

CAPITOLUL 1. ORGANIZAREA DATELOR ÎN BAZE DE DATE

1.1 Noțiuni generale

Dezvoltarea rapidă a industriei software și a calculatoarelor electronice a condus la lărgirea sferei de aplicații ale acestora necesitând stocarea unui volum tot mai mare de date. Astfel, pentru a îndeplini aceste cerințe, s-a impus perfecționarea modului de organizare și structurare a datelor care să permită accesul foarte rapid la datele stocate, în scopul regăsirii și prelucrării acestora.

Accesul rapid la date trebuie să rezolve problemele legate de viteza de regăsire a datelor, dar și de asigurarea accesului mai multor utilizatori la aceleași date (acces concurrent), pentru efectuarea unor operații diferite de prelucrare.

Problemele implicate de organizarea și regăsirea datelor se referă la:

- definirea, structurarea, ordonarea și gruparea datelor în colecții de date omogene;
- stabilirea legăturilor (relațiilor) dintre colecțiile de date;
- reprezentarea și stocarea datelor pe suporturi magnetice (organizarea fișierelor de date, a structurilor de tip index etc.).

Evoluția modului de organizare al datelor din punct de vedere teoretic și practic, s-a extins cu precădere asupra principiilor și metodelor de structurare, stocare și regăsire a datelor, în concordanță cu progresele înregistrate în domeniile hardware și software și cu cerințele tot mai diverse și exigente ale utilizatorilor acestor date. În acest context, au apărut și s-au dezvoltat concepte noi privind gestiunea datelor cu mijloace

electronice de calcul, printre care se înscriu cele de bază de date și bancă de date.

Pornind de la realitatea obiectivă, pentru a identifica datele care urmează să fie stocate și regăsite pentru prelucrare cu ajutorul calculatorului, precum și pentru a satisface cerințele specifice de organizare a acestor date este necesară modelarea acestei realități, folosind modelele de date.

Modelele de date transpun complexitatea informațiilor reale în reprezentări specifice, în scopul efectuării prelucrării cu calculatorul electronic. Pentru folosirea modelului de date adecvat este necesar să se efectueze un studiu și o analiză privind situația reală, pentru identificarea necesarului de informații și a modului de reprezentare, precum și a cerințelor de prelucrare a acestor informații. Modelele de date necesită identificarea *structurii modelului*, a *operatorilor* care acționează asupra structurilor de date și a *restricțiilor* (constrangerilor) pentru menținerea corectitudinii datelor.

Structura modelului presupune identificarea obiectelor reale din activitatea care urmează a fi informatizată (care sunt denumite entități), a caracteristicilor acestora (denumite atribute sau câmpuri de date) și a legăturilor dintre aceste obiecte (denumite relații sau asocieri).

Entitatea constituie unitatea informațională elementară din structura unei colecții de date prin care se reflectă elemente reale sau abstracte ce definesc segmente omogene ale realității obiective caracterizate prin nume unic, număr maxim de apariții (realizări) și o structură specifică acestei realități pe care o descrie.

Pe scurt, prin entitate se înțelege un conținut de sine stătător, o existență determinată (ca întindere, importanță, valoare etc).

De exemplu, într-o societate comercială pot fi definite entitățile: personal, utilaje, produse, materiale, secții, ateliere,

contracte, furnizori, beneficiari, facturi, note de intrare-recepție, etc.

Fiecare apariție reală a unei entități în cadrul unei colecții de date se numește realizare, formată, la rândul său, din elemente descriptive denumite atribute.

Atributul sau caracteristica este o proprietate a unei entități prin care aceasta poate fi descrisă din punct de vedere informațional ca o componentă a datelor. De exemplu, pentru entitatea personal se pot defini atribute ca: numele, data-nașterii, starea civilă, adresa, calificarea etc.

Fiecare atribut poate avea mai multe valori. De exemplu, pentru sexul persoanei se pot atribui valorile: bărbătesc și femeiesc, pentru starea civilă valorile: căsătorit, necăsătorit, divorțat și văduv.

Între entitățile unei colecții de date se pot stabili anumite legături (conexiuni) denumite relații.

