

Univerza v Ljubljani  
Fakulteta za računalništvo  
in informatiko

6.6.  
2016

Predmet:  
Osnove podatkovnih baz

Modul:  
Podatkovna skladišča

Gradivo:  
v.2015

Univerza v Ljubljani  
Fakulteta za računalništvo  
in informatiko

## Vsebina

- Namen podatkovnih skladišč
- Lastnosti podatkovnih skladišč
- Arhitektura podatkovnih skladišč
- Orodja, ki uporabljajo podatkovna skladišča
- Načrtovanje podatkovnih skladišč

Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

2



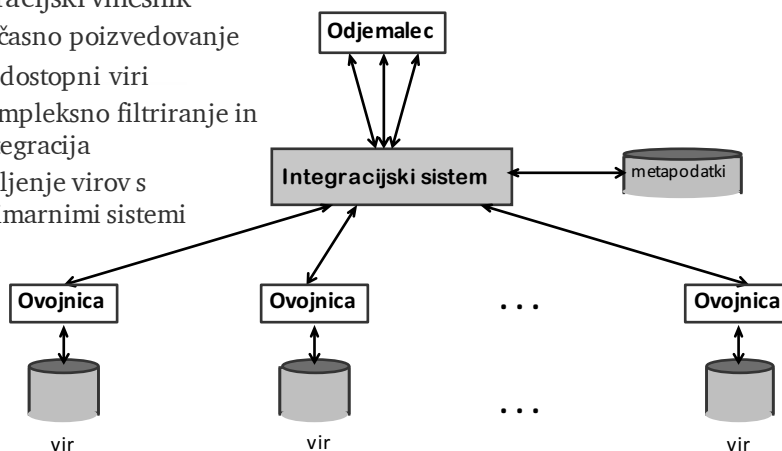
## Zakaj podatkovna skladišča?

