

INTRODUCCIÓ A LES BASES DE DADES



**Institut Rafael
Campalans**

Plaça del Remei, 1
17160 Anglès

Cicle: DAM

Curs: 2022/2023

Mòdul: 02 Bases de Dades

Objectius



- Identifica els diferents elements, objectes i estructures d'emmagatzematge físic disponibles en un SGBD corporatiu • Identifica l'estructura d'un diccionari de dades.
- Diferencia entre el nivell intern, el nivell conceptual i el nivell físic d'una base de dades.
- Diferencia entre els diferents models de bases de dades.
- Identifica les bases de dades distribuïdes: utilitat, diferències, avantatges i inconvenients, distribució de les dades, arquitectura, seguretat i recuperació.
- Identifica les bases de dades centralitzades i les bases de dades distribuïdes: utilitat, diferències, avantatges i inconvenients.

Introducció

- La feina d'informàtic és gestionar dades. Quan aquestes dades son molt nombroses necessitem fer programes informàtics que ens ajudin a automatitzar el treball sobre aquestes dades.
- En aquest curs veurem la gestió de dades des d'un punt de vist acadèmic. Les necessitats de cada projecte podran fer que alguns de les coses que veiem s'hagin d'adaptar.
- El BigData y els analistes de dades es una de les professions de futur. La base de com ser-ho ho treballarem en aquest curs.

Què és una dada?

- Una **dada** és la representació ,a través d' una seqüència de símbols, nombres o lletres, d'una variable que pot ser **quantitativa o qualitativa** i indica un valor que se li assigna a les coses.
- Les dades descriuen **fets empírics**. Per examinar-los han de ser organitzats o tabulats, ja que una dada per si mateixa no pot demostrar massa sinó que s' ha d' avaluar el conjunt per examinar els resultats.
- Per exemple:

40 **Edat**

IMPORTANT



Diferència entre dada i informació

- El concepte de dada està estretament lligada al d'informació, però hi ha una diferència fonamental entre ambdós termes. Mentre **les dades refereixen a esdeveniments o fets registrats**, la **informació** està constituïda per aquelles dades en brut que són processades de manera que **generin contingut que pugui ser interpretat pels usuaris**.
- **Les dades no tenen sentit per si mateixes**, però en ser processades i contextualitzades es converteixen **en informació** certa i disponible per conèixer un fenomen, prendre decisions o executar accions.



Pregunta: Què ens interessa tenir dades o informació?

Exemples de dades

- **Exemples de dades qualitatives:**

Nom: Francesc Barragan

Ocupació: Docent

Domicili: Avinguda Barcelona

Gènere: Masculí

- **Exemples de dades quantitatives:**

Edat: 47

Alçada: 1.84

Nombre de fills: 1

Anys en el càrrec: 2



Entitats, atributs i valors

Elements que caracteritzen les informacions:

1. **Les entitats** són els objectes del món real que conceptualitzem. Són identificables, és a dir, distingibles els uns dels altres. I ens interessen algunes (com a mínim una) de les seves propietats.
2. **Els atributs** són les propietats de les entitats que ens interessen.
3. **Els valors** són els continguts concrets dels atributs, les determinacions concretes que assoleixen. Potser que no tingui un contingut concret i és llavors quan parlem de valors **NUL**.

Exercici

- **Exemples de entitat, atributs i valors**

Considerarem que **una pel·lícula concreta és una entitat**, perquè és un objecte del món real, que hem conceptualitzat dins d'una categoria (la dels films cinematogràfics), i que al mateix temps és distingible d'altres entitats de la mateixa categoria (és a dir, d'altres films).

D'aquesta pel·lícula ens interessaran alguns aspectes, que anomenarem **atributs**, com per exemple, **el títol, el director i l'any de producció**.

