

Introducció a les bases de dades

FITXERS, BASES DE DADES I SGBD

EL MODEL CONCEPTUAL

OBJECTE, ENTITAT I ATRIBUTS



Per on comencem?

Què són els fitxers??

Fitxers

Quan la informació creixia, també creixien els arxivadors, de manera que podem tenir carpetes i subcarpetes:



Fitxers

Però ja podeu imaginar que arriba un moment en què aquest model és inviable.



Fitxers

La informàtica va copiar aquest sistema i va construir fitxers com a un sistema de carpetes i subcarpetes, fixant-se en el model físic

```
El volumen de la unidad C es ARANYA 1
El número de serie del volumen es 212A-58C9
Directorio de C:\

DOS                <DIR>          09-10-96  11:06a  DOS
COMMAND  DOS      56,565  05-31-94  6:22a  COMMAND.DOS
WINA20   386       9,349  05-31-94  6:22a  WINA20.386
CONFIG  DOS        162  02-03-97  10:39p  CONFIG.DOS
AUTOEXEC DOS       448  10-24-96  12:09a  AUTOEXEC.DOS
MTH      <DIR>          09-10-96  1:22p  MTH
AUTOEXEC BAT      572  04-19-97  12:11p  AUTOEXEC.BAT
COMMAND  COM     95,334  08-24-95  9:50a  COMMAND.COM
MSDOS    SYS     1,641  02-03-97  10:39p  MSDOS.SYS
NETLOG   TXT       547  09-10-96  1:35p  NETLOG.TXT
WINDOWS  <DIR>          09-10-96  1:35p  WINDOWS
CONFIG   SYS       162  07-21-97  11:19p  CONFIG.SYS
AE16     <DIR>          09-10-96  2:00p  AE16
MSINPUT  <DIR>          09-10-96  1:56p  MSINPUT
MRACK    <DIR>          09-10-96  4:31p  MRACK
ARCHIV"1 <DIR>          09-10-96  1:36p  Archivos de programa
WINDRV   <DIR>          09-10-96  4:40p  WINDRV
ANYWARE  <DIR>          09-10-96  5:06p  ANYWARE
VIRUS    BAT       81  11-22-96  5:26p  VIRUS.BAT
Presione cualquier tecla para continuar . . .
```

Fitxers

De fet, els sistemes d'arxius dels sistemes operatius, tenen aquest esquema:



Inconvenients dels fitxers

A més a més del volum de dades, aquest sistema basat en fitxers per mantenir la informació té diversos inconvenients:

- Redundància de dades
- Aïllament de les dades
- Problemes d'integritat
- Problemes de seguretat
- Accessos concurrents

Redundància de dades

Sovint, dins d'una organització, diferents programadors de diversos departaments creen fitxers i aplicacions durant períodes de temps llargs.

Això pot portar a **redundància de dades**, una situació que succeeix quan una determinada dada s'ha de modificar a diversos llocs.

Aquesta pràctica pot conduir a alguns problemes com ara:

- **Inconsistència** en el format de la dada.
- Que la mateixa informació estigui present a diversos llocs (fitxers), cosa que malgasta espai d'emmagatzematge i duplica esforços.
- *Inconsistència de dades*, una situació en què diverses còpies de la mateixa dada estan en conflicte.

Aïllament de les dades

Com els canvis fets per una operació són visibles a altres usuaris i sistemes concurrents.

Aquest problema es dona en situació de concurrència.

Això és un problema perquè:

- És difícil per a noves aplicacions obtenir les dades apropiades que poden estar guardades en diversos fitxers.

Problemes d'integritat

Un altre dels desavantatges dels sistemes basats en fitxers és la ***integritat de les dades***.

Es refereix al manteniment i la garantia que les dades del sistema són **correctes i consistents**.

Alguns factors a considerar quan d'adreça aquesta problemàtica són:

Els valors de les dades han de satisfer certes restriccions de consistència que estan especificades en les aplicacions.

