



Volumul VII, Numărul 13 / 2005

ISSN 1454-9980

**Data warehousing,
pilonul central al procesului de exploatare a datelor**

(pag. 31-38)

Bogdan ANASTASIEI

Volume VII, Issue 2 (13) / 2005

**Cross-cultural
Management
Journal**

DATA WAREHOUSING, PILONUL CENTRAL AL PROCESULUI DE EXPLORARE A DATELOR

Bogdan ANASTASIEI

Rezumat

Data warehousing is a technology that allows any decision maker to get access, at any moment, to all the information he needs, no matter the place where this information can be found. In a nutshell, it is a set of concepts and tools meant to provide a flexible and efficient way to use the most important asset of the organization: data and information.

Cuvinte cheie: *data warehousing, data mining, sistem informațional, depozit de date*

Una din condițiile absolut necesare pentru supraviețuirea firmei, atât în anii '90, cât și la începutul mileniului al treilea, este capacitatea de a analiza schimbările din mediul de afaceri, precum și de a reacționa la acestea cât mai rapid posibil. Pentru a face acest lucru, managerii superiori și analiștii din întreprindere au nevoie de informație în cantitate mai mare și de calitate mai bună.

Tehnologia informațională a determinat schimbări revoluționare în privința modului în care operează organizațiile din zilele noastre. Din păcate însă, deși există un număr tot mai mare de computere pe birourile angajaților, iar rețelele de comunicații împânzesc tot globul, foarte mulți manageri și decidenți nu reușesc să obțină accesul la informațiile critice care există deja în organizație.

În fiecare zi, organizațiile mari creează milioane de octeți de date cu privire la toate aspectele afacerii lor, mii de informații cu privire la clienți, produse, angajați și operații. Problema este că aceste date se găsesc stocate în sute sau mii de computere din organizație, astfel încât este deosebit de dificil pentru un analist să localizeze datele de care are nevoie la un moment dat. Experții estimează că numai o parte

foarte mică a datelor culese, procesate și stocate în întreprindere este realmente la dispoziția managerilor și decidenților. Deși tehnologiile destinate manipulării și prezentării datelor au cunoscut o dezvoltare explozivă în ultimul deceniu, abia în ultimii ani specialiștii în tehnologia informatică au ajuns la concluzia că multe din deciziile luate de responsabilii din organizații nu au la bază o cantitate suficientă de informații, tocmai din cauza problemei expuse mai sus.

Soluția la aceste probleme o reprezintă noțiunea de "data warehousing". Este vorba de o tehnologie care permite oricărui decident din organizație să acceseze în orice moment toate informațiile de care are nevoie, indiferent la ce nivel s-ar găsi aceste informații. Tehnologia *data warehousing* (pe care o vom numi aici "depozitarea datelor") reprezintă de fapt un set de concepte și instrumente create în urma tentativelor repetate ale cercetătorilor de a oferi organizațiilor o modalitate flexibilă și eficientă de accesare și utilizare a celui mai important activ de care ele dispun: datele și informațiile.

Conceptul de depozitare a datelor nu a apărut "peste noapte". El a rezultat din integrarea tehnologiilor și experienței acumulate

în industria informațională pe parcursul ultimelor două decenii. Această experiență le-a permis specialiștilor IT să identifice problemele cheie care trebuiau rezolvate și să vină cu soluțiile cele mai potrivite.

Sisteme operaționale și sisteme informaționale

Una din cele mai importante idei pe care o datorăm creatorilor conceptului de depozitare a datelor este aceea că există două tipuri principale de sisteme de informație în organizație: sisteme operaționale și sisteme informaționale.

Sistemele operaționale, așa cum arată și denumirea lor, nu fac altceva decât să ajute angajații să execute operațiunile curente din organizație. Aceste operațiuni sunt, de exemplu, înregistrarea comenzilor, verificarea stocurilor, calculul și plata salariilor, înregistrările contabile și multe altele. Având în vedere importanța lor pentru organizație, sistemele operaționale au fost primele care au constituit obiectul informatizării. De-a lungul anilor, programele care asistă funcționarea acestor sisteme au fost în continuu extinse și perfecționate, până la punctul în care au fost integrate complet în organizație. Astăzi, marea majoritate a organizațiilor mari nu ar putea funcționa fără sistemele lor operaționale și fără datele pe care aceste sisteme le înregistrează și procesează.

