

EXTENSII ALE BAZELOR DE DATE DISTRIBUITE

Asistent dr. Dan-Andrei Sitar-Tăut - Universitatea „Babeş-Bolyai” Cluj-Napoca, Facultatea de Ştiinţe Economice şi Gestiunea Afacerilor

Rezumat:

Today there are many opinions supporting database distributed systems and centralized database systems and considering them as a guaranteed solution, even if some cases prove this is not the most efficient approach. We will try to fight against to this preconceived idea through reveal of some technologies as federated database systems, which are applicable in some categories of circumstances. Both, the latter ones and distributed database systems could represent a solid platform for interfacing technologies such as: data warehouses, data marts, OLAP and OLTP systems, data mining, web, etc. This paper aims making familiar to few alternatives for the classic distributed systems, analyzing those technologies having distributed systems upstream placed, and also the causality inter-relationships that exist between them.

BAZE DE DATE FEDERATIVE

Multă lume mai crede că atunci când vrem să extragem informații despre clienții noștri, nu trebuie decât să lansăm o simplă întrebare adresată bazei de date de clienți stocată pe un singur nod de la care să primim și răspunsul. În realitate lucrurile nu ar trebui să arate așa. Datele privind clienții sunt deja stocate în diferite locații din cadrul firmei. De ce să mai duplicăm informația existentă prin aducerea ei într-un singur loc? Sistemul va distribui cererea către nodurile din jurul său și va obține în cele din urmă răspunsul. [Winter 2002]

În ultimul deceniu s-a conturat o modă – necesară am putea spune – în ceea ce privește **sistemele de baze de date multiple**, cunoscute în literatura de specialitate ca și **MultiDataBase Systems (MDBS)**.

Definiție: *Sistemele de baze de date multiple reprezintă un sistem de baze de date distribuite în care fiecare sit păstrează o autonomie totală.* [Connolly et al. 2001]

În general, rolul acestora este de a se interconecta logic mai multe sisteme de baze de date în vederea atingerii unui țel comun. Sistemele locale își păstrează controlul complet asupra bazei de date prin intermediul aplicațiilor existente. De aceea, sistemul global nu poate aduce atingeri asupra structurii sau aplicațiilor care rulează local. Dialogul dintre cele două nivele se face prin intermediul unei interfețe software și la nivelul datelor pe care situl local consimte să le partajeze. Putem să avem sisteme de baze de date multiple în care nu avem utilizatori locali. Acestea se numesc **sisteme de baze multiple nefederative**. [Sitar 2005-3]

Principiul federalizării, care s-a dovedit a fi un tipar comportamental atât în cazul organismelor biologice, dar și în comunitățile de succes, poate fi de asemenea un curs firesc de urmat și în cazul corporațiilor. Acest principiu presupune existența unor entități mici care își pot îndeplini autonom – pe cât se poate cu puțință – sarcinile, fără a se abate atât de la obiectivul individual, cât și global de sporire a profitului și de micșorare a costurilor. Dacă principiul federalizării este acceptat ca un comportament strategic în viața reală, atunci de ce nu, și structurile informatice ar putea fi supuse proiectării conform aceluiași principiu. [Schmitt & Saake 2005], [Sitar 2005-2], [Sitar 2005-3] Și sistemele colaborative multiagent se bazează pe acest principiu.

Fiecare membru al federației este și va rămâne independent atât timp cât activitatea lui urmărește și binele întregii federații. Ca principiu antreprenorial, federalismul promovează expansiunea clară și diferențiată a rețelelor și proceselor corporatiste. Din punctul de vedere

structural și organizațional, sistemul federativ protejează mai micile entități împotriva coloșilor, fără a fi private de sprijinul de care au nevoie pentru desfășurarea activității și – paradoxal – conferindu-le un nivel ridicat de libertate.

Necesitatea în continuă creștere de cooperare între diferite entități impune un acces integrat la baze de date distribuite heterogene, dar autonome. [Sitar 2005-2] Acest acces trebuie să confere imaginea unei singure baze de date și necesită interconectarea sistemelor de baze de date prin intermediul unei rețele de comunicații și suprapunând un nivel software deasupra sistemelor de gestiune a bazelor de date în vederea susținerii comunicării, să partajeze anumite date, însă să păstreze autonomia de la nivelul local. [W3C FDMS Group] La începutul anilor '80 integrarea diferitelor surse heterogene de informații a devenit un important și interesant domeniu de investigație. Până să se sedimenteze denumirea de sisteme federative, în publicațiile acelor vremuri ele figurau fie ca "next-generation gateways", "data access middleware" sau "multi-database servers". Activitățile din acest domeniu și-au propus combinarea sistemelor informatice existente. Realizarea unei integrări iscusite va permite accesul transparent la aceste sisteme heterogene.

Definiție: *Un sistem de baze de date federative (FMDBS¹) este o colecție de sisteme de baze de date colaborative care sunt autonome și posibil heterogene.* [Sheth & Larson 1990]

[Buch 2002] și [Nukpe 2001] susțin ideea că bazele de date federative reprezintă unificarea logică a unor baze de date distincte ce rulează pe servere independente, (în general descentralizate geografic) fără partajare de resurse și conectate prin intermediul unui LAN. Reprezentând un sistem transparent și integrat de metabaze de date, ele modelează universuri similare sau complementare [Bouguettaya 1998]. Cât timp bazele de date rămân autonome, federalizarea constituie o alternativă viabilă în raport cu principiul centralizării. [Wikipedia 2005]

Cea mai sugestivă exemplificare a conceptului o reprezintă însuși Internetul, care, la rândul său, reprezintă un sistem federativ de dimensiuni mari, chiar dacă au la bază tehnologii diferite.

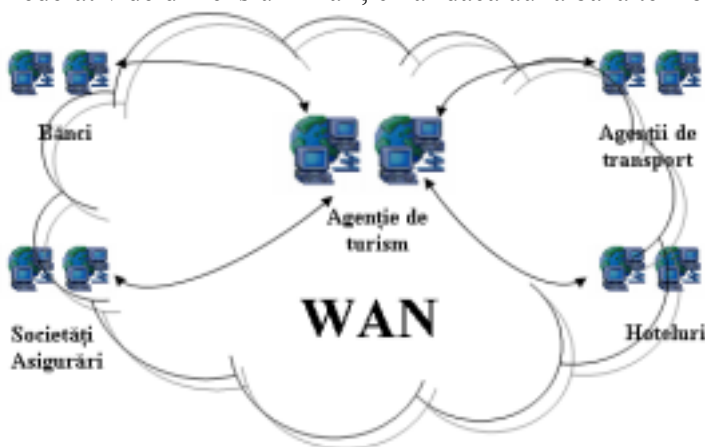


Figura 1 – Sistem de baze de date federative peste mai multe organizații

În **Figura 1** se observă că avem un sistem de baze de date federative al unei agenții de turism care interconectează sistemele unor tipuri de instituții cu domenii de activitate diferite. O agenție de turism colaborează cu agențiile de transport pentru a-i putea prezenta clientului condițiile oferite de parteneri și pentru a putea face rezervări. Tot din necesitatea de întâmpinare a nevoilor clientului, sunt necesare și comunicarea cu băncile – ofertă de instrumente comode de plată, acordarea la fața locului a unor eventuale credite pentru excursii, și chiar pentru verificarea bonității clientului sau a situației încasărilor, tranzacții bancare cu partenerii – și cu societățile de asigurări (asigurări medicale pentru străinătate sau țară, a mijloacelor de transport proprii și pentru punerea la adăpost față de situații neprevăzute ce pot rezulta din activitatea pe care o prestează). Este de la sine înțeles că activitatea partenerilor va fi autonomă, adică agenția nu va putea efectua orice și pentru