És difícil fer canvis a les aplicacions per tal de forçar el compliment de noves restriccions.

Problemes de seguretat

La seguretat pot ésser un problema en l'aproximació basada en fitxers perquè:

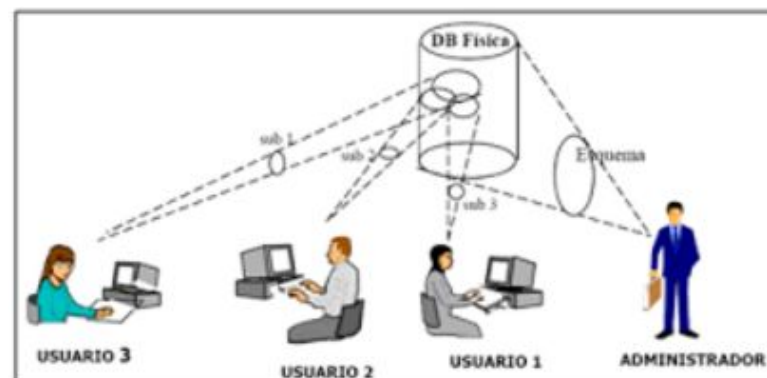
Hi ha restriccions relacionades als privilegis d'accés.

Nous requeriments de les aplicacions s'afegeixen al sistema cosa que dificulta imposar restriccions.

Accessos concurrents

La **concurrència** és l'habilitat del sistema de permetre que múltiples usuaris accedeixin a les mateixes dades sense que es doni un efecte advers al processament de transaccions.

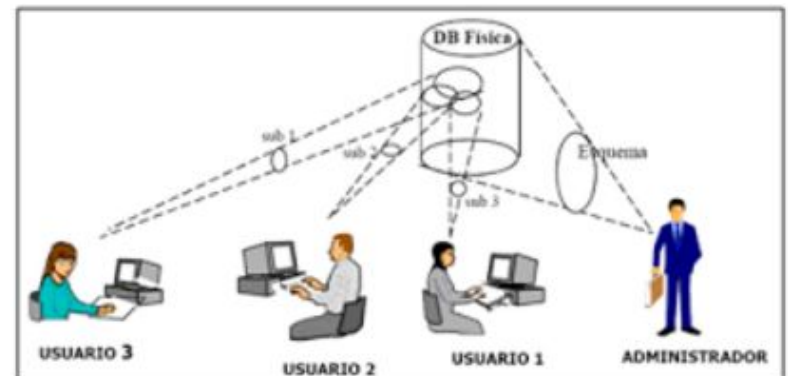
Un sistema basat en fitxers ha de gestionar, o evitar, la concurrència a través de les aplicacions. Típicament, en un sistema basat en fitxers, quan una aplicació obre un fitxer, aquest fitxer es bloqueja. Això significa que ningú més té accés a aquest fitxer al mateix temps.



Accessos concurrents

En sistemes de bases de dades, la concurrència es gestiona, permetent així que múltiples usuaris accedeixin al mateix registre.

Aquesta és una diferència important entre sistemes basats en fitxers i en bases de dades.



Fitxers vs Base de dades

A més a més de tots els inconvenients que hem enumerat, els fitxers es van quedar “curts”.

De seguida, el volum d'informació que es va voler guardar i les necessitats de fer recerques efectives van esdevenir en l'aparició dels **Sistemes Gestors de Bases de Dades**.



Les base de dades i els fitxers

Per què van aparèixer les bases de dades?

Perquè resolien molts dels problemes que plantejaven els fitxers:

	Fitxers	Bases de dades
Entitats tipus	Les entitats instància d'un fitxer pertanyen a una sola entitat tipus.	Les BD contenen entitats instància d'infinitat d'entitats tipus interrelacionades.
Interrelacions	El sistema no interrelaciona fitxers.	El sistema té previstes eines per interrelacionar fitxers.

