

Računovodstveni
Informacijski
Sustavi

Poslovna inteligencija

Prof.dr.sc. Dražena Gašpar

17.12.2015.

Literatura

Literatura:

1. Panian Ž., Klepac G. "Poslovna inteligencija", Masmedia, Zagreb, 2003
2. Panian Ž. i suradnici "Poslovna inteligencija: studije slučajeva iz hrvatske prakse", Narodne novine, 2007.
3. Klepac G., Mršić L. "Poslovna inteligencija kroz poslovne slučajeve", Lider press i Tim press, Zagreb, 2007.

ZAŠTO BI ?????

- Podatkovna eksplozija
- Poslovno okruženje – pritisci na organizacije
- Razvoj informacijsko komunikacijske tehnologije (ICT)

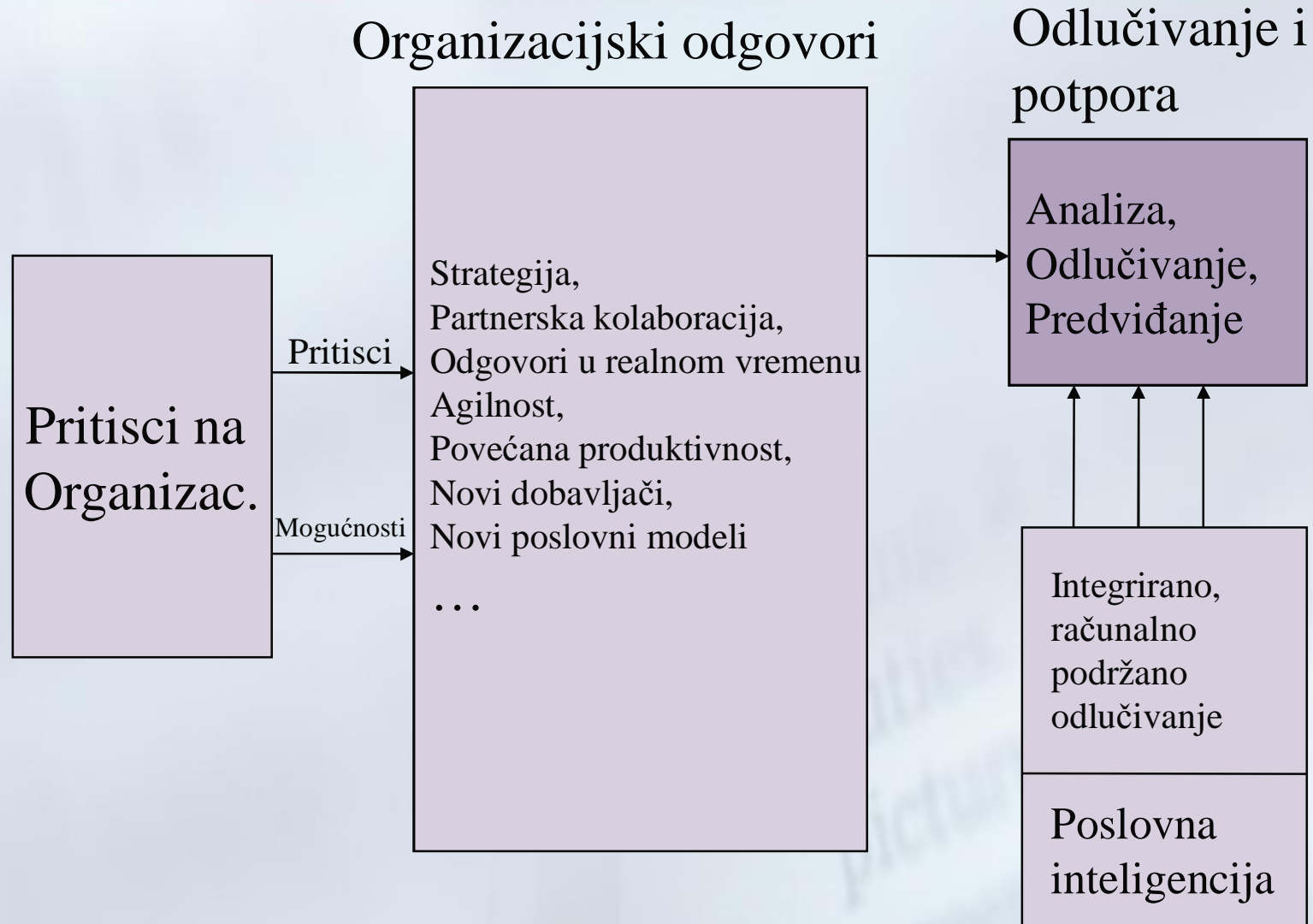
Podatkovna eksplozija

- Kvantiteta podataka, i onih koji su potrebni i digitaliziranih je enormno povećana
- Donositelji odluka trebaju pristup sve većem broju podataka i izvora podataka
- IT tj. pohranjivanje podataka omogućava izniman porast digitalno pohranjenih podataka (relativno jeftino)
- Na jednoj strani su ogromne količine podataka, dok s druge strane postoji stalni nedostatak kvalitetnih podataka

Faktori poslovnog okruženja koji kreiraju pritiske na organizacije

Tržišni	Jaka konkurencija Globalno tržište Internet i e-tržište Inovativne marketinške metode Povećane mogućnosti outsourcinga zahvaljujući IT-iju Potreba za transakcijama u realnom vremenu i na zahtjev
Zahtjevi potrošača	Traže prilagodbu Traže kvalitetu, raznolikost proizvoda, brzinu isporuke Kupci postaju sve moćniji i manje lojalni
Tehnološki	Više inovacija, novi proizvodi, usluge Povećana stopa zastarjevanja Podatkovna eksplozija (engl. <i>information overload</i>)
Društveni	Povećanje državne regulative / deregulative Raznolikija i starija radna snaga, više žena Sigurnost i teroristički napadi – primarna briga Nova zakonska regulativa za izvješćivanje Povećana društvena odgovornost kompanija

ODGOVOR NA PRITISKE ???



Poslovna inteligencija - DEFINICIJA

Stevan Dedijer (1970-tih)

"BI je korištenje kolektivnog znanja organizacije sa ciljem postizanja konkurentske prednosti“.

Razvoj IT-ija dovodi do definicije:

BI je sposobnost organizacije da pribavi informacije, istraži ih, stekne uvid i razumijevanje što dovodi do boljeg donošenja odluka.

Poslovna inteligencija - DEFINICIJA

BI je široka kategorija aplikacijskih programa i tehnologija za prikupljanje, pohranjivanje, analiziranje i omogućavanje pristupa podacima kako bi se pomoglo poslovnim korisnicima da donesu bolje poslovne odluke. BI aplikacije obuhvaćaju aktivnosti potpore odlučivanju, upite i izvješća, OLAP, statističke analize, predviđanja, data mining.

www.sauder.ubc.ca/cgs/itm/itm_glossary.html

BI je popularni, objedinjavajući pojam koji se koristi za opis skupa koncepata i metoda za unapređenje donošenja poslovnih odluka, uz uporabu na činjenicama baziranim sustavima potpore.

www.noisebetweenstations.com/personal/essays/metadata_glossary/metadata_glossary.html

BI se obično opisuje kao rezultat opsežne analize detaljnih poslovnih podataka. Sadrži baze podataka i aplikacijske tehnologije, kao i analize prakse. Ponekad se koristi kao sinonim za "potporu odlučivanju" iako je BI tehnički mnogo širi pojam, koji eventualno obuhvaća i upravljanje znanjem, planiranje resursa poduzeća, data mining i druge prakse.

it.csumb.edu/departments/data/glossary.html

Poslovna inteligencija - DEFINICIJA

BI objedinjava metodologije, tehnologije i platforme za skladištenje podataka (*Data Warehousing*), OLAP procesiranje podataka (*On-line Analytical Processing*) i rudarenje podataka (*Data Mining*) koje omogućavaju tvrtkama kreiranje korisnih upravljačkih informacija iz podataka o poslovanju koji se nalaze disperzirani na različitim transakcijskim sustavima te dolaze iz različitih internih i eksternih izvora.

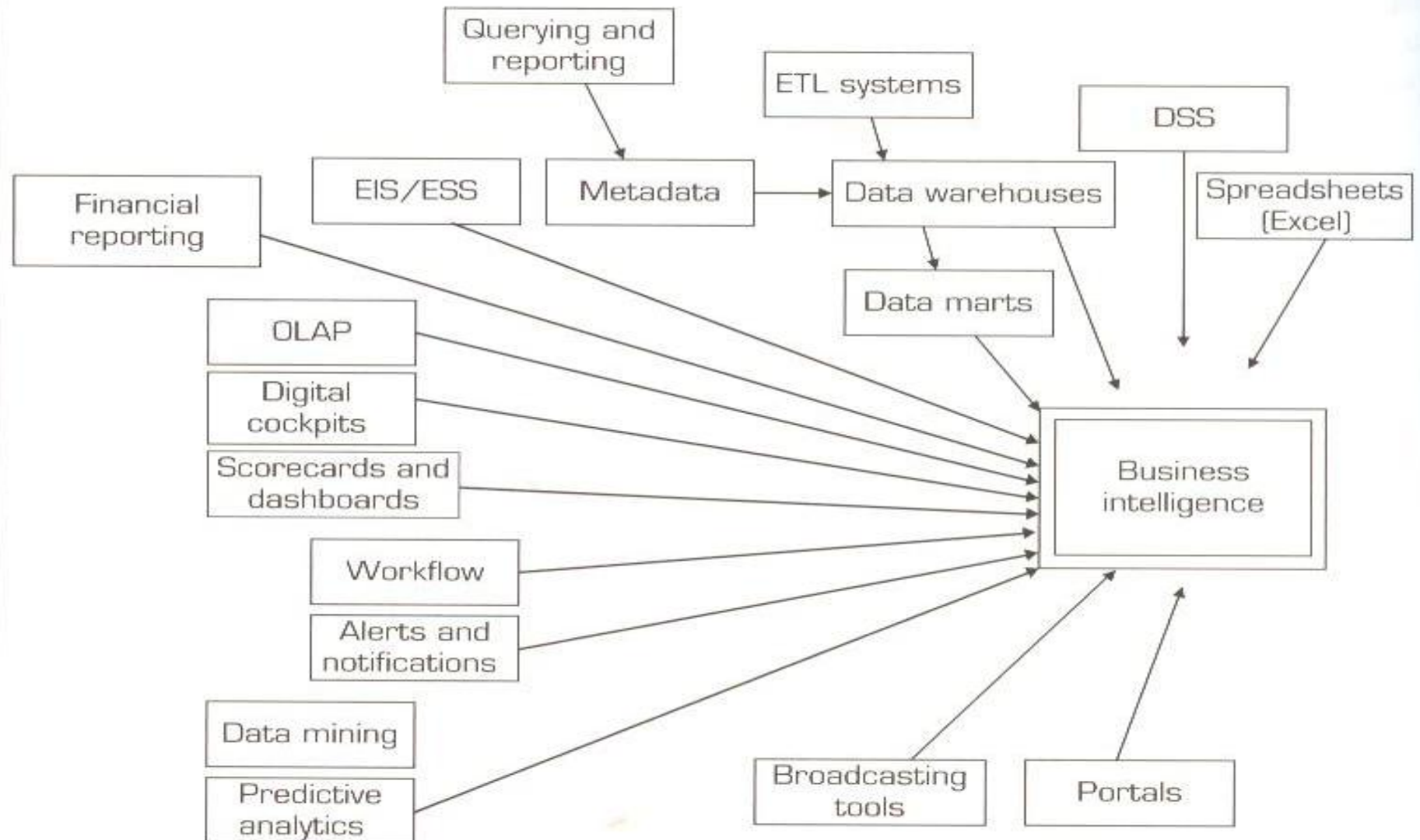
BI je znanje izvedeno na temelju analize organizacijskih informacija.

planning.ucsc.edu/irps/dwh/DWHGLOSS.HTM

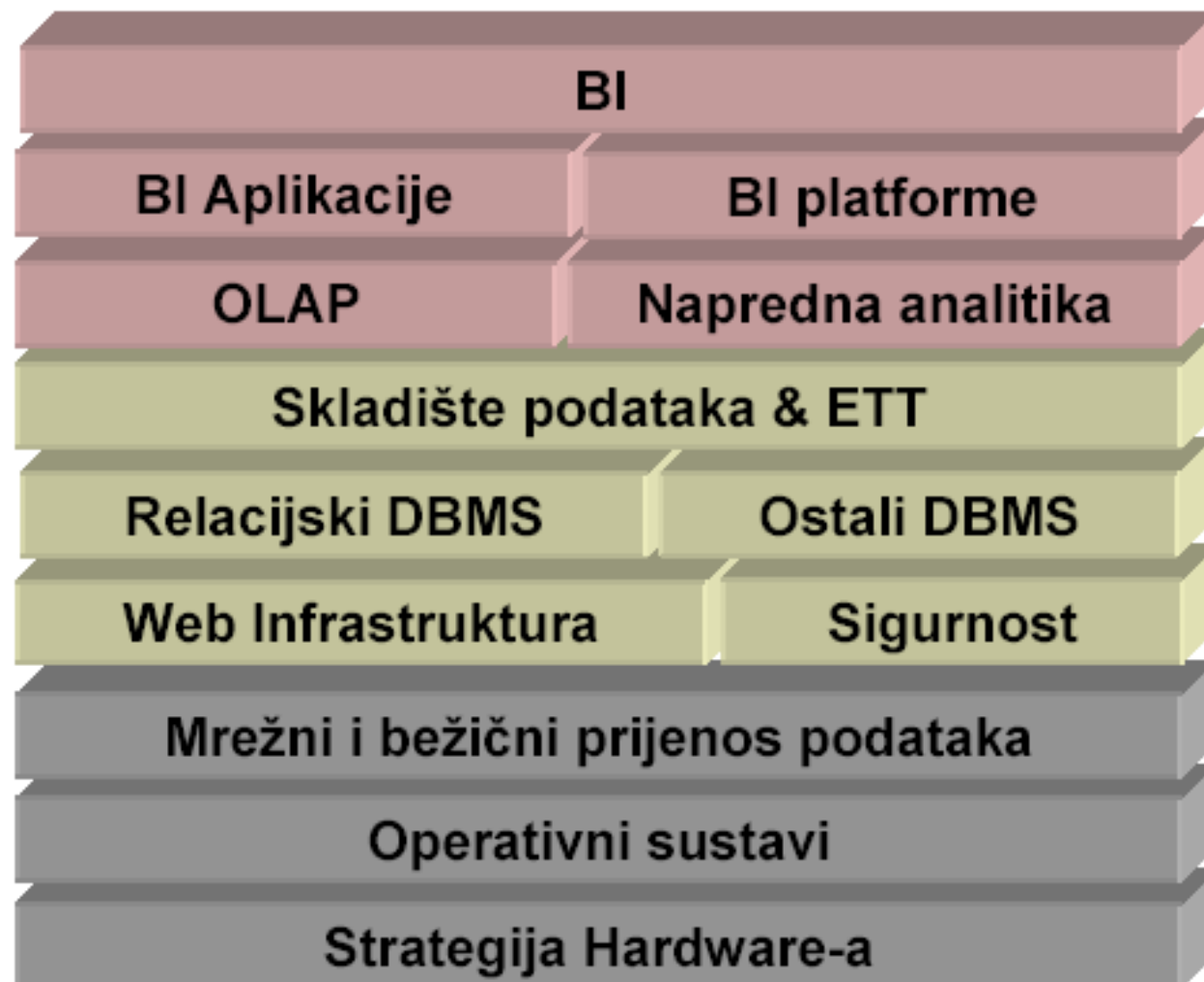
BI je proces prikupljanja informacija iz područja poslovanja. Može biti opisana kao proces transformiranja podataka u informacije, a zatim u znanje. BI ima za cilj ostvariti značajnu konkurentsku prednost i ostvariti vrijedne ključne kompetence u pojedinim slučajevima.