Relația descrie o legătură specifică, delimitată logic, între două tipuri semnificative de entități sau chiar în interiorul entităților, prin care se definește gradul de implicare a unei entități în raport cu altele. De exemplu, astfel de relații se pot stabili între entități ca: utilaje și produse, contracte și furnizori, furnizori și facturi etc.

Pentru a stabili tipul legăturii, trebuie determinate numerele minime și maxime posibile de realizări ale unei entități, care pot fi asociate cu o realizare a unei alte entități cu care se află în legătură. O astfel de implicare reprezintă cardinalitatea relației (legăturii).

După gradul asocierii (valorile maxime identificate), relațiile din cadrul unei colecții de date pot fi de diferite tipuri și anume:

- de tipul 1 - 1, prin care se semnifică conexiunea simplă;
- de tipul 1 - n și n - m prin care se semnifică conexiunea multiplă.

După numărul de entități distincte ce participă la asociere, relațiile pot fi:

- relații binare, în care se realizează legături între două entități distincte;
- relații recursive, în care se realizează legături ale entităților cu ele însele;
- relații complexe, în care se realizează legături între mai mult de două entități distincte.

În funcție de modul în care sunt definite și reprezentate elementele prezentate anterior, pot exista mai multe tipuri de modelele de date, fiecare cu caracteristicile, avantajele și dezavantajele sale. Cele mai des întâlnite modele de date în practica curentă a dezvoltării aplicațiilor informatice sunt:

- modele ierarhice sau arborescente;
- modele rețea;
- modele relaționale;
- modele orientate obiect.

Modelele ierarhice (arborescente) sunt construite astfel încât au un nod (element) unic rădăcină și un număr finit de succesori, care la rândul lor au alte noduri succesoare, până la nodurile terminale, care nu mai au alți succesori. Orice nod, exceptând nodul rădăcină, are un predecesor unic. Relațiile care se stabilesc în cadrul acestor modele de date sunt de tipul 1-n.

Modelele de tip rețea sunt modele de tip graf cu legături bidirecționale între noduri. Un nod poate avea mai mulți predecesori și mai mulți succesori. În aceste modele se stabilesc relații de tipul m-n.

Modelele de tip relațional presupun organizarea datelor în tabele bidimensionale de date, care au fost aduse prin anumite metode specializate într-o formă optimă pentru exploatare.

În astfel de modele relațiile stabilite admit orice tip de relație: 1-1, 1-n, n-m.

Modelele orientate obiect redau structuri de date complexe ale realității, încorporând atributele care descriu entitățile împreună cu operațiile care se pot executa asupra acestor atribute într-o structură unică denumită clasă de obiecte. Aceste modele de date admit orice tip de relație: 1-1, 1-n, n-m.

În realizarea unei aplicații informatice se parcurg de regulă mai multe etape. Una dintre cele mai importante etape, de care depinde în mare măsură eficiența sistemului informatic, constă în modelarea sistemului. În acest sens se folosesc o serie de tehnici de modelare și proiectare a sistemelor informatice. Pentru modelarea datelor gestionate de către sistemul informatic, cea mai cunoscută metodă de modelare poartă denumirea de modelul Entitate - Atribut - Relație (EAR).

1.2 Noțiunea de bază de date

Noțiunea de bază de date a apărut în anul 1969, cu ocazia prezentării primului raport CODASYL (CONference of DATA SYstem Language), având ca idee centrală organizarea datelor într-un mod care să asigure independența organizării datelor față de programele de aplicații și a programelor de aplicații față de organizarea datelor. Accesul oricărui utilizator la baza de date se asigură printr-o descriere globală a datelor, care conține descrierea structurii colecțiilor de date și a legăturilor dintre acestea.

O colecție de date aflate în interdependență care reflectă un anumit aspect al realității și este destinată unui anumit grup de utilizatori împreună cu descrierea structurii lor și relațiilor dintre ele, formează o **bază de date**. Există mai multe posibilități de creare și gestionare a bazelor de date: manual (de exemplu, o agendă telefonică clasică) sau cu ajutorul calculatorului (o agendă telefonică în format electronic).

