

Poglavje 7

Podatkovno skladišče

- Namen podatkovnih skladišč
- Razvoj podatkovnih skladišč
- Arhitektura podatkovnega skladišča
- Orodja, ki uporabljajo podatkovna skladišča
- Implementacija podatkovnega skladišča



Podatkovno skladišče



Kaj si bomo pogledali?

- Namen podatkovnih skladišč
- Razvoj podatkovnih skladišč
- Arhitektura podatkovnega skladišča
- Orodja, ki uporabljajo podatkovna skladišča
- Implementacija podatkovnega skladišča



Zakaj podatkovna skladišča?..

- Nosilci odločanja potrebujejo dostop do vseh podatkov v okviru IS ne glede na platformo in fizično lokacijo
- Nosilci odločanja potrebujejo pregled ne le nad trenutnimi vrednostmi posameznih (zbirnih, summariziranih) podatkov in kazalnikov, temveč pregled nad zgodovino vrednosti
- Nosilci odločanja potrebujejo tudi pregled nad trendi (rasti, padanja, stagnacija)

Zakaj podatkovna skladišča?

- Podatkovna skladišča omogočajo realizacijo podpore odločanju glede na predhodno opredeljene potrebe nosilcev odločanja
- Podatkovna skladišča:
 - Hranijo podatke iz **več različnih podatkovnih virov**
 - Hranijo **zgodovinske** podatke
 - Hranijo **agregirane** podatke



Evolucija podatkovnih skladišč..

- V sedemdesetih so podjetja v svoje IS uvajala aplikativne sisteme, ki so omogočali avtomatizacijo poslovnih in ostalih procesov na operativnem nivoju
- Posledica je bila akumulacija **večjih in večjih količin** podatkov v podatkovnih **bazah transakcijskih sistemov**
- Podjetja sedaj dajejo poudarek različnim načinom uporabe teh podatkov za podporo odločitvenim procesom z namenom pridobiti strateško prednost pred konkurenco

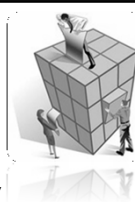
Evolucija podatkovnih skladišč..

- Nekateri transakcijski aplikativni sistemi imajo elemente, ki omogočajo podporo odločitvenim sistemom na operativnem in taktičnem nivoju (poročila, sumarni pregledi, grafi, ...)

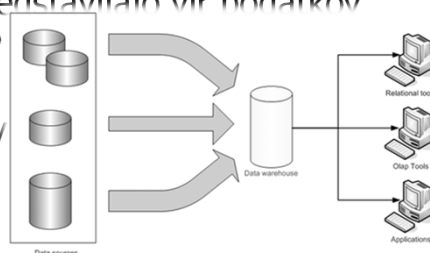


- Podjetja imajo velikokrat več transakcijskih aplikativnih sistemov, ki vsak na svoj način (včasih tudi kontradiktorno) obravnavajo iste podatke

Evolucija podatkovnih skladišč



- IS podjetij potrebujejo podatkovno skladišče,
 - ki predstavlja arhiv podatkov,
 - ki predstavlja vir znanja in
 - omogoča enoten in hkrati uporabniku prilagojen integriran in konsolidiran pogled na (sumarizirane, zgodovinske, ..) podatke
- Podatkovna skladišča predstavljajo vir podatkov za različna orodja (OLAP orodja, ..) in na ta način odločitvenim procesom v



Kaj je podatkovno skladišče?

- Definicija:
 - Podatkovno skladišče je vsebinsko organizirana, integrirana, časovno odvisna in nespremenljiva zbirka podatkov za namene podpore odločitvenim procesom.



Entitetna usmerjenost

- Organizacija podatkovnega skladišča temelji na glavnih entitetah podjetja (npr.: stranka, izdelek, regija, račun,...) in ne na funkcionalnih področjih oz. področjih, ki jih pokrivajo posamezni transakcijski sistemi

Integriranost

- Podatkovno skladišče integrira podatke iz več aplikativnih sistemov v okviru IS podjetja.
- Podatki iz različnih sistemov so večkrat med seboj nekonsistentni. Naloga podatkovnega skladišča je, da omogoči konsistenten in enoten pogled na podatke



Časovna odvisnost

- Za podatek v podatkovnem skladišču je oz. mora biti poznan čas (DD:MM:YYYY HH:MIN) ali časovni interval njegove veljavnosti oz. čas, ko je bil prenesen v podatkovno skladišče
- Časovna odvisnost je večkrat prikazana v razširjenem časovnem formatu (poleg leta še kvartal, četrletje)
- Časovna odvisnost pomeni zajem trenutnih stanj podatkov, kar omogoča opazovanje trendov

Nespremenljivost

- Podatki v podatkovnem skladišču niso podvrženi spremembam v realnem času s strani aplikacij, temveč se osvežujejo (iz transakcijskih sistemov in ostalih virov) z neko (smiselno) frekvenco
- Podatki se ob osveževanju večinoma le dodajajo v podatkovno skladišče (novi podatki)

Koristi od podatkovnih skladišč

- DW predstavlja potencialno dobičkonosno investicijo
 - Poročilo 2002: "Data warehouse and Business Intelligence systems implementation can generate a median five-year return on investment of 112% with a pay back of 1.6% on average costs of 4.5 million. Out of the organisations included in this study, 54% have a ROI of 101% or more."
- DW predstavlja konkurenčno prednost pred konkurenco
 - Dostop do podatkov, ki nosijo povsem nov tip informacije (trendi, primerjave, ...)
- Uporaba DW pomeni bistveno večjo učinkovitost nosilcev odločanja
 - Integrirani, konsistentni, zgodovinski podatki

PODATKOVNE BAZE 2 - VSP

- 436 -

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

DW in OLTP..

- Podatkovno skladišče: DW – Data Warehouse
- OLTP – On Line Transaction Processing
- Praviloma imamo za DW in OLTP različne instance podatkovne baze

PODATKOVNE BAZE 2 - VSP

- 437 -

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

DW in OLTP..

OLTP systems	Data warehousing systems
Holds current data	Holds historical data
Stores detailed data	Stores detailed, lightly, and highly summarized data
Data is dynamic	Data is largely static
Repetitive processing	<i>Ad hoc</i> , unstructured, and heuristic processing
High level of transaction throughput	Medium to low level of transaction throughput
Predictable pattern of usage	Unpredictable pattern of usage
Transaction-driven	Analysis driven
Application-oriented	Subject-oriented
Supports day-to-day decisions	Supports strategic decisions
Serves large number of clerical/operational users	Serves relatively low number of managerial users

DW in OLTP..

- DW omogoča od enostavnih do zelo kompleksnih poizvedb ob uporabi različnih orodij in tehnologij
- Orodja, tehnologije, sistemi:
 - Reporting, poizvedovanje, razvojna orodja
 - Direktorski IS (EIS - Executive information systems)
 - OLAP orodja in aplikacije
 - Orodja za Odkrivanje zakonitosti v podatkih (Data mining tools)

DW in OLTP...

▪ Primeri vprašanj za OLTP...

- Kakšen je bil celoten prihodek podjetja v prvem kvartalu 2011?
- Kakšen je bil celoten prihodek iz prodaje nepremičnin za vsak tip nepremičnine v Veliki Britaniji v letu 2003?
- Katera so tri najbolj popularna področja v velikih mestih za najem nepremičnin v 2004 in kako je to v primerjavi s preteklimi tremi leti.
- Kakšen bi bil učinek na prodajo nepremičnin v različnih regijah Velike Britanije če bi cene nepremičnin, dražjih od 100.000 funtov povečali za 3,5% obenem pa zmanjšali davke za 1,5%. £100,000?

DW in OLTP

▪ Primeri vprašanj za DW...

