

Kolegij: Sustavi za potporu odlučivanju

Materijal za čitanje uz poglavlje 4 – Skladištenje podataka

Izvor: <http://www.skladistenje.com/tekst.asp?cont=pocetnici>

ABC od podataka do informacije, dio prvi

U zadnjih nekoliko godina **Business Intelligence** područje u svijetu je jedno od najinteresantnijih područja implementacije informacijskih sustava u kojem se obrću goleme količine zelembača... U Hrvatskoj je BI još uvijek (na žalost) za širi krug korisnika Terra incognita i u ovoj dvodjelnoj kolumni ćemo se orijentirati na tehničko-tehnološki framework – kroz case study zamišljenog BI projekta vidjeti što sve na putu od podatka do informacije treba napraviti. Naravno, u cijeloj BI priči, najprije valja postaviti pitanje...

Zašto, Why, Warum, Per Que?

Pojam Business Intelligence postojao je i mnogo ranije – to smo nazivali “poslovnom špijunažom”. Sve je to počelo 1992. godine kad je “kum” skladištenja podataka Bill Inmon definirao skladište podataka kao **skup subjektivno orijentiranih, integriranih, vremenski ovisnih i nepromjenjivih podataka za podršku poslovnom odlučivanju**. Razlog zbog kojeg je cijeli koncept nastao je rasterećenje transakcijskih sustava (famoznih ERP-ova) od upita za koje oni nisu optimirani poput onih koji sumariziraju milione slogova iz više tablica s više uvjeta u nekoliko brojki zanimljivih rukovoditeljima. Zbog toga se podaci iz transakcijskih aplikacija u regularnim vremenskim intervalima prebacuju u drugi sustav optimiran za takve upite za potrebe izvještavanja i analize. I nakon toga je uslijedio pravi boom – ljudi su shvatili da se u tim podacima krije pravo blago za analizu različitih uzoraka u podacima (a i tehnologija je to mogla omogućiti), koji, ako se pravilno analiziraju i protumače, mogu donijeti kompaniji veliku kompetitivnu prednost na tržištu i financijsku korist. Nakon nekoliko godina, sve te tehnologije za analizu svih internih i eksternih informacija iz poslovnog okruženja su zajedničku kapu našle pod pojmom Business Intelligence, odnosno u zadnje vrijeme **e-Intelligence** pod utjecajem Internet revolucije...

Naš case study biti će izrada sustava za podršku poslovnom odlučivanju koji će pokriti sve potrebe kompanije za izvještavanjem, analizom, predviđanjima i planiranjem. Postoje tri stvari koje moraju biti postavljene na svoje mjesto prije početka – svaki BI projekt mora zadovoljiti barem tri osnovna uvjeta koji će mu osigurati uspješnost (naravno, ima toga još, ali ovo su ipak osnove):

1. mora imati jasno definirane ciljeve – ako to nije slučaj, projekt će se jednostavno raspliniti jer nitko neće shvaćati što i zašto se radi
2. mora biti vođen od poslovne strane i imati podršku top rukovodstva – u suprotnom će biti smatran još jednim hirom IT-a
3. mora imati osigurano odgovarajuće financiranje – takvi projekti i njima pripadajuće tehnologije su još uvijek vrlo skupe...

Projekt ima slijedeće ciljeve – omogućiti transparentno izvještavanje o prodaji u cijeloj organizaciji u skladu s ovlaštenjima, analizu podataka iz prodaje i o kupcima, predviđanje prodaje u slijedećem razdoblju i na temelju predviđanja planiranje, budžetiranje i praćenje izvršenja plana. Radi mira u kući analizu podataka s web-siteta ostavljamo za drugu fazu projekta i zanemariti ćemo dijelove projekta vezane uz definiciju izvještajnih potreba, te se koncentrirati na tehnologiju u pozadini...

Svako skladište svoju bazu ima...