¹ Federated MultiDataBase System

oricine tranzacții bancare sau să modifice salariul angajaților din hoteluri. Țelul comun al federației este de maximizare a profiturilor individuale și deci și a celui global, câștigarea de noi clienți, reducerea timpilor de desfășurare a unor tranzacții obișnuite etc. [Sitar 2005-1]

Clasa bazelor de date heterogene și a sistemelor care le gestionează poartă adjectivul „federative”, ceea ce înseamnă că ele sunt cuplate doar pentru a întruni anumite cerințe bine definite ale organizației, chiar dacă ele au fost create pentru alte aplicații specifice, au constrângeri de integritate proprii și necesită măsuri de securitate pentru a-și păstra funcționalitatea și autonomia. [Hsiao 1992] Bazele de date din cadrul acestui sistem alcătuiesc o federație.

Un sistem federativ reprezintă o îmbinare între un sistem distribuit și unul centralizat, adică un sistem distribuit pentru utilizatorii globali și unul centralizat pentru cei locali. Nu aderăm întru totul la această accepțiune pe motivul că utilizatorii sistemului local ar putea fi tot atât de bine utilizatori ai unui sistem distribuit, iar în al doilea rând un sit nu „și-ar putea permite” libertatea să-i condiționeze virtualului nod central ce date îi va pune la dispoziție, ce tranzacții sau ce interogări îi dă voie să lanseze. Un al treilea punct de vedere ar fi că dacă am avea de-a face cu o cădere a nodului central, activitatea sitului nu are de ce să fie întreruptă.

Potrivit lui Félix Saltor, există trei elemente esențiale care caracterizează sistemele federative: **autonomia, heterogenitatea și distribuirea.**²

Caracterul heterogen al datelor provenite din varii surse reprezintă o mare și continuă provocare adresată tehnicii, de mai bine de un deceniu. Să luăm drept exemplu terminologia bancară. Fiecare bancă își are propriile denumiri și caracteristici pentru instrumentele de creditare. Privite la nivel internațional, însemnele monetare și ratele de schimb sunt de asemenea diferite. Să nu uităm și faptul că în lumea financiară se constituie acele consorții bancare, sau sunt tot mai dese preluările sau absorbțiile unor bănci de către altele. Toate acestea sunt premisele necesității unei comunicări – mai de suprafață sau mai în profunzime – între acestea, fără ca heterogenitatea să devină o barieră informațională.

Așadar, federalismul este un principiu structural și organizațional a cărui componente de bază, independente și autonome își unesc forțele pentru a forma o structură bine încheată de nivel mai înalt care să combine într-o anumită proporție uniformitatea cu diversitatea indispensabile asigurării succesului organizației. [Schmitt & Saake 2005]

De multă vreme informația este considerată un factor-cheie în industrie (și nu numai), ea devenind necesară și pentru stabilirea relațiilor între sursele și serviciile deja existente (de exemplu, pentru cuplarea sectorului administrativ cu cel de producție, ambele având deja baze de date). [W3C IDC #22542] Multe organizații sunt nevoite să interacționeze cu baze de date distribuite, preexistente, heterogene și autonome. Ele trebuie să găsească soluții care să facă posibilă expansiunea bazelor de date și să promoveze partajarea datelor între utilizatori și aplicații, fără eforturi prea mari de implementare. Acest lucru presupune o tehnologie ce furnizează o integrare selectivă, dar totuși controlată a unor astfel de baze de date. Bazele de date federative sunt capabile să ofere diferite nivele de integrare, cu implicații diverse asupra gradului de control în condițiile conservării autonomiei bazelor de date individuale. [Sheth 1991]

Ideea federalizării a prins bine publicului, indiferent că este vorba de comunitatea academică, de firme dezvoltatoare de tehnologii IT, sau de grupuri economice. Sistemele de baze de date distribuite sunt de dorit, însă aria lor de aplicabilitate este relativ redusă în comparație cu sistemele federative. Atunci când pornim o afacere mare care presupune o serie de departamente, birouri, filiale, puncte de lucru dispersate geografic, ne vom construi cu siguranță un sistem distribuit. Pe parcurs însă, atunci când afacerea ta se îngemănează cu o serie de terți și își diversifică domeniul de activitate nu ne vom mai permite să reprojectăm din temelii sistemul informatic sau să-l actualizăm în consecință, în baza concepției inițiale. Reprojectarea unui sistem de baze de date distribuite necesită în afară de resurse financiare importante și o perioadă destul de îndelungată de

² [Saltor 1995], [Nukpe 2001]

implementare. Cum lumea afacerilor se mișcă și se schimbă extrem de rapid, trebuie să găsim o soluție nu doar mai ieftină, cât mai ales puțin consumatoare de timp de implementare. Cu atât mai mult cu cât relațiile cu unii terți sunt sporadice și întotdeauna se va ivi unul nou față de care nu voi avea nici un interes de a realiza o colaborare atât de profundă încât să necesite dezvoltarea unui sistem de baze de date împreună. În astfel de situații soluția ideală este cea de construire a unui sistem federativ peste sistemul nostru și cel al partenerilor de afaceri. Acest lucru nu va afecta autonomia niciunuia dintre noi, fiecare având propria afacere, de cele mai multe cazuri chiar total în domenii diferite. Aceeași abordare se impune și atunci când fuzionăm cu alte companii sau le preluăm.

Bazele de date distribuite și federative reprezintă de multe ori suportul și pentru alte tehnologii, precum depozitele, data mart-urile sau piețele de date, sisteme OLAP și OLTP, data mining, baze de date multimedia, baze de date documentare, baze de date statistice etc. Ele pot fi aplicate în orice domeniu de activitate, chiar și în cazul bibliotecilor on-line, sisteme GIS (Geo-Information System) [Tuladhar 2005], afacerilor pe Internet³, fie că vorba de simple achiziții de bunuri sau de instrumente bancare, asigurări, marketing și multe altele.

