

---

# Fonaments de bases de dades

---

PID\_00255958

Ignasi Lorente Puchades

---

Temps mínim de dedicació recomanat: 2 hores

---



**Ignasi Lorente Puchades**

# Índex

<b>Introducció</b> .....	5
<b>Objectius</b> .....	6
<b>1. Les dades i les bases de dades</b> .....	7
1.1. Els models de dades .....	7
1.1.1. Model jeràrquic .....	8
1.1.2. Model en xarxa .....	8
1.1.3. Model relacional .....	9
1.1.4. Model documental .....	11
1.2. Sistemes de gestió de bases de dades .....	11
1.2.1. Funcions principals del sistema de gestió de bases de dades .....	12
<b>2. Disseny i desenvolupament d'una base de dades</b> .....	14
2.1. Etapa de recollida i anàlisi de requisits .....	14
2.2. Etapa de disseny conceptual .....	14
2.3. Etapa de disseny lògic .....	15
2.4. Etapa de disseny físic .....	15
2.5. Etapa d'implementació i optimització .....	15
<b>Resum</b> .....	16
<b>Glossari</b> .....	17



## Introducció

Una base de dades és una col·lecció d'elements emmagatzemats de forma organitzada amb el propòsit de facilitar la seva cerca, recuperació, classificació, interrelació, organització, etc. a un conjunt d'usuaris i aplicacions.

Aquests elements que s'emmagatzemaran s'anomenen dades i poden tenir característiques molt diferents, de manera que haurem de ser capaços de crear un sistema que permeti gestionar-les malgrat la seva heterogeneïtat.

Per a aconseguir aquest propòsit, aquests elements s'emmagatzemaran en un sistema informàtic basat en arxius que afavorirà la relació entre les dades, que estaran representades de forma única, i permetrà l'accés a la informació de forma simultània a diferents usuaris mentre s'assegura la seva integritat.

Si entenem, doncs, una base de dades com la representació informàtica d'un tot (un conjunt referencial), en què els elements que en formen part són la representació dels requeriments d'una necessitat que se'ns presenta, un sistema gestor de base de dades (SGBD) serà el conjunt de programes que permetrà la seva creació, manipulació i administració.

S'ha de considerar que les bases de dades treballen sobre un univers finit, que no representa «totes les coses» i que està delimitat per l'ús que es vulgui donar de les dades emmagatzemades. Per exemple, si volem mantenir una base de dades d'una llibreria, identificarem aquesta base de dades com un conjunt d'elements com ara llibres, autors, temes, idiomes, etc. i les relacions que es donen entre aquests elements.

Així doncs, crear una base de dades consistirà en comprendre una realitat i descriure-la de manera que els elements rellevants d'aquest context i les relacions entre aquests siguin representables. A partir d'aquest disseny, podrem crear i informar la base de dades per després donar resposta a les preguntes que es formulin.

A grans trets, els tres eixos principals d'un sistema de base de dades són:

- les dades, que representaran la informació,
- les operacions, que permetran als usuaris treballar amb la base de dades, i
- el sistema informàtic, que permetrà el seu emmagatzematge.

Nosaltres, en aquesta assignatura, treballarem amb bases de dades relacionals i, per tant, amb sistemes de gestió de bases de dades relacionals.

### Sistema gestor de bases de dades

Un sistema gestor de bases de dades (SGBD; en anglès *data-base management system*) és un tipus de programari que serveix d'interfície entre la base de dades, l'usuari i les aplicacions que l'utilitzen.

## Objectius

En els materials didàctics d'aquesta unitat trobareu les eines indispensables per a assolir els objectius següents:

- 1.** Conèixer els conceptes fonamentals de les bases de dades: què són les dades, les bases de dades i els sistemes de gestió de bases de dades.
- 2.** Conèixer quins són els models de dades habituals per a dissenyar una base de dades.
- 3.** Conèixer quines són les etapes en el disseny d'una base de dades.
- 4.** Conèixer els elements bàsics que componen un sistema gestor de bases de dades i quines són les seves funcions principals.

