

SKLADIŠTA PODATAKA KAO TEMELJ POSLOVNE INTELIGENCIJE

1. Zašto skladišta podataka?

Predvidjeti budućnost u ovom kompetencijskom poslovnom svijetu najjače je oružje u borbi za opstanak odnosno prevlast na tržištu. Zato nije neobično što sve velike organizacije i tvrtke koriste i ulažu u napredak alata i tehnologija poslovnih inteligencija. Poslovna inteligencija služi za dobivanje i analizu podataka te prikazivanje poslovnih informacija čiji izvor je najčešće skladište podataka jer jedino informacija u pravo vrijeme i na pravom mjestu ostvaruje profit.

2. Poslovna inteligencija

Pojam „inteligencija“ znači sposobnost shvaćanja i laganog snalaženja u novim prilikama; sposobnost otkrivanja zakonitosti u odnosima među činjenicama i rješavanju problema. Primjena inteligencije u poslovnoj tvrtki temelji se na informacijama, podacima i znanju. Poslovna inteligencija (eng. Business intelligence, BI) odnosi se na tehnologije i aplikacije za prikupljanje podataka, integraciju, analizu i prezentaciju poslovnih informacija. Sustavi poslovne inteligencije obično se oslanjaju na skladišta podataka koja sadrže *povijesne nepromjenjive podatke* koji se izvlače iz različitih izvora. Grubo rečeno, *to su procesi gdje sirovi podaci postaju informacije*.

Postoje dva različita izvora podataka iz kojih organizacije, obavljajući svoje poslovne aktivnosti, prikupljaju informacije:

- Vanjski izvor podataka – Podaci pristižu iz okruženja tvrtke, odnosno sa tržišta u kojem dotična tvrtka djeluje
- Unutarnji izvor podataka – Podaci nastaju realizacijom poslovnih procesa unutar same tvrtke

3. Sustavi poslovne inteligencije

Zbog neodređenih granica i nedostatnog razumijevanja poslovne inteligencije i njenih dometa, ne postoje strogo definirana područja u kojima poslovna inteligencija može djelovati. Posljedica toga je što je došlo do raznolikosti stilova poslovne inteligencije koje su evoluirale u nekoliko smjerova. Ako obratimo pozornost na razvoj aplikacija i tehnologije, danas možemo govoriti o pet dominantnih područja poslovne inteligencije.

i. Poslovno izvještavanje (eng. Reporting)

Odnosi se na širok pojam izvještaja raznih formata namijenjenih financijskom i operativnom poslovnom izvještavanju te usporednih tablica rezultata poslovanja i poslovnih taktika namijenjenih primjeni od strane korisnika informacija različitih profila.

ii. OLAP sustavi (On Line Analyzing Processing)

To su sustavi čiji analitički alati omogućavaju presijecanje i raslojavanje (Slice And Dice) ograničenih skupova podataka namijenjenih menadžerima i ostalim djelatnicima koji zahtijevaju sigurno i jednostavno okruženje za provedbu osnovnih analiza podataka.

iii. Dubinska analiza podataka (eng. Data mining)

Aplikacije koje omogućavaju cjelovitu i složenu matematičku, financijsku i statističku obradu podataka za svrhe analize korelacije, analize trendova, tržišnih analiza, financijskih analiza te predviđanja budućih događaja i procesa, odnosno otkrivanje prikrivenih znanja iz raspoloživih podataka. Namijenjene su dominantno profesionalnim analitičarima.

iv. Alarmni alati i sustavi ranog obavještanja

Alati i sustavi za proaktivno, rano obavještanje i upozoravanje širokog kruga korisnika prema unaprijed utvrđenim kriterijima i događanjima koji će poslužiti kao poticaj za aktiviranje obavještajnog mehanizma. Namijenjeni su podjednako unutarnjim i vanjskim korisnicima informacija.

v. Skladište podataka (eng. Data warehouse)

Obzirom da se ostala područja poslovne inteligencije najčešće "oslanjaju" i podrazumijevaju uporabu skladišta podataka, skladište podataka može se uvrstiti u jedno od područja poslovne inteligencije. Skladište podataka je baza velike količine podataka koja sadrži nepromjenjive povijesne podatke.