- Katere vrste nepremičnin se prodajajo s ceno, ki je večja od povprečne cene nepremičnin v glavnih mestih Velike Britanije in v kakšni korelaciji je to z demografskimi podatki?
- Ali obstaja signifikantna povezava med celotnim prihodkom, generiranim v posamezni nepremičninski agenciji, in številom agentov v tej agenciji?

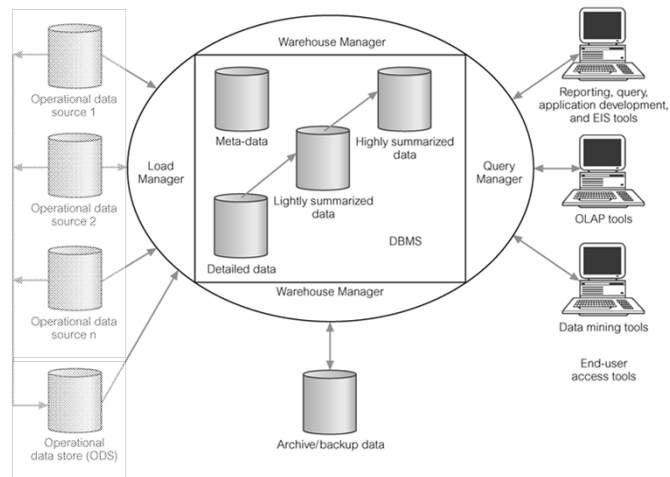
Problemi podatkovnih skladišč..

- Podcenjevanje potrebnega časa in resursov za polnjenje podatkovnih skladišč
 - Poleg polnjenja pridobivanje in čiščenje podatkov v večini primerov zahtevata veliko časa
- Skriti problemi transakcijskih in drugih sistemov, ki predstavljajo podatkovni vir (Izvorni sistemi)
- Manjkajoči podatki: V izvornih sistemih se nekateri pomembni podatki niso zajemali
- Večja zahtevnost s strani končnih uporabnikov ali nerealne zahteve

Problemi podatkovnih skladišč

- Homogenizacija podatkov in problemi pri integraciji
 - Načrtovalec DW ima včasih probleme pri združevanju podatkov iz različnih izvornih sistemov in pri tem lahko prihaja do napak
- Ogromna poraba diskovnega prostora
- Lastništvo podatkov
- Zahtevno vzdrževanje
- Izgradnja in vzpostavitev DW je praviloma dolgotrajen projekt

Arhitektura DW..



PODATKOVNE BAZE 2 - VSP

- 444 -

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

Arhitektura DW..

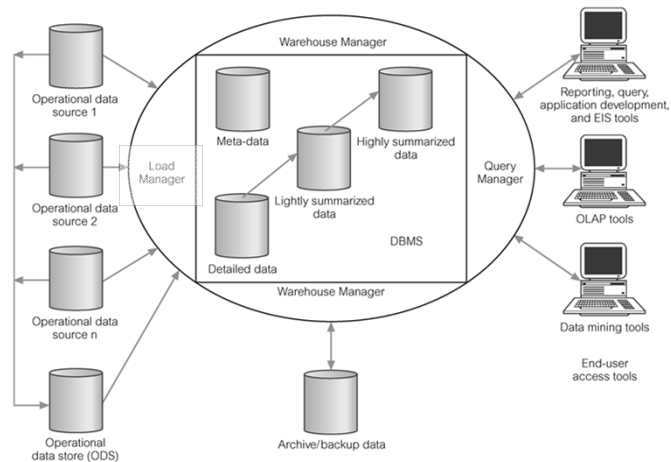
- **Izvorni sistemi (Operational Data Sources):**
 - Podatkovne baze transakcijskih aplikativnih sistemov
 - Internet
 - Ostali viri
- **Repozitorij (Operational Data Store):**
 - Repozičtorij podatkov podatkovnega skladišča
 - Večkrat ima vlogo vmesnega člana pri prenosu podatkov v DW
 - Vzpostavljen predvsem v primerih, ko so med izvornimi sistemi stari sistemi, do katerih DW ne more direktno dostopati



PODATKOVNE BAZE 2 - VSP

- 445 -

Arhitektura DW..



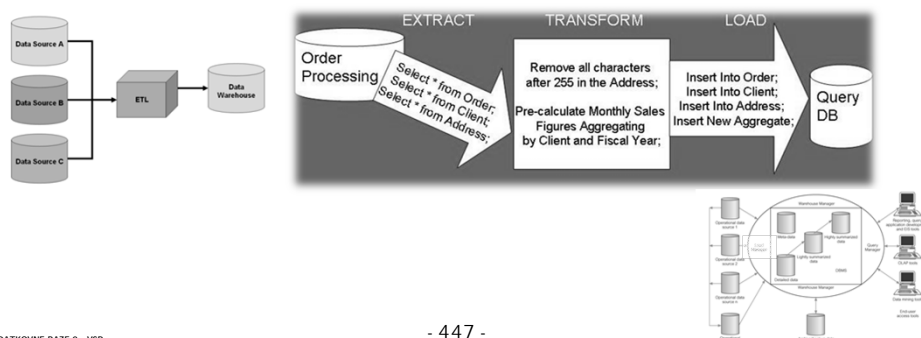
PODATKOVNE BAZE 2 - VSP

- 446 -

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

Arhitektura DW..

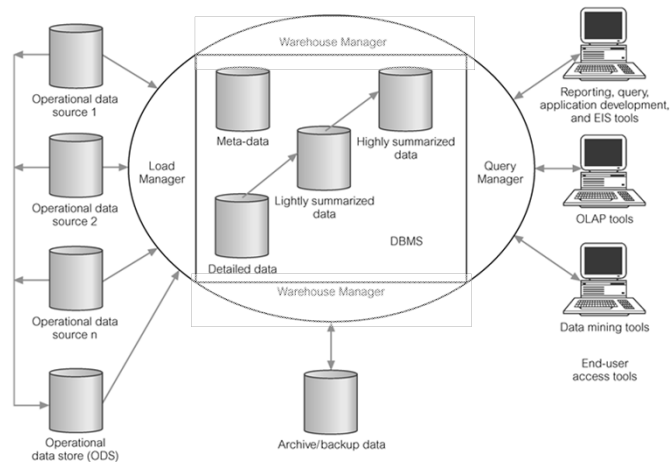
- Upravljalec prenosa podatkov (Load Manager):
 - Izvaja vse operacije na vezane ekstrakcijo podatkov in prenos podatkov v DW: transformacije in čiščenje podatkov
 - Podatke prenaša bodisi direktno iz izvornih sistemov, bodisi preko Repozitorija



PODATKOVNE BAZE 2 - VSP

- 447 -

Arhitektura DW..



PODATKOVNE BAZE 2 - VSP

- 448 -

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

Arhitektura DW..

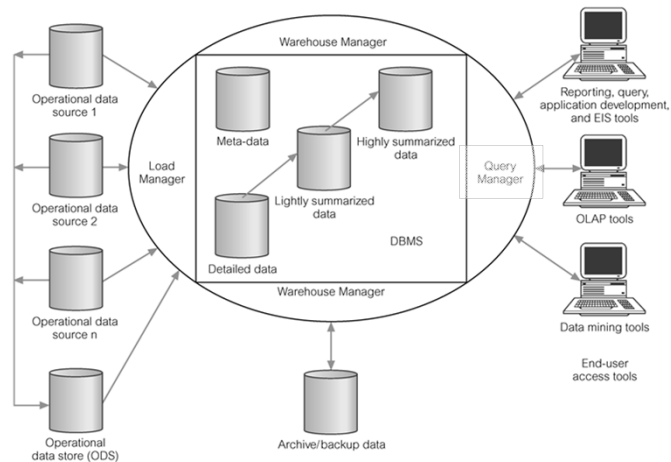
- Upravljevec DW (Warehouse Manager):
 - Upravlja s podatki v DW
 - Izvaja operacije nad podatki:
 - Analizira podatke za zagotavljanje konsistentnosti
 - Transformira podatke iz izvornih sistemov in jih združuje (integrira)
 - Kreira indekse in poglede nad tabelami
 - Generira (odkriva in kreira) denormalizacije (če je potrebno)
 - Generira agregacije (če je potrebno)
 - Izvaja arhiviranje podatkov in izdelovanja varnostnih kopij



PODATKOVNE BAZE 2 - VSP

- 449 -

Arhitektura DW..