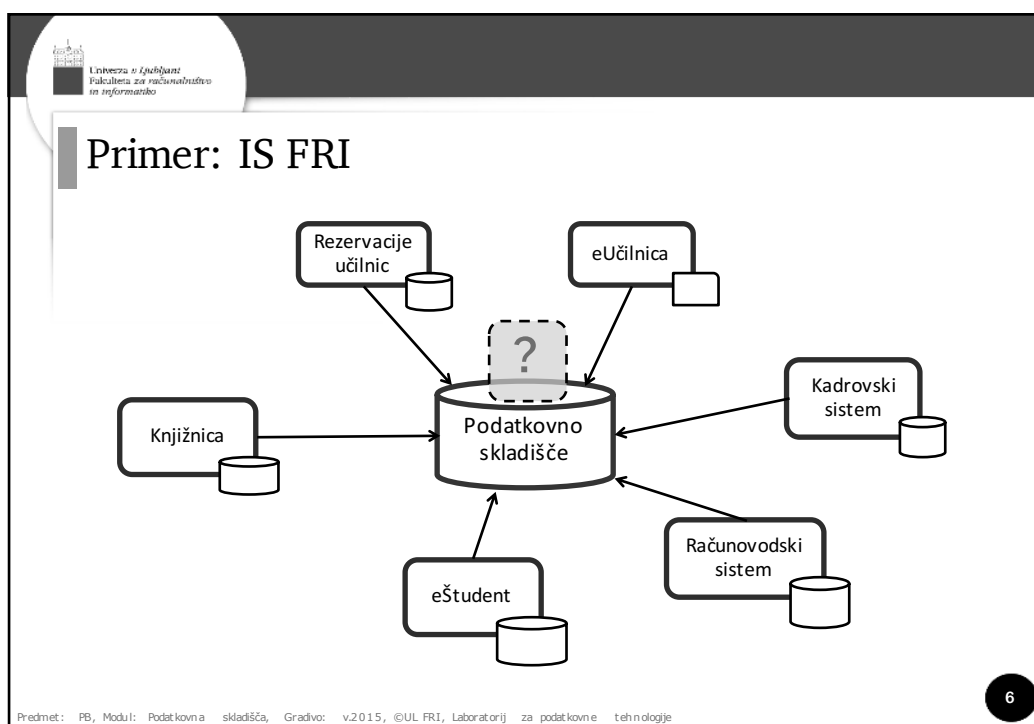
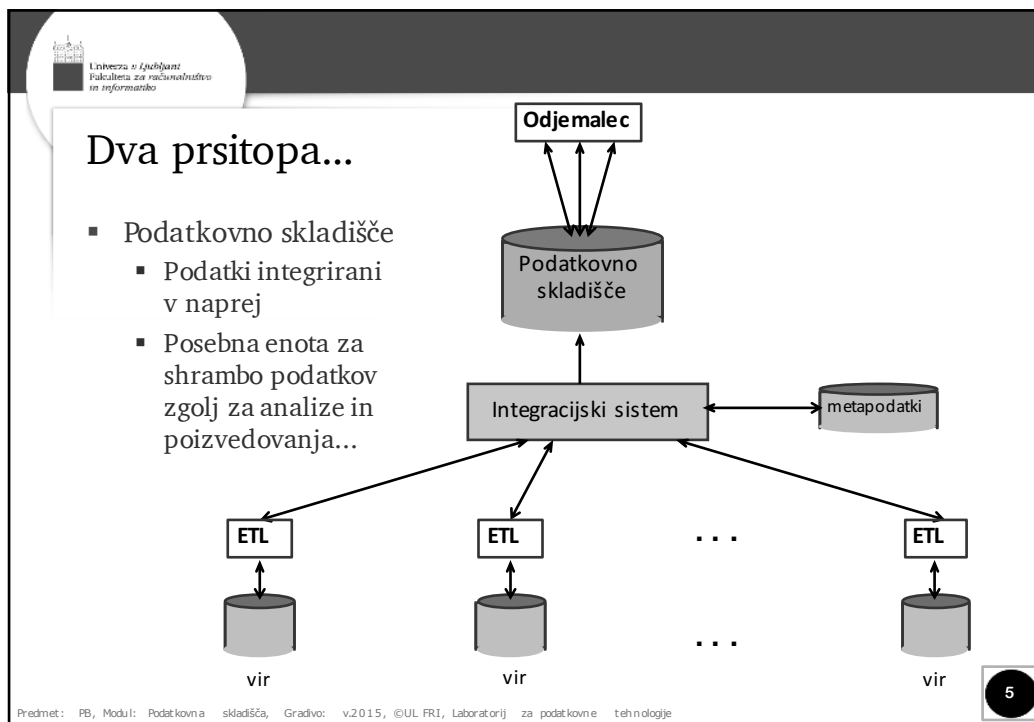
- Za poslovno odločanje so potrebni podatki:
  - iz različnih aplikacij informacijskega sistema, ne glede na uporabljeno platformo in fizično lokacijo aplikacije;
  - za daljše časovno obdobje – podatki, ki omogočajo pregled nad zgodovino vrednosti, ne le trenutne vrednosti podatkov in kazalnikov;
  - za identifikacijo trendov – podatki, ki omogočajo pregled nad trendi, kot so rast, padanje, stagnacija itn.



## Dva pristopa...

- Integracijski vmesnik
  - Počasno poizvedovanje
  - Nedostopni viri
  - Kompleksno filtriranje in integracija
  - Deljenje virov s primarnimi sistemi
  - ...







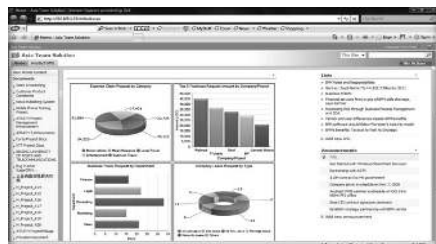
## Evolucija podatkovnih skladišč

- V sedemdesetih aplikativni sistemi → avtomatizacija procesov na operativnem nivoju
- Posledica: akumulacija velikih količin podatkov v PB transakcijskih sistemov
- Danes poudarek različnim načinom uporabe teh podatkov za podporo odločitvenim procesom z namenom pridobiti strateško prednost pred konkurenco!



## Podpora odločanju v transakcijskih IS

- Podpora odločanju na operativnem in taktičnem nivoju tudi v nekaterih transakcijskih IS → poročila, sumarni pregledi, grafi, ...



- Problem več transakcijskih IS. Področja obravnavana ločeno, po svoje...