Pe de altă parte, în organizație există o altă categorie de funcțiuni, mai complexe decât cele exemplificate mai sus. Este vorba de funcțiuni care au de-a face cu planificarea strategică, cu prognoza și managementul întreprinderii. Acestea sunt de asemenea extrem de importante pentru supraviețuirea și dezvoltarea firmei. Funcții precum “planificarea de marketing” sau “analiza financiară” au și ele nevoie de sisteme computerizate. Dar acestea sunt diferite de sistemele operaționale, iar tipurile de informații de care au ele nevoie sunt de asemenea diferite. Avem de-a face în acest caz cu sisteme informaționale, care sunt sisteme de management al cunoștințelor în adevăratul sens al cuvântului.

Sistemele informaționale sunt sisteme de analiză a datelor și de luare a deciziilor (adesea decizii majore cu privire la strategia de viitor a organizației). Sistemele informaționale nu au doar alt scop decât cel al sistemelor

operaționale, ci au și o cu totul altă dimensiune. În vreme ce sistemele operaționale se concentrează de regulă asupra unui singur domeniu (de exemplu, contabilitatea, gestiunea producției sau gestiunea resurselor umane), sistemele informaționale operează întotdeauna cu date dintr-un număr mare de domenii și au nevoie de cantități mari de date *intercorelate*. Conceptul de “data warehousing” are în vedere cu precădere sistemele informaționale, și într-o măsură mai redusă sistemele operaționale.

Rolul sistemelor operaționale (numite uneori în literatura de specialitate anglo-saxonă “*business applications*”) este aceea de a asigura desfășurarea zilnică, săptămânală și lunară a operațiunilor întreprinderii. Dacă aceste sisteme nu mai funcționează, întreprinderea își încetează activitatea.

În ceea ce privește sistemele informaționale (numite uneori și “*about-the-business applications*”), rolul lor este acela de a analiza activitățile din organizație. Ele sunt folosite pentru a interpreta evenimentele din trecut și pentru a lua decizii cu privire la acțiunile ce vor fi întreprinse în viitor. Atunci când sistemele informaționale nu mai funcționează, organizația nu-și încetează activitatea, dar competitivitatea sa pe termen lung va fi mult diminuată.

Un sistem informațional eficient posedă următoarele caracteristici:

- permite încărcarea rapidă a datelor din baza de date, cu ajutorul unor aplicații de tip OLTP (*Online Transaction Processing*) sau OSS (*Operations Support Systems*);
- permite acumularea și stocarea unei cantități foarte mari de date;
- permite salvarea și recuperarea rapidă a datelor, în cazul unor defecțiuni în funcționarea sistemului;
- facilitează prelucrarea unui mare număr de interogări ale utilizatorilor;
- dispune de facilități flexibile de import și export al datelor;
- permite utilizarea în comun a informației de către departamentele interesate;
- dispune de un modul de analiză care poate include instrumente de realizare a graficelor și diagramelor, editor de rapoarte, instrumente de modelare statistică, foi de calcul, simulatoare, limbaje de interogare, procesoare de text, aplicații pentru *desktop publishing*, software pentru managementul proiectelor, aplicații care utilizează

inteligența artificială, software pentru *data mining*, instrumente de dezvoltare a aplicațiilor (limbaje de programare) și altele.

Data warehousing: definiție și aplicații

Termenul de *data warehouse* (depozit de date) a fost folosit și definit pentru prima dată de William Inmon, în anul 1990, în cadrul articolului cu titlul “*What is a Data Warehouse?*”, apărut în revista *Prism*. Definiția pe care Inmon o dă acestui concept este următoarea: “Un depozit de date este o colecție de date structurate pe subiecte, integrate, dependente de timp și nevolatile, date care vin în sprijinul procesului managerial de luare a deciziilor”. Termenii care apar în această definiție au semnificațiile următoare:

- *date structurate pe subiecte*: date grupate astfel încât să ofere informații cu privire la un anumit subiect sau domeniu, și nu asupra operațiunilor curente ale întreprinderii;
- *date integrate*: depozitul conține date colectate dintr-un mare număr de surse și grupate într-un ansamblu coerent;
- *date dependente de timp*: toate datele dintr-un depozit de date sunt legate de un anumit moment din timp;
- *date nevolatile*: datele dintr-un depozit sunt stabile. Odată cu trecerea timpului sunt adăugate date noi, dar datele deja existente nu sunt șterse niciodată. Acest lucru îi permite managerului sau analistului să aibă în permanență un tablou de ansamblu, foarte consistent, al activității organizației.

