

Abordare comparativa Oracle 9i - SQL Server 2000 din punct de vedere al îndeplinirii regulilor lui Codd pentru sistemele OLAP

Prep. Daniela -Ioana SANDU
Catedra de Informatica Economica, A.S.E. Bucuresti

The first database management systems appeared in the late 60s as programs packages designed to solve a certain type of problem, extended then to solve problems of general interest. The final result of this mixture was a collection of complex products, difficult to understand, install, maintain and use. Access to this type of systems meant previous preparation of the users by computer specialists – database administrators and developers. The relational model and Codd's 12 relational rules appeared to remedy the lacks and deficiencies described previous. The 80s are characterized by growth of the data volume manipulated inside an economic organization, growth in the exactingness and quality of the data analysis therefore the appearance of the multidimensional OLAP systems. For these systems Codd defined a set of 12 rules used to help evaluate the efficiency of an OLAP product.

Keywords: *Multidimensional model, reality compliance, conceptual vision, transparency, accessibility, reporting, dimensions, aggregation.*

Prezentul articol își propune efectuarea unei analize comparative între două sisteme de gestiune a bazelor de date intens utilizate în dezvoltarea de software: SQL Server 2000 și Oracle 9i, din punct de vedere al îndeplinirii criteriilor de eficiență a produselor OLAP.

În cele ce urmează, se face o prezentare succintă a regulilor lui Codd privind sistemele OLAP:

R1. Viziunea conceptuală multidimensională. Realitatea, cu procesele și fenomenele ei, are caracter multidimensional. Prin urmare și viziunea conceptuală a analistului asupra acesteia trebuie să fie multidimensională. Viziunea conceptuală multidimensională este redată de modelul OLAP multidimensional care oferă o reprezentare abstractă a realității. Modelul multidimensional se spune că *este conform* cu realitatea datorită *similitudinii și echivalenței* sale cu realitatea. O viziune conceptuală multidimensională facilitează și ușurează analiza și proiectarea întrucât datorită *conformității* cu realitatea este mai ușor de înțeles și mai intuitiv decât modelele cu o singură dimensiune.

R2. Transparența unui sistem OLAP se analizează:

a. *Din punct de vedere al sistemului de front-end* în care este încorporat sistemul OLAP. Pa-

chetele de front-end ale unui sistem OLAP sau ale unui SBGD contin aplicații care permit utilizatorilor accesul la date reprezentate multidimensional respectiv relational. Transparența în acest caz presupune ca adăugarea unui sistem OLAP la pachetul de aplicații de front-end nu trebuie să influențeze - favorabil sau defavorabil - funcționarea acestor aplicații și de asemenea nu trebuie să influențeze complexitatea sistemului de front-end pe ansamblu.

b. *Din punct de vedere al utilizatorului.* Sistemul OLAP trebuie să fie transparent pentru utilizator în sensul că acesta nu trebuie să simtă și să știe că datele de intrare în sistemul OLAP provin dintr-un mediu de baze de date omogen sau eterogen

R3. Accesibilitatea. Instrumentul OLAP trebuie să-și mapeze schema logică pe sursele de date eterogene, să acceseze datele și să realizeze conversiile necesare pentru a prezenta datele într-o formă singulară, coerentă și consistentă. Mai mult, nu analistul ci sistemul OLAP trebuie să știe de unde și din ce sisteme sursa se preia datele.

R4. Performanța la raportare. Pe măsura ce numărul de dimensiuni ale sistemului dimensional sau dimensiunea bazei de date crește, utilizatorul sistemului OLAP nu trebuie să

perceapa o scadere în performanta raportarilor.

