

VISOKA TEHNIČKA ŠKOLA STRUKOVNIH STUDIJA
KRAGUJEVAC

Dr Miroslav Banković, prof.

POSLOVNA INTELIGENCIJA

Kragujevac, 2012.

1. ŠTA JE POSLOVNA INTELIGENCIJA?

Poslovna inteligencija (engl. Business Intelligence) je skup metodologija i softverskih alata za identifikaciju, ekstrakciju i analizu poslovnih podataka, koji omogućavaju:

- efikasno korišćenje podataka (najčešće iz skladišta podataka (engl. Data Warehouse))
- pretvaranje podataka u informacije potrebne za donošenje poslovnih odluka.

Cilj poslovne inteligencije je da se iz velikog obima internih i eksternih poslovnih podataka preduzeća, uočavanjem njihovih veza i zakonitosti pojavljivanja, dobiju upravljačke informacije kojima se povećava uspešnost poslovanja.

Kao **radovi i otkrića kojima je postavljena osnova** za otkrivanje znanja u bazama podataka mogu se navesti

- radovi starogrčkih matematičara Euklida i Pitagore (zaslužni za razvoj algoritama za klasterisanje podataka)
- radovi Bajesa, Paskala, Laplasa, Njutna, Lobačevskog i Gausa (razvoj teorija verovatnoće, optimizacije i dr.).

Prvi autor koji je upotrebio izraz "poslovna inteligencija" bio je IBM-ov istraživač Hans Peter Lun. U članku napisanom 1958., definisao ju je kao "sposobnost zapažanja veza među prezentiranim podacima na način koji pomaže aktivnostima koje vode ka određenom cilju".

POREKLO POSLOVNE INTELIGENCIJE U SADAŠNJEM SMISLU

Koreni poslovne inteligencije mogu se pratiti unazad do prvih obrada podataka - jednostavnih aplikacija kao što su praćenje dugovanja i potraživanja. Ove aplikacije koristile su tehnologije sa sekvencijalnim pristupom, kao što su papir i magnetne trake. Korišćenje sekvencijalnih memorijskih medija za skladištenje značilo je da je celoj datoteci moralo da se pristupi čak i ako je samo deo datoteke potreban. Feromagnetni materijal se skidao sa magnetnih traka usled kontaktnog čitanja, tako da su čitave datoteke gubljene. Ovi problemi doveli su do potrebe za novim načinom memorisanja i analiziranja informacija. U ovom kursu ćemo razmotriti kako preduzeće rešava izazove izdvajanja podataka iz sistema za onlajn obradu transakcija uz loš kvalitet podataka, kao i probleme strukturisanja podataka. Opisaćemo kako nastaje skladište podataka iz modela podataka, tabela strukture, granulacije i podrške istorijskim podacima, i kako se to uklapa u širi intelektualni koncept nazvan "korporacijska fabrika informacija", o kome će kasnije biti više reči.

Rane aplikacije obrade podataka

Sa sekvencijalnom memorijom, podaci su organizovani u tzv. master datoteke, koje su sadržale centralizovane podatke upotrebljive u više aplikacija. Bušene kartice, magnetne trake i papirni izveštaji koji su nastajali kroz aplikacije brzo su pali "u senku" memorija na magnetnim diskovima, koji omogućuju direktni i efikasan pristup podacima. Procesori su se naglo razvijali i postajali snažniji i efikasniji, dramatično je rasla brzina a opadali su troškovi obrade. Sa memorisanjem podataka na magnetne diskove, master datoteke su mutirale u baze podataka – centralizovane kolekcije podataka na disku, raspoložive za obradu u svakoj aplikaciji koja ima potrebe za njima. Sa memorijom na disku, transakcije se mogu obrađivati direktno i online iz baze podataka, što je značilo izvršenje transakcije

za 2 – 3 sekunde. Ubrzo je online obrada transakcija prerasla u online aplikacije koje koriste jedinstvenu centralnu bazu podataka.

Online aplikacije su se fokusirale na visoku raspoloživost i dobro vreme odziva. Time su ubrzo zauzele centralnu poziciju u funkcionisanju poslovnog sistema, postavši npr. suštinski deo direktnе interakcije sa kupcima. Ipak, razvoj aplikacija zasnovanih na online transakcijama doveo je i do sledećih izazova:

- zastarevanje aplikacija do puštanja u online rad
- stvarni broj aplikacija
- nedovoljna dokumentacija
- novi zahtevi
- teške izmene jednom razvijenog sistema
- krhkost sistema
- osetljivost na vreme odziva.

Dodatni problem sa mnogo online aplikacija bio je i nedostatak integralnosti. Svaka online aplikacija je razvijana po sopstvenoj specifikaciji, a svaka je nastajala u uslovima drugačijih specifičnih zahteva. Nije bilo jedinstvenog stava o tome šta predstavlja:

- korisnik
- proizvod
- transakcija
- isporučilac
- isporuka.

Zbog teško izvodljivih izmena aplikacija, nije se mogla vršiti reinterpretacija. rezultat je bio da i najveće i najsloženije korporacije nisu znale ko su njihovi korisnici. Korporacije su svakog meseca trošile gomilu novca na tehnologiju a da nisu imale suštinske informacije, npr. ko su im najbolji kupci, koji proizvodi se prodaju i koliki je prihod ostvaren u zadnjem tromesečju.

Uvođenje Extract datoteka

Prvi korak u rešavanju problema nedostatka suštinskih korporacijskih podataka bio je koncept *Extract datoteka*. Jedna aplikacija bi iz baze podataka kreirala extract datoteku i isporučila bi je drugoj aplikaciji, tako da je izgledalo da se podaci dele i da se tako kreiraju korporacijski podaci. Extract datoteke su postale vrlo popularne i brzo su se previše namnožile. Svaka nova extract dstoteka je pogoršavala probleme umrežavanja, uključujući:

Integritet podataka. Isti elementarni podatak se javlja na mnogo mesta. Na jednom mestu ima vrednost 25, na drugom 67, na trećem 135, a da niko ne zna koja vrednost je tačna.

Redundantnost podataka. Preveliko je ponavljanje istih podataka. Isti podaci se prosleđuju sa mesta na mesto, enormno se uvećava količina podataka u opticaju, opterećujući memorijske resurse i procesor.

Pravovremenost podataka. Dok se razmenjuju kroz sistem, podaci zastarevaju. U jednom danu, vrednost istog podatka može se 5 – 6 puta promeniti. Formiranje extract datoteka nije bilo u stanju da prati brzinu promene podataka.

Višestruko skladištenje podataka. Svako skladište podataka pravi se u vlastitom operativnom domenu, bez koordinacije i integracije sa drugim skladištima, što dovodi do kontradikcije i konflikata u odlučivanju među različitim poslovnim domenima korporacije.

Obrada extract datoteka zamrzava sistem koji je već pred kolapsom. Aplikacije sa online transakcijama se i inače teško menjaju u bilo kom pogledu, a nagomilavanje ekstrakcija oko njih ih dodatno konzervira.

Podaci su još nedostupniji. Obrada extract datoteka prenosi zahteve u pogledu koordinacije na okruženje, čime se izaziva nemogućnost dobijanja pouzdanih informacija, itd.

Od posebnog značaja je nedostatak istorijskih podataka. Online aplikacije određuju vrednosti tekućim podacima daju odgovor na pitanja kao što su: Koliki je saldo računa u banci ovog trenutka? Gde se isporuka robe nalazi sada? Kakav je status zahteva za osiguranje u ovom trenutku? Online aplikacije optimizuju "ovog trenutka" aspekt obrade informacija. Čim je podatak vremenski datiran, on prestaje da bude interesantan. Može se reći da gomile takvih podataka "začepljaju arterije" efikasne online obrade, zato online podaci i obrada zahtevaju što brže odbacivanje upotrebljenih podataka. Ipak, istorijski podaci imaju nesumnjivu vrednost, jer zahvaljujući njima poslovni sistem može da razlikuje "drveće i šumu", da počne da razume bazu svojih kupaca koji su bića sa navikama i običajima koje treba ispoštovati.

Zato što se nije posvećivalo dovoljno pažnje podacima integrisanim na nivou korporacije i istorijskim podacima, bilo je teško pristupiti podacima. I kad im se pristupi, nije im se moglo verovati, što je izazvalo frustraciju u pogledu sposobnosti nalaženja i obrade pouzdanih podataka. Frustracija krajnjih korisnika zbog podataka zaključanih u mrežno okruženje rezultovala je shvatanjem da je to neka druga vrsta podataka. Postojala je suštinska razlika između operacionih i informacionih podataka. Tabela 1.1 pokazuje tu razliku.

Tabela 1.1: Karakteristike sistema zasnovanih na operacijama i na informacijama

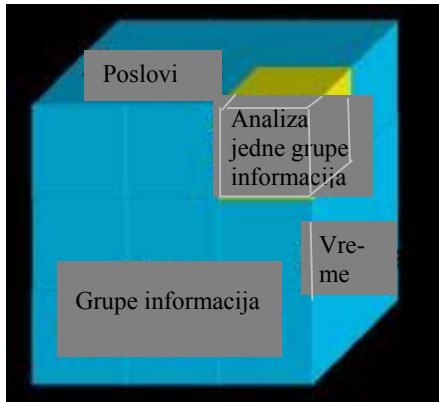
Operacioni (operativni)	Informativni / DSS
Detaljni	Sumarni
Mogu se menjati	Trenutni snimci; nisu dozvoljene izmene
Tačni samo u trenutku nastanka	Svaki zapis sadrži trenutak formiranja
Koriste ih izvršioci	Koriste ih menadžeri
Izgrađeni na osnovu zahteva	Izgrađeni bez poznavanja zahteva
Podržavaju male uniformne transakcije	Podržavaju mešovite složene poslove
Vreme odziva 2 – 3 sekunde	Vreme odziva do 24 h
Podaci dizajnirani za optimalno korišćenje memorije	Podaci dizajnirani za optimalni pristup
Trenutni tekući podaci	Prevashodno istorijski podaci
Podaci orientisani prema aplikaciji	Integrисани podaci
Podaci prema funkcionalnoj upotrebi	Podaci za specifična područja primene
Zahteva se referencijalni integritet	Ne zahteva se referencijalni integritet
Visoka raspoloživost je normalna	Visoka raspoloživost je dobra ako je moguća

Suštinska je razlika između *operativnih informacija* i *informativnih informacija*. Operativne informacije služe za podršku dnevnog odvijanja poslova. Informativne informacije su poznate kao informacije koje koriste *sistemi za podršku odlučivanju* (*Decision Support Systems* - DSS).

Osnova za DSS obradu postala je:

- online analitička obrada (Online Analytical Processing - OLAP)
- "rudarenje podataka" (Data Mining)
- koncept skladišta podataka (Data Warehousing), u kome je skladište podataka fizički odvojena instanca u odnosu na online operativne aplikacije.

Online analitička obrada (OLAP)



Slika 1.1 OLAP kocka

Online analitička obrada (*Online Analytical Processing - OLAP*) nastaje kao pristup kojim se dobija odgovor na multidimenzionalne analitičke upite u podatke. Termin je izведен iz pojma **online obrada transakcija** (*Online Transaction Processing – OLTP*), vezanog za rad sa tradicionalnom bazom podataka.

OLAP alati omogućavaju interaktivno analiziranje multidimenzionalnih podataka iz različitih perspektiva. Sastoje se od tri osnovne komponente:

- **konsolidacija** (agregiranje podataka koji se mogu kumulirati i obrađivati u jednoj ili više dimenzija, npr. zbrajanje podataka iz svih prodajnih punktova da bi se istraživali trendovi prodaje)
- **drill-down** ("svrdlo") tehnika (omogućava navigaciju od zbirnih podataka do najsitnijih detalja, npr. raščlanjavanje ukupnih podataka o prodaji na prodaju po proizvodima u nekom regionu)
- **slicing and dicing** ("rezanje sloj po sloj" i "rezanje na kockice") tehnika (*slicing* – izvlačenje iz višedimenzionalne baze 1 sloja tj. skupa podataka, npr na slici 1.1 sve grupe informacija i svi poslovi u 1 trenutku vremena; *dicing* – podskup podataka po svim dimenzijama, npr. na slici 1.1 neke grupe informacija za neke poslove u nekom intervalu vremena, čime se slojevi kombinuju i vide iz različitih perspektiva).

Jezgro OLAP sistema je multidimenzionalni model podataka koji se može predstaviti u vidu **OLAP kocke** odn. multidimenzionalne ili hiper kocke (slika 1.1). OLAP kocku čine numerički podaci (činjenice) koje se nazivaju **mere**, kategorisane u **dimenzije**. Mere se izvode iz činjenica u tabeli činjenica sa stranim ključevima a dimenzije se izvode iz tabela dimenzija (sadrži primarne ključeve i ostale attribute). Jedna tabela činjenica obično pokazuje na više tabela dimenzija, po hijerarhijskom principu.

OLAP sistem se javlja u više varijanti, od kojih će ovde biti pomente najviše korišćene.

ROLAP (Relational Online Analytical Processing) – Relaciona online analitička obrada

ROLAP alati pristupaju podacima u relacionoj bazi podataka i generišu SQL upite da bi obradili informacije na željenom nivou, onda kada je to korisniku potrebno. Kod njih nema faze prethodne obrade i dodatnog memorisanja podataka kao u drugim (npr. MOLAP) varijantama. Sa ROLAP alatima mogu se kreirati dodatne tabele u bazi podataka (najčešće sumarne tabele odn. agregacije) koje sumarizuju podatke za bilo koju željenu kombinaciju dimenzija.