Definiția lui W. Inmon rămâne valabilă și acum, la 15 ani după ce a fost formulată, deși există situații în care ea nu este respectată întru totul. De exemplu, multe din datele existente în depozitele de date de astăzi sunt volatile. Din cauza spațiului de stocare uriaș necesitat de depozitele de date (de foarte multe ori de ordinul teraocteților), doar un anumit număr de perioade istorice sunt menținute în depozit. Spre exemplu, dacă managerul decide că are nevoie de datele financiare din ultimii trei ani ca “materie primă” pentru un model de analiză, imediat ce sunt adăugate datele pentru ultima lună, datele corespunzătoare primei luni din serie (cea mai veche) sunt automat șterse.

Un alt specialist în domeniu, Ralph Kimball, a dat o definiție mai simplă noțiunii de *data warehousing*: “o colecție de date structurate special pentru a fi interogate și analizate”. Deși mai puțin detaliată decât cea a lui Inmon, definiția lui Kimball nu este mai puțin precisă.

Depozitele de date reprezintă un domeniu foarte nou, astfel încât este greu de anticipat care sunt modificările ce-l vor afecta în viitor. Una din cele mai importante pare a fi apariția așa numitor servere de baze de date paralele, care utilizează tehnologii de interogare cu performanțe superioare. Acest tip de servere vor face posibil accesul la baze de date uriașe, într-un timp mult mai scurt. O altă posibilă direcție de dezvoltare este cea a tehnologiilor care permit procesarea de date mixte: cifre, text și multimedia. Crearea unor instrumente avansate de vizualizare a datelor le va permite utilizatorilor să detecteze aspecte care le scăpaseră până acum din vedere.

Iată câteva posibile modalități de utilizare a sistemelor de *data warehousing*, pentru trei domenii diferite: comerțul en detail, telecomunicațiile și sectorul financiar-bancar.

Tabelul 1. Aplicații ale procesului de data warehousing

Modalități de utilizare a tehnologiilor de depozitare a datelor
Comerț en detail
Analiza datelor culese de scannerele de la casele de marcat
Analiza rezultatelor campaniilor de promovare (reduceri de prețuri, cupoane etc.)
Analiza stocurilor
Fundamentarea strategiilor de preț
Negocierea prețurilor cu furnizorii
Gestiunea programelor de facilități pentru clienții fideli
Analiza profitabilității
Segmentarea clientelei

<p><i>Telecomunicații</i></p> <p>Analiza:</p> <ul style="list-style-type: none"> - numărului și duratei convorbirilor; - vânzărilor de echipamente; - profitabilității clienților; - costurilor; - stocurilor. <p>Negocierea prețurilor cu furnizorii Gestiunea programelor de facilități pentru clienții fideli</p>
<p><i>Finanțe și bănci</i></p> <p>Gestiunea relațiilor cu clienții Segmentarea clientelei și targeting Analiza riscului de credit Analiza oportunităților de fuziuni și achiziții Construirea profilului clientelei Analiza performanțelor sucursalelor bancare</p>

Structura sistemelor de data warehousing

În cele ce urmează vom expune pe larg arhitectura unui sistem de depozitare a datelor, așa cum se prezintă ea în majoritatea organizațiilor. Ceea ce înțelegem aici prin “arhitectură” este o modalitate de a prezenta structura generală a comunicării, procesării și prezentării datelor, aflată la îndemâna utilizatorului final din organizație. Această structură este, după cum se va observa, destul de complexă (acesta este de fapt unul din motivele pentru care a fost necesar un timp atât de lung pentru dezvoltarea acestei tehnologii). Cunoașterea arhitecturii unui sistem de *data warehousing* este foarte importantă pentru a înțelege cum sunt legate între ele componentele implicate în strategia de construire a unui astfel de sistem.

Un sistem de depozitare a datelor este organizat în general pe opt componente, și anume:

- baza de date operațională;
- componenta de acces la informație;
- componenta de acces la date;
- directorul de date;
- componenta de procesare a datelor;
- componenta de transmisie a datelor;
- componenta de depozitare fizică a datelor;
- componenta de prezentare a datelor.