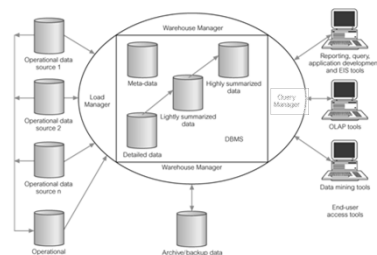
PODATKOVNE BAZE 2 - VSP

- 450 -

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

Arhitektura DW..

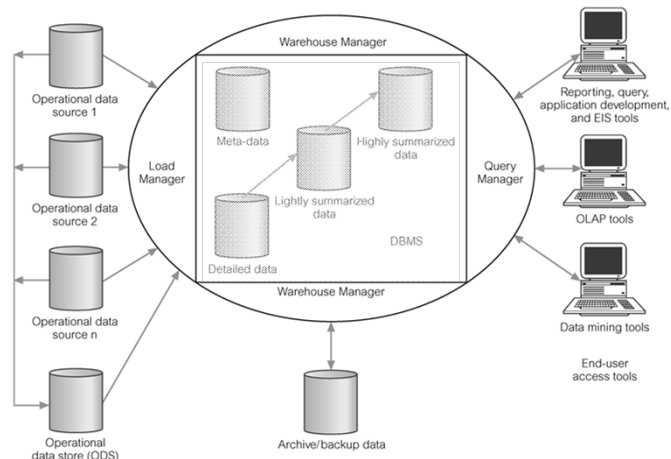
- Upravljelec poizvedb (Query Manager):
 - Izvaja vse operacije vezane na poizvedbe uporabnikov
 - Izdeluje plane izvajanja poizvedb in urnike za izvajanje poizvedb
 - Izdeluje profile poizvedb, kar omogoča Upravljalcu DW, da določi, katere indekse in agregacije potrebuje



PODATKOVNE BAZE 2 - VSP

- 451 -

Arhitektura DW..



Arhitektura DW..

- Trije nivoji "agregiranosti" podatkov:
 - Podrobni oz. transakcijski podatki (Detailed Data), ki predstavljajo vir za višja dva nivoja
 - Delno in visoko agregirani podatki (Lightly and Highly Summarized Data), ki jih generira Upravljalet DW. Cilj agregiranih podatkov je pohitrilev izvajanja poizvedb. Ta del DW se spreminja glede na spremembe profilov poizvedb

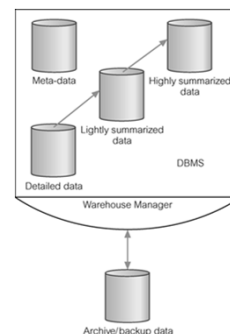
Arhitektura DW..

▪ Arhivski podatki in varnostne kopije (Archive/Backup Data):

- Za arhiviranje/hranjenje podrobnih in agregiranih podatkov
- Pomembno je izdelati varnostno kopijo agregiranih podatkov, če bomo arhivirali podrobne podatke

▪ Meta podatki (Metadata):

- Gre za ekvivalent podatkovnega slovarja oz. sistemskega kataloga pri DBMS oz. instanci podatkovne baze
- Meta podatki so podatki o podatkih, so opisi (definicij) podatkov v DW

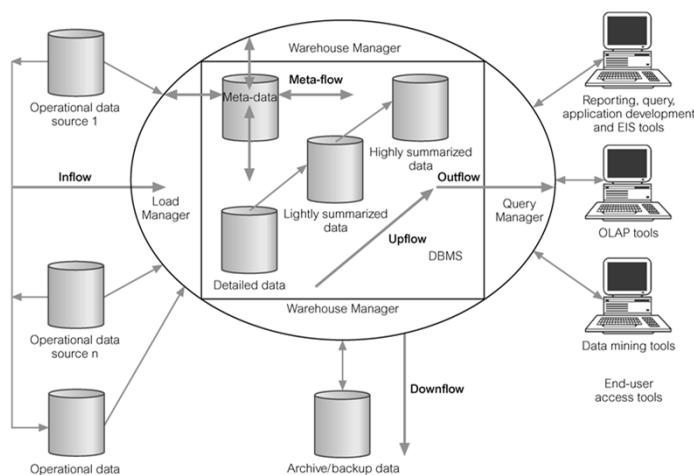


Arhitektura DW

▪ Meta podatki se uporabljajo:

- Pri ekstrakciji in polnjenju DW (polnjenje DW)
- Pri izdelavi agregiranih tabel (upravljanje DW)
- Pri poizvedbah, ko na podlagi meta podatkov Upravljelec poizvedb določi najprimernejši podatkovni vir in uporabo najprimernejšega indeksa (delovanje DW)

Podatkovni tokovi pri podatkovnih skladiščih..



PODATKOVNE BAZE 2 - VSP

- 456 -

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

Podatkovni tokovi pri podatkovnih skladiščih..

- Polnjenje (Inflow): procesi, ki so povezani z ekstrakcijo podatkov, čiščenjem podatkov in samim polnjenjem
- Agregacija (Upflow): procesi, povezani z agregacijo podatkov in s tem dodajanjem vrednosti podatkom v DW
- Arhiviranje in izdelovanje varnostnih kopij (Downflow)
- Uporaba (Outflow): procesi, ki so povezani s pripravo podatkov v obliko, ki je primerna za uporabnike in orodja, ki jih uporabljajo
- Upravljanje meta podatkov (Metaflow): procesi, ki so povezani z upravljanjem in uporabo meta podatkov

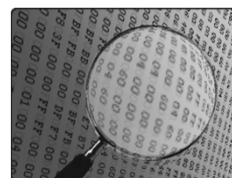
PODATKOVNE BAZE 2 - VSP

- 457 -

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

Orodja in tipi aplikacij, ki uporabljajo DW

- Generatorji poročil, orodja za kreiranje poizvedb (QBE – Query by Example)
- Razvojna orodja
- Direktorski aplikativni sistemi (EIS – Executive Information Systems)
- OLAP orodja in OLAP aplikacija
- Orodja za Odkrivanje zakonitosti v podatkih (Data Mining)



PODATKOVNE BAZE 2 - VSP

- 458 -

Generatorji poročil

- Priprava in generiranje poročil – različna orodja za pripravo standardnih poročil (pogosto desktop aplikacije) ter generatorji poročil (strežniške aplikacije).



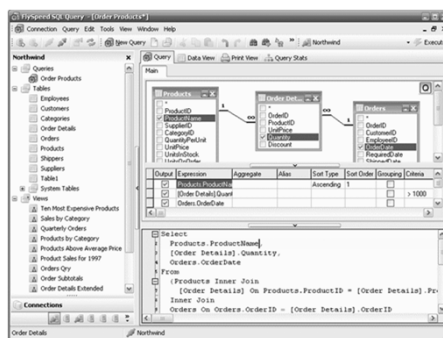
PODATKOVNE BAZE 2 - VSP

- 459 -

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

Orodja za poizvedovanje

- Orodja za poizvedovanje – orodja za poizvedovanje po relacijskih DW sprejmejo SQL na vходу ali generirajo SQL. QBE - za kompleksne poizvedbe manj primeren.