Prednosti ROLAP su:

- pogodan je za obradu velikih količina podataka, posebno tamo gde postoji visok nivo kardinalnosti (jednom nastupajućem elementu nadređenog entiteta odgovaraju stotine hiljada nastupajućih elemenata podređenog entiteta)
- zbog velikog broja alata za ekstrakciju podataka i sposobnosti preciznog prilagođavanja ETL koda konkretnom modelu podataka, vreme ekstrakcije je znatno kraće nego u drugim OLAP varijantama
- podaci su u standardnoj relacionoj bazi podataka i može im se pristupiti bilo kojim SQL reporting alatom (alat za pristup ne mora biti OLAP alat)
- pogodan je za podatke koji se ne mogu agregirati (npr. za tekstualne opise)
- razdvajanjem memorijskog prostora za podatke od multidimenzionalnog modela mogu se uspešno modelovati podaci koji se inače teško uklapaju u striktno dimenzioni model
- ROLAP može da pojača kontrolu autorizacije baze podataka kao što je sigurnost nivoa reda tabela, uz filtriranje rezultata upita prema unapred postavljenim kriterijuma (SQL uslov WHERE)

Nedostaci su:

- generalno, sporija obrada nego kod drugih varijanti
- formiranje agregiranih tabela mora se raditi sopstvenim ETL kodom; ROLAP alati tu ne pomažu, što znači dodatno vreme za razvoj ETL programa i više koda
- ako se preskoči korak kreiranja agregiranih tabela, performanse (brzina odvijanja) upita opadaju, jer se moraju čitati velike tabele detaljnih podataka; to se delimično ublažava dodavanjem agregiranih tabela unapred u model podataka, ali nije praktično da se to radi za sve moguće kombinacije dimenzija / atributa
- ROLAP se naslanja na bazu podataka opšte namene i nisu moguće neke tehnike koje npr. MOLAP nudi, npr. hijerarhijsko indeksiranje; ipak, novi SQL operatori kao CUBE ili ROLLUP, DB2 Cube Views itd., ublažavaju ovaj nedostatak
- s obzirom da su SQL alati oslonjeni na SQL za sve obrade, oni nisu podesni onda kad ima dosta kalkulacija koje se teže izvode u SQL-u, npr. budžetiranje, alokacija, finansijsko izveštavanje i sl.

Proizvodi bazirani na ROLAP:

- Microsoft Analysis Services,
- MicroStrategy,
- Oracle Business Intelligence Suite Enterprise Edition (ranije Siebel Analytics)
- Tableau Software
- Mondrian (open source ROLAP server).

MOLAP (Multidimensional Online Analytical Processing)

MOLAP podržava kao i ROLAP multidimenzionalni model podataka. Od ROLAP tehnike se bitno razlikuje, jer zahteva prethodnu obradu i memorije podatke u formi OLAP kocke. Većina MOLAP proizvoda podatke smešta u memoriju u vidu optimizovanog multidimenzionalnog niza, a ne u relacionu bazu podataka.

Prednosti MOLAP su:

- visoke performanse (brzina) upita, zahvaljujući optimizovanoj memoriji, multidimenzionalnom indeksiranju i keširanju
- manji potreban prostor na disku u odnosu na podatke u relacionoj bazi podataka, zahvaljujući tehnikama kompresije
- automatsko izračunavanje višeg nivoa agregacija podataka
- posebno je podesan za skupove podataka manjih dimenzija
- modeli nizova obezbeđuju prirodno indeksiranje
- efikasna ekstrakcija podataka kroz prestrukturisanje agregiranih podataka.

Nedostaci su:

- u nekim MOLAP rešenjima korak u kome se izvodi obrada (Load faza ETL procesa) može dugo da traje, posebno za veliki obim podataka; uobičajena pomoć u takvim situacijama je inkrementalna obrada, tj. obrada samo onih podataka koji su se promenili (obično novih podataka) umesto obrade svih podataka
- MOLAP alati pokazuju pad performansi upita na modelima podataka sa visokom kardinalnošću
- neki MOLAP proizvodi otežano vrše izmene i spori su u upitima nad modelima sa više od deset dimenzija; to ograničenje je različito izraženo, zavisno od kompleksnosti i kardinalnosti dimenzija, kao i od broja memorisanih činjenica ili mera; napredniji MOLAP proizvodi bez problema obrađuju na stotine dimenzija
- neke MOLAP metodologije uvode redundantne podatke.

Proizvodi bazirani na MOLAP:

- Cognos Powerplay,
- Oracle Database OLAP Option,
- Microsoft Analysis Services,
- Essbase,
- TM1,
- Lilith Hicare
- Daptech Keystone
- Palo (open source MOLAP server).

HOLAP (Hybrid Online Analytical Processing)

Predstavlja kombinaciju ROLAP I MOLAP implementacija OLAP koncepta. Dozvoljava da se jedan deo podataka smesti u MOLAP, a drugi deo u ROLAP memoriju, čime do izražaja dolaze prednosti obe OLAP tehnologije.- veća skalabilnost ROLAP i brža obrada MOLAP tehnologije. Npr. HOLAP server može da veliki obim podataka drži u relacionoj bazi dok agregacije smešta u odvojenu MOLAP memoriju. Nivo kontrole koji dizajner ima nad OLAP kockom varira od jednog do drugog HOLAP proizvoda.

Vertikalno particioniranje

U ovom modu HOLAP agregacije memoriše u MOLAP tehnologiji, sa siljem poboljšanja performansi, dok detaljne podatke drži u ROLAP tehnologiji, kako bi se optimizovala obrada kocke.

Horizontalno particioniranje

U ovom modu HOLAP memorije u MOLAP tehnologiji sloj podataka dobijen slicing operacijom, obično najnoviji (tj. dobijen operacijom slicing u funkciji vremena – podaci u datom trenutku) da bi upiti brže radili, a ostale podatke koristeći ROLAP, čime uvažava činjenicu da u velikoj OLAP kocki može biti "gustih" (sa mnogo podataka) i "retkih" regija (oskudnih u podacima).

Proizvodi bazirani na HOLAP:

- Holos firme Crystal Decisions (prvi proizvod koji je koristio HOLAP memoriju)
- Microsoft Analysis Services,
- Oracle Database OLAP Option,
- MicroStrategy
- SAP AG BI Accelerator
- Microsoft SQL Server 7.0 OLAP Services (podržava hibridni OLAP server).

"Rudarenje podataka" (Data Mining)

Predstavlja jedan od osnova poslovne inteligencije, jer predstavlja korak analize u procesu **otkrivanja znanja** zamemorisanog u bazi podataka (Knowledge Discovery in Database – KDD). Opšti cilj Data Mining-a je da, povezivanjem podataka u okviru baze, iz nje izvuče znanje o poslovnom sistemu i pretvori ga u strukturu informacija koju čovek lako razume i interpretira. Omogućava bolje upravljanje podacima, pripremu i modelovanje podataka, analizu veza među podacima, analizu njihove upotrebljivosti i kompleksnosti, vizuelizaciju i online izmene.

Data Mining predstavlja:

- ekstrakciju implicitnih, prethodno nepoznatih a potencijalno korisnih informacija iz podataka
- nauku o ekstrahovanju korisnih informacija iz velikog obima podataka ili baza podataka.

Ova tehnika omogućuje korisniku da pronalazi podatke polazeći od generalnih informacija, a zatim detaljnijih, dolazeći do nivoa koji mu je potreban.

Data Mining komponenta poslovne inteligencije je prevashodno namenjena menadžerima najvišeg nivoa kojima su potrebne informacije nestruktuirane prirode. Data mining se nalazi na vrhu hijerarhije organizacije poslovne inteligencije i njime se omogućava dobijanje krajnjih podataka.

Cilj Data Mininga je identifikovati vredne nove, potencijalno korisne veze i uzorce u postojećim podacima (Jackson, 2002)

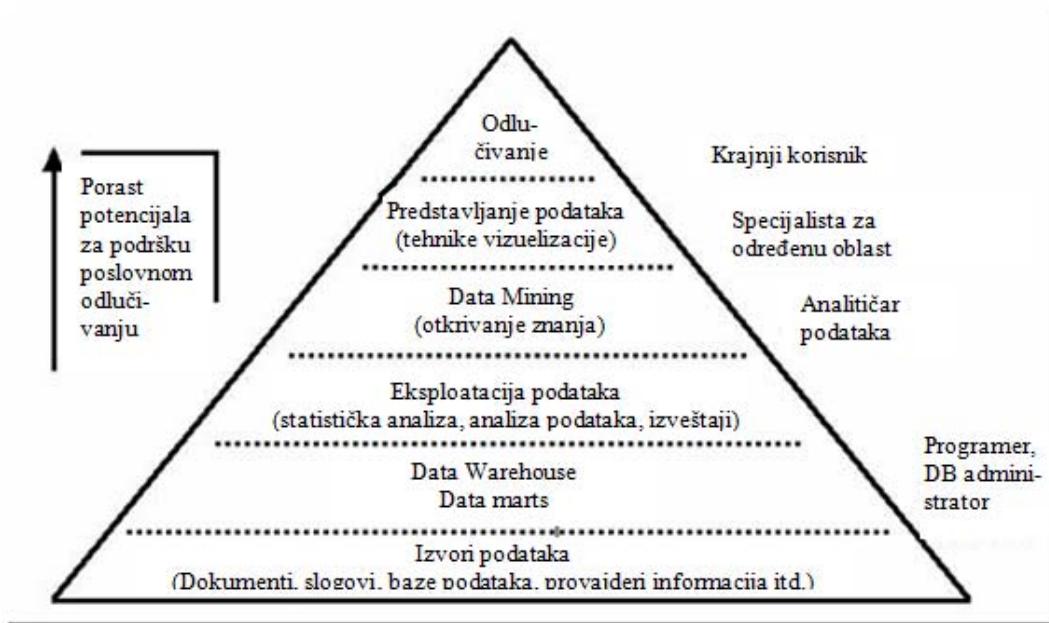
Data Mining koristi sledeće tehnike:

- Neuronske mreže (Neural Networks)
- Zaključivanje na osnovu slučajeva (Case-Based Reasoning – CBR)

- Genetički algoritmi (Genetic Algorithms)
- Stabla odlučivanja (Decision Trees)
- Asocijacijska pravila (Association Rules)
- Statističke metode: deskriptivne i vizuelizacione tehnike, klaster analize, korelacione analize, diskriminantnu analizu, faktorsku analizu, regresionu analizu, logističku regresiju i dr.)

Alati za Data Mining:

- Statistički softverski paketi (npr. SAS, Statistika, SPSS, i dr.)
- Matematički softverski paketi (npr. MathLab, Mathematica)
- Alati uključeni u skladištenje podataka (OLAP) ili sistem za upravljanje bazom podataka (npr. Microsoft SQL Server Business Intelligence – uključuje i Enterprise Miner)
- Specijalizovani alati za opšte ili poslovne primene (npr. DataMiner, IntelliMiner, i sl.)



Slika 1.2 Evolucija od podataka do znanja

Podaci se nalaze na najnižem nivou hijerarhijske strukture organizacije poslovne inteligencije. Na ovom nivou oni nemaju konkretno značenje za donosioca odluke. Ekstrahovanjem i integracijom oni prelaze u Data Warehouse, pri čemu postaju grupisani po nekom osnovu.

Podaci se mogu analizirati statističkim analizama, izveštajima, ali njihovo istraživanje putem Data Mining metoda predstavlja sofisticiraniju tehniku na višem nivou. Podaci sada imaju određeno značenje za donosioca odluke i oni postaju INFORMACIJE za njega.

Ovakve informacije se predstavljaju različitim tehnikama vizualizacije, a koje mogu biti i grafički prikazi. Ovim se dobijaju zbirni izveštaji na osnovu kojih se donosi odluka. Informacije na taj način evoluiraju u ZNANJE.

Skladište podataka

Od početka preorientacije na koncept Data Warehousinga, Data Warehouse (skladište podataka) se definiše tako da je:

Predmetno orijentisano. Podaci se organizuju oko ključnog objekta ili procesa u organizaciji. Klasični primeri su predmetno orijentisane baze podataka za kupca, materijal, dobavljača i transakcije.

Integrисано. Podaci iz različitih predmetnih područja treba da se međusobno dopunjaju i racionalizuju.

Trajно. Podaci u Data Warehousu se ne menjaju. Kad se zapis smesti u Data Warehouse, više se ne menja. To je razlika u odnosu na slog podataka u online okruženju, koji se vrlo često menja.

Promenljivo u vremenu. Zapis je aktuelan samo u nekom trenutku kada je ili periodu vremena za koji je kreiran. Posle toga vrednosti podataka nisu ažurne.

Kreirano za potrebe menadžerskog odlučivanja.

Prethodna definicija ostala je nepromenjena od uvođenja skladišta podataka. Dodatno, Data Warehouse obezbeđuje:

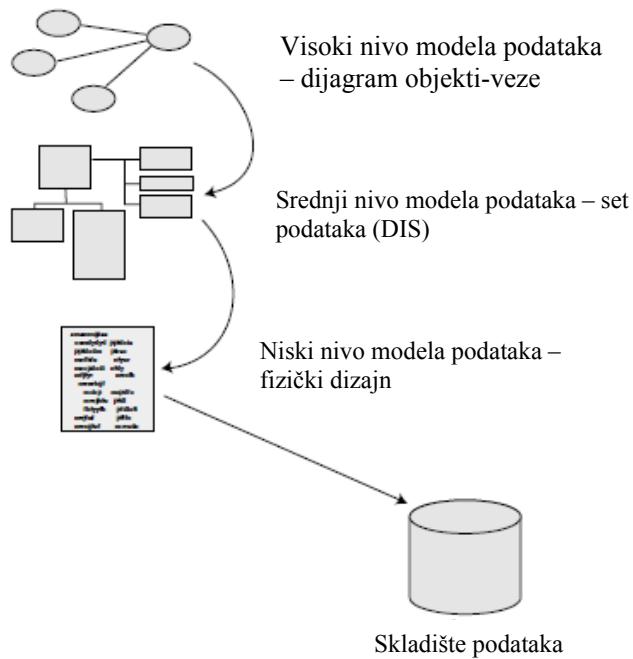
- detaljne ili granularne podatke
- integrisane podatke
- istorijske podatke
- podatke kojima se lako pristupa.

Data Warehouse se nalazi u centru okruženja za poslovnu inteligenciju. Ono (skladište podataka) predstavlja jedini izvor istinitih podataka za korporaciju i sadrži podatke na granularnom nivou. Dodatno, sadrži i veliki obim istorijskih podataka. Podržava proces prenosa podataka sa izvornih sistema, transformacije i filtriranja podataka tako da se mogu memorisati u integralnom modelu podataka na atomskom nivou granularnosti.

