

Integrarea tehnologiei relationale cu tehnologia OLAP

Lect.dr. Mihaela MUNTEAN

Catedra de Informatica Economica, A.S.E. Bucuresti

In most organizations the business intelligence applications are deployed using specialized databases (for example multidimensional databases) and operational applications are deployed using relational databases and SQL-based reporting tools. This separation between operational applications and BI applications is expensive because of the process of replicating data and the costs of maintaining additional hardware and database administrators. This paper discuss about the advantages of an integrated relational-multidimensional database. An integrated relational-multidimensional database causes a convergence of operational and business intelligence applications.

Keywords: Relational schema, star schema, multidimensional model, multidimensional database, business intelligence applications.

Conceptul de business intelligence

În pagina Web a firmei IBM, conceptul de *Business Intelligence* este definit astfel: „*BI înseamna utilizarea tuturor datelor de care dispune o firma, pentru a îmbunătăți procesul decizional. BI presupune accesul la date, analiza lor și descoperirea de noi oportunități de utilizare a lor*”. În climatul competițional al zilelor noastre este vital pentru organizații să ofere acces rapid la informații, cu costuri mici, pentru un număr cât mai mare și mai variat de utilizatori. Soluția la această problemă este un sistem BI (sistem informatic pentru inteligența afacerilor).

Principalele obiective ale unui sistem BI sunt: i) să permită soluții cu costuri scăzute care oferă avantaje firmei; ii) să permită acces rapid și ușor la informațiile firmei pentru un număr mare și variat de utilizatori; iii) să ofere suport pentru tehnologiile moderne (tehnici de analiză complexe-instrumente OLAP, instrumente de tip data mining etc); iv) să ofere un mediu de operare deschis și scalabil.

Factorii care au determinat dezvoltarea sistemelor informatice pentru inteligența afacerilor sunt:

- utilizarea depozitelor de date specific proiectate pentru a susține procesul decizional dintr-o organizație și optimizate pentru interogări on-line rapide și pentru agregări complexe;

- utilizarea Internetului ce permite unui număr mare de utilizatori să acceseze aplicațiile BI;

- creșterea substanțială a competiției între organizații; pentru a face față acestei competiții acerbe, managerii trebuie să ia decizii strategice corecte, bazate pe informații reale, informația devenind astfel un factor de producție esențial, iar utilizarea aplicațiilor BI o necesitate.

La ora actuală există o gamă variată de produse software pentru a dezvolta soluții BI.

- *Instrumentele de raportare.* Cele mai multe organizații utilizează instrumente de raportare nu foarte complexe, pentru a accesa datele stocate în depozitele de date. Aplicațiile BI construite cu astfel de instrumente: i) furnizează rapoarte statice sau parametrizate de stinătate unui număr mare de utilizatori din organizație; ii) oferă puține facilități analitice; iii) utilizează de regulă baze de date relationale (depozite de date/centre de date) și limbajul SQL.

- *Instrumentele de analiză și interogare ad-hoc.* Aplicațiile BI construite cu astfel de instrumente: i) permit un nivel ridicat de interacțiune cu utilizatorul (tehnici de navigare și selecție a datelor); ii) utilizează de regulă baze de date relationale și limbajul SQL; iii) oferă unele facilități analitice; iv) utilizează adesea modelul de date multidimensional (schema stea sau fulg de zapadă); v) permit

interogari „unidimensionale” cum ar fi „Care sunt primii 10 clienti?”.

▪ *Instrumentele OLAP (de analiza multidimensionala)*. Asa cum indica cuvintele folosite pentru a construi acronimul (on-line, analytic, processing) rolul instrumentelor OLAP într-o organizatie este de a oferi un acces interactiv si usor la resursele analitice necesare procesului decizional si de conduce-re. Instrumentele OLAP utilizeaza tehnici analitice simple (analiza multidimensionala a datelor) pentru a analiza seturi mari de date. Majoritatea specialistilor sunt de acord ca depozitele de date împreuna cu instrumentele OLAP ofera posibilitatea de a transforma cantitatile uriase de date din firme, în infor-matii utile procesului decizional. De aseme-nea, folosirea tehnicilor analitice oferite de instrumentele OLAP împreuna cu depozitele de date si facilitatile oferite de Web, permit un acces mai usor si mai rapid la informatiile necesare procesului decizional modern.