en.wikipedia.org/wiki/Business_intelligence

Evolucija BI



Poslovna inteligencija - infrastruktura



Poslovna inteligencija

BI omogućuje organizacijama sustavno promoviranje kulture razumijevanja i poduzimanja akcija kroz:

- Donošenje odluka bazirano na činjenicama
- Kvalitetu informacija
- Smislenost oblika informacija
- Kvantitetu informacija
- Dijeljenje informacija

Poslovna inteligencija

- Donošenje odluka bazirano na činjenicama

Vizija BI promovira zaokret od intuitivnog pristupa i spekulativnog nagađanja prema donošenju odluka baziranog na činjenicama.

Bolja iskorištenost ljudi i informacija – poboljšana profitabilnost, povećanje tržišnog udjela i unapređenje odnose s kupcima i korisnicima usluga.

Poslovna inteligencija

- Kvaliteta informacija

BI uporabljiv onoliko koliko su kvalitetni podaci na kojima se temelji.

Kvaliteta podataka podrazumijeva njihovu kompletnost, točnost, konzistentnost i pravovremenost.

Sve dostupne i vezane informacije (interne i eksterne) trebaju biti osigurane kako bi se omogućila šira perspektiva.

Informacije iz različitih izvora trebaju biti sintetizirane i homogenizirane, sa zajedničkim definicijama kako bi osigurale jedinstvenu polaznu osnovu za sve korisnike.

Poslovna inteligencija

- Smislenost oblika informacija

BI je usmjeren prema korisniku:

- informacije i analize se prezentiraju korisniku na način koji odražava njegovo shvaćanje vođenja poslovanja.
- nije dovoljno samo osigurati informacije, već to mora biti ostvareno i prezentirano na korisniku jednostavan način

Poslovna inteligencija

- Kvantiteta informacija

BI - zaokret od specijaliziranih analitičara prema menadžerima i stručnjacima jer oni bolje shvaćaju dinamiku poslovanja.

BI se koristi za interpretaciju povijesnih podataka, predviđanje budućih trendova kao i za određivanje i mjerenje pokazatelja poslovanja.

Istraživanjem podataka iz unutarnjih i vanjskih izvora, korisnici otkrivaju kretanje trendova kod klijenata i na tržištu te ih pretaču u nove poslovne prilike i uštede.

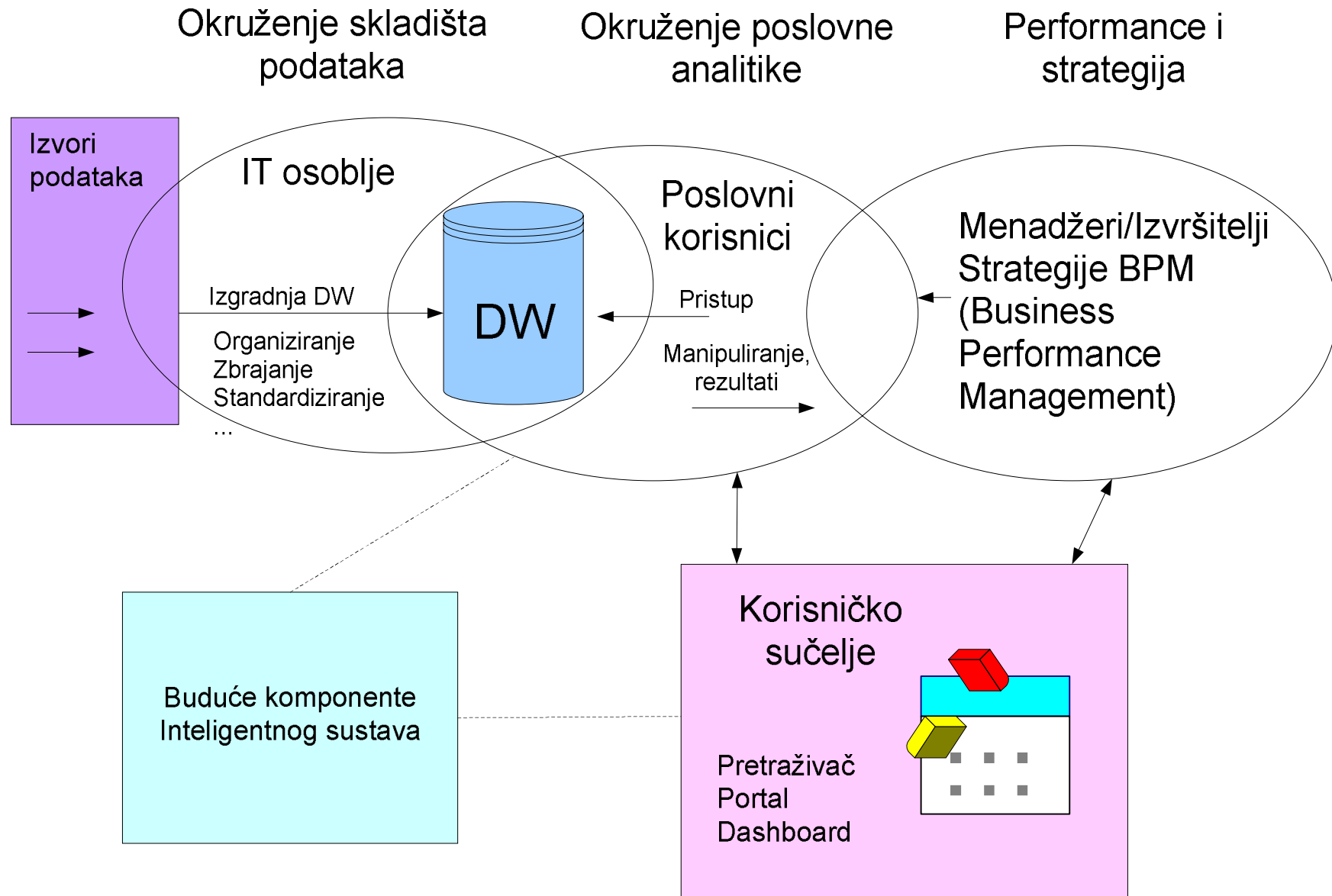
Poslovna inteligencija

- Dijeljenje informacija

Bi promovira aktivno dijeljenje informacija, pogleda, veza, pretpostavki, aplikacija i uvida na stanje unutar i izvan poduzeća.

Dijeljenje informacija reducira redundanciju poslova, a svi korisnici profitiraju od svojih rezultata, ali i od rezultata svojih kolega.

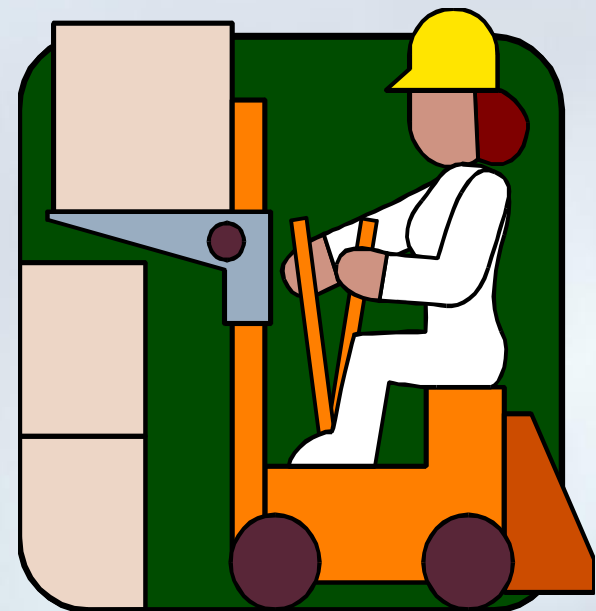
BI arhitektura



Pretpostavke uspješne primjene BI

- Razumijevanje korisnika
- Primjena paradigme broja klikova
- Napredniji korisnici trebaju biti i konzumenti i proizvođači informacija
- Uspostava i unapređenje kulture mjerenja
- Razvoj BI treba biti strateško opredjeljenje tvrtke

Tema: Skladišta podataka



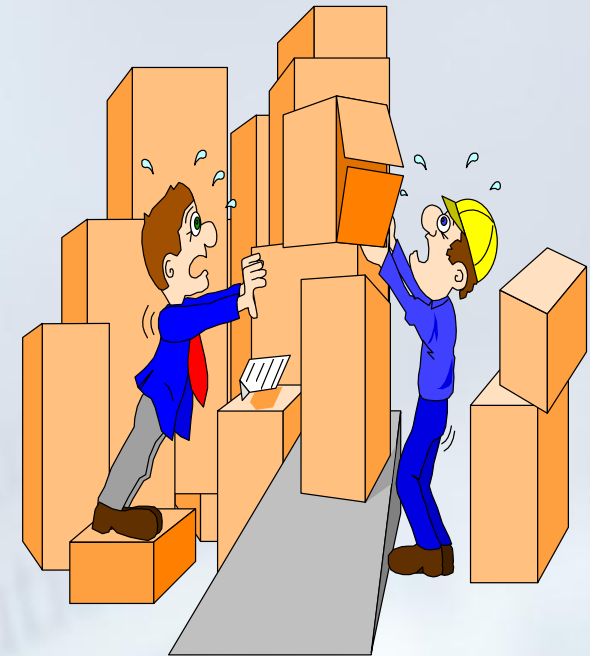
Podaci svuda, ali



- Korisnik ne može pronaći podatke kada ih treba
 - Podaci razasuti po mreži
 - Previše verzija, nejasne razlike
- ⌘ Korisnik ne može dobiti podatke kada ih treba
 - ☑ potreban ekspert da se dođe do podataka
- ⌘ Korisnik ne razumije podatke koje je pronašao
 - ☑ raspoloživi podaci slabo dokumentirani
- ⌘ Korisnik ne može koristiti pronađene podatke
 - ☑ Neočekivani rezultati
 - ☑ Potrebno transformiranje podataka iz jednog oblika u drugi

Što korisnici kažu...

- Podaci bi trebali biti integrirani u okviru poduzeća
- Zbrojni podaci imaju stvarnu vrijednost za organizaciju
- Povijesni podaci su ključni za razumijevanje podataka tijekom vremena
- Potrebne su mogućnosti tipa Što-ako analize

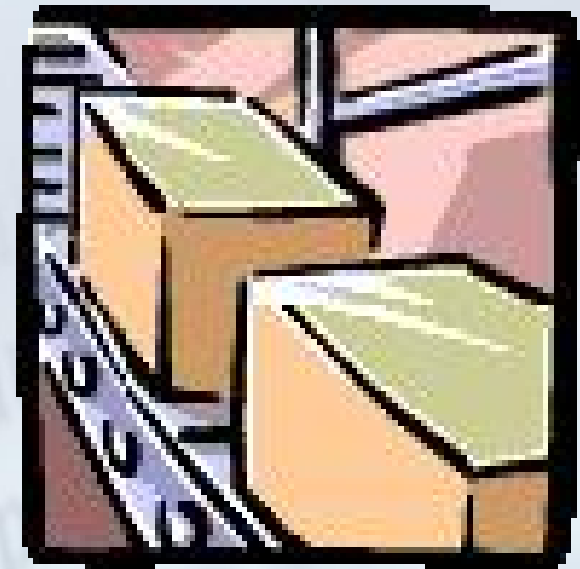


Definicija Inmon-a

Skladište podataka je predmetno orijentirani, integrirani, relativno stabilni i vremenski orijentirani skup podataka u funkciji potpore odlučivanja menadžera.

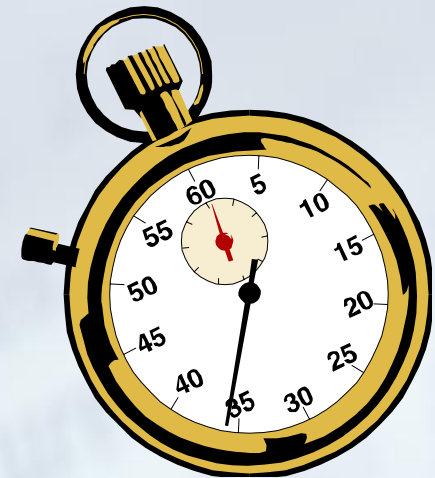
Zašto ne operativni sustavi?