PODATKOVNE BAZE 2 - VSP

- 460 -

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

Razvojna okolja

- Razvojna okolja – če uporabniških zahtev ni moč realizirati z orodji za generiranje poročil ali poizvedovalnimi orodji, potrebno razviti uporabniško aplikacijo.
- Na voljo mnoga razvojna okolja. Večina omogoča povezavo s poljubnimi SUPB ter OLAP mnogimi orodji.

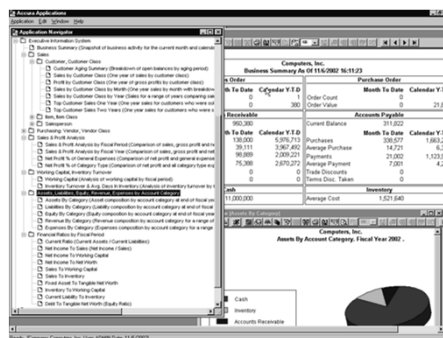
PODATKOVNE BAZE 2 - VSP

- 461 -

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

Direktorski informacijski sistemi

- Direktorski informacijski sistemi – inicialno razviti za potrebe najvišjega vodstva in podporo strateškemu odločanju. Danes podpirajo odločanje na vseh ravneh.



PODATKOVNE BAZE 2 - VSP

- 462 -

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

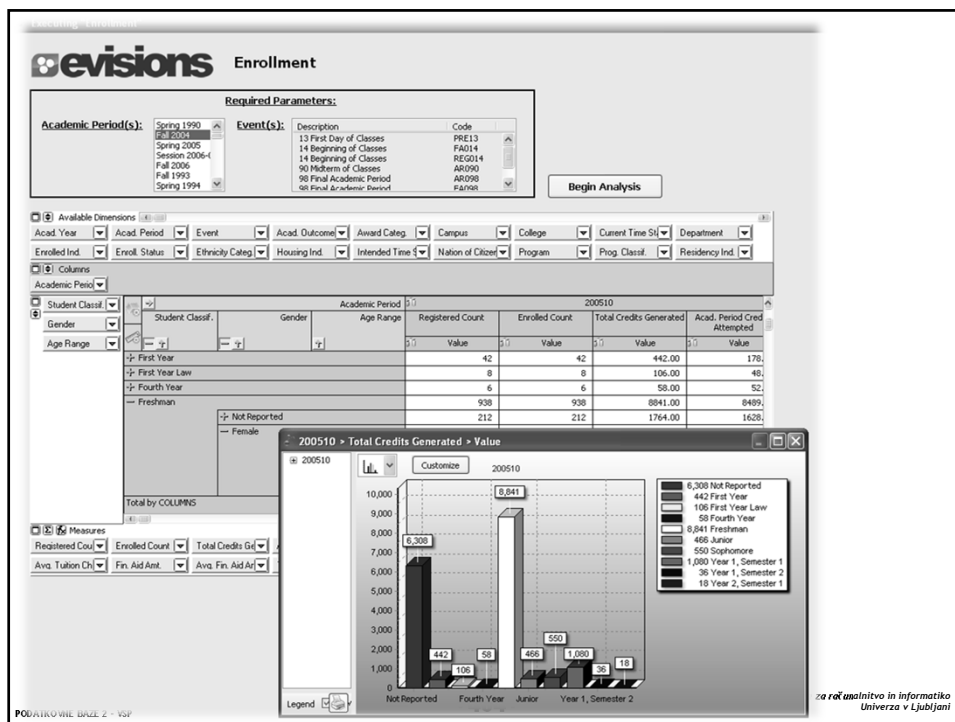
OLAP orodja

- OLAP orodja – zasnovana na konceptu več-dimenzionalnih PB. Izkušenim uporabnikom omogočajo analiziranje podatkov s pomočjo kompleksnih več-dimenzijskih pogledov.
- Primeri poizvedb: merjenje učinkovitosti marketinške akcije, napovedovanje prodaje, planiranje virov ipd.
- Kot vir za OLAP orodja potrebujemo podatke, organizirane v posebne več-dimenzijske modele.

PODATKOVNE BAZE 2 - VSP

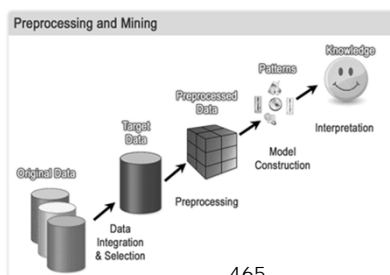
- 463 -

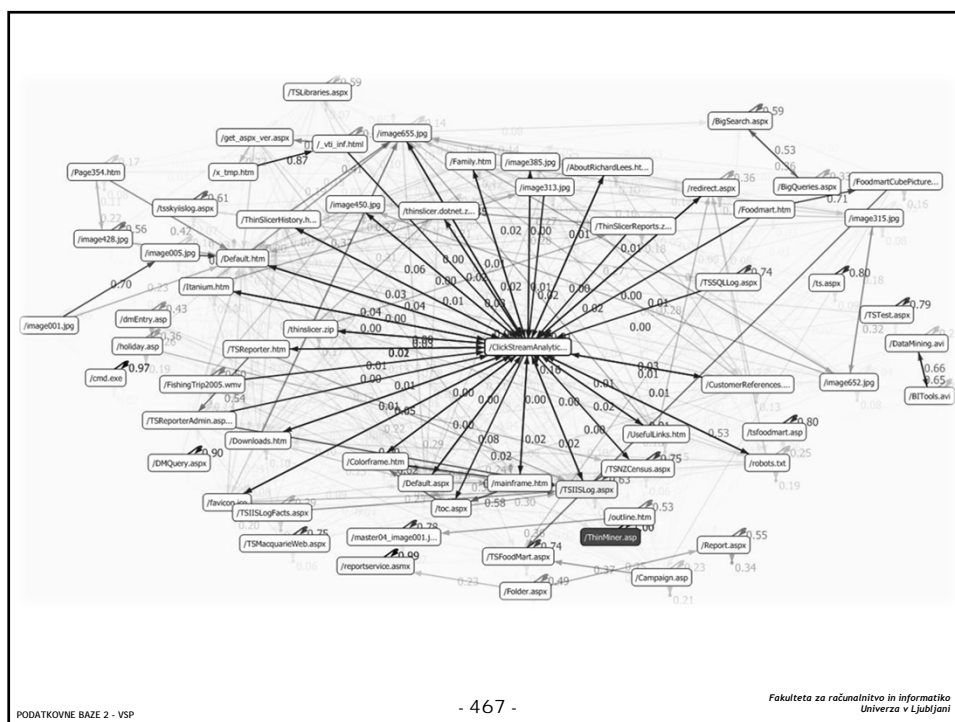
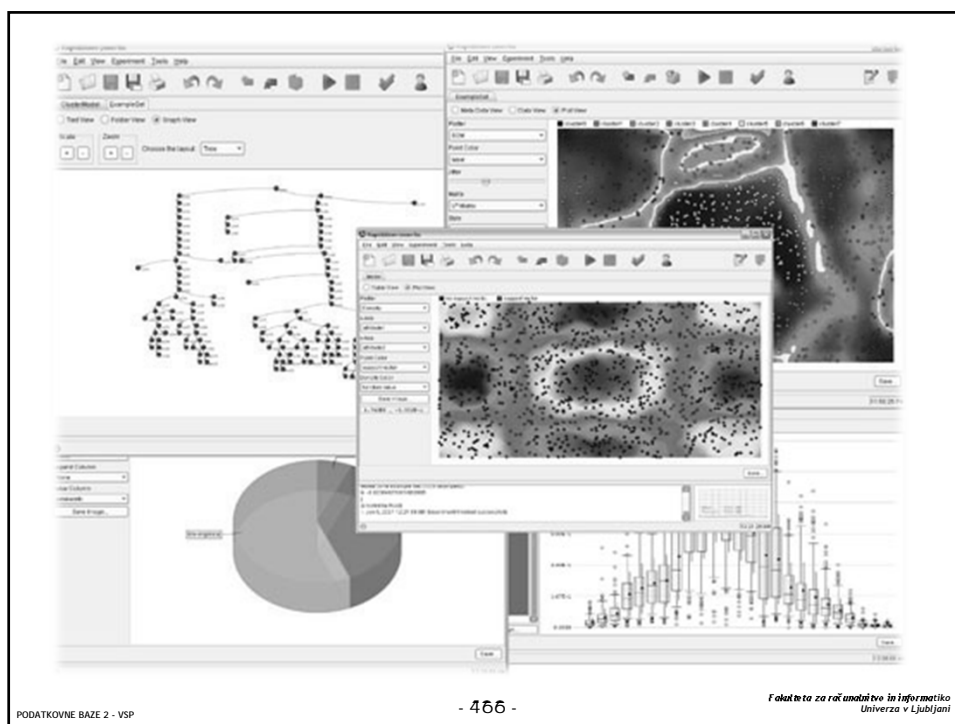
Fakulteta za računalništvo in informatiko
Univerza v Ljubljani



