

Introducció

Bases de Dades

Marta Tarrés-Puertas

Enginyeria de Sistemes TIC
Universitat Politècnica de Catalunya
<http://itic.cat>

26 de febrer de 2015

- 1 Conceptes introductoris
- 2 Motivació
- 3 Els Sistemes Gestors de Bases de Dades
- 4 Models de dades
- 5 Aspectes històrics

- Analista / Dissenyador de sistemes d'informació
- Els tres móns
 - Món real: Objectes. Observació i abstracció.
 - Món conceptual: Informació. Disseny.
 - Món de les representacions informàtiques: Dades
- Món Conceptual
 - Informació = entitat + atribut + valor + temps
 - Domini d'un atribut
 - valor nul
 - identificador, clau
 - instància, entitat genèrica
- Món de les Representacions
 - Dada
 - Representació tabular
 - Fitxers de dades. Registre, camp

- Accés a les dades: Seqüencial per Posició, Directe per Posició, Seqüencial per Valor, Directe per Valor
- Diferències fitxers/BD
 - Entitats tipus
 - Interrelacions
 - Redundància
 - Usuaris

- Top Largest Databases in the World (Dades 2010):
<http://www.comparebusinessproducts.com/fyi/10-largest-databases-in-the-world>
- <http://en.wikipedia.org/wiki/Petabyte>

- Bancs: Transaccions
- Aerolínies: reserves, programacions de vols
- Universitats: registre, graus
- Vendes: clients, productes, compres i ventes
- Vendes on-line: seguiment de comandes, recomanacions personalitzades
- Fabricació: producció, inventari, encàrregs, logística
- Recursos humans: registre d'empleats, salaris, deduccions fiscals
- Telecomunicacions: registre de trucades, generació de factures telefòniques, ...

- Els DBMS contenen informació empresarial
 - Col·lecció d'informació interrelacionada
 - Conjunt de programes per tal d'accedir a la informació
 - Un entorn amigable i eficient d'utilitzar (característiques desitjables)
- Les Bases de dades emmagatzemen gran quantitat d'informació

Com gestionar les dades? Fitxers?

Directament en els fitxers ...
No n'hi ha prou. Veiem-ne els inconvenients

Inconvenients de l'us de sistemes de fitxers per emmagatzemar informació

- Inconsistència i redundància de dades
 - Necessitat d'escriure un nou programa per tal de dur a terme cada nova tasca
- Dificultat en l'accés a la informació
- Aïllament de les dades - varietat d'arxius i formats
- Problemes d'integritat
 - Comprovació que un compte no estigui en números vermells
 - Difícil afegir noves restriccions o canviar-ne d'existents

Inconvenients de l'ús de sistemes de fitxers per emmagatzemar informació

- Atomicitat dels canvis
 - Inconsistència a fallades del sistema davant modificacions parcials dutes a terme. Per exemple: Transferència de comptes d'un origen a un destí cal que es completi totalment o no es realitzi.
- Concurrència (accés per diversos usuaris a la vegada)
 - La concurrència és necessària per garantir el rendiment
 - L'accés incontrolat duu a inconsistències. Per exemple: 2 persones llegint un balance de 100 i modificant-lo retirant diners (50) cadascú al mateix temps.
- Problemes de seguretat
 - Difícil de donar accés parcial a la informació

Els Sistemes Gestors de Bases de Dades

Objectius i serveis dels DBMS

- 1 Consultes no predefinides i complexes
- 2 Flexibilitat i independència
- 3 Problemes de la redundància
- 4 Integritat de les dades
- 5 Concurrència d'usuaris
- 6 Seguretat
- 7 Nous objectius adaptats a les noves necessitats i les noves tecnologies

Què és una BD?

És una col·lecció d'informació interrelacionada i un conjunt de programes que permet als usuaris accedir a la informació i modificar-la.

Permet als usuaris una visió abstracta de la informació. És a dir, el sistema amaga detalls de com la informació s'emmagatzema i es gestiona.