## 1. Les dades i les bases de dades

El modelatge d'una base de dades requerirà l'estudi del món que es vol descriure, reconeixent la informació i el coneixement que es vol representar i la forma com es relacionen els elements o entitats d'aquest món entre si.

Les dades són la representació física de la informació que considerem rellevant del món real. La transformació d'aquesta informació (inclòs tot el coneixement que l'envolta) a dades, representades en un suport informàtic, requereix un procés previ de modelatge de la base de dades.

Distingirem, doncs, tres àmbits diferents que ens ajudaran a definir els passos a seguir durant el procés de modelatge de la nostra base de dades:

- El món real, en què hi ha la informació i el coneixement que volem representar.
- El món conceptual i lògic, que sintetitza el coneixement en entitats i les relacions concretes entre aquestes.
- El món físic, en el qual els diferents conceptes identificats es representen mitjançant elements informàtics.

### Representació de la informació

Hi ha diferents maneres de representar una mateixa realitat. La seva eficiència vindrà donada segons l'ús que se'ls doni.

En aquesta assignatura ens centrarem en una part del disseny de la base de dades: la que permet representar i estructurar la informació rellevant de manera que després pugui ser consultada i manipulada.

Més endavant, en altres mòduls, veurem com crear el model conceptual i lògic. Ambdós són part del procés de creació d'una base de dades. El model físic, però, no es veurà en aquesta assignatura.

### 1.1. Els models de dades

Els models de dades permetran organitzar les dades mitjançant diferents tipus d'estructures. La representació de les dades es farà des d'un punt de vista lògic, no físic ni d'implementació.

Hi ha diferents aproximacions al modelatge d'una base de dades i cadascuna ofereix uns avantatges i uns inconvenients. És per això que, depenent de l'ús que vulguem donar a la informació, serà convenient utilitzar-ne una o una altra.

Entre alguns dels models lògics principals trobem:

- Model jeràrquic
- Model en xarxa
- Model relacional
- Model documental

### 1.1.1. Model jeràrquic

En el model jeràrquic, l'estructura que suporta les dades i les seves relacions és l'arbre. Així doncs, en el modelatge jeràrquic les dades s'organitzaran de manera que un element es relacioni amb els altres elements seguint l'estructura *pare - fill*, excepte l'arrel que no té cap element *pare*. Aquest tipus de modelatge, per la seva simplicitat, va ser un dels més utilitzats en els primers sistemes de gestió de bases de dades.

És important assenyalar que aquest model de dades servirà únicament per a representar relacions d'un element a molts (un pare - molts fills), per la qual cosa permetrà modelar de forma simple les relacions bàsiques.

#### Exemple de model jeràrquic

Podem pensar en el context d'una biblioteca, en què interessa desar i manipular informació sobre llibres, com ara els capítols que conté, l'autor, o els altres llibres que aquest ha escrit, etc.

En conseqüència, aquest tipus de modelatge no ens servirà per a representar relacions en què més d'un element d'un tipus concret estan relacionats amb més d'un element d'un altre tipus.

Per tant, seguint l'exemple anterior, el model jeràrquic no ens permetrà representar els casos en què un llibre tingui més d'un autor. Així doncs, aquest model de dades redueix la capacitat de representar fidelment alguns elements del món real.

### 1.1.2. Model en xarxa

El model en xarxa és una extensió del model jeràrquic que intenta donar cobertura al problema entre les relacions *molts a molts* que ens podem trobar durant el modelatge. L'estructura que suporta les dades en aquest model són els *grafs*. Aquest model es coneix com a model en xarxa per la manera que tenen les dades de representar-s'hi i relacionar-s'hi.

En aquest tipus de model, l'estructura *pare - fill* s'expandeix de manera que permet que un fill tingui més d'un pare, acceptant la representació de la redundància en les relacions entre diferents elements. Amb tot, els elements de la base de dades es relacionaran amb altres elements de forma binària.

#### Origen del model jeràrquic

El model jeràrquic va ser desenvolupat per IBM a principis dels anys seixanta del segle passat.

#### Origen del model en xarxa

L'inventor original del model en xarxa va ser Charles Bachman i es va convertir en una especificació estàndard el 1969.



