

Multidimensional data model of Oracle 9i OLAP

Lect.dr. Mihaela MUNTEAN

Catedra de Informatică Economică, A.S.E. București

This paper provides an overview to the multidimensional data model that is used by Oracle 9i OLAP.

Keywords: *multidimensional data model, dimension, relation, variable, composite.*

Introducere

Încă din anii '70-'80, s-au dezvoltat sisteme informatice ce au permis analiză multidimensională, înainte de a fi cunoscute sub numele de sisteme OLAP.

În 1972 funcțiile analitice și facilitățile de gestiune a datelor au fost integrate într-un limbaj, limbajul *Express*, un instrument multidimensional pentru aplicații de marketing. După 30 de ani, *Express* rămâne una din principalele tehnologii OLAP folosite, conceptele și modelul de date fiind neschimbate. În 1995 firma Oracle achiziționează *Express*, iar în 2002 lansează *Oracle9i Release 2 OLAP* care integrează toate facilitățile OLAP

oferite de *Express* în baza de date relațională Oracle.

Până la apariția lui Oracle 9i, firma Oracle utiliza pentru a dezvolta aplicații OLAP, *Oracle Express*. *Oracle Express* este un pachet software de tip sistem de gestiune a bazelor de date multidimensionale (SGBDMD) cu următoarele caracteristici: i) oferă un limbaj de manipulare a datelor foarte puternic; ii) utilizează modelul de date multidimensional; iii) utilizează o arhitectură client/server (Oracle Express Server) sau o arhitectură pe un singur nivel (Oracle Personal Express); iv) permite dezvoltarea de aplicații OLAP care pot fi executate utilizând un browser Web.

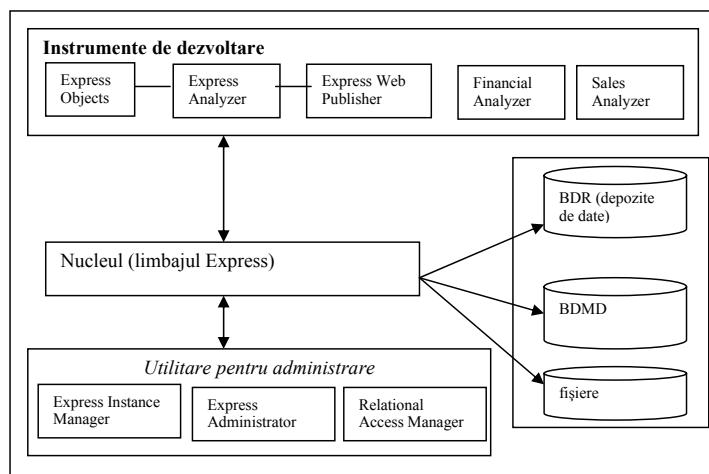


Fig.1. Arhitectura Oracle Express pe componente

Arhitectura pe componente a *SGBDMD-ului Oracle Express* este formată din (figura 1):

Utilitare pentru administrare (*Express Instance Manager*, *Express Administrator* și *Relational Access Manager*). *Oracle Express Administrator* este un utilitar pentru crearea, configurarea și instalarea bazei de date multidimensionale *Express*. Aplicațiile OLAP dezvoltate cu *Oracle Express* pot accesa di-

rect date stocate în baze de date relaționale (depozite de date) cu ajutorul lui *Relational Access Manager*. *Express Instance Manager* este un utilitar orientat Java care gestionează și configurează servicii și sesiuni de lucru.

Instrumente de dezvoltare. *Oracle Express Analyzer* și *Oracle Express Objects* permit crearea de aplicații OLAP. Aplicațiile OLAP dezvoltate cu aceste instrumente permit ana-

lize multidimensionale complexe (de exemplu analize de tip “What if” sau de tip “Top n”), au o performanță ridicată la interogare, dar sunt dificil de gestionat întrucât presupun: procese separate, metadate separate și fișiere de date separate. De asemenea, există o replicare a datelor (între depozitul de date relațional și baza de date multidimensională *Express*). Utilizatorii care vor să acceseze atât datele stocate în depozitul de date relațional, cât și datele stocate în baza de date multidimensională *Express*, trebuie să utilizeze două instrumente sau aplicații diferite (două limbaje diferite și anume limbajul SQL și limbajul *Express SPL-Stored Procedural Language*)

Elementele componente ale unei baze de date *Express* sunt: *variabilele, dimensiunile, relațiile, formulele, dimensiunile compuse (Express Server creează automat o dimensiune compusă, când se definește o variabilă împărțită), modelele, programele și seturile de valori*. Termenul de *măsură* este folosit pentru a referi variabilele, formulele și relațiile.