R5. Arhitectura client-server. În practica, datele care fac obiectul analizei OLAP sunt stocate pe sisteme mainframe, pe servere de baze de date si accesate prin intermediul statiilor client. Prin urmare, sistemele OLAP trebuie sa fie capabile sa opereze într-un mediu client-server. În acest scop, componenta server a instrumentului OLAP trebuie sa fie suficient de inteligenta astfel încât clientii sa se poata conecta cu efort minim de programare si înțelegere. Serverul OLAP trebuie sa fie capabil sa mapeze schemele logice si fizice disparate în scopul construirii unei scheme conceptuale, logice si fizice comune.

R6. Dimensionalitate generica. Manipularile si calculele de date care se pot face într-o dimensiune pot fi executate de asemenea în orice alta dimensiune. Analistii utilizatori pot crea orice dimensiune cu caracteristicile dorite care pot participa ca operanzi în calcule.

R7. Manipularea dinamica a matricelor înprastiate rare. Înprastierea, adica lipsa celulelor de valori (lipsa unui procent din celulele de valori) dintr-un set de date este o caracteristica a distributiei datelor. Incapacitatea sistemului OLAP de a se adapta distributiei setului de date si a regla setul de date duce la imposibilitatea obtinerii unor operatii rapide si eficiente si deci scadere în performanta cererilor si rapoartelor solicitate de utilizatori. Schema fizica a instrumentului OLAP trebuie sa se adapteze modelului analitic creat pentru a oferi manipulare optima a matricelor înprastiate. Pentru orice matrice înprastiată data exista o singura schema fizica optima. Aceasta schema optima ofera maximum de eficienta memoriei si operabilitate matriceala – aceasta în cazul în care setul de date nu este pastrat în memoria cache. Unitatea fizica de baza a instrumentului OLAP trebuie sa fie configurabil pentru orice subset de dimensiuni disponibile, în orice ordine, pentru operatii practice în cadrul modelelor analitice largi. Metodele de acces fizic trebuie sa contina diferite tipuri de mecanisme cum ar fi: calculatii directe, arbori B si derivati, hashing.

R8. Suport multiutilizator. Instrumentele OLAP trebuie sa ofere acces concurrent la date (în cazul regasirilor si actualizarilor), integritate si securitate datelor.

R9. Operatii interdimensionale. Oricare doua celule, indiferent de dimensiunea în care apar în cadrul unui model analitic pot fi folosite într-o formula. Formulele nu sunt restrictionate la o singura dimensiune la un moment dat.

R10. Manipularea intuitiva a datelor. Operatiile modelului multidimensional trebuie sa se execute printr-o actiune directa într-o celula a modelului fara sa fie nevoie de utilizarea unui meniu sau secvente de pasi într-un ansamblu de obiecte de interfata.

R11. Raportare flexibila.

a. rândurile, coloanele, etichetele de pagina ale rapoartelor trebuie sa permita afisarea a de la 0 pâna la n dimensiuni fiecare, unde n este numarul total de dimensiuni ale modelului;

b. orice dimensiune afisata într-un rând, coloana, eticheta de pagina trebuie sa permita afisarea oricarui subset de membri componentii dimensiunii, în orice ordine.

R12. Numar nelimitat de dimensiuni si niveluri de agregare. Un sistem OLAP trebuie sa permita definirea unui numar nelimitat de dimensiuni, iar în cadrul unei dimensiuni definirea unui numar nelimitat de niveluri de agregare. Practica a aratat ca un model multidimensional, pentru a fi fiabil trebuie sa aiba între 15 si 50 de dimensiuni.

Din punct de vedere al tipurilor de cerinte, cele 12 reguli ale lui Codd privind sistemele multidimensionale se pot clasifica astfel:

1. Reguli ale modelului dimensional: R1, R6, R9, R12.
2. Reguli de raportare: R4, R7, R10, R11.
3. Reguli de arhitectura: R5, R8.
4. Reguli de transparenta si accesibilitate: R2, R3.

În tabelul 1 se prezinta pentru componentele OLAP ale celor doua SGBD-uri care fac analiza prezentului articol, principalele componente si instrumente folosite în vederea satisfacerii celor 12 reguli prezentate anterior.

