

PREZENTARE GENERALĂ A OLAP (ONLINE ANALYTICAL PROCESSING)

BALAN Denis

Universitatea Tehnică a Moldovei

Abstract: OLAP (On-Line Analytical Processing) este o tehnologie utilizată pentru a organiza bazele de date foarte mari ale firmelor și a oferi suport decizional pentru afaceri. Bazele de date OLAP sunt împărțite în mai multe cuburi, iar fiecare cub este proiectat de un administrator de cub pentru a se potrivi cu modul în care regăsiți și analizați date, astfel încât să fie ușor să se utilizeze rapoartele PivotTable și PivotChart necesare. Bazele de date OLAP facilitează interogările de informatică decizională. OLAP este o tehnologie de baze de date care a fost optimizată pentru interogare și raportare în loc de procesare a tranzacțiilor. Datele sursă pentru OLAP sunt bazele de date OLTP (Online Transactional Processing), care se stochează, de obicei, în depozite de date.

Cuvinte cheie: OLAP, business intelligence, OLTP, non-OLAP.

1. Introducere

Conceptul de On-line Analytical Processing a apărut începând cu anii 60-70 din dorința de a modela prin funcții analitice activitățile financiare. Primul limbaj multidimensional, A Programming Language (APL) a fost dezvoltat de firma IBM și utilizat pe mainframe-uri încă din 1962, multe din conceptele acestuia fiind și astăzi implementate în unele limbaje, cum ar fi Adaytum Planning și Lex 2000.

2. OLAP Cube

Ce este business intelligence?

Un analist economic dorește deseori să obțină o imagine de ansamblu a afacerii, să vadă tendințele generale pe baza datelor agregate și să vadă acele tendințe în funcție de orice număr de variabile. Informatica decizională de afaceri (Business Intelligence sau inteligența de afaceri) reprezintă procesul de extragere a datelor dintr-o bază de date OLAP și analizarea lor pentru a găsi informații care pot fi utilizate pentru a lua decizii de afaceri în cunoștință de cauză (figura 1).

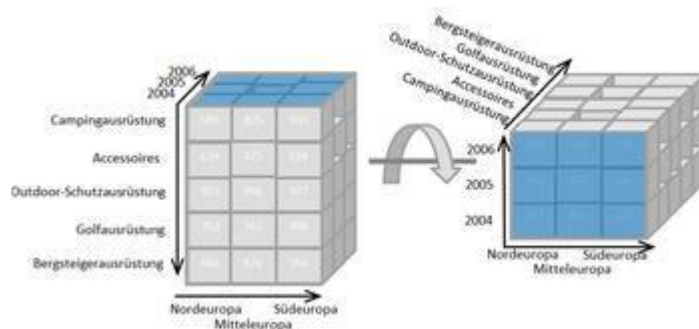


Fig. 1. Componentele OLAP

3. Caracteristici generale

Tehnologia OLAP este caracterizată de o dinamică analiză multidimensională în sprijinul utilizatorului final printr-o serie de activități:

- Aplicarea de formule și modele asupra dimensiunilor și ierarhiilor;
- Previziuni pe perioade diferite de timp;
- Analiza în adancime (drill-down);
- Extragerea unui subset de date pentru vizualizare;
- Rotații în cadrul dimensiunilor.

4. Caracteristici de bază ale sistemelor OLAP

Regula 1: O viziune conceptuală multidimensională

Viziunea conceptuală a modelelor OLAP trebuie să fie multidimensională, bazată pe viziunea sau modelul existent în organizație.

Regula 2: Manipularea intuitivă a datelor

Sistemele OLAP trebuie să permită operații intuitive și flexibile de manipulare a datelor, cum ar fi navigarea pe nivelurile ierarhiilor (operații de drill down, drill up, drill across), analize pe secțiuni din date, etc.

Regula 3: Accesibilitate

Sistemele OLAP trebuie să ofere acces la o singură viziune logică a datelor din organizație. Sursele de date, în modelul OLAP, trebuie să fie transparente utilizatorilor.

Regula 4: Surse de date variate

Un sistem OLAP trebuie să fie capabil să lucreze cu date stocate atât în baze de date multidimensionale (MOLAP), cât și în baze de date relaționale (ROLAP), sau chiar sisteme hibride (HOLAP).

Regula 5: Modele de analiză OLAP

Sistemele OLAP trebuie să suporte patru modele de analiză: explicativ, direct, contemplativ și formativ în sensul că un model trebuie să permită cel puțin realizarea rapoartelor parametrizate, analize de tip "ce se întâmplă dacă?", operații de tip drill-down/roll-up și slice/dice.

Regula 6: Arhitectura client/server

Orice sistem OLAP ar trebui să fie bazat pe o arhitectură client/server, oferind accesul utilizatorilor prin intermediul unui client, iar prelucrarea multidimensională să fie realizată de un server specializat.

Regula 7: Transparența

Accesul la sursele de date eterogene ar trebui să fie transparente pentru utilizatori, iar analiza datelor să poată fi realizată și prin intermediul diverselor instrumente client, ca: grafice, calcul tabelar, procesoare de text, etc.

