

BAZE DE DATE

CURS: FLORIN RADULESCU

Email: florin@cs.pub.ro

Lab: Cf. orar

NOTARE

1. 60% IN CURSUL SEMESTRULUI:
 - ◆ PREZENTA CURS – 10%
 - ◆ PREZENTA, ACTIVITATE SI TEST LABORATOR – 30%
 - ◆ LUCRARE LA MIJLOCUL SEMESTRULUI FARA DEGREVARE – 20%
1. 40% VERIFICARE FINALA (EXAMEN)

1.1. DEFINITII

- ◆ Bază de date
- ◆ Sistem de gestiune a bazelor de date

- ◆ **Definiție:** O **bază de date** (BD, eng. DB) este un ansamblu structurat de date înregistrat pe suporturi accesibile calculatorului pentru a satisface simultan cerintele mai multor utilizatori într-un mod selectiv și în timp util.
- ◆ **Definiție:** Un **sistem de gestiune a bazelor de date** (SGBD/DBMS) este ansamblul de programe care permit utilizatorului să interacționeze cu o bază de date.

O bază de date (BD, eng. DB) este **un ansamblu structurat de date** înregistrat pe suporturi accesibile calculatorului pentru a satisface simultan cerintele mai multor utilizatori într-un mod selectiv și în timp util.

... un ansamblu structurat de date ...

Regulile și conceptele care permit descrierea structurii unei BD formează **modelul datelor**. În timp au fost definite trei astfel de modele:

- ◆ **Modelul ierarhic** în care datele erau organizate sub forma unui arbore, nodurile constând din înregistrări iar arcele referinte către alte noduri.
- ◆ **Modelul rețea** în care datele erau organizate sub forma unui graf orientat. Nodurile și arcele au aceeași semnificație ca mai sus.
- ◆ **Modelul relațional** în care, intuitiv, datele sunt organizate sub forma de tabele.

RELAȚIE

- ◆ Termenul de **relație** (care da denumirea modelului) provine din matematica iar reprezentarea intuitivă a unei relații este o tabelă.
- ◆ În cazul modelului relațional descrierea structurii unei baze de date constă în principal din descrierea tabelelor componente:
 - ◆ denumire,
 - ◆ lista de coloane și
 - ◆ tipul datelor din acestea.