Să le prezentăm pe fiecare pe rând.

Baza de date operațională

Scopul sistemelor operaționale este acela de a procesa datele de care firma are nevoie în activitățile sale curente. Pentru acest motiv au fost create, pe parcursul vieții organizației, bazele de date operaționale. Ele au capacitatea de a procesa datele pentru un număr relativ redus de tranzacții, bine definite. Totuși, din cauza dimensiunilor limitate ale sistemelor operaționale, bazele de date create pentru a sprijini aceste sisteme au dificultăți atunci când este vorba de accesarea datelor pentru alte scopuri decât cele pentru care au fost concepute inițial. Această dificultate este amplificată de faptul că foarte multe sisteme operaționale au o vechime de 10 sau 15 ani. Din cauza acestei vechimi, tehnologia utilizată pentru accesul la datele operaționale este adesea depășită.

Scopul unui sistem de depozitare a datelor este acela de a “elibera” informațiile conținute în bazele de date operaționale și de a le utiliza în combinație cu datele obținute din alte surse, adesea externe. Tot mai multe organizații capturează informații din baze de date exterioare. Aceste informații privesc caracteristici demografice, comportamente de cumpărare, tendințele de pe piață, acțiunile concurenței și așa mai departe. Așa numita “autostradă a informației” oferă accesul la tot mai multe resurse pe zi ce trece.

Componenta de acces la informație

Aceasta este componenta pe care utilizatorul final o folosește în mod direct. De fapt, ea este reprezentată de instrumentele folosite în mod obișnuit de un analist pentru a manipula datele: Microsoft Excel, Lotus 1-2-3, Microsoft Access, SPSS, SAS etc. Această componentă include atât elemente hard, cât și soft, necesare pentru generarea și afișarea de rapoarte și grafice pentru analiza și prezentarea datelor. În ultimele două decenii, această componentă s-a dezvoltat foarte mult, ca urmare a apariției și răspândirii computerelor personale.

Astăzi, utilizatorii finali au la dispoziție foarte multe instrumente sofisticate de analiză și prezentare a datelor; totuși, accesul facil la datele brute existente în bazele de date operaționale pune încă probleme. Soluția ar putea fi inventarea unui limbaj de programare comun, care să poată fi utilizat pentru accesarea tuturor structurilor de date din organizație.

Componenta de acces la date

Această componentă face legătura între componentele precedente, permițând comunicarea între ele. Ea este așadar o interfață între bazele de date operaționale și sistemele care permit manipularea datelor (analiză și prezentare). În rețelele de astăzi, limbajul de accesare a datelor cel mai utilizat este SQL (*Structured Query Language*). La început, SQL a fost dezvoltat de IBM ca limbaj de interogare, dar în ultimii 20 a devenit standardul în materie de limbaj de accesare a datelor. Crearea unor filtre speciale de acces al datelor (de exemplu, EDA) permite acum limbajului SQL să acceseze aproape orice tip de bază de date, relațională sau nerelațională. Mai mult, ele fac posibilă interogarea unor baze de date care au o vechime de zeci de ani.

Într-un sistem de depozitare a datelor eficient, componenta de acces la date poate utiliza baze de date situate în orice tip de rețea, indiferent de protocolul acesteia, și găzduite pe orice fel de sisteme hardware, indiferent de producător. Unul din elementele cheie ale conceptului de depozitare a datelor este acela de a oferi utilizatorilor finali ceea ce se numește “acces universal la date”. În mod ideal, acești utilizatori ar trebui să fie capabili să acceseze toate datele care le sunt necesare în activitatea lor, fără nici un fel de restricții.

Directorul de date

Pentru a putea asigura accesul universal la date, este absolut necesară existența unui așa numit director de date, cunoscut și sub denumirea de “depozit de meta-date”. Meta-datele sunt datele despre datele din organizație. Ele sunt de regulă instrucțiuni specializate din limbajele de programare, care ne indică tipurile de date care se găsesc în baza de date operațională, locul unde sunt depozitate datele și modul în care pot fi accesate. Directorul de date permite utilizatorului final să acceseze o informație fără a ști neapărat unde se găsește aceasta (în care bază de date sau în care înregistrare).