Kao prvi korak, **treba odabrati platformu** na kojoj će se sve odvijati – to mora biti kvalitetan i pouzdan DBMS, budući da će u svakoj srednjoj ili većoj organizaciji skladište podataka koje služi kao podloga svemu što slijedi vrlo brzo narasti na desetke ili stotine Gigabajta... Najjednostavnije je da to bude DBMS koji već koristite, budući da ga već ionako imate licenciranog i ne trebate ga kupovati. Tu naravno ima iznimaka – mainframe platforme nisu baš pogodne za takve projekte jer zahtijevaju velike memorijske i diskovne resurse (skupo...), pa se skladišta uglavnom baziraju na DBMS-ovima na Unix i

Windows platformama. Postoje DBMS-ovi napravljeni isključivo za skladištenje podataka – najpoznatiji su NCR-ov **Teradata** i **Redbrick**, sada u vlasništvu Informixa (odnosno IBM-a) – koji omogućavaju razne zgodne i funkcionalne stvari poput kreiranja agregacija prilikom učitavanja podataka, te razne načine indeksiranja podataka radi brzih operacija nad podacima u skladištu. U zadnjih nekoliko godina i vodeći RDBMS sustavi na tom polju daju izvrsne rezultate (**Oracle**, **Informix**, **DB2** i **SQL server**), a svaki ima svoje prednosti u odnosu na konkurente i čvrsto drži svoju tržišnu poziciju. U svakom slučaju, na rasprave o prednostima i vrlinama svake od solucija potrošeno je jako puno tinte, te ćemo se time detaljno pozabaviti drugom prilikom – za sada je dovoljno da znamo da je platforma odabrana i da možemo krenuti dalje.

Projektiranje i optimizacija skladišta podataka

Slijedeće pitanje je **koje podatke** uopće trebamo u skladištu? Nakratko uzamemo definicije izvještajnih potreba i definiramo podatke koje ćemo iz transakcijskih podataka učitati u skladište. Naravno, ti podaci su u trećoj normalnoj formi, kako svaka dobar ERP sustav nalaže – takvu formu na žalost skladište najčešće ne poznaje iz jednostavnog razloga što ne bi radilo brzo kad se postavi upit na nekoliko (stotina) milijuna slogova. Dakle, prilikom učitavanja podataka radi se denormalizacija i u podacima se pojavljuje određena količina redundancije, što ionako nije bitno ako se sjetimo nepromjenjivosti iz inicijalne definicije – podaci se kreiraju i mijenjaju u ERP-u, a u skladište se samo pune... Dva osnovna oblika tablica koja se pojavljuju u skladištu su fact table ili činjenična tablica i dimension table ili dimenzijska tablica. U činjeničnoj tablici nalaze se podaci koje želimo analizirati (na primjer, tablica prodaje), dok se u dimenzijskim tablicama nalaze podaci po kojima analizirane podatke želimo grupirati (na primjer, tablica prodavaonica, tablica artikala i tablica vremenskih razdoblja). Ove dvije vrste tablica mogu se organizirati u dvije osnovne strukture – prva i jednostavnija je **star schema** ili zvjezdasta shema u kojoj su sve dimenzijske tablice primary key – foreign key vezom vezane na činjeničnu tablicu (puno redundancije), dok je druga i općenitija **snowflake schema** ili pahuljičasta shema u kojoj dimenzijske tablice imaju i međusobne veze i hijerarhije (manje redundancije).

Koliko god cijena računala padala, skladištenje podataka je strašno zahtjevan hobi – probajte zamisliti skladište od 200 GB podataka i fact table s 150 milona slogova s podacima iz maloprodaje i tri dimenzijske tablice – vrijeme, prodavaonice i artikli. Korisnik postavlja upit u kojem želi detaljne podatke za tekuću godinu za grupu artikala i za prodavaonice u Karlovcu – to je join četiri tablice s tri where filtera... Naravno, svaki rezultat koji stiže nakon više od 30 sekundi je ravan katastrofi, a to je tek jedan od 250 korisnika sustava... Postoji nekoliko načina da se takvi upiti izvode brzo – jedan je naravno moćni n-procesorski stroj s 16 GB RAM-a, a drugi je **optimizacija**. Mnogo je načina na koji se sustav može optimirati i fino tunirati, ali najbitnija rješenja nalaze se na početku – treba pažljivo definirati podatke koji ulaze u skladište (sve što je u skladištu a ne koristi se je bespotreban overhead koji samo troši resurse), zatim tablicama treba pažljivo odabrati i kreirati indekse, definirati i kreirati potrebne agregacije, te definirati odnose među tablicama i odabrati prikladnu shemu odnosa. Najsigurniji rezultat daje metoda pokušaja i mjerenja vremena odziva. Ako imate OLAP server i kažete da optimizacija skladišta nije ni toliko bitna – varate se, i kreiranje OLAP kocke često koristi brzinu indeksiranih tablica i znatno skraćuje noćne obrade, a vrlo često korisnici traže i mogućnost drill-trougha do granularnih podataka.