	Fitxers	Bases de dades
Redundàncies	És necessari crear fitxers a mida de cada aplicació, amb totes les dades necessàries, encara que estiguin repetides en altres fitxers.	Tècnicament, totes les aplicacions poden treballar amb la mateixa BD, la qual cosa evita la redundància de dades i els riscos que comporta.
Inconsistències	És possible que els valors d'unes mateixes dades en diferents fitxers no coincideixin, si els programadors no les han actualitzat degudament.	Si les interrelacions estan ben dissenyades, les dades només han d'estar emmagatzemades en la BD un sol cop. Per tant, no hi ha risc d'inconsistències.
Obtenció de dades	Si no hi ha una aplicació que obtingui les dades que volem, o bé s'ha de fer un programa a mida, o bé s'ha d'aprofitar la sortida d'un programa amb objectius similars, i fer els càlculs necessaris manualment.	Permeten obtenir qualsevol conjunt de dades, segons les necessitats, dels del seu propi entorn de treball, sense haver d'escriure, compilar i executar cap nou programa d'aplicació contra la BD.
Aïllament de dades	Les dades estan disperses i aïllades en diferents arxius, la qual cosa dificulta el desenvolupament de les aplicacions.	Totes les dades són en la mateixa BD, interconnectades, la qual cosa en facilita l'obtenció.

	Fitxers	Bases de dades
Integritat de dades	Els programes han d'implementar totes les restriccions sobre les dades, afegint el codi font corresponent. El manteniment és complicat quan la informació es conté en diferents fitxers utilitzats per diferents aplicacions.	La BD s'encarrega directament d'implementar les restriccions sobre les dades. Els programes no han d'incorporar codi font addicional per garantir-les.
Atomicitat	Alguns conjunts d'operacions sobre les dades s'han d'executar de manera indivisible (o tots o cap), independentment de les fallades que el sistema pugui presentar (com ara per un tall de subministrament elèctric). Però això és molt difícil de garantir amb un sistema d'informació basat en fitxers.	Les BD incorporen la tècnica de les transaccions per tal de garantir fàcilment l'execució atòmica d'una pluralitat de processos sobre les dades.
Accés concurrent	L'actualització simultània de dades d'un mateix fitxer per part de diferents usuaris o aplicacions en pot provocar fàcilment la inconsistència.	Amb la tècnica del bloqueig, les BD garanteixen automàticament la consistència de les dades, malgrat que més d'un usuari o més d'una aplicació les vulguin actualitzar simultàniament.

Les base de dades i els fitxers

	Fitxers	Bases de dades
Seguretat	Habitualment, cada fitxer serveix per a un sol usuari o una sola aplicació (sobretot simultàniament), i ofereix una visió única del món real. Però no sempre tots els usuaris que utilitzen un fitxer haurien de tenir accés a totes les dades que conté.	Una BD pot ser compartida per molts usuaris de diferents tipus (fins i tot, simultàniament), els quals poden tenir diferents visions (vistes) del món real, en funció del seu perfil i dels permisos que s'hagin de concedir en cada cas.

Fitxers i BBDD

- Tots utilitzem les BBDD en un format o un altre. (document, fulls,...)
- Els fitxers són totalment plans, és a dir, necessito guardar totes les dades sense poder relacionar una amb altres.

Exemple cotxes:

Imaginem que tenim un taller mecànic. Volem guardar:

- Cotxes
- Clients/usuaris
- Stock de recanvis.

Dada

- I què és una **dada**? és una representació d'un objecte o concepte.
- Les dades esdevenen informació útil qual els donem un sentit, és a dir, quan les hem processat de tal manera que ens proporcionen coneixement.
- Les dades es poden emmagatzemar de múltiples formes:
 - Arxivadors,
 - Fulls de càlculs
 - Carpetes
 - Llibres de comptabilitat
 - Llistes

BD i SGBD

Hem de diferenciar el que és:

- Base de dades
- Sistema Gestor de bases de dades (SGBD)

BD i SGBD

Base de dades: és un conjunt de dades de diferents àmbits, organitzats sistemàticament, és a dir, seguint unes certes regles.