Componenta de procesare a datelor

Rolul acestei componente este acela de a planifica (programa) diversele sarcini care trebuiesc îndeplinite pentru a construi, întreține și utiliza sistemul de depozitare a datelor. Componenta de procesare ar putea fi gândită ca un fel de “supervizor” de înalt nivel al tuturor proceselor din sistem.

Componenta de transmisie a datelor

Rolul său este acela de a transporta informația prin rețelele de calculatoare a organizației. În general, ea este constituită dintr-un ansamblu de protocoale care permit comunicarea datelor între rețele, dar uneori poate fi mai mult decât atât. Componenta de transmisie a datelor ar mai putea fi utilizată pentru a colecta tranzacții sau mesaje și pentru a le transmite la o anumită locație, într-un anumit moment din timp.

Componenta fizică de depozitare a datelor

Aici este locația în care se găsesc de fapt datele și informațiile folosite de sistemul de *data warehousing*. Există unele situații în care procesul de depozitare a datelor nu presupune stocarea fizică a datelor. În acest caz, această componentă are mai curând o natură virtuală.

În cadrul sistemului fizic de depozitare, datele operaționale și externe sunt stocate într-o formă care să permită un acces ușor și cât mai flexibil cu putință. Tot mai mult, sistemele de depozitare a datelor sunt instalate pe platforme de tip client/server, dar uneori pot fi instalate și pe computere de mare capacitate (mainframe).

Componenta de prezentare a datelor

Această componentă include toate procesele necesare pentru a încărca, selecta, edita, condensa și combina datele capturate fie din bazele de date operaționale, fie din sursele externe.

Componenta de prezentare a datelor include adesea limbaje de programare complexe, dar au fost create instrumente care ajută utilizatorul final să o folosească, chiar și dacă nu are nici un fel de cunoștințe de programare. Această componentă poate include de asemenea instrumente de analiză care permit identificarea de modele și structuri în interiorul datelor brute capturate din bazele de date operaționale.

Procesul de construire a unui depozit de date

Procesul de creare a unui depozit de date nu este unul ușor. El solicită un efort care presupune pe de o parte o înțelegere exactă a tipului de bază de date ce urmează a fi creat și a modulelor de analiză care vor fi puse la dispoziție, iar pe de altă parte un contact permanent cu potențialii utilizatori. Design-ul unui depozit de date este adesea unul iterativ. Fiecare etapă a procesului cere o atenție deosebită, întrucât odată ce sistemul este populat cu o mare cantitate de date, orice schimbare devine foarte dificilă.

În procesul de constituire a unui sistem de depozitare a datelor trebuie ținut seama de trei criterii, în funcție de care vor fi definite opțiunile cu privire la arhitectura sistemului. Aceste criterii sunt:

- dimensiunea sistemului;
- redundanța datelor;
- tipul de utilizator final.

Dimensiunea sistemului

Un sistem de depozitare a datelor poate fi conceput pentru a procesa și analiza toate datele stocate în întreprindere, de la înființarea ei și până în momentul de față, sau poate fi creat în scopul de a furniza date pentru un singur manager din firmă, pe parcursul unui singur an. Ambele sisteme pot fi numite în egală măsură sisteme de *data warehousing*. Totuși, cu cât dimensiunea este mai mare, cu atât sistemul va fi mai valoros pentru organizație. Totodată, crearea

și întreținerea unui sistem de dimensiuni mari este dificilă și costisitoare. Din acest motiv, multe întreprinderi preferă să înceapă prin a construi sisteme de *data warehousing* la nivel funcțional, departamental sau divizional, urmând apoi să le extindă la nivelul întregii organizații

Redundanța datelor

Există trei nivele de redundanță a datelor pe care organizația ar trebui să le ia în considerare atunci când definește opțiunile pentru sistemul de depozitare a datelor:

- *depozitele de date "virtuale"*. O asemenea strategie presupune ca utilizatorii finali să poată accesa direct bazele de date operaționale, folosind orice instrumente doresc. Această opțiune are avantajul unei deosebite flexibilități, reducând totodată la minimum nivelul redundanței datelor. Dezavantajul acestei abordări este riscul supraîncărcării bazelor de date cu un număr foarte mare de interogări, ceea ce ar duce la încetinirea funcționării sistemului.