Operațiile pe care utilizatorii unei baze de date le pot efectua asupra datelor se clasifică în mai multe categorii:

- Introducerea de valori noi;
- Modificarea datelor existente;
- Ștergerea unora dintre datele existente;
- Interogarea bazei de date cu scopul de a regăsi informații conform unui anumit criteriu definit de utilizator.

Entitățile, atributele și relațiile stabilite între entități, descrise printr-un limbaj specializat în scopul memorării lor pe un suport tehnic de date adresabil (memorie externă, de regula hard-discul) constituie baza de date. Aceasta poate fi exploatată cu ajutorul unor programe specializate, denumite sisteme de gestiune a bazelor de date (SGBD).

Atât baza de date cât și SGBD-ul folosit la gestionarea și exploatarea ei devin operaționale și funcționează sub controlul sistemului de operare aferent calculatorului în scopul satisfacerii cerințelor generale și specifice ale utilizatorilor.

Una dintre caracteristicile esențiale care deosebesc o bază de date de modalitatea de organizare a datelor sub forma de fișiere constă în includerea în baza de date a descrierii structurii datelor. Prin această descriere se asigură independența datelor din baza de date față de programele de aplicații și invers. Independența datelor față de programe poate fi privită la mai multe niveluri:

- independență fizică - permite ca modalitatea și tehnicile de stocare a datelor (memorare a lor pe suport de memorie externă) să poată fi modificate fără a necesita rescrierea programelor de aplicații care exploatează baza de date;
- independență logică - permite adăugarea de entități noi sau extinderea/modificarea structurii bazei de date fără a mai fi necesară rescrierea programelor de aplicații care exploatează baza de date.

Un **sistem de baze de date** (database system) este un sistem de gestionare a unei anumite activități din lumea reala cu ajutorul calculatorului electronic. Ca dimensiune, sistemele de baze de date pot fi foarte diferite: de la câteva înregistrări (exemplul precedent al agendei telefonice) pana la zeci sau sute de milioane de înregistrări (de exemplu, într-un sistem de evidenta a clientilor unei companii de telefonie mobila).

Componentele unui sistem de baze de date sunt: hardware-ul, software-ul, utilizatori si datele (Figura 1.1.)

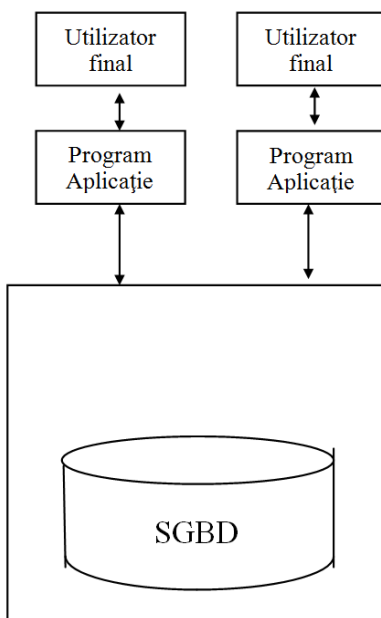


Figura 1.1. Componentele unui sistem de baze de bate

a. Hardware-ul este reprezentat de regula de calculatoare de uz general, de la calculatoare PC standard (desktop, laptop), până la stații multiprocesor cu putere de calcul foarte mare. Ceea ce interesează în mod deosebit în utilizarea unui

calculator pentru un sistem de baze de date este capacitatea memoriei externe, folosita pentru memorarea colecției de date persistente ale bazei de date. De regula, bazele de date se stocheaza pe unități de hard disc care asigura acces foarte rapid la date și capacitate mare de stocare. Pentru efectuarea unor copii de siguranță ale bazelor de date se folosesc unitățile de bandă magnetică sau discurile optice.

b. Software-ul este reprezentat în principal de Sistemul de Gestiune a Bazei de Date (SGBD). SGBD-ul este cea mai importantă componentă software a unui sistem de baze de date, dar nu este singura: orice SGBD este executat sub controlul sistemului de operare al calculatorului respectiv. Alte componente software care intra in alcatuirea unui sistem de baze de date sunt programe de proiectare si dezvoltare a programelor de aplicatii cu care se va exploata în mod curent baza de date.

c. Utilizatorii. Utilizatorii unui sistem de baze de date se pot împărți în urmatoarele categorii:

- *programatorii de aplicații* sunt cei care scriu aplicațiile de baze de date, folosind limbaje de programare de nivel înalt cu ar fi de exemplu C, C++, Java, Visual Basic;
- *utilizatorii finali*, adica acei utilizatori care accesează baza de date prin intermediul unui program de aplicație si au posibilități de execuție respectiv drepturi limitate de acces la date;
- *administratorul bazei de date* este un specialist in informatica si in special in baze de date și are ca sarcină menținerea funcționarii corecte si eficiente a bazei de date prin stabilirea drepturilor de acces a diferitelor categorii de utilizatori, indexarea/reindexarea bazei de date, efectuarea operațiilor periodice de salvare a datelor, etc.

d. Datele memorate într-o bază de date. Acestea reprezintă informațiile propriu-zise care descriu activitatea informatizată și sunt date persistente, adică ele rămân memorate pe suport magnetic, independent de execuția programelor de aplicații.

1.3 Clasificarea bazelor de date

Bazele de date și-au impus superioritatea față de fișierele clasice de date și au avut o evoluție rapidă. În prezent, practic nu există aplicație în domeniul economic care să nu folosească o bază de date. Diversitatea modurilor în care sunt utilizate bazele de date și a modelelor de organizare a datelor a condus la necesitatea clasificării bazelor de date după diferite criterii.

▪ După modelul datelor organizate în baza de date, se disting:

- baze de date de tip ierarhic sau arborescent, având o structură arborescentă în care există relații de subordonare de tip ierarhic (1 la n), astfel încât fiecare entitate subordonează una sau mai multe entități și este subordonată, la rândul ei, unei singure entități superioare. Astfel de baze de date s-au folosit în trecut, fiind prima generație de baze de date dar astăzi se întâlnesc foarte rar, având mai mult o importanță istorică;

- baze de date de tip rețea, care asigură legături de orice natură între entitățile bazei de date (1 la 1; 1 la n; n la m), astfel încât orice entitate poate fi legată direct cu alte entități în scopul reflectării unei corespondențe funcționale. Ele se bazează pe modelul rețea (un graf orientat alcătuit din noduri conectate prin arce), un model complex și dificil de folosit. Limbajele de gestiune a acestor baze de date efectuează prelucrarea secvențială a datelor, care este o prelucrare lentă. Un SGBD de tip rețea este IDMS (Integrated Database Management System), care dispune de un limbaj de descriere și

un limbaj de manipulare. Si aceste baze de date au disparut practic din utilizarea curenta;

- baze de date de tip **relațional**, care se bazează pe modelul relațional al datelor, au cunoscut o largă răspândire în prezent deoarece asigură independența organizării datelor față de programele de aplicații și invers, o redundanță minimă și controlată a datelor, un acces rapid la datele stocate și limbaje puternice de manipulare (limbaje procedurale și limbaje neprocedurale). Modelul de date relațional s-a impus prin simplitate și prin aria largă de aplicabilitate. Acest model care operează cu entități (tabele bidimensionale denumite și **relatii**), având structura formată din înregistrări (denumite și **tupluri**), **attribute** (coloanele tabelelor), domenii și valori. Numărul tuplurilor dintr-o relație reprezintă cardinalitatea relației, iar numărul valorilor dintr-un tuplu reprezintă gradul relației.

Produsul dintre cardinalitatea și gradul relației reprezintă numărul de valori reale stocate la un moment dat.

Atributul poate fi asimilat cu noțiunea de câmp de date întâlnită la fișierele de date. Numărul de attribute reprezintă un element invariabil determinat prin structura bazei de date. Atributul este identificat prin:

- nume, care, în general, corespunde semnificației valorilor din cadrul coloanei (atributului);
- tip, care reprezintă tipul de date pentru valorile din câmpurile de date;
- dimensiune, care reprezintă numărul maxim de poziții (caractere) rezervate pentru datele fiecărui câmp de date.

Suma dimensiunilor tuturor atributelor reprezintă dimensiunea relației.

La intersecția unei linii (unui tuplu) cu o coloană (un atribut) se găsește valoarea efectivă a câmpului de date. Toate valorile existente într-un tuplu formează o realizare a bazei de date.