

UTILIZAREA DE INSTRUMENTE BI (BUSINESS INTELLIGENCE) ÎN ANALIZA DECIZIONALĂ PENTRU CULTURA CUNOȘTERII

Dora Coardos
coardos@ici.ro

Vasile Constantin Coardos
vasile@ici.ro

Cornelia - Ioana Lepădatu
cornelia@ici.ro

Institutul Național de Cercetare - Dezvoltare în Informatică, ICI, București

Rezumat: În acest articol, se prezintă o parte din rezultatele cercetării științifice proprii privind utilizarea instrumentelor de analiză multidimensională, de tip OLAP, în domeniul culturii cunoașterii. Aceste cercetări au constituit unul din obiectivele principale ale proiectului CEEX, intitulat „Sisteme suport pentru cultura cunoașterii, bazate pe soluții și instrumente din domeniul BI (Business Intelligence) – SSCBT”. Sunt prezentate, de asemenea, soluțiile de business intelligence (BI), destinate îmbunătățirii proceselor de analiză a datelor. Sunt descrise modele multidimensionale prin care se pot analiza diferite probleme legate de activitatea dintr-o bibliotecă. În final, este prezentată implementarea unor astfel de modele cu ajutorul serviciilor oferite de Microsoft SQL Server 2005.

Cuvinte cheie: BI (Business Intelligence), tehnologia OLAP, Data Mining, model multidimensional, cultura cunoașterii.

Abstract: In this paper we emphasize some results of our scientific research related to the usage of multidimensional analysis tools OLAP-like in the knowledge culture. These results emerged from some objectives of the CEEX research project entitled “Decision support systems in the knowledge culture based on solutions and tools in the Business Intelligence field – SSCBT”. Are emphasized also solutions for business intelligence (BI) intended to improve the data analyzing process. Multidimensional models that are intended to analyse various problems in the management of a library are presented. Finally an implementation of these models using services offered by Microsoft SQL Server 2005 is presented.

Keywords: BI (Business Intelligence), OLAP technology, multidimensional data mining, knowledge culture.

1. Introducere

Cu trecerea timpului, organizațiile acumulează un volum tot mai mare de date legate de procesele de business și tranzacțiile comerciale realizate. Sistemele tranzacționale și soluțiile de baze de date clasice nu oferă suportul pentru realizarea unei analize rafinate și nici pentru identificarea principalelor evoluții ale indicatorilor de performanță ai afacerii. Aplicațiile de acest tip oferă mai cu seama suportul pentru gestionarea tranzacțiilor realizate, integrând funcții de bază pentru interogarea datelor. Sistemele tranzacționale și bazele de date relaționale acoperă în principal nevoile de bază ale organizației, fiind destinate mai degrabă laturii operative din cadrul companiilor.

Spre deosebire de această categorie de aplicații, soluțiile de business intelligence sunt destinate îmbunătățirii proceselor de analiză a datelor. Informațiile sunt organizate în structuri care satisfac nevoile de analiză completă și complexă a datelor. Depozitele de date, organizarea multidimensională sau instrumentele de „Data Mining” sunt doar câteva exemple de tehnologii capabile să ofere funcții de analiză detaliată.

Soluțiile de asistare a deciziilor [3] și de business intelligence pot fi considerate ca un nivel intermediar între bazele de date tradiționale și manager, oferind acestuia posibilitatea de a lua decizii pe baza studierii unei multitudini de forme de reprezentare a rezultatelor aplicării mecanismelor de analiză online a datelor.

Organizațiile utilizează o gamă largă de produse și tehnologii pentru a genera ceea ce se cheamă Business Intelligence (BI) și asistare a deciziilor.

Cele mai obișnuite instrumente - căutare și raportare, procesare analitică online, analize statistice, previziuni și analize de date - pot fi folosite într-o multitudine de feluri. Aplicațiile pot oferi acces ad-hoc la o singură zonă de informații sau pot fi motoare importante pe bază de web, care conduc procesele de afaceri. Scopul este transformarea volumului mare de date în informații utile.

Cea mai frecventă platformă utilizată pentru a obține acest lucru este baza de date. De fapt, o combinație de baze de date, numită depozit de date, este cea mai bună sursă pentru BI. Datele selectate pentru a fi utilizate în depozite, sunt reformatate și stocate printr-un proces numit ETL - extragere, translație și încărcare. Acest proces standardizează diferitele structuri de date, astfel încât să fie accesate și analizate cu mare acuratețe.

Producătorii de sisteme de gestiune a bazelor de date au integrat funcții performante de BI, care comunică eficient cu soluțiile tradiționale de baze de date. În acest mod, pe lângă suportul clasic pentru bazele de date relaționale, au fost integrate tehnologii tot mai avansate de interogare și analiză a datelor. Printre acestea se numără tehnologia OLAP (OnLine Analytical Process), precum și posibilități de structurare în cuburi și de analiză multi-dimensională a datelor. Tehnologia Data Mining („mineritul datelor”), care își are originile în domeniul statisticii, permite folosirea unor modele matematice pentru predicția fenomenelor economice pe baza informațiilor din bazele de date.

Toate soluțiile consacrate de baze de date integrează și funcții de business intelligence. Astfel, sunt de

remarcat următoarele companii ce oferă acces la instrumente performante specifice acestui domeniu: IBM (DB2 Universal Database), Microsoft (SQL Server), Oracle (bazele de date Oracle), Sybase etc.