- OLTP – OnLine Transaction Processing
- Izvode kritične poslovne aplikacije
- Traže visoku razinu performansi
- Koriste se za vođenje poslovanja!



Operativni sustavi

- Vode poslovanje u realnom vremenu
- Optimizirani za rad s velikim brojem jednostavnih čitanje/pisanje transakcija
- Optimizirani za brze odgovore na predefinirane transakcija
- Koriste ih ljudi koji rade s kupcima, proizvodima – službenicu, prodavači ...



RDBMS se koriste za OLTP

- Baze podataka se tradicionalno koriste za OLTP
 - Operativni zadaci i podaci
 - Detaljni, ažurni podaci
 - Strukturirani, ponavljajući zadaci
 - Čitanje/ažuriranje nekoliko slogova
 - Izolacija, oporavak i integritet su kritična pitanja

2 osnovna tipa podataka

Primitivni/Operativni podaci	Izvedeni/DSS podaci
Aplikacijski orijentirani	Predmetno orijentirani
Detaljni	Sumarni, redefinirani
Mogu se mijenjati	Ne mogu se mijenjati
Izvršavanje je kontinuirano	Izvršavanje je heurističko
Zahtjevi za obradu su unaprijed poznati	Zahtjevi za obradu nisu unaprijed poznati
Transakcijski orijentirani	Analitički orijentirani
Visoka raspoloživost	Raspoloživost nije kritična

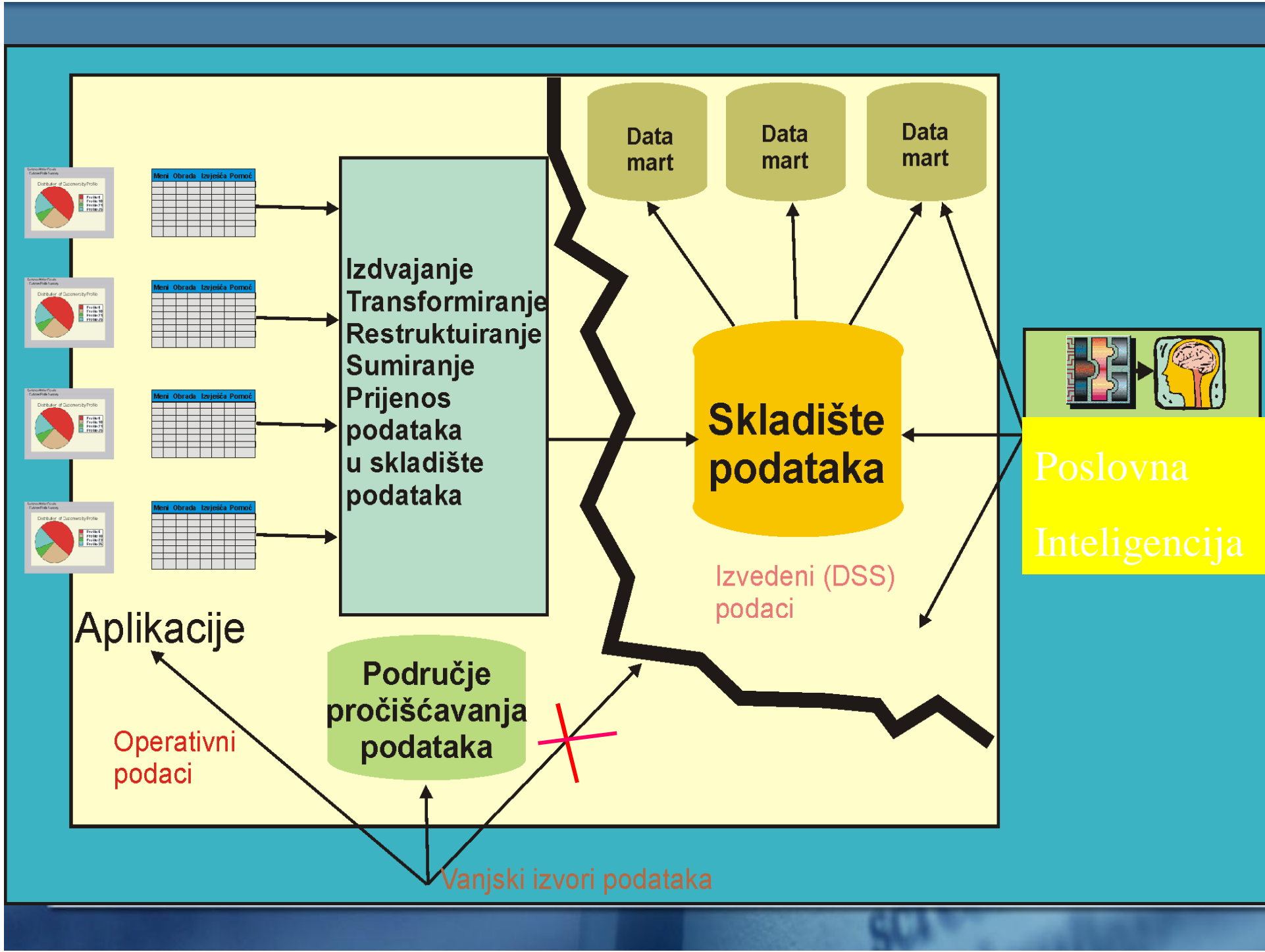
2 osnovna tipa podataka

Primitivni/Operativni podaci	Izvedeni/DSS podaci
Cjelovitost održavanja	Održavanje po podskupovima
Neredundantni	Redundancija kao životna istina
Struktura statična, sadržaj promjenjiv	Fleksibilna struktura
Obrade koriste manju količinu podataka	Obrade koriste ogromne količine podataka
Podržavaju dnevne operacije	Podržavaju potrebe menadžera
Odnose se na kraći vremenski period (najčešće jednu godinu)	Nemaju vremensku ograničenost, prate povijest podataka u poduzeću

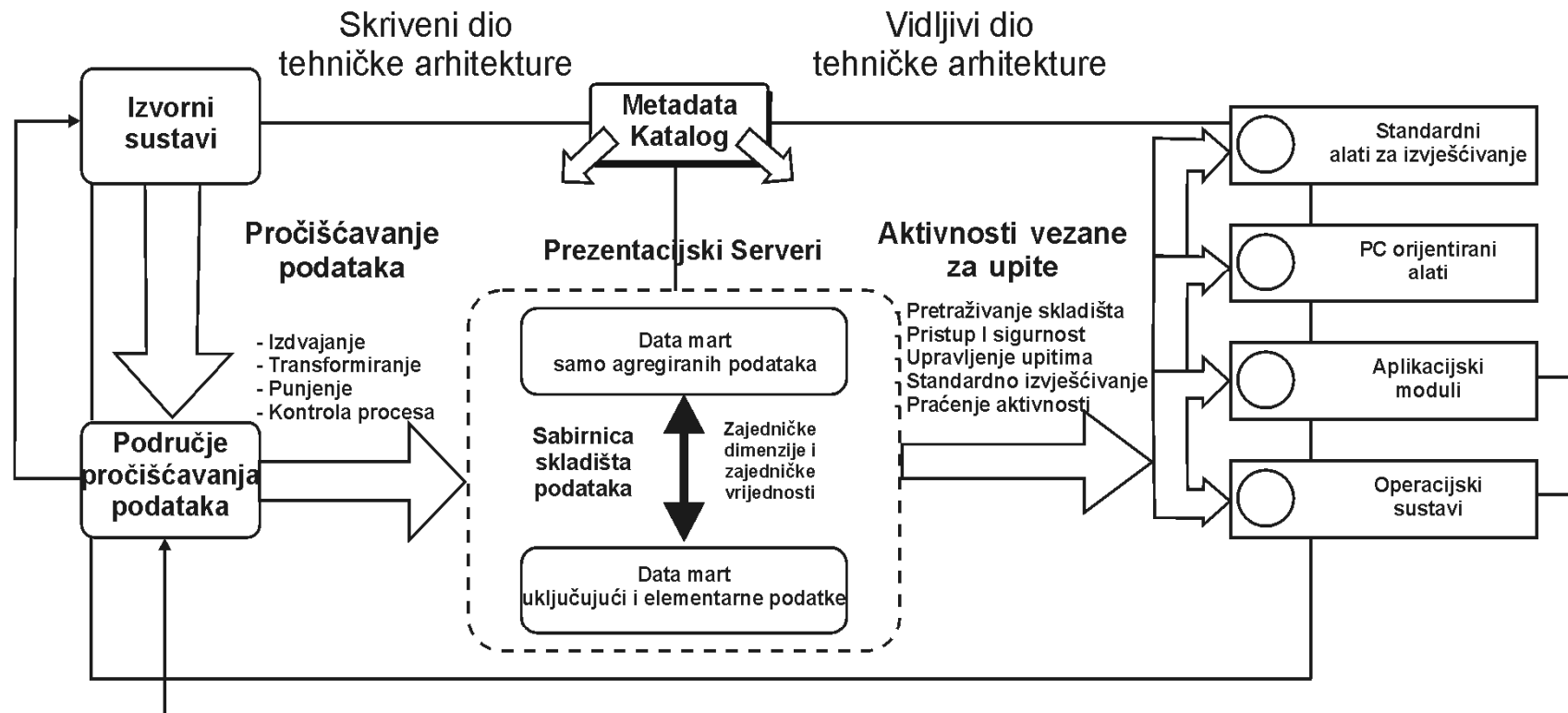
- OLTP sustavi “vode” poslovanje



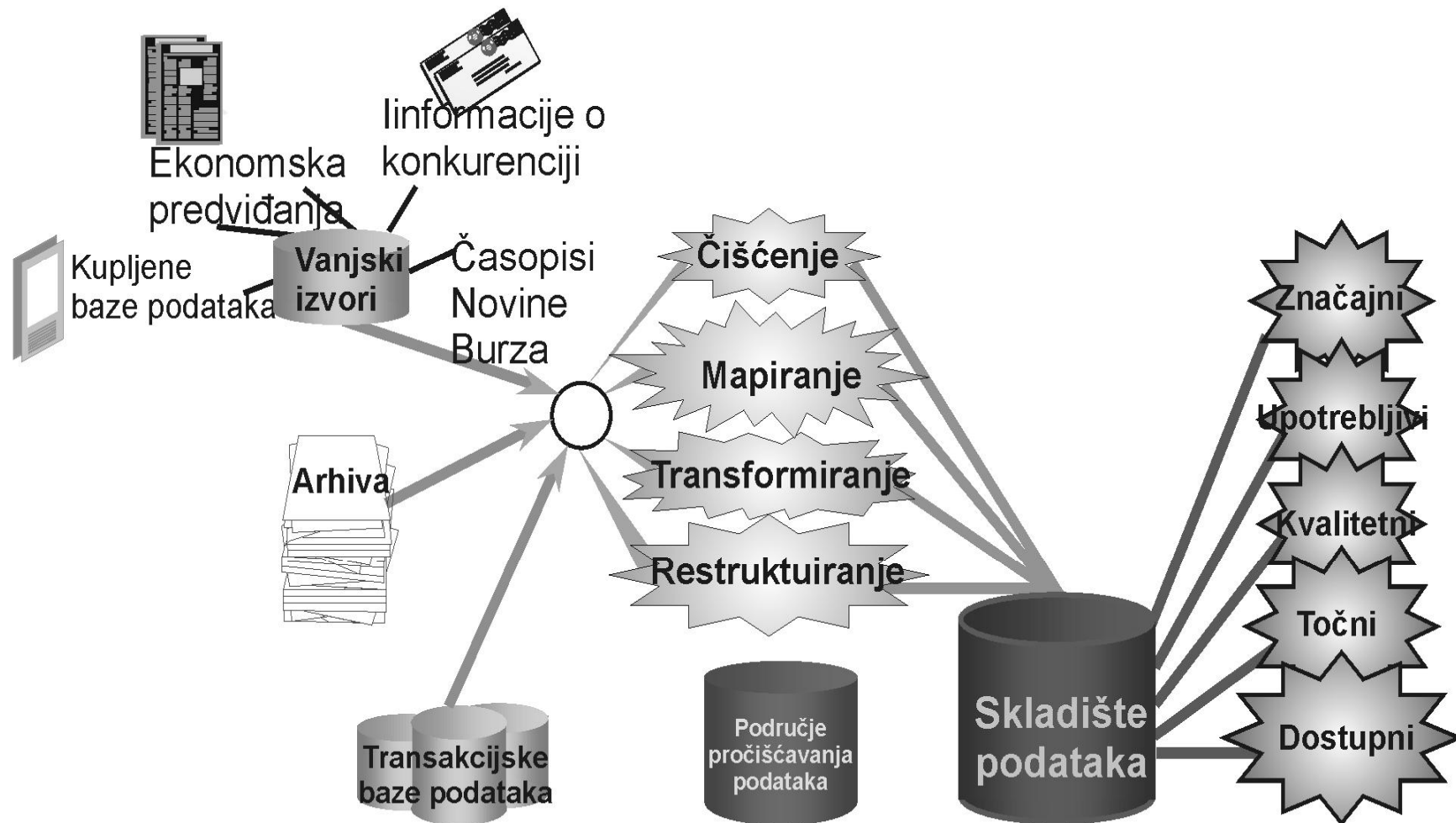
- Skladište podataka pomaže u “optimiziranju” poslovanja



Arhitektura



Područje pročišćavanja podataka



Područje pročišćavanja podataka

ETL (Extraction Transformation Loading)

- Otkrivanje promjena u izvornim podacima potrebnim za skladište podataka;
- Izdvajanje podataka iz izvornih sustava;
- Čišćenje i transformiranje podataka;
- Restrukturiranje ključeva podataka;
- Indeksiranje podataka;
- Sumiranje podataka;
- Održavanje metapodataka;
- Učitavanje podataka u skladište podataka.

