

Proiectarea modelelor de date multidimensionale conceptuale

Lect. Mihaela MUNTEAN,
Catedra de Informatica Economica, A.S.E. Bucuresti

If an organization is to achieve its goals in the current "Information Age", the work of managers must be supported by high quality information. Systems that process organizational performance data into managerial information have either multidimensional or relational data structures. This paper explores the effectiveness of multidimensional data models in providing information support for management decision making, with particular reference to a study of research performance information in universities.

Keywords: *multidimensional databases, relational databases, multidimensional model, decision support systems, research management, and information support.*

Introducere

În secolul 21, *secolul informatiilor si al afacerilor inteligente*, organizatiile vor putea sa-si îndeplineasca obiectivele numai daca managerii vor putea utiliza ca suport în procesul decizional informatii suficiente si de calitate. Cercetarea în domeniul managementului tactic (si al celui strategic în special), a confirmat ca procesele decizionale la aceste niveluri sunt în general dezordonate si nestructurate. Prezentarea informatiilor orientate pe subiecte (subiectele afacerilor) si nu dupa tipul aplicatiilor operationale, poate oferi un suport eficace procesului decizional la aceste niveluri. La ora actuala, preocuparea majora a specialistilor în domeniul sistemelor informatice este de a satisface cererea tot mai mare a managerilor pe ntru informatii care sa le permita evaluarea cât mai rapida si mai corecta a performantelor organizatiilor pe care le conduc. Sistemele suport de decizie orientate pe date pot satisface aceasta cerere de informatii de calitate si pot îmbunatati eficacitatea muncii decizionale a managerilor.

Tehnica de modelare entitate-asociere si structurarea datelor în tabele normalizate reprezinta standardul pentru specialistii de baze de date, care utilizeaza în mod frecvent bazele de date relationale pentru a stoca volumele mari de date tranzactionale existente în organizatii. Totusi, utilizarea tabelor pentru a furniza informatiile necesare procesului decizional la nivelul or-

ganizatiilor nu este întotdeauna o solutie ideala pentru manageri. Accesul la informatiile stocate în bazele de date relationale tranzactionale cere realizarea de operatii, adesea foarte complexe, între multe tabele. Din acest motiv, managerii trebuie sa apeleze la specialisti pentru a executa cererile de regasire (de exemplu în limbajul SQL). De asemenea, bazele de date relationale de dimensiuni mari, proiectate sa ofere suport aplicatiilor tranzactionale, permit cu dificultate vizualizarea informatiilor dupa subiectele de interes ale managerilor. Aparitia sistemelor suport de decizie orientate pe date (a sistemelor ROLAP sau MOLAP) a permis managerilor sa acceseze un volum mare de informatii integrate, sa le vizualizeze din diferite perspective si sa le analizeze on-line, cu scopul de a evalua cât mai obiectiv si mai exact performantele organizatiilor si de a realiza un proces decizional bazat pe analiza.

Pentru a realiza un suport decizional de calitate la nivelul organizatiilor, propunem ca *modelele conceptelor de afaceri*, care asigura o viziune consolidata asupra afacerilor, sa fie definite la nivel conceptual sub forma unor *cuturi de date n-dimensionale (hypercub)* (figura 1). Aceste modele descriu subiectele principale despre care organizatia trebuie sa colecteze informatii (ce informatii sunt necesare pentru a lua decizii mai bune?). Cerintele afacerii sunt o combinatie între cerintele

utilizatorilor, realitatile legate de sursele de date si realitatile economice. Definirea modelelor de afaceri sub forma de cuburi n-dimensionale/multicuburi permite analistilor/managerilor sa întelega mai bine datele din organizatii, date ce ar putea fi integrate, agregate, sa vizualizeze si sa

analizeze datele din multiple perspective (dimensiunile afacerii), sa puna în evidența noi aspecte ale afacerii, care ar putea oferi firmelor noi oportunitati si o îmbunătățire a eficacitatii procesului decizional si managerial (figura 2).

