

PARTEA IV.

PROIECTAREA MODELULUI DATELOR (II)

În volumul I - Analiza sistemului și identificarea soluțiilor de realizare am prezentat pe larg proiectarea modelului bazei de date relaționale pornind de la modelarea entităților prototipului cu ajutorul schemelor XSD (XML Schema Definition). Pentru o imagine completă asupra modelului proiectat, în Anexa 1 sunt prezentate schemele bazei de date relaționale.

Pornind de la modelul relațional, pentru a putea realiza analize avansate, multidimensionale, este necesar să proiectăm un depozit de date care să ofere performanța necesară implementării suportului decizional la nivel tactic și strategic pentru producătorii de energie din surse regenerabile.

Capitolul 3.

Proiectarea depozitului de date

3.1. Tehnologia depozitelor de date

În prezent, mediul economic este tot mai competitiv și mai complex, solicitând informații elaborate pentru sprijinirea deciziilor strategice. În acest context, putem privi depozitele de date ca fiind rezultatul interferenței mediului economic și al tehnologiilor informatice avansate. Creșterea volumului de informații, precum și perfecționarea tehnologiilor de exploatare a acestora au condus la o nouă calitate a folosirii datelor prin analize care pot releva conducerii organizației informații greu sau chiar imposibil de obținut pe alte căi. Se pot obține astfel informații privind preferințele clienților, profilul lor, distribuția etc. Astfel se pot furniza conducerii date, precum: în ce regiune a țării se vinde mai bine un anumit produs, care sunt preferințele unui anumit segment de piață etc. Este evident că astfel de informații nu se pot obține decât folosind anumite prelucrări cum ar fi analiza multidimensională, anumite metode statistice de

prognoză și alte metode matematice aplicate unui volum foarte mare de date din care se extrag numai datele relevante, celelalte fiind ignorate. Pentru astfel de aplicații, datele trebuie bine organizate și indexate pentru o ușoară regăsire și utilizare.

Conform [LUNG11], un depozit de date furnizează o sursă integrată și centralizată de date, separată de sistemul tranzacțional, care conține datele esențiale despre activitatea companiei din multitudinea de surse de date existente. Rapoartele obținute pe baza acestor date sunt utilizate ca un instrument de analiză strategic și competitiv, analizele rapide și corecte putând influența deciziile privind evoluția organizației pe termen mediu și lung.

Depozitele de date sunt organizate diferit față de bazele de date, fiind destinate mai mult extragerii de date decât actualizărilor repetate. Datele extrase sunt utilizate în analize dinamice care presupun schimbări de perspectivă asupra datelor și vizualizări ale acestora de la un nivel detaliat la unul sintetic, agregat și invers. Din acest motiv s-a impus o organizare specifică a datelor în care obiectele sunt structurate pe diferite niveluri care permit această analiză dinamică.

În urma cercetărilor efectuate în domeniu, în [BALU08], [LUVÉ08], [BAVE11], [LUNG11], precum și pornind de la definițiile consacrate din [INMO96], [OLAP95], [KIMB96], putem defini depozitul de date ca fiind un ansamblu de date de dimensiune foarte mare care este întreținut separat de bazele de date operaționale ale unei organizații și care este construit din date provenite din sisteme sursă prin extragere, filtrare, transformare și stocare în depozite speciale, în scopul sprijinirii proceselor decizionale. Depozitele de date sprijină prelucrarea informațiilor pentru analiză, furnizând o platformă solidă de consolidare a datelor istorice. Un depozit de date este un ansamblu de date consistente, din punctul de vedere semantic, care servește la o implementare fizică a unui model de date pentru sprijinirea deciziei și stochează informații pe care o organizație le solicită în luarea deciziilor strategice.

Un depozit de date reprezintă o modalitate de integrare și organizare a datelor din surse omogene și neomogene, provenite din sisteme tranzacționale dar și din fișiere externe, integrate după anumite criterii, supuse unui proces de extragere, transformare și încărcare, stocate agregat pe niveluri ierarhice, destinate prelucrărilor și analizelor dinamice, fiind soluția optimă de organizare a datelor pentru sistemele informatice suport de decizie și executive.

Spre deosebire de sistemele operaționale, structurile de date dintr-un depozit de date sunt optimizate pentru o regăsire și o analiză rapidă. Datele sunt istorice și sunt actualizate la intervale regulate de timp, în funcție de cerințele de raportare.

Rolul unui depozit de date este de a oferi o imagine coerentă asupra datelor, relativ la activitatea unei organizații și a contextului în care acesta acționează. Utilizarea acestei colecții poate consta din extragerea unor rapoarte (la cerere sau cu o anumită periodicitate), extragerea unor date pentru a fi utilizate de aplicațiile de birotică (programe de calcul tabelar, procesoare de text, programe de

prezentare etc.), dar mai ales pentru a fi utilizate de către aplicații specializate de analiză, precum tehnologia OLAP și tehnologiile pentru extragerea cunoștințelor din date (*data mining*).