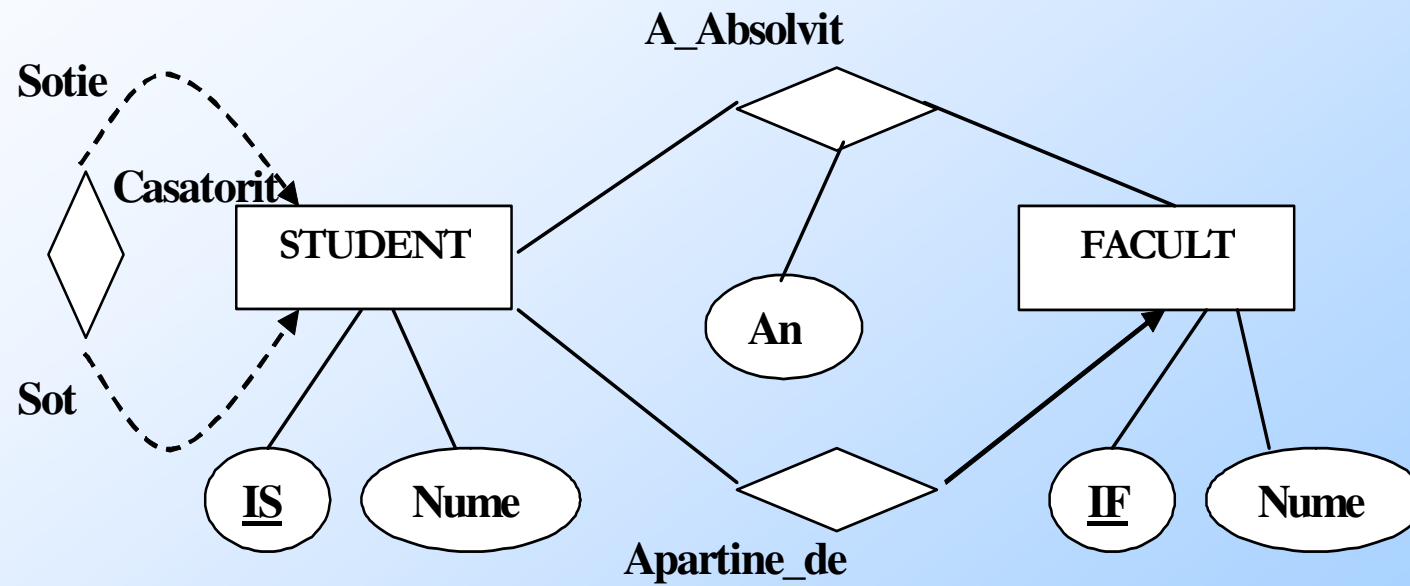
CONSTRÂNGERI DE INTEGRITATE

- ◆ De asemenea descrierea poate include informații despre elementele de identificare ale înregistrărilor dintr-o tabela (cheie primara), a coloanelor comune cu alte tabele (chei straine) și a altor restricții și verificari de corectitudine pentru datele stocate. Aceste informații sunt cunoscute sub numele generic de **constrangeri de integritate**.

MODELUL ENTITATE-ASOCIERE (EA)

- ◆ Operațiile de proiectare și descriere a structurii unei baze de date sunt facute în general cu ajutorul modelului **entitate-asociere**.
- ◆ Acesta este un model generic de descriere a datelor sub forma grafica - **diagrame entitate asociere** - din care se poate trece pe baza unor reguli de transformare în alte modele de date.

EXEMPLU



O bază de date (BD, eng. DB) este un ansamblu structurat de date **înregistrat pe suporturi accesibile calculatorului** pentru a satisface simultan cerintele mai multor utilizatori într-un mod selectiv și în timp util.

...înregistrat pe suporturi accesibile calculatorului ...

În limbajul curent se intalnesc expresii ca:
"avem aceasta informație în baza noastră de
date" și în cazurile în care datele respective
sunt de exemplu stocate sub forma unor fise
(pe hartie) sortate alfabetic sau dupa alte
criterii. În acest caz este vorba despre o
extensie a termenului de **bază de date**.

În cazul sistemelor de gestiune a bazelor de
date suporturile pe care sunt stocate datele
sunt în principal

- ◆ magnetice
- ◆ optice.

O bază de date (BD, eng. DB) este un ansamblu structurat de date înregistrat pe suporturi accesibile calculatorului **pentru a satisface simultan cerintele mai multor utilizatori** într-un mod selectiv și în timp util.

... pentru a satisface simultan cerintele mai multor utilizatori ...

- ◆ **1. Gestiunea utilizatorilor.** Un SGBD trebuie sa permita crearea, modificarea și stergerea utilizatorilor. Operația este efectuată de obicei de administratorul bazei de date.
- ◆ **2. Concurența la date.** În cazul accesului simultan al mai multor utilizatori la aceleași date un SGBD trebuie sa aiba mecanisme pentru a preintampina inconsistența datelor.

CONCURENTA - EXEMPLU

Fie o portiune a bazei de date, notata A , în care se tine numarul de locuri disponibile la o cursa aeriana și doua executii ale unui program de rezervare rulat de doua agentii de bilete diferite. Ele contin secventa:

READ A /* se citeste numarul de locuri disponibile din BD */

A = A - 1 /* se decrementează cu o unitate */

WRITE A /* se scrie noul numar de locuri disponibile în BD */

PLANIFICARE POSIBILA

MOMENT	AG. 1	AG. 2	A IN BD
T1	READ A		10
T2		READ A	10
T3	$A = A - 1$		10
T4		$A = A - 1$	10
T5	WRITE A		9
T6		WRITE A	9

Se observa ca desi s-au rezervat 2 locuri, A a scazut cu doar o unitate. O astfel de situatie trebuie preintampinata.

- ◆ In curs vom prezenta problemele legate de:
 - ◆ blocarea (acapararea) unor portiuni ale BD de catre o executie a unui program,
 - ◆ rezolvarea problemelor pe care le poate ridica asteptarea circulara pentru deblocarea acestor portiuni (deadlock),
 - ◆ executia pasilor programelor de actualizare a BD echivalenta cu o executie secventiala a programelor (serializabilitate) și
 - ◆ reguli de scriere a programelor de aplicatie pentru rezolvarea problemelor de acces concurrent.