Sistema Gestor de Base de Dades (SGBD) o Database Manager System (DBMS) és un conjunt de programes que permeten emmagatzemar i extreure la informació d'una base de dades

[DB-Engines - rànking actual utilització DBMS](#)



Beneficis d'una BD

- Organitzar i reorganitzar la informació de manera flexible.
- Emmagatzemar grans quantitats d'informació
- Obtenir i modificar dades més ràpid
- Imprimir i distribuir la informació en diferents formats
- Estalviar temps i treball

Exemples de BBDD

- Actualment trobem bases de dades en molts àmbits:
 - Sistema de nòmines d'una empresa
 - Reserves hoteleres, companyies aèries
 - Sistema d'organització d'hospitals, biblioteques, jutjats
 - Sistema d'avaluació d'una institució escolar...



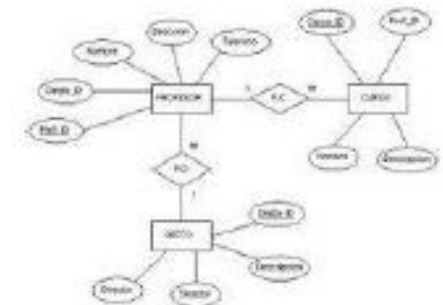
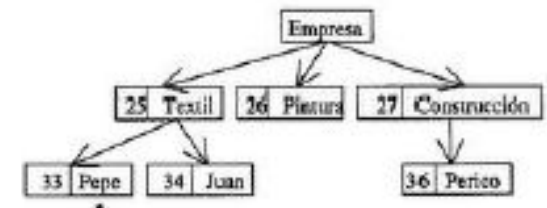
Models de Base de Dades

Les bases de dades van aparèixer per millorar les organitzacions i sobretot per millorar la recerca d'informació

Inicialment teníem models molt senzills, tabulars, jeràrquics, xarxes, etc.

Però tots ells presentaven alguns problemes

Campo	Tipo	Atributos	Nota	Predefinido	Extra
id_cliente	int(1)	UNIQUE	No		auto_increment
nombre	varchar(20)		No		
apellido1	varchar(20)		No		
apellido2	varchar(20)		No		
dni	varchar(8)		No		
telefono	varchar(10)		No		
email	varchar(50)		No		
direccion	varchar(55)		No		
usuario	varchar(10)		No		
pass	varchar(8)		No		
admin	int(1)		No		
activo	boolean("YES, NO")		No		



Model Relacional

El model de base de dades relacional va resoldre molts problemes.

- Dissenyat per IBM al 1970 (Edgar F. Codd)
- Basat en el model de **taules**
- Permet **rel·lacionar** les dades de diferents taules mitjançant “vincles” --> **descomposició de la informació en parts**
- Evita la duplicitat de la informació
- El lloc on s'emmagatzema la informació és irrellevant --> + fàcil d'entendre
- Organitzat en taules on:
 - **Columnes:** Camps
 - **Files:** Registres



Big Data: Tendència actual

Qué és BIG DATA?

"Big Data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."

Gartner, 2012

El model conceptual

“El Modelat Conceptual és l’activitat que té com objectiu obtenir i definir coneixement sobre un sistema”

El model conceptual

Què és

Cada un té un concepte diferent del
coneixement

Què és un sistema

És un conjunt d'elements relacionats entre sí i amb el seu entorn. Els sistemes poden ser naturals, filosòfics, econòmics, d'informació, etc.

