

M02. BASES DE DADES

UF1. Introducció a les bases de dades

INTRODUCCIÓ A LES BBDD

Per què emmagatzemar
informació?

1. INTRODUCCIÓ

LES DADES.

Què és una dada?

- **representació simbòlica** (numèrica, alfabètica,...) d'una característica d'una entitat
- **no té valor semàntic** (sentit) per si mateixa

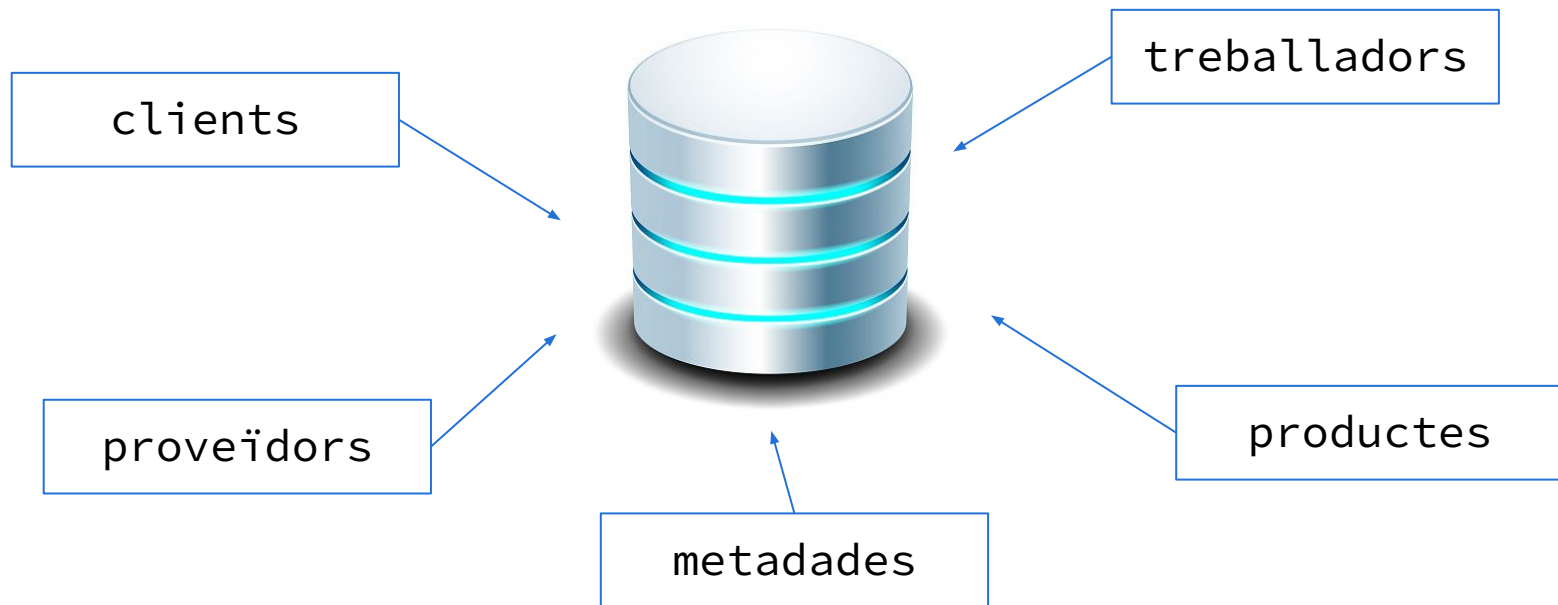
Què són les metadades?

- dades sobre les dades
- informació descriptiva (context, qualitat, condició o característiques) d'un recurs, dada o objecte
- permet de facilitar-ne la recuperació, autenticació, avaluació, preservació i/o interoperabilitat

2. GESTIÓ DE LES BASES DE DADES

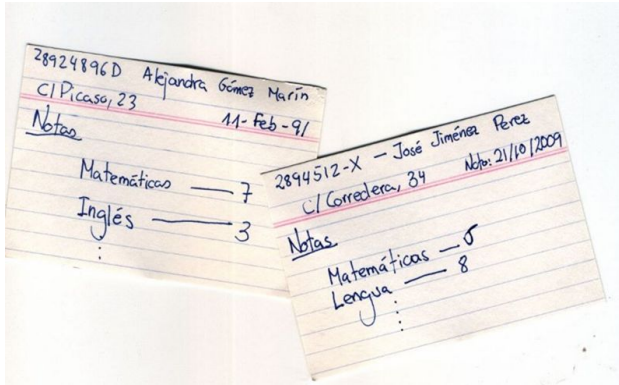
LES DADES.

Base de dades: col·lecció de dades rellevants



LES DADES. GESTIÓ DE LES BASES DE DADES

Com s'emmagatzemen les dades? On?