Stari/novi koncepti i tehnologije

- Metapodaci
- Dimenzijsko modeliranje
- Data mining

Metapodaci

- Podaci o podacima.
- Opisuju sadržaj (kontekst) i strukturu podataka.

Razlikuju se:

- > tehnički metapodaci
- > poslovni metapodaci

Metapodaci

Promatranje metapodataka kao obrazaca (engl. pattern):

- Sintaksni metapodaci

opisuju sintaksu podataka

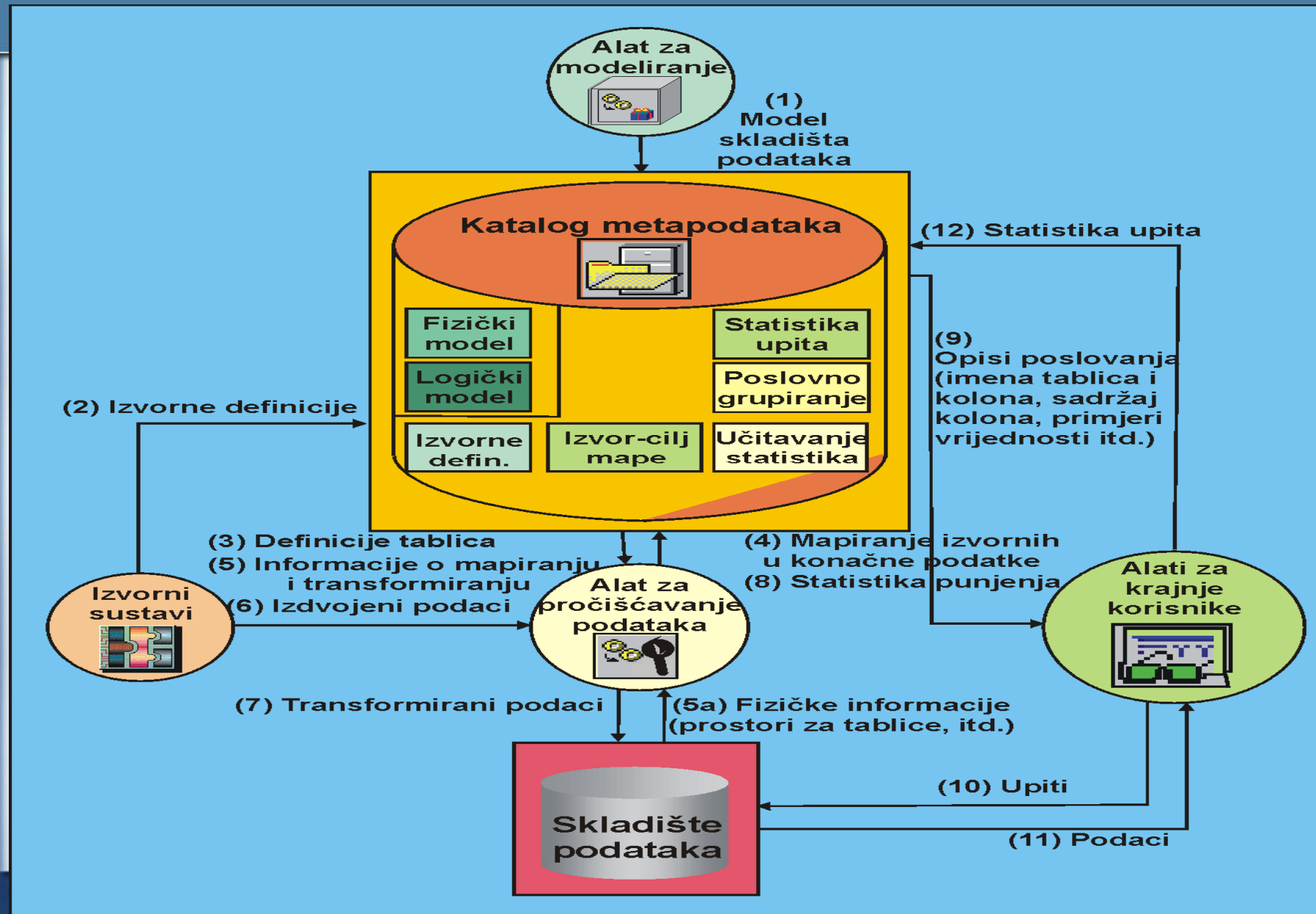
- Strukturni metapodaci

opisuju strukturu podataka

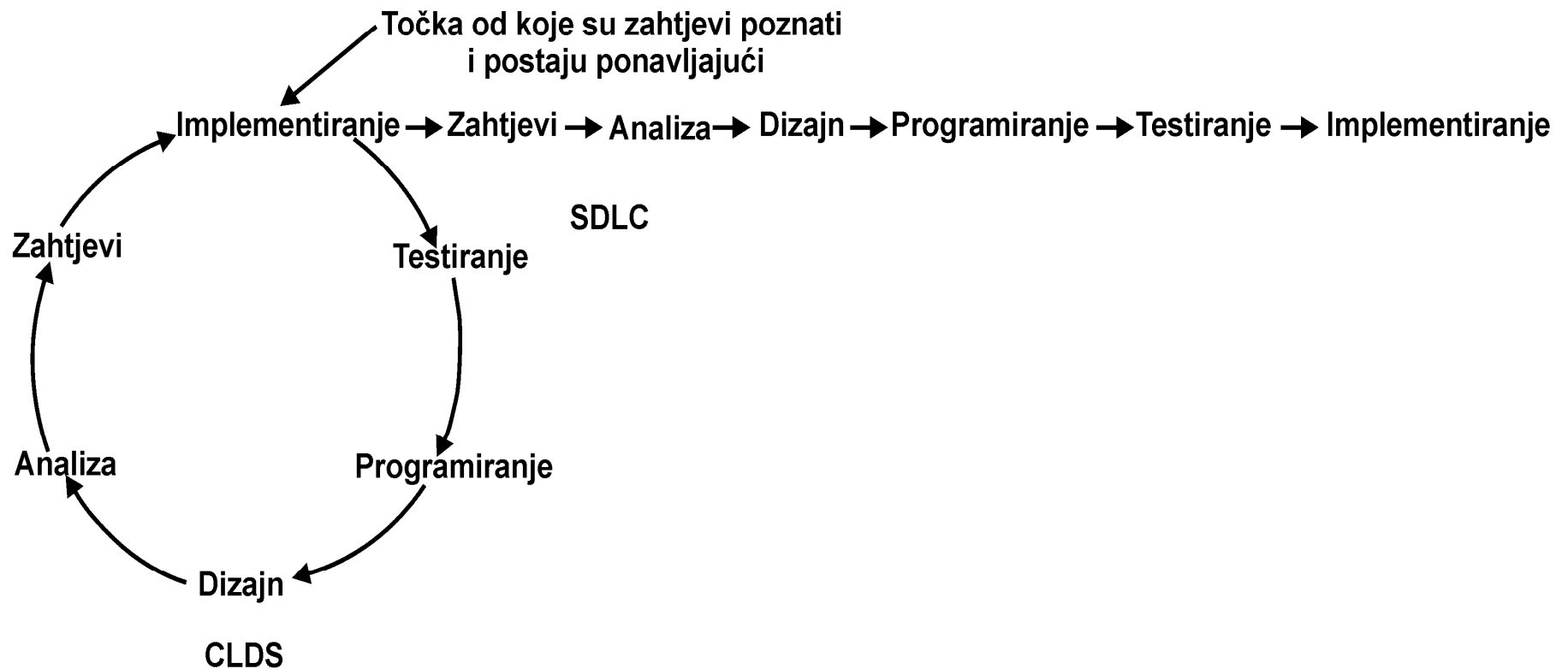
- Semantički metapodaci

opisuju značenje podataka u specifičnoj domeni

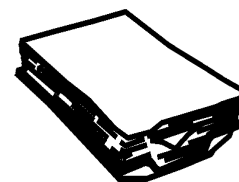
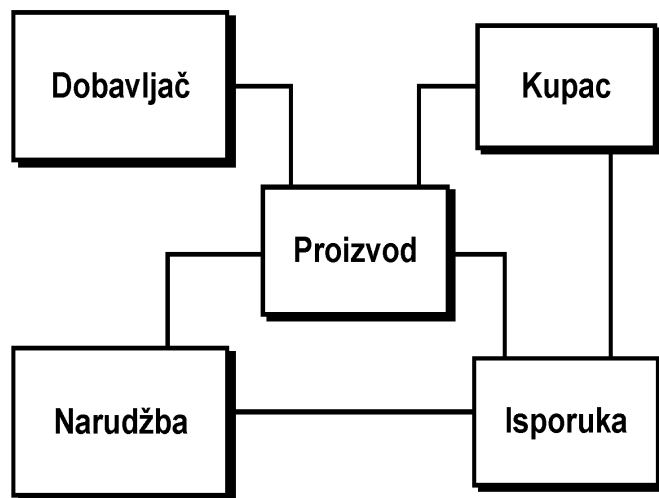
Katalog metapodataka – Holistički pristup



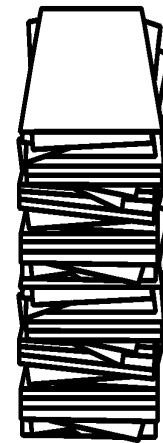
Razvojni ciklus skladišta podataka



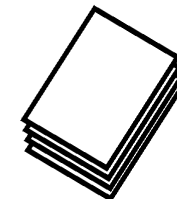
Dizajn baze za skladište podataka



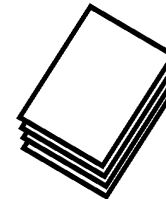
Dobavljač



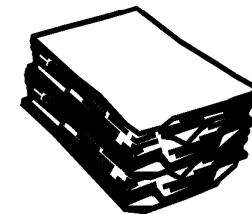
Narudžba



Isporuka



Kupac



Proizvod

Što je dimenzijsko modeliranje (DM)?

- DM je tehnika logičkog dizajna koja pokušava predstaviti podatke na standardan, intuitivan način koji omogućava visoku razinu performansi pristupa.
- Može se implementirati koristeći relacijsku ili višedimenzijsku bazu podataka.
- Svaki dimenzijski model se sastoji od jedne tablice sa složenim ključem – tablice činjenica (vrijednosna tablica, činjenična tablica, fakt tablica) i skupa manjih tablica – dimenzijskih tablica.
- Svaka dimenzijska tablica ima jednostavni primarni ključ koji odgovara točno jednom dijelu složenog ključa iz fakt tablice.
- Grafički prikaz ovog modela podsjeća na zvijezdu, pa se ova struktura naziva zvijezda spajanje (zvijezda shema).

Zvijezda shema

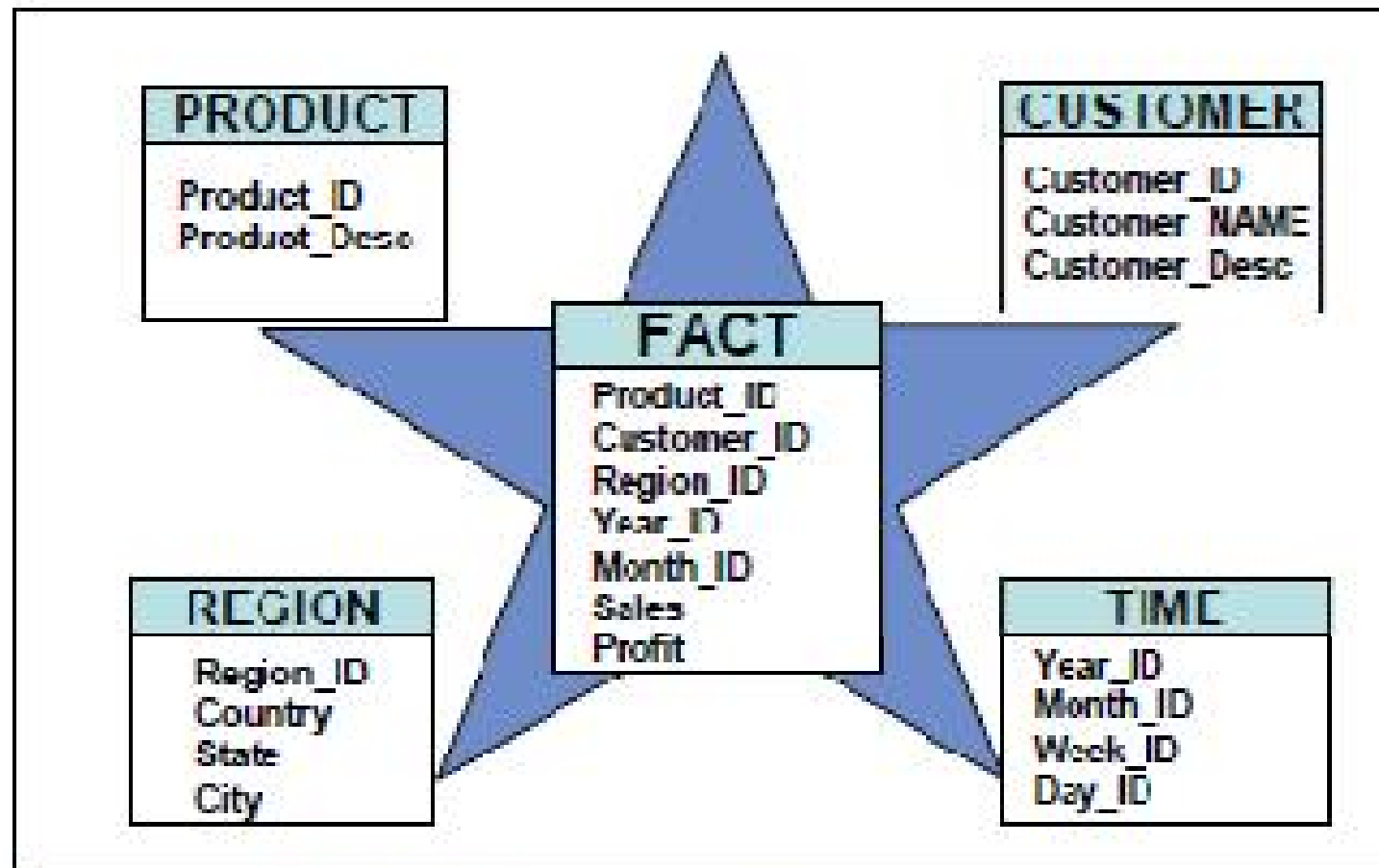


Figure 3-3 A star schema or a dimensional model

Fakt (vrijednosna) tablica

- Fakt tablica pohranjuje mjere – vrijednosti poslovanja i pokazuje na vrijednost ključa na najnižoj razini svake dimenzijske tablice. Mjere su kvantitativni ili činjenični podaci o predmetu.
- Mjere su općenito numeričke i odgovaraju na pitanje *koliko?*
- Primjeri mjera su: cijena, prodaja proizvoda, inventurna količina, prihod i sl. Mjera se može odnositi na stupac u tablici ili može biti izračunata.