Regula 8: Suport multiutilizator

Sistemele OLAP trebuie să asigure acces concurrent și distribuit la sursele de date, fiind asigurate însă integritatea și securitatea acestora (figura 2).

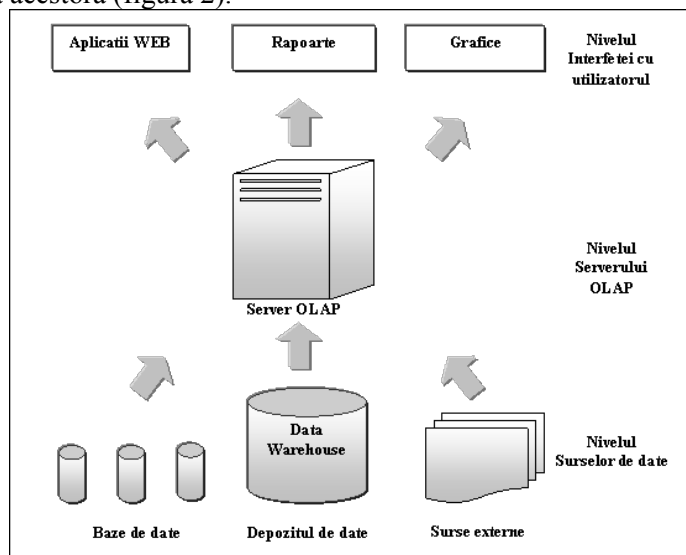


Fig. 2. Arhitectura Sistemelor OLAP

5. Structura și secțiunile componentelor

Cub - o structură de date care efectuează agregarea măsurilor după nivelurile și ierarhiile fiecărei dimensiuni pe care doriți să o analizați. Cuburile combină mai multe dimensiuni, cum ar fi timpul, geografia și linia de produse cu date sintetizate, cum ar fi cifrele de vânzări sau de inventar. Cuburile nu sunt "cuburi" în sensul

strict matematic, deoarece ele nu au în mod necesar laturile egale. Ele reprezintă un mod bun de a descrie un concept complex.

Măsură - un set de valori dintr-un cub care se bazează pe o coloană din tabelul de date și sunt de obicei valori numerice. Măsurile sunt valorile centrale din cub care sunt procesate, agregate și analizate. Exemple obișnuite includ vânzările, profiturile, veniturile și costurile.

Membru - un element într-o ierarhie care reprezintă unul sau mai multe apariții a datelor. Un membru poate fi unic sau acceptat. De exemplu, 2007 și 2008 reprezintă membri unici în nivelul de an de o dimensiune de timp, în timp ce ianuarie reprezintă membru în nivelul de lună, deoarece poate fi mai mare în dimensiune de timp în cazul în care conține date de mai mult de un an.

Membru calculat - un membru al unei dimensiuni a cărui valoare este calculată la lansarea în execuție utilizând o expresie. Valorile membrilor calculați pot fi derivate din alte valori de membri. De exemplu, un membru calculat, Profit, poate fi determinat scăzând valoarea membrului Costuri din valoarea membrului Vânzări.

Dimensiune - un set care conține o ierarhie sau mai multe ierarhii de niveluri organizate dintr-un cub, pe care utilizatorul le utilizează ca bază pentru analiza datelor. De exemplu, o dimensiune geografică poate include niveluri pentru țară/regiune, județ și oraș. O altă dimensiune poate include o ierarhie cu niveluri în funcție de an, trimestru, lună și zi. Într-un raport PivotTable sau PivotChart, fiecare ierarhie devine un set de câmpuri care poate fi extinsă sau restrânsă pentru a afișa niveluri superioare sau inferioare.

Ierarhie - o structură logică de tip arbore care organizează membrii unei dimensiuni astfel încât fiecare membru să aibă un membru părinte și niciun descendent sau mai mulți descendenți. Un descendent este un membru de nivelul inferior următor într-o ierarhie care este asociată direct membrului curent. De exemplu, într-o ierarhie Timp care conține nivelurile Trimestru, Lună și Zi, ianuarie este descendentul lui Trim1. Un părinte este un membru în următorul nivel superior într-o ierarhie care este asociată direct membrului curent. Valoarea părinte este de obicei o centralizare a valorilor tuturor descendenților. De exemplu, într-o ierarhie Timp care conține nivelurile Trimestru, Lună și Zi, Trim1 este părintele lui ianuarie (figura 3).



Fig. 3. Trei dimensiuni și tabela "fact_table"

6. De ce se utilizează OLAP Cube?

Tipic, organizațiile creează cuburi diferite pentru tipuri de date diferite. Într-o astfel de structură, pe măsură ce numărul de dimensiuni crește în timp, numărul de celule crește exponențial. Pentru a menține eficiența și viteza cererilor OLAP, datele din cub sunt presumarizate în subtotaluri (agregații). Datele presumarizate sunt stocate în agregatii, posibile la fiecare intersecție a unui nivel de dimensiune.