Cele mai multe aplicatii BI construite cu in-strumente OLAP (de exemplu, pentru analiza costurilor, analiza profitului, analiza vânzar-ilor, analize de marketing etc.) utilizeaza baze de date multidimensionale care permit ce-rieri multidimensionale complexe (de exem-ple „Care a fost procentul de crestere a vo-lumului vânzarilor anul acesta fata de anul trecut, pentru fiecare din primele 10 produse vândute?”). Utilizarea acestor baze de date este costisitoare si sunt destinate de regula managerilor si analistilor (deci unui numar restrâns de utilizatori).

La ora actuala exista doua tendinte în aplic-a-tiile BI:

1) *O corelare cât mai buna între activitatea de analiza si cea de planificare*. Aplicatiile BI trebuie: i) sa furnizeze acces rapid unui numar mare de utilizatori, distribuiti geogra-fic, la un stoc de date centralizat; ii) sa pe-r-mita agregarea complexa a datelor; iii) sa uti-lizeze facilitatile oferite de Internet. Existenta unui catalog de metadate centralizat si a unui stoc de date centralizat poate sa le ofere uti-lizatorilor o viziune comuna asupra datelor. Toti utilizatorii trebuie sa acceseze acelasi model de date si aceleasi date.

2) *Convergenta aplicatiilor BI cu aplicatiile operationale (tranzactionale)*. Activitatile o-perationale si cele de analiza constituie nu-cleul activitatii unei firme, independent de marimea ei, domeniul de activitate, forma le-gala. Informatiile despre vânzari, productie si costuri pot fi înregistrate si gestionate în una sau mai multe baze de date, folosite pentru scopuri operationale. Activitatile operation-ale se executa la un interval relativ constant. Datele sunt citite si actualizate frecvent si re-prezinta o fotografie curenta a ceea ce se în-tâmpla în firma. Fiecare cerere foloseste un volum mic de informatii, iar natura ei este în general previzibila. Monitorizarea, evaluarea, compararea, planificarea si alocarea strategi-ca a resurselor sunt exemple de activitati de analiza. Informatia generata prin activitatile de analiza este orientata pe decizie, deoarece este într-o forma care o face imediat utiliza-bila în procesul decizional. Datele sunt mai mult citite decât actualizate în aceste activi-tati. Cererile analitice folosesc date derivate si natura lor nu este întotdeauna previzibila. Ca urmare a acestor diferente, majoritatea firmelor folosesc instrumente diferite pentru cele doua tipuri de activitati: i) pentru a asi-gura eficienta maxima în ambele activitati; ii) pentru a realiza actualizare rapida în activita-tile tranzactionale si calcul rapid în activitati-le de analiza.

De asemenea, aplicatiile BI pot utiliza pentru stocarea datelor fie baze de date relationale (depozit de date/centru de date) sau baze de date specializate (de regula multidimension-ale). Stocarea datelor se face într-o baza de da-te relationala atunci când: i) volumul de date este prea mare pentru a fi duplicat; ii) datele sursa se modifica frecvent si este mai bine de a citi în timp real decât din copii; iii) integra-rea cu alte sisteme informatice relationale existente; iv) organizatia are o politica de neduplicare a datelor în alte structuri de fisie-re, pentru securitate sau alte motive, chiar daca aceasta conduce la aplicatii mai putin eficiente.

