

## Tehnologii informatice inteligente de accesare multidimensională a bazelor și depozitelor de date de marketing

Autor: **Gheorghe Orzan**

**Abstract:** În mod tradițional, una dintre cele mai importante și mai frecvente activități ale unui departament de marketing constă în gestionarea unor volume mari de date acumulate de-a lungul timpului, de consultare a variate baze de date cantitative, de grafică sau multimedia în stare offline sau online, orientate spre piață (market-oriented), date care provin din cercetările de marketing, de la alte organizații și care impun operativ consultarea interactivă și de o manieră flexibilă și consistentă prin operații de integrare, sintetizare și agregare, de reorganizare și stocare, de gestionare și procesare în raport de obiectivele de marketing. Acestea se constituie în surse de informații esențiale în procesul de fundamentare a deciziilor, atât la nivelul activităților de marketing, cât și în ansamblul organizației. Tehnicile moderne din domeniul IT&C fac însă posibilă o analiză mult mai rapidă și mai eficientă a acestor date, practic în timp real, în ciuda eventualelor distribuții geografice. Această lucrare își propune prezentare unor situații teoretice și practice în care tehnici de tip OLAP sau data mining eficientizează activitatea profesioniștilor în domeniul marketingului.

**Cuvinte cheie:** asistarea deciziilor, strategii de marketing, data warehouse, sisteme asistarea deciziilor, OLTP, OLAP, data mining.

**A. Depozitele de date** (data warehouse) reprezintă un produs al mediului economic tot mai competitiv, global, dinamic și complex în care organizația își desfășoară activitatea și care solicită informații relevante, actuale, de o manieră consistentă și foarte flexibilă în procesul de fundamentare a deciziilor de marketing iar, pe de altă parte este rezultatul evoluțiilor spectaculoase în ceea ce privește tehnologiile din domeniul IT&C.

Un depozit de date este o sumă de date consistentă din punct de vedere semantic care servește la o implementare fizică a unui model de date ca suport pentru sprijinirea deciziei și stochează informații pe care o organizație le solicită în fundamentarea deciziilor strategice. Un data warehouse (Inmon, 1996) este „o colecție de date orientată pe subiecte, integrată, atemporală, istorică și persistentă, destinată sprijinirii procesului de fundamentare a deciziei manageriale”. În sinteză definiția prezentată mai sus exprimă **caracteristicile** de bază ale depozitelor de date: **orientarea pe subiecte majore** (clienți, furnizori, concurenți, producție, vânzări), prin **integrarea** unor multiple surse eterogene de informații (baze de date relaționale, distribuite, baze de date multimedia și de tip hipermedia, fișiere și înregistrări), sunt **independente de timp** și au **caracter de istoric** atribuit prin dimensiunile spațiale și temporale ale depozitului (prin consultarea valorilor succesive ale variabilelor cercetate pentru a determina evoluția în timp și spațiu și a estima tendințele proceselor și fenomenelor investigate), cât și **persistența datelor** (datele din depozit sunt permanente și nu pot fi modificate, fiind deja istorie a

evenimentului). În mod uzual depozitele de date implică numai două operații: construirea sau încărcarea inițială a datelor (integrarea datelor, filtrarea, curățarea și consolidarea datelor) și accesul la date printr-o colecție de tehnologii inteligente de asistare a procesului de fundamentare a deciziilor și care va permite specialiștilor (manageri, analiști, consultanți, executivi) acces rapid, consistent și facil la informațiile din depozit.

O comparație între bazele de date și depozitele de date pune în valoare rolul depozitelor de date, ca model de organizare și sursă suport de informații pentru fundamentarea deciziei. Atât bazele de date cât și depozitele de date sunt structuri logice de organizare de date, exploatează tehnologii informatice de organizare și gestiune de date diferite și înmagazinează mari cantități de date.