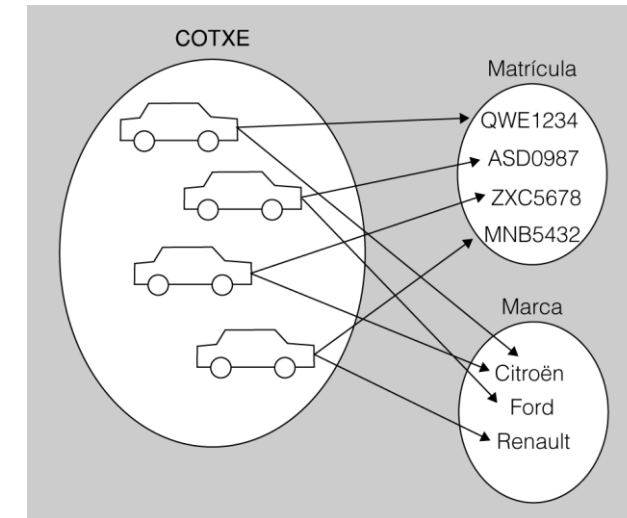
Finalment, aquests atributs adoptaran uns **valors concrets** com ara, i respectivament, Minions El Origen de Gru, Kyle Balda,.. i 2022.



**Podeu
posar més
exemples?**

Representació de les dades: tabular

- Ja sabem que les dades són **informacions representades informàticament**. Per tant, també podríem anomenar *món de les dades* el món de les representacions.
- La representació informàtica més freqüent en l'àmbit de les BD és la **representació tabular**, la qual s'implementa habitualment en fitxers que s'estructuren en **registres i camps**.



Cotxes	
Matrícula	Marca
QWE1234	Citroën
ASD0987	Ford
ZXC5678	Citroën
MNB5432	Renault

Què és una base de dades?

- Conjunt estructurat de dades que representa entitats (fets o ocurrencies del món real) i les seves interrelacions (comparteixen algun vincle entre elles).
- No necessàriament el suport ha de tenir suport informàtic.

IMPORTANT

- Exemples:

Guia telefònica
Biblioteca



Concepte i definició de base de dades

- Actualment, la majoria de base de dades estan en suport digital i són utilitzades per gran part d'aplicacions informàtiques.

Examples:

Aplicacions que utilitzen BD:

WhatsApp

Facebook

Amazon

Cercador de Google

Empreses de tot tipus:

Instituts

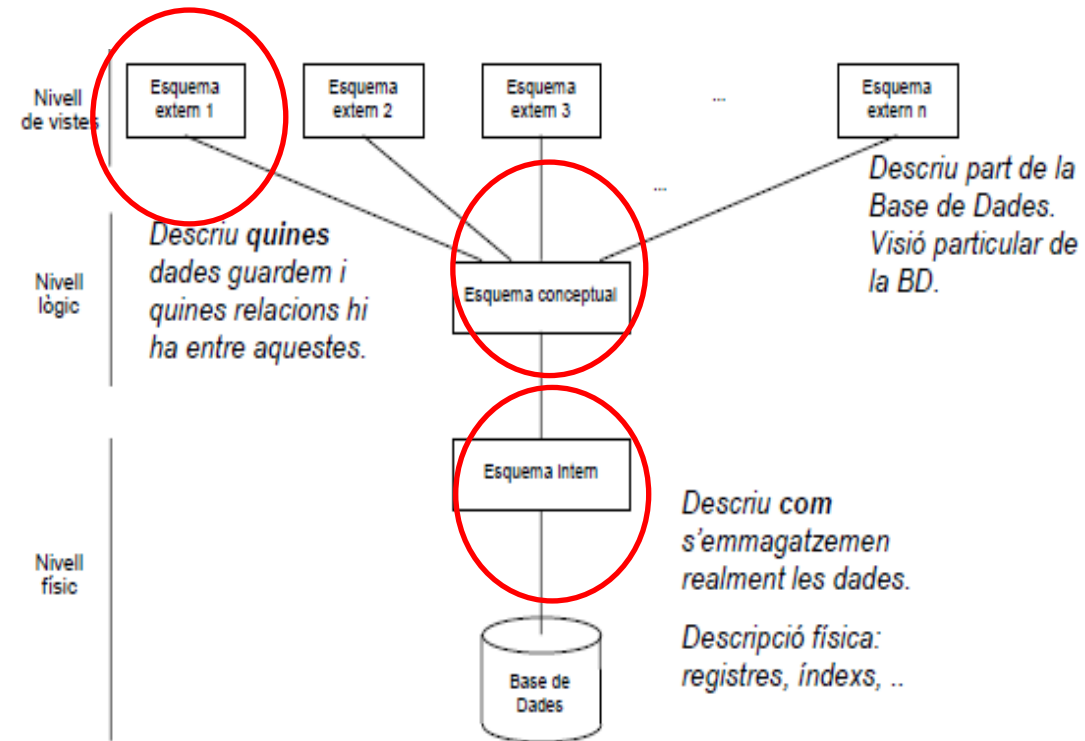
Universitats

Gestors de continguts: Wordpress, Joomla, Drupal, Prestashop



Arquitectura d'un SGBD (ANSI-SPARC)