Mnogi faktori utiču na dizajn skladišta podataka i strukturu u kojoj se podaci memorišu.

Model podataka

Dizajn skladišta podataka počinje sa modelom podataka. Na najvišem nivou, model podataka se može predstaviti dijagramom objekti-veze (*entity relationship diagram – ERD*). Dijagram objekti-veze predstavlja apstrakciju granularnih podataka koji se nalaze u skladištu. Za potrebe dizajna skladišta podataka model objekti-veze sadrži samo granularne podatke, ne i izvedene, čime se znatno ograničava veličina i kompleksnost modela podataka. On se sastoji od entiteta i relacija. Svaki entitet predstavlja neko važno predmetno područje u korporaciji (npr. kupac, proizvod, transakcija...). Dalje, svaki entitet se na nižem nivou modelovanja podataka definiše kao set podataka (*data item set – DIS*), koji uključuje ključeve i attribute kao i njihovu event. strukturu. Sledeći niži nivo dizajna je fizički dizajn koji razbija setove podataka definišući fizičke karakteristike podataka (dužina, format...). Slika 1.3 ilustruje taj proces.



Slika 1.3 Skladište podataka se dizajnira iz modela podataka

Po završenoj fazi fizičkog dizajna, skladište podataka je specificirano i definisano tako da ga sistem za upravljanje bazom podataka prepoznaće. Definisani su i ostali aspekti fizičkog dizajna – partitioniranje, punjenje, indeksiranje, memorijski medijum i označavanje vremena nastanka.

Različite fizičke tabele

Skladište podataka se pravi od povezanih tabela ili fizičkih baza podataka. U skladištu su različite tabele koje predstavljaju različita predmetna područja ili njihove podskupove. Jedna tabela se povezuje sa drugom preko deljivog ili stranog ključa. Skladište najčešće obuhvata pet predmetnih područja:

- kupac
- proizvod
- isporuka
- snabdevač
- porudžbina.

Svako predmetno područje ima posebnu fizičku tabelu ili bazu podataka. Zajedno, te tabele sa svojim relacijama formiraju skladište podataka.

Integracija i transformacija

Jedan od najvažnijih i najtežih aspekata razvoja i implementacije skladišta podataka je prebacivanje i konverzija podataka iz operativnog izvornog okruženja. Procenjuje se da se za prvu iteraciju razvoja zahteva najmanje 75% ukupnog utroška resursa. Ekstrakcijom se podaci izvlače iz operativnog okruženja i prebacuju u okruženje skladišta podataka. Podaci

se izvlače iz različitih izvora, npr. iz sistema za unos porudžbina, sistema za upravljanje proizvodnjom, iz obračuna plata itd. Pri izvlačenju se vrši i transformacija podataka, npr.:

- konvertovanje podataka u zajednički (opšti) format
- reformatiranje podataka
- ponovno dodeljivanje kodiranih vrednosti
- restrukturisanje podataka
- dodeljivanje default vrednosti
- zbrajanje
- promena redosleda
- konverzija ključeva
- konverzija iz jednog sistema za upravljanje bazom podataka u drugi
- konverzija iz jednog operativnog sistema u drugi
- konverzija iz jedne hardverske arhitekture u drugu
- spajanje različitih tipova slogova
- kreiranje meta podataka koji opisuju aktivnosti konverzije
- editovanje podataka
- dodavanje oznake vremena nastanka.

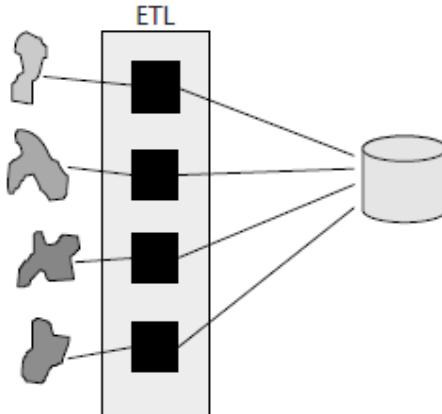
U ranim danima Data Warehousinga nije bilo drugog načina za kreiranje interfejsa između operativnog okruženja i skladišta podataka sem pisanja programa. Sa razvojem tehnologije javlja se **ETL (extract/transform/load) softver** koji automatski kreira interfejs potrebne za prenos podataka u svet skladišta podataka. Slika 1.4 pokazuje mesto ETL softvera između izvornih sistema i skladišta podataka.

Jedna od stvarnih prednosti ETL obrade je to što podaci ulaze u ETL proces u modu aplikacije a izlaze iz njega u integrисаном корпорацијском моду.

Programski alati za ETL vrše:

- **ekstrakciju** (Extract) – omogućavaju korisniku izbor relevantnih podataka iz baze podataka koji će se smestiti u skladište (koji podaci o klijentima, računima, proizvodima, narudžbama, prodaji, zalihami itd.)
- **transformaciju** (Transform) – prenos podataka u format koji odgovara skladištu (reformatiranje, usklađivanje i čišćenje podataka iz baze) – oduzimaju 80% vremena celog ETL procesa
- **punjjenje** (Load) skladišta – incijalno punjenje, punjenje istorijskih podataka i inkrementalno punjenje.

Kad se završi incijalno punjenje i punjenje istorijskih podataka, programi za inkrementalno punjenje se pokreću periodično (dnevno ili nedeljno ili mesečno) i stalno pune skladište podataka novim relevantnim podacima iz baze.



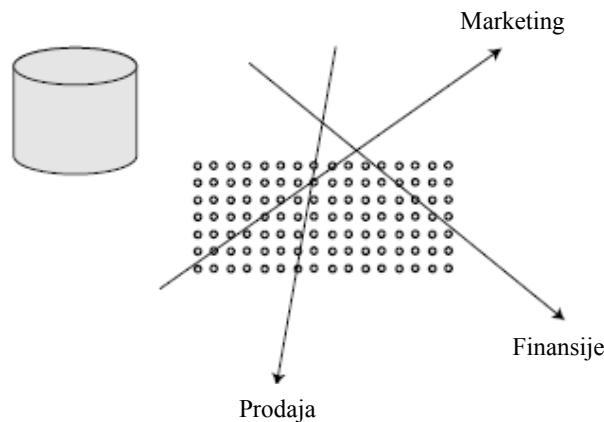
Slika 1.4 ETL obradom se vrši konverzija i integracija podataka

Granularni podaci

Podaci koji se nalaze u skladištu podataka su vrlo granularni. To znači da su podaci u skladištu na vrlo niskom nivou detaljnosti. Oni se mogu nekom aplikacijom preoblikovati tako da se vide na drugačiji način.

Ponekad nazivani i "atomski podaci korporacije", granularni podaci čine "jedinu verziju istine" koja je osnova usaglašavanja i usklađivanja za obradu informacija.

Posedovanje granularnih podataka u jezgru skladišta podataka pruža mnoge prednosti. Primarna korist je to što isti podaci mogu da se vide na različite načine. Na slici 1.5 vidi se da funkcija marketinga vidi podatke na jedan način, funkcija prodaje na drugi, a finansijska funkcija na treći način, iako sve tri funkcije imaju isti izvor usaglašavanja.

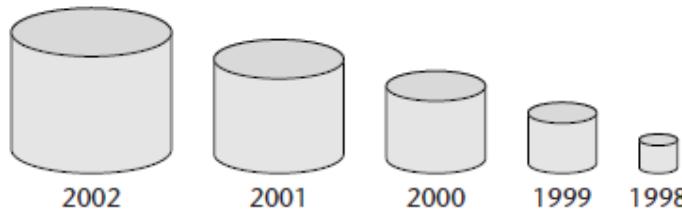


Slika 1.5 Granularni podaci omogućavaju da se isti podaci vide na različite načine

Istorijski podaci

Jedna od najvažnijih osobina skladišta podataka je da sadrži veliki obim istorijskih podataka. Na slici 1.6 prikazano je skladište koje sadrži petogodišnji istorijat, što je tipično. Neka skladišta mogu da sadrže više ili manje istorijskih podataka, zavisno od poslovnih potreba korporacije.

Iako istorijski podaci imaju mnogostruku primenu, verovatno je najkorisnija mogućnost pretraživanja unazad u vremenu i pravljenje "šta-ako" analiza. Takav pristup nam pruža uvide i saznanja koji se na drugi način ne mogu dobiti.



Slika 1.6 Skladište podataka sadrži veliki obim istorijskih podataka

Vremensko označavanje

Jedinice podataka smeštenih u skladište podataka imaju vremensku oznaku, tako da svaka jedinica podataka u skladištu ima neki element vremena dodeljen slogu podataka. Vremensko označavanje skladišta podataka vezuje podatke za vreme nastanka.

Uopšte, dva su načina na koje se slog smešta u skladište podataka: diskretno ili kontinualno (slika 1.7). U diskretnom načinu je data jedna instanca vremena u kojoj je slog važeći. U kontinualnom načinu daje se vremenski period u kome važi slog. Ovako označeni slogovi formiraju širu definiciju informacije u vremenu. Diskretni slogovi se obično koriste za veliki broj promenljivih koje se često menjaju. Kontinualne vremenske oznake koriste se za mali broj promenljivih koje se menjaju sporo i kod kojih se traži dobijanje informacija u nekom periodu.

Slogovi u skladištu podataka imaju karakterističnu strukturu koja obuhvata:

- vremensku oznaku
- ključ
- primarne podatke
- sekundarne podatke.

Relacije među podacima

Različiti tipovi podataka koji se nalaze u skladištu uspostavljaju međusobne relacije preko stranih ključeva koji pokazuju na primarne ključeve. Na primer, kupac ABC ispostavlja porudžbinu. Formira se slog za kupca ABC kao i poseban slog za porudžbinu. Slog porudžbine ima strani ključ koji pokazuje na kupca ABC.

Relacije među podacima u skladištu podataka su specifične zato što su vremenski ograničene. Kada se u skladištu uspostavi relacija, ona je važeća samo za moment vremena indiciran vremenskim oznakama u slogovima koji učestvuju u relaciji. Ova interpretacija relacije je znatno drugačija od referencijalnog integriteta s kojim se susrećemo u online okruženju.

diskretni slogovi podataka	kontinualni slogovi podataka
[] June 14	[] Jan 1 to Jan 13
[] June 23	[] Jan 14 to Feb 8
[] June 30	[] Feb 9 to Feb 23
[] July 3	[] Feb 24 to Mar 12
[] July 3	[] Mar 13 to Jul 20
[] July 6	[] Jul 21 to Dec 17

Slika 1.7 Slogovi sa vremenskim oznakama su diskretni ili kontinualni

Generički ili specifični podaci

Pitanje koje se javlja u dizajnu svakog skladišta podataka je kako obezbediti u isto vreme generičke i specifične podatke. Generički podaci obuhvataju sve instance predmetnog područja, dok se specifični podaci odnose samo na neka pojavljivanja predmetnog područja.

Generička baza podataka memoriše npr. informacije o kupcima zajedno sa tabelama sa kojima je tabela kupaca u relaciji, uključujući npr. kupce veleprodaje, kupce u Evropi, kupce po dugoročnim ugovorima i prioritetne kupce. Svaka od spoljašnjih tabela sadrži informacije specifične samo za klasu tabela koja zadovoljava kriterijume. Npr., prioritetni kupac veleprodaje ima podatke u generičkoj tabeli kupaca, u tabeli prioritetnih kupaca i u tabeli kupaca veleprodaje.

Na taj način se podaci različitog tipa mogu efikasno predstaviti u skladištu podataka.

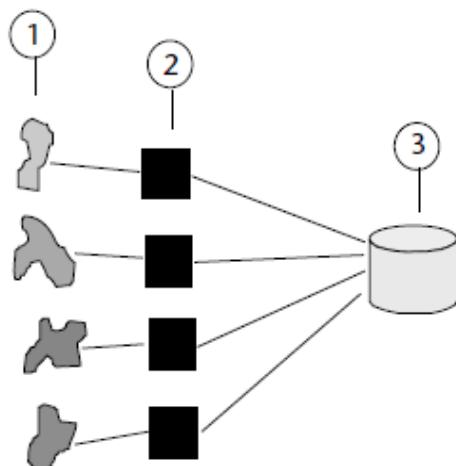
Kvalitet podataka

Kvalitet podataka je važno pitanje za okruženje sa skladištem podataka.

Kao što se vidi na slici 1.8, kvalitet podataka dolazi do izražaja na 3 mesta:

1. na mestu unosa podataka u operativnom okruženju
2. na mestu ETL obrade
3. pri smeštanju podataka u skladište.

Unos sirovih podataka vrši se u operativnom okruženju. Insistira se na brzom unosu, nema mnogo vremena za proveru kvaliteta, pa se ona sprovodi na niskom nivou.



Slika 1.8 Tri mesta na kojima kvalitet podataka dolazi do izražaja

Najveći deo aktivnosti vezanih za kvalitet podataka vrši se u fazi izvođenja ETL operacija. ETL ne zahteva da se manipuliše starim aplikacijama ni da se one na bilo koji način menjaju. Podaci koji dolaze iz operativnih aplikacija mogu se izolovati. Takođe, mogu se integrisati podaci koji dolaze iz različitih aplikacija. Podaci su svakako u prolazu, tako da je ETL proces idealno mesto za ispitivanje i kontrolu podataka, kao i vršenje promena po potrebi.

Treće mesto na kome se javlja pitanje kvaliteta podataka je na ulasku podataka u skladište. Podaci se u vremenu menjaju, tako da ono što je bilo pouzdano i odgovarajuće npr. u

jednoj godini to ne mora biti i u sledećoj godini. I u slučajevima kada su podaci idealno uneti u skladište potrebno je periodično prilagođavanje podataka promenama uslova poslovanja koje se dešavaju u vremenu.

Količine podataka

Količine podataka u skladištu rastu preko svih očekivanja. Nekada su terabajtna skladišta podataka bila nezamisliva; danas su ona realnost. Sa naglim rastom količina podataka menjaju se i tehnike i pristupi upravljanju podacima. Jedna od najvažnijih karakteristika rasta skladišta je pojava latentnih, neupotrebljenih podataka, koji samo zauzimaju prostor i izazivaju troškove. Dok su skladišta bila mala, korišćeni su svi ili skoro svi podaci u njima. Sa naglim rastom veličine skladišta podataka, rastuće količine podataka leže u skladištu neiskorišćene.