Modelul multidimensional utilizat de opțiunea OLAP din Oracle 9i

Opțiunea OLAP din *Oracle 9i Database Enterprise Edition* realizează integrarea tehnologiei relaționale cu tehnologia OLAP. Oracle 9i utilizează baza de date relațională atât pentru aplicații operaționale cât și pentru aplicații OLAP.

Opțiunea OLAP utilizează un *model de date multidimensional logic* ușor de înțeles de către utilizatori și care simplifică procesul de definire a calculelor complexe. Elementele principale ale *modelului multidimensional logic* sunt *variabilele, dimensiunile* (cu nivelurile și ierarhiile corespunzătoare) și *atributele* (figura 2). Acest *model multidimensional logic* poate fi transformat fie în *schema stea* a unui depozit de date relațional (tabelele de fapte și tabelele de dimensiuni) sau într-un *spațiu de lucru analitic*.

Dimensiunile (dimensions). Dimensiunile furnizează informații descriptive despre fiecare indicator (variabilă), conțin în general date statice și sunt esențiale pentru analiză. O

dimensiune conține mai mulți *membri*. De exemplu, toate lunile, trimestrele și anii formează dimensiunea *Timp*. Un membru poate aparține la una sau mai multe ierarhii (este un *nivel al ierarhiei*) sau poate să nu fie inclus într-o ierarhie (independent). De exemplu, în dimensiunea *Produce* membru “culoare” nu este inclus în nici o ierarhie (este de fapt un *atribut*).

Cele mai multe dimensiuni au o *structură multi-nivel sau ierarhică*. De exemplu, dimensiunea *Timp* este o dimensiune ierarhică (ore, zile, săptămâni, luni, trimestre și ani). Într-o dimensiune pot exista mai multe ierarhii. Alegerea nivelurilor intermediare de grupare sau agregare este importantă pentru înțelegerea și abilitatea de a lua decizii, deoarece datele au numai un nivel de detaliu, un nivel de agregare complet, dar multe niveluri intermediare. Conceptul de *nivel într-o ierarhie* este foarte important pentru a determina ce tipuri de navigări se pot executa în dimensiuni și ce tipuri de calcule suportă dimensiunile.

O dimensiune ce conține valori la toate nivelurile de agregare este cunoscută ca *dimensiune cu totaluri incluse* (embedded total dimension). De exemplu dimensiunea *Timp* cu totaluri incluse ar putea conține următorii membri: 2004, Q1.04, Q2.04, Q3.04, Q4.04, Jan04, Feb04, Mar04, Apr04, May04, Jun04, Jul04, Aug04, Sep04, Oct04, Nov04, Dec04.

Variabilele (measures). O variabilă este un indicator de performanță prin care se poate analiza performanța activității modelate. Tipul de dată al unei variabile indică datele pe care le conține. Fiecare variabilă este organizată sau dimensionată după una sau mai multe dimensiuni. Nu este necesar ca toate variabilele să aibă aceeași dimensionalitate (să utilizeze același set de dimensiuni). De exemplu, variabila *Vânzări* este dimensionată după dimensiunile *Produce, Locație* și *Timp*.

Pentru fiecare combinație posibilă între dimensiuni (produsul cartezian al dimensiunilor), există sau nu o valoare corespunzătoare a variabilei. Opțiunea OLAP asigură automat integritatea referențială între dimensiuni și variabile.

Un set de variabile ce utilizează aceleași dimensiuni de identificare (au aceeași dimensionalitate) formează un *cub n-dimensional*. Cum se pot afișa patru sau mai multe dimensiuni logice (ale cubului n-dimensional) în trei dimensiuni fizice (rând, coloană, pagină)?

Răspunsul este de a combina multiple dimensiuni logice în aceeași dimensiune fizică. De exemplu, vânzările pot fi afișate pe rând împreună cu produsele și locațiile, timpul pe coloane și clienții pe pagină.