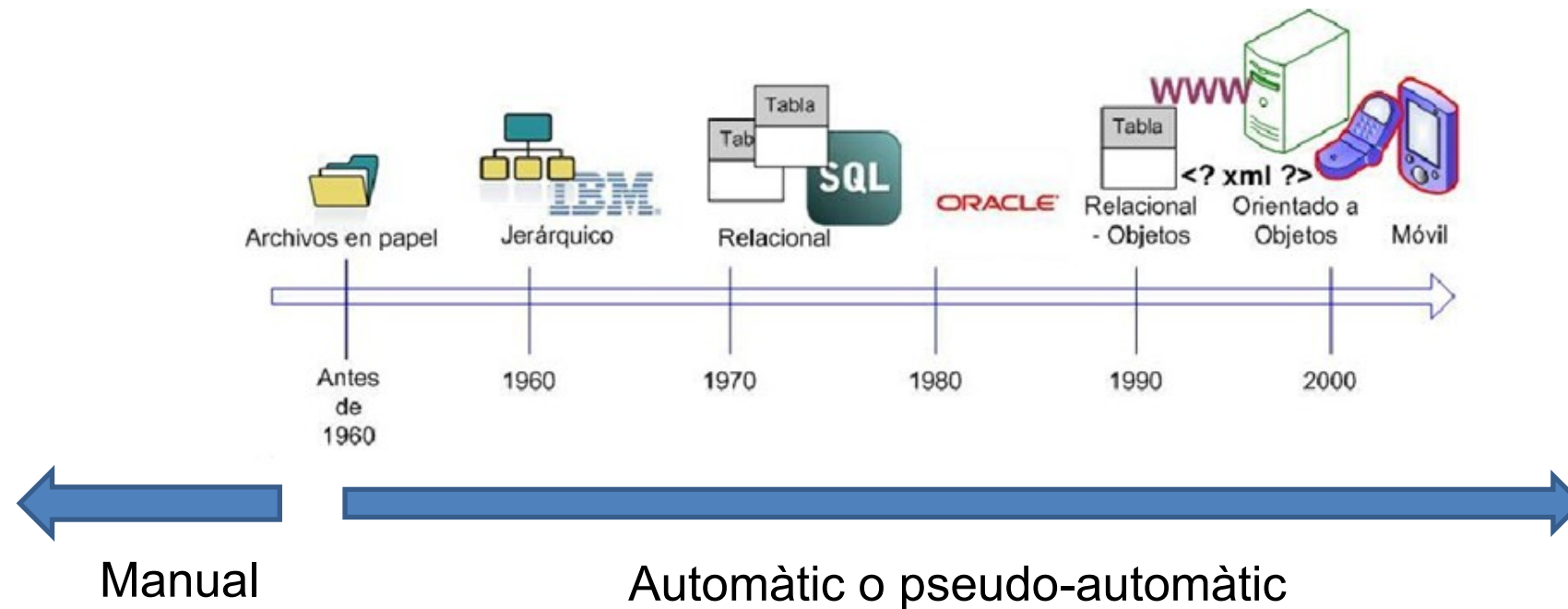
El 1975, el comitè ANSI/X3/SPARC va proposar una arquitectura per als SGBD estructurada en tres nivells d'abstracció (**intern, conceptual i extern**), que resulta molt útil per separar els programes d'aplicació de la BD considerada des d'un punt de vista físic.



https://es.wikipedia.org/wiki/Arquitectura_ANSI-SPARC

Evolució de les Bases de Dades

- La història de les bases de dades comença, quan s'utilitzà aquest terme per primera vegada l'any 1963 als Estats Units; no obstant, segles abans ja s'aplicava la base de dades per recol·lectar informació amb la diferencia que es feia de **manera manual**.