Ako je skladište veličine oko 100 GB, latentni podaci čine 10 – 20% njegovog sadržaja, međutim, kad se dostigne terabajtna veličina skladišta, učešće latentnih podataka može porasti i na 50 – 75%. Neracionalno je ulaganje u veličinu skladišta ako se ne deluje u pravcu smanjenja učešća latentnih podataka. Latentni podaci degradiraju i performanse rad sa skladištem podataka.

Uklanjanje latentnih podataka

Latentne podatke treba periodično skloniti sa diska i arhivirati na drugim medijima. Aktivni podaci su na disku i obrađuju se na normalni način, dok se neaktivni podaci fizički sklanjaju na alternativnu odn, priručnu memoriju. Time se znatno smanjuju troškovi skladišta podataka i povećava brzina pristupa podacima. Ako se želi pristup i aktivnim i izdvojenim latentnim podacima, može se koristiti tehnologija *cross-media storage manager* (CMSM) koja predstavlja interfejs između memorije na disku (sa aktivnim podacima) i alternativne memorije, sa zadatkom da upravlja saobraćajem između dve memorije tako da korisnik ima jedinstven pogled na podatke u skladištu.

Meta podaci (Metadata)

Meta podaci su jedan od najvažnijih aspekata skladišta podataka. Meta podaci su informacije o sadržaju na kome se zasniva koncept *korporacijske fabrike informacija* (*corporate information factory* – CIF). Svaka aplikacija ima svoje meta podatke koji se distribuiraju kroz celu arhitekturu sistema. Meta podaci imaju dve funkcije: da opišu podatke (tip podataka, opisi atributa, opisi domena, nazive, veličinu i dozvoljene vrednosti) koji se nalaze u nekoj komponenti arhitekture i da se razmenjuju sa drugim komponentama. Oni integrišu dolazeće podatke, služe za redefinisanje i ažuriranje delova skladišta podataka.

Meta podaci u skladištima podataka igraju više uloga. Jedna uloga opisuje koji podaci se gde nalaze za normalnu upotrebu. Meta podaci deluju i kao koordinator među različitim servisima od ETL do pristupa informacijama. Različiti servisi u arhitekturi imaju vrlo različite osnove i funkcije. Neki rade pod sistemom za upravljanje bazom podataka, neki pod drugim sistemom za upravljanje bazom podataka. Neki servisi rade pod jednim tipom multidimenzione tehnologije, neki pod drugim. Funkcije servisa se znatno razlikuju. Da bi servisi radili usklađeno, mora među njima postojati koordinacija. Koordinacija se postiže meta podacima koji se predaju od jednog sloja arhitekture ka drugim.

Razlikuje se više vrsta meta podataka, kao što su tehnički meta podaci, operativni meta podaci i poslovni meta podaci. **Tehnički meta podaci** opisuju strukturu i sadržaj različitih tipova podataka. Skladište se u okviru rečnika podataka i repozitorijuma na duži rok.

Operativni meta podaci stvaraju u svakodnevnom radu skladišta podataka. Informacije o slogovima prenetih od jedne do druge softverske komponente, dužini rada programa, broju slogova u bazi podataka itd. čine operativne meta podatke. **Poslovni meta podaci** su u obliku razumljivom za učesnike u poslovanju. Poslovne definicije, poslovne formule i uopšte uslovi poslovanja čine poslovne meta podatke. Sva tri tipa meta podataka potrebna su za upravljanje radom skladišta podataka. Važan je aspekt integriteta meta podataka. Da bi se održavala kontrola ispravnosti meta podataka koji se prenose između mnogih različitih komponenti, zahteva se određeni protokol. Da bi se očuvao integritet meta podataka u distribuiranom okruženju, svaka jedinica meta podataka mora biti jedinstvena i mora imati vlasnika. Jedini vlasnik jedinice meta podataka je osoba ili organizacija koji ima pravo da je kreira, modifikuje ili briše. Svako drugi može da deli meta podatke bez izmene. Pri prenosu meta podataka sa jednog na drugi čvor u mreži mora se pažljivo pratiti vlasništvo nad njima.

Karakteristike meta podataka:

- predmetna orijentisanost (zasnivaju se na apstrakciji realnih entiteta, npr. projekat, kupac, proizvod,...)
- definišu način na koji će se transformisani podaci interpretirati (npr. 11/04/12 kao 11. april 2012.)
- pružaju informacije o srodnim podacima u skladištu podataka
- predviđaju vreme odziva, prikazujući broj slogova koji treba obraditi u upitu
- čuvaju izračunate vrednosti i formule po kojima se vrši izračunavanje, kako bi se izbegla pogrešna interpretacija.

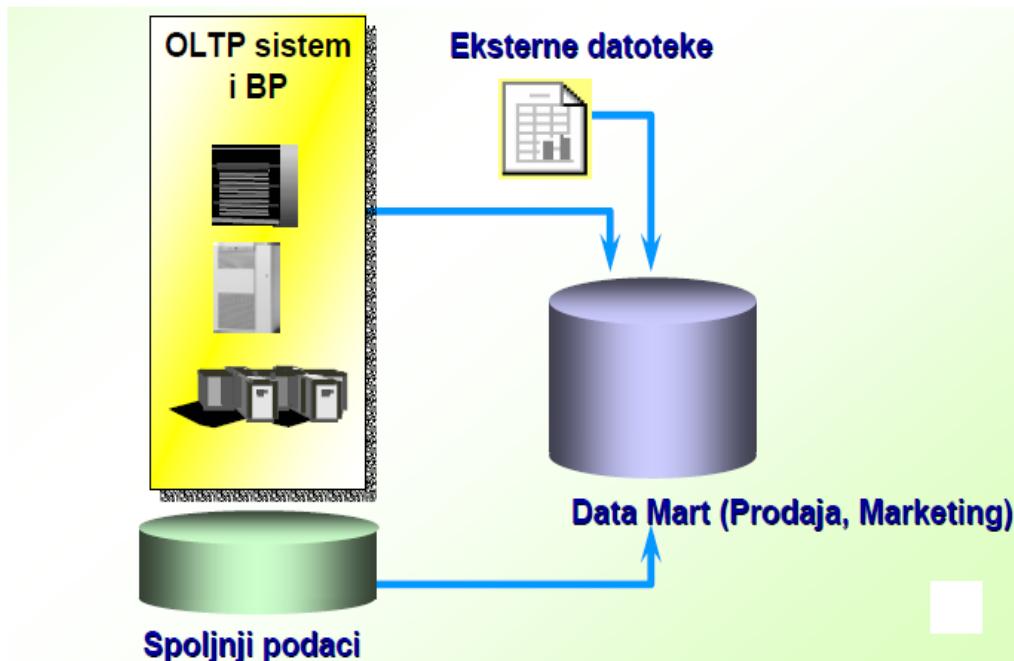
Sa stanovišta administratora skladišta podataka, meta podaci su skladište podataka i dokumentacija o sadržaju i procesima koji se odvijaju u skladištu podataka. Korisnik vidi meta podatke kao mapu koja omogućava kretanje kroz informacije. Kao takvi, meta podaci su vezivno tkivo skladišta podataka i sistema poslovne inteligencije.

Data mart ("prodavnica podataka")

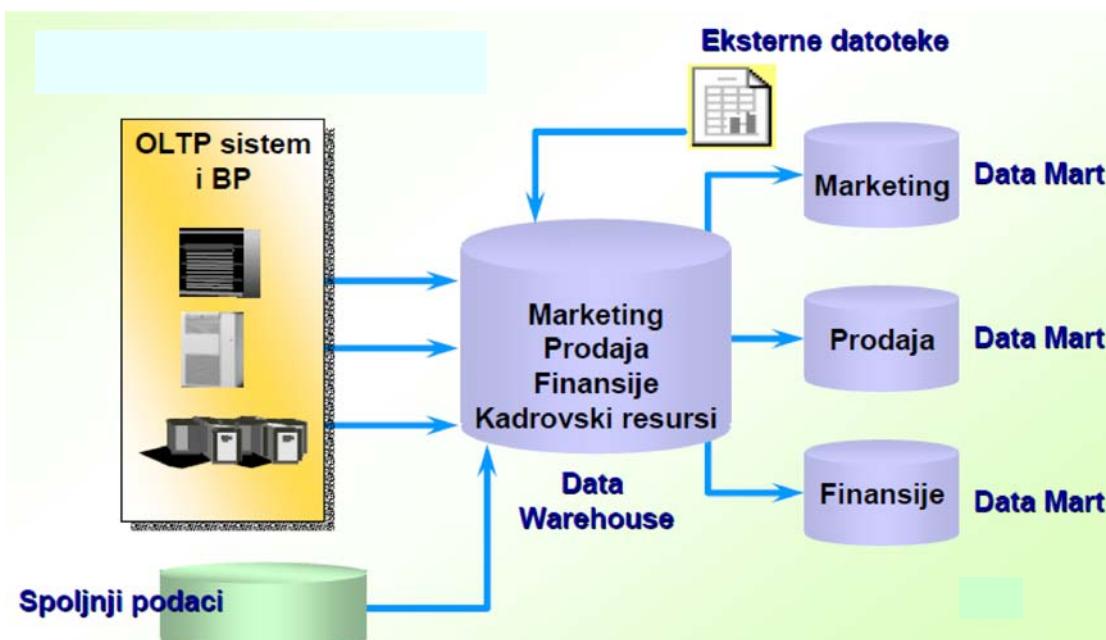
Data mart se može definisati kao pristupni sloj skladišta podataka koji se koristi za isporuku podataka korisnicima (u njemu se odvija najviše analitičkih aktivnosti u BI okruženju). Takođe, to je podskup skladišta podataka koji je obično kreiran i orijentisan ka konkretnoj poslovnoj liniji (oblasti) ili timu. U istom preduzeću može postojati više data martova namenjenih i dodeljenih jednoj ili većem broju poslovnih jedinica koje im pristupaju, uz mogućnost izmene podataka u "svojim" data martovima (ne i u "tuđim" data martovima i u skladištu podataka). Najčešći oblik data marta je multidimenzionalan, što omogućava lak pristup, brzu i kvalitetnu analizu podataka. Problem koji se može pojaviti u organizaciji koja je implementirala nekoliko data martova pre implementacije celovitog skladišta podataka je njihova integracija u jedinstven sistem. Zahteva se uravnotežavanje težnje da se data martovi kreiraju kao odvojeni i potrebe za uspešnim funkcionisanjem skladišta na globalnom korporacijskom nivou.

Data martovi mogu biti **nezavisni** (izolovani od drugih DW sistema) ili **zavisni** (oslonjeni na druge DW sisteme).

Međusobno usklađeni i koordinirani data martovi predstavljaju **supermartove**.



Slika 1.9 Nezavisni Data Mart



Slika 1.10 Zavisni Data Mart

Svaki data mart se sastoji iz niza tabela činjenica, čiji je ključ sastavljen od više spoljnih ključeva u tabelama dimenzija. **Konformisana dimenzija** (Conformed Dimension) je ona koja ima potpuno isto značenje u svakoj tabeli činjenica sa kojom je povezana. Ta dimenzija je identična u svakom data martu. Na taj način je moguća integracija data martova, a međusobne veze se uspostavljaju preko **deljenih dimenzija** (Shared Dimensions – npr. Kupac, Proizvod). Skladište podataka se može posmatrati kao kolekcija zasebno implementiranih supermartova, povezanih arhitekturom koja se zasniva na konformisanim dimenzijama i standardizovanim činjenicama.

U arhitekturi skladišta podataka razlikuju se dve vrste data martova:

- atomski (Atomic Data Marts) – sadrže multidimenzionalne podatke na najnižem nivou
- agregirani (Aggregated Data marts) – skladište podatke u skladu sa poslovnim procesima.

Razlozi za kreiranje data marta su:

- lak pristup često zahtevanim podacima
- kreiranje zajedničkog pogleda na podatke namenjene grupi korisnika
- poboljšanje vremena reagovanja krajnjih korisnika
- jednostavnost kreiranja
- niži troškovi nego kad se koristi celo skladište podataka
- lakše utvrđivanje potencijalnih korisnika nego u radu sa celim skladištem podataka.

Dok Enterprise Data Warehouse pokriva celokupno poslovanje (obuhvata podatke vezane za sve teme poslovanja, razvija se, uobičajeno, inkrementalnim pristupom, predstavlja jedan izvor podataka za celokupan menadžment poslovnog sistema, "sinhronizuje" podatke iz svih izvora podataka poslovnog sistema i može biti osnova za izgradnju pojedinačnih Data mart sistema), Data mart pokriva samo jedan segment poslovanja (obuhvata podatke vezane za jednu temu poslovanja i može predstavljati pilot projekat za realizaciju obimnijeg Data Warehouse sistema).

Sledeća tabela pokazuje osnovne razlike između skladišta podataka i Data Marta.

Osobina	Data Warehouse	Data Mart
Pokrivena oblast	Poslovni sistem	Sektor (organizaciona celina)
Pokrivenе teme	Više	Jedna
Izvori	Više	Manji broj (nekoliko)
Tipična veličina	100 GB → 1 TB	< 100 GB
Vreme implementacije	Meseci – godine	Meseci

Arhitektura Data Warehouse sistema

Arhitektura sistema skladišta podataka opisuje elemente i usluge koje pruža skladištenje, uz detaljni prikaz integracije i optimizacije komponenti, kao i potencijalnog razvoja. Zbog fokusiranja na različite segmente i elemente, javila su se različita rešenja i metodi implementacije arhitekture.

