

O noua abordare în proiectarea sistemelor suport de decizie orientate pe date

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

Because Decision Support Systems serve many different functions and are quite diverse in terms of the software used for their development and the provision of mathematical models or database access, design and development is an important topic. Choosing an appropriate design and development approach or methodology for building DSS has been a popular and controversial topic in the IS literature. This paper proposes a framework for designing and developing Data-Driven Decision Support Systems.

Keywords: data-driven DSS, framework, methodology, multidimensional models, conceptual models.

Alegerea unei anumite metodologii de proiectare și realizare a unui sistem suport de decizie orientat pe date (SSDOD) depinde de volumul de date și de sursele de date utilizate, de numărul de utilizatori ai sistemului, de instrumentele analitice folosite și de modelele utilizate. Multe SSDOD-uri specializate, de mică complexitate, sunt realizate folosind metodologia de dezvoltare a SSD-urilor de către utilizatori. SSDOD-urile complexe, la nivel de întreprindere, sunt realizate folosind instrumente complexe și metodologii de proiectare structurate a sistemelor informatice. Crearea unor aplicații SSDOD la nivel de întreprindere rămâne o activitate complexă și evolutivă. Un astfel de SSDOD devine inevitabil o parte integrată a infrastructurii informaționale a firmei. Indiferent de metodologia utilizată, toate SSDOD-urile au componente tehnice similare și au același obiectiv principal și anume de a oferi suport în procesul decizional. Specialistii informaticieni trebuie să stabilească un “schimb social” semnificativ cu utilizatorii, să fie receptivi la cererile utilizatorilor.

Metodologii de proiectare a sistemelor suport de decizie orientate pe date

Proiectarea modelului multidimensional conceptual este etapa centrală în proiectarea unui sistem suport de decizie orientat pe date (sisteme OLAP și sisteme cu depozite de date). Firmele și mediul academic au acordat puțină atenție problemelor legate de modelarea

multidimensională conceptuală. Tehnicile de modelare conceptuale existente nu pot fi direct aplicate la caracteristicile modelului de date multidimensional. Ca o consecință, multe proiecte industriale elimină etapa de modelare conceptuală și încep cu proiectarea logică (de exemplu proiectarea unei scheme stea sau fulg de zapadă). În [SBHD98] se consideră că procesul de proiectare a modelului multidimensional conceptual al unui SSDOD depinde foarte mult de cerințele utilizatorilor și de valabilitatea și structura datelor din sistemele operationale. Cele mai multe proiecte folosesc o metodologie evolutivă. Proiectele încep cu un prototip, care este apoi modificat în concordanță cu cerințele utilizatorilor. Procesul de proiectare a modelului multidimensional conceptual este executat iterativ. Există două motive principale pentru acest comportament dinamic: i) tehnologia de analiză multidimensională interactivă este nouă pentru analist. Aceasta înseamnă că este imposibil pentru el să stabilească de la început toate cerințele aplicației; ii) procesele de afaceri pe care analistul trebuie să le modeleze, se modifică frecvent. Aceste modificări trebuie reflectate în cerințele de analiză (noi tipuri de cereri). Întrucât modelul multidimensional determină facilitățile de analiză posibile, noile cerințe conduc la modificări ale schemei bazei de date. În [SBHD98] se consideră că procesul de proiectare a sistemelor suport de decizie orientate pe date este ciclic și cuprinde următoarele etape (figura 1):

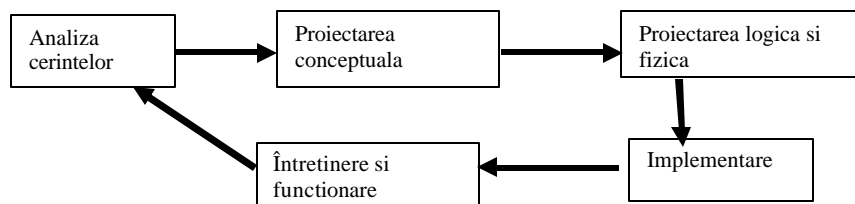


Fig. 1. Procesul ciclic de proiectare a SSDOD

►► *Analiza cerintelor*. În aceasta etapa sunt identificate cerintele utilizatorilor (de exemplu granularitatea datelor, structura si calitatea lor). Rezultatul este un set de viziuni multidimensionale (scheme externe).

