
 - 4) Vigilància de la integritat i de la qualitat de les dades.
 - 5) Disseny físic, estratègia de camins d'accés i reestructuracions.
 - 6) Control del rendiment i decisions relatives a les modificacions en els esquemes i/o paràmetres de l'SGBD i de l'SO, per a millorar-lo.
 - 7) Normativa i assessorament pels programadors i usuaris finals sobre la utilització de la BD.
 - 8) Control i administració de la seguretat: autoritzacions, restriccions, etc.
 - Els **SGBD** proporcionen **eines** molt útils per l'ABD: diccionari de dades, monitors de rendiment, monitors de la seguretat, reorganitzadors, gestors de les còpies de seguretat, etc.
-