Els 3 mons

- **El món real**

- Constituït pels objectes (materials o no) de la realitat que ens interessen i amb els quals haurem de treballar

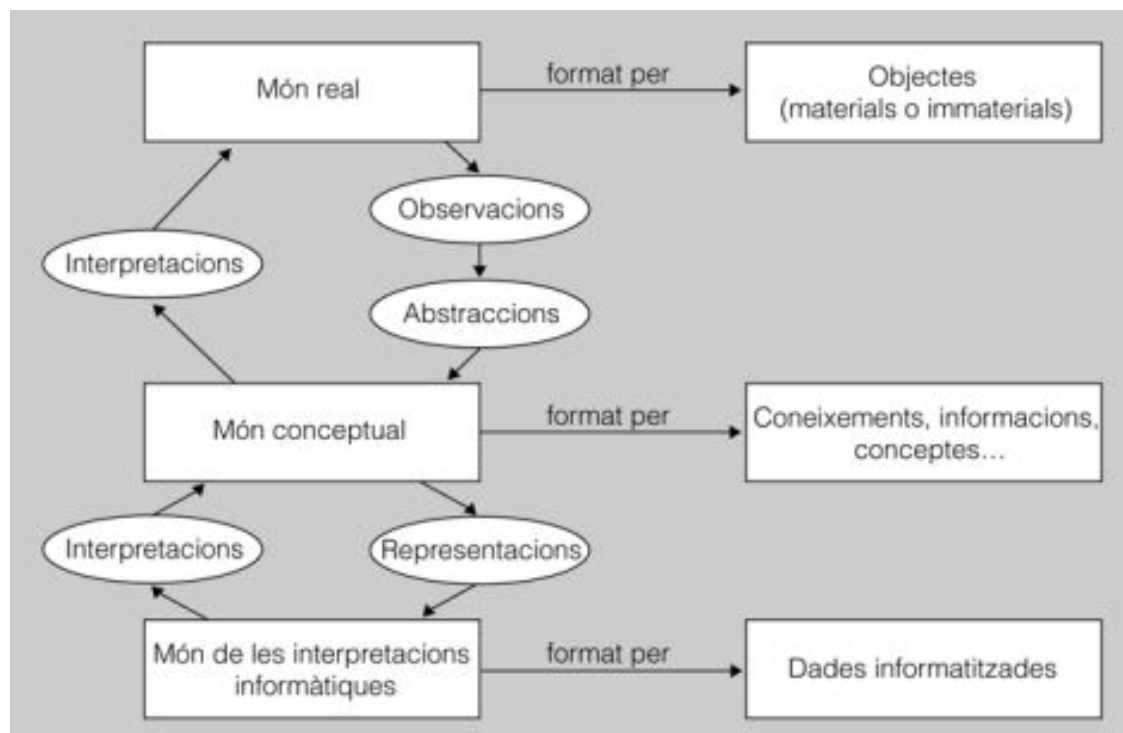
- **El món conceptual**

- És el conjunt de coneixements o informacions obtinguts gràcies a l'observació de la part del món real que ens interessa.

- **El món de les representacions**

- Està format per les representacions informàtiques, o dades, del món conceptual, necessàries per poder treballar.

Els 3 mons



Modelar un sistema

Modelar conceptualment un sistema implica la especificació d'un domini, és a dir, la **identificació de les propietats rellevants** del sistema en un determinat moment, així com el seu comportament.

Aquestes propietats rellevants són una abstracció de la realitat, segons el punt de vista del observador.

EXEMPLES

- Modelat conceptual a la interpretació del Genoma Humà (BIOINFORMÀTICA)
- Modelat d'una Empresa (la INFORMATICA aplicada al negoci, aparició dels ERP (Enterprise Resource Planning))
- Modelat d'una xarxa informàtica en un centre d'estudis com el nostre.