Data mining

- Data mining – proces odkrivanja korelacij med podatki, vzorcev in trendov z analizo velike količine podatkov in uporabo matematičnih in statističnih metod ter tehnik umetne inteligence.
- V primerjavi z OLAP orodji data mining ponuja tudi predikcijo in ne le restrospektive!





Orodja in tehnologije področja DW

- Pri izgradnji in vpeljavi DW se večkrat uporabi orodja različnih proizvajalcev, kar predstavlja dodaten element v kompleksnosti procesa razvoja in vpeljave DW
- Doseči "sodelovanje" več orodij je večkrat velik izziv

DBMS za DW..

- Standardni DBMS imajo posebne dodatne funkcionalnosti oz. karakteristike, ki omogočajo učinkovite instance PB, v katerih lahko vzpostavimo DW
- V nadaljevanju so predstavljena področja, ki so še posebej pomembna za DBMS za DW oz. predstavljajo zahteve, ki jih mora izpolnjevati DBMS za DW

DBMS za DW..

- Karakteristike polnjenja
 - Visoke zahteve za polnjenje, zahteva po sposobnosti polnjenja več milijonov vrstic na uro
 - Brez omejitev navzgor
- Procesiranje pri polnjenju
 - Konverzije tipov, transformiranje, filtriranje, preverjanje konsistentnosti (lokalne in globalne)
 - Indeksiranje
 - Ažuriranje meta podatkov

DBMS za DW..

- Odzivnost poizvedb
- Upravljanje z enormnimi količinami podatkov
 - DBMS ne sme imeti omejitev za količino podatkov
- Možnost sočasne uporabe s strani veliko uporabnikov
- DW kot distribuirana PB

DBMS za DW

- Podpora več-dimenzionalnim podatkovnim strukturam (za potrebe OLAP orodij)
- Razširjen nabor funkcionalnosti pri poizvedbah

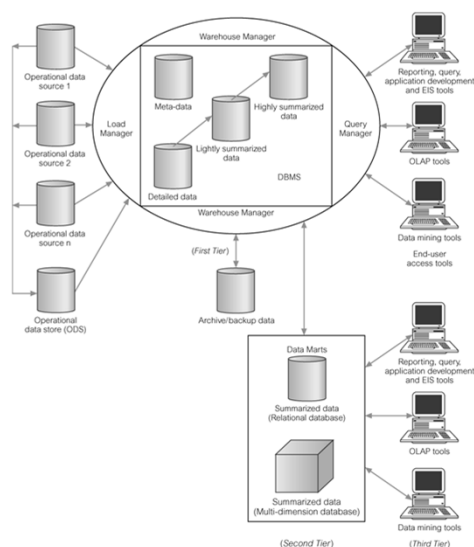
Meta podatki..

- Uporabljajo se za različne namene in upravljanje meta podatkov je ključnega pomena za integriran DW
- Področja, za katera potrebujemo meta podatke:
 - Za področje transformacije in polnjenja
 - Za področje upravljanja DW
 - Za področje upravljanja poizvedb
- Za področje transformacije in polnjenja podatkov mora biti znan izvorni objekt vsakega polja (lokacija, tabela, polje) in vse transformacije, ki izvorno polje pretvorijo v ponornega (opazovanega)

Meta podatki

- Za področje upravljanja DW mora biti opisan vsak podatek in objekt v DW. Po eni strani gre za ekvivalent sistemskemu katalogu, po drugi strani pa ima DW dodatne zahteve. Na primer: podatke o agregaciji
- Za področje upravljanja poizvedb, predvsem njihovega izvajanja, je potrebno spremljati in beležiti izvajanje posameznih poizvedb in na podlagi teh podatkov omogočiti učinkovitejše izvajanje teh poizvedb v bodoče

Data Mart..



Data Mart..

- Data mart je del (podmnožica) DW, ki pokriva zahteve določenega oddelka ali funkcionalnega področja
- Značilnosti Data mart:
 - Fokussiran je na zahteve določenega oddelka ali funkcionalnega področja
 - Praviloma ne vsebuje podrobnih transakcijskih podatkov, temveč le sumirane podatke

Data Mart

- Razlogi za kreiranje Data mart:
 - Dati uporabnikom na voljo le podatke, ki jih najpogosteje analizirajo
 - Omogočiti posameznikom ali skupinam, da opazujejo podatke (oddelka ali funkcionalnega področja) formatirane na način, na katerega so navajeni
 - Izboljšati odzivne čase zaradi manjše količine podatkov

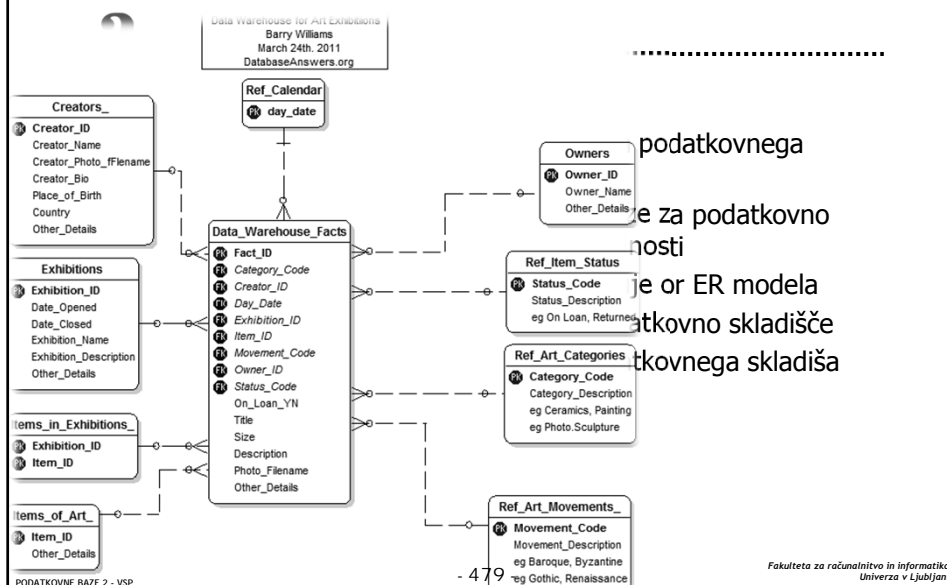
Poglavje 7

Načrtovanje podatkovnega skladišča

- Vprašanja povezana z načrtovanjem podatkovnega skladišča
- Tehnika načrtovanja podatkovne baze za podatkovno skladišče – modeliranje dimenzionalnosti
- Kako se dimenzionalni model razlikuje od ER modela
- Metodologija načrtovanja PB za podatkovno skladišče
- Kriteriji za ocenjevanje dimenzij podatkovnega skladišča



Načrtovanje podatkovnega skladišča



Načrtovanje podatkovnega skladišča...