Microsoft dispune de o gamă de servicii complete de business intelligence integrate cu baza de date SQL Server. Spre exemplu, Microsoft Reporting Services este o soluție completă de raportare care permite prezentarea informațiilor analizate de soluția BI prin intermediul Web-ului. Microsoft Analysis Services este un alt serviciu performant SQL Server, care permite realizarea analizelor OLAP și Data Mining [1], [2].

2. Soluții de business intelligence

O facilitate esențială a sistemelor BI este capacitatea acestora de a se conecta simultan și coerent la mai multe surse de date, care pot fi sisteme operaționale diferite, rezultate ale unor cercetări de piață sau orice poate avea relevanță pentru organizația-beneficiar, în general, informații structurate, dar aflate în formate variate, de la fișiere text până la structuri stocate în depozite de date.

Din punct de vedere al instrumentelor utilizabile de management, soluțiile BI se împart în două categorii:

- soluții de Raportare Dinamică (cunoscute și ca OLAP);
- soluții de Data Mining.

Soluțiile de Raportare Dinamică

Cunoscute și ca soluții OLAP (OnLine Analytical Processing), soluțiile de raportare dinamică permit neutilizatorilor de IT să acceseze în mod interactiv, coerent și sintetic structuri complexe de date colectate din mai multe sisteme informaționale. De cele mai multe ori, soluțiile de acest fel se bazează pe puterea de stocare și de procesare a unui server pentru a rula generarea raportului, pe mașina-client rulând un program de tip „client” sau doar un browser web.

În ultimii ani, se observă o diversificare a modalităților de abordare a raportării dinamice. Domeniul raportărilor dinamice mai este cunoscut și ca Fast Analysis of Shared Multidimensional Information (FASMI).

Soluțiile de Data Mining

Din cauza volumului de date tot mai mare și, mai ales, a complexității datelor și relațiilor dintre ele, posibilitatea ca omul - chiar dotat cu cele mai eficiente instrumente de raportare și vizualizare - să descopere legături între diverse evenimente încapsulate în datele pe care le avem înregistrate în sistemele noastre, devine tot mai mică. Aici intervin soluțiile de Data Mining, care descoperă și verifică automat sau semiautomat legături între evenimente corelate.

Platforme de BI, structuri clasice de BI

În abordarea clasică, datorită volumului mare de date și a necesităților de procesare uriașe, atât soluțiile de raportare dinamică, cât și cele de Data Mining au nevoie de o infrastructură informatică puternică, pentru a avea capacități masive de stocare, interconectare, procesare și corelare. Această infrastructură se referă nu numai la capacitățile hardware, cât mai ales la cele software, inclusiv parametrizările specifice fiecărei organizații.

Depozitarul central, care înmagazinează totalitatea informațiilor colectate din sistemele informatice, este componenta centrală a sistemului. Denumirea consacrată este chiar Data Warehouse (depozitul de date).

Pentru a colecta în Data Warehouse date din mai multe structuri informatice, care să poată fi ulterior și corelate (chiar dacă provin din sisteme diferite), de obicei, este necesar un instrument informatic, destinat colectării, filtrării și transformării datelor primare ale sistemelor operaționale. De această sarcină se ocupă, de obicei, serviciile ETL (Extract, Transformation, Load) ale sistemelor BI.

Primele sisteme de raportare dinamică - sau cele cu o structură simplificată din zilele noastre - pun între utilizatorul final și Data Warehouse interfața de interogare, care de obicei propune un mod grafic, interactiv de navigare în date. Această structură este răspândită la sistemele OLAP clasice.

Principalele dezavantaje ale acestei soluții OLAP constau în timpul de răspuns prea mare - datorat timpului de procesare al unui volum de date prea mare - precum și în modalitățile relativ greoaie de a defini informația căutată.

Pentru a ameliora aceste aspecte, generația următoare de soluții a oferit OLAP multidimensional. În această abordare, între interfața de interogare și Data Warehouse a apărut o structură nouă de date, numită hipercub de date. Rolul acestuia este să colecteze, de obicei, într-un singur tabel, toate informațiile stocate în Data Warehouse, care sunt relevante pentru un anumit tip de analiză, și să aibă gata calculate toate detaliile și combinațiile de bază posibile.

Construirea soluției de tip OLAP multidimensional duce, însă, la creșterea semnificativă a timpului și a resurselor necesare dezvoltării rapoartelor dinamice. În plus, interogările și raportările sunt limitate la cele pe care proiectantul cubului le-a prevăzut. Orice altă informație, disponibilă în Data Warehouse, dar neîncărcată în cub, rămâne indisponibilă până la construirea unui nou cub, mai cuprinzător. Implicit, este nevoie de un timp semnificativ mai mare alocat analizei de sistem inițială.

În plus, capacitățile de stocare necesare platformei BI cresc vertiginos atât datorită redundanțelor multiple între Data Warehouse și diversele hipercuburi, cât și datorită renunțării la structurile relaționale în cadrul hipercuburilor.

Problema accesului la informațiile nestocate în hipercuburi a rezolvat-o următoarea generație de platforme BI, care propun soluții de tip OLAP hibrid, ce conțin mecanisme de suplimentare a dimensiunilor stocate în hipercuburi atunci când utilizatorii solicită acest lucru (mecanisme de tip drill-through - explorare/căutare elemente de tipul: foldere, fișiere sau componente relaționale). Din păcate, complexitatea interogărilor face aceste soluții relativ neprietenoase, în special pentru utilizatorul de business, cu un background IT limitat.