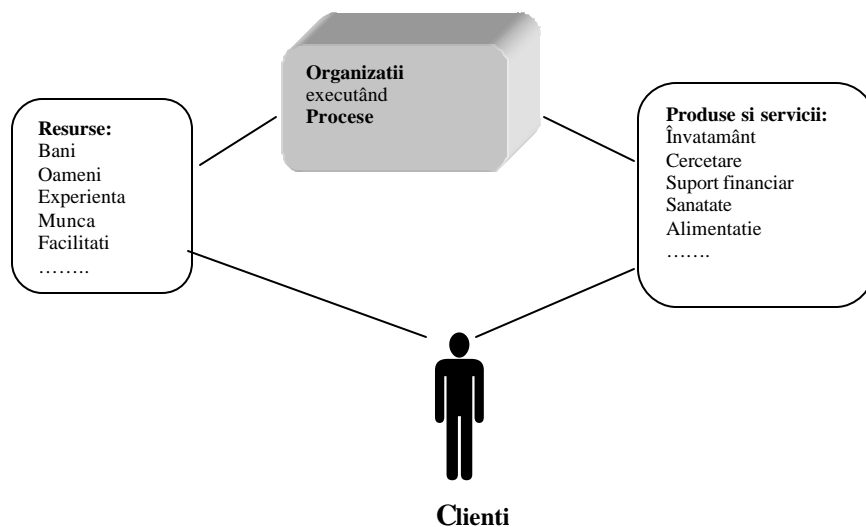


Fig.1. Modelul conceptelor de afaceri sub forma unui cub n-dimensional

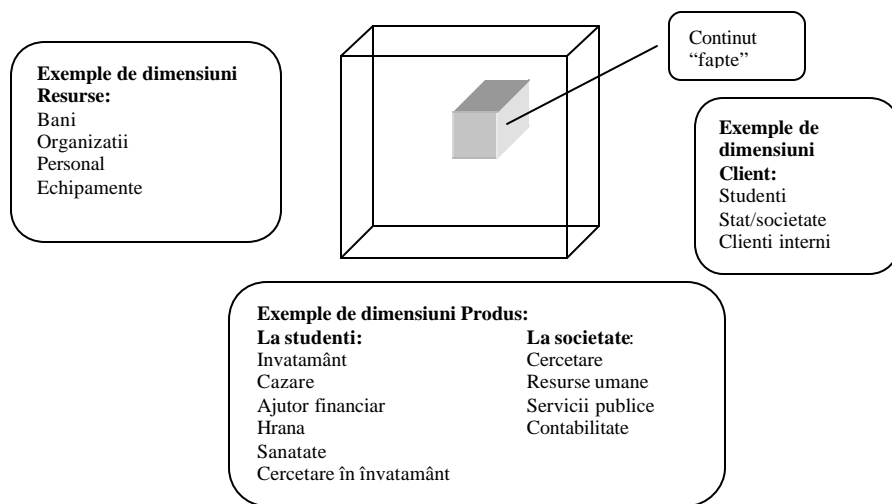


Fig.2. Procesul (problema analizata) sub forma unui "cub" de informatii

În figura 3 este prezentat un prototip de model de afaceri (sub forma unui multicub de informatii) pentru managementul si evaluarea activitatii de cercetare stiintifica în învățământul superior. Cerintele prototipului au fost determinate pe baza identificării indicatorilor sintetici ai activitatii de cercetare, precum si a problemelor le-

gate de managementul cercetarii, obiectivele si strategiile activitatii de cercetare stiintifica în învățământul universitar. Asigurarea calitatii în învățământul universitar presupune identificarea unor indicatori cantitativi complecsi care sa permita o evaluare mai buna a performantei cercetarii.

