

James A. Senn's
Informaciona tehnologija

Poglavlje 7

Višekorisničke baze podataka i skladišta podataka



Ciljevi

- ▶ Objasniti razliku između objektno-orijentisanih i relacinih baza podataka.
- ▶ Objasniti pojmove: šema baze, pogled (prikaz) i indeks.
- ▶ Analizirati prednosti klijet/server konfiguracije.
- ▶ Objasniti razliku između deljenih (centralizovanih) i distribuiranih baza podataka.



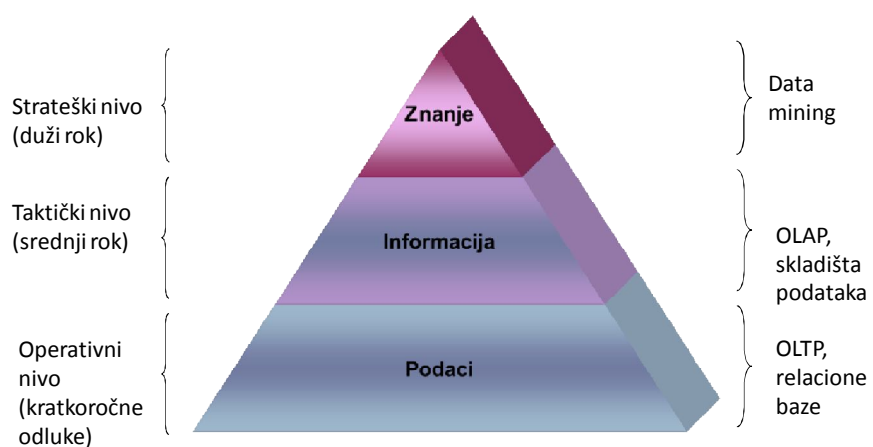
Ciljevi

- ▶ Objasniti šta su to skladišta podataka i po čemu se razlikuju od relacionih baza podataka.
- ▶ Objasniti pojmove: OLAP, data mart, data mining.



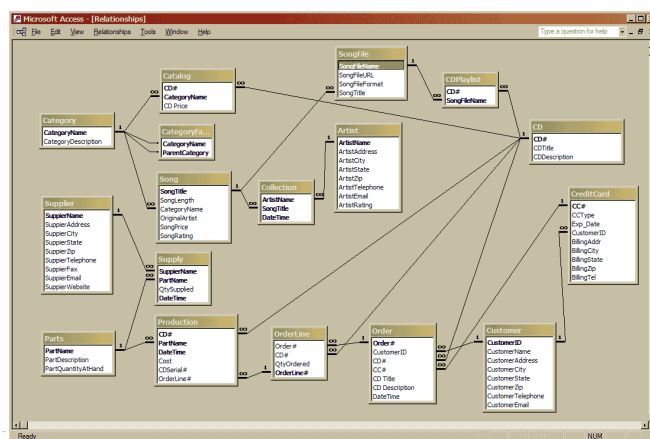
Uticaj informacionih tehnologija na odlučivanje

Osnova procesa odlučivanja su *podaci*, *informacije* i *znanje*.



Relacione baze, OLTP

- ▶ OLTP – *online obrada transakcija*. Omogućava istovremeni unos podataka kao i njihovo preuzimanje iz relacione baze podataka.



Skladišta podataka

Definicija

- ▶ Skladište podataka je baza podataka za procese podrške odlučivanju u kojoj su podaci:
 - ▶ **subjektno orijentisani** – odlikavaju poslovne procese,
 - ▶ **integrisani** – baza podataka konsoliduje podatke iz različitih sistema koji koriste razne vrste kodovanja, mernih jedinica itd. i obezbeđuje konzistentnost podataka,
 - ▶ **vremenski zavisni** – svi podaci su u vezi sa nekim vremenskim trenutkom na osnovu kojeg se podaci mogu i porediti,
 - ▶ **nepromenljivi** – podaci se, najčešće, pridodaju već postojećim umesto da ih zamenjuju
- ▶ Skladište podataka je **analitička** baza podataka u kojoj su omogućeni složeni, unapred nepredviđeni (ad-hoc) pristupi velikom broju različitih podataka.