Alegerea platformei optime pentru implementarea unui sistem de Business Intelligence într-o organizație presupune luarea în considerare a unui set de criterii relevante, care să folosească drept referință în selectarea soluției BI. Dintre aceste criterii amintim:

- timpul de implementare;
- flexibilitatea;
- ușurința în utilizare a platformei;
- capacitatea de procesare și scalabilitatea.

Un factor determinant în alegerea platformei ar trebui să fie timpul necesar adoptării sistemului BI de către echipa tehnică a organizației-beneficiar, cea care va asigura mentenanța și dezvoltarea sa ulterioară.

O dată ce un sistem de Business Intelligence a fost ales, se deschid numeroase domenii în care acesta poate fi folosit pentru suportul deciziilor de business.

Depozite de date (data warehousing)

Depozitele de date (data warehouse) au fost definite în foarte multe moduri, astfel încât este destul de dificil de formulat o definiție riguroasă.

Conform definiției date de W. H. Inmon [5], unul dintre cei mai importanți proiectanți de sisteme warehouse, „un depozit de date este o colecție de date orientate pe subiecte, integrate, istorice și nevolatile, destinată asistării procesului de luare a deciziilor manageriale”. În sinteză, definiția prezentată mai sus exprimă caracteristicile principale ale depozitelor de date: orientare pe subiecte, integrare, caracter istoric și persistența datelor.

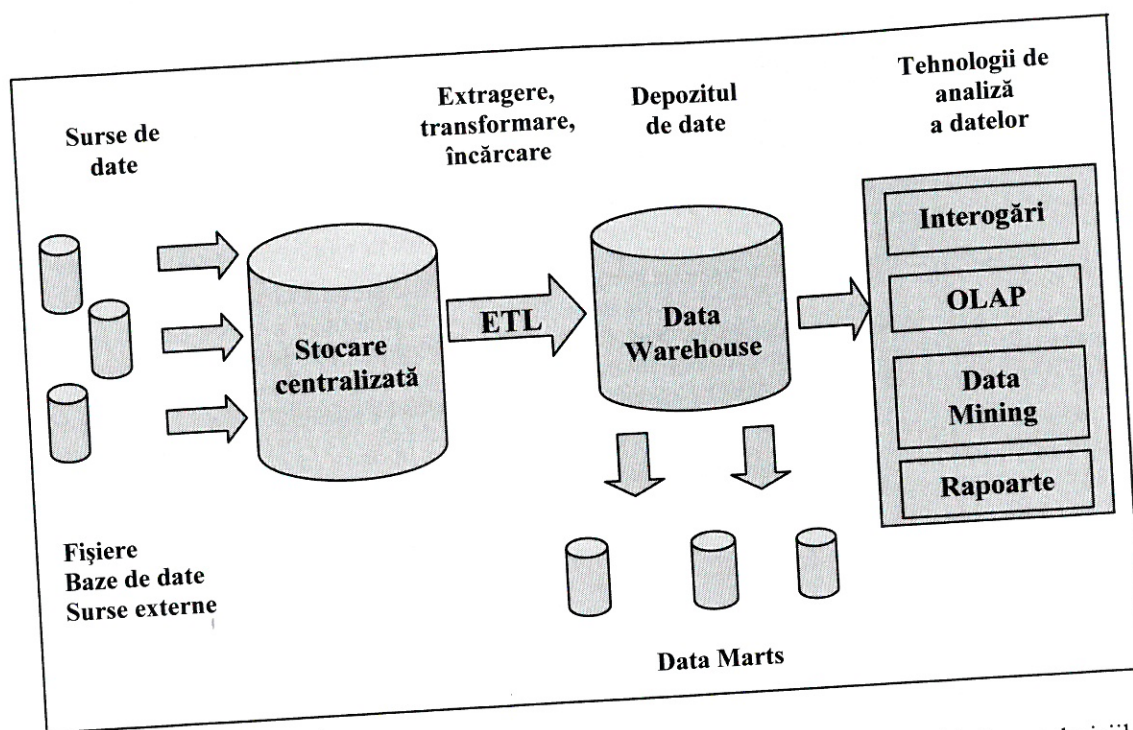
Aceste caracteristici fac distincția între data warehouse și alte depozite de date, cum ar fi sistemele de baze de date relaționale și sistemele de prelucrare a tranzacțiilor.

O noțiune frecvent utilizată în legătură cu depozitele de date este cea de „data warehousing” care desemnează procesul de construire și utilizare a depozitelor de date (data warehouse). Construirea unui depozit de date necesită integrarea datelor, curățirea datelor (data cleaning) și consolidarea datelor. Utilizarea unui depozit de date necesită, adesea, o colecție de tehnologii de asistare a deciziilor. Acestea permit managerilor și specialiștilor (de exemplu, analiști, consilieri etc.) să utilizeze depozitul pentru a obține rapid și convenabil datele necesare și să ia deciziile bazate pe informațiile din depozit.

Esența unui depozit de date constă într-o bază de date de dimensiuni foarte mari, conținând informațiile pe care le pot folosi utilizatorii finali [4].

În depozitul de date întâlnim mai multe tipuri de date care corespund diferitelor cerințe informaționale ale utilizatorilor: date detaliate, date agregate, metadata. Metadatale descriu datele conținute în depozitul de date și modul în care ele sunt obținute și stocate. Prin metadata se precizează structura datelor, proveniența lor, regulile de transformare, de agregare și de calcul.