LES DADES. GESTIÓ DE LES BASES DE DADES

Anys 60 - 70

- Aplicacions de propòsit específic i processament per lots
- Diferents tipus de suports per a emmagatzemar les dades (cintes magnètiques, targetes microperforades, fitxers)
- DB redundants: a cada inserció/modificació es creava un nou fitxer (màster i còpies)



[The CDC 6600 supercomputer, circa 1964](#)
[/ Photo by Arnold Reinhold](#)

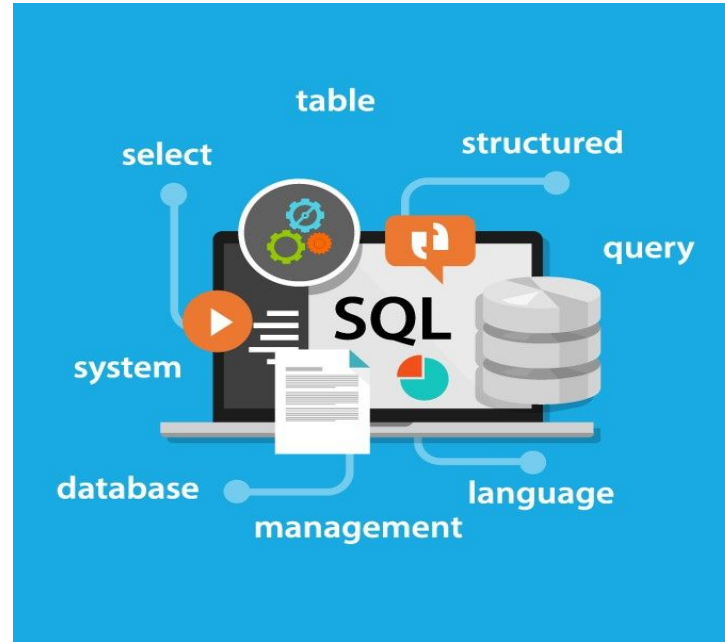
LES DADES. GESTIÓ DE LES BASES DE DADES

Anys 70 - 80

- L'aparició dels PCs (Personal Computers) estén la informàtica (empreses i institucions)
- Model E-R (entity-relationship) per Peter Chen (1976)
- Model de base de dades relacional proposat per E. F. Codd al 1970
 - DB relacionals
 - Primers SGBD comercials (DB2/IBM, Oracle)
- SQL (Structured Query Language) per a gestionar DB relacionals (inicis dels anys 80)

LES DADES. GESTIÓ DE LES BASES DE DADES

Anys 70 - 80



Característiques principals de SQL

LES DADES. GESTIÓ DE LES BASES DE DADES

Anys 90

- Boom dels DBMS open source (MySQL, PostgreSQL)
- 4GL pel desenvolupament d'aplicacions basades en DB
- Accés online a les DB (entorn client/servidor)
- Bases de dades distribuïdes



LES DADES. GESTIÓ DE LES BASES DE DADES

I avui?

- TAD (Tipus de Dades Abstractes)
- XML DB per a emmagatzemar tipus de dades complexes
- DB orientades a objectes (OODB)
- NoSQL per a Big Data i aplicacions web en temps real
- Cloud DB (distribució i accés al núvol)



LES DADES. GESTIÓ DE LES BASES DE DADES

DBMS (*Database management system, Sistema gestor de bases de dades*)