Fakt tablica

- Najkorisnije činjenice u fakt tablici su numeričke i po njima se može zbrajati.
- Zbrajanje je bitno iz razloga što DW aplikacije skoro nikada ne rade s jednim retkom fakt tablice, već obično rade sa stotinama, tisućama ili čak milijunima slogova (redaka) u jednom trenutku. Vrlo često, najskorisnije što se može učinit s tolikom brojem slogova je zbrojiti ih.

Fakt tablica

- Prije dizajna fakt tablice – usitnjenost (*granularnost*) fakt tablice mora biti određena.
- Usitnjenost odgovara definiciji pojedinačnog sloga najniže razine u toj fakt tablici.
- Usitnjenost se može odnositi na pojedinačnu transakciju, dnevni snapshot ili mjesečni snapshot.

Dimenzijska tablica

Dimenzija predstavlja jedan skup objekata ili događaja u stvarnom svijetu.

Svaka dimenzija koju korisnik identificira za model podataka implemetira se kao dimenzijska tablica.

Dimenzije su kvalifikatori koji daju značenje mjerama iz fakt tablice, zato što daju odgovor na što, koji, kada, koliko i gdje pitanja. Na primjer:

Koji *kupci* su ostvarili najveći promet prošle godine?

Koliki je bio naš profit po *dobavljačima*?

Koliko je prodano jedinica svakog *proizvoda*?

Dimenzijska tablica

- Dimenzijska tablica sadrži informacije o tome kako organizacija želi analizirati činjenice:
 - “Prikazati prodaju (činjenica) za prošli tjedan (vrijeme) za crvene šalice (proizvod) u zapadnoj hercegovini (zemljopisni pojam)”
- Dimenzijska tablica najčešće sadrži opisne tekstualne informacije “crvene šalice”, “zapadna hercegovina”
- Dimenzijski atributi se koriste kao izvor najzanimljivijih “ograničenja” u DW upitima, oni su uvijek zaglavlja redaka u SQL odgovoru na upit.

Dimenzije

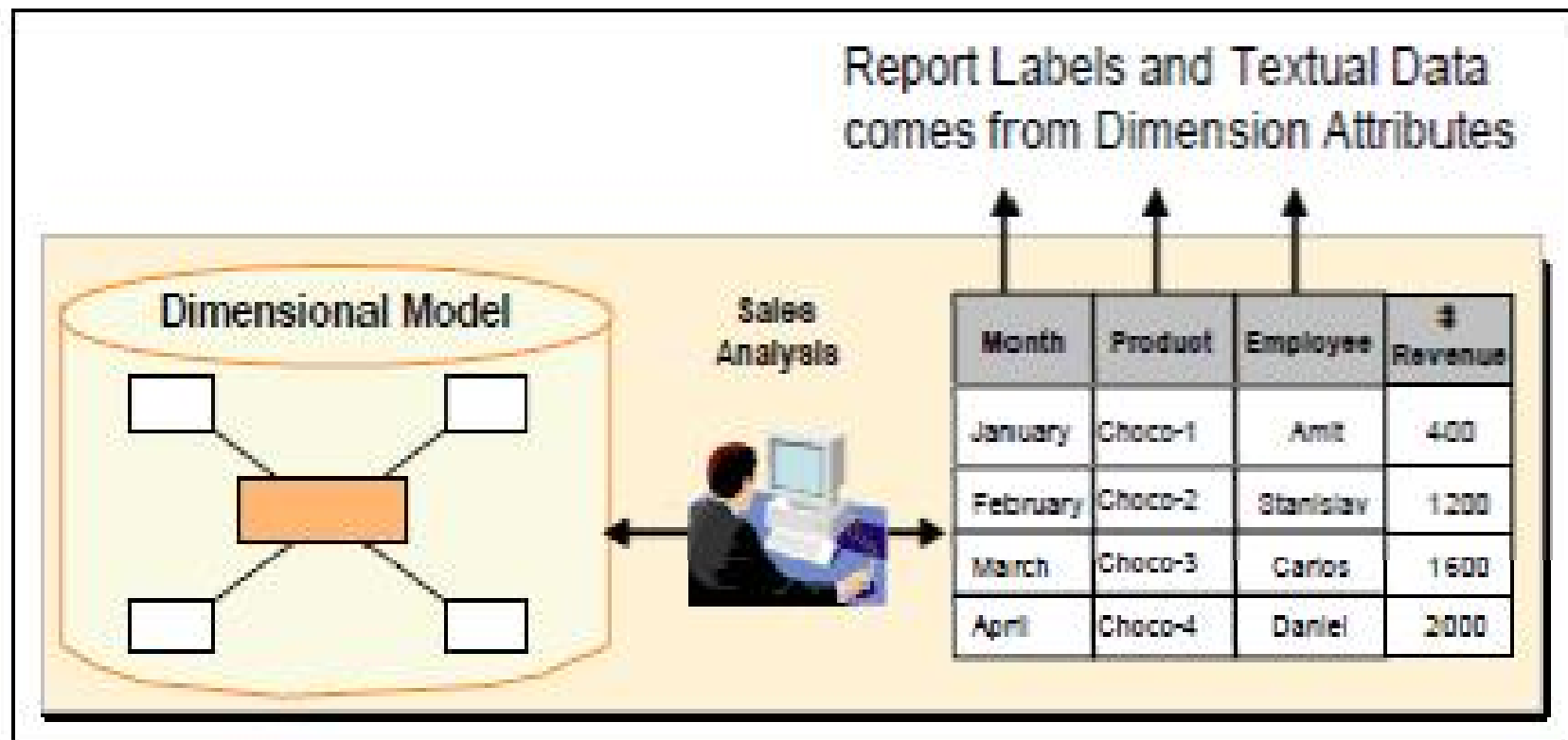


Figure 3-6 The textual data in the report comes from dimension attributes

Dimenzijski model podataka

Osnova za razumijevanje razlike između relacijskog i dimenzijskog modela podataka jeste način pohranjivanja i predstavljanja podataka.

Podaci su u relacijskoj bazi podataka pohranjeni u obliku tablica gdje svaki red tablice predstavlja jedan slog, a svaki slog je podijeljen na polja (stupci u tablici) u kojima su pohranjeni podaci.

U multidimenzijskoj bazi podataka osnovni način predstavljanja podataka jeste matrica podataka ili ukrštena tablica koja čini dimenzijsko područje (engl. *array*). Područje je osnovna komponenta multidimenzijske baze podataka.

Dimenzijski model podataka

Multidimenzijsko područje predstavlja višu organizacijsku razinu od relacijske tablice. Ovakva struktura samo po sebi sadrži vrijednu “inteligenciju” s obzirom na odnose između podataka jer su određene “perspektive” izravno ugrađene u strukturu kao dimenzije, umjesto da se pohrane u polja. Ovakva “inteligentna” struktura prezentira informacije krajnjem korisniku na znatno organiziraniji način, osiguravajući usput i značajne prednosti u performansama izvršavanja zahtjeva. Naime, za razliku od relacijske strukture koja mora proći kroz svaki pojedinačni slog kako bi pronašla sve odgovarajuće slogove, multidimenzijska struktura “presreće upite na pola puta” time što predstavlja podatke na unaprijed organiziran način.

Dimenzijski model podataka

Relacijski prikaz

Model	Boje	Prodaja
Karavan	Plav	6
Karavan	Crven	5
Karavan	Bijel	4
Sportski	Plav	3
Sportski	Crven	5
Sportski	Bijel	5
Limuzina	Plav	4
Limuzina	Crven	3
Limuzina	Bijel	2

Multidimenzijski prikaz

Dimenzija

MODEL

Karavan	6	5	4
Sportski	3	5	5
Limuzina	4	3	2

Plava Crvena Bijela

BOJA

Dimenzija

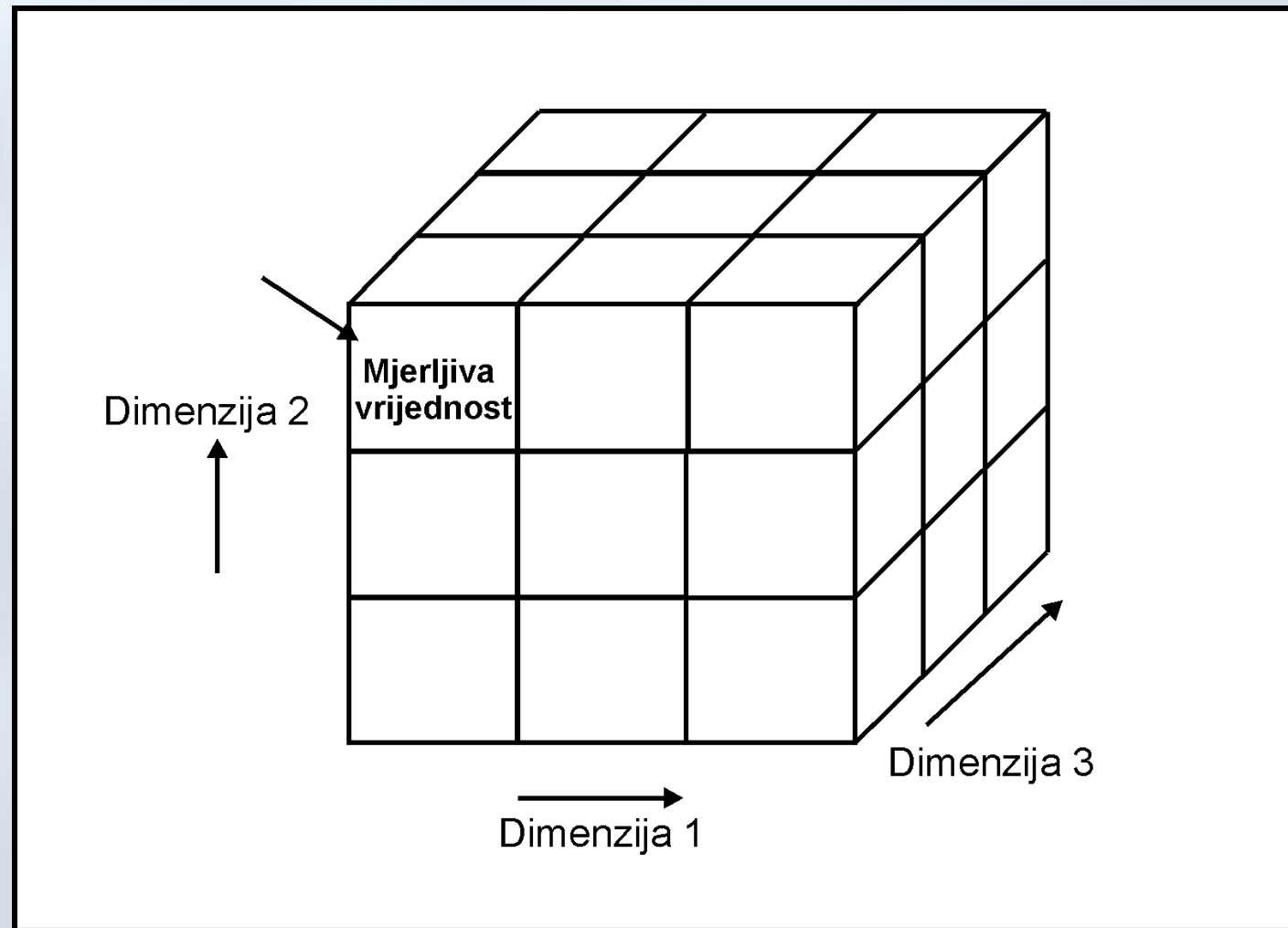
PRODAJA

Dimenzijski model podataka

Osnovna ideja multidimenzijske baze podataka jeste da se skoro svaki tip poslovnih podataka može predstaviti kao dio *kocke* (engl. *cube*) podataka gdje ćelije kocke sadrže mjerljive vrijednosti, a rubovi kocke određuju prirodne dimenzije podataka.

Podrazumijeva se više od tri dimenzije u modeliranju tako da bi se kocka trebala zvati *hiperkocka*, iako su termini *kocka* i *podatkovna kocka* postali uobičajeni.