Obiectivul principal al organizării datelor în depozite de date îl reprezintă asigurarea suportului pentru analize complexe și dinamice asupra datelor istorice și curente ale organizației. Pornind de la această premisă, am identificat o serie de obiective specifice, evidențiate în continuare și detaliate pe larg în [LUNG11]:

- *depozitul de date trebuie să asigure accesul la datele organizației*, într-un timp cât mai scurt, în funcție de cerințele utilizatorilor;
- *depozitul de date trebuie să permită utilizarea datelor direct în analize*, fără alte prelucrări suplimentare, ceea ce presupune existența unor procese de transformare și corectare a datelor în vederea asigurării calității acestora;
- *depozitul de date trebuie să stocheze date istorice*, deoarece acestea sunt esențiale pentru efectuarea de analize economice. Orizontul de timp acoperit trebuie să fie de cel puțin cinci ani, ajungând uneori la zece ani, în funcție de dinamica evoluției pieței și, deci, de relevanța datelor cu caracter istoric pentru nevoile analizei;
- *depozitul de date trebuie să stocheze date orientate pe subiectele importante ale procesului economic* (clienți, furnizori, produse, activități) și nu orientate pe aplicații, precum în cazurile sistemelor operaționale (care utilizează baze de date sau fișiere).

Pornind de la obiectivele menționate se desprind câteva caracteristici importante ale depozitelor de date prezentate în **Tabelul 1**:

Tabelul 1. Caracteristici ale depozitelor de date

Caracteristică	Explicație
proveniența datelor	<ul style="list-style-type: none"> • date importate din sistemul informatic operațional; • date de arhivă (în perioada de constituire a depozitului); • date externe (baze de date publice, date demografice, date statistice, date de prognoză economică, date obținute în urma unor sondaje de opinie etc.).
consistența datelor	<ul style="list-style-type: none"> • la două solicitări ale aceluiași set de informații, utilizatorul trebuie să primească aceleași date, chiar dacă ele au fost cerute la momente de timp diferite • dacă datele nu au fost complet încărcate atunci utilizatorul va fi avertizat cu privire la acest lucru și este sfătuit să aștepte până ce vor fi complet încărcate.

calitatea datelor	<ul style="list-style-type: none"> • este un factor determinant pentru procesul de analiză • se întâlnește frecvent situația în care datele nu sunt de bună calitate sau nu sunt extrase în întregime sau au un caracter incert din punctul de vedere al conținutului ceea ce face ca analiza ulterioară să conducă la rezultate eronate.
redundanța datelor	<ul style="list-style-type: none"> • în sistemul operațional redundanța este eliminată (prin procesul de normalizare) pentru a evita anomaliile de actualizare • în depozitul de date redundanța este creată în mod intenționat prin denormalizare și agregare pentru a permite un acces mai rapid la date.
integrarea datelor	<ul style="list-style-type: none"> • este o consecință importantă a realizării depozitului de date și, în cele din urmă, rațiunea pentru care acesta este creat • datele sunt încărcate pentru a răspunde nevoilor informaționale ale întregii organizații, asigurând faptul că rapoartele generate pentru diverse compartimente vor conține aceleași rezultate • integrarea datelor provenind din aceste sisteme dar și din alte surse se referă la aspecte precum: modalități unice de codificare a datelor, sistem de unități de măsură unitar, sistem stabil de reprezentare fizică a datelor, convenții clare privind modul de reprezentare a datelor, convenții unice privind denumirile câmpurilor de date.

Multe aplicații operaționale (tranzacții) presupun *actualizarea continuă* a colecțiilor de date (actualizare, modificare, ștergere). La depozitele de date, *actualizarea este foarte rară*, adică dinamica lipsește. Actualizarea se realizează aici doar prin adăugarea periodică a unor date extrase din sistemele operative sau din alte surse de date.

Pentru a obține informațiile dorite, depozitele de date sunt supuse unor *prelucrări complexe*, cu ajutorul unor metode specifice, cum ar fi: analiza multidimensională a datelor, metode statistice superioare de prognoză, metode matematice aplicate unui volum foarte mare de date. Aceste metode presupun folosirea unui *software* specializat deosebit de complex, bazat pe noi tehnologii informatice, precum cele prezentate anterior: extrageri de cunoștințe din date (*data mining*), OLAP (*Online Analytical Processing*), concentrări de date (*data mart*).