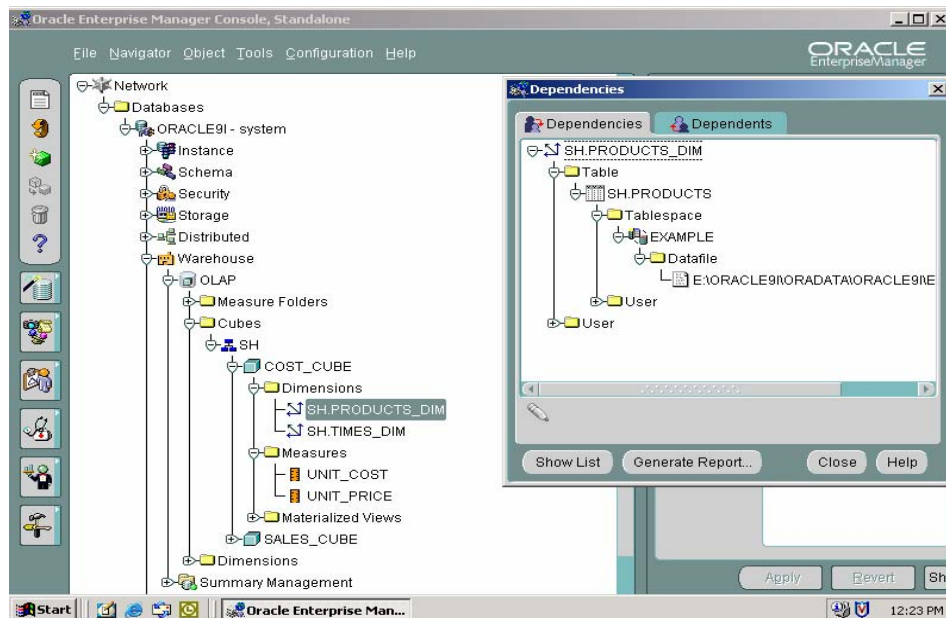


Fig. 2. Modelul de date multidimensional logic

În Oracle 9i termenul de bază de date se referă numai la baza de date relațională. Elementele componente ale *bazei de date multidimensionale Express* au fost incluse în Oracle 9i sub denumirea de *spațiu de lucru analitic (analytic workspaces-AW)*. La fel ca o schemă stea, un *spațiu de lucru analitic* este organizat după subiect (de exemplu vânzările). Un *spațiu de lucru analitic* conține un număr de obiecte multidimensionale și anume: *dimensiuni, variabile, relații, programe, formule, modele, seturi de valori etc.* Un *spațiu de lucru analitic* poate fi gândit ca o schemă multidimensională (sau *model multidimensional*) care :

- asigură integritatea referențială. De exemplu, dacă o variabilă are trei dimensiuni, atunci pentru fiecare valoare a variabilei corespunde un membru din fiecare dimensiune.
- asigură consistența datelor prin reutilizarea dimensiunilor cu mai multe variabile.

În timpul unei sesiuni de lucru, utilizatorul poate schimba ordinea membrilor unei di-

mențiuni sau poate selecta numai anumiți membri (selecție curentă). Selecția curentă se poate modifica prin adăugarea unor noi valori ale dimensiunii sau ștergerea altor valori.

Dimensiunile sunt primele obiecte care trebuie create într-un *spațiu de lucru analitic*.

O variabilă este un obiect din *spațiu de lucru analitic* ce stochează date, o matrice multidimensională ale cărei celule stochează date. De exemplu variabila *Vânzări* poate fi definită cu OLAP DML astfel:

```
DEFINE vanzari decimal <produs,
locatie, timp>
```

Relațiile (relations). O relație descrie corespondența între membrii unei dimensiuni și membrii altei dimensiuni sau ai aceleași dimensiuni (numită *relație cu ea însăși*). Se pot utiliza relațiile și pentru a defini ierarhiile în dimensiuni. Există patru tipuri de relații: i) *relații de tip părinte* (parent relation); ii) *relații de tip nivel* (level relation); iii) *relații de tip familie* (family relation); iv) *relații de tip atribut* (attribute relation)

O relație de tip părinte asociază membrii dintr-o dimensiune cu membru părinte corespunzător (din ierarhie). Relațiile de tip părinte sunt utilizate de motorul multidimensional pentru selecția datelor (de exemplu, pentru a găsi părintele sau copiii unui membru) și pentru agregări și alocări. De exemplu, o relație de tip părinte în dimensiunea *Timp* ar putea conține următoarele valori:

Timp	rel_parinte
2004	all times
Q1.04	2004
Q2.04	2004
Q3.04	2004
Q4.04	2004
Jan04	Q1.04
Feb04	Q1.04
Mar04	Q1.04
Apr04	Q2.04
May04	Q2.04
Jun04	Q2.04
Jul04	Q3.04
Aug04	Q3.04
Sep04	Q3.04
Oct04	Q4.04
Nov.04	Q4.04
Dec04	Q4.04