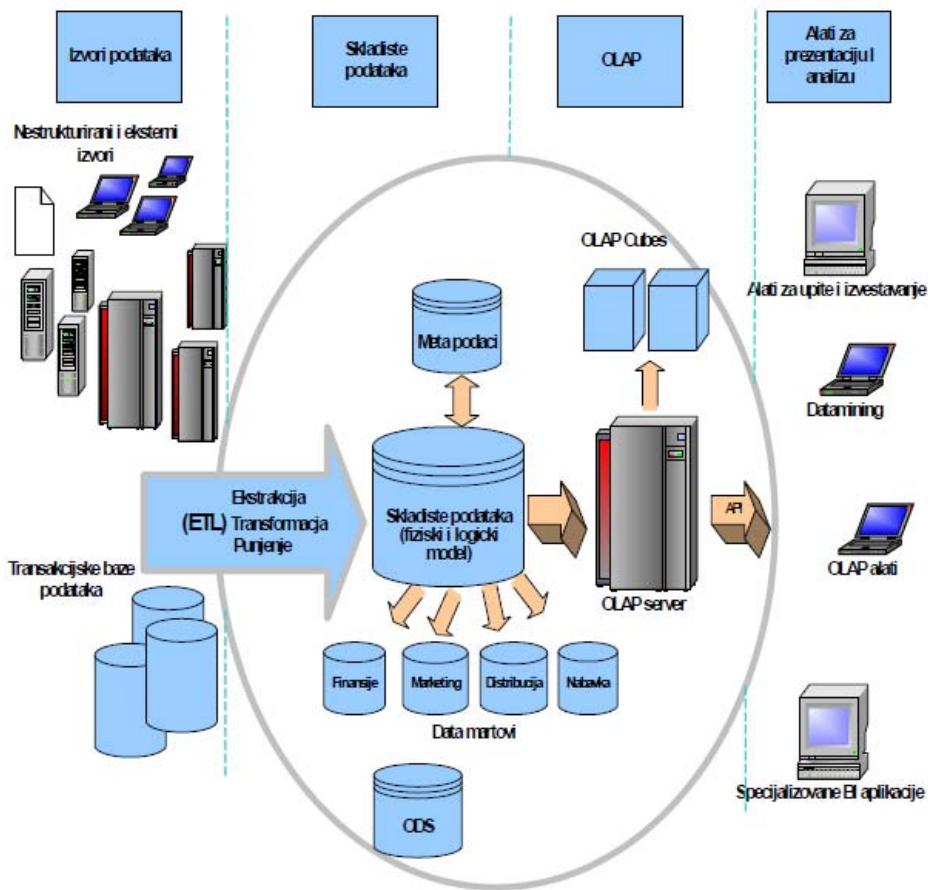
Dva najčešća pristupa organizovanju podataka u skladištima su:

- dimenzionalni (uveden i predstavljen od strane Ralfa Kimballa) i
- normalizovan (od autora Bila Inmona).

U dimenzionalnom pristupu transakcioni podaci su podeljeni ili u činjenice (numeričke podatke – specifične vrednosti) ili u dimenzije (referentne informacije koje svakoj transakciji daju njen kontekst). Npr., u transakciji prodaje činjenice su broj naručenih proizvoda i cena, a dimenzije su datum, kupac, proizvod i lokacija. Prednost ovog pristupa je lakoća upravljanja i brzina, ali u slučaju promena nastaju poteškoće.

U normalizovanom pristupu, tabele se grupišu po predmetnim oblastima koje odslikavaju definiciju podataka (npr. kupac, proizvod, finansije). Promene su luke, ali su performanse slabije nego u dimenzionalnom pristupu.

Na slici 1.11 prikazana je arhitektura Data Warehouse sistema.



Slika 1.11 Arhitektura Data Warehouse sistema

Delovi arhitekture su:

- izvori podataka (Data Sources)
- ETL (Extract/Transform/Load) procesi
- model baze podataka (logički i fizički)
- OLAP server (OLAP kocka)
- meta podaci (Metadata)
- skladište operativnih podataka (Operational Data Storage)
- data martovi (Data Marts)
- alati za izveštavanje i analizu (Reporting and Analytical Tools).

Izvori podataka mogu biti *spoljašnji* i *unutrašnji*. *Unutrašnji podaci* pripadaju kompaniji i generiše ih transakcioni sistem. Opisuju aktivnosti koje se događaju u preduzeću – u finansijama, u logistici, prodaji, proizvodnji itd. *Spoljašnji podaci* se pribavljuju izvan kompanije, i pomoću njih organizacija uočava povoljne mogućnosti i pretnje u okruženju. Spoljašnji podaci mogu se odnositi na:

- konkurentnost (proizvodi, usluge, promene kod konkurenčije,...)

- ekonomski domen (fluktuacija valuta, politički faktori, kretanje kamata, berzanski podaci,...)
- struku (tehnološki trendovi, marketinški trendovi,...)
- ekonometrijske parametre (prihodi pojedinih grupa, ponašanje kupaca,...)
- psihometriju (profilisanje kupaca,...)
- demografiju i marketing.

Spoljašnji podaci mogu biti na strukturisanim medijima (tabele, spreadsheetovi) ili nestrukturisanim (tekstualni fajlovi, fotografije, multimedijalni sadržaji).

ETL procesi imaju za cilj vađenje (extract), preoblikovanje (transform) i punjenje (load) odnosno unošenje podataka iz jednog ili više transakcionih sistema u skladištenje podataka. Pre početka ETL procesa treba obaviti pripremu (reformatiranje, usklađivanje i čišćenje podataka). Podaci u stvarnom svetu su nekompletни (nedostaju vrednosti atributa, tačne vrednosti važnih atributa), neprecizni (pogrešni), nekonistentni (neslaganje u imenima ili šiframa, nejednak i nesinhronizovan ciklus ažuriranja). Izvorne podatke iz raznih datoteka i baza podataka treba unificirati, tj. prikazati u jedinstvenom formatu. Usklađivanjem se izbegava redundantnost podataka. Čišćenjem se uklanjuju podaci koji su posledica ranijih grešaka u radu informacionih sistema.

Zbog velikog obima ulaznih podataka, u zadnje vreme se razvijaju alati bazirani na paralelnom procesiranju, sa jednim od tri osnovna tipa paralelizma:

- *Data* – ulazna datoteka deli se u manje datoteke koje se obrađuju istovremeno
- *Pipeline* – dopušten je simultani rad nekoliko SW komponenti na aistom toku podataka
- *Component* – simultani rad višestrukih procesa na različitim tokovima podataka u okviru istog posla (npr. sortiranje jedne ulazne datoteke dok se druga obrađuje).

Obično se u okviru istog posla kombinuju sva tri tipa paralelizma.

Model baze podataka (Data base Model) obuhvata:

Fizički model baze podataka – Za skladište podataka vezuju se relacioni i višedimenzionalni model. Dok je relacioni model dvodimenzionalan i može se predstaviti tabelama u kojima redovi i kolone stvaraju celije sa detaljnijim podelama (konkretnim vrednostima atributa), višedimenzionalni model prestavlja n-dimenzionalni prostor u kome svaki objekat (npr. region, kupac, vreme,...) postaje dimenzija. Objekti se predstavljaju na koordinatnim osama i za svaki novi objekat sistem dobija novu koordinatu. Dobijanje informacija vezanih za više dimenzije predstavlja "isečanje" (slicing/dicing) odgovarajućih delova (skupova podataka) iz baze. Kako su podaci predstavljeni grafovima, gde se veze trenutno identifikuju, upiti i izveštaji dobijaju se znatno brže, skraćuje se vreme pristupa i održavanja u odnosu na relacioni model koji traži indeksiranje i druge metode sortiranja. Za uzvrat, zahtevaju se specijalizovani API (Application Programming Interface) programi. Višedimenzionalni model je podesan kada postoje zajedničke i kompleksne veze među elementima kao i velika količina podataka (ako želimo spisak radnika, u kome su sva imena i matični brojevi različiti, ipak je pogodniji relacioni model)

Logički model baze podataka – Logička organizacija je izuzetno važna u procesu izgradnje skladišta podataka. Po Kimbalu, u izgradnji skladišta treba razmotriti sledećih 9 pitanja, odnosno proći 9 tačaka odlučivanja:

1. Koje poslovne procese treba modelovati, odn. koje su tabele činjenica?
2. Šta je suština svake tabele činjenica?
3. Koje su dimenzije svake tabele činjenica?
4. Koje su činjenice bitne?

5. Koji su atributi dimenzija?
6. Kako pratiti dimenzije koje se sporo menjaju?
7. Koje aggregacije treba koristiti, koje tabele su heterogene, koji su načuni upita?
8. Koliko dugo treba čuvati podatke?
9. S kojom učestanošću se podaci ekstrahuju i učitavaju u skladište?

Odluke treba donositi u navedenom redosledu. Ovo je metodologija od vrha ka dnu (Top-Down), jer počinje utvrđivanjem ključnih procesa, a u centar pažnje se stavljuju karakteristike dimenzionog modelovanja. Podaci se dele na mere i kontekst. Numeričke mere su činjenice, a kontekst je tekst koji opisuje situaciju u trenutku nastanka činjenica. Kontekst treba podeliti u logičke grupe, npr. u e-trgovini kontekst se deli na Proizvod, Vreme, Kupac, itd. Te grupe su dimenzije.

Najzastupljeniji oblici šema baza podataka su zvezdasta (star join) i pahuljičasta (snowflake). Kriterijumi za izbor šeme su: složenost problema, veličina tabela, performanse i memoriski prostor.

Skladište operativnih podataka (Operational Data Storage – ODS) je integrisana baza operativnih podataka, koja sadrži "žive" podatke a ne preglede u nekom trenutku. Tradicionalna arhitektura skladišta podataka nije u skladu sa potrebama menadžera za trenutno važećim podacima, radi odlučivanja u realnom vremenu. Zato kad performanse postanu kritične potrebna su "živa", operativna skladišta podataka, odn, tzv. "report server" ili "ogledalo baze".

Osnovne karakteristike skladišta operativnih podataka su:

- predmetna orijentisanost – ODS se projektuje i organizuje oko glavnih predmeta interesovanja kompanije, kao što su npr. kupac ili proizvod – na jednom mestu su osnovni podaci o kupcu i sve njegove transakcije.
- integrisanost – ODS predstavlja integriranu sliku predmetno orijentisanih podataka izvučenih iz bilo kog operativnog sistema.
- orijentisanost na trenutnu vrednost – ODS odslikava trenutni sadržaj njegovih izvornih sistema
- promenljivost – ODS se menja onoliko često koliko je potrebno za prikaz trenutnog stanja
- detaljnost – granularnost ODS-a može ali ne mora da se razlikuje od granularnosti svog izvornog operativnog sistema.

Alati za izveštavanje i analizu (Reporting and Analytical Tools) su aplikacije koje omogućuju analizu, izveštavanje, postavljanje upita i sl. Neki od njih su:

- Business Intelligence alati
- Izvršni informacioni sistemi (poznati kao Dashboards)
- OLAP alati
- Analitičke aplikacije
- Data Mining.

Sa stanovišta krajnjeg korisnika, ovo je najvažniji sloj u arhitekturi skladišta podataka. Korisnici se prema potrebnim alatima dele na:

- "moćne korisnike" (oni koji imaju potrebe da poznaju strukturu skladišta podataka međuzavisnosti u njemu i da složenim alatima rade sopstvene kompleksne upite i analize)
- "povremene korisnike" (nisu direktno zainteresovani za detalje skladišta podataka, zauzeti svojim poslovima, a povremeno im trebaju informacije)

- korisnike koji imaju potrebu za statičkim informacijama (precizno definisanim podacima u određenom vremenskom intervalu)
- korisnike koji zahtevaju dinamičke ili ad hoc upite i analitičke mogućnosti alata (analitičari kojima u bilo kom trenutku može zatrebatи bilo koja informacija iz skladišta podataka, uz visoke performanse i drill-down mogućnosti).

Različite vrste korisnika zahtevaju različite prezentacione alate, ali svi oni mogu da pristupaju zajedničkom skladištu podataka.

ETL alati – ETL proces se može isprogramirati korišćenjem ETL alata ili bez njih, standardnim programima (hand-coded ETL). Hand-coded ETL procesi su jednostavniji za održavanje i izmene, ali ne obezbeđuju metapodatke o transakcionim izvorima podataka i šemama skladišta podataka koji korisnicima pomažu da razumeju poreklo informacija u skladištu podataka.

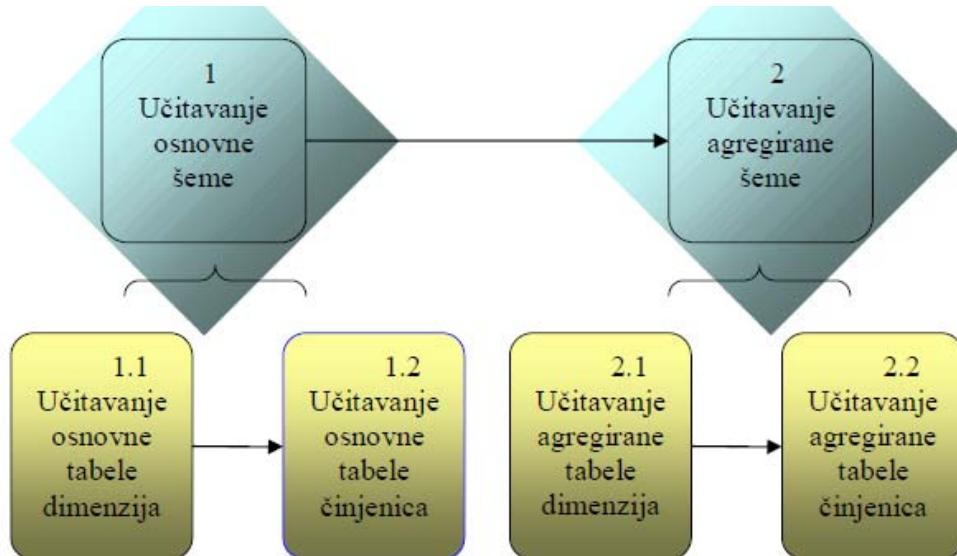
Paket ETL softvera obično podržava skladište informacija o izvornim podacima i šemama skladišta podataka. Grafički ETL alati omogućavaju razvoj softvera bez pisanja koda.

ETL alati se mogu podeliti u dve grupe:

- alati za generisanje koda (code generators) – mainframe-based kao npr. COBOL, RDBMS-based (storage procedure) ili se izvršavaju na serveru (npr. Java)
- server-based alati, što je danas uobičajeno; uključuju mašine koje izvode ETL proces i maju direktnu interakciju sa izvorom i odredištem podataka; uključuju integracioni softver, softver za poruke, alate za sortiranje, aplikacije za čišćenje podataka itd.