O bază de date (BD, eng. DB) este un ansamblu structurat de date înregistrat pe suporturi accesibile calculatorului pentru a satisface simultan cerintele mai multor utilizatori **intr-un mod selectiv** și în timp util.

... într-un mod selectiv ...

Orice SGBD are mecanisme prin care diversilor utilizatori sau categorii de utilizatori li se asociază **drepturi de acces** specifice la obiectele bazei de date.

În acest mod fiecarui utilizator i se da dreptul de a efectua doar operațiile specifice activității sale și doar pe acea porțiune a bazei de date care este necesară pentru acestea.

MOTIV

- ◆ Blocarea accesului unor categorii de utilizatori la date pe care **nu trebuie** sa le acceseze. În acest fel este asigurata una dintre funcțiunile de baza ale unui SGBD și anume **confidentialitatea datelor**.
- ◆ Blocarea accesului unor categorii de utilizatori la date de care **nu au nevoie** în activitatea lor, minimizându-se astfel riscul distrugerii accidentale a datelor prin operații necorespunzatoare.

O bază de date (BD, eng. DB) este un ansamblu structurat de date înregistrat pe suporturi accesibile calculatorului pentru a satisface simultan cerintele mai multor utilizatori într-un mod selectiv și în **timp util**.

... și în timp util (1)...

- ◆ În cazul bazelor de date de dimensiuni mari este evident ca orice cautare care s-ar baza pe o parcurgere secventiala a înregistrarilor din tabele ar duce la timpi de raspuns inadecvat de mari.
- ◆ De asemenea, operații mai complicate prin care se regasesc date stocate în mai multe tabele legate între ele prin coloane comune pot duce în lipsa unor algoritmi specifici la timpi de executie inacceptabili.

... și în timp util (2)...

De aceea orice SGBD are mecanisme prin care minimizează timpul de raspuns, mecanisme bazate în special pe indecși și modalități specifice de organizare fizica a datelor.

1.2. Funcțiile unui SGBD

1. → Descrierea datelor
2. Utilizarea datelor
3. Integritatea datelor
4. Confidentialitatea datelor
5. Accesul concurent la date
6. Siguranta în funcționare

Descrierea datelor (1)

- ◆ Un SGBD trebuie sa includă posibilitatea descrierii stucturii obiectelor care formează baza de date. În cazul bazelor de date relaționale aceasta consta în principal în posibilitatea creerii și modificării structurii tabelelor și constrangerilor de integritate asociate acestora.
- ◆ Limbajul prin care se realizează aceste operații se numeste **Limbaj de Descriere a Datelor** (LDD) și în cazul primelor sisteme de gestiune el era implementat sub forma unor module separate. În sistemele relaționale bazate pe SQL aceste operații au fost incluse în limbaj sub forma comenzilor de tip CREATE (pentru creare) sau ALTER (modificare).

Descrierea datelor (2)

- ◆ O caracteristica importanta a sistemelor relaționale este aceea ca descrierea structurii bazei de date este stocata în același mod ca și baza de date în sine.
- ◆ Exista pentru aceasta o serie de tabele gestionate direct de sistemul de gestiune numite **cataloge** care contin descrierea obiectelor care formează baza de date.

Descrierea datelor (3)

- ◆ Aceste tabele sunt accesate de sistem pentru a regasi informațiile necesare executiei cererilor dar pot fi accesate și de utilizatori pentru a obtine informații privind structura bazei de date.
- ◆ Astfel toate informațiile de sistem necesare unui SGBD relațional sunt stocate în același mod, inclusiv descrierea utilizatorilor și a drepturilor de acces ale acestora.
- ◆ Aceasta abordare a fost necesara pentru a nu avea mecanisme diferite de acces la datele de sistem și la datele stocate în baza de date.

1.2. Funcțiile unui SGBD

1. Descrierea datelor
2. → Utilizarea datelor
3. Integritatea datelor
4. Confidentialitatea datelor
5. Accesul concurent la date
6. Siguranta în funcționare

Utilizarea datelor (2)

Aceasta funcție include operațiile de lucru cu datele înregistrate într-o bază de date.