- Pred začetkom izdelave podatkovnega skladišča, je potrebno odgovoriti na več vprašanj:
 - Katere so najpomembnejše zahteve uporabnikov?
 - Katere podatke najprej uporabiti?
 - Ali s podatkovnim skladiščem pokriti celoten poslovni sistem ali le določeno poslovno področje?
 - Če začnemo z manjšim obsegom, ali naj bo infrastruktura, na kateri se bo razvilo skladišče, pripravljena tudi za končno podatkovno skladišče?
 - ...



Načrtovanje podatkovnega skladišča...

- Izdelava DW za celoten poslovni sistem težka naloga – veliko področij, številne uporabniške zahteve itn.
- Pogosto odločitev za razvoj več področnih podatkovnih skladišč (Data marts) in na koncu združitve v enotno skladišče.

Načrtovanje podatkovnega skladišča...

- Zelo malo načrtovalcev se je pripravljenih zavezati izdelati načrt za podatkovno skladišče celotnega poslovnega sistema, ki mora takoj ustrezati vsem zahtevam uporabnikov.
- Kljub vmesnim rešitvam – izdelava Data Mart-ov, ostaja končni cilj nespremenjen: izdelava podatkovnega skladišča, ki podpira zahteve poslovnega sistema.

Načrtovanje podatkovnega skladišča

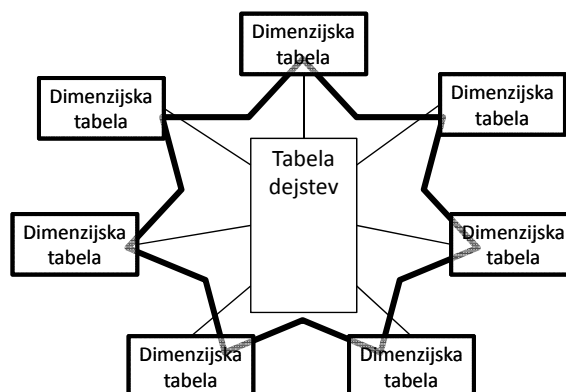
- Aktivnosti:
 - Zajem zahtev in
 - Analiza
- Intervjuji v fazi zajema zahtev:
 - odgovorni za posamezna poslovna področja (marketing, prodaja, finance, vodstvo...) – uporabniške zahteve
 - administratorji podatkovnih virov – kateri podatki obstajajo?
- Na ta način se identificira najpomembnejšo množico zahtev, ki jim mora podatkovno skladišče ustreči.

Modeliranje dimenzij...

- Tehnika logičnega načrtovanja, ki omogoča predstaviti podatke na način, da so hitro dostopni...
- Tehnika uporablja koncepte ER modeliranja z nekaj pomembnimi omejitvami.
- Vsak model dimenzij vsebuje:
 - eno tabelo, ki vsebuje kompozitni primarni ključ – tabela dejstev in
 - množico manjših dimenzijskih tabel.
 - Vsaka dimenzijska tabela ima enostaven ključ – komponenta kompozitnega ključa iz tabele dejstev

Modeliranje dimenzij...

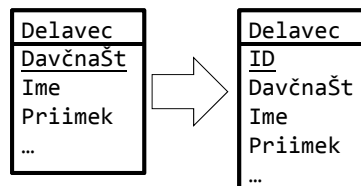
- Omenjene tabele tvorijo ti. zvezdno shemo.
 - Primarni ključ tabele dejstev je sestavljen iz dveh ali več tujih ključev posameznih dimenzijskih tabel.



Modeliranje dimenzij...

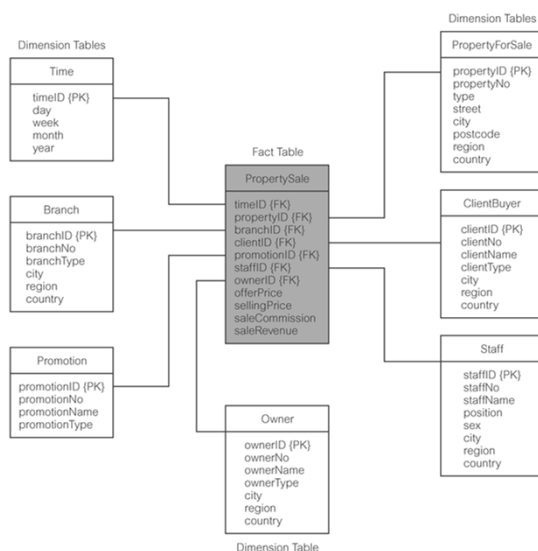
- Vse naravne primarne ključe se nadomesti s surogati: vsak stik med tabelo dejstev in dimenzijsko tabelo temelji na surogatih.

- Surogat: umetni ključ



- Surogati omogočajo neodvisnost med podatki DW in OLTP!

Modeliranje dimenzij...



Modeliranje dimenzij...

- Zvezdna shema:
 - logična struktura, ki vsebuje v središču tabelo z dejstvi in je obkrožena z dimenzijskimi tabelami z referenčnimi podatki, ki so lahko denormalizirani.
- Dejstva se generirajo na podlagi dogodkov, ki so se zgodila v preteklosti – spremembe so malo verjetne

Modeliranje dimenzij...

- Večina podatkov v podatkovnem skladišču se nahaja v tabeli dejstev → lahko so ekstremno obsežne tabele.
- Podatki v tabelah dejstev
 - referenčni podatki – namenjeni branju,
 - skozi čas se ne spreminjajo,
 - numerične vrednosti.
- Najbolj uporabna dejstva v tabeli dejstev so numerična → lahko se seštevajo, kar omogoča hitro agregiranje podatkov.

Tabela dejstev

M1	M2	M3	M4	M5

Numerične vrednosti (dogodki)

Modeliranje dimenzij...

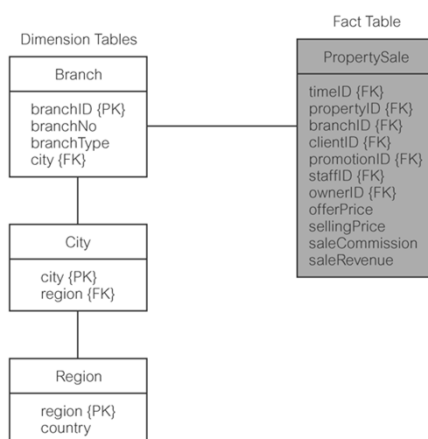
- Podatki v dimenzijskih tabelah
 - navadno opisni tekstovni podatki.
 - atributi uporabni kot omejitve (filter) v poizvedbah po DW.
- Povečanje hitrosti poizvedb v zvezdnih shemah: uporaba denormaliziranih podatkov v posameznih dimenzijah.

Modeliranje dimenzij...