**Diferențele semnificative** între cele două modele de organizare logică de date pot fi structurate astfel:

- Bazele de date din sistemele informatice conțin date curente, detaliate, care sunt actualizate și interogate în aceleași perioade cu culegerea și procesarea lor și fac obiectul sistemelor informatice de prelucrare a tranzacțiilor (TPS). Depozitele de date sunt construite special ca sisteme suport de asistare a deciziei și au ca obiectiv regruparea și consolidarea datelor, agregarea și sintetizarea lor, reorganizarea și stocarea informațiilor provenite din surse diverse și variate de date inclusiv date multimedia, hipermedia și de localizare spațială și fac obiectul sistemelor informatice inteligente – sisteme suport de asistare a deciziei (DSS). Ele se aplică asupra unor volume foarte mari de date, eterogene, constituite ca depozite de istoric/arhivă de date și presupun calcule complexe (analiză de tendință, corelații între evenimente, asociații între anumite fapte, secvențe, tipare de comportament, modele de cunoaștere) ;
- Sistemele de gestiune a bazelor de date sunt adecvate aplicațiilor curente de gestiune și servesc la crearea și întreținerea sistemelor de baze de date tranzacționale OLTP (On-Line Transaction Processing) și au ca obiectiv execuția în timp real a tranzacțiilor zilnice (aprovizionare, vânzări, stocuri, producție, decontări, plăți, activități de marketing și management, contabilitate) cât și a procedurilor de interogare (Query). Sistemele de gestiune a depozitelor de date, pe de altă parte, servesc top managerii, managerii, utilizatorii sau specialiștii în domeniul analizei și fundamentării deciziei, fac obiectul sistemelor informatice OLAP (On-Line Analytical Processing) și oferă tehnologii de agregare a datelor stocate în depozitele de date într-o abordare multidimensională, cu acces rapid la informațiile necesare, într-o manieră consistentă, interactivă și foarte flexibilă;
- Un sistem OLTP este centrat pe client (customer oriented) și este utilizat pentru procesarea tranzacțiilor și interogărilor din bazele de date constituite operațional, baze care pot oferi răspunsuri la întrebări de tipul **Cine?, Ce?, Unde?, Când?** prin analiza datelor de tip numeric sau statistic. Un sistem OLAP este orientat spre piață (market oriented) și este utilizat de manageri, analiști, specialiști prin instrumente FASMI (Fast Analysis Shared Multidimensional Information) de analiză rapidă a informației multidimensionale distribuită în locații multiple și accesibilă în același timp unui număr mare de utilizatori pentru analize complexe prin agregare, sintetizare, consolidare și care permit construirea de scenarii, prin posibilitatea de a răspunde la întrebări de tipul **Ce ar fi dacă?** pentru a descoperi corelații

între evenimente, asociații între anumite fapte, secvențe și tipare de comportament;

- Tehnologiile OLAP utilizează bazele de date multidimensionale, construite de regulă din date de istoric sau date care provin de la diferite organizații, integrând informații din surse eterogene, de tipuri și naturi diferite. Tehnologiile OLTP utilizează baze de date relaționale, care sunt bidimensionale prin definiție, fiind focalizate pe datele curente dintr-o organizație sau departament fără a referi date istorice sau date din alte organizații.

**B. Tehnologia OLAP** (OnLine Analytical Processing) este un instrument puternic, performant și ușor accesibil de agregare a datelor stocate în arhive mari de date, care printr-o abordare multidimensională, asigură acces rapid la informații, iar printr-o analiza multidimensională oferă răspunsuri concrete, consistente, corecte și în timp real la întrebările analiștilor de marketing privind fenomene, procese și activități de marketing. Tehnica OLAP permite o modelare dimensională, prin utilizarea cubului OLAP, o structură multidimensională, un hiper cub, care modelează aspectul complex al unui proces sau fenomen, atât prin reprezentările sale cantitative, măsurabile ale activității, în strânsă legătură cu aspectele calitative sau de context. Context este definit prin dimensiunile/parametrii săi: când s-a desfășurat, unde s-a desfășurat, cine a inițiat, cine a executat-o, cine a beneficiat de ea,, cu ce resurse, cu cine a fost efectuată, de la cine și către cine s-a desfășurat. Trăsătura esențială a acestor întrebări este caracterul lor multidimensional. Există totuși câteva tipuri uzuale de întrebări, care pot arunca o lumină asupra complexității instrumentelor care trebuie să furnizeze răspunsuri:

- **rapoarte multidimensionale**, ca de exemplu: vânzările totale ale firmei din produsul X, și respectiv ponderea vânzărilor realizate prin magazinele din orașul Y în perioada .... ?
- **comparații**, ca de exemplu: care este media abaterii procentuale de la planul de vânzări în primul semestru al acestui an comparativ cu vânzările din primul semestru al anului trecut ?
- **clasificări și profiluri statistice**, ca de exemplu: care este volumul vânzărilor și media adaosului pentru primii 20 % dintre distribuitori și care este contribuția acestora la totalul vânzărilor pe trimestrul trecut ?
- **agregări libere**, ca de exemplu: care sunt veniturile realizate în ultimele două trimestre de filialele județene din Muntenia ?
- **evaluări What – If**, ca de exemplu: în ce măsură ar influența profitul total al firmei o creștere cu 15 % a vânzărilor în județele din Moldova ?

Pentru oricine care a formulat interogări în bazele de date multidimensionale (engl. Query), este evident că exprimarea unor asemenea cereri depășește posibilitățile oricărui instrument de acces și raportare.

Printre calitățile pe care trebuie să le îndeplinească un bun instrument OLAP se numără:

- să poată susține analize sofisticate;
- să poată fi utilizate eficient de diverse categorii de utilizatori;
- să fie scalabile la volume oricât de mari de date;
- să permită accesul concurent al unui mare număr de utilizatori;
- să fie ușor de întreținut și de configurat;

- să fie bazate pe o arhitectură deschisă deoarece evoluția tehnologiei informației poate aduce schimbări radicale în structura sistemului informatic care, însă, nu trebuie să afecteze instrumentația utilizată pentru analiză.