Kad je sustav u funkciji, vrlo je važno mjeriti korištenost pojedinih dijelova, kako bi se mogle dodati agregacije koje se često traže a nisu prethodno kreirane ili kako bi se podaci koji se rijetko traže mogli prebaciti na jeftiniji i sporiji medij. Morate voditi i evidenciju o tome tko koliko koristi sustav i da li postoje pokušaji neovlaštenog pristupa.

Nakon definiranja potrebnih podataka, agregacija i strukture skladišta treba napraviti inicijalno učitavanje.

A tko će sve to napuniti i počistiti?

Skladište zahtijeva **initial load** ili inicijalno punjenje s postojećim historijskim podacima – ono se naravno radi odmah nakon definicije strukture i traje relativno dugo. Podaci koji se učitavaju u skladište u pravilu dolaze iz nekoliko različitih sustava s različitim platformi, te se učitavaju preko različitih sučelja, bilo native drivera ili putem *ODBC*-a. Kad je skladište u funkciji, podaci u njemu se u redovnim vremenskim razmacima – dnevno, tjedno ili mjesečno – moraju nadopunjavati ili osvježavati,

što se naiva **incremental load**. U tu svrhu pišu se procedure za učitavanje koje su praktično uvijek automatizirane i schedulerane u noćno doba, kako bi podaci uvijek bili svježiji. Naravno, nakon svakog osvježavanja i agregacije u skladištu moraju se obnoviti, što može trajati i po nekoliko sati...

Koliko u ERP sustavu ima artikala koji nemaju definiranog dobavljača, koliko ima kupaca s upisanim krivim imenom ili adresom, koliko je krivih knjiženja? Takvi problemi rješavaju se također prilikom učitavanja na nekoliko načina – ili se ignoriraju (najlošije), ili se automatski korigiraju (bolje) ili se identificiraju, odvajaju, ispravljaju u ERP-u i ispravni učitavaju u slijedećoj iteraciji (najbolje). Međutim, s kontrolom kvalitete podataka procedure učitavanja postaju složene i međuzavisne i prijete da se otrgnu kontroli. Rješenje se zove **Extraction, Transformation and Loading Tool (ETL)** – specijalizirana grupa alata koji služe automatiziranju procedura za ekstrakciju, transformaciju i čišćenje podataka, te punjenje istih u skladište. Često su to vrlo skupe, moćne i sofisticirane platforme, od kojih su najpoznatije **Ardent DataStage** i **Informatica**, ali također postoje i jeftinije i jednostavnije solucije – no to je ponovo zasebna tema o kojoj se može dugo raspravljati.

Nakon što smo odabrali ETL tool, uspješno ga implemetirali i riješili problem brzog učitavanja i kvalitete podataka, problem na kojem su se mnogi prije njega spotakli i na kojem je mnogi BI projekt pokleknuo, te se sada našli pred slijedećim pitanjem – OK, podaci su u skladištu, a što sad? Moramo ih nekako iskoristiti... Tko će imati pristup kojim podacima? Što će tko s podacima raditi? Kako ćemo informacije distribuirati? Na sva ta pitanja ćemo pokušati odgovoriti u drugom dijelu kolumne za mjesec dana... (D.O.)

