

1. CONCEPTE ȘI PROBLEMATICĂ

În acest moment termenul de **bază de date** a intrat în limbajul curent fiind folosit uneori și în alte accepțiuni decât cea de proveniență. În acest capitol sunt definite conceptele de **bază de date** (BD) și **sistem de gestiune a bazelor de date** (SGBD) și se face o trecere în revistă a elementelor continute în aceste definiții. Sunt prezentate apoi **funcțiile** pe care trebuie să le asigure un SGBD și **categoriile de utilizatori** ale unui astfel de sistem. Capitolul se încheie cu prezentarea celor **trei nivele de descriere** a unei baze de date și a conceptului de **independența date-program**.

1.1. Bază de date și Sistem de gestiune a bazelor de date

Literatura de specialitate conține mai multe definiții pentru conceptele de **bază de date** și **sistem de gestiune a bazelor de date**. În continuare sunt prezentate cele din [DeAd82] pentru valoarea lor descriptivă.

Definiție: O bază de date este un ansamblu structurat de date înregistrat pe suporturi accesibile calculatorului pentru a satisface simultan cerințele mai multor utilizatori într-un mod selectiv și în timp util.

Definiție: Un sistem de gestiune a bazelor de date este ansamblul de programe care permit utilizatorului să interacționeze cu o bază de date.

Aceste definiții conțin majoritatea elementelor importante ale problematicei bazelor de date prezentate în capitolele următoare:

... un ansamblu structurat de date ...

Nu orice colecție de date este o bază de date. O cerință primordială este aceea a organizării acestora după anumite reguli. 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 referințe 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.

Dezavantajul principal al primelor două modele constă în operarea greoie cu datele: orice operație necesită navigația printr-un arbore sau o rețea pentru localizarea datelor cautate

sau a locului în care se înserează noile înregistrari. De aceea ele nu se mai folosesc în realizarea sistemelor de gestiune moderne, acestea fiind bazate pe modelul relațional al datelor. Popularitatea acestui model este datorată simplității sale (din punct de vedere al utilizatorului) și a posibilității de definire a unor limbaje neprocedurale de descriere și manipulare a datelor.

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.

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

Capitolul 3 prezintă bazele teoretice ale modelului relațional iar Capitolele 4 și 5 limbajul de cereri SQL care s-a impus ca standard pentru sistemele relaționale.

Operațiile de proiectare și descriere a structurii unei baze de date sunt făcute în general cu ajutorul modelului **entitate-asociere**. Acesta este un model generic de descriere a datelor sub forma grafică - diagrame entitate asociere - din care se poate trece pe baza unor reguli de transformare în alte modele de date. Capitolul 2 prezintă două versiuni ale acestui model iar în cadrul Capitolului 3 sunt descrise regulile de transformare din model entitate-asociere în model relațional.

...înregistrat pe suporturi accesibile calculatorului ...

Dacă ansamblul de date nu este înregistrat pe suporturi accesibile calculatorului acesta nu se poate numi bază de date în accepțiunea lucrării de față.

În limbajul curent se întâlnesc expresii ca: "avem această informație în baza noastră de date" și în cazurile în care datele respective sunt de exemplu stocate sub forma unor fișe (pe hartie) sortate alfabetic sau după 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 și optice.

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

Funcțiile unui SGBD relative la accesul utilizatorilor la baza de date sunt următoarele:

1. Gestiunea utilizatorilor. Un SGBD trebuie să permită crearea, modificarea și ștergerea 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 să aibă mecanisme pentru a preîntâmpina inconsistența datelor. Iată un exemplu de caz:

Fie o porțiune a bazei de date, notată A, în care se ține numărul de locuri disponibile la o cursă aeriană și două execuții ale unui program de rezervare rulat de două agenții de bilete diferite. Dacă programul este de tipul:

```
READ A      /* se citește numărul de locuri disponibile din BD */  
A = A - 1    /* se decrementează cu o unitate */  
WRITE A     /* se scrie noul număr de locuri disponibile în BD */
```

atunci este posibil ca pașii celor două execuții să fie întretreși în felul următor:

Moment de timp	Agentia 1	Agentia 2	A în 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 observă că deși s-au rezervat 2 locuri, A a scăzut cu doar o unitate. O astfel de situație trebuie preintampinată.

Capitolul 6 prezintă problemele legate de blocarea (acapararea) unor porțiuni ale BD de către o execuție a unui program, rezolvarea problemelor pe care le poate ridica așteptarea circulară pentru deblocarea acestor porțiuni (deadlock), execuția pașilor programelor de actualizare a BD echivalentă cu o execuție secvențială a programelor (serializabilitate) și reguli de scriere a programelor de aplicație pentru rezolvarea problemelor de acces concurrent.