Dimenzijski model podataka



Tipični operatori

- Roll up (drill-up): zbrajanje (agregiranje) podataka
 - *Po hijerarhiji prema gore ili po dimenziji*
 - Drill down (roll down): suprotno od roll-up [svrdlanje]
 - *od više razine agregiranja prema nižoj tj. prema detaljnim podacima, ili uvođenje novih dimenzija*
- Slice and dice [raslojavanje i presijecanje]:
 - *Projekcija i selekcija*
- Pivot (rotate) [rotiranje] :
 - *Zaokretanje kocke, vizualizacija, 3D prema seriji 2D.*

Drill down i roll up

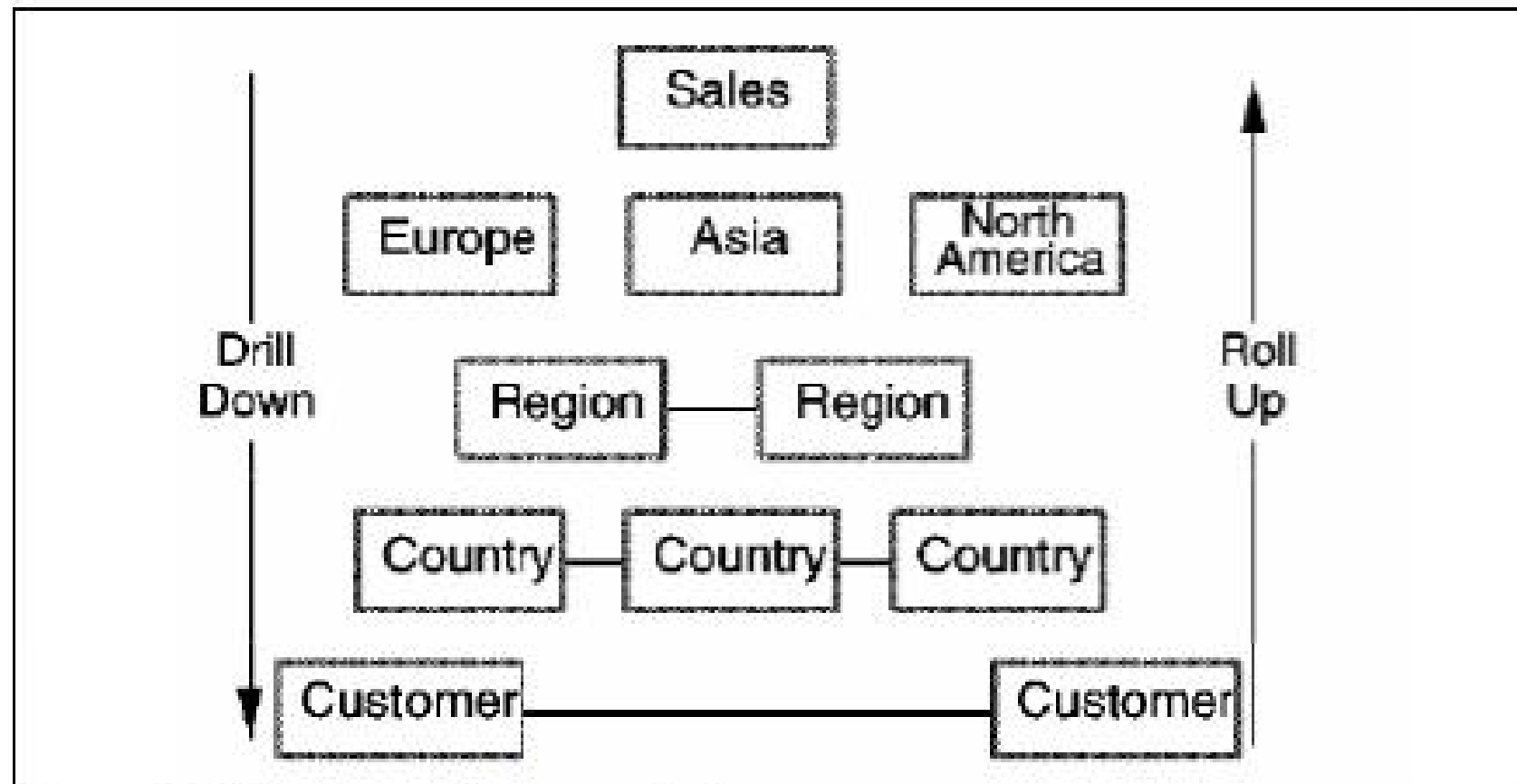
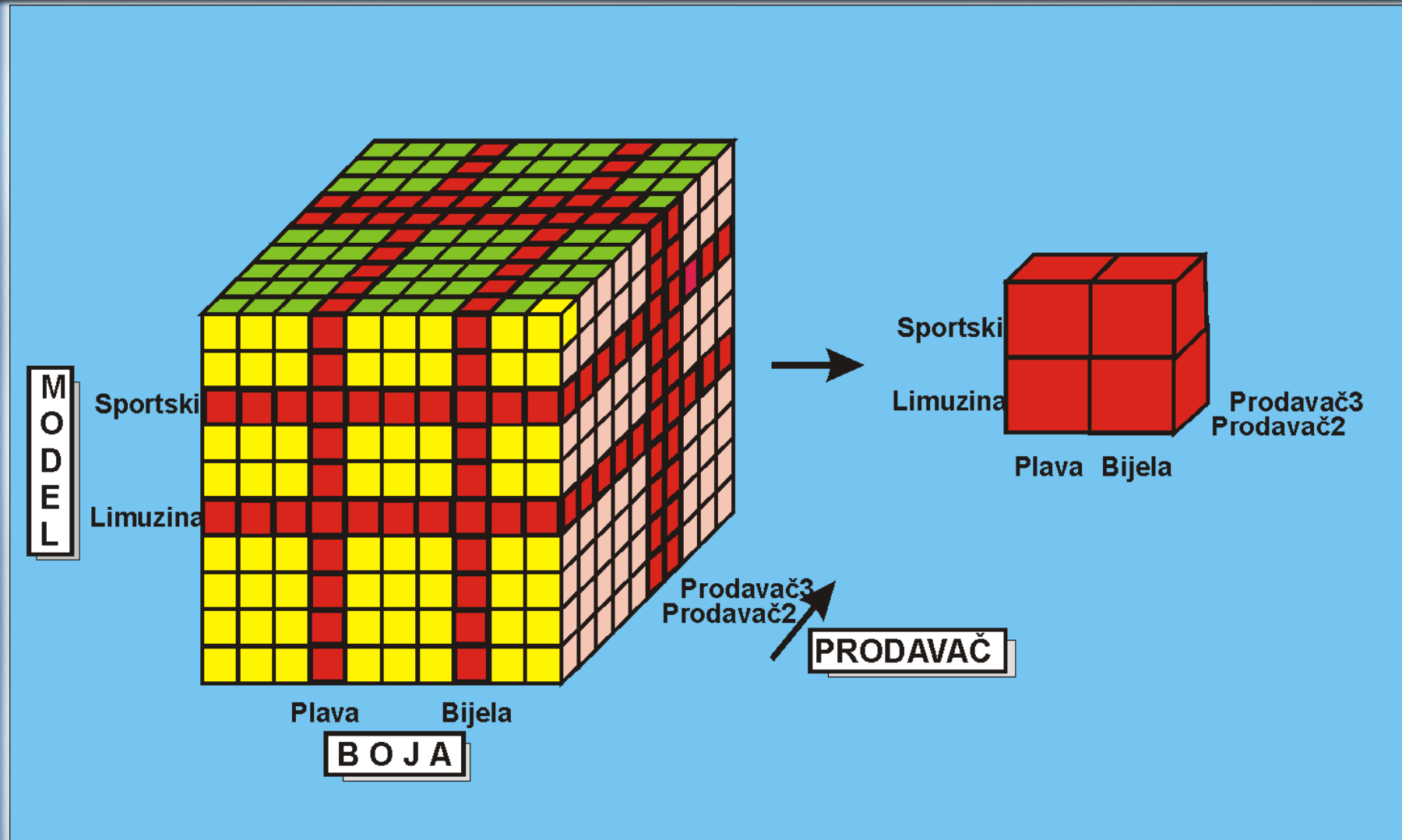


Figure 4-7 Drill-down and roll-up analysis

Slice and dice [raslojavanje i presijecanje]

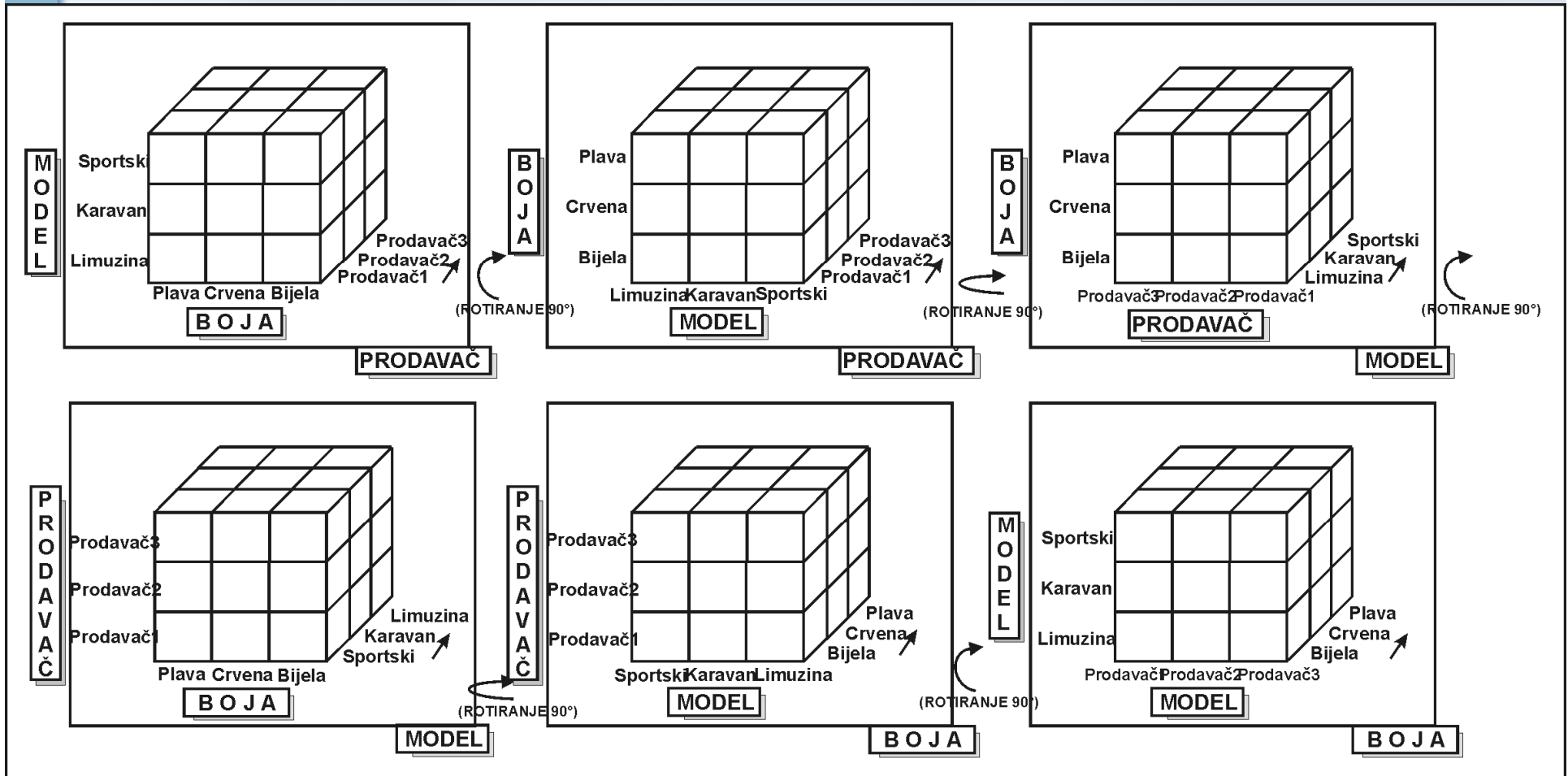


Pivoting [rotiranje]

Pivoting – zamjena stupaca i redaka

Analiza istih podataka iz različitih perspektiva

Pivoting [rotiranje]



Dimenzijski model podataka

Prodaja

M Sportski	6	5	4
O Karavan	3	5	5
E Limuzina	5	3	2

Crvena Plava Bijela
BOJE

Starosna dob zaposlenog

Jurković				21					
Štern								19	
P Štencel	63								
e Jurić					31				
Z Zovko						27			
i Marić							56		
m Bošnjak		45							
e Rener									41
Karačić			19						
	31	41	23	01	14	54	03	12	33

Zaposlen #

Dimenzijski model podataka

Prednosti:

- Poslovni način promatranja podataka kroz dimenzije i vrijednosti
- Jednostavnija automatizacija – dio znanja o podacima čini sastavni dio modela

Ograničenja:

- Potrebna je korelacija između podataka
- Broj dimenzija koje ljudski um može pojmiti je ograničen

Korištenje skladišta podataka



Istraživači, Farmeri i Turisti

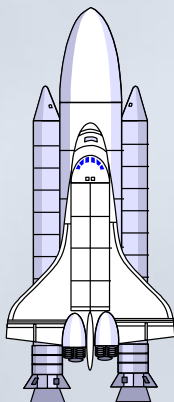


Turisti: Pretražuju informacije dobivene of farmera

Farmeri: Sakupljaju informacije Kroz poznate pristupe



Istraživači: Pronalaze nepoznate i neočekivane informacije skrivene u detaljnim podacima



Alati za korisnički pristup skladištu podataka – BI alati

BI ili BA (Business Analytic)

MicroStrategy klasifikacija:

- Izvješćivanje iz IS (engl. *enterprise reporting*)
- Analiza kocke (engl. *cube analysis*)
- Ad hoc upiti i analiza -Alati za izravnu analitičku obradu – OLAP alati
- Statistička analiza i *data mining*
- Distribucija izvješća i upozoravanje (engl. *report delivery and alerting*)

Alati za korisnički pristup skladištu podataka – BI alati

- Izvješćivanje iz IS (engl. *enterprise reporting*) – predefinirana, formatirana i statična izvješća namijenjena za širu distribuciju većem broju korisnika (operativno izvješćivanje i dashboard)
- Analiza kocke (engl. *cube analysis*)
Na kockama bazirani BI alati s jednostavnim OLAP mogućnostima (slice-and-dice)
- Ad hoc upiti i analiza -Alati za izravnu analitičku obradu – OLAP alati
Relacijski OLAP za naprednije korisnike, postavljanje svih tipova upita (slic-dice, drill down/up)

Alati za korisnički pristup skladištu podataka – BI alati

- **Statistička analiza i *data mining***

Statistički, matematički i data mining alati za prediktivnu analizu, otkrivanje uzročno-posljedičnih veza između dviju metrika, financijska analiza, predviđanje ..