O bază de date multidimensională utilizată de tehnologiile OLAP este formată din două structuri:

- Structura **datelor** prin care se descriu structurile proprii ale bazelor de date referite, cu câmpurile, atributele și caracteristicile asociate;
- Structura **metadatelor** definită prin dimensiunile, structurile ierarhice ale dimensiunilor, membrii și multipli acestor structuri.

Dimensiunile OLAP sunt identificate în bazele de date tranzacționale ca fiind câmpurile ce conțin caracteristicile unei tranzacții, datele de identificare a tranzacției respective și care în cele mai multe cazuri sunt cheile externe utilizate în bazele respective. Dimensiunile reprezintă parametri în care s-a desfășurat o activitate, nu se intersectează cu alte dimensiuni și sunt de natură complet diferită (timp, spațiu, produse, furnizori, clienți, resurse financiare, resurse materiale, resurse umane). Mulțimea valorilor posibile pe care le poate lua dimensiunea respectivă poartă numele de membrii dimensiunii. Prin regruparea valorilor se generează multipli de membrii, grupe de valori ale dimensiunii respective cu o caracteristică comună. De exemplu se poate defini localitatea ca multiplu de membrii formați din toți clienții organizației din localitatea respectivă. Dimensiunile și multipli lor formează structuri arborescente cunoscute în OLAP sub numele de ierarhii. Modelarea dimensională oferă un model conceptual, **cubul OLAP**, care permite agregarea datelor într-o structură ierarhică, simplă și foarte flexibilă, fără a modifica structura de bază a colecțiilor încorporate în depozit, cu păstrarea legăturilor cu sursele inițiale de date și cu posibilitatea de descompunere a datelor centralizate pe niveluri ierarhice inferioare până la setul de tranzacții inițiale.

Motorul OLAP folosește ca mod de vizualizare a datelor în depozitul de date, o reprezentare într-un spațiu n dimensional, prin care axele dau dimensiunea logică a cubului.

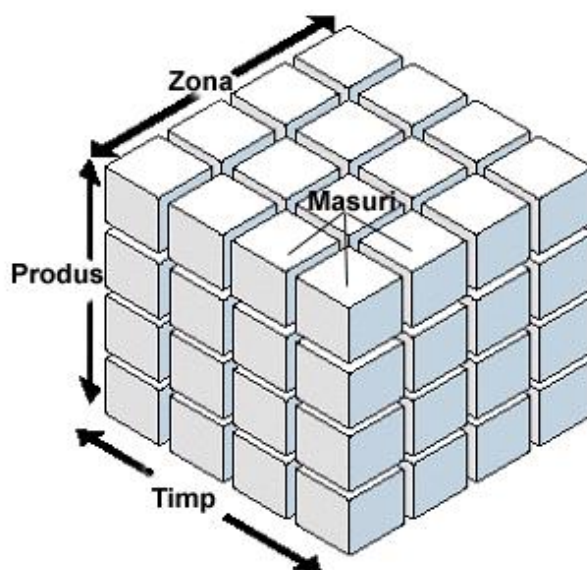


Figura 1: Cubul de date

Dacă luăm ca exemplu dimensiunile: Produsul, timpul regiunea (zona), cubul măsurilor cantitative sau valorice ar avea reprezentarea din figura 1.

Asupra cuburilor n dimensionale motorul analitic execută o serie de operații care se integrează în aplicații OLAP manevrate de utilizatorul final. Pentru execuția acestor operații este necesară „navigarea liberă” de-a lungul nivelelor ierarhice definite pentru depozitul de date.

Principalele operații OLAP care se pot efectua sunt:

- **Specificarea criteriilor de selecție** este primul pas în orice analiză. Utilizatorul trebuie să poată exprima cu ușurință criterii simple, bazate pe valori ale atributelor și/sau pe valori ale metricilor. Aceste criterii simple trebuie să poată fi apoi combinate prin operatori logici și trebuie să poată fi salvate în bibliotecă pentru eventuale reutilizări.
- **Rotațiile** sunt operații care permit utilizatorilor să găsească perspectiva care-l interesează specificând dimensiunile și direcțiile de rotație sau indicând un pivot.
- **Schimbarea nivelului de agregare** permite găsirea nivelului de agregare optim pentru analiză. Se poate adânci analiza spre nivele de detaliu prin operații de schimbare a sensului de navigării de-a lungul nivelelor unei ierarhii; prin drill down se poate naviga pe nivelele cu un grad de detaliu mai mare, iar prin roll up se va naviga pe nivelele ierarhice superioare ca nivel de agregare.
- **Specificarea modului de prezentare** trebuie să permită analistului să găsească modalitățile optime de valorificare vizuală a datelor extrase. În afară de posibilitățile grafice tipice pentru prezentare, este important ca utilizatorul să poată vizualiza date multidimensionale într-o manieră tabelară. În acest sens se pot utiliza tabele complexe, care să poată grupa coloane și linii exprimând dimensiuni diferite (de pildă timpul și dispunerea în spațiu) și nivele de agregare diferite. Perspectiva multidimensională asupra datelor se face prin operații care permit obținerea de vederi (view) bidimensionale dintr-un cub n dimensional, care se mai numesc **slice-uri**, felii de informații care servesc unor anumite scopuri și decizii manageriale.