La ora actuala exista un numar mare de *indicatori de performanta* care pot fi utilizati pentru evaluarea performantei cercetarii stiintifice finantate si nefinantate în învățământul universitar, dintre care putem enumera urmatoorii:

i) numarul de carti publicate de fiecare cadru didactic (în editura ASE, în tara, în strainatate);

ii) numarul de capitole din carti publicate de fiecare cadru didactic;

iii) numarul de comunicari la manifestarile stiintifice (nationale si internationale) publicate de fiecare cadru didactic;

iv) numarul de articole publicate de fiecare cadru didactic (în tara, în strainatate);

v) alte rezultate si activitati creative importante pe fiecare cadru didactic.

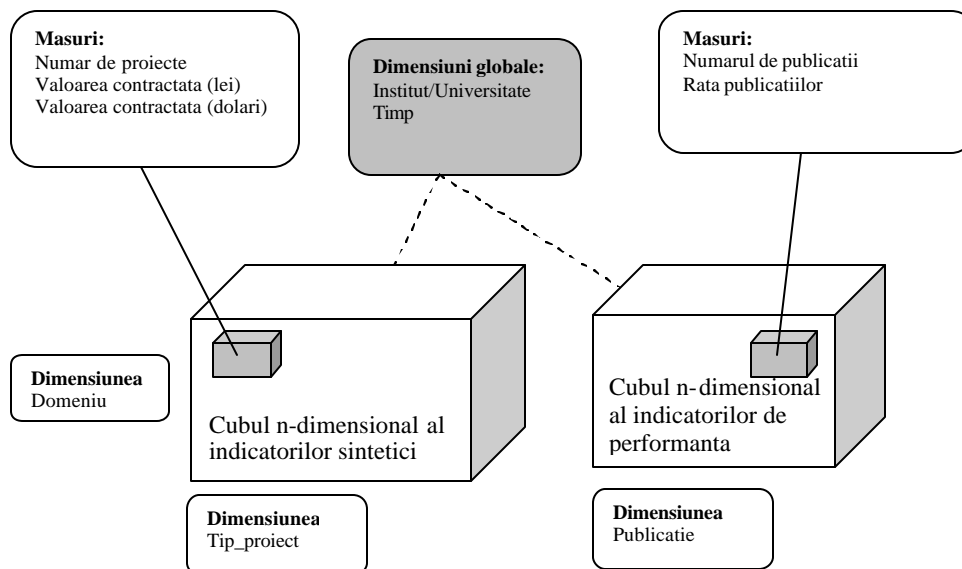


Fig.3. Multicubul de informatii pentru activitatea de cercetare stiintifica în învățământul universitar

Ca indicatori sintetici se vor utiliza:

- i) numarul de proiecte de cercetare;
- ii) personalul implicat în activitatea de cercetare (numarul de cadre didactice, numarul personalului TESA, numarul de doctoranzi, numarul de studenti);
- iii) valoarea contractata (lei si dolari)

Indicatorii prezentati mai sus sunt adecvati pentru obiectivele si strategiile universitatii.

Proiectarea modelului multidimensional conceptual

Exista doua metode pentru proiectarea modelului multidimensional conceptual *bottom-up* si *top-down*. În cazul activitatii de cercetare în învățământul superior se va utiliza metoda de tip *top-down*. Aceasta metoda este mai adecvata, deoarece în etapa anterioara s-au pus în evidenta indi-

catorii de performanta corespunzatori si cuprinde urmatoarele etape:

► *Identificarea variabilelor*. Pe baza studiului si analizei activitatii pentru care se construiesc sistemul, se identifica variabilele sau indicatorii de performanta ai activitatii respective (se specifica si granularitatea acestor variabile) si anume: i) personalul implicat în activitatea de cercetare: numarul de cadre didactice antrenate în cercetare la nivel de catedra si an (*nrcadre*), numarul de cadre TESA antrenate în cercetare la nivel de catedra si an (*nrtesa*), numarul de doctoranzi antrenati în cercetare la nivel de catedra si an (*nrdoct*), numarul de studenti antrenati în cercetare la nivel de catedra si an (*nrstud_cer*); ii) numarul de proiecte pe tip de proiect, an, domeniu, catedra (*nrproiecte*); iii) numarul de publicatii la nivel de catedra, tip de publicatie si

an (*nrpub*); iv) numărul de studenți la nivel de facultate (*nrstud*) și an, numărul de cadre didactice la nivel de catedra și an (*nrprof*); v) valoarea contractată în dolari (*valoare_dolari*) pe tip de proiect, an, domeniu, catedra; vi) valoarea contractată în lei, pe tip de proiect, an, domeniu, catedra (*valoare_lei*). Pentru fiecare indicator/variabilă se vor stabili factorii de care depind sau în funcție de care variază (dimensiunile cubului) și anume: i) variabilele *nrcadre*, *nrtesa*, *nrdoct*, *nrstud_cer*, *nrprof*, *nrstud* au ca dimensiuni: Timp, Institut (universitate); ii) variabilele *nrproiecte*, *valoare_lei*, *valoare_dolari* au ca dimensiuni: Timp, Institut, Domeniu, Tip_proiect; iii) variabila *nrpub* are ca dimensiuni: Timp, Institut, Publicație. De asemenea, fiecare variabilă va fi analizată, în scopul de a determina dacă este aditivă, semiaditivă sau neaditivă. Toate variabilele definite sunt aditive.

► **Identificarea dimensiunilor și a ierarhiilor.** Se identifică următoarele dimensiuni ierarhice: Publicație, Institut (universitate) și Tip_proiect. De exemplu, dimensiunea Institut este o dimensiune cu structură ierarhică cu nivelurile: Institut, Facultate și Catedra. Dimensiunea Timp nu are o structură ierarhică.

► **Definirea cuburilor n-dimensionale/structurii multicub.** Pentru activitatea de cercetare în învățământul superior se definește o structură multicub cu două cuburi n-dimensionale (cubul indicatorilor sintetici și cubul indicatorilor de performanță) și două dimensiuni globale (Timp și Institut).

► **Rafinarea modelului multidimensional.** Structura multidimensională identificată în etapa anterioară se poate rafina prin adăugarea sau stergerea dimensiunilor. Pentru a elimina valorile fără semnificație și a reduce explozia datelor derivate în cub, se pot combina două sau mai multe dimensiuni într-o singură dimensiune. Este necesar a se stabili și dimensiunile care se modifică frecvent și care este rata de modificare a fiecărei dimensiuni. Unele dimensiuni nu se modifică în timp, în spe-

cial dacă modelul este utilizat pe termen scurt. Cele mai frecvente modificări apar în dimensiunea Tip_proiect, iar rata ei de modificare poate fi de cel puțin un an. Deoarece dimensiunea Tip_proiect este mică (are puțini membri) se va utiliza ca metodă de modelare a modificărilor care vor apărea: metoda de a păstra o singură dimensiune în cub, ce reprezintă reuniunea versiunilor. O componentă cheie a modelării multidimensionale este definirea formulelor și în special a formulelor de agregare. Se vor defini următoarele formule de calcul: i) rata publicațiilor (numărul mediu de publicații pe cadru didactic); ii) încărcare cadre didactice (*nrstud/nrprof*); iii) procentul de participare la cercetare. De asemenea, se vor realiza agregări după dimensiunile: Tip_proiect, Institut, Publicație. Pentru fiecare dimensiune se stabilesc nivelurile de granulație și anume: i) dimensiunea Institut are nivelul de granulație: catedra; ii) dimensiunea Publicație are nivelul de granulație: subcategorie de publicație; iii) dimensiunea Timp are nivelul de granulație anii calendaristici (2000, 2001, 2002 etc); iv) dimensiunea Tip_proiect are nivelul de granulație tipurile de proiecte finanțate (de exemplu granturi anuale de tip A, proiecte master/doctorat etc). Pentru a reduce fenomenul de împrăștiere s-a ales o structură multicub cu două cuburi n-dimensionale.