- **Distribucija izvješća i upozoravanje (engl. *report delivery and alerting*)**

Za slanje čitavih izvješća ili upozorenja većem broju korisnika (internih ili vanjskih), a bazira se na pretplatama, rasporedima ili bitnim događajima u bazi podataka

BI alati - OLAP

OLAP – OnLine Analytical Processing

OLAP alati omogućavaju modeliranje, analizu i vizualizaciju velikih količina podataka, bilo pohranjenih u DBMS ili DW sustavima, baziran na dimenzijskom modelu podataka.

BI alati - OLAP

OLAP alat omogućavaju korisnicima da iskoriste sve prednosti multidimenzijskog modela podataka:

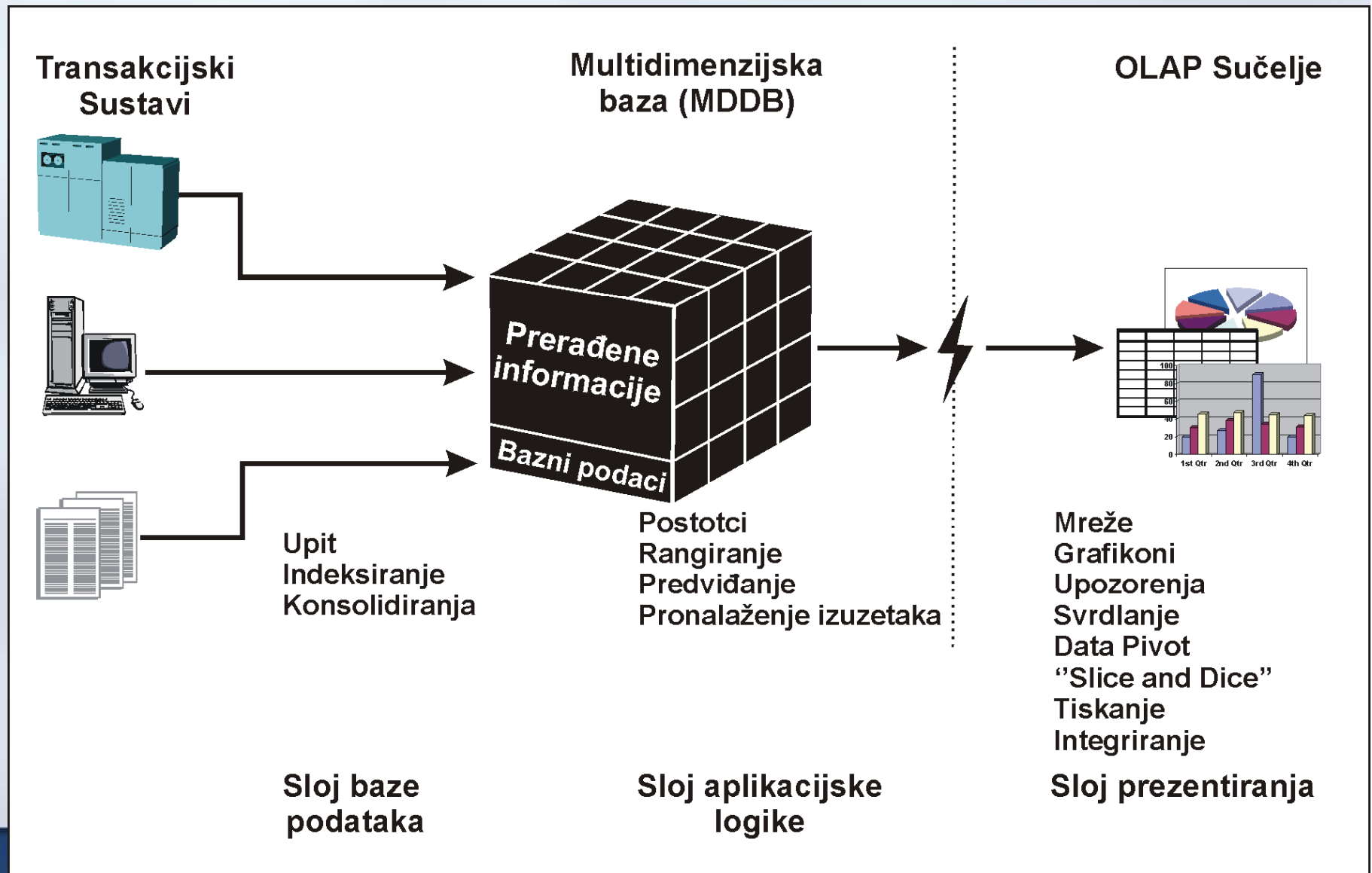
- promatranje podataka na različite načine (kao mreže ili ukrštene tablice koje se mogu jednostavno okretati što omogućava analiziranje podataka iz različitih kutova)
- interaktivno pretraživanje podataka (korisnik može slijediti tijekom svojih razmišljanja i promatrati informaciju na agregatnoj razini (na primjer: Prodajna regija) i zatim ići (“svrdlati”) do detaljnijih informacija (na primjer: grad, prodavaonica i sl.).

BI alati - OLAP

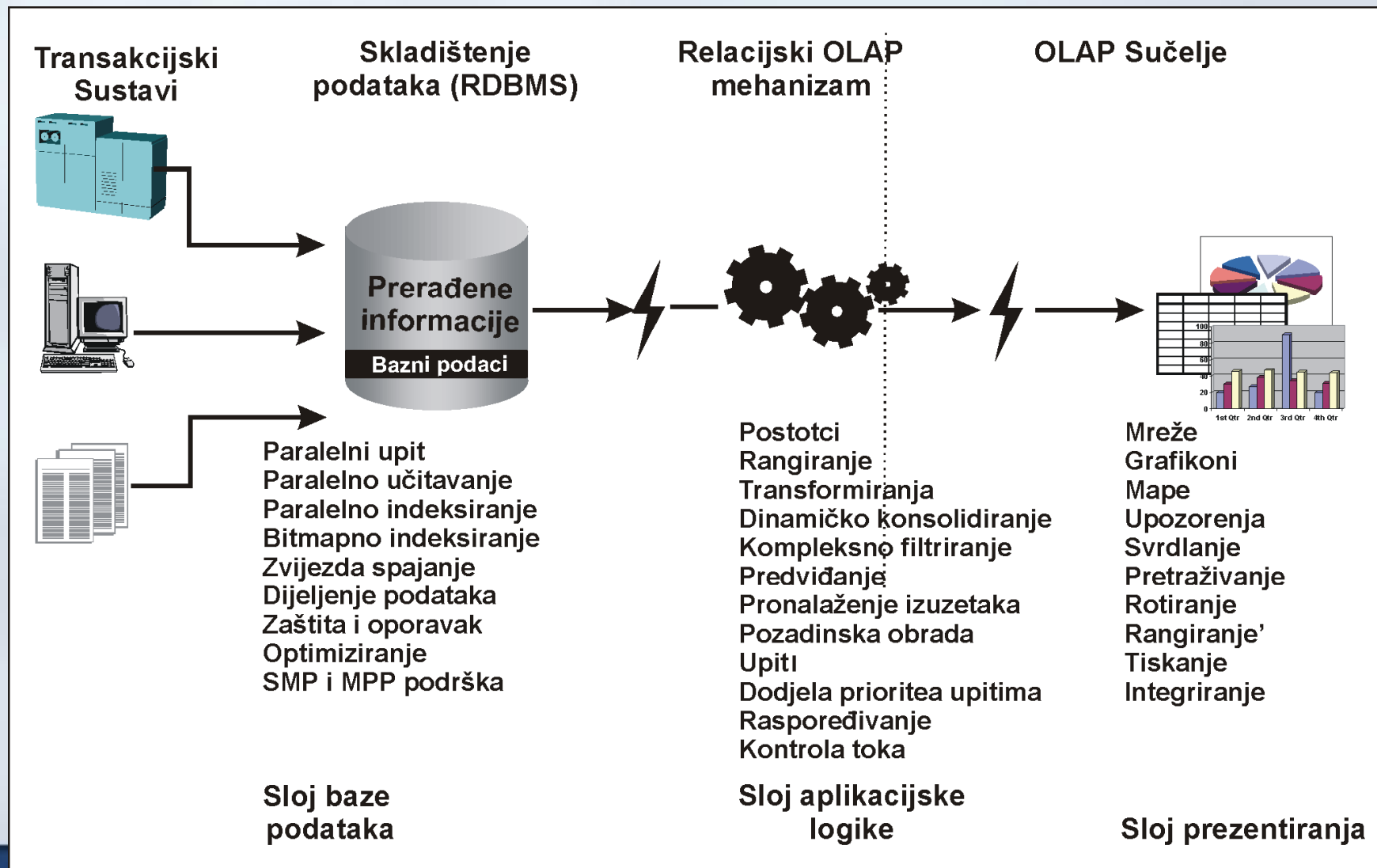
Osnovne vrste OLAP-a:

- Multidimenzijски OLAP – MOLAP
- Relacijski OLAP – ROLAP
- OLAP baze podataka
- Web OLAP
- Desktop OLAP – jeftini, jednostavni OLAP alati za lokalnu višedimenzijску analizu i prezentiranje (klijent strana)

BI alati - MOLAP



BI alati - ROLAP



BI alati – Data mining

DATA MINING - rudarenje podataka

Definicije:

-pronalaženje zakonitosti među podacima sadržanim u bazama podataka, raznim tekstualnim podacima, nestrukturiranim podacima ili podacima organiziranim u vremenske serije.

-sustavan, interaktivan i iterativan (ponavljajući) proces izvođenja i prikazivanja korisnoga, implicitnog i inovativnog znanja iz podataka.

-Otkrivanje znanja u podacima

BI alati – Data mining

Razvoj:

Analitička statistika je jezgra procesa za otkrivanje znanja.

Prvotno se termin data mining koristio za opis procesa kroz koji se otkrivaju unaprijed nepoznati obrasci unutar skupa podataka.

Iz statističke perspektive, rudarenje podataka se može opisati kao računalski automatizirana istraživačka analiza podataka iz (obično) velikih i složenih baza podataka s različitih platformi, lokacija, operacijskih sistema i softvera

BI alati – Data mining

TEMELJNI PREDUVJETI:

- Velike količine kvalitetnih podataka (DW)
- Stručnost i kompetencija u tumačenju dobivenih rezultata

BI alati – data mining

Prednost:

Primjenom metoda i alata za rudarenje podataka mogu se otkriti ponekad iznenađujući obrasci ponašanja nekih osoba ili neočekivani odnosi među pojavama.

BI alati – Data mining

Uspješnost primjene data mininga ovisi prvenstveno o stručnosti i poslovnoj kompetenciji onih koji tumače dobivene rezultate. Upravo te osobe svojim znanjem i iskustvom mogu biti sposobne neki naizgled besmislen uzorak interpretirati na poslovno korektan i smislen način i pretvoriti ga u vrijednu informaciju.

BI alati – Data minig

DVA osnovna ograničenja:

- Količina i kvaliteta podataka
- Stručnost i kompetencije u analizi rezultata

BI alati – data mining

Rudarenje podataka je izrazito multidisciplinarno područje i obuhvaća: baze podataka, skladišta podataka, ekspertne sustave, teoriju informacija, statistiku, matematiku, logiku, te čitav niz pridruženih područja.

Područja u kojima se rudarenje podataka može uspješno primjenjivati su raznolika, primjerice, poslovanje poduzeća, ekonomija, mehanika, medicina, genetika itd.

Rudarenje podataka primjenjivo je u svim onim područjima gdje se raspoložuje velikom masom podataka na osnovu kojih se žele otkriti određene pravilnosti, veze i zakonitosti.

BI alati

Čitav niz faktora može utjecati na ishod nekog događaja, a zadatak je rudarenja podataka otkriti najznačajnije među njima i njihove karakteristike s obzirom na ciljana stanja.

Bez obzira na područje primjene dobro iskorištene metode rudarenja podataka sposobne su otkriti zakonitosti iz velike mase podataka pri čemu područje primjene ostaje u drugom planu što zapravo jest snaga primjene metoda rudarenja podataka – naglasak je na podacima a ne na području provođenja analiza.

BI alati – data mining

Rudarenjem je moguće utvrditi sljedeće vrste informacija:

- **klase**, postupkom klasificiranja prema unaprijed definiranim klasama,
- **klasterne** odnosno kategorije, postupkom klasificiranja bez unaprijed zadanih klasa,
- **asocijacije**, koje su uvjetovane događajima (npr. kupci koji kupuju proizvod A u 65% slučajeva kupuju i proizvod B),
- **sekvence**, koje ustanovljuju događaje koji u određenoj vjerojatnosti slijede jedan za drugim,
- **prognoze**, kojima se prognozira budućnost iz postojećih podataka.

BI alati data mining

Klasifikacija alata prema strukturi podataka koju algoritmi koriste:

- **Statističke metode** - linearna i nelinearna regresija, metoda glavnih komponenata, analiza vremenskih serija, korelacija, klaster analiza
- **Stablo odlučivanja** (*decision tree*) klasifikacijska i regresijska stabla (CART), hi-kvadrat automatska detekcija interakcija (CHAID).
Rastavljaju problem na diskretne skupove, korijen slijedi čitav niz čvorova (grananja)
- **Zaključivanje na bazi slučajeva** (*case-based reasoning*) koristi povijesne podatke za prepoznavanje obrazaca.
- **Neuralne mreže**. Uporaba povezanih čvorova koji operiraju po principu sličnom neuronima ljudskog mozga.

BI alati data mining

Klasifikacija alata prema strukturi podataka koju algoritmi koriste (nastavak):

- **Inteligentni agenti** – polu-autonomni programi koji inteligentno pomažu korisnicima u radu s računalnim aplikacijama, učeći kroz na primjerima baziranom zaključivanju i tako poboljšavajući svoje performanse.
- **Genetički algoritmi** - rade na principu proširenja mogućih izlaznih rezultata. Za dani fiksni broj mogućih izlaza, genetički algoritmi traže i definiraju nova i bolja rješenja. Koriste se kod klasteriranja i asocijativnih pravila.
- **Ostali alati** – induktivna pravila, vizualizacija, uglavnom se odnose na Web mining.