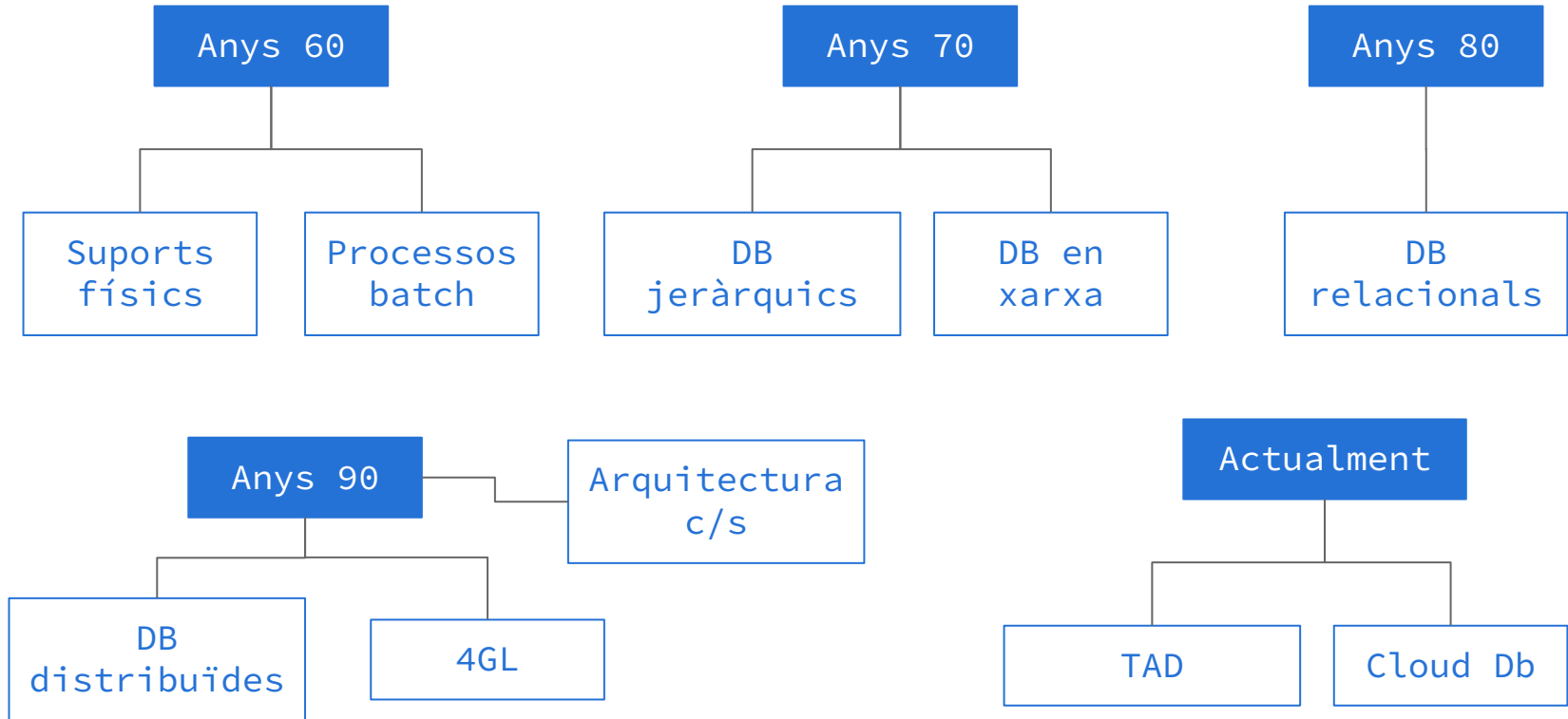
Un *database management system* (DBMS) és un programari o conjunt d'aplicacions que permet accedir a les dades i operar amb elles (creació, modificació, eliminació...) així com atendre les sol·licituds d'accés a la base de dades que fan els usuaris i/o aplicacions a la base de dades. Els DBMS gestionen grans quantitats d'informació, garantint la seva fiabilitat.

LES DADES. GESTIÓ DE LES BASES DE DADES

IS (Information Systems, Sistemes d'Informació, SI)

Un sistema d'informació (IS, Information Systems), en canvi, és un sistema format per persones, dades, activitats, i en definitiva, el conjunt de recursos que processen la informació d'una organització. Els Sistemes d'Informació informàtics són el camp d'estudi de les Tecnologies de la Informació i la Comunicació (TIC).

LES DADES. EVOLUCIÓ DELS DBMS



4. REPRESENTACIÓ DE LES DADES

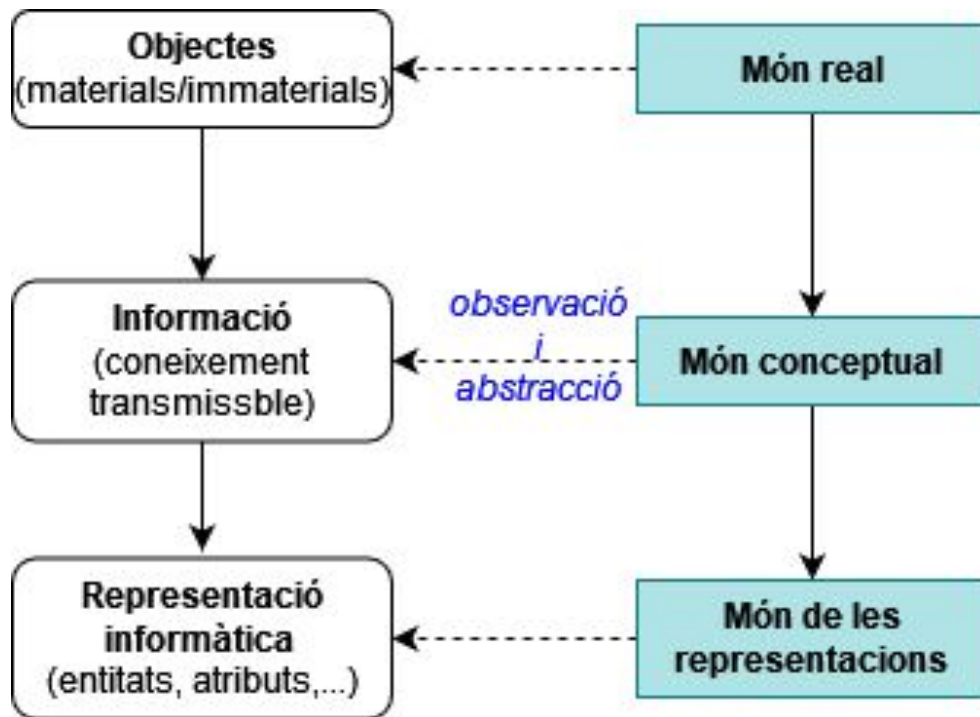
LES DADES. COM ES REPRESENTEN?