Així doncs, seguint aquest model, un llibre podrà tenir múltiples autors, però no podrem establir una relació entre els llibres, els autors i el gènere literari.

Les relacions entre els diferents elements estaran fixades sense tenir en compte el valor de la informació, per la qual cosa no s'assegura la coherència de les dades. A més, el fet que els camins entre els elements de la base de dades estiguin fixats, no afavorirà un accés flexible a les dades.

### 1.1.3. Model relacional

El model relacional és el més utilitzat en l'actualitat per a modelar problemes reals i administrar dades de forma dinàmica, i és el model que estudiarem en aquest curs. És un model amb base matemàtica, basat en la teoria de conjunts i que presenta una sèrie de restriccions, entre aquestes les d'integritat referencial.

#### Origen del model relacional

El model relacional va ser plantejat per Edgar Frank Codd, d'IBM, el 1970. Per la seva flexibilitat es va consolidar ràpidament com a paradigma de modelatge de bases de dades.

La **integritat referencial** és la propietat que ens assegurarà que, per cada dada d'un element que faci referència a una certa dada d'un altre element, aquesta dada existirà i serà vàlida, reforçant d'aquesta manera la coherència del model de dades.

En aquest model els diferents elements o entitats, que representen els conceptes del món real que volem modelar i les seves interrelacions, s'emmagatzemaran en estructures de dades tabulars, és a dir, en *taules* que s'interrelacionaran les unes amb les altres mitjançant associacions entre les seves dades.

De forma similar al model en xarxa, aquestes interrelacions permetran connectar i creuar les dades de diferents entitats (taules) en relacions de tipus *un a molts* i *molts a molts*. Cada entitat representada en una taula estarà formada per un conjunt d'atributs (columnes) que la diferenciaran de la resta d'entitats i, tanmateix, diferenciaran les diferents ocurrències d'una mateixa entitat (files).

Per tant, seguint l'exemple anterior, un autor serà definit per un nom, uns cognoms, una data de naixement, un lloc de naixement, etc. Alguns d'aquests atributs, a més, permetran relacionar diferents entitats entre si de manera que, per exemple, puguem obtenir informació sobre un autor (entitat) a partir del nom (atribut) de la ciutat (entitat) on ha nascut.

Aquestes relacions, entre les diferents taules a partir dels atributs que comparteixen, permetran mantenir el que anomenem la *integritat referencial*. Aquest mecanisme ens facilitarà la consistència entre les dades de les diferents taules, de manera que no es puguin alterar els valors de les ocurrències d'una taula sense que les taules relacionades també es vegin afectades per aquest canvi.

Per a aconseguir aquest propòsit, cada taula tindrà un conjunt d'atributs que permetran identificar de forma unívoca les seves ocurrences i que anomenarem claus.

Una **clau** és un atribut o un conjunt mínim d'atributs, que permet identificar entitats que representen objectes de forma unívoca. D'entre tots els possibles conjunts, se n'escollirà un, el qual anomenarem clau primària.

Per a ser considerats com a clau, aquests conjunts de dades hauran de ser tan petits com es pugui, evitant atributs sobrats i no serà estrany trobar més d'un conjunt possible. A més, els atributs que faran referència a les taules relacionades hauran de contenir dades vàlides en correspondència amb les claus primàries que permetin mantenir les referències.

Totes aquestes dades s'hauran de poder consultar o manipular. Per això disposem d'un llenguatge formal, l'àlgebra relacional, que permet descriure la forma com volem realitzar les consultes sobre les diferents taules de la base de dades. Com que no és executable, tenim un altre llenguatge per a accedir a les dades, l'SQL (*structured query language*).

El llenguatge SQL (per les seves sigles de l'anglès *structured query language*) és el més comú per a construir consultes i representa un estàndard sobre el qual s'han definit diferents implementacions segons el motor de la base de dades relacional.

Com veurem al llarg del curs, el llenguatge SQL està dividit en subllenguatges que ens permetran emmagatzemar i manipular les dades, i també gestionar la manera com interactuaran els diferents usuaris amb el sistema gestor de base de dades.