Există patru categorii de operații principale:

- ◆ **Inserarea** de noi date. Aceasta se concretizează prin adăugarea de noi linii în tabelele care formează baza de date.
- ◆ **Stergerea** de linii din tabele.
- ◆ **Actualizarea** datelor, însemnând modificarea conținutului unor linii existente în tabele.
- ◆ **Regăsirea** datelor după anumite criterii de căutare

Utilizarea datelor (3)

Pentru implementarea acestei funcții fiecare SGBD are un **Limbaj de Manipulare a Datelor (LMD)** care poate fi un modul separat sau inclus în limbajul sistemului cum este în cazul SQL.

Utilizarea datelor (4)

Operația de **regasire** este esentiala, ea fiind necesara și în cazul celorlalte trei operații:

1. În cazul inserarii, dacă aceeași informație nu poate fi prezenta de mai multe ori în baza de date, este necesara o cautare prealabila pentru a verifica acest lucru.
2. În cazul stergerii și actualizarii este necesara localizarea datelor care vor fi eliminate sau modificate.

Executia unei cereri - etape

1. Consultarea cataloagelor de sistem pentru a verifica existenta obiectelor implicate și a afla structura și proprietățile acestora precum și pentru a verifica daca utilizatorul care a lansat comanda are dreptul de a efectua operația respectiva
2. Crearea unui plan de executie optim pentru operația respectiva, luând în considerație existenta unor eventuale structuri de cautare rapida (indecsi).
3. Executia operației și verificarea faptului ca rezultatul nu violează constrangerile de integritate existente
4. Executarea altor operații care pot fi declansate automat de operația de baza (în cazul în care sistemul suporta astfel de operații).

1.2. Funcțiile unui SGBD

1. Descrierea datelor
2. Utilizarea datelor
3. → Integritatea datelor
4. Confidentialitatea datelor
5. Accesul concurent la date
6. Siguranta în funcționare

Integritatea datelor (1)

- ◆ Majoritatea sistemelor de gestiune permit definirea unor reguli pe care datele stocate trebuie sa le verifice numite **constrangeri de integritate**.
- ◆ În cazul în care o operație are ca rezultat violarea acestor restricții aceasta este automat rejectata și nu are efect în baza de date.
- ◆ În felul acesta este asigurata o mai mare siguranta în ceea ce priveste corectitudinea datelor.

Integritatea datelor (2)

- ◆ Sa consideram o bază de date
continând informații despre
 - ◆ facultați,
 - ◆ studentii inscrisi și
 - ◆ notele obtinute de acestia.

Integritatea datelor (3)

- ◆ Introducerea accidentală de valori incorecte în baza de date. Exemplu: restricție privind intervalul de valori admisibil pentru note: 0-10.
- ◆ Lasarea necompletată a unor informații necesare. Exemplu: numele studentului.
- ◆ Alocarea accidentală a aceluiași număr matricol pentru mai mulți studenți (constrângere de unicitate a valorilor).
- ◆ Introducerea accidentală a unui student cu un cod incorect al facultății la care acesta este înscris (constrângere referențială)

Integritatea datelor (4)

- ◆ Definierea de constrangeri de integritate nu previne însă total erorile accidentale de operare: de exemplu introducerea din greșeală a unei note de 4 în loc de 5 nu va fi semnalată, ambele valori fiind în intervalul admisibil.

1.2. Funcțiile unui SGBD

1. Descrierea datelor
2. Utilizarea datelor
3. Integritatea datelor
4. → Confidentialitatea datelor
5. Accesul concurent la date
6. Siguranta în funcționare

Confidentialitatea datelor (1)

- ◆ Accesul la date este permis doar **utilizatorilor înregistrați** și doar în măsura drepturilor de acces alocate.
- ◆ Un utilizator este identificat printr-un **nume-utilizator** și o **parola**.
- ◆ Fiecarui utilizator i se permite accesul doar la o **portiune** a bazei de date și doar pentru a efectua anumite tipuri de operații.
- ◆ Toate aceste informații relative la ce, cum și de unde poate accesa datele un utilizator reprezintă **drepturile de acces** asociate acestuia
- ◆ Drepturile de acces sunt stocate în **cataloge**le sistemului.