Univerza v Ljubljani  
Fakulteta za računalništvo  
in informatiko

## Lastnosti podatkovnega skladišča

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



Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©U.L.F.R.I., Laboratorij za podatkovne tehnologije

9



Univerza v Ljubljani  
Fakulteta za računalništvo  
in informatiko

## Vsebinska organizacija podatkov

- Organizacija podatkovnega skladišča temelji na glavnih entitetah podjetja, npr.:
  - študent,
  - izpit,
  - profesor,
  - predmet,
  - račun...
- Ne temelji na funkcionalnih področjih oz. področjih, ki jih pokrivajo posamezni transakcijski sistemi!

Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©U.L.F.R.I., Laboratorij za podatkovne tehnologije

10



## Integriranost

- Podatkovno skladišče integrira podatke iz več aplikativnih sistemov v okviru IS podjetja.
- Podatki večkrat nekonsistentni. Naloga skladišča je, da omogoči konsistenten pogled na podatke!



## Časovna odvisnost

- Podatki v skladišču veljavni za določen čas oziroma interval.
- Za podatek v podatkovnem skladišču moramo poznati čas prenosa in čas veljavnosti.
- Časovna odvisnost večkrat prikazana v razširjenem časovnem formatu (poleg leta še kvartal, četrtletje).
- Časovna odvisnost omogoča opazovanje trendov.



## Nespremenljivost

- Podatki v podatkovnem skladišču niso podvrženi spremembam v realnem času s strani aplikacij!
- Osveževanje (iz transakcijskih sistemov in ostalih virov) z neko (smiselno) frekvenco
  - Osveževanje: večinoma le dodajanje novih podatkov v podatkovno skladišče.



## OLAP in OLTP..

- OLAP – On Line Analytical Processing (vir podatkovno skladišče)
- OLTP – On Line Transaction Processing (vir transakcijska baza)
- Za OLAP in OLTP praviloma različne instance podatkovne baze!



## OLAP in OLTP..

OLTP	OLAP
Obdelujejo sveže (trenutne) podatke	Obdelujejo zgodovinske podatke
Baza vsebuje podrobne podatke	Baza vsebuje podrobne in sumarizane podatke
Podatki so dinamični	Podatki so statični
Postopki uporabe se ponavljajo - transakcije so predvidljivi	Postopki uporabe (analize) nepredvidljivi
Zelo veliko število transakcij na časovno enoto	Majhno število transakcij
Transakcijsko usmerjeni sistemi	Analitično usmerjeni
Aplikacijsko naravnani	Predmetno naravnani
Podpora vsakodnevnim odločitvam	Podpora taktičnim in strateškim odločitvam
Veliko število operativnih uporabnikov	Manjše število vodstvenih uporabnikov



## OLAP vs 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.





Univerza v Ljubljani  
Fakulteta za računalništvo  
in informatiko

## OLAP vs OLTP..

- Primeri vprašanj za OLAP
  - 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?
  - 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?

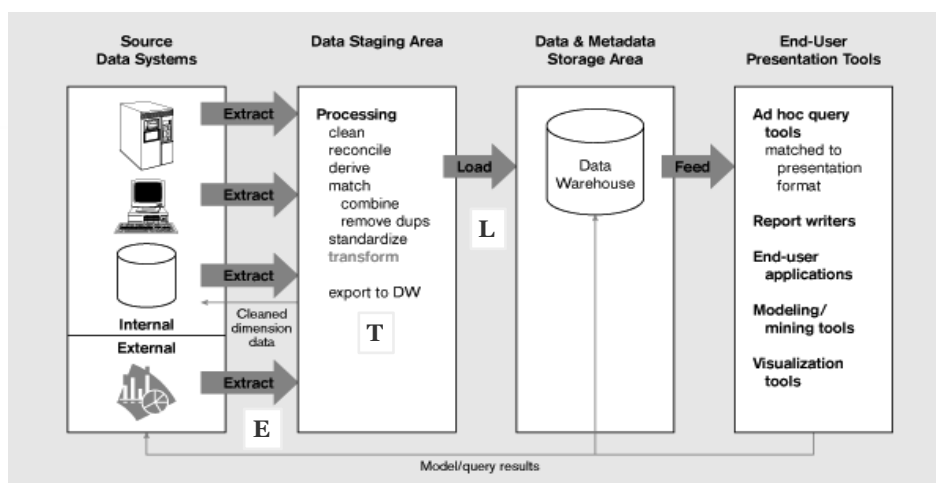
Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