De asemenea, depozitele de date pot fi acce-sate de o mare varietate de instrumente de raportare si interogare ce utilizeaza limbajul SQL. Totusi limbajul SQL standard nu per-

mite operatii multidimensionale. Specialistii ofera trei solutii pentru a adauga functionalitate multidimensionala la datele stocate în baze de date relationale: i) integrarea procesarii multidimensionale în sistemul de gestiune a bazei de date relationale, fie prin extinderea limbajului SQL sau prin adaugarea functionalitatii multidimensionale în nucleul SGBD-ului (de exemplu Oracle 9i cu optiunea OLAP); ii) executarea în mai multi pasi a comenzilor SQL (de exemplu Microstrategy); instrumentul OLAP realizeaza o serie de comenzi SELECT, în care iesirile comenzilor anterioare sunt stocate în tabele temporale, care sunt apoi utilizate de comenzile urmatoare; iii) încarcarea datelor relevante din tabelele corespunzatoare pe un server de aplicatie intermediar, unde este realizata procesarea multidimensionala.

Bazele de date multidimensionale ofera facilitati analitice complexe, o performanta mai buna la interogare, dar întretinerea unei astfel de baze de date este costisitoare. O baza de date multidimensionala cere replicarea datelor, un proces costisitor care poate cauza si o întârziere semnificativa în utilizarea datelor de catre analisti. De asemenea, se cere modelarea separata a datelor (modelarea multidimensionala) si multe baze de date multidimensionale nu ofera facilitatea de recuperare a erorilor si alte facilitati specifice bazelor de date relationale. SGBDMD cer antecalcularea tuturor agregatelor posibile, astfel sunt adesea mai performante decât SGBDR traditionale, dar mai dificil de actualizat si administrat.

Stocarea fizica a datelor multidimensionale, precum si fenomenul de împrastiere sunt preocupari majore, în domeniul bazelor de date multidimensionale. O tehnica de stocare a datelor optima ar trebui sa tina cont de multi factori dinamici si anume: profilul datelor si volumul lor (numarul de dimensiuni si membrii ai dimensiunilor, tipuri de date etc), fenomenul de împrastiere, frecventa de modificare în sursele de date (cât de des vor fi actualizate bazele de date multidimensionale), frecventa de modificare în datele

multidimensionale (de exemplu pentru analiza de tip „what if”), frecventa de modificare în modelul multidimensional, accesul concurent etc.

În concluzie, utilizarea celor doua tehnologii separat implica: i) doua baze de date gestionate separat; ii) dictionare de metadate separate; iii) modele de securitate separate pentru BDR si BDMD; iv) instrumente de administrare separate; v) mai mult personal pentru administrare; vi) o interfata OLAP API special proiectata pentru a accesa baze de date multidimensionale; vii) de regula, aplicatiile care utilizeaza limbajul SQL nu pot interoga datele din bazele de date multidimensionale; viii) replicarea datelor din baza de date relationala în baza de date multidimensionala; ix) actualizarea datelor se face în doua etape, mai întâi depozitul de date, apoi baza de date multidimensionala; (acest proces este consumator de resurse si timp); de asemenea, creste perioada de timp între momentul actualizarii surselor de date si momentul actualizarii bazei de date multidimensionale; x) din perspectiva unui administrator de baza de date se utilizeaza trei tehnologii diferite (instrumentul de integrare a datelor, tehnologia relationala si tehnologia OLAP), trebuie gestionate trei dictionare de metadate si exista doua stocuri de date (depozitul de date si baza de date multidimensionala).

De exemplu, aplicatiile BI, construite pâna acum cu produse Oracle, utilizau: i) *Oracle Express Server/Oracle Personal Express*, un server OLAP (datele sunt stocate într-o baza de date multidimensionala); ii) *Oracle Express Administrator* un utilitar grafic pentru definirea, întretinerea si popularea bazelor de date multidimensionale Express; iii) *Oracle Express Relational Access Manager (RAM)* care permite ca *Express Analyzer* sa acceseze datele stocate într-o baza de date relationala (configurata cu *Oracle Relational Access Administrator-RAA*); iv) *Oracle Warehouse Builder* pentru a crea depozitul de date si pentru a se încarca date în depozitul de date. *Oracle Warehouse Builder* include si instrumentele ETI (figura 1).