<http://www.skladistenje.com/jedan.asp?ID=179>

ABC od podataka do informacije, dio drugi

U prvom dijelu priče o Business Intelligence projektu pokrenuli smo projekt, definirali poslovne potrebe, dizajnirali i napunili Data Warehouse s kvalitetnim i pročišćenim podacima. U drugom dijelu ćemo se koncentrirati na tehnologije i metodologije kreiranja informacija iz postojećih podataka – platforme za reporting, OLAP i Data Mining. Pokušati ćemo odgovoriti na slijedeća pitanja: Kako ćemo podatke pretvoriti u korisne informacije? Kako osigurati odgovarajuća ovlaštenja pristupa informacijama? Koji softver koristiti za prikaz informacija korisnicima?

Metapodaci

Ono što zaista omogućava da proces kreiranja informacije od nepreglednih količina podataka su *metapodaci*, odnosno podaci o podacima. U metapodacima piše od kuda se koja tablica učitava, što sadrži, s kojim se tablicama i kako kombinira u slijedećim koracima, tko je njen vlasnik i/ili kreator, koliko često i u koje doba se nadopunjava, tko je smije koristiti. Za svaku proceduru piše tko i kada je može pokrenuti, koji su joj ulazni, a koji izlazni parametri. Za svakog korisnika piše koje izvještaje i podatke smije vidjeti, učitati ili promijeniti. Jednostavno rečeno – u metapodacima piše sve bitno za cijeli proces...

Naravno tu postoji i problem – svaki dobavljač ima svoj način kreiranja i spremanja metapodataka, te se metapodaci između produkata različitih dobavljača mogu razmjenjivati samo ako postoji odgovarajući *interface*, koji ponekad postoji, a ponekad i ne, što predstavlja veliki problem. Postojale su dvije grupe koje su pokušale kreirati standard metapodataka koje su se nedavno integrirale i sada se pod okriljem *Object Management Group* (OMG koji je definirao CORBA standard) razvija *Common Warehouse Metamodel*, zajednički standard metapodataka koji bi za koju godinicu trebali podržavati svi aktualni proizvodi na tržištu, što će korisnicima život učiniti puno lakšim...

Thick client ili thin client, pitanje je sad...

Na ovo pitanje na svu sreću nije bilo teško odgovoriti. Iako Windows klijenti još uvijek imaju puno veće mogućnosti od klijenata koji se baziraju na korištenju browsera, razlika u funkcionalnosti se sve brže smanjuje. činjenica je i da će većina korisnika koristiti samo dio funkcionalnosti koju Windows klijent može pružiti, te da su uglavnom puno složeniji za upotrebu (to naravno zahtijeva i pripadajuću edukaciju) od *thin clienata* koji zahtijevaju od korisnika uglavnom poznavanje klikanja mišem i gledanja

u ekran. Na strani *thin clienta* je puno jednostavnija centralna administracija sustava i nepostojanje bilo kakve instalacije ili konfiguracije kod klijenata, što je izuzetno važan argument u fazi eksploatacije i održavanja sustava. Točan odgovor je, dakle, otprilike ovakav – za prosječnog korisnika sve potrebe zadovoljava *thin client*, dok za nekolicinu kreatora sustava i analitičara podataka treba osigurati punu funkcionalnost Windows klijenta.

Kod *thin client* arhitekture, *enterprise intelligence portal* postao je praktično nezaobilazni dio koncepta – on omogućava da klijent personalizira izgled sučelja sustava prema vlastitim preferencama, da ima dostupnost s jednog mjesta do svega do čega ima pravo dostupa, da “svoje” izvještaje publicira i daje na uvid drugim korisnicima (u skladu sa svojim pravima), te da se pretplaćuje na izvještaje na koje ima pravo. Time se administracija sustava znatno smanjuje, a fleksibilnost korištenja i zadovoljstvo korisnika znatno povećavaju.