BI alati – data mining

Data mining alati se mogu rabiti za modeliranje velikog broja različitih problema:

- **Klasifikacija i regresija.** Predstavlja najširu skupinu problema na koje se ovi alati trenutno primjenjuju tako što kreiraju modele za predviđanje klase (klasifikacija) ili vrijednosti (regresija), a uključuju različite klasifikacijske i regresijske tehnike od koji su najpoznatije: stabla odlučivanja, neuronske mreže, Naïve-Bayes i slično.
- **Asocijativnost i redoslijed** (engl. *Association&Sequencing*). Često se nazivaju i *analiza tržišne košarice* jer su usmjerene na otkrivanje pravila ponašanja potrošača (na primjer: ukoliko kupe jedan proizvod, što će uz njega još kupiti i slično).
- **Klasteriranje.** Deskriptivna tehnika koja grupira slične entitete zajedno i odvaja različite entitete po grupama. To je tipičan primjer primjene *data mining* alata gdje korisnik nema određenog plana već se nada da će alat pronaći neku značajnu strukturu. Tehnike klasteriranja sadrže posebne tipove neuronskih mreža poznatih kao *Kohonen mreža*, demografske algoritme i sl..

BI alati – data minig

Data mining je dio većeg iterativnog procesa poznatog pod nazivom pronalaženje znanja (engl. *knowledge discovery*). Osnovni koraci procesa pronalaženja znanja:

- **Definiranje problema.** Podrazumijeva definiranje ciljeva projekta za pronalaženje znanja, kao i identificiranje podataka koji će se upotrebljavati.
- **Prikupljanje, čišćenje i pripremanje podataka.** Odnosi se na prikupljanje neophodnih podataka iz različitih internih i eksternih izvora, rješavanje razlika i konflikata unutar podataka, spajanje podataka iz različitih tablica kako bi se kreirao homogeni izvor, kreiranje agregata i slično. Ovaj dio preuzima DW.
- **Data minig.** Podrazumijeva odabir *data mining* alata, generiranje uzoraka (ako je neophodno) za obuku, testiranje i provjeru modela, kao i uporabu alata za izgradnju, testiranje i odabir modela.
- **Provjera modela.** Testira preciznost modela na neovisnom skupu podataka, onom koji se nije rabio za kreiranje modela, utvrđuje osjetljivost modela, te probno testira njegovu uporabljivost.
- **Aktiviranje modela.** Može zahtijevati izradu kompjuteriziranog sustava koji preuzima odgovarajuće podatke i generira predviđanje u realnom vremenu tako da donositelj odluka može primijeniti predviđanje.
- **Nadzor (praćenje) modela.** Zahtijeva stalnu provjeru modela na novim podacima kako bi se utvrdilo da li taj model još uvijek odgovara realnoj situaciji.

Data mining

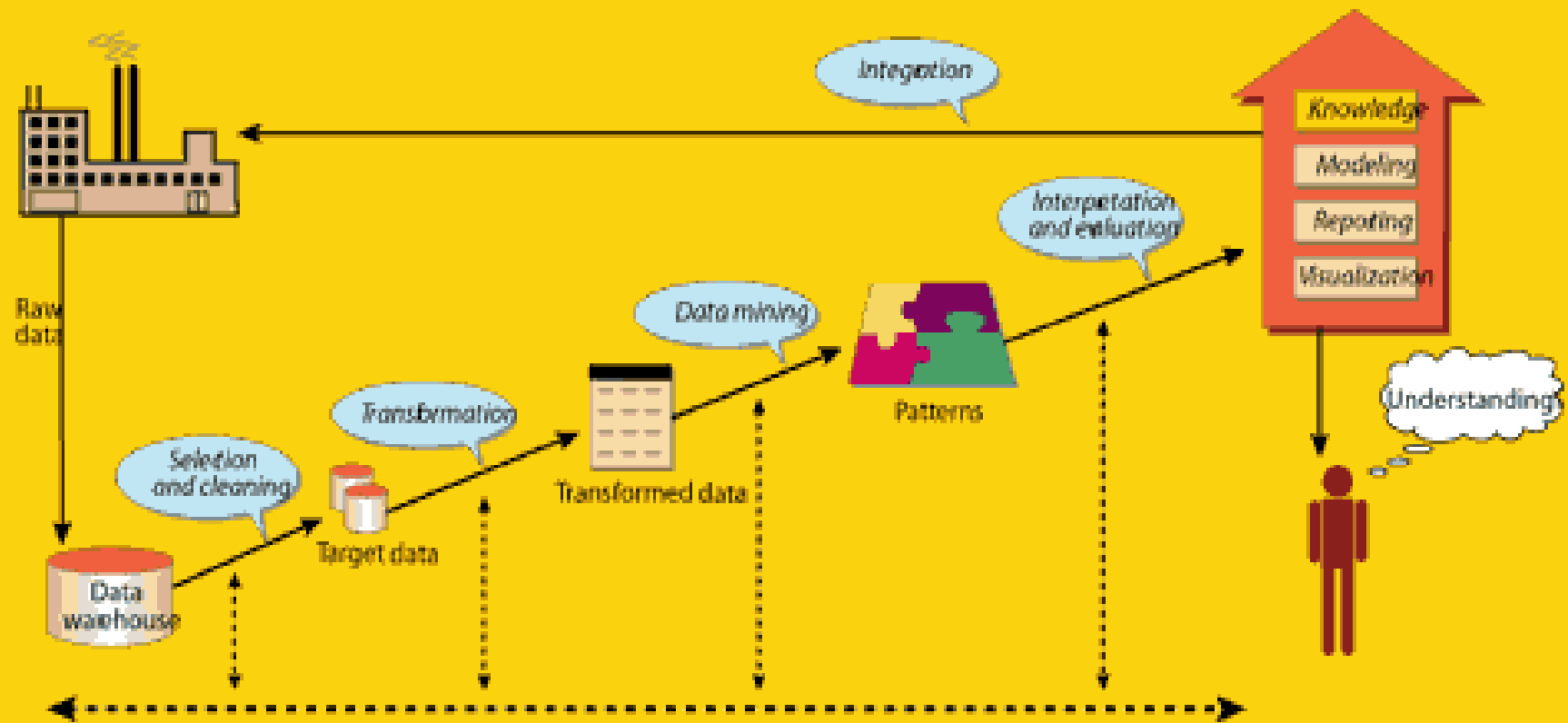


Figure 2. The KDD process.

Vrednovanje informacija

Odabrana su DVA pristupa:

1. Hammingov vremenski pristup
2. Liautaudov poslovno-pragmatički pristup

Vrednovanje informacija

1. Hammingov vremenski pristup

Richard Hamming polazi od koncepta životnog ciklusa informacije (*Information Life Cycle*), koji pretpostavlja da svaka informacija u određenom trenutku nastaje, neko vrijeme traje, da bi nakon isteka tok vremena nestala.

Vrijednost informacije je funkcija vremena.

Vrijednost nije postojano, već vremenski promjenjivo obilježje svake informacije.

Vrednovanje informacija

1. Hammingov vremenski pristup

Funkciju vrijednosti informacije čine tri područja:

1. Područje prognoziranja – informacija se dobiva i prije no što je potrebno donijeti neku odluku. Vrijednost je tada najveća i ona s vremenom nelinearno opada.

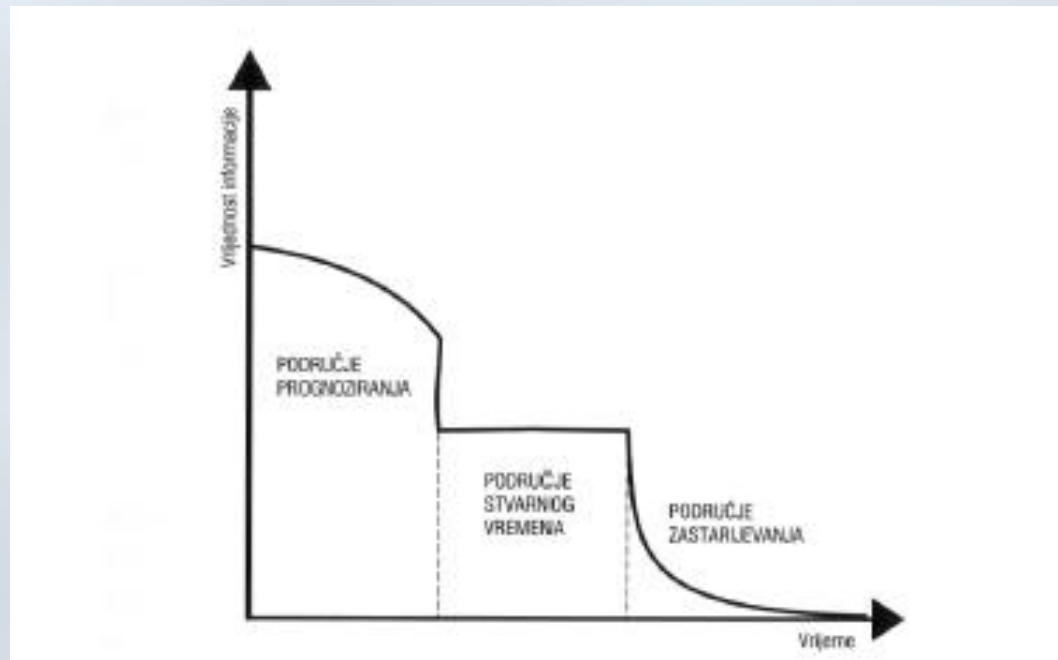
2. Područje stvarnog (realnog) vremena - informacija se dobiva (tj. nastaje) upravo onda kada treba donijeti odluku. **Vrijednost informacije je konstantna.**

3. Područje zastarijevanja – informacija se dobiva prekasno tj. onda kada odluka više nije potrebna ili nije relevantna za daljnji tijek događaja (procesa). U ovome području vrijednost informacije strmo eksponencijalno pada.

Vrednovanje informacija

1. Hammingov vremenski pristup

Prihvaćen u teoriji menadžmenta.



Vrednovanje informacija

1. Hammingov vremenski pristup

Suvremeni menadžment zahtijeva raspoloživost informacije u području prognoziranja, odnosno u području stvarnog vremena, kada njena vrijednost nadmašuje ili je uravnotežena s "težinom" odluke koju treba donijeti.

Naglasak je na istraživanju mogućih načina primjene informacijske tehnologije koji će osigurati isporuku informacija donositeljima odluka onda kada je njena vrijednost najveća.

Vrednovanje informacija

2. Liautaudov poslovno-pragmatički pristup

Po Bernardu Liautaudu - vrijednost određene informacije proteže se u kontinuumu.

Uporaba informacije širi se (propagira) diljem poduzeća, pa čak i izvan njega, prema klijentima i partnerima.

=> vrijednost informacija može se zadovoljavajuće precizno definirati kao funkcija broja korisnika koji mogu pristupati tim informacijama i analizirati ih i broja poslovnih područja kojima korisnici pripadaju:

vrijednost (informacija) = korisnici² x poslovna područja

Vrednovanje informacija

2. Liautaudov poslovno-pragmatički pristup

Što više ljudi dijeli iste informacije, oni se bolje razumiju, jednostavnije komuniciraju, tješnje surađuju i donose bolje odluke.

Vrijednost informacije raste skokovito onda kada neko novo poslovno područje tvrtke ostvari mogućnost online pristupa istim informacijama.

=> stvaranje nove organizacijske inteligencija.

Vrednovanje informacija

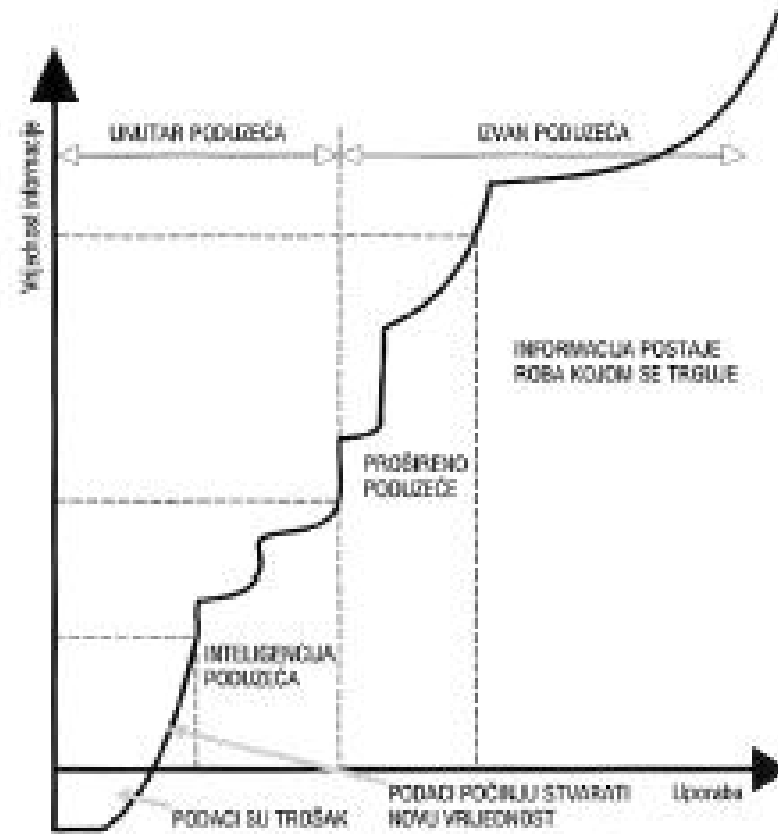
2. Liataudov poslovno-pragmatički pristup

Kontinuum vrijednosti informacije ima pet zona:

1. zona u kojoj podaci predstavljaju trošak
2. zona u kojoj podaci počinju stvarati novu vrijednost
3. zona inteligencije poduzeća
4. zona proširenog poduzeća
5. zona u kojoj informacija postaje roba kojom se trguje

Vrednovanje informacija

2. Liautaudov poslovno-pragmatički pristup



Sretan Božić i Nova Godina



Pitanja ????