Inkrementalni ETL proces – Punjenje skladišta podataka sastoji se iz dva procesa: inicijalnog punjenja i inkrementalnog punjenja. Inicijalnim punjenjem se u skladište unose početne vrednosti svih podataka, a inkrementalnim se unose samo izmenjeni i novi podaci. Inkrementalno radi sa manjom količinom podataka ali je složenije od inicijalnog i zahteva poštovanje pravila koja proističu iz načina na koji se obrađuju podaci koji se menjaju u izvornom sistemu.

Proces učitavanja podataka u skladište podataka može se predstaviti kao na slici 1.12:



Slika 1.12 Proses učitavanje podataka u skladište podataka

a) Učitavanje osnovne šeme

Osnovna šema sadrži podatke koji su direktno učitani iz izvornog sistema; ona predstavlja originalnu šemu. Agregirana šema je izvedena iz osnovne šeme – sadrži manji broj redova koji se dobijaju izvršavanjem nekog upita.

Tabela činjenica sadrži podatke povezane sa procesom i spoljašnji ključ pomoću koga se određuje sa kojom tabelom dimenzija je povezana. Činjenice su pokazatelji za ocenu procesa – bez dimenzija su beskorisne. Dimenzije daju činjenicama sadržaj i smisao. Skup atributa određuje jednu tabelu u kojoj jedna kolona predstavlja surrogat ključ kao jedinstveni identifikator jednog sloga u skladištu podataka. Slična terminologija važi i u agregiranoj šemi.

Učitavanje podataka u osnovnu šemu deli se na dva procesa: učitavanje osnovne tabele dimenzija i učitavanje osnovne tabele činjenica; tabela dimenzija mora se napuniti pre tabele činjenica, jer tabela činjenica ima primarni ključ tabele dimenzija kao spoljašnji ključ. Pri učitavanju tabele dimenzija moraju se uzeti ulazni podaci, sakupiti vrednosti dimenzija, obraditi novi zapisi dimenzija, izvršiti promene tipa 1 i tipa 2, kao u sledećem primeru:

Tabela Proizvod ima sledeće atribute:

Proizvod
Šifra_proizvoda(SK) ID(NK)
Naziv_proizvoda(1)
Opis_proizvoda(1)
Jedinica_mere(1)
Naziv_brenda(1)
Brand_code(2)
Brand_menadžer(2)
Naziv_kategorije(1)
Category_code(2)

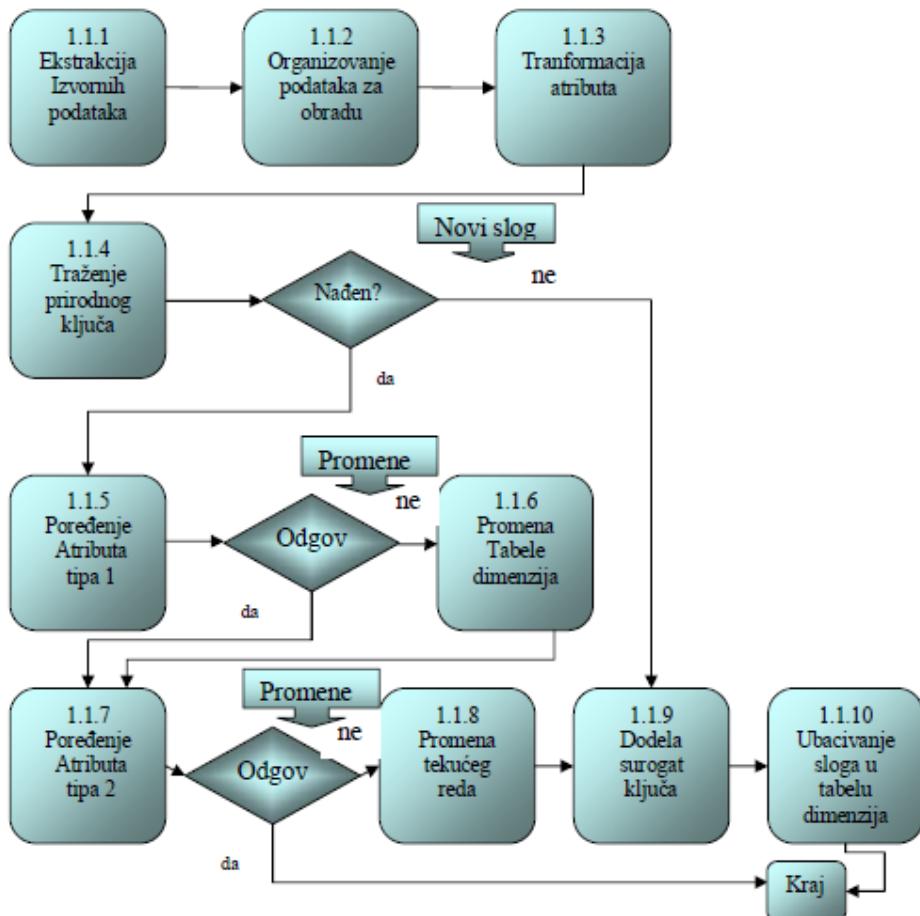
Prirodni ključ (NK) je jedinstveni identifikator sloga u izvornom sistemu, ali se u skladištu podataka usled izmena pojedinih atributa može javiti više slogova sa istim prirodnim ključem. Zato se mora uvesti surrogat ključ (SK) Šifra_proizvoda kao atribut koji će jedinstveno identifikovati svaki red u tabeli.

U tabeli se javljaju atributi označeni kao atributi tipa 1 i tipa 2. Kada je atribut tipa 1, promena tog atributa dovodi do kreiranja novog sloga u skladištu podataka; kad se menja atribut tipa 1, dolazi do promene odgovarajućeg sloga u skladištu.

Na slici 1.13 prikazani su osnovni koraci učitavanja podataka u tabelu dimenzija.

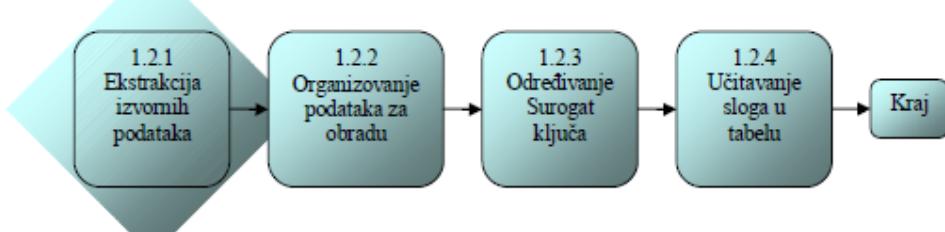
Prvi korak je ekstrakcija podataka. Podaci se mogu dobiti na više načina, npr. relacioni podaci korišćenjem SQL servera, eksterni podaci on line itd. Po ekstrakciji se svi atributi koji se odnose na jedan slog moraju organizovati u jedan red (korak 1.1.2 na slici 1.11). U sledećem koraku se svaki red izvornih podataka prevodi u odgovarajući atribut tabele dimenzija (pravila: polja sa višestrukim atributima mogu biti podeljena tako da se dobije niz atributa, polja sa null vrednostima mogu se zameniti razumljivim tekstrom kao "nepoznato", može se podesiti stanje i format atributa itd). Posle ovoga je set potencijalnih slogova spreman za obradu. Sledeći koraci određuju kako se upravlja ovim slogovima. U

koraku 1.1.4 traži se prirodni ključ (ako se prirodni ključ ne nađe u postojećoj tabeli, slog je nov). U koraku 1.1.9 novom slogu se dodeljuje surogat ključ, dok se u koraku 1.1.10 novi slog unosi u tabelu.



Slika 1.13 Koraci u učitavanju osnovne tabele dimenzija

Učitavanje podataka u tabelu činjenica može se predstaviti sledećom šemom (slika 1.14):



Slika 1.14 Učitavanje u osnovnu tabelu činjenica

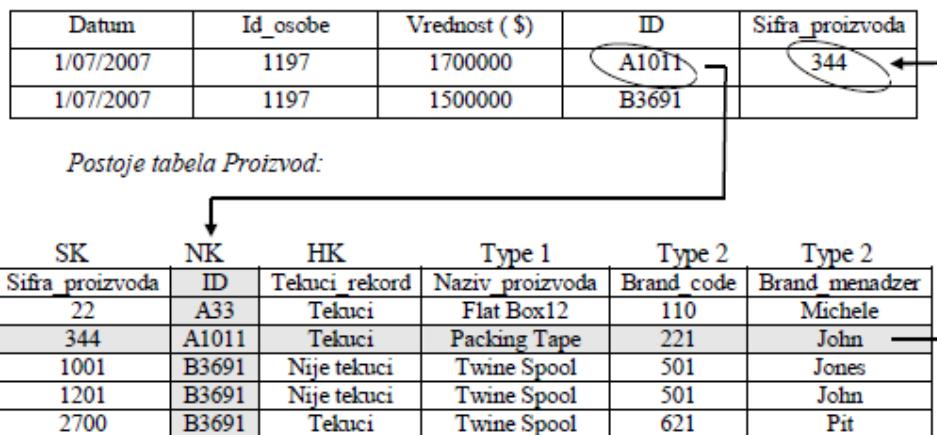
Prvi korak (1.2.1) je ekstrakcija izvornih podataka, koje ponekad treba rekonstruisati tako da može biti učitan jedan red odvojeno od drugih (korak 1.2.2). Za svaku dimenziju povezану sa tabelom činjenica mora se odrediti odgovarajući surogat ključ, korišćenjem prirodnog ključa (ovaj korak se ponavlja za svaki slog – u tabeli dimenzija se traži prirodni ključ dobijen od izvornog sistema i kad se nađe, slogu koji se uključuje u tabelu činjenica se dodeljuje odgovarajući surogat ključ).

Primer određivanja surogat ključa:

Prema slici 1.15, slog koji treba učitati je slog čiji ID je A1011. Ova vrednost se traži u tabeli dimenzija. Proizvod sa traženim prirodnim ključem postoji i njegova šifra je 344. Ova vrednost se dodeljuje slogu koji se učitava.

U procesu učitavanja novog sloga treba ispitati i promene atributa tipa 2. Radi olakšavanja ovog procesa, u tabelu Proizvod uvodi se nova kolona Tekuci_record. Kad se promeni atribut, prethodni "Tekuci" record prelazi u "Nije tekuci". Npr., u tabeli Proizvod, atributi Brand_code i Brand_menadzer su atributi tipa 2. Ako se neki od ovih atributa promeni, za istu vrednost prirodnog ključa, u tabelu Proizvod biće učitan novi slog. U gornjoj tabeli na slici 1.13 vidimo da je drugi slog koji treba učitati slog čiji je ID B3691. Program mora da nađe slog sa traženim prirodnim ključem a da važi da je Tekuci_record = "Tekuci". Slog koji zadovoljava te uslove je slog čija je šifra 2700. Ova vrednost se dodeljuje slogu koji se učitava.

Slogovi koje treba učitati u tabelu činjenica



Slika 1.15 Određivanje surogat ključa

b) Učitavanje agregirane šeme

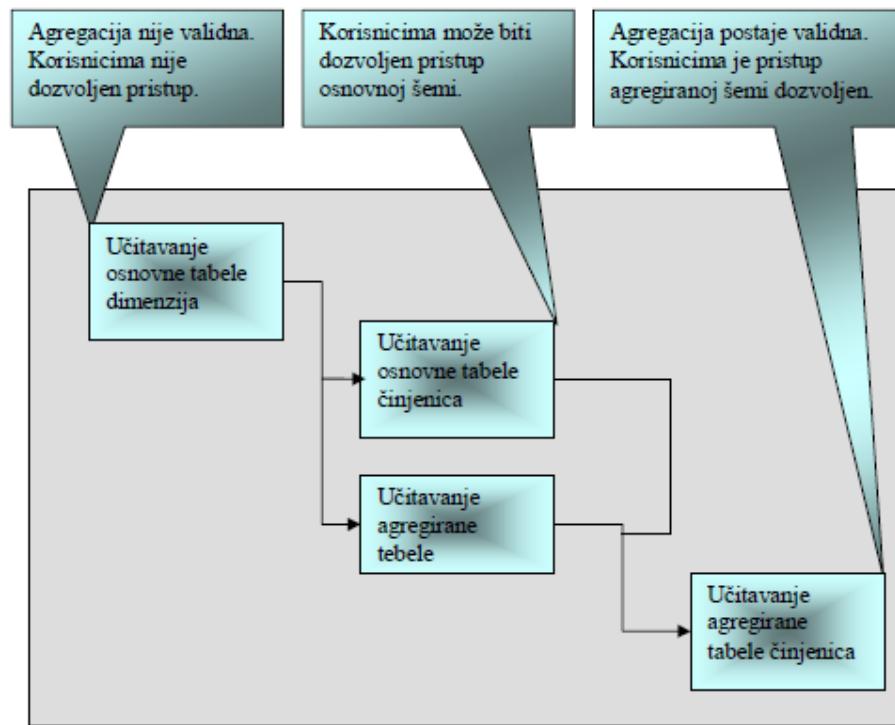
Ovaj proces je sličan procesu učitavanja osnovne šeme. U punjenju osnovne šeme projektovani su odvojeni procesi za svaku tabelu. Odvojeni procesi se koriste i u procesu učitavanju podataka u agregiranoj tabelo činjenica i agregiranoj tabeli dimenzija. Time se olakšava održavanje, a nezavisni procesi daju veću efikasnost u automatizaciji procesa učitavanja podataka (posebno bitno za agregaciju, jer je ona podložnija promenama od osnovne šeme). Aggregirana tabela činjenica može biti odbačena, aли aggregirana tabela dimenzija ne, jer može biti uključena u novu agregaciju.

Da agregacija ne bi postala nevažeća, potrebno je da se njen učitavanje vrši istom frekvencijom kao i učitavanje osnovne šeme, inače aggregirana šema neće biti validna ako dođe do promena u osnovnoj šemi.

Na slici 1.16 prikazan je redosled učitavanja podataka.