Autentikacija i sigurnost pristupa

Slijedeće važno pitanje je sigurnost pristupa podacima – kako osigurati da onaj koji informacije treba vidjeti zaista vidi sve informacije koje su mu potrebne, a da onaj koji ih ne smije vidjeti ih zaista ne vidi? Do nedavno se autorizacija pristupa uglavnom temeljila na *database* ovlastima ili na rješenjima koje su korisnici interno razvijali. Na sreću, pojavio se *Lightweight Directory Access Protocol (LDAP)* i postao praktično svjetski standard za autorizaciju, te svi bitniji noviji produkti na tržištu podržavaju integraciju s nekim od vodećih LDAP servera. Princip integracije je jednostavan – kod instalacije samo navedete ime ili IP adresu postojećeg LDAP servera, a instalacija sama instalira potrebne ekstenzije i administratorske alate i sustav je vrlo brzo spreman za korištenje. Većina dobavljača ima i pripadajući *ticket server* koji svakom korisniku kod prijave izdaje *ticket* koji vrijedi određeno definirano vrijeme – na taj način *password* od klijenta do servera putuje samo jednom, a ako se netko i dokopa ticketa, nakon prolaska definirane količine vremena isti postaje nupotrebljiv.

Enterprise BI Suite ili best-of-breed koncept

Kad imate podatke u skladištu podataka, te odgovarajuće procedure za redovno dopunjavanje, postavlja se pitanje kako te podatke iskorištavati. Naravno, nema skladišta podataka koje može dati odgovor na sva pitanja – novija istraživanja provedena među velikim američkim kompanijama govore da većina tih kompanija ima četrdeset ili više skladišta podataka s više stotina terabajta podataka – stoga se alati za transformaciju podataka u informacije i distribuciju odabiru u ovisnosti o onome što korisnik treba.

Postoje dva osnovna koncepta koji imaju svoje prednosti i mane. Prvi koncept se zasniva na korištenju *Enterprise Business Intelligence Suite (EBIS)*, platforme istog dobavljača koja sadrži nekoliko alata za reporting, OLAP, vizualizaciju i Data Mining, te portalsku aplikaciju za pristup informacijama. Prednosti ovakve solucije su očigledne – jednostavna integracija alata istog proizvođača i uklapanje u cjelokupnu sliku, jednostavnija administracija sustava, jedan standard metapodataka koji se jednostavno razmjenjuje među dijelovima suite, *single sign-on* standardno podržan od platforme i jednostavno za implementaciju. Mane? Da li ste ikada išli na jutarnji jogging u Nike tenisicama, majici, trenirci i kapi? Vjerojatno niste, budući da imate Adidas tenisice, Nike trenirku, Reebok majicu i Lotto kapu – stvari koje ste kupili jer su vam se najviše sviđale i jer ste ih smatrali najboljima. Tako je i s EBIS platformama – u nečemu su bolje a u nečemu lošije, što nas vodi drugom konceptu – *best-of-breed* solucijama.

Garner grupa definira Enterprise BI Suitu kao paket alata koji omogućava kreiranje upita, izvještaja, vizualizaciju i OLAP, uz ostvarenje određenih uvjeta skalabilnosti, iskoristivosti i upravljivosti. Trenutačno u svijetu postoji svega nekoliko kompanija koje takvo integrirano i zaokruženo rješenje mogu ponuditi. Na vrhu su prilično neprikosnovene dvije kompanije – Cognos i Business Objects sa svojim BI platformama. Njih u stopu slijede MicroStrategy i Brio Technology, dok su u utrci još i Crystal Decisions (donedavno Seagate Software) i Hummingbird, i to bi otprilike bili svi. Naravno, postoji gomila drugih dobavljača s gomilom drugih dobrih proizvoda, ali oni ne zadovoljavaju osnovno EBIS pravilo – jednostavna integracija s ostalim komponentama i princip izgradnje sustava sličan slaganju lego-kockica...