Arhitectura complexă a unui depozit de date este prezentată în figura de mai jos:



Depozitele de date sunt destinate managerilor, analiștilor și specialiștilor angrenați în luarea deciziilor strategice privind dezvoltarea și viitorul organizațiilor. Pentru aceasta, ei au nevoie de instrumente performante de accesare și utilizare a datelor din depozite, instrumente asigurate prin software-ul asociat depozitului de date. Pe de o parte, regăsim instrumentele necesare utilizatorilor care au nevoie de acces rapid, de informații punctuale care includ un limbaj de interogare gen SQL sau de generatoare de rapoarte (Report Writers) ce transpun informațiile în formate adecvate. Pe de altă parte, sunt instrumentele specializate pentru asistarea deciziilor, care transformă informațiile în forma cerută de decidenți (grafice, diagrame, organigrame) sau oferă posibilitatea analizei tendințelor, corelațiilor și interpretarea acestora. În această categorie se încadrează instrumentele OLAP și Data Mining.

3. Modele multidimensionale ale datelor în domeniul culturii cunoașterii

Depozitele de date și instrumentele OLAP sunt bazate pe modele multidimensionale de date. Aceste modele vizualizează datele sub forma unui cub de date (data cube).

Cubul de date permite modelarea și vizualizarea datelor în dimensiuni multiple. El este definit prin dimensiuni și fapte. În termeni generali, dimensiunile exprimă perspectivele în care o anumită organizație dorește să păstreze înregistrările privitoare la tranzacțiile desfășurate.

De exemplu, organizația Biblioteca A poate crea un depozit de date pentru achiziții, care conține înregistrările, luând în considerare următoarele dimensiuni: timp, domeniu și zonă. Aceste dimensiuni permit memorarea achizițiilor lunare pe domenii și zone. Fiecare dimensiune poate avea un tabel asociat, numit tabel dimensiune, care descrie dimensiunile. De exemplu, un tabel dimensiune pentru domeniu poate conține atributele: nume-domeniu, cod-domeniu, tip. Tabelele dimensiune pot fi specificate de utilizatori sau de experți sau pot fi generate în mod automat și adaptate în funcție de distribuția datelor.

Considerăm că un model multidimensional de date este organizat în jurul unei teme centrale, achiziții. Această temă este reprezentată de tabelul de fapte. Faptul are o măsură numerică. El exprimă măsurile prin care dorim să analizăm relațiile între dimensiuni. De exemplu, faptele din depozitul de date achiziții includ: valoare (achiziții în lei), cantitate (numărul de documente achiziționate), total achiziții planificate. Tabelul de fapte conține numele faptelor sau măsurile, precum și cheile pentru fiecare din tabelele dimensiune aflate în legătură.

Deși, în mod obișnuit, ne bazăm pe cuburi 3D, în depozitele de date cubul este n-dimensional. Pentru definirea cubului de date și a modelului multidimensional de date, pornim de la un exemplu de cub de date 2D care poate fi, în principiu, un tabel sau o foaie de calcul privind achizițiile unei organizații de tip bibliotecă. În particular, este vorba de datele referitoare la achizițiile Bibliotecii A pe domeniile achiziționate trimestrial în municipiul București, după cum se vede în tabelul următor.

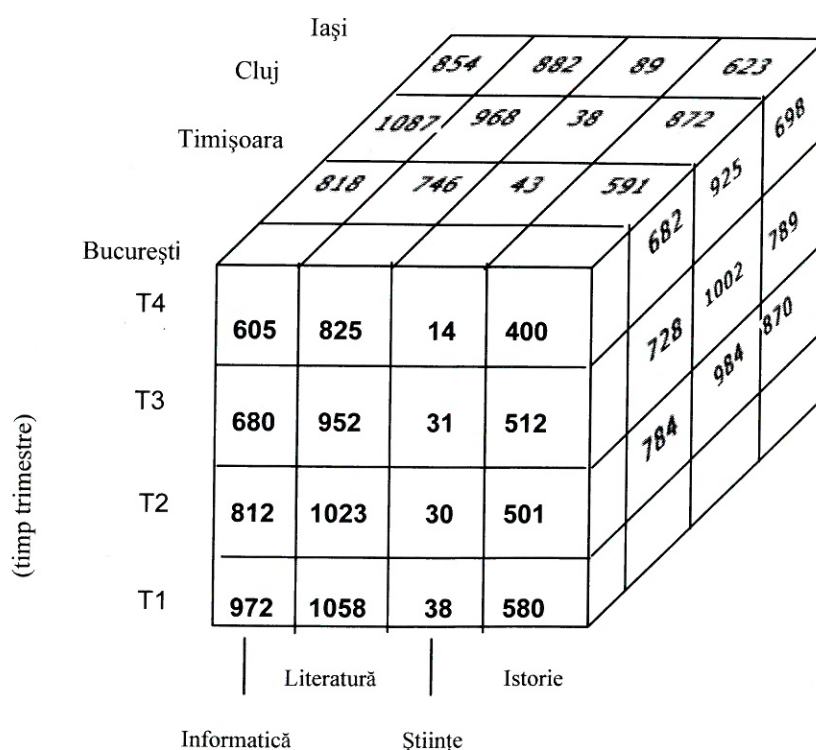
În reprezentarea 2D, achizițiile din București sunt prezentate luând în considerare dimensiunea timp (organizată pe trimestre) și dimensiunea domeniu (organizată pe tipuri de domenii achiziționate). Valorile documentelor achiziționate din domeniile alese sunt afișate în lei.