Da bi agregacija postala validna, najpre se učitavaju podaci u osnovnu tabelu, nakon čega se podaci istovremeno učitavaju u tabelu činjenica i aggregiranu tabelu dimenzija. Učitavanje aggregirane tabele činjenica počinje po završetku svih ostalih učitavanja. Korisnici mogu da pristupe osnovnoj šemi kad se završi učitavanje osnovne tabele

činjenica. Po učitavanju agregirane tabele činjenica korisnici dobijaju pravo pristupa celom skladištu podataka.



Slika 1.16 Redosled učitavanja tabela

Sledeća tabela pokazuje neke bitne razlike između klasičnih (OLTP) baza podataka i skladišta podataka.

Karakteristika	Operativna BP (OLTP)	Skladište podataka
Tipične operacije	Operacije ažuriranja	Operacije upita
Kritične transakcije	Transakcije ažuriranja	Transakcije upita
Ažuriranje BP	Veliki broj DML operacija	Punjene i periodično osvežavanje
Frekvencija upita	Niska / srednja	Visoka
Kompleksnost upita	Niska	Visoka
Količina podataka / transakc.	Mala / srednja	Velika
Očekivano vreme odgovora	Do nivoa sekunde	Nekoliko sekundi – više sati
Vremenska diskretizacija podataka	Dan - sekunda	Dan - godina
Aktuelnost podataka	Do jedne godine	Više godina
Obim baze podataka	MB - GB	GB - TB
Povećanje obima BP	Linearno	Polinomijalno / Eksponencijalno
Granularnost podataka	Elementarni podaci	Agregirani podaci
Nivo agregacije	Nizak	Visok
Šema BP	Normalizovana, kompleksnija	Denormalizovana, manje kompleksna
Izvori podataka	Operativno poslovanje	Operativna BP, interni i eksterni izvori
Organizacija podataka	Prema funkcijama	Prema temama
Podrška poslovnim procesima	Operativno poslovanje	Analiza i odlučivanje
Forme za prikaz podataka	Statičke, retko promenljive	Kontekstno zavisne, promenljive
Intenzitet korišćenja BP u vremenu	Uniforman	Neuniforman, mogući "udarni" termini

Evolucija obrade informacija

Iz mnogo razloga, nekada je bilo skoro nemoguće iz aplikacija dobiti prave informacije za odlučivanje. Korporacijske aplikacije nisu bile integrisane, nisu sadržale istorijske podatke i nalazile su se na tehnologijama kojima nije jednostavno pristupiti. Rezultat je bila frustracija krajnjih korisnika.

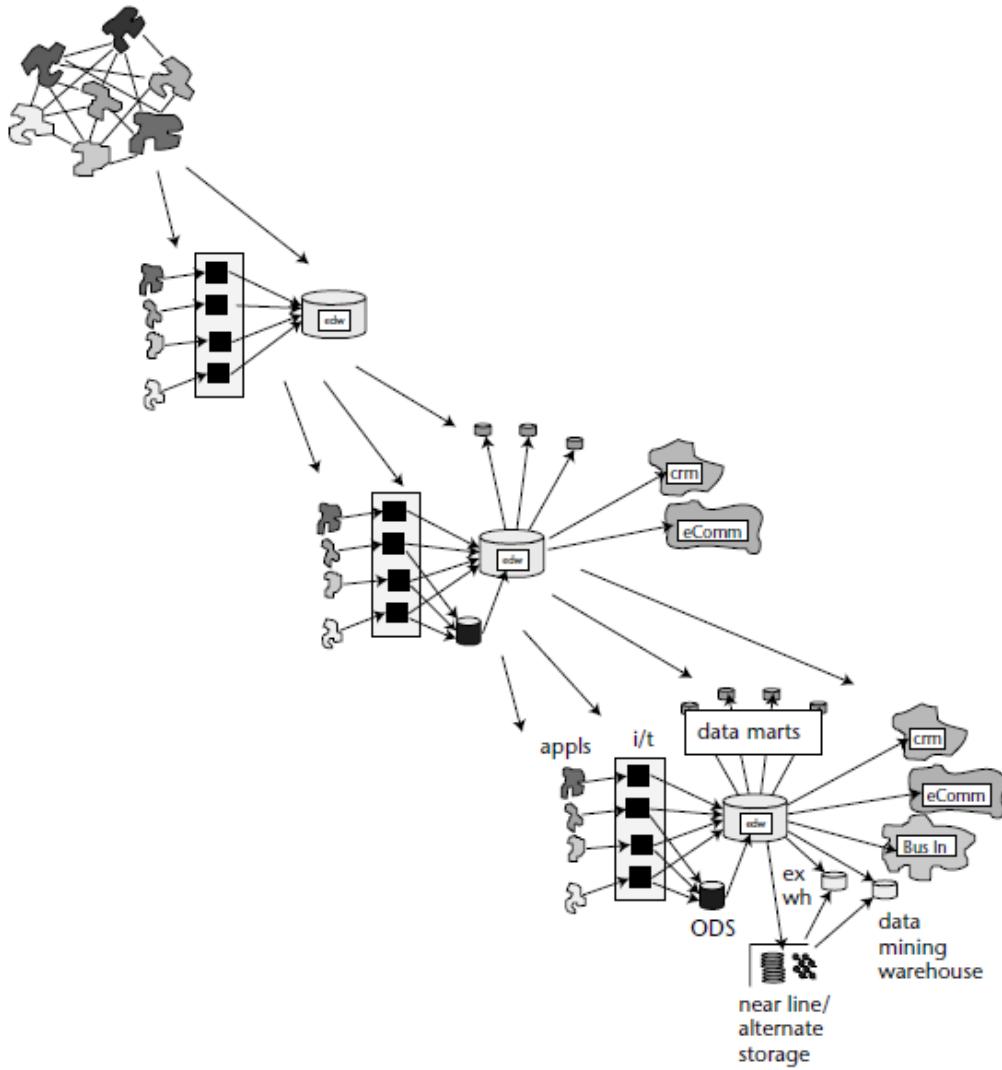
Frustracija krajnjih korisnika dovela je do ideje skladišta podataka. Skladište podataka je bilo ozbiljno odstupanje od teorije baze podataka koja je zahtevala da se svi podaci okupe u jednu bazu podataka. Koncept skladišta podataka fokusirao se na različite vrste baza podataka za različite svrhe. Operativna obrada transakcija podržana je jednim tipom baze podataka a obrada informacija drugim. Skladište podataka je izazvalo da se podaci integrišu i memorišu u vremenu u fizički odvojenu tehnologiju baze podataka koja je optimalna za pristup i analizu informacija. Iz skladišta podataka razvili su se Data Mart moduli, pojavile su se aplikacije sistema za podršku odlučivanju, a skladišta podataka su porasla u veličini do tačke kada obim podataka u skladištu za nekoliko redova veličine prevaziđa veličinu ranijih baza podataka. Rane online baze podataka smatrane su velikim sa veličinom od 10 GB, skladišta podataka se smatraju velikim sa 10 TB, što je za tri reda veličine veće od 10 GB. Sa skladištima podataka javila se potreba za drugim oblicima informacionih tehnologija.

U vezi sa skladištima podataka javio se niz različitih struktura u arhitekturi sistema. Ubrzo su postali operativni oblici kao *operational data stores* (ODS), *data mining* i moduli za pretraživanje, alternativni oblici memorije itd. Skladište podataka je omogućilo različite oblike okruženja u kojima se odvijaju vrlo različiti vidovi obrade informacija. U središtu tih struktura bilo je skladište podataka. Skladište podataka obezbedilo je granularne podatke koji su preoblikovani u različite forme da bi zadovoljili različite vidove obrade za podršku odlučivanju.

Arhitekturna mreža koja je proistekla iz ovoga nazvana je "korporacijska fabrika informacija" (corporate information factory – CIF). Na slici 1.17 prikazan je rast sveta obrade informacija i skladišta podataka.

Na čelu obrade informacija je skladište podataka i široka arhitektura oko skladišta podataka – CIF. Često je arhitektura koja je nastala nazivana "hub-and-spoke" (glavčina i paoci) arhitektura. Slično planovima leta i strategijama koje koriste komercijalne aviokompanije, gde 1 aerodrom igra ulogu glavčine točka iz koje polaze "paoci" – letovi u različitim pravcima, skladište podataka predstavlja glavčinu a različite analitičke aplikacije i data mart-ovi predstavljaju destinacije. Proces isporuke informacija na zadate destinacije je analogan aviolinijama ili paocima.

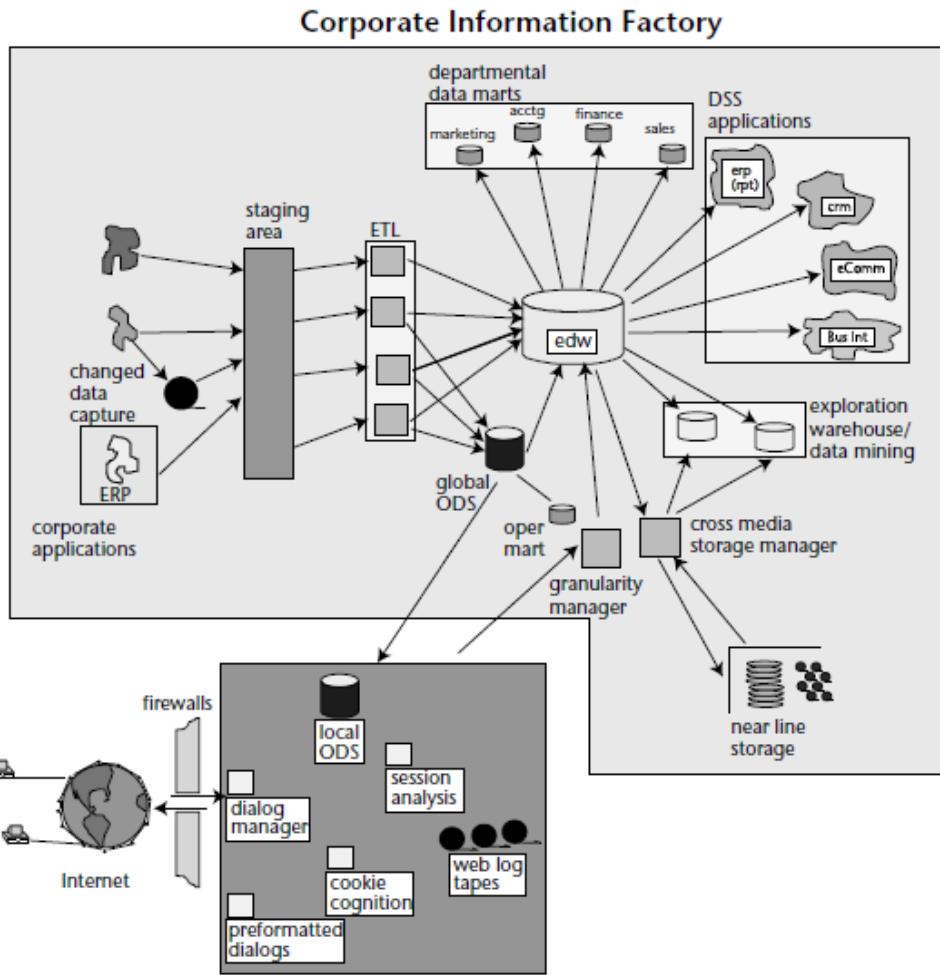
CIF predstavlja napredak u razvoju koji je nastao sa obradom informacija, uporedno sa razvojem Enterprise Resource Planning (ERP) sistema.



Slika 1.17 Evolucija CIF

Po mnogo čemu CIF liči na plan grada. Plan grada zahteva za svoj razvoj godine i decenije. U slučaju velikih gradova koji se grade planski potrebne su godine i decenije za realizaciju plana. Isto važi i za CIF.

Na slici 1.18 prikazan je CIF u razvijenom obliku.



Slika 1.18 CIF i okruženje za elektronsko poslovanje

Postavljanje osnove za poslovnu inteligenciju

Kad je u upotrebu uvedeno skladište podataka, time je postavljena osnova za poslovnu inteligenciju. Danas ima više različitih oblika poslovne inteligencije. Kao što se vidi na slici 1.19, poslovna inteligencija se pojavljuje kroz pretraživanje podataka, data mining, data mart-ove, podršku e-poslovanju i sisteme za podršku odlučivanju (DSS).

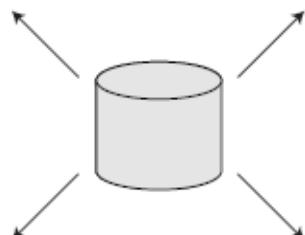
Skladište podataka postaje infrastruktura na koju se oslanja poslovna inteligencija. Kad se izgradi skladište podataka, postaje lako i prirodno da se nad njim izgradi i sistem poslovne inteligencije.

Pretraživanje podataka/
data mining

- testiranje hipoteze
- analiza uzoraka
- prediktivno modelovanje
- neuronske mreže
- stabla odlučivanja

data marts

- KPI
- redovno merenje
- rad sa KPI varijablama
- redovna sumarizacija
- podaci oblikovani prema zahtevima
- OLAP multidimenziona obrada
- tabele činjenica
- tabele dimenzija
- vizuelizacija podataka



podrška e-poslovanju

- stvaranje portala
- filtriranje i redukcija podataka
- integracija podataka
- prodaja, promocije, posebni događaji

DSS aplikacije

- CRM
- bodovanje kredita
- upravljanje online kupcima
- analiza elastičnosti

Slika 1.19 Vidovi poslovne inteligencije

2. MULTIDIMENZIONALNA ANALIZA PODATAKA

2.1 Multidimenzioni izrazi (MDX)

Multidimenzionalni izrazi (MultiDimensional eXpressions – MDX) predstavljaju jezik upita za OLAP baze podataka, slično kao što SQL predstavlja jezik upita za relacione baze podataka. Pored upita, omogućavaju i manipulisanje i kalkulacije (sa sintaksom sličnom onoj kod spreadsheet jezika, npr. Excel-a) nad multidimenzionalnim podacima memorisanim u vidu OLAP kocke. Neki od MDX izraza mogu se direktno prevesti u tradicionalni SQL, ali najčešće se zahteva sinteza više nezgrapnih SQL izraza čak i da bi se dobili vrlo jednostavni MDX izrazi. MDX su prihvatili mnogi prodavci OLAP sistema, i može se reći da su postali standard u poslovnoj inteligenciji.