Acest prototip de model multidimensional conceptual poate fi implementat atât în baze de date relaționale cât și în baze de date multidimensionale și permite realizarea următoarelor tipuri de analize: i) evoluția grafică a personalului implicat în activitatea de cercetare (cadre didactice și personal tesa); ii) evoluția grafică a numărului de studenți implicați în activitatea de cercetare; iii) evoluția grafică a numărului de doctoranzi implicați în activitatea de cercetare; iv) evoluția grafică a numărului de proiecte de cercetare (pe domeniu de cercetare, pe facultate, pe perioadă 2000-2003); iii) topul primelor trei facultăți în funcție de indicatorul sintetic ales; iv)

evolutia grafica a valorii contractate (lei/dolari); v) încărcarea cadrelor didactice; vi) rata publicatiilor; vii) evolutia grafica a numarului de publicatii; viii) procentul de participare la activitatea de cercetare; ix) evolutia grafica a numarului de studenti la nivel de facultate etc.

Concluzii

Utilizarea modelului multidimensional si a unei baze de date multidimensionale permite managerilor: i) sa manipuleze datele fara intermediari (cum ar fi în cazul interogarii unui sistem relational). Prezentarea multidimensională a informatiilor permite managerilor sa manipuleze si sa interpreteze mai usor aceste informatii. Se reduce astfel timpul de acces si efortul depus; ii) sa identifice mult mai usor erorile si datele care lipsesc, precum si sa le corecteze. Toti managerii care utilizeaza sistemul suport de decizie vor vizualiza informatiile din aceleasi perspective (modelul multidimensional este proiectat pe baza tipurilor de cereri si a scenariilor de analiza stabilite de manageri). Cererile realizate pe bazele de date relationale au diferite rezultate în functie de modul cum sunt ele formulate.

Principalele dezavantaje ale utilizarii unei baze de date multidimensionale sunt: i) dimensiunea Tip_proiect se poate modifica în functie de timp; ii) informatiile oferite au o implicatie politica, deoarece ar putea intensifica competitia între facultati pentru resurse; iii) se observa de asemenea, ca baza de date multidimensională nu permite cercetatorilor individuali sa extraga o lista cu propriile publicatii. Baza de date multidimensională stocheaza numai date agregate (de exemplu numărul anual de publicatii de un anumit tip, la nivelul unei facultati) si nu date atomice. Din acest motiv se utilizeaza o baza de date relatională sursa care stocheaza datele de detaliu integrată cu baza de date multidimensională care stocheaza datele agregate. Cele doua sisteme (tranzactional si suport de decizie) se completeaza reciproc, oferind astfel managerilor informatii com-

plete si curente necesare procesului decizional.

Chiar daca este unanim acceptata, ideea ca indicatorii de performanta pot fi utilizati pentru evaluarea activitatii de cercetare, totusi este greu de stabilit un set de indicatori corespunzatori, pentru toate institutiile de cercetare si toate disciplinele, astfel ca analizele comparative la nivel global sa fie corecte si relevante pentru manageri. De asemenea, apar probleme legate de colectarea datelor, analiza si prezentarea informatiilor, acuratetea, relevanta lor, precum si daca rezultatul final justifica efortul depus. Asigurarea informatiilor de calitate poate îmbunătăti procesul decizional, deciziile pot fi luate mai usor, se pot identifica, probleme care până atunci au fost ignorate sau necunoscute.

Bibliografie

- [ET96] - E. Thomsen, *OLAP Solutions: Building Multidimensional Information Systems*, John Wiley&Sons, New York, 1996
- [MG99] - G. Marakas, *Decision Support Systems in the 21st Century*, Prentice Hall, 1999
- [BI2001] - Academia de Studii Economice, Departamentul de Cercetari Economice, *Buletin Informativ*, 2001, 2002