Tabelul 1 – Conditii de satisfacere a regulilor lui Codd

Regula	SQL Server 2000	Oracle 9i
R1	Analysis Services	Warehouse Builder
R2	Analysis Services ofera ferestre, meniuri si itemuri de interfata independente de baza de date SQL Server 2000.	Warehouse Builder ofera ferestre, meniuri si itemuri de interfata independente de baza de date Oracle 9i.
R3	Data Transformation Services	Warehouse Builder
R4	Rapoarte OLAP disponibile în Analysis Services.	Rapoarte OLAP disponibile în Oracle Server 9i - OLAP Services Instance Manager.
	Indexurile utilizate pentru cresterea eficientei raportarilor : Arbori B, indexuri bitmap, indexuri partitionati, indexuri domeniu	Arbori B
R5	SQL Server200 SQL Client 2000	Oracle Server 9i Oracle Client 9i
R6	Restrictiile în definirea dimensiunilor, atributelor sau operatiilor care se pot face cu atributete tin de limitarile induse de numarul maxim de atribute, coloane ce se pot defini într-o tabela:	
	1024	1000
R7	Arbori B	Arbori B si derivati, hashing.
R8	Ofera acces multiutilizator la server prin componenta client.	
R9	Prin comenzile SQL care implementeaza legaturile dintre tabele, se pot face operatii cu doua atribute care provin din dimensiuni diferite. SGBD-uri au si limbaje proprii:	
	T-SQL	PL/SQL
R10	Crearea, manipularea si vizualizarea datelor se face prin simplu click de mouse.	
R11	Rapoarte OLAP disponibile în Analysis Services.	Rapoarte OLAP disponibile in Oracle Server 9i - OLAP Services Instance Manager. Pentru crearea de rapoarte, Oracle dispune de componenta Forms 9i si Graphs 9i.
	Rândurile, coloanele si etichetele de pagina afiseaza o singura dimensiune. Lungimea denumirii de coloana este restrictionata astfel:	
R12	128 caractere	30 caractere
	Numarul de dimensiuni si niveluri de agregare este limitat de spatiul fizic.	

Solutii OLAP (si nu numai) se pot dezvolta atât cu SQL Server cât si cu Oracle 9i, eficienta, stabilitatea si robustetea aplicatiilor dezvoltate depinzând mai mult de experienta dezvoltatorilor si a administratorului bazei de date decât de furnizorul bazei de date. Alegerea unuia dintre cele doua SGBD-uri se face tinând cont de raportul pret/performanta, de facilitatile si limitarile fiecaruia. În tabelul 2 se prezinta o scurta comparatie între cele doua SGBD-uri având în vedere aceste caracteristici.

Tabelul 2 – Comparatie între SQL Server 2000 si Oracle 9i

Caracteristica	SQL Server 2000	Oracle 9i
Lungimea numelui bazei de date	128	8
Lungimea numelui coloanei	128	30
Lungimea indexului	900	749
Numar maxim coloane pe index	16	32

Numar maxim coloane pe tabela	1024	1000
Lungime maxima rând tabela	8036	255000
1 licenta Enterprise Edition pentru 1 procesor	40000 USD. Versiunea care include si OLAP si DataMining ajunge la 60000-80000 USD.	20000USD
Platforme	Windows	Orice platforma

Din analiza datelor prezentate se deduce ca SQL Sever 2000 este un SGBD mai ieftin decât Oracle 9i, este usor de instalat, utilizat si administrat, dar Oracle 9i suporta toate platformele (nu numai cele Windows) si are un limbaj propriu mai puternic.

Bibliografie

1. E.F.Codd "Providing OLAP to User-Analysts: An IT Mandate".
2. Microsoft Developer Network (MSDN) – January 2002 (3 CDs).
3. Oracle 9i – On Line Documentation (2 CDs).
4. SQL Server 2000 Books Online
5. www.databasejournal.com