Evolució de les Bases de Dades(II)

- **Anys 50** s'utilitzaven **fitxers**:
 - Una aplicació treballava amb pocs fitxers per generar un nou fitxer (processos batch)

Exemple: entitat bancària:

No Integritat
No Accés concurrent
No Seguretat
Poca flexibilitat (calia reescriure les aplicacions si hi havia canvi en les dades, en les consultes...)
...

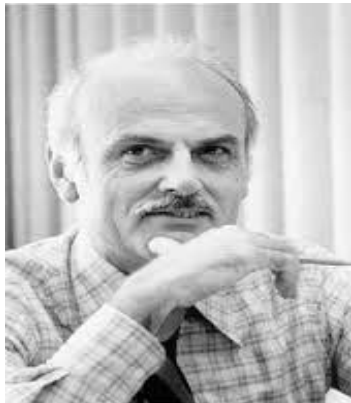
Redundància!

CompteCorrent	Número_compte DNI Nom_client Cognoms_client Adreça_client Telèfon Saldo_actual Ingressos
Préstecs	Número_préstec Número_compte DNI Nom_client Cognoms_client Adreça_client Telèfon Data_alta_préstec Data_venciment

Evolució de les bases de dades(III)

➤ Anys 60

- Sistemes centralitzats: gran ordinador (mainframes) i terminals “tontos”
- Van aparèixer les primeres bases de dades basades en el **Model de dades jeràrquic i en xarxa**.
- ✎ • Les **BD jeràrquiques** emmagatzemen la informació en una estructura jeràrquica que podem imaginar amb una forma d'arbre invertit, on cada node pare pot tenir diferents fills. El node superior, que no té pare, es coneix com a *arrel*. I els nodes que no tenen fills s'anomenen *fulles*.
- ✎ • Les **BD en xarxa**, semblant al model jeràrquic, amb registres interrelacionats mitjançant una estructura en forma d'arbre invertit, però més flexible, ja que permetia que els nodes tinguessin més d'un sol pare.



➤ Anys 1970-72

- Edgar Codd (IBM) proposa el **Model de Dades Relacional**. es basa en la lògica de predicats i en la teoria de conjunts (àrees de la lògica i de les matemàtiques). Actualment, és el sistema més àmpliament utilitzat per modelitzar dades

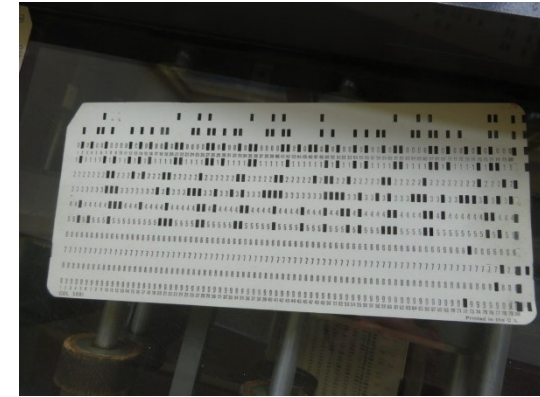


Evolució de les bases de dades (IV)

➤ Anys 1970:

El programari de gestió de fitxers era massa senzill:

- Tractament d'interrelacions no estava suportat.
- Varis usuaris no podien actualitzar dades a l'hora.



Durant la 2^a meitat dels anys 70 sorgeix un programari més sofisticat: els **DataBase Management System** (DBMS) o **Sistemes Gestors de Bases de Dades** (SGBD) que permetia resoldre aquestes limitacions.



Evolució dels SGBD (V)

➤ Anys 80:

- Es consolida el model relacional com la base teòrica
- dels SGBDs Comercials. (Oracle va néixer finals dels 1970)



➤ Finals anys 80:

- Estandardització del llenguatge SQL (*Structured Query Language*) per la manipulació dels SGBD Relacionals

➤ Anys 90:

- Els fabricants de SGBD consoliden els seus productes i proporcionen extensions pel suport de nous tipus de dades (imatges, multimèdia,...) SGBD són el pilar de molts sistemes Client / Servidor a través d'Internet

Evolució dels SGBD (VI)

➤ Anys 90



- SGBD Relacionals són part fonamental de sistemes de gestió empresarial i de la majoria d'aplicacions.
- Nous paradigmes: Objecte-Relacionals, Orientació a Objectes, XML, documentals, Datawarehouse, Geolocalització...