Skladišta podataka

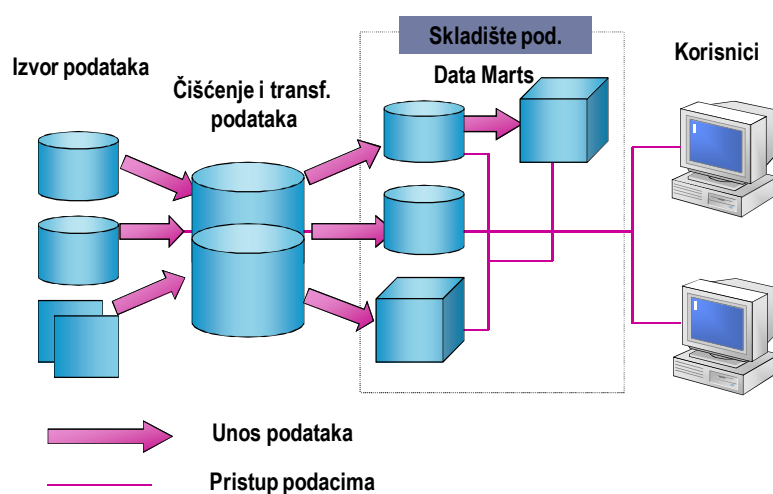
Definicija

► Skladišta podataka (Data Warehouse):

- Velike baze podataka, dizajnirane za potrebe izvršavanja velikog broja upita, koje u sebi sadrže podatke iz prethodnih vremenskih perioda.
- Podaci u skladištima podataka su organizovani po **dimenzijama** ili temama (kupci, dobavljači, proizvodi, regioni,...) i sadrže informacije koje su relevantne za analizu podataka i proces odlučivanja.
- Skladišta podataka koriste **višedimanzionalnu** strukturu podataka (čuvaju podatke u više od dve dimenzije), za razliku od relacionih baza koje čuvaju podatke u dvodimenzionalnim tabelama.
- Podaci u skladištima su **denormalizovani**.

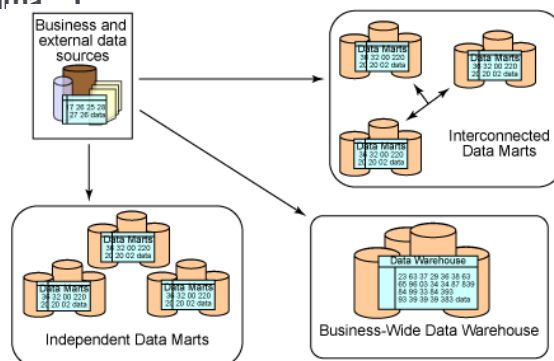


Postupak izgradnje skladišta podataka



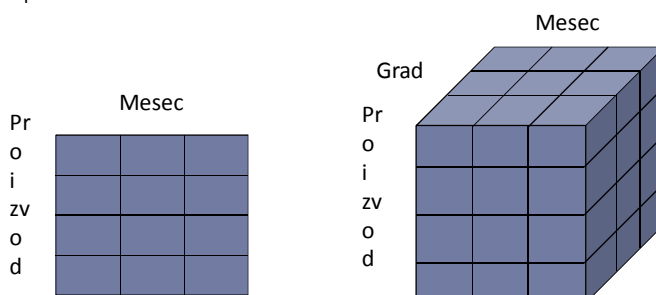
Data mart

- ▶ **Lokalna skladišta podataka (Data Marts):**
 - ▶ Mala skladišta podataka projektovana za potrebe određene grupa krajnjih korisnika (npr. lokalno skladište za računovodstvo, za odeljenje marketinga, upravljanje ljudskim resursima)



Primeri dvodimenzionih i trodimenzionih modela podataka

- a) Podaci o prodaji za svaku oblast se nalaze u različitim tabelama
- b) Svi podaci smešteni su u trodimenzioni niz