... 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 fiecărui utilizator i se dă dreptul de a efectua doar operațiile specifice activității sale și doar pe acea porțiune a bazei de date care este necesară pentru acestea.

Mecanismul de drepturi de acces are ca obiective principale:

- Blocarea accesului unor categorii de utilizatori la date pe care **nu trebuie** să le acceseze. În acest fel este asigurată una dintre funcțiile de bază ale unui SGBD și anume **confidențialitatea 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 necorespunzătoare.

Fiecare tip de SGBD are propriile sale mecanisme de descriere a drepturilor de acces bazate în principal pe acordarea sau neacordarea dreptului de a citi sau scrie diverse porțiuni ale bazei de date. O prezentare detaliată a acestei probleme se găsește în Capitolul 7.

... și în timp util ...

În cazul bazelor de date de dimensiuni mari este evident că orice căutare care s-ar baza pe o parcurgere secvențială a înregistrărilor din tabele ar duce la timpi de răspuns inadecvat de mari. De asemenea, operații mai complicate prin care se regăsesc date stocate în mai multe tabele legate între ele prin coloane comune pot duce în lipsa unor algoritmi specifici la timpi de execuție inacceptabili.

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

1.2. Funcțiile unui SGBD

O definiție alternativă a conceptului de bază de date este următoarea [ElNa94]: o bază de date este o colecție de date interconectate. Deci ea reprezintă depozitul de date al oricărei aplicații de gestiune, partea sa statică. Operațiile asupra datelor sunt efectuate de sistemul de gestiune a bazelor de date. El este cel care asigură structurarea datelor, accesul concurrent al utilizatorilor, selectivitatea accesului și timpi de execuție normali pentru cereri. Dar acestea sunt doar o parte din operațiile pe care acesta trebuie să le asigure. Funcțiile unui sistem de gestiune a bazelor de date sunt următoarele:

1.2.1. Descrierea datelor

Un SGBD trebuie să includă posibilitatea descrierii structurii obiectelor care formează baza de date. În cazul bazelor de date relaționale aceasta constă în principal în posibilitatea creerii și modificării structurii tabelor și constrângerilor de integritate asociate acestora.

Limbajul prin care se realizează aceste operații se numește 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).

O caracteristică importantă a sistemelor relaționale este aceea că descrierea structurii bazei de date este stocată în același mod ca și baza de date în sine. Există pentru aceasta o serie de tabele gestionate direct de sistemul de gestiune numite **cataloge** care conțin descrierea obiectelor care formează baza de date. Aceste tabele sunt accesate de sistem pentru a regăsi informațiile necesare execuției cererilor dar pot fi accesate și de utilizatori pentru a obține 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.

Această abordare a fost necesară pentru a nu avea mecanisme diferite de acces la datele de sistem și la datele stocate în baza de date.

Din cele de mai sus rezultă că nu orice program care gestionează date este un SGBD. Să luăm exemplul unui program scris în limbajul C care gestionează niște fișiere de înregistrări a caror structură este cablată în program. Acesta nu respectă funcția de descriere a datelor necesară unui SGBD deoarece utilizatorul:

- Nu poate adăuga noi fișiere de date.
- Nu poate modifica structura fișierelor de date gestionate (adaugări sau ștergeri de câmpuri ale înregistrărilor fișierelor, schimbarea tipului acestora).
- Structura bazei de date este cablată în program și nu este stocată în baza de date.

Un SGBD trebuie să permită deci definirea și modificarea de către utilizatorii autorizați a structurii bazei de date, limitările fiind doar de natură tehnică sau privind tipurile de obiecte care se pot înregistra: dimensiune maximă pentru o tabelă, număr maxim de coloane sau linii dintr-o tabelă, tipul unei coloane ales dintr-o mulțime predefinită de tipuri posibile, etc.

1.2.2. Utilizarea datelor

Această funcție include operațiile de lucru cu datele înregistrate într-o bază de date. Cele patru categorii de operații principale sunt următoarele:

- 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

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.

Operația de regăsire este esențială, ea fiind necesară și în cazul celorlalte trei operații:

- În cazul inserării, dacă aceeași informație nu poate fi prezentă de mai multe ori în baza de date, este necesară o căutare prealabilă pentru a verifica acest lucru.
- În cazul stingerii și actualizării este necesară localizarea datelor care vor fi eliminate sau modificate.