➤ Panorama actual

- Tot i que a dia d'avui els sistemes relacionals encara són els que predominen en el mercat i la gran majoria d'aplicacions els utilitzen, en la última dècada estan prenent força un nou tipus de **base de dades anomenades NoSQL** (MongoDB, Hadoop/Hbase, Cassandra, CouchDB, Riak...)
- Capacitat de tractar grans volums de dades de forma òptima
- Escalabilitat
- Exemple: Facebook, Amazon, Google...
- Un altre paradigma que ha aparegut recentment es el cloud computing, on tenim no fixed costs of ownership data and software
- BigData



Evolució dels SGBD (VII)

➤ Panorama actual NoSQL Databases:

El terme NoSQL sobre l'any 1998 s'utilitzava per referenciar a bases de dades relacionals de codi obert que no utilitzava llenguatge de consulta SQL. A partir del 2009 s'utilitza aquest terme per bases de dades distribuïdes de codi obert no relacionals.

Dintre de totes les implementacions actuals de base de dades NoSQL es troben algunes que no utilitzen SQL com MongoDB que utilitza JSON, per hi ha altres que continuen utilitzant-ho com BigTable(GQL)

El terme NoSQL en bases de dades, intenta resoldre les limitacions dels models relacionals en qüestió d'emmagatzemament massiu de dades on es necessita tenir servidors molts potents i amb carrega balancejada.

Evolució dels SGBD (VIII)

Exemples de NoSQL Databases:

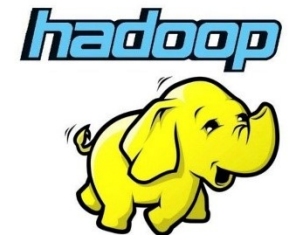
MongoDB

- És la base de dades NoSQL líder i permet a empreses ser més àgils i escalables.
- Ha estat creada per brindar rendiment i gran disponibilitat.
- Està orientada a documents, els quals són emmagatzemats en BSON, que és una representació binària de JSON.
- Utilitzat per companyies de contes com SourceForge, foursquare, MTV, Disney, IGN, Guardian, NYTimes, Doodle.



Hadoop/Hbase

- Hadoop és un framework que permet el processament de grans volums de dades a través de clusters.
- És un sistema distribuït que utilitza una arquitectura Master-Slave, usant per emmagatzemar el seu Hadoop Distribuite File System (HDFS) i algoritmes de MapReduce per fer càlculs.
- HBase és la base de dades de Hadoop, la qual no admet SQL i no segueix l' esquema relacional. La utilitzen Amazon, Adobe, AOL, Ebay, Facebook, IBM, New York Time, Microsoft, Twitter, Yahoo.



Evolució dels SGBD (IX)

➤ Exemples de NoSQL Databases:

Cassandra.

- És una base de dades de codi obert la principal característica de la qual és que fusiona Dyaname, d'Amazon amb BigTable, de Google.
- Permet resoldre la problemàtica relacionada amb el rendiment del motor de recerca.
- Es va dissenyar perquè les configuracions d' explotació fossin altament escalables, horitzontals i econòmiques.
- Entre les empreses que les fan servir hi ha Facebook, WebEx, Symantec, IBM, Netflix, Cisco, Twitter.



Evolució dels SGBD (X)

EL BIGDATA

S'anomena **Big Data**, dades massives o macrodades, als conjunts d'informació el volum de les quals sobrepassa les capacitats de les eines convencionals de processament de dades.

També es refereix a la recopilació de totes aquestes dades i la nostra capacitat d'utilitzar-les al nostre favor en una àmplia gamma d'àrees, des dels negocis fins a la política.



Resum Models de Base de dades

En resum **els models de dades més utilitzats al llarg del temps** han estat els següents, exposats en ordre d'aparició:

1. Model Jeràrquic
2. Model En xarxa
3. Model Relacional
4. Model Orientat a Objectes
5. **Model amb UML**



Pregunta

- Perquè creieu que van aparèixer les bases de dades i la necessitat d'organitzar la informació?



Sistemes gestors de base de dades

- Per poder treballar amb la informació continguda en una base de dades (crear la seva estructura, afegir, modificar, eliminar dades...consultar dades) hi ha aplicacions especialitzades en aquesta tasca anomenats Sistemes Gestors de Base de Dades (SGBD) o Database Management System (DBMS) en anglès.
- Més formalment: És una col·lecció de dades interrelacionades i un conjunt de programes per gestionar-les.