►► *Proiectarea conceptuala*. Scopul principal al acestei etape este de a transforma viziunile multidimensionale stabilite în etapa anterioara într-un model conceptual multidimensional. În timpul primei iteratii este creat un model conceptual initial. În timpul celorlalte iteratii, modelul este modificat progresiv pentru a îndeplini noile cerinte ale utilizatorilor. Proiectarea conceptuala este etapa cea mai importanta în procesul de modelare a datelor.

►► *Proiectarea logica si fizica*. În aceasta etapa se stabilesc ce produse software si arhitecturi se folosesc sau care sunt tehnicile de optimizare utilizate.

►► *Implementarea*. În aceasta etapa este inclusa încarcarea initiala cu date a bazei de date (pentru prima iteratie a ciclului).

►► *Întretinerea si functionarea sistemului*. În aceasta etapa sunt încarcate noi date în baza de date si utilizatorii analizeaza datele. Daca exista un volum mare de noi cerinte, începe o noua iteratie a ciclului de proiectare.

La ora actuala nu exista nici o metodologie unanim acceptata, dar exista un consens general ca proiectarea unui SSDOD este un proces complex si evolutiv. Cele mai cunoscute metodologii sunt:

►► *Metodologia lui Cabibbo si Torlone* [CT98a][CT98b] are ca punct de pornire modelul entitate-asociere al bazelor de date relationale existente pentru domeniul analizat si consta din patru etape: i) identificarea faptelor si dimensiunilor; ii) restructurarea modelului entitate-asociere; iii) derivarea unui graf dimensional; iv) transformarea grafului în modelul multidimensional.

►► *Metodologia lui Golfarelli* [GR99] contine sase etape: i) analiza sistemului informatic existent; ii) specificarea cerintelor; iii) proiectarea modelului multidimensional conceptual; iv) validarea modelului multidimensional; v) proiectarea logica; vi) proiectarea fizica.

►► *Metodologia lui Chuck Ballard, Dirk Herreman, Don Schau* [BHS98] are urmatoarele etape: i) identificarea cerintelor informationale ale utilizatorilor; ii) analiza cerintelor; iii) validarea cerintelor; iv) modelarea cerintelor; v) proiectarea modelului multidimensional; vi) construirea, validarea si integrarea solutiei.

►► *Metodologia lui Erik Thomsen* [ET96] are urmatoarele etape: i) analiza cerintelor; ii) definirea modelului multidimensional conceptual.

Se observa ca *proiectarea modelului multidimensional conceptual* reprezinta o *etapa comuna* (se regaseste în toate metodologiile, indiferent ca se proiecteaza un sistem cu depozite de date sau un sistem OLAP). De asemenea, este si o *etapa centrala* în proiectarea unui SSDOD. În figura 2 se prezinta o sinteza a etapelor de proiectare a modelului multidimensional conceptual, pentru fiecare metodologie prezentata, cu scopul de a identifica caracteristicile comune ale acestor metodologii.

În figura 3 se propune un cadru generalizat al etapelor de proiectare si realizare a unui SSDOD (se considera ca aceste etape ar trebui executate indiferent ca se proiecteaza un sistem cu depozite de date sau un sistem OLAP). Procesul de proiectare a modelului multidimensional conceptual depinde foarte mult de cerintele utilizatorilor, de valabilitatea si structura datelor din sistemele operationale sursa.