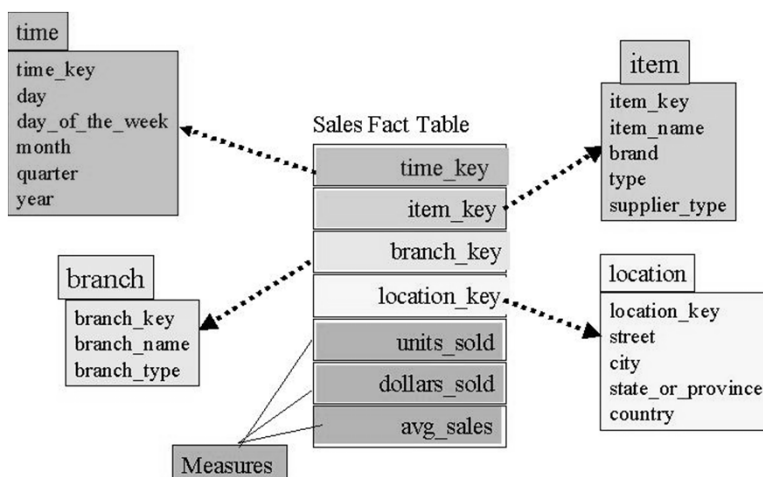
- Snežna shema (ang. snowflake) –
 - shema brez denormaliziranih podatkov
- Zvezdna snežna shema (ang. starflake) –
 - hibridna struktura, ki vsebuje tako zvezdno (denormalizirano) kot sneženo (normalizirano) shemo.
 - shema omogoča predstavitev dimenzij na oba načina – ustreza različnim performančnim zahtevam poizvedb.

Modeliranje dimenzij...

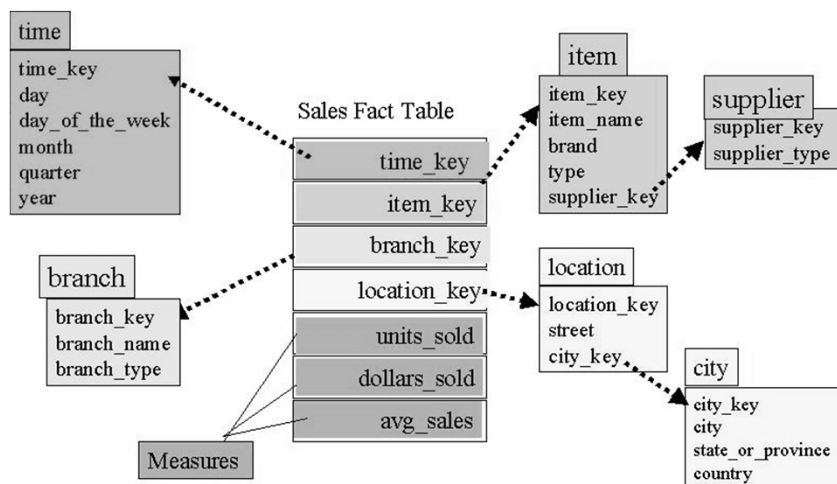
- Normalizirana varianta dimenzije Branch:



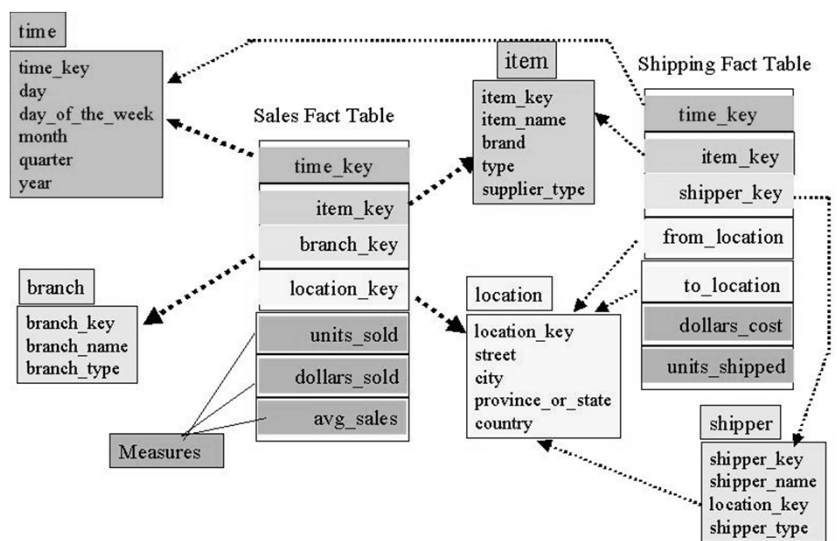
Zvezdna shema



Snežna shema



Sozvezdna shema



Modeliranje dimenzij...

- Zvezdna shema prinaša pomembne prednosti (ne glede na uporabljeno različico):
 - Učinkovitost,
 - Zmožnost prilagajanja različnim zahtevam,
 - Razširljivost,
 - Zmožnost modeliranja splošnih poslovnih situacij,
 - Predvidljivo procesiranje poizvedb.

Primerjava med zvezdno shemo in ER modelom

- En ER podatkovni model se normalno dekomponira v več zvezdnih shem.
- Omenjenih več zvezdnih shem se medsebojno povezuje preko dimenzij v "skupni rabi".

Metodologija načrtovanja DW

■ Metodologija načrtovanja po Kimballu:

- K1: Izbira procesa
- K2: Določitev zrnatosti
- K3: Izbira dimenzij
- K4: Identifikacija dejstev
- K5: Shranjevanje izračunanih vrednosti v tabeli dejstev
- K6: Podroben opis dimenzij
- K7: Določitev dolžine zgodovine podatkov v tabeli dejstev
- K8: Obvladovanje počasi spreminjajočih se dimenzij
- K9: Opredelitev prioriternih poizvedb

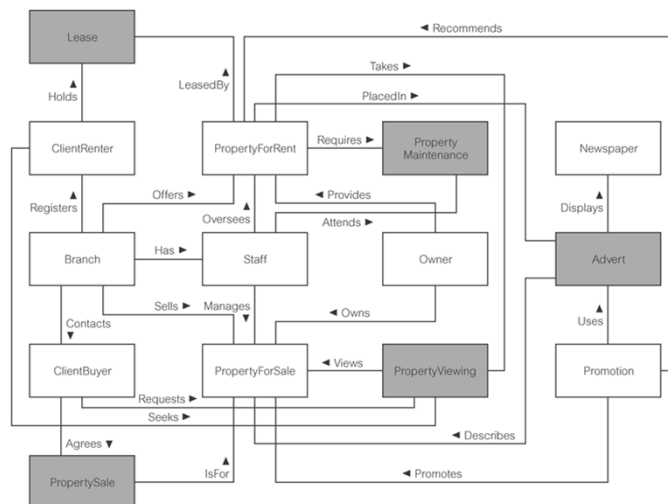
Metodologija načrtovanja DW

■ K1: Izbira procesa (poslovne funkcije)

- Izbira vsebine za področno skladišče.
- Pomembno:
 - izdelava v predvidenem času in finančnem obsegu,
 - pokrivanje ključnih poslovnih vprašanj (tipično prodaja, finance),
 - podatkovna zasnova za nadaljnja področja (ponovno uporabne dimenzije).