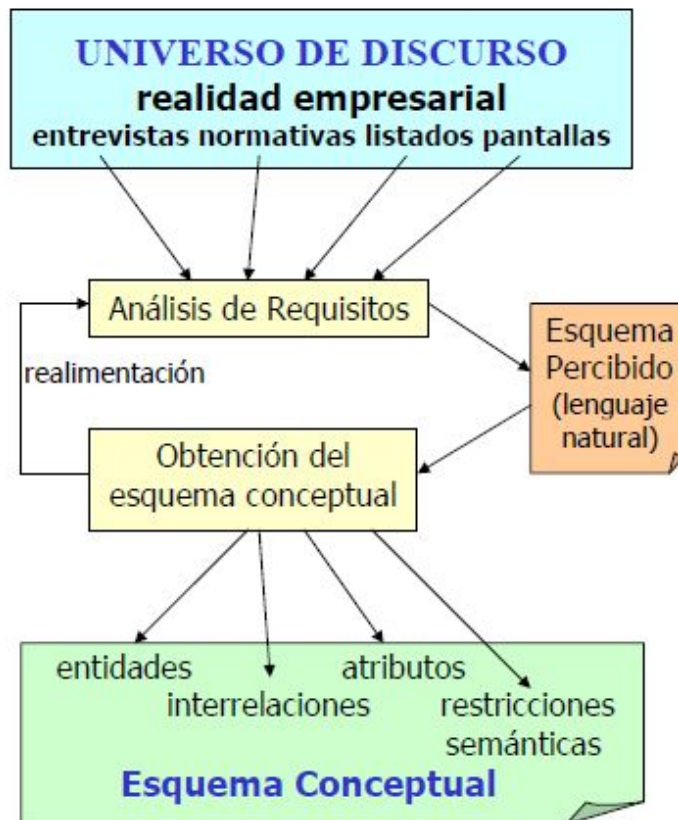
Disseny conceptual

- El modelat o disseny conceptual és la primera fase del desenvolupament d'una base de dades.
- És una activitat en la qual es construeixen esquemes conceptuais d'una realitat. Representen doncs el “món real”.
- Solen tenir una gràfica associada.
- Podem dividir-lo en dos etapes:
 - Anàlisi de requeriments
 - Generar el esquema conceptual

Disseny conceptual



Etapes del disseny



- **Entradas:**
 - Realidad empresarial: entrevistas, normativas, listados, pantallas, etc.
- **Salidas:**
 - Esquema conceptual: entidades, interrelaciones, atributos, restricciones semánticas, etc.

Anàlisi de requeriments

- **Percepció**, identificació i descripció dels elements del “mon real” a analitzar.
- Respon a la pregunta:
“Què es vol representar?”
- Es vol elaborar un **esquema descriptiu** de la realitat.
- S'utilitza un llenguatge natural per reduir problemes de comunicació entre usuaris i analistes.

Relació entre usuaris i analistes

Com veuen els usuaris als analistes

- No entenen el negoci
- Intenten dir-nos com fer la nostra feina
- No aconsegueixen de manera acceptable les especificacions del sistema
- Diuen NO als nostres suggeriments
- Sempre demanen més pressupost
- Sempre es retrasen
- No responen de manera ràpida i satisfactòria els canvis necessaris del sistema.

Com veuen els analistes als usuaris

- No saben el que volen
- Tenen moltes necessitats “polítiques”
- Ho volen tot “ja”
- No tenen prioritats entre les necessitats
- Rebutgen totes les responsabilitats sobre el sistema
- Són incapaços de respectar la planificació

Nous conceptes

- Una de las cosas més complexes és traslladar el “llenguatge natural” a un “llenguatge conceptual”.
- Nosaltres parlem de persones, de coses, de dades concretes (un número de telèfon, una marca de cotxe), però el llenguatge conceptual està ple d'Objectes, Entitats i Atributs.

Definicions

- Una **entitat** és un objecte de dades que té unes propietats. Els noms solen ser entitats. Ex:
 - A la universitat existeixen professors que tenen nom, cognom, dni, adreça,... PROFESSORS és una entitat
 - Quan s'ha d'esborrar un estudiant d'una BD. ESTUDIANTS és una entitat
- Un **atribut** és un objecte de dades que representa una propietat. Ex:
 - Es pot consultar el nom d'un professor.