Confidentialitatea datelor (1)

- ◆ Prin mecanismul de drepturi de acces este asigurata funcția de **confidentialitate** a datelor.
- ◆ Este prevenita posibilitatea de acces la date a persoanelor care **nu** au acest drept.
- ◆ Exemplu: în cazul sistemului informatic al unei organizații, acces selectiv la salariile angajaților, detalii ale proceselor tehnologice, datele contabile etc.

1.2. Funcțiile unui SGBD

1. Descrierea datelor
2. Utilizarea datelor
3. Integritatea datelor
4. Confidentialitatea datelor
5. → Accesul concurent la date
6. Siguranta în funcționare

Accesul concurent la date (1)

- ◆ ***Facilități de blocare a unor porțiuni ale bazei de date.*** Aceasta înseamnă că o execuție a unui program poate capata un acces exclusiv la o porțiune a bazei de date, denumită și **articol**, moment în care orice altă execuție trebuie să aștepte deblocarea acesteia pentru a continua.
- ◆ Dacă notăm operațiile de blocare și deblocare cu **LOCK** și **UNLOCK**, programul de rezervare de locuri poate fi rescris

Accesul concurent la date (2)

- ◆ **LOCK A** /* se blochează articolul A */
- ◆ **READ A** /* se citește A din BD */
- ◆ **A = A - 1** /* se decrementează A */
- ◆ **WRITE A** /* se scrie A în BD */
- ◆ **UNLOCK A** /* se deblochează A */

Accesul concurent la date (3)

- ◆ Blocarea articolelor rezolva o serie de probleme dar crează premisele apariției interblocajului, situație în care două execuții de programe asteapta deblocarea unui articol blocat de cealaltă (**deadlock**)

Accesul concurent la date (4)

Program P1

LOCK A

LOCK B

.

UNLOCK B

UNLOCK A

Program P2

LOCK B

LOCK A

.

UNLOCK A

UNLOCK B

Accesul concurent la date (5)

- ◆ ***Executia serializabila.*** În cazul mai multor executii simultane care accesează baza de date se considera ca efectul lor este corect daca rezultatul final este identic cu executia lor succesiva.
- ◆ Aceasta proprietate se numeste **serializabilitate** și un SGBD trebuie sa puna la dispozitie mecanisme de asigurare a serializabilității executiilor programelor de acces.

1.2. Funcțiile unui SGBD

1. Descrierea datelor
2. Utilizarea datelor
3. Integritatea datelor
4. Confidentialitatea datelor
5. Accesul concurent la date
6. → Siguranta în funcționare

Siguranta în funcționare (1)

- ◆ Desi nu este legata direct de cele prezentate pana acum, siguranta în funcționare este o caracteristica esentiala pentru un SGBD
- ◆ Contine acele elemente care exclud sau minimizează posibilitatea de pierdere a datelor datorata incidentelor software sau hardware.

Siguranta în funcționare (2)

- ◆ ***Salvarea datelor.*** Implementarea operațiilor de salvare este mult mai sofisticata
- ◆ Este dificila efectuarea unei de copii de siguranta consistente ale bazei de date în conditiile în care aplicația rulează non-stop și operarea nu poate fi oprita pentru efectuarea salvării.

Siguranta în funcționare (3)

- ◆ ***Restaurarea dupa incident.*** La distrugerea bazei de date este necesara efectuarea operației de restaurare din copie.
- ◆ Multe tipuri de SGBD au posibilitatea de a înregistra toate schimbarile aparute în baza de date sub forma unor ***fișiere jurnal.***
- ◆ În acest fel operația de restaurare dupa incident se face restaurând ultima copie de siguranta și reefectuând toate prelucrarile consistente înregistrate în jurnal dupa momentul în care a fost efectuata copia.