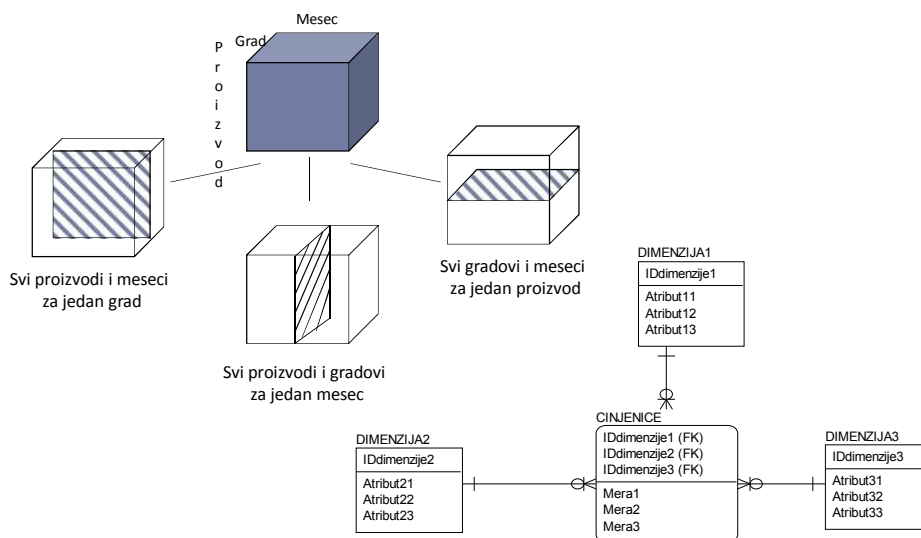


a) Dvodimenzioni model podataka

b) Trodimenzioni model podataka

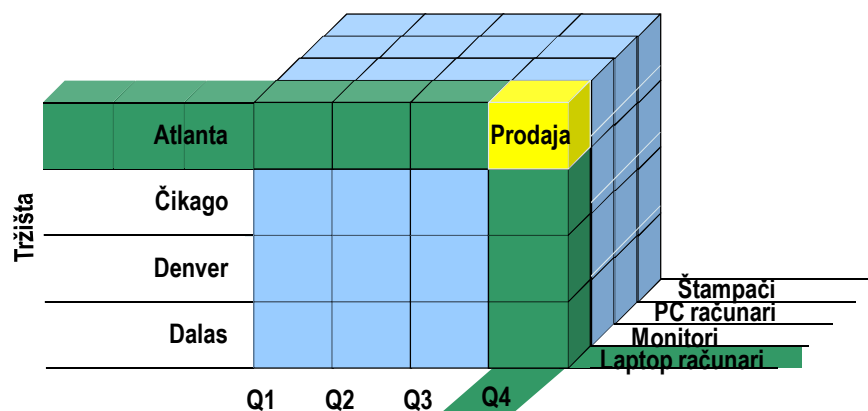


Višedimenzionalna baza podataka



OLAP

- ▶ **analitička obrada podataka u realnom vremenu (OLAP):**
 - ▶ **Analiza akumuliranih podataka u skladištima podataka u realnom vremenu.**