Da bi skup alata bio suite, treba imati slijedećih pet nivoa funkcionalnosti:

1. *single sign-on user interface* – praktično svi sada već imaju neku vrstu portala, te su *thin client* rješenja definitivno dominantna
2. *BI services* nivo – serverske aplikacije koje omogućuju dinamičko posluživanje informacija klijentima (reporting, OLAP, ad-hoc query i mining serveri)
3. *Metadata* nivo – središnji nivo koji kreira, standardizira i objedinjava metapodatke svih komponenata sustava
4. *Creation* nivo – nivo za učitavanje i dostup do potrebnih podataka
5. *Security* nivo – nivo koji se brine za sigurnost dostupa i raspolaganja podacima i procedurama u sustavu za sve ostale nivoe

Prema potrebi, periferni nivoi (najčešće je to *Creation* nivo) mogu biti nadomješteni nekim drugim proizvodima, dok se u *BI services* nivou mogu koristiti svi ili samo neki dijelovi – možda negdje postoji potreba za OLAP analizama, dok drugdje postoji potreba samo za običnim relacijskim izvještajima...

U *best-of-breed* konceptu za svaki vid transformacije i korištenja informacija koristi se ponajbolji alat koji se može naći na tržištu. Mnoge kompanije se nađu u *best-of-breed* soluciji bez namjere – pokreću se manji projekti po raznim odjelima na različitim platformama, koji zadovoljavaju dio potreba, te se kasnije integriraju u cjeloviti sustav. Prednosti tog koncepta su kvaliteta pojedinih komponenti, a mana je složena integracija i administracija, te nepostojanje globalnog standarda metapodataka.

Nakon odabira alata koji će se koristiti za transformaciju podataka i iskorištavanje informacija, treba pristupiti iskorištavanju informacija – SQL reportingu, OLAP-u i Data Minigu.

Report generatori

Prvi i prilično standardan zahtjev za korištenje informacija je kreiranje statičnih izvještaja – to nije jako sofisticiran i zahtjevan posao ako niste suočeni s tablicama od više desetaka miliona redaka, s nekoliko stotina korisnika i nekoliko stotina izvještaja koje dnevno trebate isporučiti. U slučaju da niste suočeni s tim problemom, zadovoljiti će vas Microsoft Access ili Crystal Reports za one naprednije. U slučaju da jeste, morati ćete posegnuti za snažnijom platformom – ako koristite Oracle to će biti vjerojatno Oracle Reports, a čak i u tim uvjetima se Crystal Reports jako dobro drže. Nabrajati ostale dobavljače reporting platformi moglo bi potrajati u nedogled, ali korisno je napomenuti da praktično sve EBIS platforme imaju vrlo kvalitetne report generatore, te omogućavaju i jednostavnu http distribuciju gotovih izvještaja. Vrlo je važno napomenuti da obavezno birate alat koji izvještaje publicira u *.pdf formatu (u tom slučaju garantirano nema problema s ispisom na štampaču) i koji omogućava jednostavan export podataka iz izvještaja u Excel (to valjda ne treba objašnjavati zašto...)

OLAP - Ulaznica u višedimenzionalni svijet

OLAP (On-Line Analytical Processing) je kao pojam nastao prilično davno, ali tek u zdanjih pet godina je postao prilično "vruć". Ideja počiva na tome da je poslovanje kompanije i funkcioniranje ljudskog uma u biti višedimenzionalno – ako želimo analizirati npr. prodaju, onda je želimo analizirati u vremenu, po regijama, po prodajnim mjestima, po artiklima, po kupcima... Kad bismo za svaku tu dimenziju radili zasebnu sumarijacijsku tablicu za potrebe izvještavanja, to je dugačak i mukotrpan posao. Ideja OLAP-a se zasniva na *engine-u* koji takve multi-dimenzionalne sumarijacije kreira automatski i stavlja ih korisniku na raspolaganje, čime se stvara mogućnost kombiniranja najrazličitijih uvjeta (prodaja cipela br.43 u Rijeci svaki četvrtak popodne...) koji neovisno o svojoj smislenost, korisniku daju odgovor na postavljeni upit u realnom vremenu, bez velikog potrebnog predznanja. U početku su se OLAP tehnologije dijelile na dva osnovna pravca. Prvi je bio MOLAP (*Multidimensional OLAP*) koji je od izvorne relacije napravio multi-dimenzionalnu kocku – njegova prednost bila je brzina, a mana ogromno zauzeće prostora u slučaju postojanja većeg broja dimenzija. Drugi je bio ROLAP (*Relational OLAP*) koji je na postojeću relaciju dodao samo sumarijacije – njegova prednost je bio mali prostorni *overhead*, što je plaćano smanjenom brzinom. Na kraju je prevladao hibridni HOLAP koncept (*Hybrid OLAP*), koji samo sumarijacije drži u multi-dimenzionalnoj kocki, dok elementarni nivo podataka drži u izvornoj relaciji i njima pristupa pomoću *drill-through* procedura – na taj način objedinjene su i velika brzina pristupa i relativno malo zauzeće prostora.