7. Diferențe de caracteristici între surse de date OLAP și non-OLAP

Regăsirea datelor - un server OLAP returnează datele noi pentru Excel de fiecare dată când se necesită modificarea aspectului raportului. De asemenea, există alte opțiuni pentru reîmprospătarea raportului.

Tipuri de câmp - datele sursă OLAP, câmpurile dimensiune pot fi utilizate doar ca rând (serie), coloană (categorie), sau câmpuri de pagină. Câmpurile de măsură pot fi utilizate doar ca câmpuri de valori. Pentru alte tipuri de surse de date, toate câmpurile pot fi utilizate în orice parte a unui raport.

Calculare - serverele OLAP furnizează valori rezumate direct pentru un raport, astfel încât să nu se poată schimba funcțiile de rezumare pentru câmpuri de valori. Pentru alte tipuri de surse de date, se pot modifica funcțiile de rezumare pentru un câmp valoare și se pot utiliza mai multe funcții rezumat pentru același câmp valoare. Nu se pot crea câmpuri calculate sau elemente calculate în rapoarte cu sursa de date OLAP.

8. Compatibilitatea cu limbajul SQL

Prin definiție, limbajul PL/SQL reprezintă extensia procedurală a limbajului SQL. PL/SQL suportă toate comenzile limbajului de manipulare a datelor (DML) ale limbajului SQL, precum și comenzile pentru controlul tranzacțiilor, funcțiile și operatorii SQL. Limbajul PL/SQL nu suportă comenzile de definire a datelor, comenzile pentru controlul sesiunii și comenzile pentru controlul sistemului ale limbajului SQL (Figura 4).

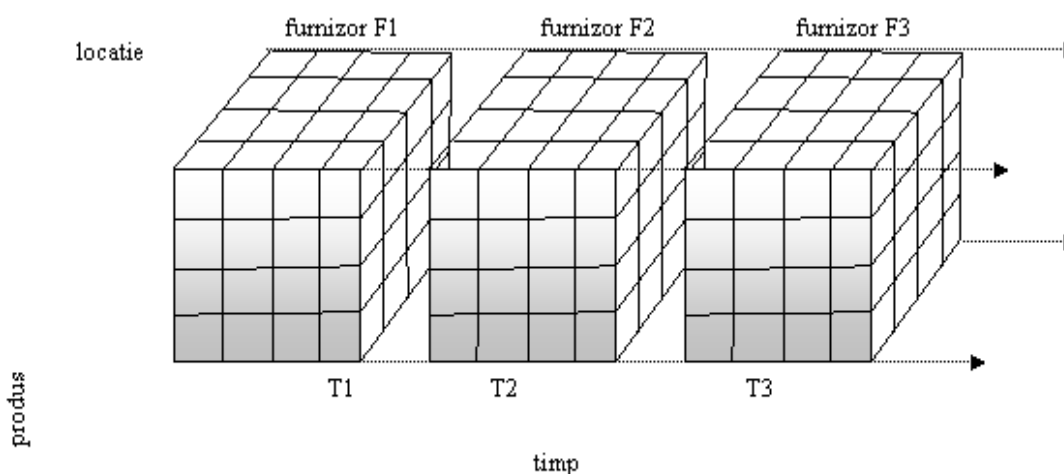


Fig. 4. Cub de date cu patru dimensiuni

Concluzii

Necesitatea folosirii modelării multidimensionale apare acolo, unde încetează „puterea” modelului relațional.

Deși bazele de date relaționale sunt extrem de flexibile în prezentarea informației, utilizatorii interesați să extragă informații care să-i ajute în adoptarea deciziilor vor să primească o imagine de ansamblu asupra activității, și nu să vizualizeze datele la nivel de înregistrări. Bineînțeles că SGBD relaționale pot oferi o imagine de ansamblu, dar consumul de timp și de resurse conduce la o slabă eficiență în cazul utilizării acestei soluții.

Bibliografie

1. *Providing OLAP to User Analysis: An IT Mandate*, [Regim de acces]: <https://support.office.com/ro-ro/article/Prezentare-general%C4%83-a-OLAP-Online-Analytical-Processing>
2. *Orlikowski w. J. - The Duality of Technology. Rethinking the concept of technology in organization, Sloan School of Management Working paper, No. 3141, MIT 1990*, [Resursă electronică]. Regim de acces: <http://www.scribub.com/stiinta/informatica>
3. *Power D.J. - Decision Support Systems: Concepts and Resources, Cedar*, [Resursă electronică]: Regim de acces: [https://ro.wikipedia.org/wiki/Cub_\(Data_Warehouse\)](https://ro.wikipedia.org/wiki/Cub_(Data_Warehouse))
4. *Thomsen E. - OLAP Solutions: Building Multidimensional Information Systems, John Wiley&Sons, New York, 2002, second edition*, [Resursă electronică]: Regim de acces: <https://dssresources.com/dssbook>