■

Metodologija načrtovanja DW

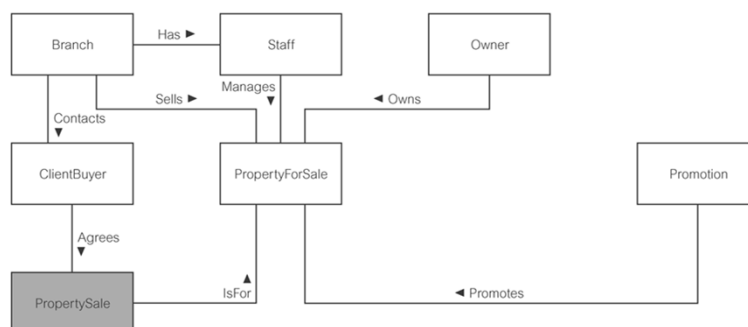


PODATKOVNE BAZE 2 - VSP

- 500 -

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

Metodologija načrtovanja DW



PODATKOVNE BAZE 2 - VSP

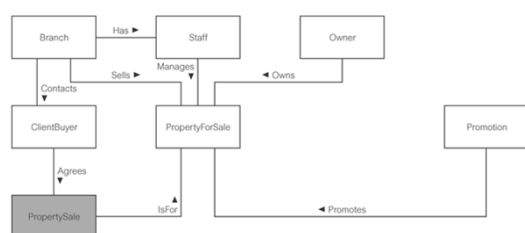
- 501 -

Fakulteta za računalništvo in informatiko
Univerza v Ljubljani

Metodologija načrtovanja DW

■ K2: Izbira zrnatosti

- Kaj bo zajemal zapis v tabeli dejstev
- Priporočilo: najvišja možna zrnatost
- Z izbiro zrnatosti določimo možne dimenzije ter njihovo zrnatost.

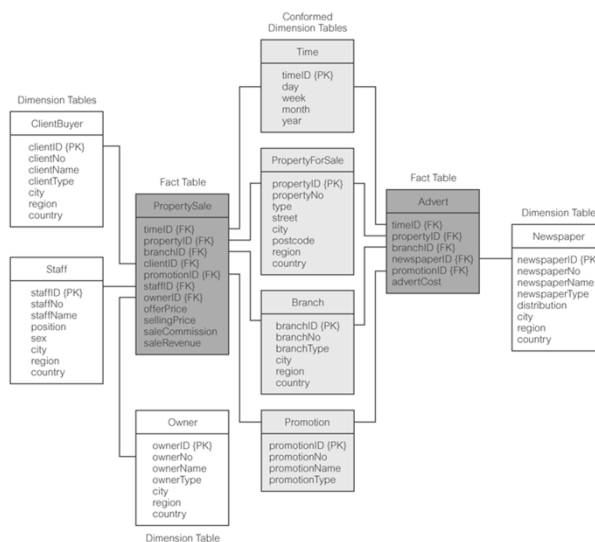


Metodologija načrtovanja DW

■ K3. Identifikacija in prilagajanje dimenzij

- Kako bomo pregledovali tabelo dejstev?
- Slabo izbrane dimenzije zmanjšajo uporabnost skladišča.
- Skladnost dimenzij pomembna za integracijo področnih skladišč.

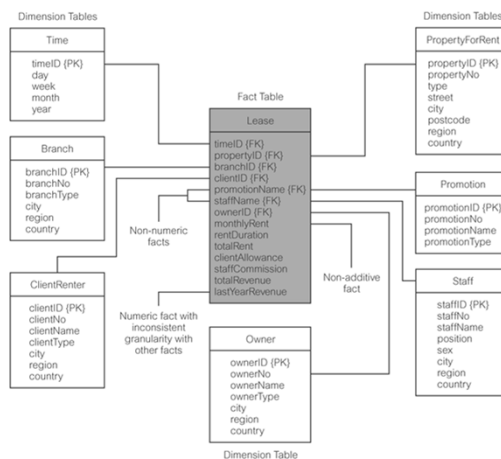
Metodologija načrtovanja DW



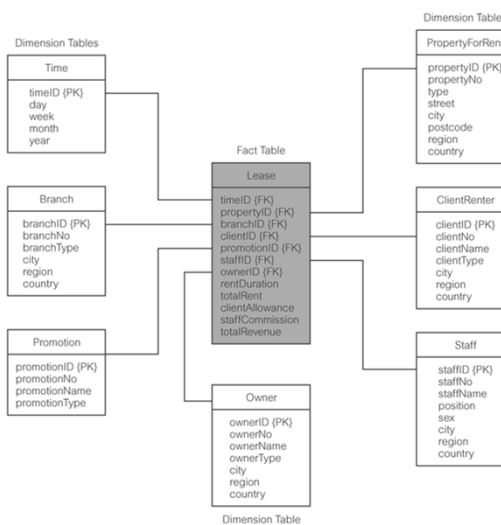
Metodologija načrtovanja DW

- K4. Identifikacija dejstev
 - Dejstva odvisna od izbrane zrnatosti
 - Dejstva naj bodo numerična in seštevna

Primer: slabo strukturirana tabela dejstev



Primer: pravilno strukturirana tabela dejstev



Metodologija načrtovanja DW

- K5. Shranjevanje izračunanih vrednosti v tabeli dejstev
 - Za vsako izbrano dejstvo je potrebo preveriti, ali bi bilo smiselno njegove vrednosti izračunati vnaprej
 - Primer:
 - $\text{totalRevenue} = \text{totalRent} - \text{clientAllowance} - \text{staffCommission}$
 - totalRevenue torej lahko izračunamo iz ostalih dejstev, ali pa ga hranimo ločeno
 - Kadar gre za ključne vrednosti, ki se pogosto uporabljajo, jih pogosto hranimo ločeno in ne izračunavamo sproti. Na ta način tudi olajšamo poizvedovanje s strani uporabnikov.



Metodologija načrtovanja DW

- K6. Podroben opis dimenzij
 - V tej fazi so dimenzije dokončno opredeljene, zato pripravimo podrobne opise le teh.
 - Namen opisov je uporabnikom podatkovnega skladišča omogočiti pravilno razumevanje posameznih dimenzij.

Metodologija načrtovanja DW

▪ K7. Določitev dolžine zgodovine podatkov v tabeli dejstev

- Dolžina zgodovine podatkov se opredeli na podlagi potreb uporabnikov.
- Za analitične potrebe podjetja navadno najbolj potrebujejo pogled 1 do 2 leti nazaj.
- Zaradi zakonskih in drugih zahtev pa se v skladišču pogosto hrani daljša zgodovina, tuid več kot 5 let nazaj.
- Težave v primeru dolge zgodovine:
 - Pridobivanje podatkov za prvo polnjenje, ki so shranjeni v starih formatih, medijih
 - Problem, da dimenzije starih podatkov niso nujno enake (glej K8)

Metodologija načrtovanja DW

▪ K8. Obvladovanje počasi spreminjajočih se dimenzij

- Problem počasi spreminjajočih dimenzij nastane, ko je potrebno za zgodovinske transakcije zaradi sprememb v podatkih uporabljati drugačne dimenzije kot za novejšje transakcije.
- Trije tipi počasi sprminjajočih dimenzij:
 - T1: spremenjeni atribut dimenzije se zamenja (prepiše)
 - T2: spremenjeni atribut dimenzije zahteva dodajanje nove vrednosti v dimenzijsko tabelo
 - T3: spremenjeni atribut dimenzije povzroči nastanek alternativnega atributa v dimenzijski tabeli, tako da je mogoče hkrati dostopati do starih in novih vrednosti znotraj istega zapisa dimenzijske tabele

Metodologija načrtovanja DW

- K9. Opredelitev prioriternih poizvedb
 - V tej fazi so naslavljamo vprašanja fizičnega načrtovanja podatkovnega skladišča:
 - Fizična ureditev zapisov tabele dejstev na disku
 - Odločitev za shranjevanje agregatov
 - Ostale odločitve povezane z administracijo, varnostnimi kopijami, optimizacijo indeksov in varnostjo.

Metodologija načrtovanja DW

- Rezultat načrtovanja je data mart, ki:
 - Podpira potrebe izbranega poslovnega procesa
 - Omogoča enostavno integracijo z drugimi data marti ter posledično oblikovanje celovitega podatkovnega skladišča

Metodologija načrtovanja DW

- Integracija data martov v sozvezdno shemo podatkovnega skladišča
- Ostale data marte prikazane na shemi smo načrtovali po enakem postopku kot data mart s tabelo dejstev Lease