Organitzem el nostre entorn en “mons”:

- **món real:** objectes (materials o no) de la realitat que ens interessen i amb els quals haurem de treballar
- **món conceptual:** coneixements o informacions obtinguts mitjançant l'observació de la part del món real que ens interessa.
- **món de les representacions:** representacions informàtiques, o dades, del món conceptual, necessàries per poder treballar

LES DADES. COM ES REPRESENTEN?

Conceptualització del nostre entorn



LES DADES. COM ES REPRESENTEN?

Podem transformar les observacions del món real en dades informàtiques de manera automàtica?

- **fase de disseny lògic:** es treballa amb el model abstracte de dades resultant de finalitzar l'etapa de disseny conceptual, per tal de traduir-ho al model de dades utilitzat pel DBMS amb el qual es vol implementar i mantenir la futura DB.
- **fase de disseny físic:** optimització de l'esquema lògic obtingut en la fase de disseny anterior, per tal d'incrementar l'eficiència en algunes de les operacions a fer amb les dades.

LES DADES. EL MÓN CONCEPTUAL

La informació és la conceptualització obtinguda a partir de l'observació del món real. Es caracteritza, fonamentalment, per tres elements: entitats, atributs i valors.

- **entitats:** objectes del món real que conceptualitzem. Identificables (distingibles els uns dels altres), ens interessen com a mínim una de les seves propietats (atributs).
Exemple: COTXE, ALUMNE, BANC, MOBLE
- **atributs:** són les propietats que ens interessen de les entitats.
Exemple: color, edat, oficina, material
- **valors:** són els continguts concrets dels atributs, les determinacions concretes que assoleixen.
Exemple: vermell, 17, BCN-003, fusta

LES DADES. EL MÓN CONCEPTUAL. ENTITATS

- **Entitat tipus (template)**: tipus genèric d'entitat (abstracció), que fa referència a una classe de coses (cotxe, arbre, alumne,...)
- **Entitat instància**: conceptualització d'un objecte concret del món real (*Ford Mustang Vermell*), distingible dels altres objectes del mateix tipus, gràcies a alguna propietat (com podria ser el valor de l'atribut *matrícula*).

LES DADES. EL MÓN CONCEPTUAL. ATRIBUTS

- **tipus de dada:** defineix un conjunt de valors amb unes característiques comunes que els fan compatibles i també defineix una sèrie d'operacions admissibles sobre aquests valors.
Exemple: enters, caràcters, booleans,...
- **domini:** conjunt de valors vàlids per a l'atribut en qüestió. Pot ser predefinit (enters, reals,...) o definit per l'usuari (rang d'edat)
- **atribut identificador:** és el que permet distingir inequívocament cada entitat instància de la resta. Valor és únic (no es repeteix en les diferents entitats instància)
- **clau:** atribut o conjunt d'atributs que permeten identificar una entitat tipus

LES DADES. EL MÓN DE LES REPRESENTACIONS

Una **dada** és una representació simbòlica (numèrica, alfabètica, ...) d'una característica d'una entitat (objecte de la vida real). La forma més habitual per a representar un conjunt de dades és la **representació tabular**.

student ID	first name	last name	subject	grade
1	María	García	PB	9
2	Miguel	Fernández	BD	6
3	Susana	Guerrero	ISO	7
4	Ernesto	López	PAX	4

← registre

camp

valor

LES DADES. EL MÓN DE LES REPRESENTACIONS

Com accedim a les dades?

	P - per posició	V - per valor
S - seqüencial	SP	SV
D - directe	DP	DV

SP (accés seqüencial per posició): un cop s'ha accedit a un registre que es troba en una posició determinada, s'accedeix al registre que ocupa la posició immediatament posterior.

DP (accés directe per posició): s'obté directament un registre pel fet d'ocupar una posició determinada.

SV (accés seqüencial per valor): un cop s'ha accedit a un registre que té un valor concret, s'accedeix al registre que ocupa la posició immediatament posterior.

DV (accés directe per valor): s'obté directament un registre pel fet de tenir un valor determinat en un dels seus atributs (o més).