DEPOZITE DE DATE

Derularea fenomenului economic asistat de calculator în cadrul organizațiilor a condus la acumularea unor cantități din ce în ce mai mari de date. Unele companii au considerat că menținerea datelor arhaice constituie nici mai mult nici mai puțin decât un balast pentru derularea în bune condiții a activității economice, pe motivul că o bază de date supradimensionată cauzează deficiențe în exploatare. Această afirmație este adevărată în condițiile în care rata de înnoire a sistemelor de calcul și a celor informaționale este una lentă. O altă categorie de firme a înțeles că înnoirea instrumentelor de susținere a activității este necesară. Odată cu acestea, dimensiunea bazei de date nu mai reprezenta o povară. Așa că, mai mult sau mai puțin intenționat s-au creat premisele constituirii unui depozit de date. Cei care au înțeles că o analiză a datelor mai vechi este o cale spre succesul afacerii au câștigat o luptă mută, dar importantă, cu concurența ignorantă. Învățarea din greșelile trecutului prin nerepetarea lor, precum și descoperirea unor șabloane comportamentale care garantează succesul constituie un avantaj.

Definiție: *Depozitele de date (DW⁴) sunt colecții de date nevolatile, orientate spre subiect, integrate, variabile în timp care sprijină managementul firmei în procedeul de luare a deciziilor.* [Inmon 1996]

Spunem că datele sunt nevolatile deoarece ele sunt reactualizate sporadic și menținute în baza de date. Rareori avem de-a face cu actualizări de conținut, depozitele bazându-se pe inserări. Orientarea spre subiect se referă la faptul că depozitul gravitează în jurul „subiectelor” unei organizații – clienți, produse, vânzări – și nu înspre domenii de aplicație (facturare, aprovizionare, desfacere etc.). Integrarea presupune înmagazinarea în același loc, într-o formă unitară și coerentă a datelor provenite de la diverse subiecte sau activități. Variabilitatea în timp se referă la caracterul permanent dinamic al conținutului bazei de date, lucru ce conferă o instabilitate în timp a conținutului, neputând afirma că între două momente de timp mai îndepărtate nu s-a întâmplat nimic semnificativ. [Connolly et al. 2001]

Un aspect deosebit de important este acordat memorării și gestionării metadatelor, adică a datelor despre date. Rolul metadatelor este unul foarte însemnat. Cu ajutorul datelor putem să studiem fără probleme proveniența și chiar întreaga istorie a unui articol. Astfel putem afla compartimentul de la care provine, data intrării în depozit, modificările suferite până în prezent, atât valorice, cât și structurale etc.

Un depozit de date este reprezentat din structuri de date ce sunt optimizate în vederea

³ pentru detalii se poate consulta referința [Sitar & Sitar 2001]

⁴ Data Warehouse

distribuirii lor în cadrul unei rețele de calculatoare. Acesta colectează și înmagazinează seturi de date istorice din multiple sisteme operaționale, alimentând între timp și una sau mai multe piețe de date.⁵ Nu subscriem întru totul la această ipoteză deoarece modalitățile de elaborare a unui depozit pot să difere de modalitatea expusă anterior.

Definiție: *Piețele de date (DM⁶) sunt submulțimi ale depozitelor de date, particularizate în vederea îndeplinirii nevoilor unui departament sau care acoperă un subiect. [Inmon 1996]*

Piețele de date sunt depozite de date cu caracter mai specific, care conțin date sumarizate într-o anumită măsură, putând astfel să răspundă unor probleme formulate de utilizatorii unui domeniu.

Există mai multe modalități de abordare a constituirii depozitelor de date, respectiv a datelor. O variantă ar fi aceea că odată cu construirea depozitului se construiesc și piețele de date, depozitul furnizând informații sumarizate specializate pe un domeniu. O altă abordare se referă la faptul că piețele de date preexistă, dorindu-se integrarea lor într-un depozit de date. O altă variantă se referă la constituirea premiselor creării unui depozit, în paralel creându-se și câteva piețe de date de importanță vitală în desfășurarea activității organizației. [Connolly et al. 2001] Se va alege varianta optimă în funcție de condițiile de organizare a companiei, de necesitățile imediate și nu în ultimul timp, în funcție de cheltuielile și efortul ocazionat în raport cu așteptările. Putem să avem depozite care să coste de la câteva zeci de mii de dolari la câteva milioane. Dimensiunile lor pot atinge valori inimaginabile, de ordinul Petaocteților. Evident, la astfel de dimensiuni nu se poate pune problema stocării bazei de date pe un singur server, fiind imperios necesară utilizarea unui sistem distribuit, mai mult sau mai puțin omogen.

Atât pentru realizarea unui depozit de date, cât și pentru exploatarea acestuia, un sistem relațional trebuie să îndeplinească o serie de cerințe referitoare la: performanțele de încărcare, prelucrarea încărcării, administrarea calității datelor, performanțele interogărilor, scalabilitatea la nivel de teraocteți sau chiar superioare, scalabilitatea la nivelul masei de utilizatori, distribuirea depozitului în rețea, administrarea depozitului de date, analiza dimensională integrată, funcționalitatea interogărilor avansate. [Connolly et al. 2001] Datorită dimensiunilor mari pe care le are un astfel de depozit, și sistemele SGBDOO ar putea furniza rezultate bune.

Beneficiile înmagazinării datelor sunt: o potențială rată ridicată de întoarcere a investițiilor, avantajul competitiv pe care-l pot oferi și sporirea productivității organelor decizionale. Dezavantajele lor ar fi: subestimarea resurselor necesare pentru încărcarea datelor, probleme ascunse ale sistemului sursa, eșuarea capturării unor date, cerințe sporite din partea utilizatorului final, uniformizarea datelor, cerințe mai mult decât modeste privind resursele, proprietatea asupra datelor, gradul înalt de întreținere, inițierea unor proiecte de lungă durată și complexitatea integrării. Chiar dacă problemele ce pot fi întâlnite sunt mai multe decât avantajele, o valorificare corespunzătoare a avantajelor oferă de cele mai multe ori satisfacțiile dorite.

Înmagazinarea datelor, susținută de tehnologii precum bazele de date distribuite, baze de date federative, tehnologii Web etc. constituie un suport solid pentru instrumentele de raportare și interogare, dezvoltare de aplicații, sisteme EIS⁷, OLAP, data mining și altele. Nu întotdeauna utilizarea depozitelor de date constituie un scop nobil sau etic. Descrierea unei aplicabilități „diabolice” – precum proiectul Echelon al organizației UKUSA – a acestor coloși de date apare în referința [Sitar & Sitar 2002-3]. Vă asigurăm că nu este singura realizare întunecată din domeniul depozitelor de date.

⁵ [W3C ODC]

⁶ Data Mart

⁷ Executive Information System

ELEMENTE PRIVIND PRELUCRAREA ANALITICĂ ȘI DE EXTRAGERE A CUNOȘTINȚELOR

Procesul de înmagazinare al datelor nu trebuie privit ca o obligație sau modă. În aceste condiții informația strânsă cu atâta râvnă reprezintă o povară – mai mult chiar, una scumpă – pentru organizație. Așa cum se sublinia și în lucrarea „*Abordarea informației prin prisma procesului decizional al unei organizații*” [Sitar & Popa 2002], informația reprezintă cel mai de preț lucru pentru procesul decizional. În anii '80, în SUA s-a estimat că aproximativ 70% din timpul de lucru al angajaților este destinat procurării, utilizării sau difuzării de informație. Constituirea depozitelor de date trebuie să se facă pentru progresul companiei într-o manieră care să permită câștigarea unui avantaj competitiv.