Zona = „B”				
Trimestre	Domenii			
	Informatică	Literatură	Științe	Istorie
1	605	825	14	400
2	680	952	31	512
3	812	1023	30	501
4	927	1058	38	580

Presupunem acum că dorim vizualizarea datelor despre achiziții cu o a treia dimensiune - zonă. Cele trei dimensiuni sunt: timp, domeniu și zonă (B, T, C, I). Varianta 3D este prezentată în tabelul de mai jos. Datele 3D sunt reprezentate ca serii de table 2D.

	Zona = „I”				Zona = „C”				Zona = „T”				Zona = „B”			
	Domeniu				Domeniu				Domeniu				Domeniu			
Trimestre	Informatică	Literatură	Științe	Istorie	Informatică	Literatură	Științe	Istorie	Informatică	Literatură	Științe	Istorie	Informatică	Literatură	Științe	Istorie
1	854	882	89	623	1087	968	38	872	818	746	43	591	605	825	14	400
2	493	890	64	698	1130	1024	41	925	894	769	52	682	680	952	31	512
3	952	924	59	789	1034	1048	45	1002	940	795	58	728	812	1023	30	501
4	659	992	63	870	1142	1091	54	984	978	864	59	784	927	1058	38	580

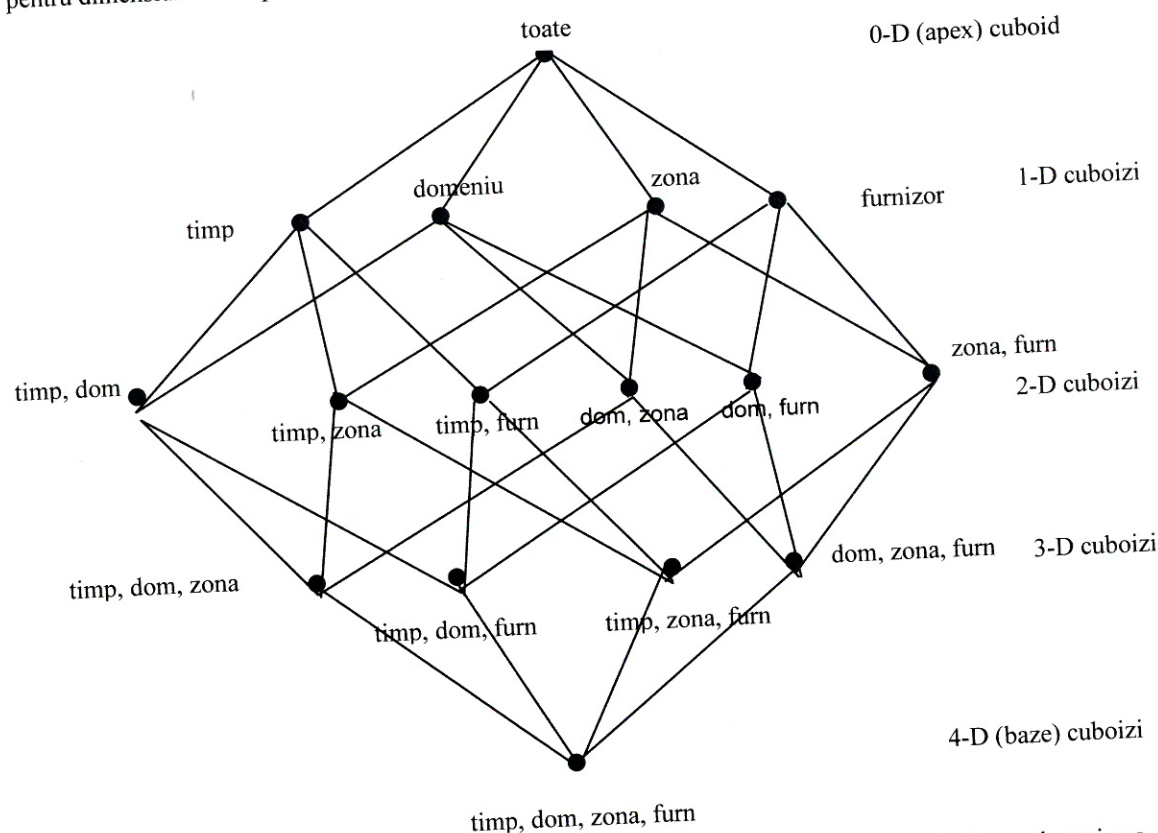
Conceptual, putem reprezenta aceleași date în formatul cub de date 3D ca în figura de mai jos:



Un cub de date este un set de date, organizat și sumarizat într-o structură multidimensională printr-un set de dimensiuni și măsuri. Cubul de date furnizează un mecanism simplu pentru interogarea datelor cu un timp de răspuns foarte scurt. Fiecare cub are o schemă reprezentată de setul de tabele din depozitul de date. Tabelul central este tabelul de fapte și este sursa de măsuri din cub, iar tabelele dimensiune sunt sursele de dimensiuni.

Presupunem acum că dorim să vizualizăm achizițiile adăugând a patra dimensiune, cum ar fi furnizorii. Vizualizarea în 4D (cubul 4D) se poate realiza ca serii de cuburi 3D. Astfel, putem afișa orice n-D date ca serii de (n-1)D cuburi.