Objectius dels SGBD

- ❖ Proporcionar una forma d'emmagatzemar i recuperar la informació d'una Base de Dades de manera pràctica i eficient.
- ❖ Oferir als usuaris una visió abstracte de les dades. El sistema oculta certs detalls en la forma en que s'emmagatzemen i es mantenen les dades.
- ❖ Facilitar als desenvolupadors la creació i manipulació d'informació.



Avantatges d'utilitzar un SGBD

- 1 Realitzar consultes no predefinides i complexes**

Proporcionen mecanismes per manipular dades (alta, baixa, modificació i consulta) i definir-ne de noves.
- 2 La independència física de les dades**

Un canvi sobre l'esquema físic no afecta ni als esquemes lògics ni als externs.
- 3 Minimitzar la redundància de dades**

Definim que volem guardar a nivell lògic i el SGBD ens ajuda a definir com s'acaba guardant a nivell físic. Tria la forma d'enregistrar les dades com el manteniment de les relacions, entre aquestes, sovint van orientades a minimitzar la redundància (característica no desitjable).
- 4 Vetllar per la integritat de les dades**

Podem definir regles semàntiques sobre determinades dades, el SGBD exigeix en tot moment el compliment d'aquestes per mantenir la integritat de les dades.
- 5 La independència lògica de les dades**

Un canvi sobre l'esquema lògic no afecta ni als esquemes físics ni als externs.
- 6 Accés concurrent**

Implementa diferents mecanismes per facilitar l'accés concurrent a les dades (múltiples usuaris treballant sobre la mateixa BD).
- 7 Seguretat**

Proporciona mecanismes per la definició d'usuaris i permisos sobre la BD. L'encarregat de gestionar aquests permisos és un usuari especial anomenat Administrador de la BD. Encriptació

Propietats ACID

ACID (Atomicity, Consistency, Isolation, Durability)

Atomicity

És l'habilitat d'un SGBD de garantir que són executades totes les tasques que intervenen en una transacció, o bé no se n'executa cap.

Consistency

La consistència es refereix al fet que la base de dades ha d'estar en un estat quan acaba una transacció que ha de ser coherent amb l'estat que tenia en començar la transacció.

Isolation

L'aïllament es refereix a l'habilitat de les aplicacions de fer que les operacions en una transacció estiguin aïllades de totes les altres operacions.

Durability

La definitivitat o durabilitat es refereix a la garantia que un cop un usuari ha estat notificat de l'èxit d'una transacció, aquesta persistirà, no pot ser desfeta.

Llenguatges

SQL (Structured Query Language) és el llenguatge més utilitzat per interaccionar amb un SGBD relacional. Està compost de dos llenguatges:

- Llenguatges de definició de dades (DDL): per la definició de la estructura de la BBDD.
- Llenguatges de manipulació de dades (DML): per la consulta, inserció ,modificació i eliminació de les dades a la bd.

A part existeixen llenguatges de quarta generació que permet el accés a les dades d'un SQL

- Llenguatges 4GL (C#, Python, Java...)
- Interfícies visuals
- Llenguatge hostatjat (precompiladors: C, COBOL)



Usuaris

Administradors de base de dades

Són uns usuaris especials que realitzen tasques d'administració i control centralitzat de les dades, i gestionen els permisos d'accés concedits als diferents usuaris i grups d'usuaris, per tal de garantir el funcionament correcte de la BD.



Programadors d'aplicacions

Són professionals informàtics que creen els programes que accedeixen als SGBD i que, posteriorment, són utilitzats pels usuaris que hem anomenat *externs*.



Usuaris especialitzats

Interactuen directament amb el sistema, sense utilitzar les interfícies proporcionades per programes intermediaris. Formulen les consultes en un llenguatge de BD (normalment, SQL)



Usuaris d'aplicacions

Són usuaris, que no interactuen directament amb el sistema, sinó mitjançant alguna aplicació informàtica desenvolupada prèviament per altres persones amb aquesta finalitat.