17



Univerza v Ljubljani  
Fakulteta za računalništvo  
in informatiko

## Generična dvo-nivojska arhitektura



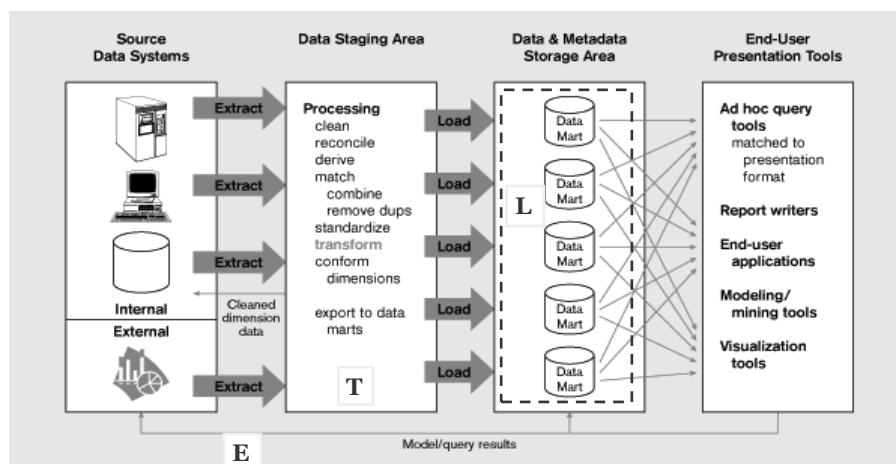
Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

18



Univerza v Ljubljani  
Fakulteta za računalništvo  
in informatiko

## Neodvisna področna skladišča



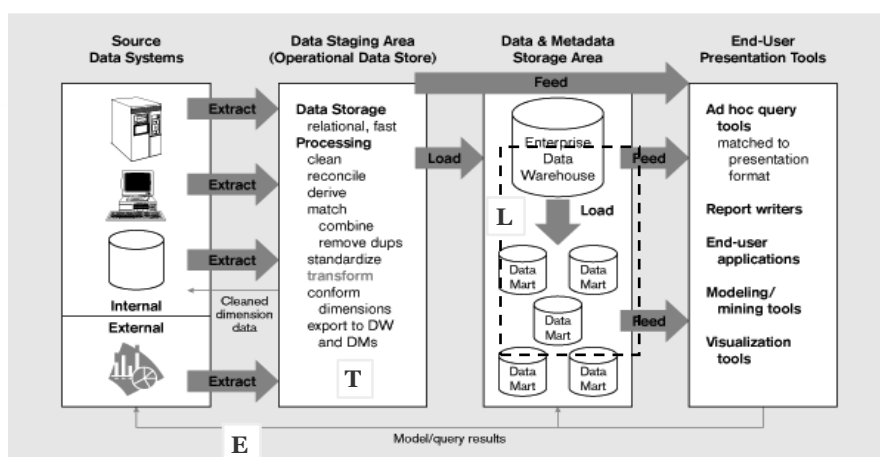
Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

19



Univerza v Ljubljani  
Fakulteta za računalništvo  
in informatiko

## Odvisna področna skladišča



Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

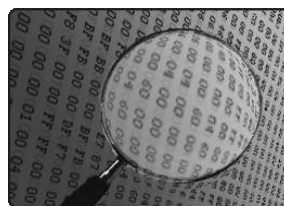
20



Univerza na Ljubljani  
Fakulteta za računalništvo  
in informatiko

## Orodja za dostop do DW

- DW mora podpirati učinkovite ad-hoc in rutinske analize.
  - Visoka učinkovitost dosežena s pred-planiranjem povezav, agregacij, periodičnih poročil itn.
- Tipična orodja končnih uporabnikov:
  - Orodja za poročila in poizvedovanje,
  - Razvojna orodja,
  - Direktorski IS,
  - OLAP orodja,
  - Data mining orodja.



Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©U.L.F.R.I., Laboratorij za podatkovne tehnologije

21



Univerza na Ljubljani  
Fakulteta za računalništvo  
in informatiko

## Generators

- Priprava in standardizacija poročil (s



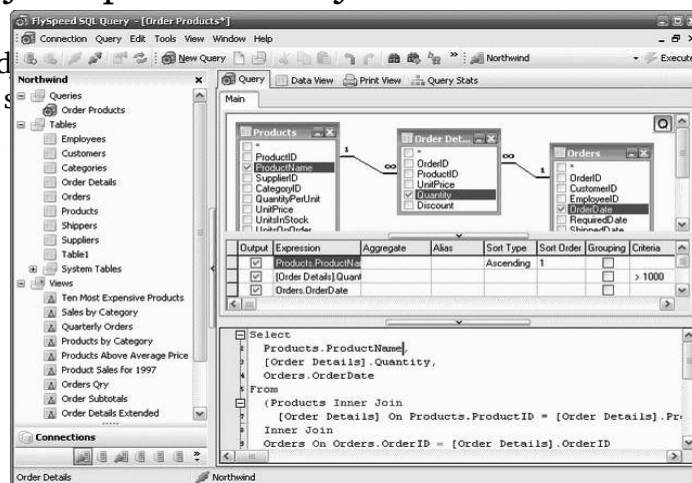
pravo  
generatorji

Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©U.L.F.R.I., Laboratorij za podatkovne tehnologije

22

## Orodja za poizvedovanje

- Orodja za poizvedovanje: Orodja za generiranje poročil ali poizvedovalni orodja, potrebna za razvoj uporabniških aplikacij



Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

23

## 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 mnogimi OLAP orodji.

Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

24

Univerza na Kijevu  
Fakulteta za računalništvo  
in informatiko

## Dir

Application Navigator

Business Summary (Snapshot of business activity for the current month and calendar year)

Customer Aging Summary (Breakdown of open balances by aging period)

Sales by Customer Class (One year at sales by customer class)

Profit by Customer Class (One year at gross profits by customer class)

Sales by Customer Class by Month (One year sales by month with calendar year)

Sales by Customer Class by Year (Sales for a range of years encompassing sales)

Top Customer Sales One Year (One year sales for customers who were sold)

Top Customer Sales Two Years (One year sales for customers who were sold)

Item Item Class

Subcategory

Inventory Summary (Vendor Class)

Sales & Profit Analysis

Sales & Profit Analysis by Fiscal Period (Comparison of sales, gross profit and net profit)

Sales & Profit Analysis by Fiscal Year (Comparison of sales, gross profit and net profit)

Net Profit % of Calendar Period (Comparison of net profit and gross profit)

Net Profit % of Category Type (Comparison of net profit and all category type sales)

Working Capital, Inventory Turnover

Working Capital (Analysis of working capital by fiscal period)

Inventory Turnover & Avg. Days in Inventory (Analysis of inventory turnover by fiscal period)

Assets By Category (Assets categorized by account category at end of fiscal year)

Liabilities By Category (Liabilities categorized by account category at end of fiscal year)

Equity By Category (Equity categorized by account category at end of fiscal year)

Revenue By Category (Revenue categorized by account category for a range of years)

Expenses By Category (Expenses categorized by account category for a range of years)

Financial Ratios by Fiscal Period

Current Ratio (Current Assets / Current Liabilities)

Net Income To Sales (Net Income / Sales)

Net Income To Working Capital

Net Income To Net Worth

Sales To Working Capital

Sales To Inventory

Fixed Asset To Total Net Worth

Inventory To Working Capital

Current Liability To Inventory

Debt To Tangible Net Worth (Equity Ratio)

Business Summary As Of 11/8/2002 16:11:23

Order

Month To Date

Calendar Y-T-D

Order Count

Order Value

Purchase Order

Month To Date

Calendar Y-T

Order Count

Order Value

Accounts Payable

Current Balance

Month To Date

Calendar Y-T

Purchases

Average Purchase

Payments

Average Payment

Trade Discounts

Terms Disc. Taken

Inventory

Average Cost

Assets By Category

Computers, Inc.