Aquests subllenguatges són:

- **Llenguatge de definició de dades** (LDD o DDL, en anglès). Aquest llenguatge permetrà definir l'estructura de la base de dades indicant quines seran les seves taules, els seus atributs, dependències entre unes taules i les altres, etc. Aquestes definicions s'emmagatzemaran en un arxiu especial anomenat diccionari de dades.
- **Llenguatge de manipulació de dades** (LMD o DML, en anglès). Aquest llenguatge servirà per a gestionar la informació continguda en la base de dades com ara les operacions per a inserir, recuperar, eliminar o modificar les dades.

#### Nota

Tot atribut identificador és una clau, però no tota clau és un atribut identificador.

- **Llenguatge de control de dades** (LCD o DCL, en anglès). Aquest llenguatge inclourà requeriments per a gestionar l'accés que els diferents usuaris tindran a la base de dades o les operacions que aquests podran realitzar sobre les dades.

#### 1.1.4. Model documental

Els models basats en documents ens permetran gestionar grans quantitats de dades. Al contrari del model relacional, la integritat referencial no serà el concepte sobre el qual pivotarà el seu disseny sinó que ho seran les dades en si, agrupades en col·leccions de documents.

De forma similar al concepte d'objecte en el paradigma de la programació orientada als objectes, les dades estaran emmagatzemades en documents estructurats que seguiran un esquema *clau-valor*.

Aquesta manera d'estructurar les dades ens facilitarà la possibilitat de realitzar operacions lògiques amb elements que, malgrat que no estiguin definits de la mateixa manera, comparteixen característiques similars.

Aquests sistemes gestors de bases de dades permetran, a més, un creixement del volum de dades que es durà a terme d'una manera molt simple, sense implicar un gran cost en el disseny i que es podran executar en sistemes de cost reduït, ja que no requerirà una gran capacitat de computació.

Les bases de dades que segueixen el paradigma NoSQL són aplicacions directes d'aquest model lògic de bases de dades.

#### 1.2. Sistemes de gestió de bases de dades

L'objectiu principal d'un sistema de gestió de bases de dades (SGBD) serà el de proveir mecanismes que assegurin un bon rendiment del sistema, garantint la disponibilitat de les dades i certificant la seva validesa en tot moment.

Un **SGBD** és una aplicació informàtica que permet la interacció dels usuaris amb una base de dades, altres aplicacions i els mateixos mòduls que la conformen mitjançant una interfície.

Un SGBD basat en el model relacional constarà, principalment, dels elements següents:

#### Esquema clau-valor

L'accés a les dades seguint l'esquema clau-valor és aquell en què s'accedeix a la dada mitjançant un valor únic, que anomenarem clau.

- El controlador d'arxius, que gestiona l'assignació de l'espai físic que ocuparan les dades en un o diferents sistemes d'emmagatzemament (disc, cintes, núvol, etc.).
- El motor de la base de dades, que gestiona l'accés a les dades i serveix d'interfície entre les aplicacions externes i les dades, permetent les operacions d'alta, baixa, modificació i eliminació.

### 1.2.1. Funcions principals del sistema de gestió de bases de dades

L'objectiu d'un SGBD serà, doncs, el de proveir d'un conjunt d'aplicacions necessàries per a treballar amb les dades. El disseny d'aquestes aplicacions serà el de possibilitar, en tot moment, l'emmagatzematge i la recuperació de la informació de la manera més eficient i segura possible.

Les funcions principals de l'SGBD són:

- Permetre l'emmagatzematge de les dades tenint en compte els diferents tipus d'informació que es guardarà com ara cadenes de text, nombres, dates, etc. optimitzant l'espai en el disc segons sigui la necessitat de cadascun d'aquests tipus. A més, haurà de permetre emmagatzemar text en diferents codificacions i formats de forma simultània.
- Permetre la manipulació i l'accés a les dades per part de l'usuari de manera eficient i segura. La seguretat de la base de dades estarà determinada pel control de l'accés a la informació, sia a persones o altres programes informàtics, amb l'objectiu de protegir les dades d'un accés no autoritzat. Per a garantir aquesta seguretat en l'accés a les dades, un SGDB haurà de proveir els mecanismes per a registrar qui (o què) accedeix en cada moment a les dades amb vista a facilitar auditories posteriors.
- Oferir mecanismes que permetin la recuperació fàcil de les dades en cas que es produeixin caigudes en el sistema de la base de dades. Aquests mecanismes han de ser capaços de restablir l'estat original de la base de dades (és a dir, l'estat abans del problema).
- Oferir mecanismes per a recuperar la base de dades fins a un estat estable en el supòsit que algun problema hagi pogut corrompre les dades. Aquests mecanismes també permetran facilitar la migració de la base de dades a nous entorns, de manera que es pugui replicar la informació o canviar el maquinari en què està instal·lat l'SGDB.
- Permetre l'accés concurrent a les dades per a mantenir la consistència de la informació, de manera que es concedeixi que diferents usuaris actualitzin i accedeixin a les dades de forma simultània. Per a garantir aquesta integritat en les dades es defineix, entre d'altres, el concepte de transacció.

Una **transacció** és un conjunt d'operacions considerades com una unitat única, per tant, en el cas que hi hagi un error durant l'execució d'aquestes operacions, les dades involucrades tornaran a prendre el valor previ a la seva execució.

## 2. Disseny i desenvolupament d'una base de dades

El procés de disseny d'una base de dades consisteix en definir l'estructura lògica i física d'una base de dades per a respondre a les necessitats i funcionalitats que els propietaris de les dades volen oferir als seus usuaris, sia persones o bé un conjunt concret d'aplicacions informàtiques, facilitant l'accés i la manipulació de la informació.

Aquest procés de disseny es dividirà en diferents etapes seqüencials, en què el resultat de la finalització de cadascuna servirà de punt de partida de l'etapa següent. Aquest procés podrà ser cíclic i en molts casos serà habitual refer el disseny d'una etapa prèvia a partir de les necessitats detectades en les fases posteriors.

El resultat de la darrera etapa serà el disseny final de la base de dades a implementar. Quan plantegem el procés de desenvolupament de la base de dades, hem d'incloure una primera fase de recollida i anàlisi dels requisits, i una darrera fase d'implementació i optimització de la base de dades. Tot seguit, veurem en què consisteixen cadascuna d'aquestes etapes.

### 2.1. Etapa de recollida i anàlisi de requisits

El primer pas en el desenvolupament d'una base de dades serà la recollida i l'anàlisi dels requisits, restriccions i necessitats que tenen els propietaris de les dades.

Per a obtenir aquesta informació caldrà, a més a més, mantenir converses amb els diferents usuaris de la futura base de dades i de les aplicacions que hi accediran per a tenir una visió global de com s'haurà d'estructurar i emmagatzemar la informació.

Serà, doncs, tan important obtenir el coneixement conceptual de la informació que s'haurà de gestionar com el coneixement de la manera en què els usuaris la tractaran.

### 2.2. Etapa de disseny conceptual

El segon pas serà el del disseny conceptual. En aquesta etapa, a partir dels requisits recollits a la primera etapa, es crearà un esquema o mapa conceptual d'alt nivell.

En aquest procés es representaran les dades mitjançant un model de dades que visualment permeti identificar els elements que formaran part de la base de dades, la seva caracterització en termes d'atributs o propietats i com es relacionen entre si. Aquest model s'anomena esquema conceptual.

**Model ER**

En aquests materials farem servir el model Entitat-Relació (ER), també anomenat entitat-interrelació, per a representar els models conceptuals de les dades.

### 2.3. Etapa de disseny lògic

El tercer pas en el desenvolupament de les bases de dades és el disseny lògic. En aquest punt s'ha de determinar el model de dades de la base de dades que s'utilitzarà d'entre els diferents models possibles que hem presentat.

En aquesta etapa l'esquema conceptual es transformarà en un esquema lògic adequat al model de dades que s'utilitzarà.

En aquests materials, atès que treballarem el model relacional i partirem d'un esquema ER, estudiarem el disseny lògic basat en la transformació de diagrames ER al model relacional.