LES DADES. EL MÓN DE LES REPRESENTACIONS

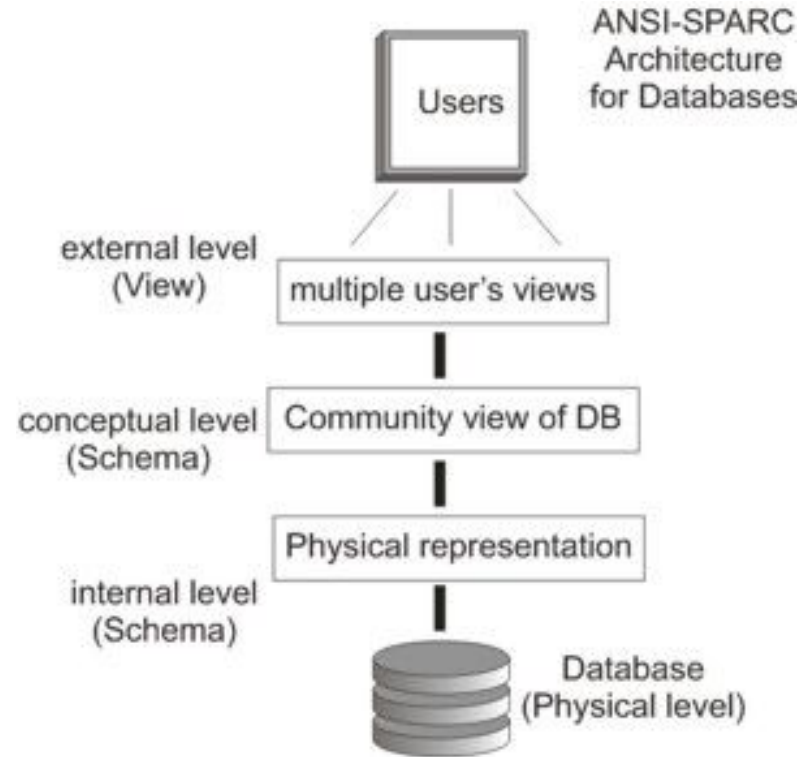
Nivells de representació

- **nivell lògic:** es treballa segons la conceptualització de les dades, sense necessitat de saber com s'emmagatzemen de manera física
- **nivell físic:** requereix un coneixement a baix nivell de la implementació física de l'organització de les dades i l'accés a aquestes

LES DADES. EL MÓN DE LES REPRESENTACIONS

Arquitectura ANSI/X3/SPARC

- **Nivell Extern o Lògic (User Views):** part de la base de dades rellevant per a l'usuari
- **Nivell Conceptual:** descriu les dades i les interrelacions, sense especificar com s'emmagatzemen físicament
- **Nivell Intern o Físic:** implica com es representa físicament la base de dades al sistema informàtic



LES DADES. EL MÓN DE LES REPRESENTACIONS

Arquitectura ANSI/X3/SPARC

L'avantatge d'aquesta arquitectura en nivells és que proporciona independència lògica i física de les dades respecte a les aplicacions:

- **Independència lògica:** es poden fer canvis en el nivell conceptual (afegir taules o atributs) sense que sigui necessari reescriure totes les aplicacions.
- **Independència física:** és possible modificar la ubicació dels fitxers que contenen les dades sense que es vegin afectades les aplicacions.

LES DADES. EL MÓN DE LES REPRESENTACIONS

Diccionari de dades

Un model de dades és un model abstracte que organitza elements de dades i estandarditza com es relacionen entre si i amb les propietats de les entitats del món real.

Per exemple, un model de dades pot especificar que l'element de dades que representa un cotxe es compon d'altres elements que, al seu torn, representen el color i la mida del cotxe i defineixen el seu propietari.

MODEL ENTITAT-RELACIÓ

Per què emmagatzemar
informació?

EL MODEL E-R

A l'hora de dissenyar una base de dades, simplifiquem el procés dividint-lo en etapes.

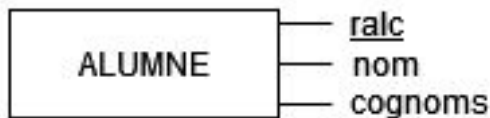
Les etapes són les següents:

- **etapa del disseny conceptual**: s'obté una estructura de la informació de la DB, independent de la tecnologia a utilitzar. El resultat és un model de dades d'alt nivell. En el nostre cas, el model entitat-relació (E-R)
- **etapa de disseny lògic**: s'ajusta el model E-R (resultat del disseny conceptual) al model DBMS amb el qual s'implementarà la db.
- **etapa del disseny físic**: es transforma l'estructura obtinguda a l'etapa del disseny lògic de manera que s'assoleixi una major eficiència (mida de buffers, pàgines,...) que dependran del DBMS escollir.