1.3. Categoriile de utilizatori

1. Utilizatori privilegiiati

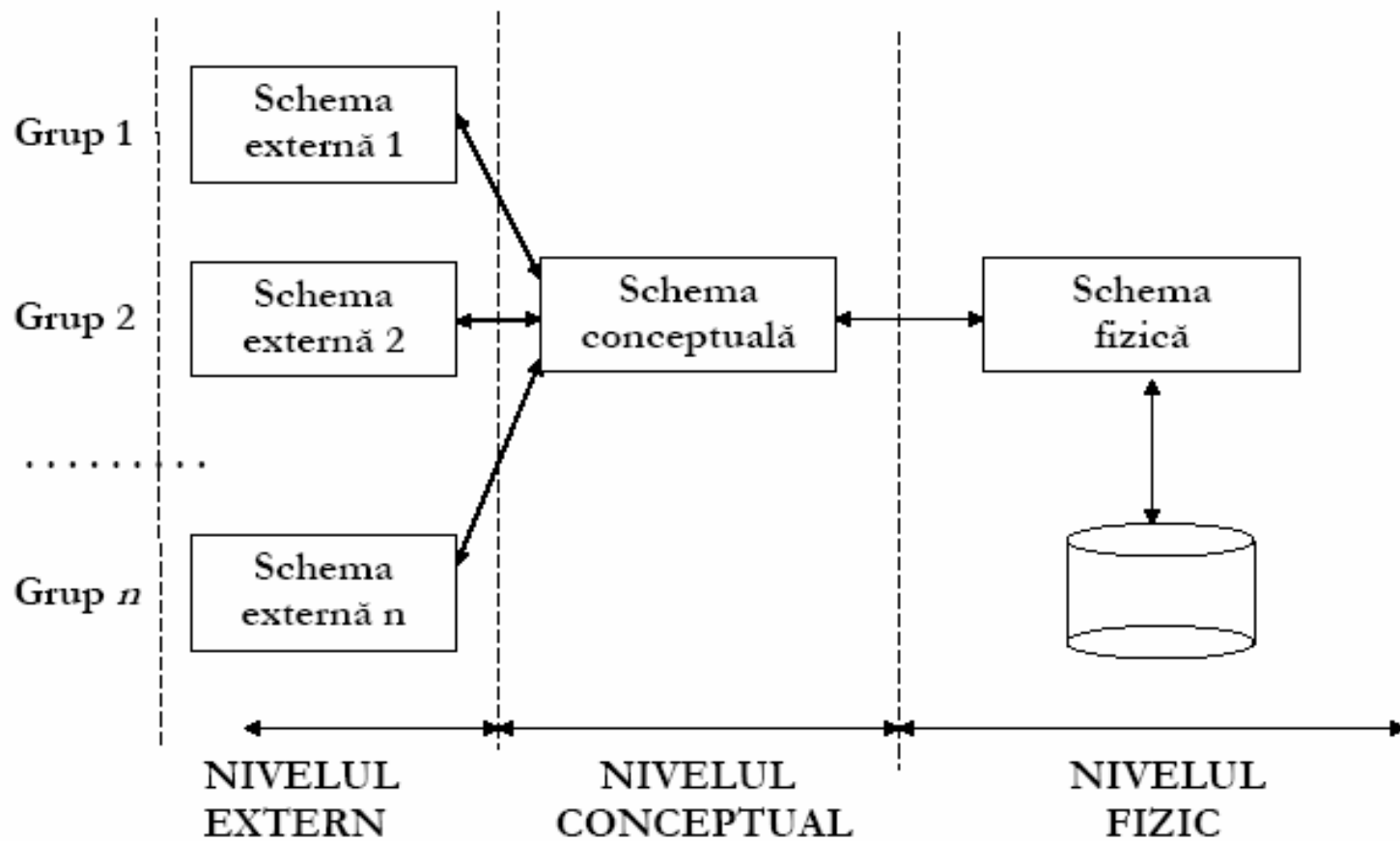
- ◆ Administratorul bazei de date

2. Utilizatori neprivilegiiati

- ◆ Utilizatorul este o persoana fizica (exemplu: în Oracle o persoana care utilizeaza **sqlplus**).
- ◆ Utilizatorul este o entitate generica (aplicatie software)

1.4. Nivele de reprezentare

- ◆ O aceeași bază de date poate fi privită din diverse perspective rezultând descrieri diferite.
- ◆ Termenul consacrat pentru descrierea structurii unei baze de date este acela de **schema**.
- ◆ În literatura de specialitate există o clasificare pe trei nivele a acestor descrieri: ***fizic, conceptual și extern***.
- ◆ Gestionarea schemelor bazei de date pentru fiecare nivel este una dintre atribuțiile **administratorului** bazei de date.