4. Što je skladište podataka

Mnoge tvrtke ovise o trenutnim podacima koji se pohranjuju u operacijskim bazama podataka. U takvim bazama pohranjuju se podaci o svakodnevnim poslovnim operacijama, odnosno *trenutne vrijednosti*. Poslovna inteligencija se izvodi iz podataka koji *nastaju tijekom redovitog poslovanja tvrtke*. Obzirom da je količina tih podataka velika, potrebno je organizirati ih na odgovarajući način. To se ostvaruje skladištenjem podataka. Svako će skladište definirati sljedeća obilježja:

- ✓ Tematska usmjerenost
- ✓ Sadržajna nepromjenjivost
- ✓ Integriranost podataka
- ✓ Vezanost uz vrijeme

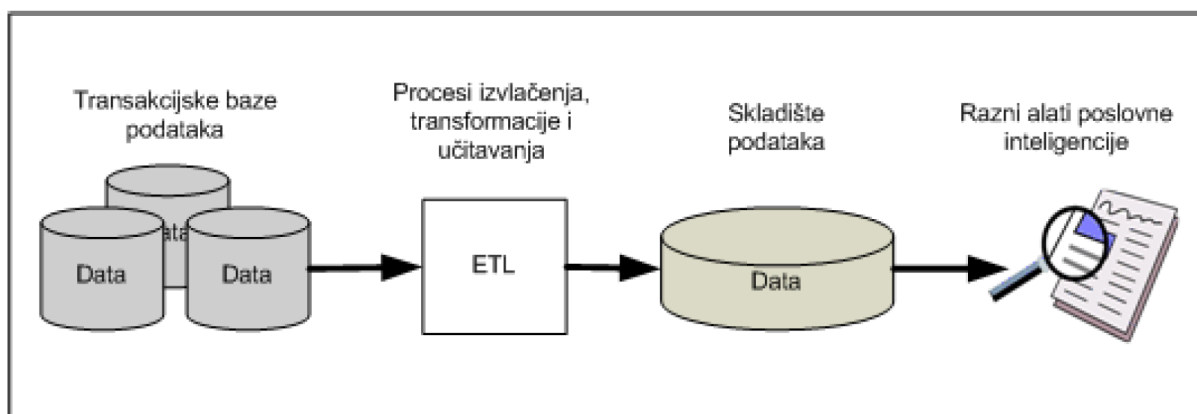
Tabela 1. Razlike između transakcijskog sustava i skladišta podataka

Sadržaj podataka	Trenutne vrijednosti, detaljni podaci	Povijesni podaci, detaljni i sumarni podaci
Vrijednost podataka	Vrlo promjenjiva	Postojana
Namjena podataka	Vođenje poslovnog sustava, dnevne operacije	Izvrješćivanje o stanju poslovnog procesa, analiza
Jedinica obrade	Transakcija	Upit
Korisnici	Službenici	Analitičari i menadžeri
Raspoloživost	Izuzetno važna	Manje važna
Izmjena podataka	Polje po polje	Nema direktne izmjene
Radne karakteristike	Čitanje/pisanje	Čitanje
Interakcija korisnika	Predodređena	Ad-hoc
Pristup zapisima	Deseci zapisa istovremeno	Milijuni zapisa istovremeno
Fokus	Spremanje podataka	Dobivanje informacija

U transakcijskom sustavu (sustav u kojem se nalazi operacijska baza podataka), zapisi su kraći i najčešće uključuju istovremenu povezanost dvije do tri tablice - brzina je važan faktor. Ažuriranje, unošenje i brisanje se odvija zapis po zapis dok kod skladišta podataka vrijeme nije glavni faktor tako da se procesiraju upiti koji za svoj rezultat koriste i po desetke tisuća zapisa. Konačne razlike između sustava skladišta podataka i transakcijskog sustava (još se i naziva OLTP sustav – *On-Line Transaction Processing*) su prikazane u tablici.