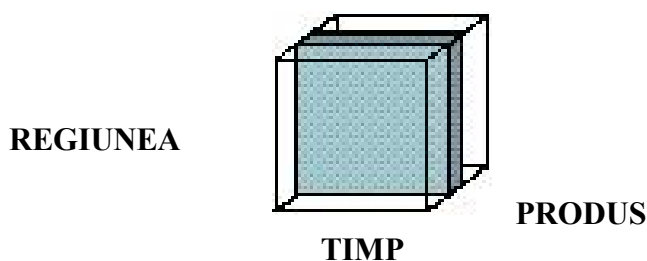


Figura 2: Slice pentru viziunea managerului de produs

De exemplu viziunea managerului de produs interesează doar vânzările anumitor produse, în toată regiunea și în toată perioada de timp considerată

Alte tipuri de slice-uri pot felia cubul pe toate cele trei coordonate carteziene așa cum se poate vedea în figura de mai jos.

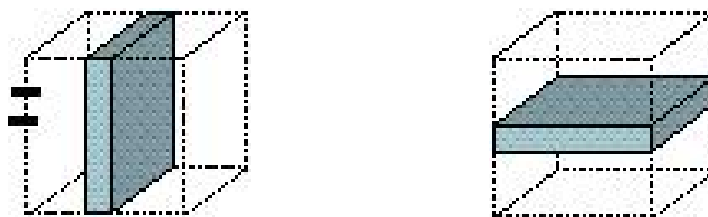


Figura 3: Alte tipuri de slice-uri

Pentru prezentarea convenabilă a rezultatelor se fac racordări ale vederilor bidimensionale cu baza de date multidimensională, prin mapări de date, coroborate cu utilizări de grafice și tabele complexe

Fiind o tehnologie relativ nouă, modelul de arhitectură OLAP (figura 4) care s-a impus pentru sistemele orientate spre analiză multidimensională este unul de tip client/server în trei straturi.

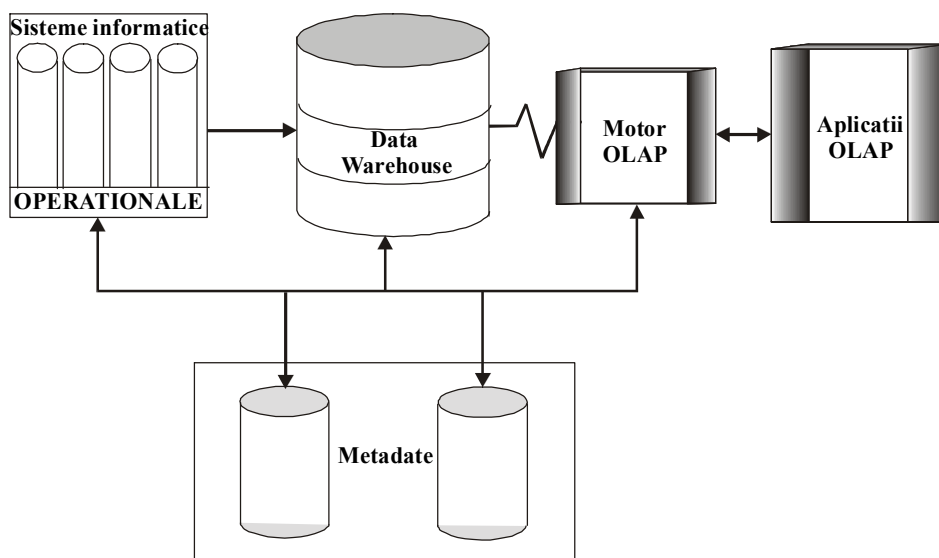


Figura 4: Arhitectura unui sistem OLAP

a) **Data Warehouse** formează nivelul cel mai de jos, responsabil cu stocarea și regăsirea datelor. De regulă aplicațiile tranzacționale utilizează sisteme relaționale dar pentru depozitele de date se folosesc sisteme multidimensionale. Dat fiind volumul mare de date, este recomandabil ca SGBD-urile folosite să ofere suport pentru prelucrări paralele și distribuite, să dispună de mecanisme performante de indexare și de optimizare, să ofere un înalt nivel de siguranță;

b) **Motorul analitic OLAP** (OLAP engine) are sarcina de a prelua cererile exprimate de utilizatori și, pe baza consultării metadatelor, să genereze cererile necesare pentru obținerea datelor ce vor fi redirectate către clienți. În plus, datelor obținute li se vor aplica la acest nivel o serie de prelucrări de generarea de interogări, manipulări de date și de sintetizarea rezultatelor;

c) **Metadatele** – în care sunt stocate dimensiunile, membrii și multipli lor, precum și structurile ierarhice ale dimensiunilor, informații care apar pe axele cuburilor și sunt prezentate utilizatorului ca nume de rînduri sau de coloane (pivot table);

d) **Aplicațiile OLAP** sunt reprezentate de instrumente mînuite de utilizatorul final. Există aplicații generale care răspund suficient de bine nevoilor unei categorii largi de utilizatori, există instrumente specializate pe domenii (cum ar fi de pildă analize regresionale, analize ANOVA, analize SWOT, analiza financiară) și, în fine, există posibilitatea de a dezvolta instrumente foarte specifice.