### 2.4. Etapa de disseny físic

El quart pas en el desenvolupament de les bases de dades és el disseny físic. En aquest punt s'adaptarà l'esquema lògic a les necessitats específiques d'un SGBD concret.

En aquesta etapa es tindran en consideració tots els requisits tecnològics necessaris per a la implementació de la base de dades que es vol crear. En aquests materials veurem exemples de creació de bases de dades mitjançant el MySQL.

El MySQL és un sistema gestor de bases de dades de programari lliure molt popular sota les condicions de la llicència GPL que ofereix diverses interfícies de programació d'aplicacions (API en anglès) que permeten a les aplicacions escrites en diversos llenguatges de programació accedir a les bases de dades MySQL.

### 2.5. Etapa d'implementació i optimització

La darrera etapa és la implementació i optimització de la base de dades. L'objectiu d'aquesta etapa serà carregar les dades i ajustar alguns paràmetres del model físic i de la base de dades amb l'objectiu d'optimitzar-ne el rendiment.

## Resum

En aquest mòdul hem fet una introducció als conceptes fonamentals de les bases de dades i els sistemes gestors de bases de dades. Hem vist una definició de què són les dades i com podem representar-les d'acord amb un marc referencial.

També hem explicat els diferents models de dades que permeten representar el que ens interessa modelar del món real per a donar resposta a algun problema informàticament.

En aquesta assignatura ens centrarem en el model relacional que permet representar els objectes del món real, que anomenarem entitats, en taules; les seves propietats o característiques, que anomenarem atributs, en camps i les relacions entre aquestes mitjançant claus que les relacionen.

Hem comentat l'existència d'un llenguatge d'accés estandard a les dades, l'SQL, i els subllenguatges que el conformen.

Hem revisat les funcions principals dels SGBD actuals i alguns dels serveis que ens donen per a aconseguir-los. És especialment important el concepte de seguretat en la integritat de les dades i el de transacció.

Finalment, hem enumerat les diferents etapes del procés de desenvolupament d'una base de dades, incloent les tres etapes del procés de disseny de les bases de dades (conceptual, lògic i físic), des de la recollida de les dades fins a la selecció de l'SGBD, passant per la representació del model conceptual en diagrames entitat-relació.



## Glossari

**base de dades** *f* Conjunt de dades emmagatzemades per a la seva recuperació i manipulació posterior, independentment de l'aplicació que evita la redundància, garanteix la integritat de les dades i tolera les fallades.

**clau** *f* Atribut o conjunt d'atributs que permet identificar els objectes de forma unívoca.

**clau primària** *f* Atribut o conjunt d'atributs clau escollit com a principal.

**dada** *f* Nom que rep la informació en el món de les representacions informàtiques.

**ER** *m* El model *entitat-relació*, de l'anglès *entity-relationship* és un model conceptual d'alt nivell que permet identificar els elements que componen una base de dades.

**integritat referencial** *f* Mecanisme que facilita la consistència entre les dades de la base de dades.

**LDD** *m* Llenguatge de definició de dades. Permet la descripció de les dades per a un SGDB.

**LMD** *m* Llenguatge de manipulació de dades. Permet la interacció de les aplicacions amb un SGBD per a recuperar, modificar i eliminar informació d'una base de dades.

**model relacional** *f* Model de dades en què les dades que descriuen un segment de la realitat s'emmagatzemen en estructures en forma de taules que s'interrelacionen les unes amb les altres mantenint la integritat referencial.

**MySQL** *m* Sistema gestor de bases de dades de programari lliure molt popular que ofereix interfícies de programació d'aplicacions per a accedir i manipular les dades.

**SGBD** *m* Sistema gestor de bases de dades. Aplicació informàtica que permet la interacció dels usuaris amb una base de dades, altres aplicacions i els mateixos mòduls que la conformen mitjançant una interfície.

**SQL** *m* Llenguatge estructurat de consultes, de l'anglès *structured query language*. Llenguatge que permet construir consultes i representa un estàndard sobre el qual s'han definit diferents implementacions segons el motor de la base de dades relacional.

**taula** *f* Estructura dins d'una base de dades relacional en què s'emmagatzemen les dades que representen un concepte particular d'un segment de la realitat.