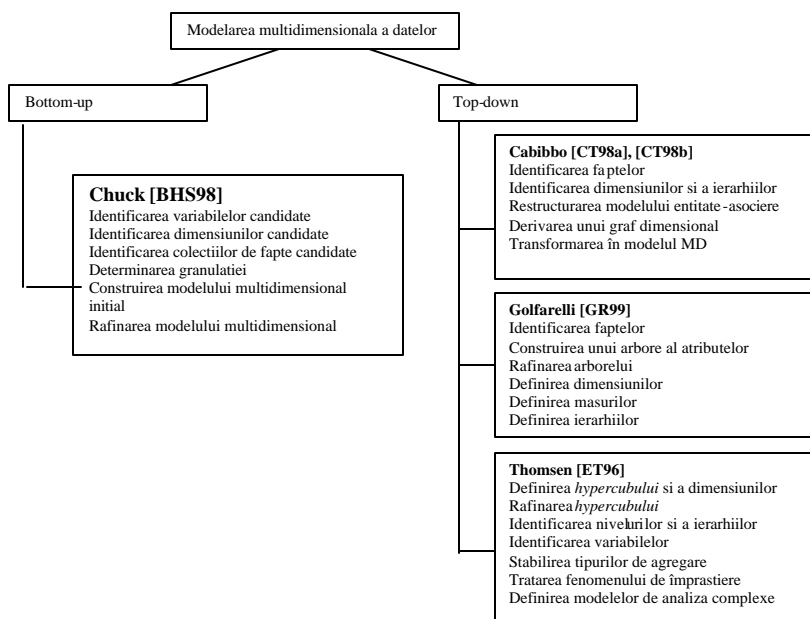


Fig. 2. Metode de proiectare a modelului multidimensional conceptual

Cadrul generalizat din figura 3 încearcă să sintetizeze aceste cerințe și să le structureze în două categorii și anume: *cerințe orientate pe proces* și *cerințe orientate pe informații*. Pentru SSDOD-uri la nivel de înțelegere este foarte dificil de a identifica, în etapa de studiu și analiză a cerințelor informaționale, toate cerințele utilizatorilor. De aceea se consideră a fi utilă și importantă această structurare, pentru a putea fi identificate mai ușor cerințele.

De asemenea, nu există la ora actuală nici un standard pentru modelarea multidimensională conceptuală. Există totuși un consens general că modelarea entitate-asociere nu este potrivită pentru proiectarea unui SSDOD. Unele din metodele și tehnicile de modelare multidimensională sunt specifice sistemelor cu depozite de date. Acestea utilizează în general ca punct de pornire modelul entitate-asociere al bazelor de date operationale existente, pentru domeniul analizat, pe care apoi îl transformă în modelul multidimensional [CT98a] [GR99]. În absența surselor de date (și a modelelor) se vor identifica pe baza studiului și analizei activității pentru care se construiește SSDOD-ul, indicatorii de performanță ai activității respective (variabilele *hypercubului n-dimensional/multicubului*). Singura metodologie specifică sistemelor OLAP este

cea propusă de Erik Thomsen care utilizează o abordare *top-down* pentru proiectarea modelului multidimensional conceptual inițial.

Bibliografie

- [BHS98] C. Ballard, D. Herreman, D. Schau, *Data Modeling Techniques for Data Warehousing*, 1998, <http://www.redbooks.ibm.com>
- [ET96] E. Thomsen, *OLAP Solutions: Building Multidimensional Information Systems*, John Wiley&Sons, New York, 1996
- [GR99] M. Golfarelli, S. Rizzi, *Designing the Data Warehouse: Key Steps and crucial issues*, Journal of Computer Science and Information management, vol2, nr.3, 1999
- [SBHD99] C. Sapia, M. Blaschka, G. Höfling, B. Dinter, *Extending the E/R model for the multidimensional paradigm*, Proc. International Workshop on Data Warehouse and Data Mining in conjunction with the ER'98, nr 1552, in LNCS, Springer, 1999
- [CT98a] L. Cabbibo, R. Torlone, *From a procedural to a Visual Query Language for OLAP*, 10th IEEE International Conference on Scientific and Statistical Database Management, 1998,
- [CT98b] L. Cabbibo, R. Torlone, *A logical approach to multidimensional databases*, Proc. of EDBT'98, Springer, 1998