Istorijat

MDX je prvi put uveden u upotrebu 1997. god., u sastavu OLE DB for OLAP od strane grupe SQL Server inženjera Microsoft-a. U komercijalnu upotrebu uveden je 1998. kroz Microsoft OLAP Services 7.0, zatim kroz Microsoft Analysis Services.

MDX su brzo prihvatili i drugi prodavci OLAP sistema, kako za aplikacije na strani servera (Applix, icCube, MicroStrategy, NCR, Oracle Corporation, SAS, SAP, Teradata, Whitelight) tako i za klijent aplikacije (Panorama Software, PoweOLAP, XLCubed, Proclarity, AppSource, Jaspersoft, Cognos, Business Objects, Brio Technology, Crystal Reports, Microsoft Excel i Microsoft Reporting Services).

Sa uvođenjem programa *XML for Analysis*, MDX je standardizovan kao jezik za upite i još šire je prihvaćen. *XML for Analysis* se odnosi na detalje MDX jezika upita u okviru OLE DB for OLAP specifikacije. U okviru Analysis Services 2005, Microsoft je dodao nove MDX ekstenzije, npr *subselect*, koje primenjuju neki spreadsheet programi kao Microsoft Excel 2007.

MDX tipovi podataka

U MDX postoji 6 primarnih tipova podataka, i to:

- **Scalar** – skalar je broj ili niz znakova (string). Može se navesti kao literal, npr. broj 5 ili kao niz, npr. "OLAP" ili ga može vratiti MDX funkcija, npr Aggregate (broj), UniqueName (niz), .Value (broj ili niz) itd.
- **Dimension/Hierarchy** – *Dimension* predstavlja dimenziju OLAP kocke. Dimenzija je primarni pojam o oređivanju informacija o atributima u kocki. MDX ne zna niti pretpostavlja veze među dimenzijama – one su međusobno nezavisne. Dimenzija sadrži neke članove (members – objašnjeno u nastavku) organizovane u nekoj hijerarhiji po nivoima, koja se može specificirati jedinstvenim imenom, npr. [Time], ili je vraća MDX funkcija, npr. *Dimension*. *Hierarchy* označava hijerarhiju dimenzija OLAP kocke. Ona se može specificirati jedinstvenim imenom, npr. . [Time] . [Fiscal], ili je vraća MDX funkcija, npr. *Hierarchy*. Hijerarhije se sadrže u dimenzijama (OLE DB for OLAP MDX specifikacija ne razlikuje dimension i hierarchy tipove podataka, dok ih Microsoft Analysis Services tretira odvojeno).
- **Level** - Level označava nivo u hijerarhiji dimenzija. Može se navesti kao jedinstveno ime, npr. . [Time] . [Fiscal], . [Month] ili ga vraća MDX funkcija, npr. . Level.
- **Member** – Member je član u hijerarhiji dimenzija. Može se navesti po svom

jedinstvenom imenu, npr. [Time] . [Fiscal] . [Month] . [August 2006], po kvalifikovaom imenu, npr. [Time] . [Fiscal] . [2006] . [Q2] . [August 2006], ili ga može vratiti neka MDX funkcija, npr. . PrevMember, . Parent, . FirstChild i sl. Svi članovi su vezani za hijerarhiju. Ako je isti proizvod član dveju različitih hijerarhija, npr. ([Product] . [ByManufacturer] and [Product] . [ByCategory]), mogu se koristiti dva člana koje treba koordinirati kroz setove i n-torke.

- **n-torka** – n-torka je uređeni skup sastavljen od jednog ili više članova iz različitih dimenzija. n-torke se mogu specificirati nabranjem članova, npr. ([Time] . [Fiscal] . [Month] . [August], [Customer] . [By Geography] . [All Customers] . [USA], [Measures] . [Sales]) ili vraćanjem iz MDX funkcije, npr. . Item.
- **Set** – set je, u Microsoft tumačenju, uređen skup n-torki iste dimenzije ili istog položaja u hijerarhiji. Može se navesti nabranjem n-torki, npr. {[Measures] . [Sales], [Time] . [Fiscal] . [2006]}, {[Measures] . [Sales], [Time] . [Fiscal] . [2007]}) ili kao povraćaj iz neke MDX funkcije ili operatora, npr. Crossjoin, Filter, Order, Descendants i sl.
- **Drugi tipovi podataka** – osobine članova odgovaraju atributima u kontekstu skladišta podataka. Mogu se pretraživati po imenu, korišćenjem klauzule axis PROPERTIES u upitu. Skalarna vrednost osobine člana može se dobiti u MDX izrazu ili imenovanjem osobine (npr. [Product] . CurrentMember . [Sales Price]), ili korišćenjem specijalne funkcije za pristup (npr. [Product] . CurrentMember.Properties ("Sales Price")). U ograničenoj meri, MDX prihvata i druge tipove podataka – npr. Array (niz) se može koristiti u funkciji SetToArray da specificira niz koji se ne obrađuje u MDX već se predaje kao korisnička funkcija u ActiveX biblioteku. Objekti drugih tipova podataka predstavljaju se kao skalarni nizovi koji ukazuju na imena objekata kao što je grupno ime mere u Microsoftovoj MeasureGroupMeasures funkciji ili KPI ime u Microsoftovim KPIValue ili KPIGoal funkcijama.

Primer upita

Sledeći primer, uzet iz "SQL Server 2000 Books Online", pokazuje jednostavan MDX upit koji koristi komandu SELECT. Upit kao rezultat vraća set koji sadrži obim prodaje za 2002. i 2003. za prodavnice u Kaliforniji.

```
SELECT
    { [Measures].[Store Sales] } ON COLUMNS,
    { [Date].[2002], [Date].[2003] } ON ROWS
FROM Sales
WHERE ( [Store].[USA].[CA] )
```

U ovom primeru, upit definiše sledeći rezultujući setinformacija:

- SELECT uslov postavlja osu po kojima se vrši upit kao što je član Store Sales u dimenziji Measures i članovi 2002 i 2003 dimenzije Date.
- FROM uslov ukazuje da je izvor podataka OLAP kocla pod nazivom Sales.
- WHERE uslov definiše "slicer" osu (osu preseka – član California u dimenziji Store).

U MDX upitu se može definisati do 128 osa za upit.

2.2 Upotreba PI alata

PI alati su vrsta aplikativnog softvera dizajnirana za pretraživanje, analizu i prikaz podataka. Čitaju podatke koji su prethodno memorisani, najčešće (mada ne i obavezno) u skladištu podataka ili data martu.

Tipovi PI alata

Softver za poslovnu inteligenciju deli se u sledeće kategorije:

- ***Spreadsheet softver***

Prikazuje podatke obično u vidu dvodimenzionalne matrice sastavljene od redova i kolona. Svaka ćelija (presek reda i kolone) sadrži alfanumerički tekst, numeričke vrednosti ili formule. Formula određuje kako se sadržaj konkretnе ćelije izračunava na osnovu sadržaja drugih ćelija (pojedinačnih ili njihovih kombinacija) i održava ažurnost pri svakoj izmeni sadržaja ćelija uključenih u formulu. Treća dimenzija se simulira upotrebom matrica na više nivoa.

Savremeni spreadsheet programi obuhvataju više worksheet-ova (listova) koji čine workbook (knjigu). Ćelija jednog worksheet-a može da referencira ćelije u drugim worksheet-ovima, kao i u drugim workbook-ovima.

Tipičan predstavnik je ***Microsoft Office Excel***.

- ***Softver za upite i izveštavanje***

To su alati koji izdvajaju, sortiraju, sumarizuju i prezentuju selektovane podatke. U velikom broju alata koji spadaju u ovu kategoriju, pomenimo ovde **SQL Server Reporting Services (SSRS)**, Microsoft-ov program za generisanje interaktivnih ili štampanih izveštaja. Uključen je u Developer, Standard, i Enterprise verzije Microsoft SQL Server-a kao opcija pri instalaciji. Razvijen je 2004. godine i uključen u SQL Server 2000, da bi u novijim verzijama bio standardna komponenta SQL Server-a.

Za izradu izveštaja koristi se Report Definition Language (RDL). Izveštaji se mogu dizajnirati u novijim verzijama Microsoft Visual Studija sa instaliranim Business Intelligence Projects ili Report Builder plug-inom. Izveštaji definisani u RDL-u mogu biti u različitim formatima (Excel, PDF, CSV, XML, TIFF i drugi grafički formati, kao i XML Web Archive. SSRS za SQL Server 2008 može da napravi izveštaje i u .DOC formati za Microsoft Word.

SQL Server Reporting Services podržava i ad hoc izveštaje: dizajner razvija šemu izveštaja i postavlja je na server za izveštavanje, dok korisnik može da izabere relevantna polja podataka i da generiše izveštaje, pa da ih lokalno daunlouduje.

- ***OLAP alati***

OLAP alati pomažu korisnicima da interaktivno analiziraju multidimenzionalne podatke iz različitih perspektiva. OLAP (On Line Analytical Processing) se sastoji od tri bazne analitičke operacije: konsolidacija (roll-up), drill-down i slicing/dicing operacije. Konsolidacija uključuje agregiranje podataka koji se akumuliraju i upoređuju u jednoj ili više dimenzija (npr. prikupljanje podataka o prodaji u svim prodajnim punktovima radi anticipacije trenda prodaje). Drill-down je tehnika koja korisnicima omogućuje navigaciju kroz detalje podataka (npr. kada se traži prodaja po proizvodima u različitim regionima). Slicing i dicing omogućuju korisnicima da iz OLAP kocke izvuku (slicing) konkretni skup podataka i da ga vide (dicing) sa različitih aspekata.

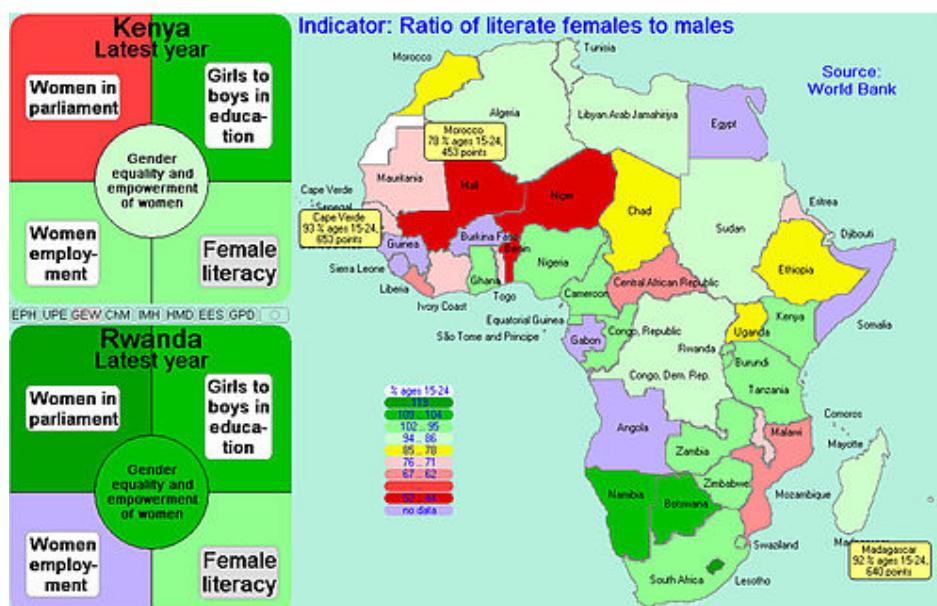
Prvi proizvod koji je izvodio OLAP upite bio je *Express* razvijen još 1970. a prihvaćen od Oracle Co. 1995. Krajem 90-tih godina prošlog veka OLAP je doživeo ekspanziju, sa nizom komercijalnih softvera. Microsoft je 1998. razvio svoj prvi OLAP Server – Microsoft Analysis Services, koji danas prestavlja standard.

- **Digital Dashboards**

Predstavljaju jednostavne real-time korisničke interfejse koji daju grafičku prezentaciju trenutnog stanja i istorijske trendove posmatranog poslovnog parametra. Korisnici vide podatke višeg nivoa a do nižeg stižu na drill-down način (klikom na deo slike). Omogućuju:

- vizuelno predstavljanje posmatranih podataka
- identifikaciju i korekciju negativnih trendova
- merenje efikasnosti / neefikasnosti
- generisanje detaljnih izveštaja za prikaz novih trendova
- donošenje pouzdanih odluka
- usklađivanje strategije sa poslovnim ciljevima
- uštede u vremenu u odnosu na analizu brojnih posebnih izveštaja
- potpunu vidljivost sistema u trenutku
- brzu identifikaciju veza među podacima.

Primer: Ravnopravnost žena u Africi



- **Data mining softver**

Namenjen je za automatsku ili poluautomatsku analizu velikog obima podataka sa ciljem da se iz njega izvuku do tada neuočene interesantne veze među podacima u vidu grupa slogova podataka (klaster analiza), neuobičajeni slogovi (otkrivanje anomalija) i zavisnosti (traganje za pravilom povezivanja). Obično koriste tehnike baze podataka kao što je prostorno indeksiranje, i metode: detekcija anomalija, nalaženje pravila povezivanja, klasifikacija, klaster analiza, stable odlučivanja, faktorska analiza, neuronske mreže, regresiona analiza, struktura analiza podataka, anaiza sekvenci i analiza teksta.

Najpoznatiji komercijalni data mining softveri i aplikacije su:

- IBM InfoSphere Warehouse – data mining platforma u bazi podataka (IBM)

- IBM SPSS Modeler – data mining softver (IBM)
- KXEN Infinite Insight – data mining softver (KXEN)
- **Microsoft Analysis Services** – data mining softver (Microsoft)
- SAS Enterprise Miner – data mining softver (SAS Institute)
- STATISTICA Data Miner – data mining softver (StatSoft)
- Oracle Data Mining – data mining softver (Oracle)
- LIONsolver – integrisana aplikacija za data mining, poslovnu inteligenciju i modelovanje, koja koristi LION (Learning and Intelligent OptimizatioN) pristup.