Sistemele care lucrează cu depozite de date dispun de o mare *flexibilitate*, ceea ce înseamnă o *conectivitate* la nivelul întregii organizații, astfel încât servere provenind de la furnizori diferiți să se poată conecta simultan la depozitul deja existent. Este, de asemenea, deosebit de important să se aleagă o arhitectură care să se adapteze ușor la modificările de performanțe, capacitate și conectivitate. Procesele de configurare, optimizare și administrare a sistemului, inclusiv procedurile de salvare-restaurare, precum și păstrarea în tot acest timp a funcționalității sistemului, pot deveni operații dificile dacă trebuie repetate la fiecare adăugare a unor noi servere în sistem.

Depozitele de date sunt destinate managerilor și analiștilor angrenați în luarea deciziilor strategice privind dezvoltarea și viitorul organizațiilor. Pentru aceasta, ei pot utiliza interfețe performante de accesare și analiză a datelor din depozite, adică *produse software* asociate depozitului de date.

Interfețele de tip Data Mining asigură extragerea și transformarea datelor în cunoștințe, de aceea uneori se consideră termenul *data mining* sinonim cu termenul *Knowledge Discovery in Databases* (KDD). Se utilizează tehnici ale analizei statistice superioare și de Inteligență Artificială care permit descoperirea de corelații, reguli, cunoștințe utile sprijinirii deciziilor.

3.2. Arhitectura pe componente a depozitelor de date

Arhitectura pe componente evidențiază componentele unui depozit de date și legăturile dintre ele: *depozitul de date, sursa de date, interfețele de analiză (Figura 1)*.

Depozitul de date conține mai multe tipuri de date care corespund diferitelor cerințe informaționale ale utilizatorilor:

- *datele detaliate* sunt cele relativ recente, livrate utilizatorilor, de regulă la nivel de execuție;
- *datele agregate*, deși determină o creștere a redundanței datelor, sunt necesare în depozitul de date deoarece în acest fel se poate asigura un timp mediu de răspuns cât mai redus. Aceste date presupun un grad de prelucrare prealabilă, astfel încât să fie pregătite pentru suport decizional și analize avansate: consolidare, totalizare, agregare, împachetare (în formate accesibile interfețelor de analiză utilizate). Tot aici se găsesc date având o anumită vechime (câțiva ani), în formă detaliată;
- *metadatele* (dicționarul de date) descriu datele conținute în depozitul de date și modul în care ele sunt obținute și stocate. Prin metadate se precizează structura datelor, proveniența lor, regulile de transformare, de agregare și de calcul. Ele sunt utilizate ori de câte ori se utilizează depozitul de date: la încărcarea datelor, la consultare, la actualizare, adică pe parcursul întregului ciclu de viață al depozitului.

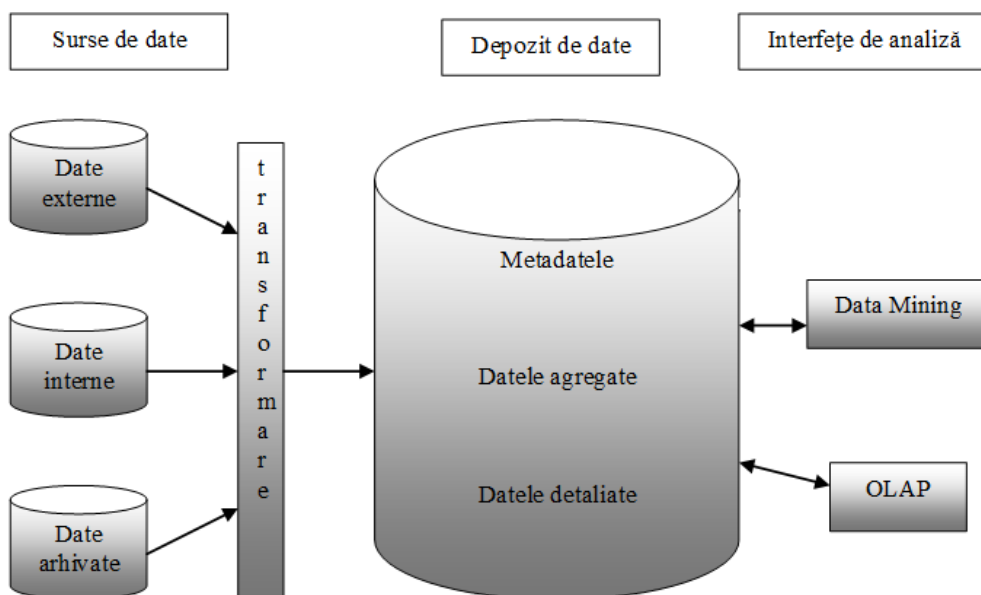


Figura 1. Arhitectura pe componente a depozitelor de date (sursa: [LUNG11])

Sursele de date pentru depozitul de date sunt: datele operaționale curente (baze de date și/sau fișiere din sistemul informatic operațional al organizației), datele vechi arhivate, datele externe (baze de date și fișiere din sistemele informatice ale altor organizații).