EL MODEL E-R. ENTITATS I ATRIBUTS

Entity-relationship (E-R) , per Peter Chen el 1976

- Es coneix també amb el nom de model entitat-interrelació
- És independent del tipus de DBMS (jeràrquic, en xarxa, relacional,...).
- **Univaluat**: un atribut és univaluat si té un sol valor per a cada ocurrència d'una entitat
- **Clau**: atribut que permet identificar l'entitat (cada instància té una clau diferent)
- **Clau Candidata**: conjunt de claus de l'entitat (atributs que identifiquen l'entitat)
- **Clau Primària**: atribut seleccionat d'entre totes les claus candidates com a identificador de l'entitat



Entitat i atributs

EL MODEL E-R. INTERRELACIONS

Interrelacions



Es defineix com una associació entre entitats. Una interrelació pot tenir atributs, que també han de ser

- univaluats
- valors atòmics
- domini

EL MODEL E-R. GRAUS DE LES INTERRELACIONS

Una interrelació pot associar dues o més entitats.

El **grau de la interrelació** indica el nombre d'entitats que associa una interrelació.

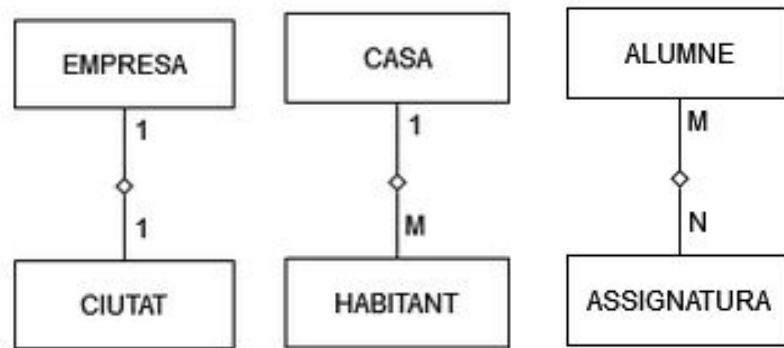
- **Interrelacions binàries**: interrelacions de grau dos (associa dues entitats)
- **Interrelacions n-àries**: interrelacions de grau més gran de dos

EL MODEL E-R. CONNECTIVITAT DE LES INTERRELACIONS

La connectivitat d'una interrelació expressa el tipus de correspondència que hi ha entre les ocurrències d'entitats associades amb la interrelació.

En el cas de les interrelacions binàries:

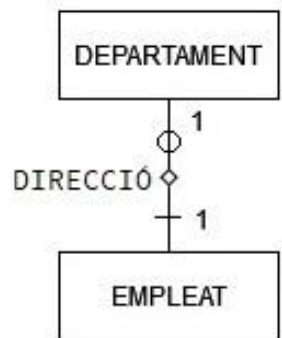
- Connectivitat un a un (1:1)
- Connectivitat un a molts (1:N)
- Connectivitat molts a molts (M:N)



EL MODEL E-R. CONNECTIVITAT DE LES INTERRELACIONS

Dins la connectivitat, podem indicar també la dependència de les entitats:

- **entitat obligatòria**
- **entitat opcional**



- L'entitat empleat és obligatòria a la interrelació direcció: no pot existir un departament que no tingui un empleat que exerceixi com a director de departament.
- L'entitat departament és opcional a la interrelació direcció: pot ser que hi hagi un empleat que no està interrelacionat amb cap departament: pot haver-hi empleats que no són directors de departament.

EL MODEL E-R. CAS PRÀCTIC

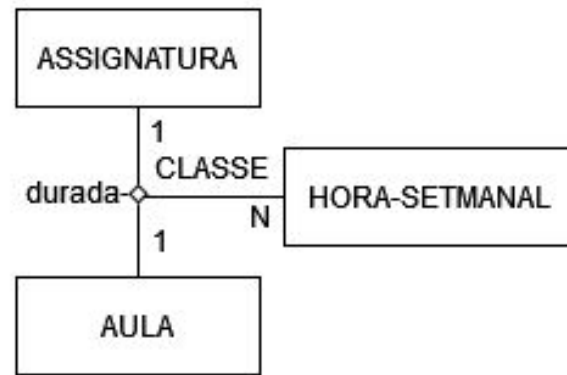
Base de dades de colònies

EL MODEL E-R. CONNECTIVITAT DE LES INTERRELACIONS

En el cas de les interrelacions n-àries:

- Connectivitat un a un a un (1:1:1)
- Connectivitat un a un a molts (1:1:N)
- Connectivitat un a molts a molts (1:M:N)
- Connectivitat molts a molts a molts (M:N:P)

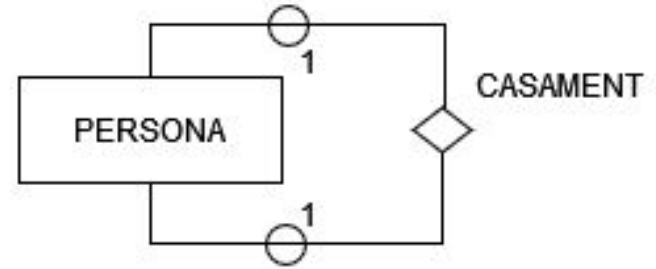
Una interrelació n-ària pot tenir $n + 1$ tipus de connectivitat, ja que cadascuna de les n entitats pot estar connectada amb “un” o amb “molts” a la interrelació.