Acumularea unui volum din ce în ce mai mare de date și stocarea pe un singur sistem de calcul este o ipoteză de neconceput. În primul rând, chiar și cu explozia tehnologică din ultima vreme, este greu de înțeles că ar putea exista o unitate de stocare care să înmagazineze experiența de zeci de ani a unei companii de dimensiuni mai mari, cu activitate frecventă și complexă. Un alt motiv ar fi că riscul memorării datelor într-un singur loc ar fi prea mare, o defecțiune minoră putând să pună capăt unei activități care datează de decenii. Mai mult, fără existența unor copii de siguranță orice atac informatic, act de sabotaj sau calamitate naturală ar fi fatal. Așadar, chiar dacă volumul de date este imens, valoarea datelor necesită replicare. Abordarea cu o bază de date locală sau centrală ar genera regăsiri mari consumatoare de timp datorate dimensiunii mari a bazei de date și a operațiilor de arhivare/dezarhivare, și totodată ar limita numărul utilizatorilor care să efectueze operații concurente.

Toate aspectele menționate în paragrafele anterioare nu pot decât să conducă la ideea că abordarea depozitelor de date, a procesării analitice și a extragerii cunoștințelor, nu pot să fie îndeplinite fără sprijinul sistemelor de baze distribuite și de baze federative. Înclinăm să credem că bazele de date federative sunt mai potrivite acestui scop, pentru că este puțin probabil ca activitatea de înmagazinare a cunoștințelor să se execute asupra unui singur sistem distribuit care să existe de zeci de ani, fără a suferi modificări majore. Chiar dacă compania a pornit de la bun început cu un sistem distribuit, acesta trebuia schimbat deoarece un sistem informatic nu are ciclul de viață infinit, ci doar de câțiva ani. Așadar avem nevoie de sisteme federative care să acapareze datele furnizate de la sistemele care și-au făcut veacul în instituția respectivă. Nu are rost să discutăm de necesitatea utilizării chiar pentru activitatea curentă a unei baze de date federative. Bazele de date federative sunt recomandate a se folosi și atunci când au loc absorbții, fuziuni și diverse colaborări cu caracter mai mult sau mai puțin temporar cu alte instituții, adică cu sistemele lor informatice sau federative. Rezultatul va fi tot un sistem federativ deoarece organizarea datelor diferă în mod sigur, pentru că ambele instituții au autonomie și pentru că fiecare partajează doar anumite informații, nu toată baza de date. Sistemele distribuite sprijină aceste tehnologii doar pe perioade relativ scurte de timp, sau dacă operațiunile de înmagazinare și cele ulterioare acestuia se desfășoară de puțin timp, adică în perioada de „domnie” a aceluiași sistem.

Instrumentele OLAP

După cum am văzut, bazele de date ale organizațiilor sunt într-o continuă creștere. Pentru a face față nu numai operațiunilor de constituire a acestor volume de date, ci și de exploatare viitoare trebuie să fim preocupați atât de instrumentele hardware, cât și de cele software. La volume imense de date de complexitate din ce în ce mai mare, anumite produse – evident și sistemele relaționale – dau bătăi de cap atunci când se doresc răspunsuri prompte.

Aplicațiile din domeniul afacerilor, precum analiza pieței sau prognoza financiară, se bazează pe scheme de baze de date orientate spre interogări. Acestea trebuie să confere o imagine multidimensională, ușor de interpretat și cât mai natural reprezentată. Aplicațiile de acest gen au ca scop regăsirea unui număr mare de înregistrări dintr-un imens bagaj de date și posibilitatea de sumarizare din mers a acestora. Suportul pentru acest gen de operații este furnizat de către instrumentele OLAP. Termenul „OLAP” a fost inventat de E.F. Codd în 1993.

Definiție: Prelucrarea analitică on-line reprezintă sinteza, analiza și consolidarea dinamică a unor volume vaste de date multidimensionale. [Connolly et al. 2001]

OLAP este o tehnologie ce suportă facilități analitice asupra uneia sau mai multor surse de date. În general, sistemele OLAP implică analiza datelor ce-și au originea în baze de date tradiționale, dar au fost ulterior transformate în structuri multidimensionale pentru o vizualizare și o analiză mai facilă. De aceea, majoritatea instrumentelor OLAP au ca platformă sistemele de gestiune a bazelor de date multidimensionale (MDDDBMS⁸), interfața cu utilizatorul fiind personalizată

Acestea sunt cele mai intuitive, mai fiabile, dar și cele mai costisitoare material instrumente de prelucrare analitică. Sunt cunoscute în literatura de specialitate sub acronimul MOLAP⁹. Ele sunt destinate în rezolvarea unor interogări care furnizează răspunsuri multiple, facilitând utilizatorului capacitatea de analiză, sinteză și comparație. Atunci când vrem să aflăm care sunt mediile generale ale studenților facultății pe ani de studiu, semestriale și anuale, pe specialități în vederea comparării rezultatelor cu situațiile din anii precedenți, cel mai bine e să apelăm la baze de date multidimensionale. Ulterior putem să facem comparații cu rezultatele obținute de facultățile de profil din țară, de stat sau particulare, în cadrul cărora întâlnim secții similare. Sau, putem să agregăm rezultatele în vederea creării premiselor de comparabilitate pe o anumită perioadă a rezultatelor cu celelalte facultăți ale universității. Dacă interogarea trebuie să returneze doar rezultate simple: „Care este media generală a studenților de la secția de Informatică economică, anul 4 de studiu pe semestrul întâi al anului curent?”, atunci nu se impune utilizarea bazelor de date multidimensionale. Pentru aceste tipuri de interogări se pot folosi baze de date relaționale. Totuși, bazele de date relaționale nu fac față cu succes unor volume de date mari. După [Connolly et al. 2001], un sistem relațional obișnuit poate efectua prelucrarea a câtorva sute de înregistrări pe secundă, în timp ce unul multidimensional tipic efectuează mai mult de 10.0000 de grupări pe secundă. Astfel, la dimensiuni mari ale bazei de date, pentru sistemele relaționale componenta „On-Line” a instrumentelor OLAP își va pierde semnificația.