Poređenje relacionih baza podataka, skladišta podataka i OLAP-a

Karakteristike	OLTP sistemi	Skladište podataka	OLAP sistemi
Tipične operacije	ažuriranje	izveštavanje	analiza
Analitički zahtev	nizak	srednji	visok
Ekрани	nepromenljivi	definiše korisnik	definiše korisnik
Količina podataka po transakciji	mala	srednja	velika
Nivo podataka	detalji	detalji i sumarni podaci	sumarni podaci
Starost podataka	tekući podaci	tekući i istorijski podaci	tekući, istorijski i projektovani podaci
Namena	transakcionim podacima	rad sa istorijskim podacima	analiza
Tip pristupa	čitanje i pisanje	samo čitanje	čitanje i pisanje
Karakteristike odziva	brzo ažuriranje, promenljivo trajanje vremena odziva sistema	dugo vreme odziva sistema	kratko vreme odziva sistema
Nivo detaljnosti podataka	transakcioni podaci	delimično sumarni podaci	sumarni i agregacioni podaci
Struktura podataka	normalizovana (zapis i)	normalizovana ili denormalizovana	dimenziona i hijerarhijska
Količina podataka	gigabajti podataka	gigabajti/terabajti podataka	gigabajti podataka
Adaptivnost sistema	ograničena, uz značajnu upotrebu resursa	slaba	jednostavnost modifikacije
Brzina uvođenja sistema	mala (reda godine)	mala (reda godine)	velika (reda dana ili nedelje)



Data mining (“prekopavanje” podataka)

- ▶ **Data Mining:** Algoritam za pronalaženje skrivenih informacija u podacima.
- ▶ **Data mining tehnike:**
 - ▶ neuronske mreže
 - ▶ stablo odlučivanja
 - ▶ genetski algoritmi
 - ▶ metoda najbližeg suseda.



Objektno-orijentisane baze podataka

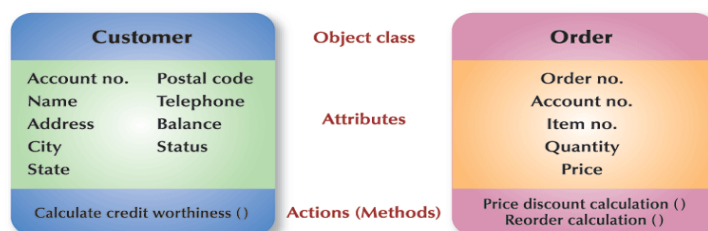
- ▶ Šema objektno-orijentisane baze podataka
 - ▶ Objektno-orijentisana struktura baza podataka razvijena je sa ciljem da se prevaziđu neki nedostaci relacionog modela, a to je pre svega nemogućnost da se kompletne informacije o objektu sačuvaju u bazi podataka (i stanje objekta i njegovo ponašanje). Objektno-orijentisani pristup modeliranju baza podataka podrazumeva da se realno okruženje sagledava preko objekata i interakcija između njih (odnosno njihovih veza).
 - ▶ Objektno-orijentisana baza podataka: Baza podataka koja čuva stanja i ponašanja o objektima.



Objektno-orijentisane baze podataka

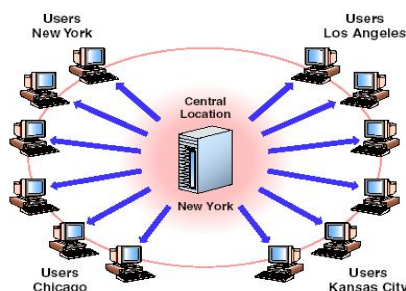
Objekti su entiteti koji su sposobni da čuvaju svoja *stanja* zajedno sa metodama koje definišu njihovo *ponašanje*. Interakcija između dva objekta naziva se **veza između dva objekta**.

Skup objekata istog tipa naziva se **klasa** (npr. jednu klasu čine kupci, drugu narudžbine,...)



Centralizovane baze podataka

- ▶ svi podaci i celokupna baza podataka nalazi se na jednoj lokaciji, jednom serveru.
- ▶ Podacima se pristupa preko terminala.
- ▶ Centralizovane baze podataka na velikim mainframe računarima, bile su decenijama najzastupljenije zbog velikih operativnih troškova, alternativnih rešenja.