Assets By Account Category, Fiscal Year 2002

Cash

Inventory

Accounts Receivable

Ready (Company: Computers, Inc. User: ADMIN Date: 11/8/2002)

Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

25

Univerza na Kijevu  
Fakulteta za računalništvo  
in informatiko

## OLAP orodja

Roll up (drill-up)  
Drill down (roll down)  
Slice and dice

Products

Months

Measures

Shoes

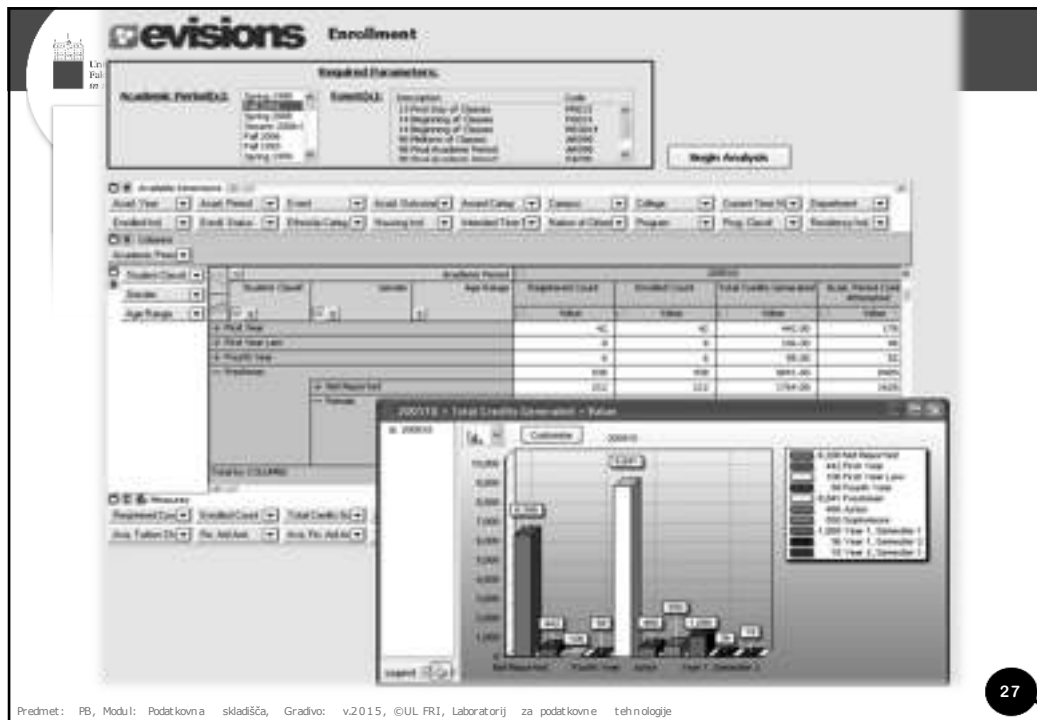
Measure

	Units	Revenue	Cost
January	250	1564	1020
February	200	1275	875
March	350	1800	1275
April	400	1935	1500
May	485	2000	1560