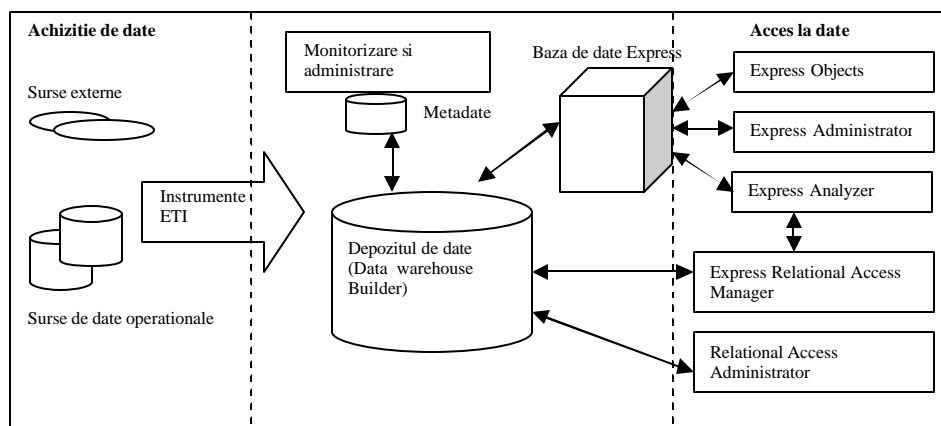


Fig.1. Arhitectura pe componente a unei aplicatii BI construita cu produse Oracle

Avantajele unui mediu integrat relational-multidimensional

Integrarea tehnologiei relationale (utilizata frecvent în aplicatiile operationale) cu tehnologia OLAP (utilizata frecvent în aplicatiile BI) ar permite realizarea de aplicatii cu o functionalitate ridicata si mai putin costisitoare, dar care ar oferi un suport adecvat atât pentru activitatea de analiza cât si pentru activitatile operationale.

Va exista un singur stoc de date de gestionat si un singur dictionar de metadate. De asemenea, trebuie implementata o singura politica de securitate. De exemplu, *Oracle 9i cu optiunea OLAP* realizeaza integrarea tehnologiei relationale cu tehnologia OLAP (figura 2). Oracle 9i utilizeaza baza de date relationala atât pentru aplicatii operationale cât si pentru aplicatii BI. Instrumentul *Oracle Enterprise Manager* si limbajul PL/SQL sunt utilizate pentru administrarea bazei de date Oracle.

Datele necesare analizelor sunt valabile 24 de ore din 24. Toti utilizatorii vad aceeasi versiune a datelor imediat dupa actualizare.

Datele pot fi stocate în tabele relationale (tabele de fapte si tabele de dimensiuni) sub forma de schema stea sau fulg de zapada sau în variabile multidimensionale. Dezvoltatorii de aplicatii pot utiliza OLAP API sau limbajul SQL pentru a accesa datele (stocate în tabele relationale sau variabile multidimensio-

nale). De exemplu *Oracle Reports* si *Oracle Discoverer* acceseaza datele prin SQL, iar *BI Beans* sau *JDeveloper* prin *OLAP API*. De asemenea, datele (stocate în tabele relationale sau variabile multidimensionale) sunt stocate fizic în fisiere de date Oracle.

Motorul multidimensional poate executa calcule multidimensionale (previziuni, scenarii „what-if”, manipularea seriilor de timp etc.) pe datele stocate în *spatiul de lucru analitic (analytic workspaces-AW)*. Limbajul de manipulare a datelor OLAP (OLAP DML) este echivalentul „multidimensional” al lui PL/SQL si extinde facilitatile analitice ale limbajului SQL. Atât OLAP API cât si PL/SQL pot include comenzi din limbajul de manipulare OLAP.