Diccionari de Dades: Introducció

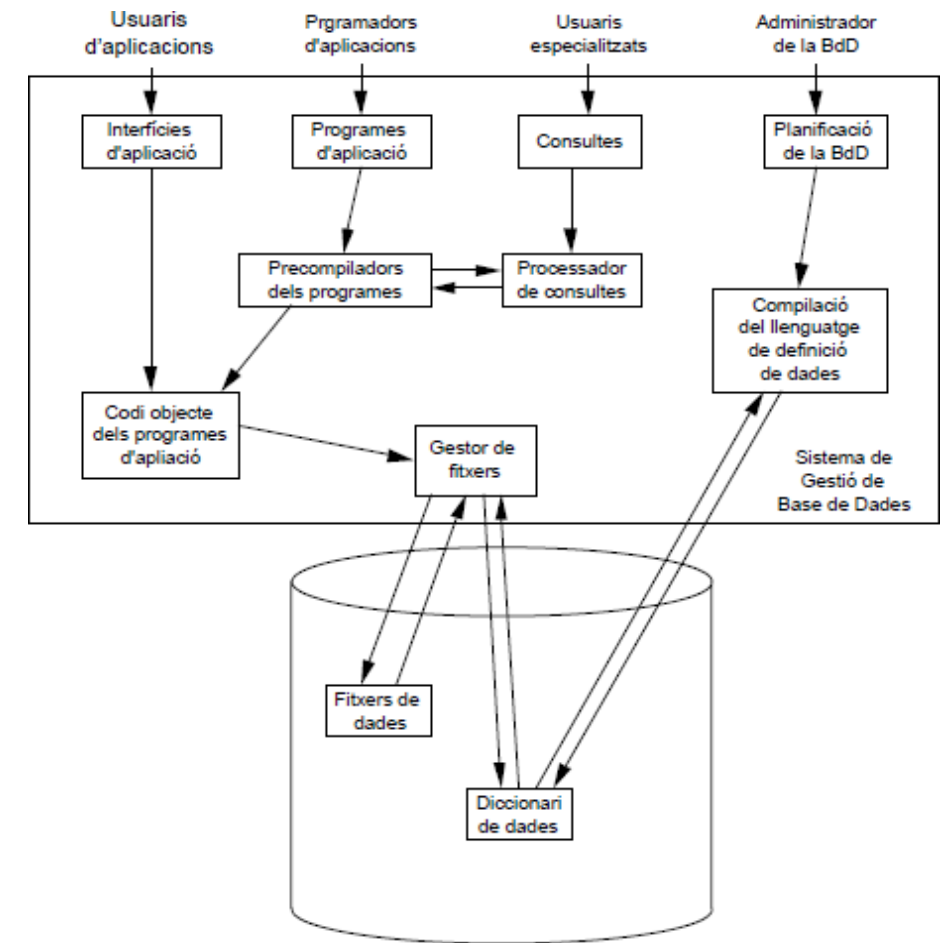
- Un diccionari de dades d'una base de dades és el conjunt de metadades que proporcionen informació sobre el contingut i l'organització de la base de dades.
- Un diccionari de dades, segons IBM Dictionary of Computing es pot definir com un “repositori centralitzat d'informació sobre dades com ara el significat, relacions amb altres dades, origen, ús i format.”
- De vegades el concepte de diccionari de dades o metadades també és conegut amb el nom de catàleg del sistema o, també, com a repositoris de metadades.

Components funcionals d'un SGBD

1).- Gestor d'emmagatzemament

- **Gestor d'autoritzacions i d'integritat.** Comprova que se satisfacin tant les restriccions d'integritat com les autoritzacions dels usuaris per accedir a les dades.
- **Gestor de transaccions.** Assegura que la BD es mantingui en un estat de consistència malgrat les fallades del sistema, i també que les transaccions concurrents no s'interfereixin entre elles.
- **Gestor d'arxius.** Gestiona la reserva d'espai d'emmagatzemament en disc i les estructures de dades utilitzades per representar la informació emmagatzemada en disc.
- **Gestor de memòria intermèdia.** Transfereix les dades des del disc a la memòria principal, i decideix quines dades s'han de tractar en memòria cau. Permet al sistema tractar amb dades de grandària molt superior a la memòria principal.