Time	All_times	Year	Quarter	Month
2004	all times	2004	NA	NA
Q1.04	all times	2004	Q1.04	NA
Q2.04	all times	2004	Q2.04	NA
Q3.04	all times	2004	Q3.04	NA
Jan04	all times	2004	Q1.04	Jan04
Feb04	all times	2004	Q1.04	Feb04
Mar04	all times	2004	Q1.04	Mar04
Apr04	all times	2004	Q2.04	Apr04
etc				

Relațiile de tip atribut stochează atribute neierarhice pentru un membru. De exemplu, o relație de tip atribut în dimensiunea *Timp* ar putea conține următoarele valori:

Timp	luna_din_an
Jan03	Jan
Feb03	Feb
Mar03	Mar
Apr03	Apr
Jan04	Jan

O relație de tip nivel este utilizată pentru a identifica un nivel de agregare dintr-o ierarhie (de exemplu să se selecteze toți membrii de la nivelul *luna*). De exemplu, o relație de tip nivel în dimensiunea *Timp* ar putea conține următoarele valori:

Timp	rel_nivel
2004	an
Q1.04	trimestru
Q2.04	trimestru
Q3.04	trimestru
Q4.04	trimestru
Jan04	luna
Feb04	luna
Mar04	luna
Apr04	luna
May04	luna
Jun04	luna
etc	

O relație de tip familie (family relation) conține toți strămoșii pentru fiecare ierarhie dintr-o dimensiune. De exemplu, o relație de tip familie în dimensiunea *Timp* ar putea conține următoarele valori:

Feb04	Feb
Mar04	Mar
Apr04	Apr
etc	

Dimensiuni compuse (composite). O dimensiune compusă este o dimensiune artificială ce combină membrii a două dimensiuni împreună și este utilizată pentru a gestiona fenomenul de împrăștiere a datelor și a reduce spațiul de stocare. În general, datele stocate într-un spațiu de lucru analitic sunt foarte

împrăștiate (există multe celule NULL, în special la nivelurile de detaliu ale modelului de date). De exemplu, variabila *Vânzări* poate fi dimensionată după *Locație*, *Produce*, *Client* și *Timp*. Se consideră că sunt 60 de luni, 2 magazine, 70 de produse și 100 de clienți, iar datele sunt 0.1 % dense. La nivel de bază ar trebui să fie 840000 de celule (60*2*70*100) din care 839160 ar avea valori nule. Se poate crea o dimensiune compusă din *Locație*, *Produce* și *Client*. Dimensiunea *Timp* și dimensiunea nou creată pot dimensiona variabila *Vânzări*:

```
DEFINE cl_prod_loc composite <client
produs locatie>
Define Vânzări variable decimal
<time cl_prod_loc <client produs
locatie>>
```

Motorul multidimensional gestionează automat dimensiunea compusă (de exemplu dimensiunea compusă se modifică automat dacă una din dimensiunile de bază sunt modificate).

Formulele. O formulă este un obiect utilizat pentru a stoca definițiile diferitelor expresii de calcul în *spațiile analitice*. Formulele sunt calculate la momentul execuției.

Programele. Un program conține secvențe de comenzi OLAP DML (un limbaj de programare utilizat pentru a crea, manipula și accesa datele multidimensionale dintr-un *spațiu de lucru analitic*, de fapt limbajul *Express SPL* îmbunătățit). Programele OLAP pot fi apelate din SELECT-SQL și PL/SQL.

Modelele. Un model este un obiect din *spațiul de lucru analitic*, care stochează un set de ecuații interdependente ce utilizează valorile unor variabile sau membrii unor dimensiuni. Modelele sunt folosite atunci când calculele sunt complexe sau variabilele nu sunt aditive. Ordinea ecuațiilor în model nu este importantă, întrucât opțiunea OLAP va determina automat dependențele și va stabili ordinea calculelor.

Bibliografie

- [BUEN02] Bud Endress, Oracle Corporation, *SQL Access to multidimensional data types in Oracle 9i Database release 2*, 2002
- [OOUG02] Oracle Corporation, *Oracle 9i OLAP User's Guide*, <http://otn.oracle.com>, 2002
- [OEOG99] Oracle Corporation, *Oracle Express Objects Getting Started*, Release 6.3.2, 1999
- [OEAG99] Oracle Corporation, *Oracle Express Analyzer Getting Started*, Release 6.3.2, 1999
- [OEDA99] Oracle Corporation, *Oracle Express Database Administrator Guide*, Release 6.3.4, 1999
- [MUNT04] Mihaela Muntean, *Inițiere în tehnologia OLAP. Teorie și practică*, Editura ASE, București, 2004