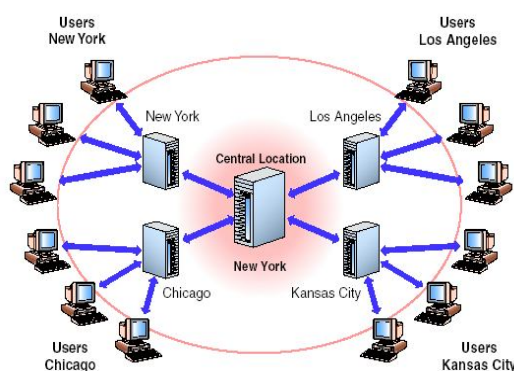


Centralizovane baze podataka

- ▶ **Prednost centralizovanih rešenja:**
 - ▶ konsistentnost podataka,
 - ▶ jedan administrator sistema,
 - ▶ bezbednost pristupa podacima se definiše na jednom pristupnom mestu,
 - ▶ backup podataka je centralizovan,
 - ▶ oporavak sistema nakon kraha, je lakši na centralizovanim lokacijama.
- ▶ **Mane centralizovanih rešenja:**
 - ▶ u slučaju pada sistema korisnici na svim lokacijama nisu u mogućnosti da nastave rad, koji je vezan za pristup bazi.
 - ▶ Pristupa podacima je spor kada veliki broj korisnika pristupa istovremeno.

Distribuirane baze podataka

- ▶ Kompletna kopija baza podataka ili jedan njen deo, nalazi se na više lokacija, na više lokalnih servera. Pristup serverima sa PC računara.
- ▶ Te lokacije su u blizini korisnika baze podataka.
- ▶ Na lokalnim serverima obavlja se obrada podataka za lokalne i regionalne ogranke, kancelarije, pogone.



Distribuirane baze podataka *Particionisanje i replikacija*

- ▶ Baze podataka mogu biti distribuirane na dva načina: *particionisanjem i replikacijom*.
- ▶ **Particionisanje: Različiti delovi baze podataka smešteni su na različitim čvornim tačkama na mreži.**
- ▶ **Replikacija (sinhronizacija):**
 - ▶ **Replicirane** baze podataka imaju **kompletnu kopiju** centralne baze na više lokacija.
 - ▶ Na lokalnim serverima se ažuriraju samo neki delovi baze podataka, koji se potom moraju postupkom replikacije usaglasiti sa centralnom bazom podataka.
 - ▶ Replikacija je prenos ažuriranih podataka sa lokalnih servera na centralni server.

Strategije distribucije

► *Strategija geografske distribucije:*

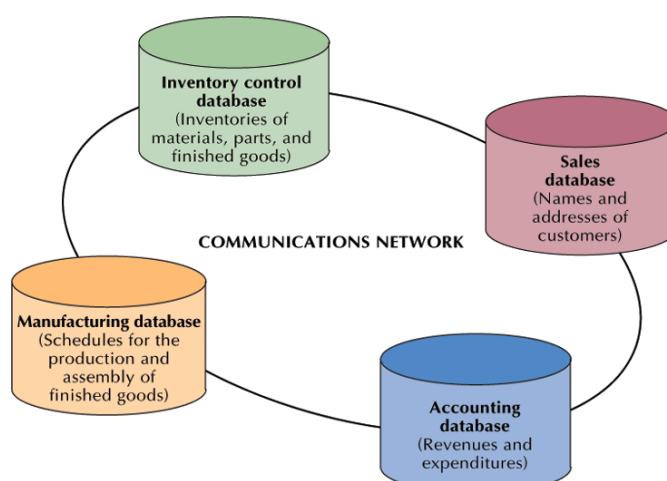
- strategija koja podrazumeva da baza podataka bude smeštena u regionu u kome se podaci i informacije iz nje najintenzivnije koriste.

► *Strategija funkcionalne distribucije:*

- distribucija baze vrši se u skladu sa poslovnim funkcijama date organizacije.



Strategija distribucije



Dizajniranje distribuirane baze

- ▶ **Direktorijum baze podataka:**

- ▶ komponenta svake deljene baze podataka koja pamti lokacije svih podataka i informacija u bazi.

- ▶ **Faktori koji utiču na dizajn:**

- ▶ Troškovi skladištenja
- ▶ Troškovi obrade
- ▶ Troškovi komunikacije
- ▶ Preuzimanje i obrada podataka
- ▶ Pouzdanost
- ▶ Učestalost ažuriranja i postavljanja upita