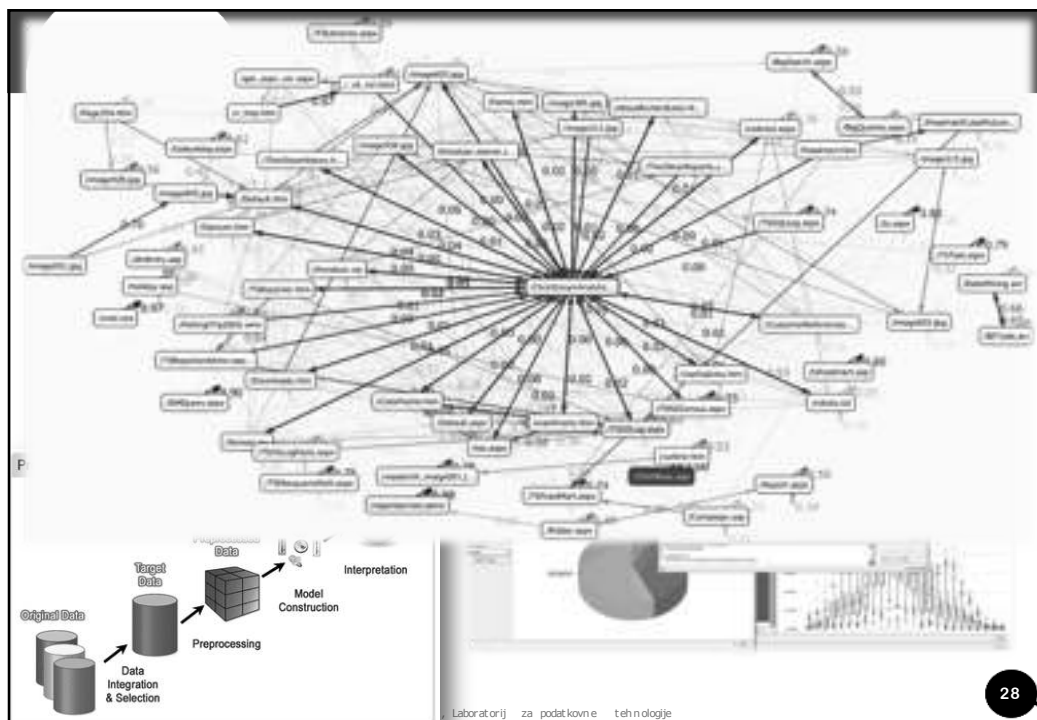
Product: Shoes

Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

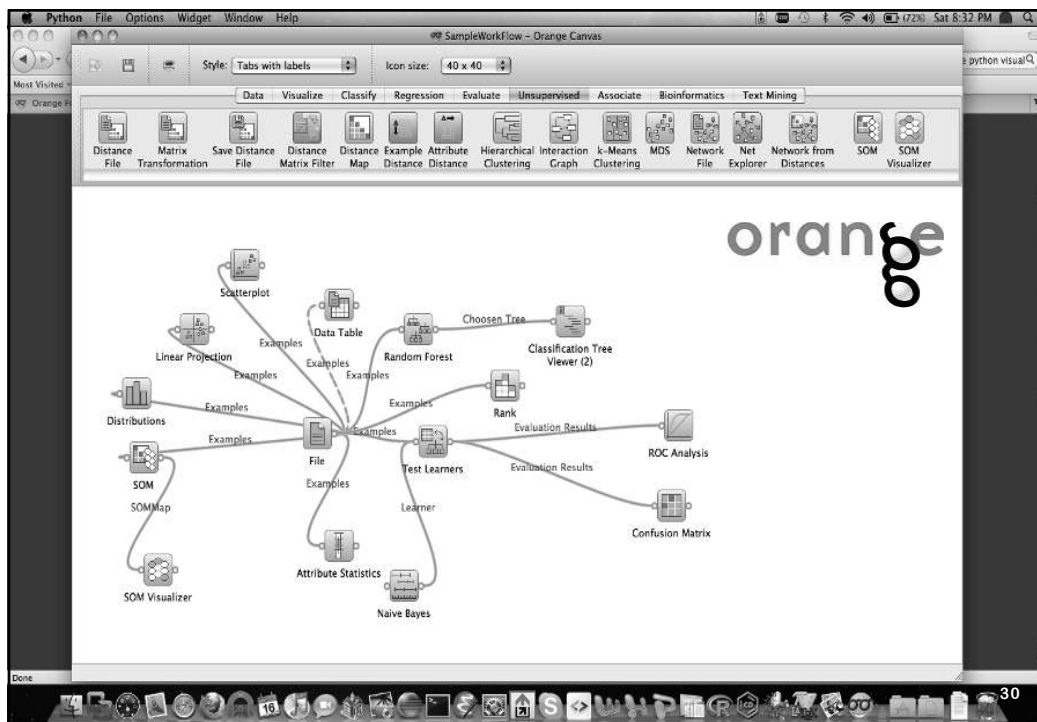
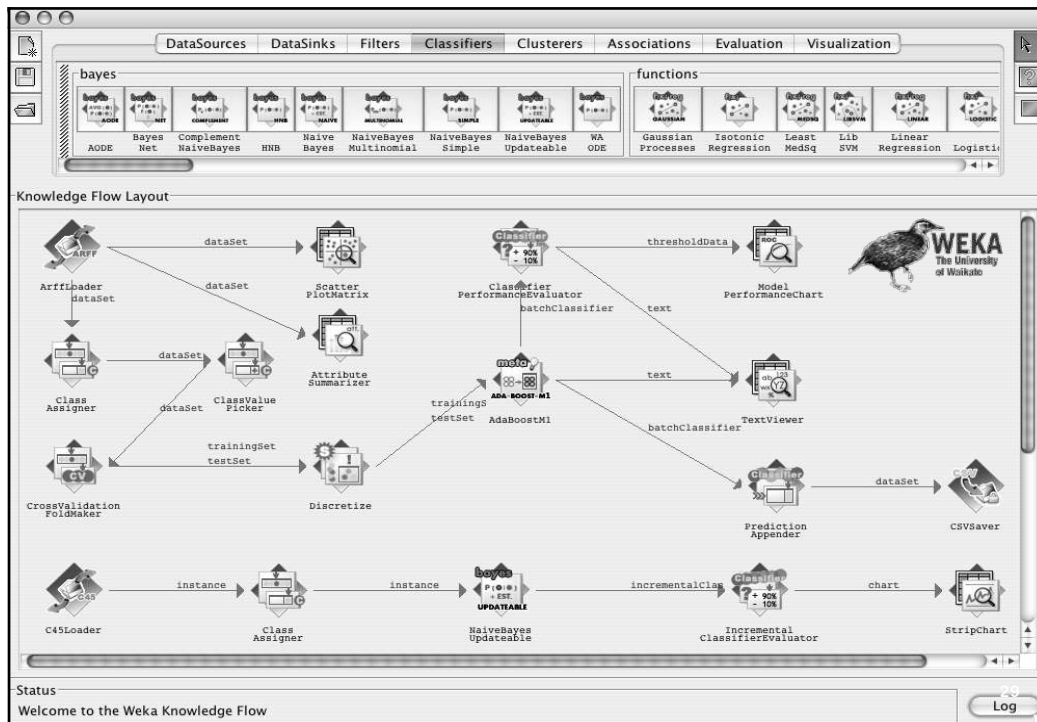
26