Construirea depozitului de date, pornind de la sursele de date, presupune parcurgerea unor *etape* în cadrul unui proces de *transformare*:

- *extragerea* datelor din datele operaționale sau din surse externe, urmat de copierea lor în depozitul de date. Acest proces trebuie, cel mai adesea, să transforme datele în structura și formatul intern al depozitului;
- *filtrarea* datelor, pentru a exista certitudinea că datele sunt corecte și pot fi utilizate pentru luarea deciziilor;
- *încărcarea* datelor corecte în depozitul de date;
- *agregarea* datelor: totaluri precalculate, subtotaluri, valori medii, sume etc., care se preconizează că vor fi cerute și folosite de utilizatori. Aceste agregări sunt stocate în depozitul de date împreună cu datele importate din sursele interne și externe.

Interfețele de analiză sunt produse *software* care implementează tehnologii informatice pentru extragerea și analiza datelor din depozitul de date: *data mining*, *OLAP*.

3.3. Proiectarea depozitului de date

Pornind de la schema bazei de date prezentată anterior, am proiectat schema de tip fulg de nea a depozitului de date. Aceasta activitate se realizează în următoarele etape: 1) *proiectarea dimensiunilor*; 2) *proiectarea tabelor de fapte*; 3) *realizarea schemei depozitului de date*.

1) Proiectarea dimensiunilor

În funcție de cerințele de analiză ale factorilor de decizie, se identifică principalele entități structurale care vor avea rol de organizare și selectare a datelor pentru explorare multidimensională. Acestea vor fi ierarhizate pe mai multe niveluri, astfel încât să fie permise operații tipice sistemelor OLAP: rotații, secțiuni, navigări în detaliu și agregări.

Dimensiunea DIM_INVESTITORI (**Figura 2**) caracterizează investitorii care au aplicat pentru avizele de racordare și care în urma contractării sunt proprietarii centralelor în funcțiune. Dimensiunea are la bază tabela virtuală D_INVESTITOR, care este creată pornind de la tabelatele relaționale T_INVESTITORI și T_LICENTA_COMERCIALE:

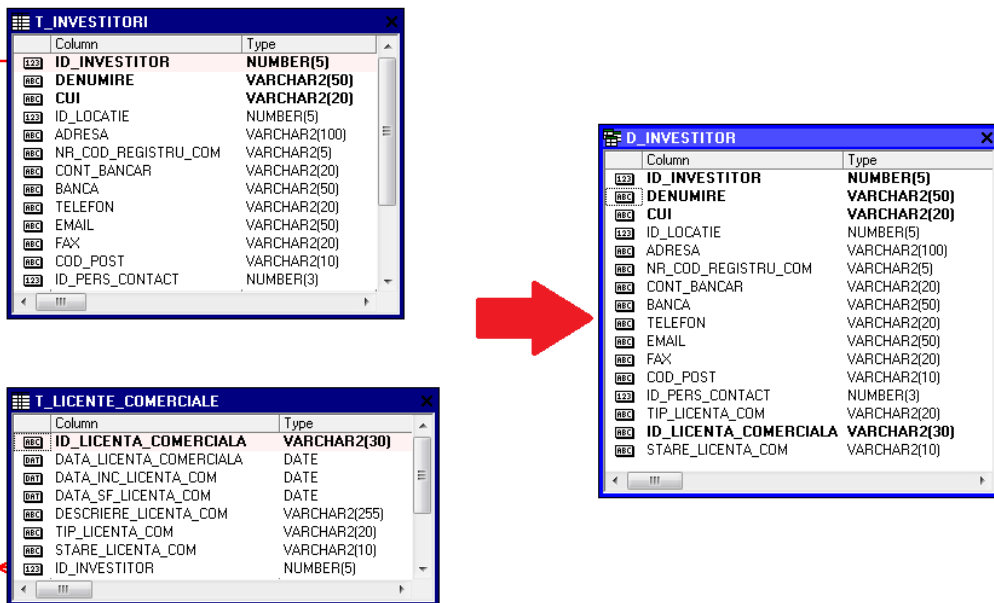


Figura 2. Tabela virtuală D_INVESTITOR

```
create or replace view d_investitor as
select I.ID_INVESTITOR,
       DENUMIRE,
       CUI,
       ID_LOCATIE,
       ADRESA,
       NR_COD_REGISTRU_COM,
```