Exemple1

Una base de dades per a un **zoo**

Tenim “**objectes**”, en el món real: animals, persones (els treballadors, els visitants....), espais, ... etc

Alguns d'aquests objectes els podem fer correspondre amb “**entitats**” per a la base de dades: per exemple, l'entitat treballadors

Cada entitat té unes propietats “**atributs**”. En el cas de l'entitat treballadors podem pensar com a atributs:

nom, cognom, lloc_de_treball, edat, any_de_contractació...

El **domini** és el conjunt de valors possibles que pot prendre l'atribut.
Ex: Si el zoo va obrir al 2007 no pot haver valors anteriors.

Exemple1

Una **entitat tipus**, és un tipus genèric d'entitat o d'abstracció que fa referència a una classe de coses, per exemple de cotxes.

en el moment en que els atributs d'una entitat prenen un valor concret es converteix en una **entitat instància**.

Ex:

ANIMAL (nom, família, categoria, any_arribada...)

Si busquem un animal concret, tenim una entitat-instància.

Maguila, Gorila, Homínido, 2003,...

Característiques

- Els atributs poden tenir valors nuls, excepte si són identificadors o clau.
- A vegades costa distingir entre un atribut identificador i un atribut clau.
- Tot atribut identificador, és un candidat a ser clau, però una clau formada per diferents atributs, no té perquè estar formada per atributs identificadors

Ex: Imagina una entitat ALUMNE. El DNI identifica i és clau.

Imagina una entitat NOTES on es guarda la nota de cada alumne en cada matèria. El DNI ni la matèria, en aquest cas, ni tan sols identifica, ja que una entitat-instància pot tenir el mateix DNI (un alumne matriculat en dues matèries). En canvi, la combinació DNI-Materia-Curs pot ser una clau.

Exercici 1

Volem fer el model conceptual per crear una base de dades que per tal de tenir emmagatzemades les dades d'un institut.

- *- Pensa en 3 entitats i en 3 atributs per a cada una de les entitats.
- *- Intenta també pensar en 5 preguntes que li faries al client, relacionades amb la base de dades que voleu crear.

Exercici 2

Volem fer el model conceptual per crear una base de dades que controli un **aeroport** que es dedica a vols comercials.

Redacta 5 preguntes que li podries fer al client en un qüestionari.

Pensa en 3 objectes que poguessis convertir en entitats.

Pensa en 5 atributs per a cada objecte. Identifica cada un i dona una clau per a cada entitat.

Conceptes que han de quedar clars sobre el model conceptual

Què és el món conceptual

Dades /Informació

Entitats

Objectes

Atributs

Valors dels atributs

Domini

Tipus d'entitats: Entitat tipus i entitat instància

Motivació de les BD

Imaginem que volem crear una empresa online de venda d'articles artesanals.

Volem registrar tots els articles, anotar vendes i compres, actualitzar stocks, extreure informes i/o realitzar consultes com ara:

- * Quin va ser l'article més venut al Febrer?
- * Quins són els beneficis obtinguts el darrer mes?

Entitats i taules

- Penseu que quan volem guardar dades, el més fàcil és una distribució tabular on:
 - cada entitat instància s'anomena **registre** (una fila de la taula)
 - cada atribut s'anomena **camp** (columna de la taula)
 - Cada intersecció d'un registre i d'un camp emmagatzema el **valor** que tingui el camp del registre de què es tracti.

Entitat tipus
ALUMNE

Atributs

NOM	COGNOM	DNI	E-MAIL
David	Gómez	65765568G	David.G34@hotmail.com
Sandra	Girau	45767672D	Sandrita21@gmail.com

Entitat instància

Valor

Propòsit general

Gestió ineficient i caòtica



Base de dades



Gestió eficient i eficaç







MIF-NEW EM



C01-381-C

amazon

The Amazon logo, featuring the word "amazon" in a bold, black, sans-serif font. A yellow curved arrow, resembling a smile, is positioned below the letters "a" and "z", pointing from the "a" to the "z".



Propietats Big Data