A part utilitza unes estructures de dades **com arxius de dades, diccionari de dades i índexs.**



2).- Processador de consultes

Diccionari de Dades: Elements

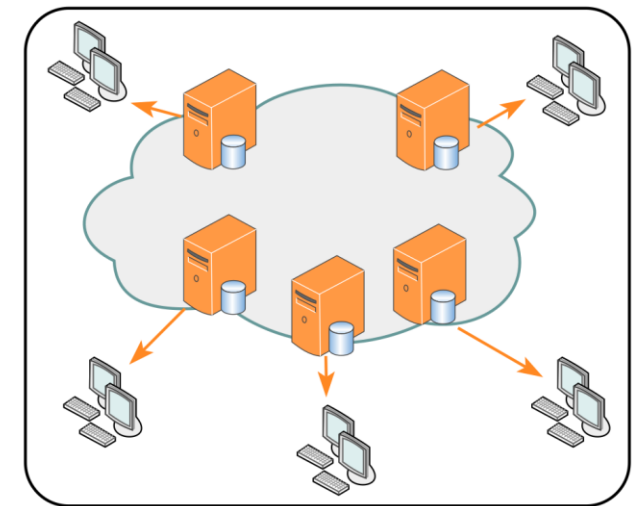
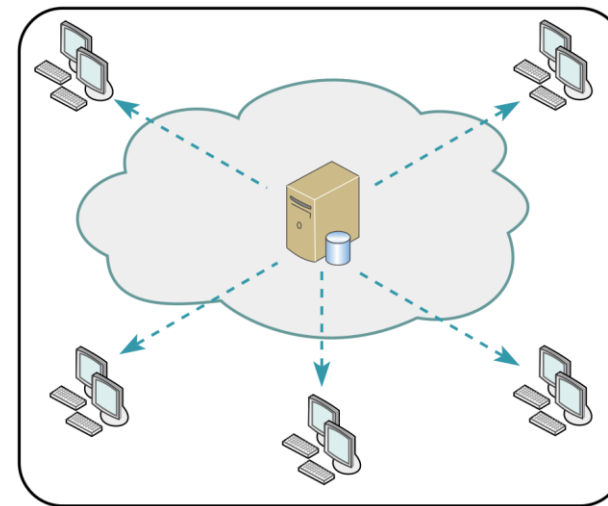
- Els elements que es troben habitualment a un diccionari de dades inclouen:
 - Definicions de l'esquema de la base de dades.
 - Descripcions detallades de taules i camps, així com de tots els objectes de la base de dades (vistes, clústers, índexs, sinònims, funcions i procediments, triggers, etc.).
 - Restriccions d'integritat referencial.
 - Informació de control d'accés, com ara noms d'usuari, rols, i privilegis.
 - Paràmetres d'ubicació de l'emmagatzemament.
 - Estadístiques d'ús de la base de dades.

Bases de dades distribuïdes

Un dels sectors informàtics on més s'està evolucionant darrerament, tot integrant el desenvolupament tecnològic amb la innovació metodològica, és el relatiu als sistemes distribuïts d'informació.

Podem tenir tres tipus diferents d'arquitectures de BD:

- 1) Centralitzades
- 2) Client-servidor
 - Servidors de dades
 - Servidors de consultes
- 3) Descentralitzades
 - 1) Sistemes paral·lels
 - 2) Sistemes distribuïts



CDN Network

Pregunta

- Perquè creieu que van aparèixer les CDN?



Bases de dades distribuïdes

- Per molt dimensionat que tinguem un servidor si anem afegint usuaris arribarà un moment que per aspectes físics no podrà donar servei. És per això que s'utilitzen tècniques de BD distribuïdes com:
 - Multiplicació.
 - Divisió.
 - Distribució amb node principal.
 - Distribució amb duplicacions en nodes seleccionats.
- Que permeten distribuir la càrrega de feina entre més nodes.

Activitat A01

Conèixer diferents SGBD



- Descarrega't la pràctica A01 del moodle del institut
- Temps 60m.
- Recorda:
 - Utilitza el model de plantilla per fer la entrega
 - Entrega en format PDF
 - Nom del Fitxer afegiu la vostra inicial del nom+cognom.

WEBGRAFIA

- Batini, C.; Ceri, S.; Navathe, S.B. (1992). Conceptual Database Design: An Entity-Relationship Approach. Reading, Massachusetts: Addison Wesley.
- Teorey, T.J. (1999). Database Modeling & Design. The Fundamental Principles (3a ed.). San Francisco: Morgan Kaufmann Publishers, Inc.