Din punct de vedere al utilizatorului final, aplicația pe care o folosește trebuie să-i asigure două funcționalități importante: **navigarea liberă** prin depozitul de date în căutarea informațiilor relevante și **posibilități diverse de prezentare a datelor**. Aceste funcționalități sunt strîns legate între ele și este greu de spus care operație este de navigare și care este de prezentare.

Cerințele de administrare și dezvoltare pentru OLAP, deși similare cu cele pentru instrumentele de interogare și raportare, sunt în general mult mai complexe. Punerea în funcțiune a unui sistem OLAP și a softului de acces la date necesită o înțelegere clară a modelului de date al întreprinderii și a funcțiilor analitice cerute de conducerea executivă și strategică. Produsele comerciale pot fi de mare folos, dar rareori există soluții „la cheie” pentru OLAP; arhitectura trebuie reglată astfel încât să suporte sursele de date folosite și să facă față cerințelor. În schimb, odată ce sistemul OLAP este funcțional, suportul tehnic pentru utilizator este minimal.

**C. Tehnologiile Data mining** (mineritul în date) reprezintă, într-o accepțiune simplă, un mod automat de detectare într-o bază de date a unor tipare relevante. Data mining utilizează o serie de tehnici statistice și de inteligență artificială ce dau posibilitatea construirii de modele ce pot previziona comportamentul clienților. Tehnologia își sporește calitățile prin integrare cu depozitele de date comerciale și cu noile modalități de prezentare și raportare.

Data mining își datorește numele similarității dintre căutarea de informații valoroase într-o bază de date mare și săparea unor galerii în munte pentru detectarea unor zăcăminte valoroase.

**Data mining** este un proces de descoperire a cunoștințelor (Knowledge discovery -KD), de extragere a informației necunoscută anterior din baze de date foarte mari.

Procesul descoperirii de corelații semnificative, modele și tendințe se asigură prin explorarea unor mari cantități de date stocate în depozite de date, utilizând tehnologii de recunoaștere a modelelor, precum și tehnici statistice și matematice.

Conform unui raport din 1997 al Grupului Gartner: „Data mining și inteligența artificială se află între primele cinci tehnologii cheie care vor avea în mod sigur un impact major asupra unui mare număr de industrii în următorii ani”. Gartner situează data mining între primele 10 tehnologii în care firmele vor investi în următorii 5 ani.

Studiile arată că în ultima perioadă s-a înregistrat o explozie dramatică a nivelului de interes privind data mining, în condițiile în care utilizatorii au dorit să profite din instrumentele oferite de această tehnologie pentru a obține un avantaj competitiv inteligent în plan concurențial.

Anumite produse software de vîrf în domeniul data mining, provenite de la companii ca SAS sau IBM, reprezintă acum mai mult decît simple motoare de modelare bazate pe algoritmi complecși. Acestea se adresează unei categorii mai largi de

probleme tehnice și economice și se integrează în mediile actuale de tehnologie informațională.

Deși se credea că data miningul va elimina nevoia de specialiști în crearea de modele statistice, lipsa experienței umane și a intuiției între niște corelații relevante și unele nerelevante, lipsă ce caracterizează softurile, infirmă această previziune.

Tradițional sunt avute în vedere mai multe tipuri de analize statistice: **analize predictive, analize descriptive, analize confirmatorii și analize exploratorii**. Un exemplu tipic de analize predictive și descriptive sunt cele specifice acțiunilor și fenomenelor de marketing. În cazul analizelor confirmatorii, având o ipoteză formulată aceasta se acceptă sau se respinge. În analizele exploratorii, se urmărește găsirea de ipoteze, care apoi se acceptă sau se resping. În acest punct sistemul preia „inițiativa” în procesul analizei datelor sistemul gândește singur ipotezele acestea nefiind formulate de utilizator. În prezent termenul de data mining se referă la procesul automat de analiză a datelor în care sistemul preia inițiativa de a genera modele.

Din punct de vedere al procesului există trei clase de activități data mining: **descoperire, modelare predictivă și analiza excepțiilor** (figura 5).