Important! Per a les interrelacions ternàries hi ha quatre tipus de connectivitat possibles.

EL MODEL E-R. INTERRELACIONS RECURSIVES

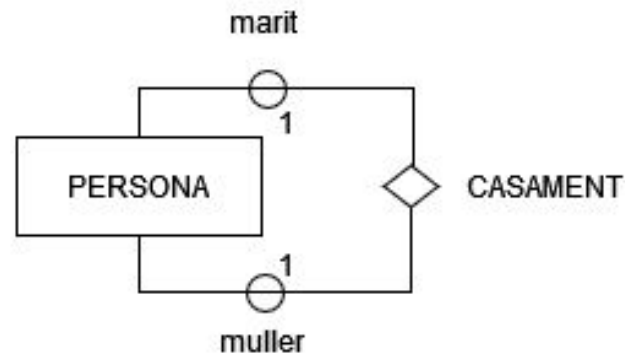
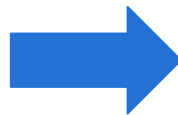
- Una **interrelació recursiva** és una interrelació a la qual alguna entitat està associada més d'una vegada.
- Permet distingir els diferents papers que una mateixa entitat té a la interrelació (es pot etiquetar cada línia de la interrelació amb un rol)



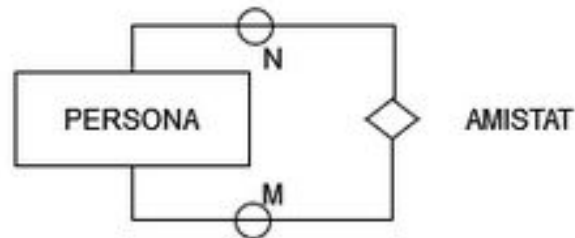
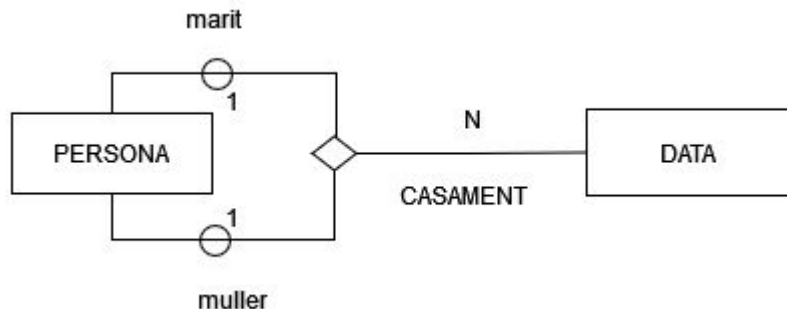
EL MODEL E-R. INTERRELACIONS RECURSIVES

Interrelació recursiva binària

- amb rols diferents
- no-diferència de rols



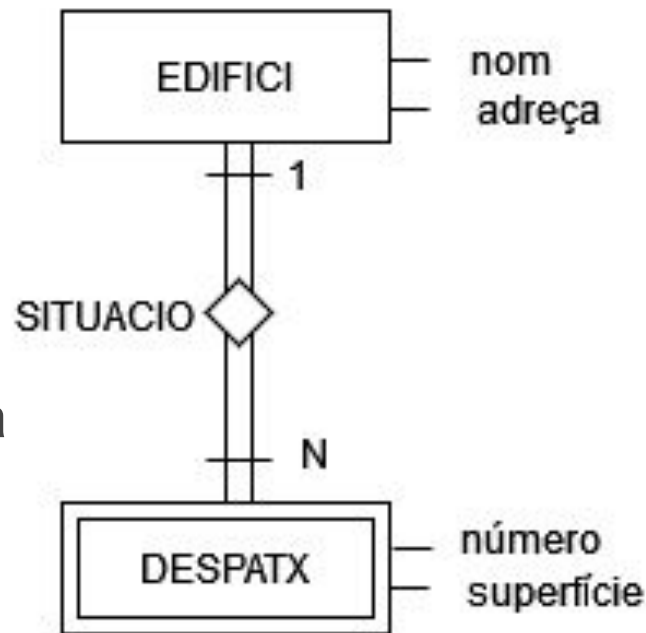
Interrelació recursiva n-ària



EL MODEL E-R. ENTITATS DÈBILS

- Els atributs d'una entitat dèbil no la identifiquen completament (identificació parcial)
- Ha de participar en una interrelació que ajuda a identificar-la.