Cubul de date este o metaforă pentru datele stocate multidimensional. Stocarea la nivel fizic în mediile actuale depinde de reprezentarea logică. Este important de reținut că acest cub de date este n-dimensional, și nu se limitează la 3D. Tabele anterioare prezintă date cu diferite grade de sintetizare. În literatura de date warehouse, cubul de date este denumit „cuboid”. Stabilind setul de dimensiuni, putem construi matricea cuboidului, fiecare prezentând datele cu diferite niveluri de sintetizare sau grupare. (Matricea cuboidului este referită tot ca un cub de date. Figura următoare exprimă matricea cuboidului, formând cuburi de date pentru dimensiunile: timp, domenii, zonă, furnizor.



Cuboidul situat la nivelul cel mai de jos este numit cuboid de bază. De exemplu, în figura de mai sus, cuboidul 4D este cuboid de bază pentru dimensiunile timp, domeniu, zonă și furnizor. Cuboidul 0D ne arată cel mai înalt nivel de sintetizare și este numit apex cuboid (cuboid vârf). În exemplul prezentat, acesta este totalul achizițiilor sau însumarea achizițiilor în lei pentru toate cele 4 dimensiuni. Cuboidul vârf este, în mod obișnuit, denumit cu all (toate).

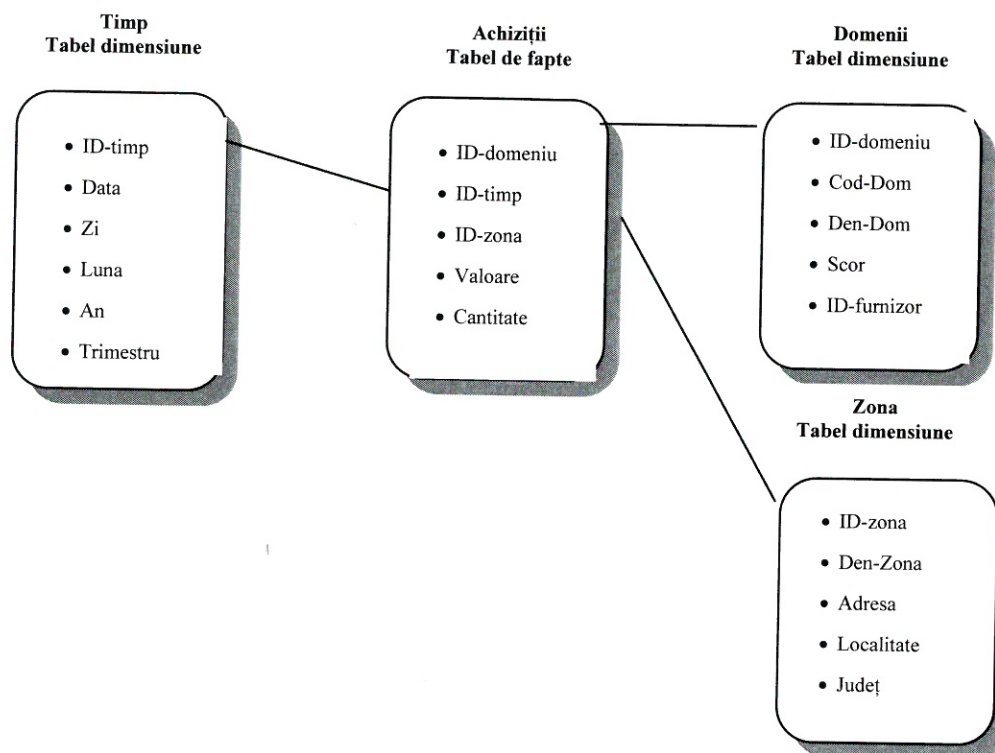
Un depozit de date necesită o schemă concisă, orientată pe subiecte care facilitează analiza on-line a datelor.

Cel mai utilizat model pentru depozitele de date este modelul multidimensional. Acesta poate fi în formă de stea, de fulg de zăpadă sau de constelație.

Schema „stea” este cel mai comun model de date, în care depozitul de date conține un tabel central (tabelul de fapte) și un set de tabele însoțitoare (tabelele dimensiune) pentru fiecare dimensiune. Tabelul de fapte cuprinde, fără redundanțe, cea mai mare parte a datelor. Graficul asociat seamănă cu o stea în care tabelele dimensiune sunt afișate radial în jurul tabelului de fapte central.

Un exemplu de schemă „stea” este prezentat în figura de mai jos. Achizițiile sunt considerate cu 3

dimensiuni, numite timp, domeniu și zonă. Schema cuprinde un tabel central pentru achiziții, care conține chei pentru fiecare din cele 3 dimensiuni înaintea celor două măsuri: valoare și cantitate.



În această schemă „stea”, fiecare dimensiune este reprezentată printr-un singur tabel și fiecare tabel conține un set de atribute.

4. Proiectarea sistemului SSCBI

Sistemul suport pentru cultura cunoașterii bazat pe soluții și instrumente din domeniul BI - SSCBI - are la bază conceptul de depozit de date (data warehouse).

Acest sistem este compus din două subsisteme: un sistem informatic tranzacțional și un sistem informatic analitic.

Sistemul informatic tranzacțional va furniza informații atât despre aparițiile editoriale pe diferite subiecte (domenii), cât și despre consultarea subiectelor (domeniilor) de interes solicitate de cititorii bibliotecii.

Toate informațiile stocate în sistemul tranzacțional sunt, ulterior, migrate în sistemul informatic analitic în vederea realizării următoarelor tipuri de analize:

- gradul de interes manifestat pentru diverse subiecte (domenii) de către cititorii bibliotecii;
- gradul de credibilitate al autorilor, editurilor și al publicațiilor.

Această analiză va permite factorilor de decizie ai unei biblioteci să stabilească o politică adecvată de achiziții de publicații.