- Models de Dades
- Esquemes

Una BD és una representació de la realitat (Model de la realitat). Tot model de BD proporciona les següents eines:

- Estructures de dades amb les quals es pot construir la BD
- Diferents tipus de restriccions (o regles) d'integritat que el DBMS haurà de fer complir a les dades: dominis, claus, etc.
- Tot un seguit d'operacions per treballar amb les dades

Col·lecció d'eines conceptuals per descriure la informació, els relacions, la semàntica, la consistència i les restriccions. Proveeix la manera de descriure el disseny d'una BD en el nivell físic, lògic i de vista.

Els models de BD més utilitzats en Sistemes d'Informació

- Model relacional (SQL)
- Model Entitat-Relació (Disseny de BD)
- Model relacional amb objectes (Object-oriented and Object-relational)
- Model semiestructurat (XML)
- Models obsolets
 - Model en xarxa
 - Model jeràrquic

La Taula Professorat

codiProducte	nomProducte	tipus
descripcio	preu	
22222	Einstein	Física
11111	Mozart	Música

- Taules = relacions = associacions entre entitats
- Atributs = Capçaleres de les columnes
- Tuples = Files
- Relació = Membres

- Esquema= Nom de la relació + llista atributs
 - Membres(Id, nom, dept) o bé Membres(id: int, nom: string, dept: string)
- Base de Dades = col·lecció de relacions
- Esquema de la BD = conjunt del tots els esquema de les relacions de la BD

Exemple BD Relacional

La Taula Professorat

Id	Nom	Dept
22222	Einstein	Física
11111	Mozart	Música

La Taula Departament

Dept	Edifici	Planta
Física	MN2	1
Música	MN1	2

- Data Definition Language (DDL): Permet especificar l'esquema de la BD
- Data Manipulation Language (DML): Permet realitzar consultes i modificacions

SQL = Data Definition Language + Data Manipulation Language

Definició de l'esquema + propietats addicionals de la informació

Exemple per la definició de l'esquema de la BD

```
create table professor(  
  id char(5),  
  nom varchar(10),  
  dept varchar(10))
```

Els valors d'una BD han de satisfer restriccions de consistència. Per exemple, el saldo d'un departament no pot ser mai negatiu. La BD ha de chequejar sempre que es faci una modificació, que el saldo no esdevingui negatiu.

Restriccions testeables amb overhead mínim

- Restriccions de domini
- Integritat referencial. Ex. El valor Dept de qualsevol professor ha d'aparèixer a la taula Departament. Les modificacions poden portar a incompliment de restriccions d'integritat.
- Assercions: Són condicions que la BD sempre ha de complir. Per exemple, cada departament ha d'oferir com a mínim 5 cursos cada quadrimestre.
- Autorització: Cal diferenciar entre els usuaris i també els diferents tipus d'accés associat: accés per lectura, inserció, modificació, esborrat (i combinació d'aquests)

El DDL

- A partir d'un conjunt d'instruccions d'entrada genera una sortida que es guarda en el data dictionary.
- El data dictionary conté metadata (informació sobre la informació). És un tipus especial de taula al que únicament hi té accés de lectura i modificació el sistema de BD (no un usuari regular). El sistema de BD consulta el data dictionary abans de llegir o modificar la informació actual.
 - Esquema de la BD
 - restriccions d'integritat
 - Clau Primària (Id identifica de manera única el professorat)
 - Integritat referencial: Ex. El valor Dept de qualsevol professor ha d'aparèixer a la taula Departament
 - Autoritzacions

- Obtenir informació emmagatzemada en la BD
- Inserir nova informació
- Esborrar informació
- Modificar informació

2 tipus

- DMLs procedural. Especificar què obtenir i com.
- DMLs declaratiu. Especificar què es vol obtenir sense especificar el com.

Una consulta (query)

Alguns exemples:

- Troba el nom del professor amb identificador 22222

```
select name  
from professor  
where professor.id= '22222'
```

- Troba l'identificador i edifici dels professors del departament de física

```
select professor.id, departament.edifici  
from professor, departament  
where professor.dept=departament.dept  
      and departament.dept= 'Física'
```


Quin problema veieu en aquest disseny?