În cadrul sistemelor multidimensionale, reprezentarea intuitivă a datelor se face prin intermediul cuburilor de date. Atunci când numărul de dimensiuni devine mai mare de 3, avem de-a face cu cuburi n-dimensionale, sau simplu: hipercuburi. Reprezentarea intuitivă a acestora va fi însă prea dificil de realizat. Cuburile de date sunt ușor de extins pentru noi dimensiuni. Navigarea prin intermediul cuburilor este facilă. Cuburile de date permit nivele ierarhice de agregare, astfel că un cub de date poate reprezenta doar o celulă din cadrul unui cub mai general. Așadar un cub presupune existența unor cuburi de granularități diferite. În cazul utilizării sistemelor de baze de date tradiționale, aceleași date pot fi reprezentate mai puțin intuitiv prin relații. Oricum, la un exemplu atât de simplu și cu date atât de puține nici n-ar avea rost să apelăm la baze de date multidimensionale.

Serverele multidimensionale OLAP acceptă următoarele operații analitice:

- **Consolidarea** – se referă la gruparea datelor și oferă un anumit grad de sintetizare a informațiilor;
- **Parcurgerea în jos** – este operația inversă consolidării și sporește gradul de detaliu al informațiilor analizate;
- **Tranșarea și tăierea** – este cunoscută și sub denumirea de pivotare. Aceasta presupune analiza datelor dintr-un anumit punct de vedere, adică nu după toate criteriile care le are la bază construcția cubului. Pur și simplu se studiază secțiuni de granularitate diferită din cadrul cubului. [Connolly et al. 2001]

În 1993 Ted Codd a formulat 12 reguli pentru identificarea instrumentelor OLAP. Acestea sunt: vederea conceptuală multidimensională, transparența, accesibilitatea, performanțe coerente la

⁸ Multi-Dimensional DataBase Management System

⁹ Multidimensional OLAP

efectuarea rapoartelor, arhitectura client-server, dimensionalitate generică, tratare dinamică a matricelor rare, suportul multiutilizator, operații inter-dimensionale nerestricționate, manipularea intuitivă a datelor, raportare flexibilă și dimensiuni și nivele de grupare nelimitate. [Codd et al. 1993]

Instrumentele ROLAP¹⁰ au cunoscut cea mai puternică creștere dintre toate instrumentele OLAP. Pentru a se conforma regulilor instrumentelor OLAP, tehnologia ROLAP solicită definirea unui strat de metadate care îi conferă o imagine multidimensională dinamică.

Instrumentele mediului de interogare gestionat oferă capacități de analiză relativ limitate, fie că este vorba de servere ROLAP, fie MOLAP. Prin intermediul lui serverele furnizează datele solicitate de utilizator sub forma unor cuburi de date, supuse apoi local analizei. Implementarea lui se face relativ simplu și cu costuri mici.

Dintre instrumentele MQE¹¹ cunoscute, amintim: PowerPlay, Pablo, Mercury Project, CrossTarget, Media etc. [Connolly et al. 2001]

Dintre domeniile înrudite cu instrumentele OLAP, sau pe care acestea se bazează, menționăm:

- Data mining, extragerea datelor sau a cunoștințelor, mineritul datelor, care încearcă să descopere tendințe și șabloane în cadrul datelor;
- Înmagazinarea datelor, despre care s-a vorbit, care-și propune să integreze toate sursele de date dintr-o organizație într-o singură – din punct de vedere logic – bază de date atotcuprinzătoare;
- Instrumente de transformarea datelor, care furnizează mijloacele de curățare și manipulare a datelor în timpul transferării lor dintr-o locație într-alta. [W3C ODC]

DATA MINING

Am construit depozite de date. Se pune întrebarea firească: „Acum ce facem cu ele?”. Neutilizarea potențialului oferit de către bazele de date de dimensiuni mari reprezintă o cale sigură spre eșecul afacerii. Însă, utilizarea acelor informații pentru ceva mai mult decât urmărirea istoricului afacerii, începe să aibă sens într-un mediu de afaceri concurențial rațional. Am văzut că instrumentele OLAP ne oferă informații pe care interogările simple ale bazelor de date nu le relevă. Totuși noi țintim mult mai sus decât a crea analize și comparații multicriteriale.

Posibilitatea de a oferi consumatorilor acele tehnologii care le vor permite să interogheze eficient bazele de date și să extragă efectiv surse de informații va deveni un factor determinant în alegerea tehnologiei noastre în raport cu cea a concurenței. Mineritul datelor reprezintă un imens potențial pentru tehnologia bazelor de date, precum și un instrument profitabil în lumea modernă a afacerilor.

Definiție: *Extragerea datelor reprezintă descoperirea automată a unor tipare netriviiale, anterior necunoscute și potențial folositoare, bine înrădăcinate în bazele de date.*¹²

Data mining, extragerea de cunoștințe sau „mineritul” datelor, constă într-o analiză automată a datelor electronice structurate, precum în cadrul unui depozit de date, cu intenția de a descoperi tipare anterior necunoscute și relații dintre date [W3C ODC] pentru a putea fi folosite în adoptarea unor decizii importante necesare eficientizării activității economice.

Extragerea datelor diferă față de instrumentele OLAP și alte forme de analiză a datelor conduse prin interogări, prin aceea că șabloanele sau tiparele comportamentale sunt determinate de către sistem prin utilizarea unor algoritmi statistici, astfel încât să se dezvăluie relații pentru care utilizatorul nu este capabil să dezvolte interogări. Evident, factorul uman trebuie să efectueze analiza tiparelor statistice pentru a determina motivele care au scos la iveală relațiile descoperite.

Cu sprijinul larg al facilităților pe care computerele de azi pot să le ofere, al instrumentelor

¹⁰ Relational OLAP

¹¹ Managed Query Environment

¹² [W3C Koundourakis]

de colectare automată a datelor în cazul procesării tranzacțiilor derivate din activitatea economică, cantități impresionante de date sunt colectate și stocate în baze de date de dimensiuni mari. Atât este de ajuns pentru ca tehnicile de data mining să-și poată intra în rol. Clasificarea și descoperirea regulilor de asociere sunt două componente ale extragerii cunoștințelor ce pot sprijini marketingul de exemplu, procesul de luare a deciziilor și conducerea afacerilor. [W3C Meadowcroft]

Extragerea datelor poate furniza avantaje importante organizațiilor care au deja propriile lor depozite de date. Importanța „explorării minelor” pentru extragerea de cunoștințe este subliniată și de către [Oprea 2001]. Chiar dacă mineritul este un domeniu relativ nou aplicat tehnologiei bazelor de date, procesul de data mining este deja aplicat cu succes în diferite domenii, printre care: **comerțul cu amănuntul/marketing, bancar, asigurări, medicină etc.**

Connolly prezintă principalele operații de extragere a datelor. Acestea sunt:

- **Modelarea predictivă.** Modelul acestei operații este experiența de învățare a omului. Omul observă aspectele principale ale unui fenomen, lucru, proces, pentru a putea să efectueze generalizări și integrări ale experiențelor viitoare.