Sistemul informatic tranzacțional este constituit din două subsisteme: APARIȚII și CONSULTĂRI. Fiecare dintre aceste subsisteme se va constitui în sursă de date pentru depozitul de date al sistemului SSCBI.

Pentru crearea bazei de date BDAC (Apariții - Consultări) a sistemului informatic tranzacțional, se utilizează componenta SQL Server Management Studio a produsului SQL Server 2005 [6].

În cadrul subsistemului APARIȚII, sunt gestionate informații referitoare la:

- editarea publicațiilor de tip „periodice” (beta-Documente), informații de genul: titlu publicație, editura, data și locul publicării, subiecte de interes tratate în cuprinsul ei;
- articolele conținute de publicații (alfa-Documente), cum ar fi: titlul articolului, autorul, subiecte de interes tratate în cadrul articolului;

- tip „măsură” obținute prin prelucrări statistice (credibilitate autor, editură sau grad de interes manifestat pentru un subiect).

Această bază de date constituie sursa de date pentru încărcarea magaziei de date APARIȚII și este formată din următoarele tabele:

Denumire tabelă	Conținut
A-BetaDOCUMENTE	Informații referitoare la publicații (periodice)
A-AlfaDOCUMENTE	Informații referitoare la articole
D-AGENTI	Informații referitoare la edituri
D-SPATIU	Informații referitoare la locul publicării (oraș, regiune, țară)
D-TIMP	Informații referitoare la data publicării
D-SUBIECTE	Informații referitoare la subiectele (domeniile) tratate
D-AUTORI	Informații referitoare la autorii articolelor
A-REFBIBLIO	Informații referitoare la bibliografia articolelor

Conținutul informațional al bazei de date APARIȚII este asigurat din surse de date externe bibliotecii cum ar fi: cataloage și planuri editoriale, referințe bibliografice de pe Web. Pentru testarea modelului experimental al sistemului SSCBI s-au folosit ca surse de date referințe bibliografice existente pe Web, referințe ce conțin informații suficiente pentru încărcarea bazei de date.

În cadrul subsistemului CONSULTĂRI, sunt gestionate informații referitoare la:

- publicațiile de tip „periodice” (beta-Documente) aflate într-o bibliotecă, informații de genul: titlu publicație, editura, data și locul publicării, subiecte de interes tratate în cuprinsul ei;
- articolele conținute (alfa-Documente), cum ar fi: titlul articolului, autorul, subiecte de interes tratate în cadrul articolului;
- tip „măsură” obținute prin prelucrări statistice (credibilitate autor, editură sau grad de interes manifestat pentru un subiect);
- cititorii bibliotecii;
- sistemul de consultări al fondului de periodice.

Sursa de date pentru încărcarea magaziei de date CONSULTĂRI este baza de date a subsistemului CONSULTĂRI.

Această bază de date este formată din următoarele tabele:

Denumire tabelă	Conținut
C-BetaDOCUMENTE	Informații referitoare la publicații (periodice)
C-AlfaDOCUMENTE	Informații referitoare la articole
D-AGENTI*	Informații referitoare la edituri
D-SPATIU*	Informații referitoare la locul publicării (oraș, regiune, țară)
D-TIMP*	Informații referitoare la data publicării
D-SUBIECTE*	Informații referitoare la subiectele (domeniile) tratate
D-AUTORI*	Informații referitoare la autorii articolelor
C-CITITORI	Informații referitoare la cititori
D-PROFESII	Informații referitoare la profesia cititorilor
C-SOLICITARI	Informații referitoare la relația cititor-document solicitat

Informațiile ce vor fi încărcate în baza de date CONSULTARI provin din catalogul online OPAC, furnizat de sistemul ALEPH, și din aplicația de gestionare a împrumuturilor.

Pentru încărcarea informațiilor în depozitul de date SSCBI, s-au folosit ca surse de date atât surse existente, cât și surse nou create.

Sistemul informatic analitic: Esența unui depozit de date constă într-o bază de date de dimensiuni foarte mari, conținând informațiile pe care le pot folosi utilizatorii finali (cititori, autori, edituri etc.).

În depozitul de date, întâlnim mai multe tipuri de date care corespund diferitelor cerințe informaționale ale

utilizatorilor: date detaliate, date agregate, metadata. Metadatale descriu datele conținute în depozitul de date și modul în care ele sunt obținute și stocate. Prin metadata se precizează structura datelor, proveniența lor, regulile de transformare, de agregare și de calcul. Metadatale joacă un rol esențial în alimentarea depozitului cu date. Ele sunt utilizate în toate etapele de încărcare a datelor și sunt consultate și actualizate pe parcursul întregului ciclu de viață a depozitului. Includerea datelor agregate în depozit, deși determină o creștere a redundanței datelor, este necesară deoarece, în acest fel, se poate asigura un timp mediu de răspuns cât mai redus.

Pentru proiectarea, crearea și gestionarea structurilor multidimensionale care conțin date agregate din diferite surse de date, cum ar fi bazele de date relaționale, SQL Server 2005 pune la dispoziția proiectanților de aplicații BI componenta Analysis Services.