Per a tota entitat dèbil, **sempre** hi ha d'haver una **única interrelació que en permeti completar la identificació**. Aquesta **interrelació** ha de ser **binària** amb connectivitat 1:N i l'entitat dèbil ha de ser al costat N.



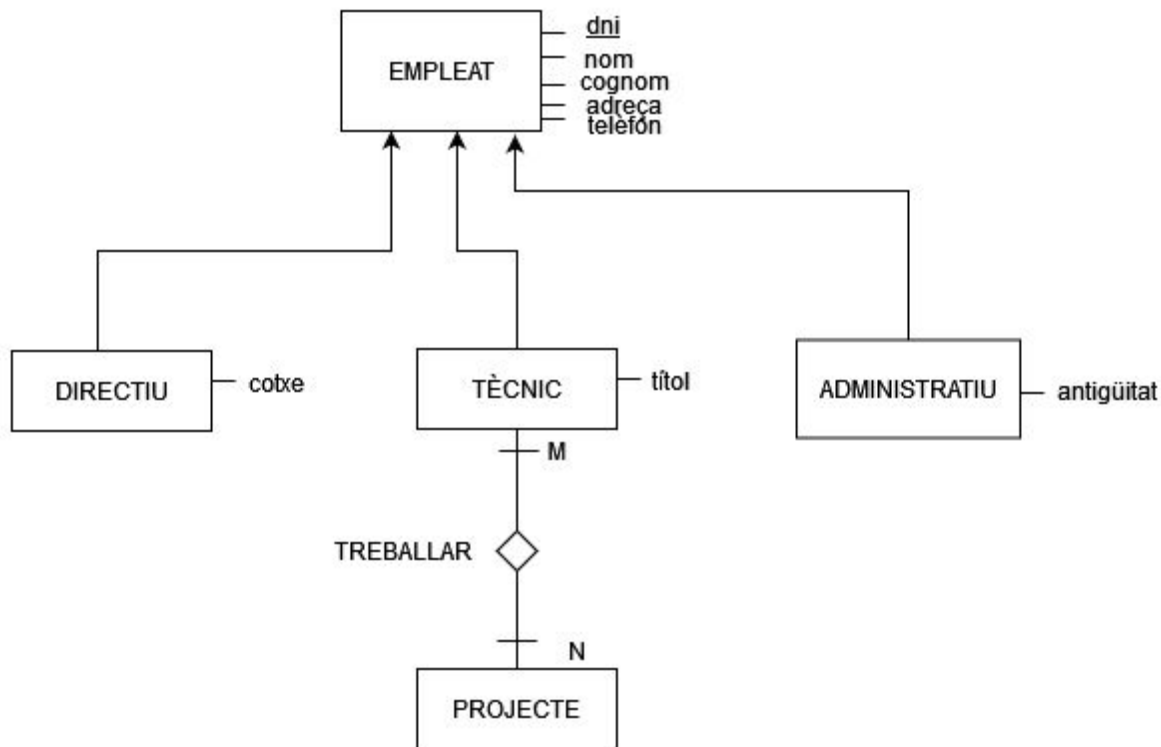
EXTENSIONS DEL MODEL E-R.

La **generalització/especialització** permet reflectir el fet que hi ha una entitat general, que anomenem entitat superclasse, que es pot especialitzar en entitats subclasse:

- L'entitat **superclasse** ens permet modelitzar les característiques comunes de l'entitat vista a un nivell genèric.
- Les entitats **subclasse** ens permeten modelitzar les característiques pròpies de les seves especialitzacions.

Cal que es compleixi que tota ocurrència d'una entitat subclasse sigui també una ocurrència de la seva entitat superclasse.

EXTENSIONS DEL MODEL E-R.



Herència de propietats:
Les característiques (atributs o interrelacions) de l'entitat superclasse es propaguen cap a les entitats subclasse.

EXTENSIONS DEL MODEL E-R.

Tipus d'especialització / Generalització:

- **Disjunta (D)** o exclusiva: no pot passar que una mateixa ocurrència aparegui a dues entitats subclasse diferents. Es denota gràficament amb l'etiqueta D.
- **Encavalcada (S)** o inclusiva: no hi ha la restricció anterior. Es denota gràficament amb l'etiqueta S.
- **Total (T)**: tota ocurrència de l'entitat superclasse ha de pertànyer a alguna de les entitats subclasse. Es denota amb l'etiqueta T.
- **Parcial (P)**: no cal que això passi. Es denota amb l'etiqueta P.