Un *spatiu de lucru analitic* contine un numar de obiecte multidimensionale (dimensiuni, variabile, relatii, programe, formule, modele etc.) si poate fi gândit ca o schema multidimensională. Într-o baza de date se pot crea mai multe *spatii de lucru analitice*. Un *spatiu de lucru analitic* apartine unui utilizator si alti utilizatori pot avea acces la el. De asemenea, un *spatiu de lucru analitic* poate fi temporar (pe durata sesiunii de lucru) sau poate fi persistent (stocat ca LOB în tabelele bazei de date). *Spatiile de lucru analitice* sunt o alternativa pentru stocarea datelor agregate ca si viziunile materializate.

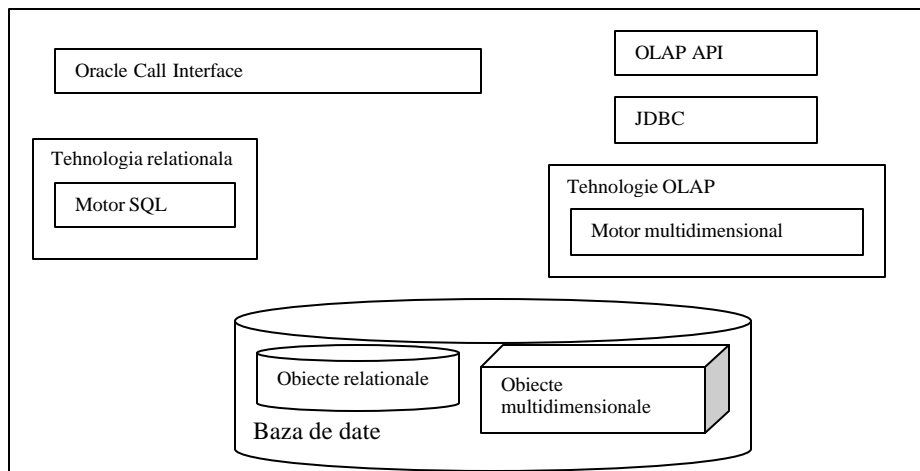


Fig. 2. Integrarea tehnologiei relationale cu tehnologia OLAP în Oracle9i

Datele nu trebuie replicate în doua stocuri de date, astfel se reduce intervalul de timp între momentul actualizării surselor de date și momentul când datele pot fi analizate. De asemenea, toți utilizatorii au acces la aceeași versiune de date.

Oracle OLAP API este o interfață Java API ce permite conexiunea la baza de date, operații de navigare (drill down, pivot etc.), selecția datelor, facilitati de analiza a datelor etc. *OLAP API* realizează conexiunea la baza de date prin *JDBC*. Pentru a accesa datele, *OLAP API* utilizează dicționarul OLAP. *Oracle 9i OLAP* transformă toate cererile OLAP API în cereri SQL. *Clienții SQL* se pot conecta la baza de date prin *Oracle Call Interface (OCI)* sau prin *JDBC*.

Oracle 9i OLAP utilizează un model de date multidimensional. Elementele modelului multidimensional sunt: măsurile, dimensiunile și atributele. Metadatele ce descriu modelul multidimensional utilizat de *Oracle 9i OLAP* sunt salvate în dicționarul OLAP și sunt create cu instrumentul *OLAP Management* integrat în *Oracle Enterprise Manager*. Aplicațiile pot accesa opțiunea OLAP prin: *Java OLAP API*, limbajul SQL sau comenzi din limbajul de manipulare OLAP.

Bibliografie

- [OOUG02] Oracle Corporation, *Oracle 9i OLAP User's Guide* release 9.2.0.2, 2002
 - [OODG02] Oracle Corporation, *Oracle 9i OLAP Developer's Guide to the OLAP DML*, release 9.2.0.2, 2002
 - [CNCS01] I. Lungu, M. Velicanu, M. Muntean, S. Ionescu, *Studiul și analiza comparativă a metodelor, tehnicilor și instrumentelor utilizate în domeniul "Inteligenței afacerilor"*, *Soluții economice și informatice pentru inteligența afacerilor*, CNCSIS, 2001
- <http://www.oracle.com>