5. Skladištenje podataka

Skladištenje podataka je proces integracije podataka o poslovanju neke organizacije u jednu bazu podataka, iz koje krajnji korisnici – analitičari, menadžeri, donositelji poslovnih odluka – mogu raditi izvješća, postavljati upite i analizirati podatke. Pojam skladištenje podataka općenito se odnosi na kombinaciju mnogih različitih baza podataka instaliranih širom poduzeća.



Slika 1. Proces skladištenja podataka

ETL postupak

1. **Extraction** - Vađenje podataka i izbor (potpunih!) podataka iz raznorodnih najčešće transakcijskih izvora podataka. Obzirom da se podaci najčešće dohvaćaju iz raznih izvora, kako ne bi došlo do redundancije, potrebno je odabrati one koji će se koristiti u aplikacijama poslovne inteligencije.
2. **Transformation** - Transformacija podataka u odgovarajući format, uključujući čišćenje, agregaciju i filtriranje. Transformacija podataka podrazumijeva pažljiv i dugotrajan proces te ga je potrebno savjesno i u cjelini izvršiti.
3. **Loading** - Učitavanje transformiranih i ujedinjenih podataka u skladište podataka. Podrazumijeva inicijalno punjenje te naknadno učitavanje novih podataka u skladište.

6. OLAP (On Line Analytical Processing)

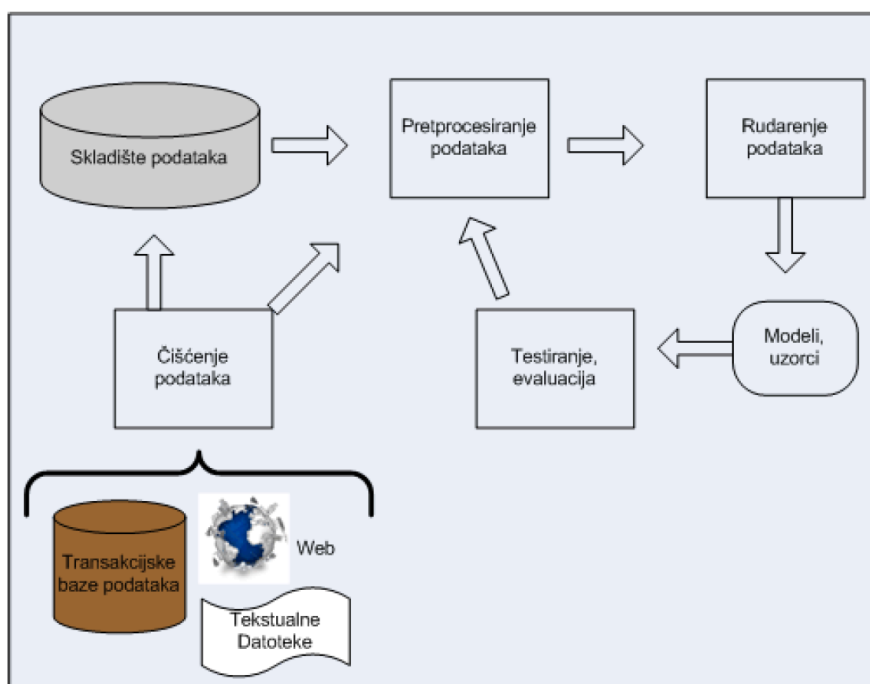
OLAP je područje poslovne inteligencije koje se odnosi na analizu i predviđanje događaja i procesa koji bi se mogli dogoditi u budućnosti. Obzirom da su podaci u skladištu podataka oblikovani u

dimenzijskim strukturama odnosno dimenzijskim tablicama, OLAP se zasniva na metodologiji multidimenzijske analize, što znači da se podatci mogu gledati kroz veći broj filtera koji se u stručnoj terminologiji nazivaju dimenzijama. U nastavku je dan pregled metoda obrade dimenzijskih podataka:

1. **Agregiranje** (eng. *Drill Up*) – Odnosi se na zbrajanje vrijednosti niže razine. Agregati povećavaju brzinu obrade dimenzijskih podataka jer nije potrebno prilikom svakog upita ponovno izvršavati operaciju agregiranja. Kod transakcijskih baza podataka ovo bi bio potencijalni problem narušavanja normalizacije odnosno pojave redundancije podataka, no skladište ovo tolerira.
2. **Detaljiziranje** (eng. *Drill Down*) – Obrnut postupak od postupka agregiranja. Primjenjiv na dimenziju vremena gdje se iz datuma izvlači podatak o mjesecu ili danu.
3. **Raslojavanje i presijecanje** (eng. *Slice & Dice*) – Odnosi se na uzimanje isječka podataka po jednoj ili više dimenzija (ako promatramo višedimenzionalnu strukturu odnosno kocku).
4. **Rotiranje** (eng. *Rotation*) – Višedimenzionalni model se rotira te pri svakoj poziciji rotacije dobivaju se nove informacije.

7. Dubinska analiza podataka

Dubinska analiza podataka (eng. Data mining) temelji se na pronalaženju zakonitosti i veza među podacima koji mogu biti organizirani u baze podataka. Drugi naziv za dubinsku analizu je i rudarenje podataka. Zakonitosti se pronalaze primjenom metoda koje svoje korijene nalaze u različitim područjima statistike, matematike, baza podataka, teorije informacije, vjerojatnosti...



Slika 2. Proces dubinskog analiziranja podataka

Ovaj stil poslovne inteligencije namijenjen je naprednijim korisnicima, prvenstveno profesionalnim analitičarima koji redovito obavljaju analize korelacija i trendova, te stvaraju određene projekcije budućih kretanja i događanja, što omogućuje poslovnom svijetu donošenje proaktivnih odluka temeljenih na znanju (engl. Knowledge-driven decisions). Alati za dubinsku analizu podataka daju odgovore na poslovna pitanja za čija je rješavanja u tradicionalnom pristupu trebalo puno više vremena. Oni pretražuju izvore baza podataka tražeći skrivene modele i *pronalaze predvidljive informacije* koje međutim mogu promaknuti stručnjacima zato što "leže" izvan njihovih očekivanja.

Za uspješno provođenje procesa dubinske analize podataka potrebno je precizno formulirati cilj i problem koji se želi riješiti, a uspješnost same analize ovisit će i o kvaliteti raspoloživih podataka. Pri tome se primjenjuju sljedeće metode:

1. **Metoda segmentiranja podataka** - podrazumijeva postupak kojim se provodi grupiranje ili pridruživanje objekata sličnih karakteristika, gdje je cilj pronaći grupe (clusters) koje se razlikuju jedna od druge dok su istovremeno članovi unutar grupa dosta slični jedni drugima. Najčešće korišteni algoritmi koji se koriste za ovu metodu su algoritam K-srednjih vrijednosti (eng. K-mean clustering) i hijerarhijsko klasteriranje.
2. **Neuronske mreže** - predstavljaju gusto isprepletene mreže međusobno povezane računskim elementima koji se nazivaju neuroni. Neuronske mreže vrlo se dobro ponašaju na težim klasifikacijskim problemima, kod kojih je teško ili nemoguće koristiti druge metode. Jedan od nedostataka neuronskih mreža jest relativno spor i zahtjevan proces 'učenja' modela te ne postoji jedinstveni model neuronske mreže koji bi se primjenjivao na sve vrste problema.
3. **Stabla odlučivanja** - predstavljaju način klasificiranja atributa u odnosu na zadanu ciljnu varijablu. Stablo se razgranava za svaku vrijednost testnog atributa, s time da se koraci ponavljaju rekurzivno sve dok se ne dostigne neki od kriterija koji zaustavlja rekurziju.
4. **Metode za analizu asocijacija** (tzv. metoda potrošačke košarice) – Analizu asocijacija karakterizira opisni pristup istraživanju podataka koji može pomoći pri identifikaciji odnosa među vrijednostima u bazi podataka. Drugim riječima, ova metoda služi za analizu kombinacija proizvoda koji se često kupuju zajedno.
5. **Genetički algoritmi** - ne koriste se u rudarenju podataka za prepoznavanje uzoraka, već se uglavnom koriste kao tehnike za rješavanje problema optimizacije.