- **Descoperirea** este procesul de căutare în baza de date pentru a găsi modele, fără a avea o idee predeterminată sau ipoteza asupra ceea ce pot fi modele. Cu alte cuvinte programul preia inițiativa în găsirea a ceea ce sunt modelele interesante, fără a fi necesar ca utilizatorul să se gândească la întrebările relevante în prealabil. În marile baze de date există atât de multe modele încât utilizatorul nu ar putea niciodată practic să se gândească la toate întrebările care ar trebui puse. Problema cheie în acest caz constă în bogăția de modele care pot fi găsite și exprimate, precum și în calitatea informației livrate – elemente care determină puterea și utilitatea tehnicii de descoperire.
- **Modelare predictivă** modelele descoperite din baza de date sunt folosite pentru a face previziuni. Modelarea predictivă permite astfel utilizatorului să prelucreze înregistrări ce au câmpuri valorice necunoscute, iar sistemul va intui valorile necunoscute pe baza unor modele anterioare din baza de date.
- **Analiza excepțiilor** reprezintă procesul prin care se aplică modelele extrase pentru a găsi anomalii sau elemente de date neobișnuite. Pentru a descoperi anomalii, mai întâi aflăm ceea ce e normal, apoi detectăm acele articole care deviază de la normă în cadrul unui interval dat. De exemplu odată ce am observat că 90% dintre cumpărători au sub 50 de ani, ne putem înteba în legătură cu cei 10% cumpărători care au peste 50 de ani și sunt în baza de date. Se observă că descoperirea ne poate ajuta să găsim „cunoștințe uzuale”, vreme în care analiza excepțiilor caută cazurile neobișnuite și specifice.



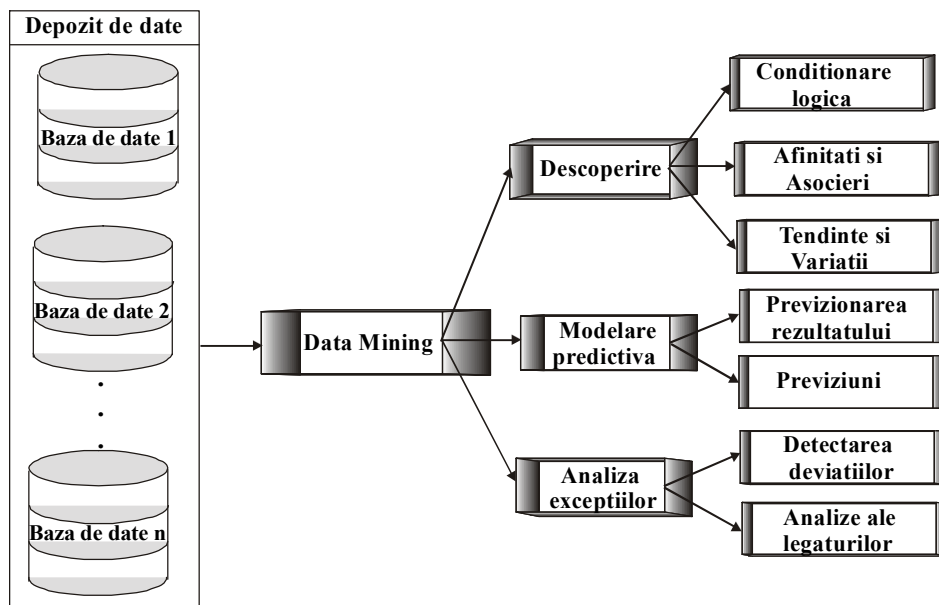


Figura 5: Clase de activități Data mining

Fiecare din aceste procese pot fi clasificate la rândul lor după regulile If/Then, asocieri etc. În timp ce IF/THEN presupune: Dacă – condiție, atunci regula 1, astfel regula 2, regulile de asociere se referă la gruparea articolelor (de ex. când cineva cumpără un produs la magazin, el poate cumpăra alt produs în același timp – un proces numit de obicei analiza coșului de cumpărături).

Din punct de vedere al tehnicilor utilizate de data mining (figura 6) identificăm:

- **Rețelele artificiale neurale** – modele de previzionare nonliniare care sunt capabile de acumulare de cunoștințe prin antrenament;
- **Algoritmi genetici** – tehnici de optimizare ce se bazează pe procese cum ar fi combinații genetice, mutații și selecție naturală. Sunt utilizate în modele ce folosesc concepte de evoluție naturală;
- **Arbori de decizie** – structuri pe trei nivele ce reprezintă seturi de decizii. Aceste decizii generează reguli de clasificare a seturilor de date. Printre modelele bazate pe arborii de decizie se include Arborii de regresie și clasificare (CART) și Detectarea automată a interacțiunilor de tip  $\chi^2$  (CHAID);
- **Metoda celui mai apropiat vecin** – o tehnică ce permite clasificarea fiecărei înregistrări din baza de date a k clase de înregistrări, cele mai similare;
- **Inducția unor reguli** – extragerea unor reguli utile de tip dacă-atunci pe baza semnificației statistice;
- **Vizualizarea datelor** – interpretarea vizuală a relațiilor complexe pentru date multidimensionale.