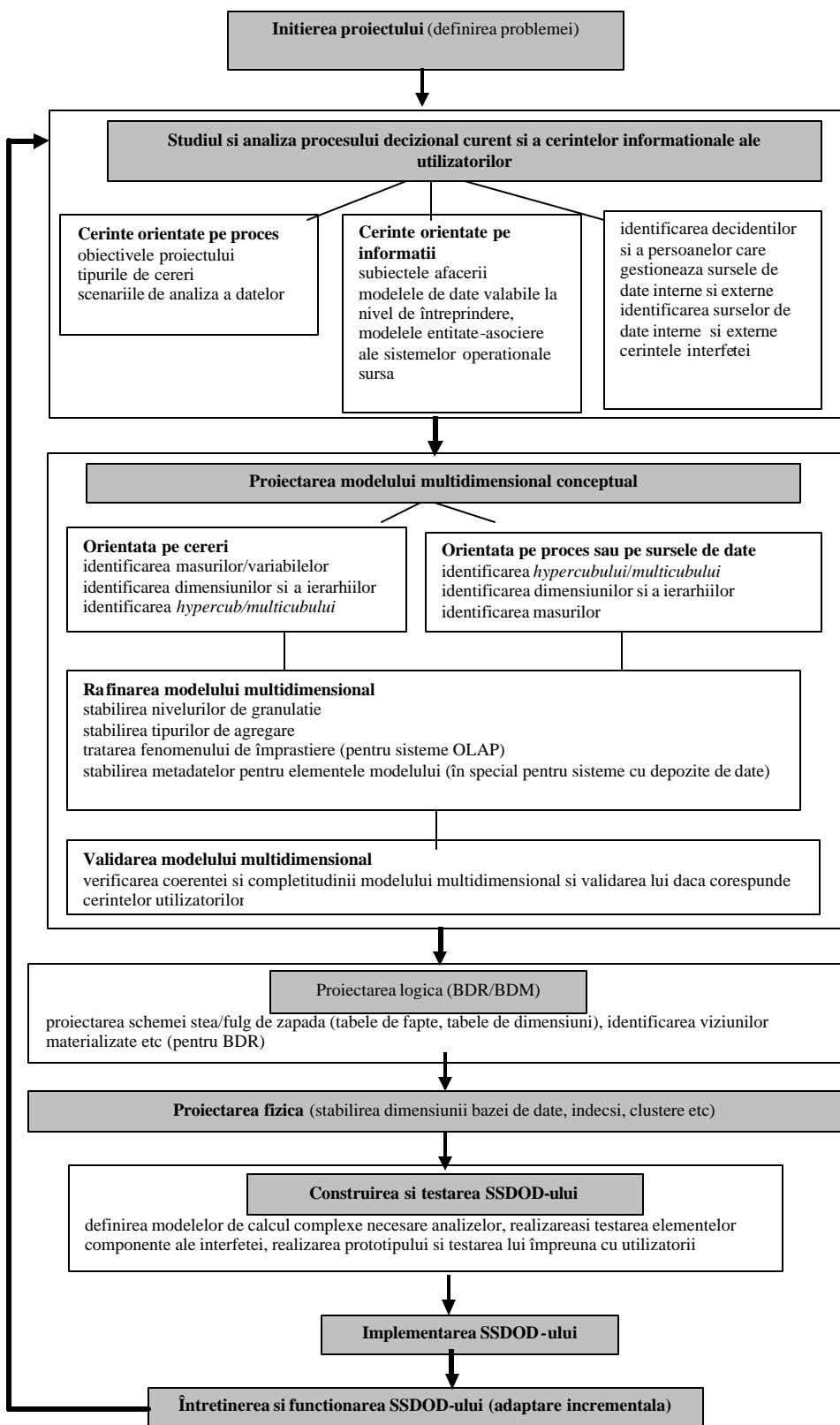


Fig. 3. Cadru generalizat al etapelor de proiectare si realizare a unui SSDOD