Vrlo važnu ulogu u prihvaćanju OLAP tehnologija odigrao je Microsoft koji je prvi sa svojim SQL severom 7 krajem 1998. godine isporučio OLAP server kao sastavni dio bez dodatne cijene - do tada

su OLAP serveri koštali više desetaka tisuća dolara... Taj primjer su, htjeli ili ne, morali slijediti i ostali *database* dobavljači, pa tako u novim verzijama DB2 i Oracle za Win NT/2000 platformu imaju uključen OLAP server.

Ostaje otvoreno pitanje OLAP klijenta kojeg će koristiti prosječan korisnik – na tom tržištu ponuda je zaista izdašna, a najčešće ste vezani za korištenje klijenta koji podržava određeni server (vidi pod metapodatke...). Zbog širine ponude, odabir OLAP klijenta ne bi trebao predstavljati veći problem – bitno je paziti da podržava isti standard autorizacije kao i ostale komponente sustava.

Kopanje podataka ili Data Mining

Najzanimljiviji i najsofisticiraniji oblik korištenja podataka u skladištu je kopanje podataka (Data Mining) što podrazumijeva otkrivanje na prvi pogled nevidljivih korelacija među podacima korištenjem sofisticiranih statističkih modela. Primjer po kojemu je kopanje podataka postalo u svijetu poznato je korelacija između prodaje piva i pelena petkom popodne. Za one koji slučajno ne znaju, u to doba tate idu s posla kući, kupuju pelene po maminoj narudžbi, pa kad su već tamo, uzmu za sebe i ekipu pivu da ne bi utakmice preko vikenda gledali "na suho"... Iako je taj primjer doživio nevjerojatnu eksploataciju do nivoa banalizacije, pomoću *data mining* metoda i tehnologija, kompanije zaista mogu doći do vrijednih podataka o ponašanju svojih kupaca kojih na prvi pogled nisu bile svjesne, te podatke iskoristiti da bi korisnike segmentirale, osmislile proizvode i usluge prikladne pojedinom segmentu, korisnike učinile zadovoljnijima, a svoje prihode povećale... Ulaženje u detalje pojedinih modela odlučivanja u *data mining* procesima tema je za zasebnu analizu, ali nedvojbeno je činjenica da model ovisi o pouzdanosti i sveobuhvatnosti statističkog modela koji mu stoji u podlozi, te stoga ni ne čudi da je predvodnik na tržištu *data mining* alata SAS, kompanija koja dolazi iz područja statističke analize, čiji je produkt *Enterprise Miner* u svijetu trenutačno praktično neprikosnoven. Naravno, za manje sofisticirane potrebe i alati ostalih dobavljača poslužiti će svrsi. Kopanjem podataka se u organizaciji praktično uvijek bavi najviše nekoliko analitičara, tako da taj dio ne mora biti nužno integriran s reportingom i OLAP-om...

Sretan završetak ili tek početak?

Zaista, nakon implementacije svega ovoga, postignit je vrlo visok stupanj "informiranosti" u organizaciji, te je projekt napravio iskorak u sasvim novi način promišljanja poslovanja – organizacija se sve više počela fokusirati na klijente kao centar poslovnog svijeta! Jednog dana će se trebati pozabaviti i novotarijama i preostalim *buzzwordima* poput *Customer Relationship Managementa* (CRM), *Clickstream* analize, *Operational Data Storea*, *Balanced Scorecardinga* i mnogim, mnogim drugim. U tim temama u slijedećim nastavcima... (D.O.)