Această strategie este utilizată la început în multe organizații unde există un număr mare de utilizatori finali având fiecare o gamă largă de necesități, iar frecvența probabilă a solicitărilor bazei de date este redusă. Depozitele de date "virtuale" sunt foarte utile pentru a ne putea face o idee inițială asupra tipurilor de date care sunt solicitate cel mai frecvent de diversele categorii de utilizatori finali.

- *depozitele de date centrale* reprezintă o singură bază de date, ce conține toate datele necesare pentru o anumită arie funcțională, sau pentru un anumit departament ori divizie. Această opțiune este utilizată atunci când avem de-a face cu un număr mare de utilizatori finali, ale căror computere sunt conectate la o unitate centrală, și care au nevoie în general cam de aceleași informații. De regulă, depozitele de date centrale conțin date care privesc sisteme operaționale multiple.

Depozite de date de acest tip sunt "reale". Ele sunt construite pe structura unor baze de date relaționale sau cu ajutorul unui server de baze de date multidimensional.

- *depozite de date dispersate*. Exact așa cum arată denumirea lor, aceste depozite conțin o serie de componente care sunt dispersate (distribuite) în mai multe baze de date fizice. Întrucât în organizațiile mari

responsabilitatea adoptării deciziilor este tot mai mult împinsă către nivelele de jos ale organizației, este nevoie ca informațiile necesare să fie transmise operativ către aceste nivele. Depozitele de date dispersate servesc foarte bine acestui scop. Dezavantajul lor este acela că implică o foarte mare redundanță a datelor și, drept consecință, o durată mai mare de încărcare și actualizare a datelor.

Tipul de utilizator final

Așa cum există foarte multe modalități de organizare a unui depozit de date, există și foarte multe categorii de utilizatori finali. În general acești utilizatori sunt împărțiți în trei mari categorii:

- manageri;
- utilizatori “avansați” (analști, ingineri etc.);
- utilizatori “secundari” (este vorba aici în special de personalul administrativ).

Fiecare din aceste categorii are propriile sale cerințe în ceea ce privește tipul de date, modul de acces, flexibilitatea și ușurința utilizării.

Se pot descrie patru etape ale procesului de construirea unui sistem de *data warehousing*:

- conceperea unei strategii;
- conceperea unei arhitecturi a sistemului;
- proiectarea sistemului;
- gestiunea depozitului de date.

Etapa 1: Conceperea unei strategii pentru crearea unui sistem de depozitare a datelor

În această etapă urmează a se răspunde la o serie de întrebări esențiale: pentru cine va fi construit sistemul, ce dimensiune va avea, ce tip de sistem va fi creat?

O posibilă modalitate de definire a strategiei este aceea de a crea în primă instanță un depozit de date “virtual”. Aceasta presupune instalarea unei baze de date în cadrul organizației, instruirea potențialilor utilizatori și monitorizarea utilizării bazei de date. În acest fel se poate determina ce informații sunt cerute mai des, de către cine și în ce scop. Această metodă este simplă și rapidă. Există totuși și potențiale riscuri: dacă datele din bază sunt de calitate slabă, sau dacă sistemul de acces la date nu este suficient de bine pus la punct, concluziile trase în final s-ar putea dovedi eronate.

O altă modalitate constă în formarea unui eșantion de utilizatori din organizație și chestionarea lor, pentru a afla care sunt problemele cu care se confruntă, de ce date au nevoie și cum doresc să le prelucreze. Pe baza rezultatelor acestei anchete este construit un prototip, pe care potențialii utilizatori sunt invitați să-l experimenteze și să-l evalueze. În urma experimentelor, utilizatorii vor aduce sugestii pentru îmbunătățirea și adaptarea sistemului. Datele necesare pentru funcționarea prototipului pot fi datele existente în întreprindere și/sau datele culese din surse externe. Dacă este cazul, pot fi create componente de acces la informație care să le permită utilizatorilor să folosească instrumentele lor de analiză preferate.

Nici o strategie de construire a unui sistem de *data warehousing* nu este singura potrivită în toate situațiile. Contextul fiecărei organizații este diferit, nevoile sunt diferite. Mai mult decât atât, tehnologiile de creare a depozitelor de date sunt în continuă dezvoltare; așadar, pot oricând apărea noi soluții care să le depășească pe cele existente din punctul de vedere al performanțelor.