Id	Nom	Dept	Edifici	Planta
22222	Einstein	Física	MN2	1
11111	Mozart	Música	MN1	2
33333	Gold	Física	MN2	1
44444	Brand	Música	MN1	2

- Repetició de la informació.
- Impossibilitat de representar certa informació. Exemple: No es podrà crear un departament nou fins que no s'hi associi un professor.

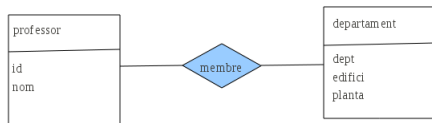
- Teoria de la normalització
- Model entitat relació

Model Entitat Relació

Modelització d'una empresa com un conjunt d'entitats i relacions

- Entitat i atributs
- Relacions: associació entre diverses entitats

Es representa per mitjà de diagrames entitat-relació (Model ER. Diagrama de Chen)



Què ha passat amb el nom de departament del professor?

- Què cal fer si el sistema falla?
- Què passa si més d'un usuari està actualitzant concurrentment la mateixa informació?

Per a tractar els accessos concurrents, els DBMS fan servir el concepte de transacció de BD. Anomenem transacció un conjunt d'operacions simples que s'executen com una unitat. Els DBMS han d'aconseguir que el conjunt d'operacions d'una transacció mai no s'executi parcialment. O s'executen totes o no se n'executa cap (ACID).

Propietats de les transaccions: ACID

- Atomicitat: all-or-nothing execution
- Consistència: La transacció ha de garantir que la BD continuarà sent consistent. Exemple: la restricció que els saldos dels comptes no siguin negatius quan acabi la transacció.
- Aïllament (Isolation): Cal garantir que cada transacció s'executarà com si cap altra s'executés al mateix temps
- Durabilitat: Un cop s'ha realitzat la transacció, cal que l'efecte no es perdi en la BD.

- Mantenir la integritat de les dades: COMMIT, ROLLBACK
- Evitar interferències: LOCK

- Centralitzat
- Client-Servidor
- Paral·lel (multi-processor)
- Distribuït

- 1950-1960
 - Cintes magnètiques amb accés seqüencial
 - Entrada de dades a través de targetes perforades

- 1960-1970: sistemes centralitzats
 - Aplicacions per lots (batch). Ús de fitxers amb accés seqüencial (fitxers mestre), problemes de redundància. Exemple: Bill of materials o Parts explosion (indústria automobilística), Data Banks (fitxers interrelacionats)
 - OLTP: aplicacions on-line transaccionals
 - Dependència del nivell físic
 - Baix rendiment. Calia millorar el temps de resposta i les transaccions per segon

- 1970-1980 DBMS relacionals. Codd.
 - Independència del nivell físic
 - Estandardització dels DMBS relacionals: llenguatge SQL com a estàndard industrial
 - Personal data bases (dBase, Access)
 - DB2, Ingres, Informix, Microsoft SQL Server, Oracle, Sybase.

- 1980-90: distribució, C/S i 4GL
 - Bases de dades OO relacionals
 - Bases de dades distribuïdes: arquitectura client/servidor (C/S)
 - Flexibilitat per a construir, fer modificacions i fer créixer la configuració informàtica global de l'empresa
 - Fourth Generation Languages (4GL)

- Tendències actuals: Data Warehouse, multimèdia, orientació a l'objecte, internet i el web. Big Data, cloud Computing
 - Incorporació de tecnologies multimedia: imatge, so
 - Estàndards XML i XQuery
 - Administració automatitzada de Bases de Dades
 - Data Warehouse
 - Sistemes d'emmagatzemament de la informació gegants: Google BigTable (petabytes of data), Yahoo PNuts, Amazon, Facebook Cassandra, Youtube, etc.

- <http://adrem.ua.ac.be/current-trends-databases>
- DB Trends conference 2012
<http://indico.cern.ch/conferenceDisplay.py?confId=192081>

- Capítol I i Capítol II, Sistac et al, “Bases de Dades”.
- Capítol I, D.J. Date, “An Introduction to DataBase Sytems. 8th edition”.
- Capítol I, Garcia-Molina, H., Ullman, Jeffrey, Widom, J. “DataBase Systems. The Complete Book. 2nd edition”.
- Capítol I, Silberschatz, Korth and Sudarshan, “Database system Concepts. 6th edition”.