NIVELE DE DESCRIERE PENTRU O BAZA DE DATE

Nivelul fizic

- ◆ La acest nivel baza de date este descrisa din perspectiva stocarii sale pe dispozitivele fizice: identificarea discurilor și a cailor unde este stocata, numele fișierelor care formează baza de date, structura fizica a acestora, etc.
- ◆ Descrierea bazei de date la acest nivel poarta numele de **schema fizica**

Nivelul conceptual (1)

- ◆ Descrierea bazei de date la acest nivel poarta numele de **schema conceptuala** (numita uneori și schema logica) a bazei de date.
- ◆ Ea consta intr-o descriere abstracta dar exacta a structurii acesteia, lasând la o parte detaliile fizice de implementare.

Nivelul conceptual (2)

- ◆ Tabelele care formează baza de date
- ◆ Structura (coloanele) fiecărei tabele
- ◆ Tipul de date asociat coloanelor
- ◆ Elementele pe baza carora se realizează interconectarea tabelelor (coloane comune)
- ◆ Constrangeri de integritate
- ◆ Operații declansate automat la modificarea unor elemente ale bazei de date

Nivelul extern

- ◆ Diferitele categorii de utilizatori ai unei baze de date au nevoie în activitatea lor doar de porțiuni specifice ale acesteia.
- ◆ Descrierea acestor porțiuni poartă numele de **scheme externe**.
- ◆ O bază de date are deci asociată o singură schemă fizică și o singură schemă conceptuală dar mai multe scheme externe.

Independenta datelor (1)

Cablarea informațiilor structurale în program constituie o bariera în calea dezvoltării și modificării acestora.

Orice operație de modificare a bazei de date a unei aplicații se masoara și prin prisma costurilor materiale

- ◆ *Independenta logica*
- ◆ *Independența fizică*

Independenta logica (1)

- ◆ ***Independenta logica*** reprezinta posibilitatea de schimbare a schemei conceptuale a bazei de date fara modificarea schemelor externe. Conditia este ca modificarea sa nu elimine nici unul dintre elementele necesare translatiei de la schema externa la schema conceptuala.

Independenta logica (2)

Permite:

- ◆ Adaugarea de noi tabele în baza de date
- ◆ Adaugarea de noi coloane în tabelele existente
- ◆ Modificarea numelor tabelor și coloanelor existente
- ◆ Adaugarea de noi constrangeri de integritate
- ◆ Modificarea în anumite limite a tipului datelor stocate în baza de date
- ◆ Restructurarea bazei de date prin mutarea unor coloane în alte tabele, fragmentarea unei tabele sau reunirea mai multor tabele.

Exemplu (1)

Schema conceptuala initiala:

- ◆ Tabela Student

- ◆ Coloanele tabelii:

CodStudent, Nume, CodFacultate, Medie

Scheme externe initiale:

- ◆ Vederile Student_i

- ◆ Coloanele vederilor:

CodStudent, Nume, CodFacultate, Medie

- ◆ Liniile vederilor:

Liniile din tabela Student care pe coloana CodFacultate contin valoarea i.

Exemplu (2)

Daca adaugam o noua coloana
"CodSpecializare" in tabela Student,
modificam doar definitia vederilor
Student_i:

Valorile de pe coloanele CodStudent,
Nume, CodFacultate și Medie din liniile
din tabela ***Student*** care pe coloana
CodFacultate contin valoarea ***i***.

Independența fizică (1)

- ◆ ***Independența fizică*** reprezintă posibilitatea de schimbare a schemei fizice a bazei de date fără modificarea schemei conceptuale și implicit a schemelor externe.
- ◆ Aceasta da posibilitatea reorganizării fizice a bazei de date fără afectarea aplicațiilor care o folosesc.

Independenta fizica (2)

Permite:

- ◆ Schimbarea dispozitivelor fizice pe care este stocata baza de date
- ◆ Schimbarea numelor fișierelor fizice în care este stocata baza de date sau a directoarelor unde acestea sunt plasate
- ◆ Adaugarea de noi structuri de cautare rapida (indecsi) pentru cresterea vitezei de executie
- ◆ Schimbarea în anumite conditii a structurii fizice a fișierelor bazei de date
- ◆ Schimbarea unor parametri ai sistemului de gestiune care afectează modul în care datele sunt stocate la nivel fizic

Sfârșitul primului capitol