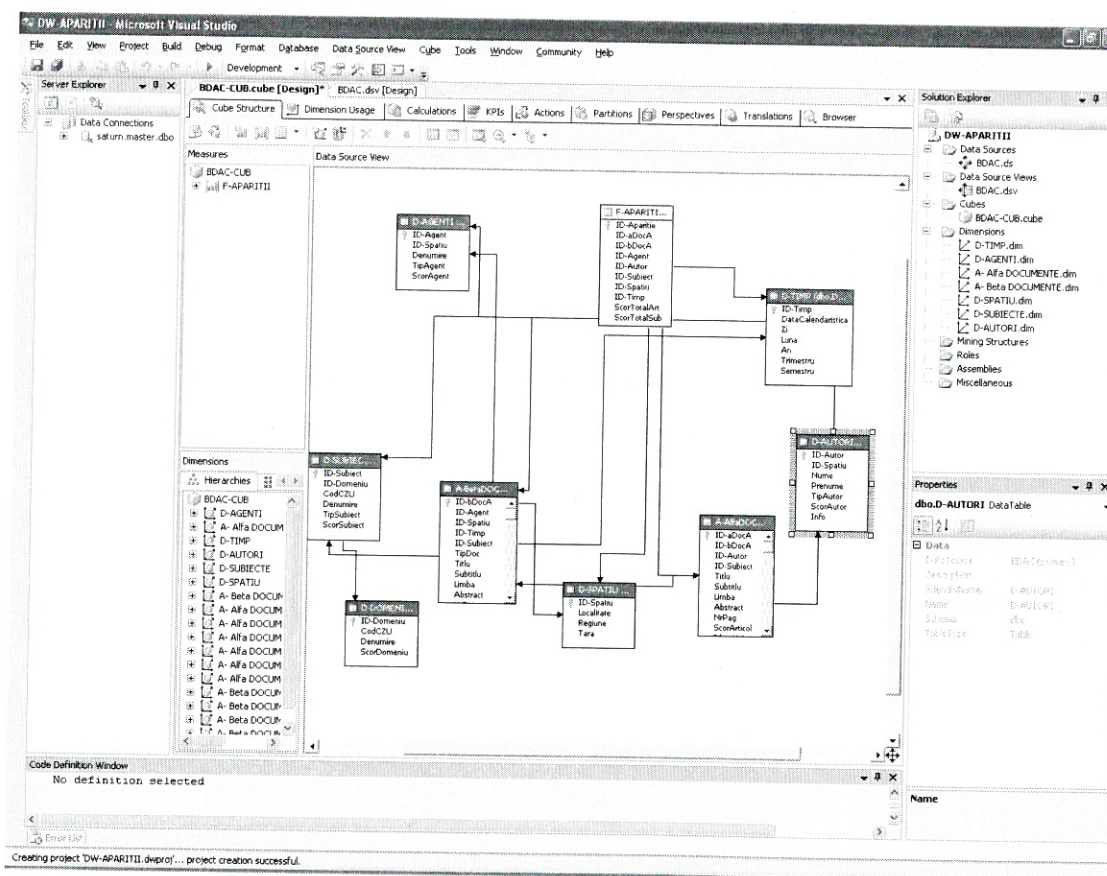
Crearea unui depozit de date implică reorganizarea datelor stocate în tabelele relaționale din OLTP (On Line Transaction Processing) în date stocate în cuburi multidimensionale de tip OLAP (On Line Analytical Processing). Definirea obiectelor în cadrul unui model multidimensional de tip OLAP se realizează în cadrul unui proiect (Analysis Services project).

Se utilizează componenta SQL Server Business Intelligence Development Studio pentru definirea de surse de date, view-ere, cuburi multidimensionale.

Rezultatele vor fi afișate în fereastra DW APARIȚII. Acestea vor constitui toate elementele sistemului analitic:

- măsuri;
- dimensiuni;
- diagrama sursei de date;
- cubul BDAC.

Un astfel de exemplu se poate vedea în figura de mai jos.



5. Concluzii

Problemele de utilizare a resurselor cultural științifice și, în special, procesul de diseminarea informațiilor din cadrul unei biblioteci necesită luarea de decizii adecvate.

Complexitatea problemelor face ca luarea deciziei să devină din ce în ce mai dificilă. Progresele realizate în cadrul teoriei deciziei și sistemelor suport de decizie au determinat apariția de metode ce pot ușura decidentului de azi luarea deciziilor celor mai adecvate, pentru probleme cu grad de complexitate ridicat.

Totuși, elaborarea unor modele adecvate, care să ia în considerare procesul de diseminarea informațiilor în cadrul unei biblioteci, precum și proiectarea și realizarea de modele multidimensionale reprezintă o sarcină dificilă.

Este important de a identifica și organiza cantitatea mare de date și informații existente în cadrul unei biblioteci, cu scopul de a le analiza și prelucra (prin OLAP, Data Mining etc.).

Bibliografie

1. **MELOMED, E., I. GORBACH, AL. BERGER, PY BATEMAN:** Microsoft SQL server 2005, Analysis Services, Sams Publishing, 2007.
2. **WHALEN, E., M. GARCIA, B. PATEL, S. MISNER, V. ISAKOV,** Microsoft SQL Server 2005, Administrator's Companion, Microsoft Press, 2007.
3. **FILIP, F. G.:** Decizie asistată de calculator, Editura tehnică, 2005.
4. **INMON, W.:** What is Data Warehouse? Prism Solutions, Inc., 1995.
5. **MUNDY, J., W. THORNTHWAITE, R. KIMBALL:** The Microsoft Data Warehouse Toolkit, Wiley Publishing, Inc, 2005.
6. * * *: Solid Quality Learning: Microsoft SQL Server 2005, Database Essentials, Microsoft Press, 2007.
7. * * *: <http://www.microsoft.com/sql/solutions/bi/default.mspx>