Cea mai importantă dihotomizare a tehnologiilor data mining se referă la păstrarea datelor; adică, se mai păstrează sau mai este nevoie de date odată ce au fost minerite. În abordările inițiale mare parte din seturile de date se păstrau pentru viitoare analize. În mod evident, o astfel de tehnică se poate aplica numai în cazul sarcinilor de tip modelare predictivă și analiză a excepțiilor, nefiind necesară în cazul descoperirii de cunoștințe întrucât nu mai sunt alte modele de distilat.

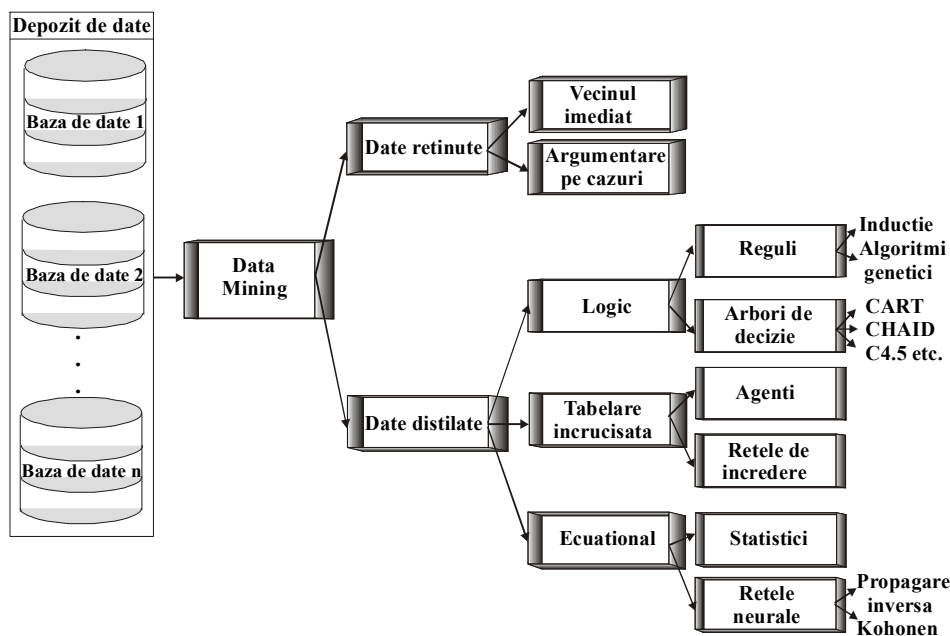


Figura 6: Tehnologiile data mining

Așa cum ușor se poate bănuși, abordările bazate pe păstrarea datelor pot avea ușor probleme din cauza dimensiunilor foarte mari ale unor astfel de seturi de date.

În timp ce în cazul distilării de modele se analizează datele, se extrag modelele, iar apoi se renunță la date, în abordările bazate pe păstrare, datele sunt reținute pentru analogiile de modele. Atunci când noi seturi de date sunt avute în vedere, atunci aceste seturi sunt comparate cu seturi anterioare de date.

Un exemplu cunoscut al abordării bazate pe reținerea datelor este metoda „vecinului imediat”. În acest caz setul de date este păstrat (de obicei în memorie) pentru comparații cu noi înregistrări. Când o nouă înregistrare este supusă analizei, este găsită „distanța” dintre acestea și înregistrările similare din setul de date și „cele mai similare” (sau vecinii cei mai apropiați) sunt identificați.

Tehnologiile bazate pe distilarea modelelor extrag modelele dintr-un set de date, apoi folosesc aceste modele pentru diverse scopuri. În mod evident, este necesar ca aceste modele să fie exprimate într-o anumită formă și limbaj. Această opțiune a dat naștere la trei abordări distincte: **logică, ecuații și tabelare încrucișată**. Fiecare dintre aceste abordări își are originile istorice în anumite secțiuni ale matematicii.

- **Logica** formează baza celor mai multe limbaje scrise. Modelele exprimate în limbaje logice se disting prin două principale caracteristici: pe de o parte sunt citibile și inteligibile, iar pe de altă parte sunt excelente pentru reprezentarea grupurilor de elemente de date. Sistemele logice condiționate pot fi separate în două grupuri distincte: reguli și arbori de decizie. Regulile condiționale pot fi implementate prin inducție sau algoritmi genetici. Există mai multe abordări și pentru generarea arborilor de decizie (de ex., CART, CHAID, C4.5).
- **Tabelarea încrucișată** este o formă simplă de analiză a datelor, bine cunoscută în statistică, și folosită pe larg în rapoarte. Un tabel încrucișat bidimensional este similar unei foi de calcul, cu capetele, atât de rânduri cât și de coloană, fiind valori ale atributelor. Celulele din foaia de calcul reprezintă o operație agregată, de obicei numărul de co-apariții ale

valorilor atributelor. Multe tabele încrucișate sunt efectiv echivalente cu un grafic tridimensional care înfățișază o numărătoare a co-aparițiilor. Termenul de „agent” este folosit uneori pentru a desemna tabelele încrucișate care sunt reprezentate grafic într-o rețea și permit folosirea unor conjuncții AND, OR etc. Rețelele de încredere, uneori numite rețele cauzale, sunt un graf direcționat, constituit din noduri (variabilele reprezentate) și legături între noduri (coeficienți probabilistici de încredere).