- **Segmentarea bazei de date.** Această tehnică urmărește fragmentarea bazei de date în segmente (clustere) de înregistrări similare. În interiorul unui segment avem un grad mare de omogenitate, caracter care dispare la exterior, adică între segmente diferite. Acest mecanism este caracteristic învățării nesupervizate. Segmentarea bazei de date este cea mai puțin precisă metodă dintre toate cele 4, ea nesesișând diferența între caracteristici redundante și irelevante. Această operațiune este asociată tehnicilor de comasare demografică sau neurală.

- **Analiza legăturilor.** Aceasta urmărește stabilirea unor legături între înregistrări, respectiv între seturi de înregistrări. Așadar, această metodă urmărește descoperirea afinităților dintre înregistrări. În ceea ce privește analiza studenților, s-ar putea constata de pildă că studenții care au luat note mari la materia Bazele informaticii, sunt predispuși să ia note mari și la Baze de date și programarea calculatoarelor. O importanță deosebită o reprezintă descoperirea tiparelor secvențiale și presupune aranjarea înregistrărilor în baza de date în funcție de evoluția în ordine cronologică a evenimentului respectiv. Astfel se pot previziona și descifra comportamentul indivizilor pe termen lung.

- **Detectarea deviațiilor.** Este o operație relativ nouă, utilă în depistarea deviațiilor față de un fenomen deja cunoscut. Metoda poate să releve uneori descoperiri importante. Ea detectează abaterea datelor de la tipare considerate cunoscute. [Connolly et al. 2001] Un avantaj semnificativ al acestei metode este reprezentarea vizuală deosebit de sugestivă.

Printre instrumentele de extragere a datelor se numără: Clementine, DataCrusher, Intelligent Miner, MineSet, Data Mining Suite, SAS System, Thought etc.

Modelele de baze de date relaționale și multidimensionale deservesc multor scopuri. Astfel, în timp ce primele se pretează la operații de citire-scriere, tranzacții, activități cu volum ridicat, celelalte se potrivesc pentru activități de volum scăzut, interogări complexe și care se întind peste cantități enorme de date ce pot fi procesate simultan. În timp ce bazele de date relaționale se bazează ca limbaj de interogare pe SQL, sistemele MOLAP nu au încă un limbaj stabilit de comun acord. Unul dintre candidații cu succes mai mare la public ar putea fi limbajul MDX, dezvoltat de Microsoft. [W3C ODC]

Limbajul MDX este un limbaj declarativ care își propune să extindă procesul de interogare a bazelor bidimensionale, care utilizează operatori relaționali, către procesări de baze de date n-dimensionale, prin utilizarea operatorilor hipercubici. [Zaharie & Albescu 2002]

Mineritul datelor reprezintă descoperirea automată a inteligenței din date brute stocate pe diferite sisteme de calcul. Acest instrument este folosit pentru detectarea comportamentului fraudulos în utilizarea cărților de credit sau a cartelelor telefonice, solicitărilor de despăgubire suspecte, determinarea tiparelor de clienți, segmentarea categoriilor de clienți etc. Ținând cont de realitatea existenței datelor murdare în cadrul bazei de date, instrumentele de data mining trebuie să fie foarte precaute. Toleranța la erori poate constitui un criteriu important în selectarea algoritmilor

utilizați la extragerea datelor. Până nu demult, algoritmi de data mining se mulțumeau să descopere tipare și tendințe anterior necunoscute, dar de masă. Noua generație de algoritmi este preocupată și de descoperirea unor comportamente și tendințe excepționale, mai ales în condițiile actuale când se urmăresc riscurile de securitate, suspectarea posibililor teroriști, transferuri suspecte de fonduri etc. Așadar, noii algoritmi trebuie să fie înzestrați cu capacități de rețele neuronale, astfel încât să poată fi antrenați și în condițiile anterior amintite: date murdare, respectiv comportamente excepționale. Fără a se ține cont și de astfel de considerente, în ziua de azi, astfel de instrumente nu pot fi considerate a fi cu un grad înalt de fidelitate. **[Kim 2002]**

Ambele tehnologii se aplică asupra unor depozite sau piețe de date. Cu cât dimensiunea bazei de date relevă mai multe tipuri de evenimente, „experiența” (volumul de date și perioada de înmagazinare) este mai bogată, iar datele sunt cât mai curate, cu atât gradul de încredere al rezultatelor obținute este mai mare. Acuratețea acestuia permite extrapolări de calitate. Oricât de evaluate ar fi mijloacele moderne de stocare sau arhivare a datelor, nu putem să concepem existența domeniilor de prelucrare analitică on-line și de descoperire a cunoștințelor, în absența unui mediu distribuit cu grad de heterogenitate variabil și cu sferă de întindere cât mai mare.

UTILIZAREA BAZELOR DE DATE PE INTERNET

După cum ziceam ceva mai devreme, Internetul este cel mai mare sistem de baze de date federative. De fapt, Internetul reprezintă o rețea mondială de calculatoare interconectate, în timp ce prin termenul Web desemnăm conținutul, adică însăși baza de date federativă. Interesantă asocierea, atâta timp cât rețeaua Internet a apărut și s-a dezvoltat exploziv fără să aibă legături prea mari cu ceea ce reprezintă bazele de date sau cu sistemele lor de gestiune.

Internetul își are rădăcinile la sfârșitul anilor 1960 și începutul anilor 1970, odată cu construirea rețelei militare ARPANET, care avea ca scop menținerea conexiunii la rețea chiar și în cazul unor dezastre, precum un bombardament atomic.

Evoluția Internetului, popularitatea, dezvoltarea exponențială a siturilor conectate la această rețea, numărul tot mai mare de atacuri informatice, invazia de virusuri, a făcut ca unele organizații să simtă nevoia protejării informațiilor. Totuși, tentația de a păstra lucrurile bune pe care Internetul le poate oferi – cum ar fi independența de platformă, tehnologiile dezvoltate, precum și contactul cu sute de milioane de potențiali clienți – era prea mare pentru a renunța la tot ce înseamnă Internet. Astfel au apărut formele de Internet „semiizolat”, intranetul, prin care firmele își partajează informațiile doar pentru uz intern, protejându-se de mediul exterior, comunicațiile realizându-se tot prin mecanismele furnizate de tehnologiile Web. Membrii unui **intranet** împiedică doar accesul celor din exterior, fără ca ei să fie limitați în a accesa informațiile din exterior. O altă formă de protejare a intimității se referă la crearea de **extranet**, adică un fel de internet deschis parțial exteriorului doar pe bază de autentificare.

Limbajul HTML este cel mai utilizat instrument în descrierea paginilor Web, fiind folosit și de editoarele de pagini de Web (precum Front Page, Page Maker, Netscape Composer etc.). Chiar dacă la început doar simpla utilizare a limbajului HTML era suficientă pentru construirea siturilor, necesitatea unor pagini din ce în ce mai complexe a solicitat sprijinul unei alte tehnologii, pentru că:

- Necesitatea introducerii paginilor dinamice. Limbajul HTML poate simula existența unor astfel de pagini, însă pentru fiecare mică modificare a unei pagini se solicită înlocuirea ei cu o nouă pagină HTML, existentă pe server;
- În cazul siturilor complexe numărul de pagini HTML este foarte ridicat, astfel încât problemele de administrare, precum și de dezvoltare de aplicații și întreținerea numelor de pagini, devine o sarcină deosebit de complexă. Mai mult, datorită existenței unei mari redundanțe informaționale, spațiul de memorare ar trebui lărgit. Mai mult, o serie de probleme nu-și pot găsi rezolvarea datorită acestor limitări.