Pentru efectuarea unei operații de acest tip un SGBD execută mai mulți pași:

1. Consultarea cataloagelor de sistem pentru a verifica existența obiectelor implicate și a afla structura și proprietățile acestora precum și pentru a verifica dacă utilizatorul care a lansat comanda are dreptul de a efectua operația respectivă
2. Crearea unui plan de execuție optim pentru operația respectivă, luând în considerație existența unor eventuale structuri de căutare rapidă (indecsi).
3. Execuția operației și verificarea faptului că rezultatul nu violează constrângerile de integritate existente
4. Executarea altor operații care pot fi declanșate automat de operația de bază (în cazul în care sistemul suportă astfel de operații).

Și din punct de vedere al acestei funcții nu toate programele de gestiune de date sunt SGBD. În cazul programului scris în limbajul C descris anterior, dacă operarea se face prin intermediul unui arbore de meniuri care permite efectuarea unui set prestabilit de operații funcția de utilizare a datelor nu este îndeplinită. Adăugarea unei noi opțiuni în program descriind o operație care nu a fost prevăzută nu se poate face de utilizator ci doar prin modificarea și recompilarea programului.

O situație specială este aceea a programelor dedicate care au însă la bază un sistem de gestiune a bazelor de date. În acest caz SGBD-ul dispune de toate facilitățile de descriere și utilizare a datelor dar acestea nu sunt puse la dispoziția utilizatorului aplicației decât în măsura în care au fost cuprinse în meniurile acesteia.

1.2.3. Integritatea datelor

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 rejectată și nu are efect în baza de date. În felul acesta este asigurată o mai mare siguranță în ceea ce privește corectitudinea datelor.

Să considerăm o bază de date conținând informații despre facultăți, studenții înscriși și notele obținute de aceștia. Impunerea unor constrangeri de integritate poate preveni în acest caz:

- Introducerea accidentală de valori incorecte în baza de date. De exemplu pentru tabela conținând notele studenților se poate defini o restricție privind intervalul de valori admisibil pentru acestea: 0-10. Orice operație de introducere a unei noi note sau de modificare a uneia existente care nu respectă acest interval de valori va genera o eroare și nu va fi dusă până la capăt.
- Lăsarea necompletată a unor informații necesare. În cazul introducerii unui nou student sau a modificării informațiilor despre unul deja înregistrat, o restricție de acest tip asociată coloanei sau coloanelor pe care este memorat numele studentului va evita apariția în baza de date a unor înregistrări în care numele este necompletat.
- Alocarea accidentală a aceluiași număr matricol pentru mai mulți studenți în cazul în care pentru coloana conținând acest număr s-ar defini o constrângere de unicitate a valorilor.
- Introducerea accidentală a unui student cu un cod incorect al facultății la care acesta este înscris. Acest lucru se poate realiza definind o constrângere de integritate prin care se verifică existența în tabela de facultăți a valorii respective înainte de scrierea ei în tabela de studenți.

Definirea 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.

1.2.4. Confidentialitatea datelor

În cazul unui SGBD accesul la date este permis doar utilizatorilor înregistrați și doar în măsura drepturilor de acces alocate. Paragraful 1.3. conține o prezentare detaliată privind tipurile de utilizatori ai unui sistem de gestiune și rolurile acestora.

Un utilizator este identificat printr-un nume-utilizator și o parolă. Fiecarui utilizator i se permite accesul doar la o porțiune a bazei de date și doar pentru a efectua anumite tipuri de operații. În cazul bazelor de date accesate în rețea se pot defini de asemenea locațiile de la care utilizatorul poate interacționa cu baza de date.

Toate aceste informații relative la ce, cum și de unde poate accesa datele un utilizator reprezintă drepturile de acces asociate acestuia și sunt stocate în cataloagele sistemului.

Asa cum s-a menționat anterior, prin mecanismul de drepturi de acces este asigurată funcția de confidentialitate a datelor. Este prevenită astfel posibilitatea de acces la date a persoanelor care nu au acest drept: de exemplu, în cazul sistemului informatic al unei organizații, accesul la datele privind salariile angajaților, detalii ale proceselor tehnologice, datele contabile sau alte informații sensibile nu vor putea fi accesate decât de categorii restrânse de utilizatori.

Din acest punct de vedere nu orice pachet de programe de gestiune a datelor poate fi numit SGBD deoarece unele dintre acestea fie nu contin nici un fel de facilitate privind gestiunea utilizatorilor și asigurarea confidentialității fie aceste facilități sunt rudimentare și nu pot fi considerate ca atare.

O alta posibilitate de asigurare a confidentialității este aceea a accesului la datele din baza de date doar prin intermediul unor programe de aplicație. Utilizatorii acestor programe nu sunt în același timp și utilizatori înregistrați ai SGBD-ului care gestionează datele iar porțiunea din baza de date la care au acces este cablata în program.