- **Abordările ecuaționale** folosesc de obicei un set de ecuații pentru a defini „o suprafață” (reprezentare grafică a unei funcții de orice grad) în cadrul unui spațiu iar apoi măsurarea distanțelor de la această suprafață pentru a face estimări.

O problemă a tehnicilor data mining a reprezentat-o necesitatea păstrării sau nu a datelor obținute din minerit. Au apărut **tehnologii cu păstrarea datelor** și **tehnologii fără păstrarea datelor**. Păstrarea datelor se poate aplica numai în cazul sarcinilor de tip modelare predictivă și analiză a excepțiilor, nefiind necesară în cazul descoperirii de cunoștințe întrucât nu mai sunt alte modele de distilat.

În timp ce în cazul distilării de modele se analizează datele, se extrag modelele, iar apoi se renunță la date, în abordările bazate pe păstrare, datele sunt reținute pentru analogiile de modele. Atunci când noi seturi de date sunt avute în vedere, atunci aceste seturi sunt comparate cu seturi anterioare de date.

Rezultatele obținute prin data mining trebuie să fie operaționale și să poată fi aplicate pentru obținerea unui anumit avantaj, fapt dependent în mare măsură și de abilitățile decidentului. Data Mining și OLAP sunt instrumente ale afacerilor inteligente bazate pe tehnologia IT&C.

Cererile OLAP se adresează unei baze de date existente generând ipoteze și relații, OLAP fiind în esență un proces deductiv. Data mining utilizează datele pentru a descoperi modele printr-un mecanism predictiv.

În concluzie tehnologia OLAP și tehnologia Data Mining se completează reciproc contribuind la utilizarea eficientă a depozitelor de date.

Literatura de specialitate prezintă numeroase și variate situații care impun utilizarea depozitelor de date și a tehnicilor moderne de accesare: insuficienta partajarea a resurselor, grupuri distincte de manageri produc rapoarte contradictorii, procesul de creare a rapoartelor este foarte anevoios, rapoartele nu sunt dinamice și nu favorizează stilul de interogarea online, rapoarte care necesită date de istoric sunt dificil de realizat, nivelul de pregătire și de înțelegere a personalului de decizie, a analistului sau specialistului lasă de dorit..

Instrumentele data mining vor continua să se maturizeze și din ce în ce mai multe organizații vor adopta acest gen de tehnologie, încorporat în instrumente software din ce în ce mai performante. Inițiativele data mining provin cel mai adesea din zona departamentelor de marketing și vânzări care dețin baze de date cu volume foarte mari, sunt generate cu instrumente software care accentuează și garantează calitatea datelor din depozitele de date iar evoluțiile din ultima perioadă impun analize detaliate, analitice, pe perioade de timp, pe zone și segmente diferite pentru a oferi răspunsuri operative și eficiente

pe o piață globalizată și pentru un client individualizat și personalizat.

Tehnologiile date warehouse , OLAP, data mining sunt influențate semnificativ de popularitatea soluțiilor inteligente bazate pe Internet și Intranet, suportă facilitățile oferite de aplicațiile WEB, prin hypertext, hyperlink și hypermedia, iar unele organizații deja folosesc aceste instrumente pentru prelucrarea pe Internet a informațiilor din depozitele proprii de date distribuite sau pun aceste informații la dispoziția partenerilor strategici ai organizației

### **Bibliografie:**

- Andone, I. Tugui, A.**, s.a. (2001) - *Dezvoltarea sistemelor inteligente în economie*, Ed. Economică, București, 2001;
- Cătoiu, I.** (coordonator) (2002) - *Cercetări de Marketing*, Ed. Uranus, București, ;
- Cipolla, E. T.** (1995) „Data Mining: Techniques to Gain Insight Into Your Data”, *Enterprise Systems Journal*, Decembrie 1995;
- Harmon, P. și King, D.** (1985) *Expert Systems: Artificial in Business*, John Wiley & Sons, New York.
- Inmon, W.H.** (1996) *Buiding the Data Warehouse*, John Wiley & Sons, New York.
- Orzan G.(2001)**, - *Sisteme informatice de marketing*, Ed. Uranus, București;
- Turban, E., Jay Aronson J.** (2001)– *Decision Support Systems and Intelligent Systems*, Ed. Prentice Hall, New Jersey, USA;
- Zaharie, D., Albescu, F.**, ș.a (2001) – *Sisteme informatice pentru asistarea deciziei*, Ed. DuAl Tech București, 2001,
- Wallinger, J.** (1997) *Database Marketing Guide*, Ed. Dryden Press, New York.