Astfel, a apărut necesitatea integrării tehnologiei Web cu cea a bazelor de date. Acum, cu un

singur șablon de pagină HTML și o bază de date relativ modestă pot fi afișate sute sau mii de pagini diferite. În abordarea tradițională am fi avut sute sau mii de fișiere HTML, în locul celor două.

În vederea colaborării celor două tehnologii, mai ales în condițiile unui mediu economic, au fost desemnate mai multe arhitecturi. Indiferent de arhitectură, există 4 componente principale care trebuie să fie luate în calcul: baza de date, logica tranzacțiilor, logica aplicației și interfața cu utilizatorul. În cadrul sistemelor centralizate acestea se aflau într-un singur loc. Totuși, în condițiile descentralizării treptate, o arhitectură cu un singur etaj era de neconceput. Așadar, s-a construit **arhitectura client-server**, considerată și ca arhitectura tradițională. Clientul are ca responsabilitate prezentarea datelor către server, iar acesta din urmă furnizează serviciile de date către client.

La mijlocul anilor '90 această arhitectură a fost reconsiderată, deoarece clientul era considerat unul „tare” și necesita existența unor resurse se calcul importante pe calculatorul local. În jurul anului 1995 a apărut **arhitectura pe 3 etaje**, care degreva clientul de problema gestionării logicii afacerii și a prelucrării datelor. Acest lucru presupune interpunerea unui nou etaj între cele două existente. Noul etaj este format dintr-un server de aplicații care va prelua cele două sarcini. Clientul devine unul „subțire”, nemai necesitând resurse importante. În această nouă arhitectură se creează premisele accesării aplicației și a bazei de date concomitent de către mai mulți utilizatori decât în cazul precedent.

Arhitectura cu trei etaje se potrivește cel mai bine cu arhitectura Web conținând un browser Web în ipostaza de client subțire și un server Web în calitate de server de aplicații.

Există și **arhitecturi multi-etajate** provenită din această arhitectură cu 3 nivele. Etajele intermediare vor partaja sarcinile pe care le are momentan etajul intermediar. [Connolly et al. 2001]

Limbaajul XML¹³

Există o curiozitate legată de limbaajul HTML, și anume că, lucrul pentru care este apreciat cel mai mult este și lucrul care i se reproșează cel mai mult. Simplitatea este ceea ce a făcut ca să devină cel mai popular limbaaj utilizat în rețeaua Internet.

Date evidențiază însă două dintre problemele reale pe care acest limbaaj le are, și anume:

- Limbaajul HTML nu reușește să facă distincție dintre metadatele structurale, semantice și cele de formatare;
- HTML-ul nu este sensibil la încălcarea unor reguli de semantică. Faptul că un document, considerat de autor „corect” din punct de vedere semantic, apărea așa cum editorul său dorea pe un anumit browser, iar pe altul apărea în mod diferit, vina era dată pe browser, nu pe sintaxa utilizată în document. [Date 2005]

Cea de-a doua afirmație nu vrea să discolpe particularitățile existente între browsere, ci vrea doar să atragă atenția asupra sintaxei prea îngăduitoare a limbaajului HTML.

Ca urmare a deficiențelor menționate, un comitet de revizuire a limbaajului SGML – din care a fost extras și limbaajul HTML – a creat în 1996 limbaajul XML.

Așadar, XML este o submulțime a limbaajului SGML, deci compatibil atât cu acesta cât și cu limbaajul HTML. Scopul declarat de autorii lui este de a permite ca toate documentele XML/SGML să beneficieze pe Web de același tratament de care se bucură momentan documentele HTML, conferind acestora din urmă mai mult dinamism și făcându-le mai atractive.

Limbaajul XML permite utilizatorilor să creeze etichete personalizate. O altă diferență față de HTML este că permite legături cu destinații multiple, nu doar una.

Limbaajul SGML, generic vorbind, presupune separarea logică a conținutului unui document față de structură. Definierea structurii se face într-un fișier DTD¹⁴ care are și extensia „.dtd”.

Modul de integrare a celor două tehnologii nu este universal valabil. Există o serie de criterii care pot să influențeze strategia de implementare. Acestea ar putea fi specificul aplicației,

¹³ EXtensible Markup Language

¹⁴ Document Type Definition

complexitatea și dimensiunea bazei de date, strategiile cunoscute de către implementator, calitatea serviciului furnizat, cheltuielile ocazionate și experiența proiectantului.

Connolly amintește câteva din soluțiile de interfațare a celor două tehnologii:

- CGI;
- Server-side includes;
- HTTP cookies;
- Extensii ale serverelor Web: NSAPI, ISAPI etc.;
- Limbajele Java, JSQL, JRB și conexiunile la baze de date prin JDBC;
- JavaScript, VBScript și alte limbaje script;
- Platforma activă Microsoft;
- Arhitectura de calcul în rețea (NCA) folosită de Oracle etc. **[Connolly et al. 2001]**

Integrarea tehnologiei Web cu cea a bazelor de date are următoarele avantaje: simplitate, independență de platformă, interfață grafică cu utilizatorul, standardizare, suport pentru platforme multiple, transparență la accesul în rețea, desfășurare scalabilă, inovația; în timp ce dezavantajele ar fi: fiabilitate, securitate, cost, scalabilitate, funcționalitate limitată a limbajului HTML, lipsă de stare, lărgime de bandă, performanțe, imaturitate în dezvoltarea instrumentelor. **[Connolly et al. 2001]**

Integrarea tehnologiei bazelor de date cu tehnologiile Web este unul din pașii importanți pe care i-a făcut tehnologia modernă. Paginile sunt acum mai atractive, mai dinamice și mai ușor de administrat. Se reduce foarte mult redundanța, chiar dacă din punctul de vedere al stocării unui număr enorm de fișiere HTML, accesarea unei astfel de pagini se face mai rapid decât prin interogarea unei baze de date.

Apariția limbajului XML este un pas revoluționar în integrarea Web-SGBD. Mai există multe lucruri de îndreptat în legătură cu acesta, însă orientarea e promițătoare. Am subliniat avantajele și dezavantajele acestui limbaj pe baza cunoștințelor teoretice acumulate și a experienței practice din ultimii ani.

Atât avantajele cât și dezavantajele constituie o stare de fapt temporară. E posibil să existe și alte avantaje sau dezavantaje neinventariate. La fel unele dintre acestea pot să dispară. Cazul cel mai fericit ar fi transformarea dezavantajelor în avantaje. Însă nici operațiune inversă nu este scosă din calcul.