Etapa 2: Conceperea arhitecturii sistemului de depozitare a datelor

Într-una din secțiunile anterioare am prezentat o descriere a componentelor arhitecturii unui sistem de *data warehousing*. Acea prezentare oferă numai un cadru general, orientativ, care ne poate ajuta să înțelegem modalitatea de “asamblare” a unui astfel de sistem. Organizațiile sofisticate vor putea pune la punct un sistem complet încă din start, în timp ce organizațiile mai mici și cu resurse financiare mai puține vor trebui să se mulțumească cu tehnologii mai simple, cel puțin pentru început.

Una din caracteristicile-cheie ale unui sistem de depozitare a datelor este *flexibilitatea*. Există un lucru foarte important care nu trebuie uitat: cu cât un asemenea sistem se dovedește a fi mai util, cu atât utilizatorii vor dori să i se adauge noi opțiuni sau noi funcționalități. Arhitectura sistemului, precum și tehnologiile folosite, ar trebui să permită cu ușurință acest lucru.

Etapa 3: Proiectarea sistemului de depozitare a datelor

Procesul de proiectare a unui depozit de date este foarte diferit de cel de proiectare a unui sistem operațional tradițional. În primul rând, dacă utilizatorii unui sistem informațional știu foarte bine de la început ce doresc de la acest sistem, nu același lucru se poate spune despre potențialii utilizatori ai unui depozit de date. În al doilea rând, proiectarea unui depozit de date necesită un mod de gândire în termeni mult mai largi și mult mai greu de definit. Proiectanții unui sistem de *data warehousing* vor trebui să facă eforturi apreciabile pentru a-i ajuta pe potențialii utilizatori să-și “vizualizeze” nevoile, să le conștientizeze. Existența unor prototipuri robuste, funcționale, este foarte importantă în această privință.

Etapa 4: Gestiunea sistemului de data warehousing

Depozitele de date nu sunt sisteme magice – folosirea lor cere foarte multă muncă. În multe cazuri, un sistem de depozitare a datelor este văzut ca o soluție miraculoasă, care le permite managerilor și angajaților să-și ducă sarcinile la îndeplinire într-un timp scurt și fără nici un fel de efort. În realitate, un astfel de sistem solicită eforturi serioase, atât de management cât și de marketing, la nivelul întregii organizații. El constituie o investiție bună numai dacă utilizatorii finali pot obține și procesa

informațiile importante mai ușor și mai rapid decât cu ajutorul tehnologiilor actuale. Managementul organizației trebuie să răspundă în detaliu la următoarele întrebări:

- cum vrem ca angajații noștri să folosească sistemul? ce performanțe așteptăm de la ei la acest capitol?
- cum îi vom informa pe utilizatorii finali cu privire la existența și avantajele sistemului și cum îi vom instrui astfel încât să-l folosească la parametrii optimi?

În sfârșit, managementul organizației trebuie să înțeleagă faptul că un sistem de depozitare a datelor necesită eforturi de întreținere, la fel ca oricare activ sau aplicație din organizație. De fapt, experiența a arătat că acest sistem devine în scurt timp unul din cele mai folosite sisteme din întreprindere.

Data warehousing este un domeniu foarte complex, din cauza numeroaselor tehnologii disponibile și a existenței unui număr mare de furnizori care încearcă să impună pe piață propriile soluții tehnologice. Întrucât prețul unei asemenea soluții este de regulă foarte ridicat, organizația trebuie să evalueze cu atenție toate alternativele înainte de a lua decizia de achiziție. În ciuda acestor dificultăți, beneficiile potențiale pe care le aduce un sistem de depozitare a datelor fac acest concept imposibil de ignorat.

Bibliografie

1. Blackwood, Ph. – *11 Steps to Successful Data Warehousing*, <http://www.databasedadvisor.com>
2. Gilbrough, P. – *An Introduction to Data Mining*, <http://www.mgmt-info-tools.com>
3. Reed, M. – *A Definition of Data Warehousing*, <http://idm.internet.com>
4. Wasserman, M. – *Mining Data*, Regional Review, Quarter 3, 2000, vol. 10, no. 3



Bogdan Anastasiei (n. 1970) este conferențiar universitar doctor la Facultatea de Economie și Administrarea Afacerilor, Catedra de Management și Marketing. Preocupările sale se situează în domeniile cercetărilor de marketing, marketingului turistic și managementului riscului. A publicat 4 cărți și peste 30 de articole științifice în reviste și volume de specialitate.