Concluzii

Bazele de date distribuite nu reprezintă panaceul universal al modelării statice și dinamice a unei organizații. O serie de alte tehnologii derivate din cea a sistemelor distribuite permit sau completează realizarea imaginii virtuale a companiei, contribuind în același timp și la conferirea unui avantaj competitiv al acesteia. Fără de acesta, trecerea de la figurant la actor principal pe tumultoasa și cruda scenă a mediului de afaceri, nu se întrevide la orizonturi.

BIBLIOGRAFIE:

1. [Bouguettaya 1998] BOUGUETTAYA, A., BENTALLAH, B., ELMAGARMID, A. – "Interconnecting Heterogeneous Information Systems", Kluwer Academic Publishers, 1998
2. [Buch 2002] BUCH, Vineet – "Database Architecture: Federated vs. Clustered", Oracle White Paper, 2002
3. [Codd et al. 1993] CODD, E.F., CODD, S.B., SALLEY, C.T. – "Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate", Arbor Software Corporation, 1993
4. [Connolly et al. 2001] CONNOLLY, T., BEGG, C. și STRACHAN, A. – „Baze de date. Proiectare. Implementare. Gestionare”, Editura Teora, București, 2001, ISBN 973-20-0601-3
5. [Date 2005] DATE, Chris J. – „Baze de date”, Ediția a VIII-a, Editura Plus, București, 2005

6. [Hsiao 1992] HSIAO, David K. – *"Federated Databases and Systems: Part I --- A Tutorial on their Data Sharing"*, The VLDB Journal — The International Journal on Very Large Data Bases archive, Volume 1 , Issue 1, Pag. 127 – 180, Springer-Verlag New York, Inc, Iulie 1992, ISSN:1066-8888
7. [Inmon 1996] INMON, W. H. – *"Building the Data Warehouse"*, John Wiley & Sons, 1996
8. [Kim 2002] KIM, Won – *"On Database Technology for US Homeland Security"*, *Journal of Object Technology*, Vol. 1, Nr. 5, pag. 43-49, Noiembrie-Decembrie 2002
9. [Nukpe 2001] NUKPE, Philip K. – *"Heterogenous Distributed Databases. Issues in Distributed Data Management"*, 2001
10. [Oprea 2001] OPREA, Dumitru – *"Managementul proiectelor: teorie și cazuri"*, Editura Sedcom Libris, Iași, 1999, ISBN 973-8028-82-5
11. [Saltor 1995] SALTOR, Felix – TUTORIAL on *"The BLOOM Model for Database Interoperability"*, 1995
12. [Schmitt & Saake 2005] SCHMITT, Ingo, SAAKE, Gunter – *"A Comprehensive Database Schema Integration Method Based on the Theory of Formal Concepts"*, *Acta Informatica*, Vol.41, Iss. 7-8; pg. 475, Heidelberg, Jun 2005, ISSN: 0015903
13. [Sheth 1991] SHETH, Amit P. – *"Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases"*, Proceedings of the 17th International Conference on Very Large Data Bases, Barcelona, Septembrie 1991
14. [Sheth & Larson 1990] SHETH, Amit P. , LARSON, James A. – *"Federated Database Systems for Managing distributed, Heterogeneous, and Autonomous Databases"*, *ACM Computing Surveys (CSUR)*, Volume 22 , Issue 3, Special issue on heterogeneous databases, pag. 183 – 236, New York, Septembrie 1990, ISSN:0360-0300
15. [Sitar 2005-1] SITAR-TĂUT, Dan-Andrei, *"An Overview on Federated Database Systems"*, The Impact of European Integration on the National Economy, Pag. 494–503, Editura Risoprint, Octombrie 2005, Cluj-Napoca, ISBN: 973-651-007-0
16. [Sitar 2005-2] SITAR-TĂUT, Dan-Andrei, Maria-Liana STANCA, *"Federated Database Integration"*, Collaborative Support Systems in Business and Education, Pag. 376-383, Editura RISOPRINT, Octombrie 2005, Cluj-Napoca, ISBN: 973-651-008-9
17. [Sitar 2005-3] SITAR-TĂUT, Dan-Andrei, *"Baze de date distribuite"*, Editura RISOPRINT, Cluj-Napoca, 2005, ISBN: 973-651-038-0
18. [Sitar & Sitar 2002-3] SITAR-TĂUT, Dan-A., SITAR-TĂUT, Adela-V., – *"Coloșii de date"* , pag. 99-106, Workshop: „Informatica economică și societatea informațională”, 10-11 mai 2002, Editura Mirton, Timișoara, Mai 2002, ISBN 973-585-691-3
19. [Sitar & Popa 2002] SITAR-TĂUT, D.-A., POPA, S.C. – *"Abordarea informației prin prisma procesului decizional al unei organizații"*, Tehnologiile informaționale și schimbările organizaționale, vol. 2, pag. 314-317, Sesiune jubiliară ISIS 2002: „Globalizare și educație economică universitară”, Editura Sedcom LIBRIS, Iași, Octombrie 2002, ISBN 973-8353-33-5
20. [Tuladhar 2005] TULADHAR, Arbind Man et al. – *"Federated Data Model to Improve Accessibility of Distributed Cadastral Databases in Land Administration"*, Pharaohs to Geoinformatics, Cairo, 2005
21. [Winter 2002] WINTER, Richard – *"Playing Every Tune: Federated Databases"*, *DB2 Magazine*, 2002
22. [Zaharie & Albescu 2002] ZAHARIE, Dorin, ALBESCU, Felicia – *"Operatori hipercubici pentru interogarea bazelor de date multidimensionale"* , pag. 23-29, Workshop: „Informatica economică și societatea informațională”, 10-11 mai 2002, Editura Mirton, Timișoara, Mai 2002, ISBN 973-585-691-3
23. [W3C Meadowcroft] http://www.benmeadowcroft.com/reports/Database_Research_Exercise.htm – MEADOWCROFT, Ben, Database research

24. [W3C Koundourakis] <http://timelab.co.umist.ac.uk/publications/theses/1997.html> - Classification and Association Tempo Miner Project, Koundourakis, George, 1997
25. [W3C ODC] <http://www.newhoo.com/Computers/Software/Databases/desc.html> - Open Directory Computers: Consultants: Databases, 1999
26. [W3C IDC #22542] <http://www.idc.com>, Report No. 22542, "*Enterprise Database Management Systems Market Forecast and Analysis, 2000-2004*"
27. [Wikipedia 2005] http://en.wikipedia.org/wiki/Federated_database_system/Federated_database_system - *Wikipedia, the free encyclopedia.htm* – Federated database system, *Junie 2005*
28. [W3C FDMS Group] [http://www-lsi.upc.es/bloom/Federated Database Management System Research Group.htm](http://www-lsi.upc.es/bloom/Federated_Database_Management_System_Research_Group.htm) – Federated Database Management System Research Group, Catalonia Subgroup