27



28





Univerza v Ljubljani  
Fakulteta za računalništvo  
in informatiko

## Načrtovanje podatkovnega skladišča...

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



Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

31



Univerza v Ljubljani  
Fakulteta za računalništvo  
in informatiko

## Načrtovanje podatkovnega skladišča

- Ključni koraki:
  - Izbira področja/vsebine podatkovnega skladišča
  - Določitev zrnatosti podatkov v tabeli dejstev
  - Izbira dimenzij
  - Identifikacija dejstev
  - Določitev atributov dimenzij
- Pomembno: pri zajemu zahtev se pogovorimo z:
  - odgovornimi za posamezna poslovna izbrana poslovna področja (npr. marketing, prodaja, finance, vodstvo...),
  - administratorji podatkovnih virov – kateri podatki obstajajo?

Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

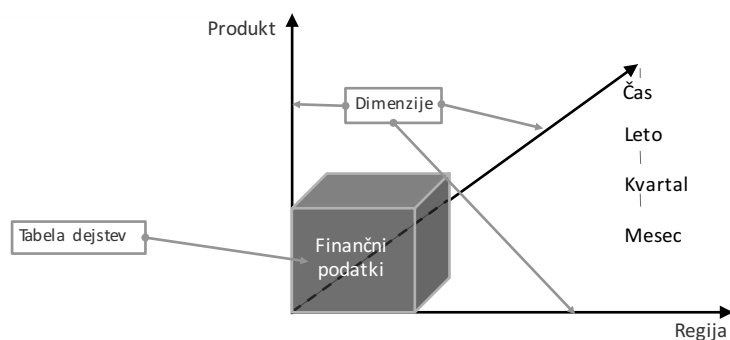
32





## Modeliranje dimenzij...

- Logično načrtovanje podatkov na način, da bodo hitro dostopni...

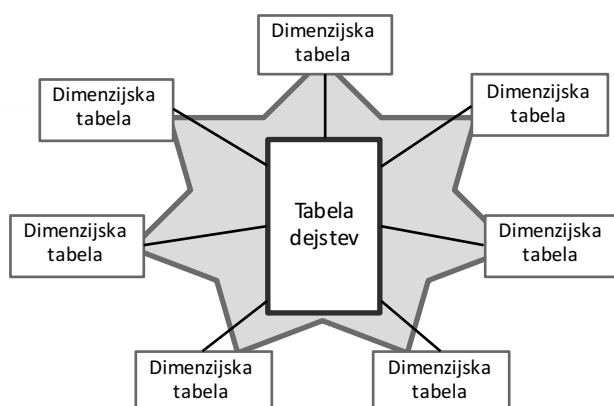


## Modeliranje dimenzij...

- 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
- Dejstva se generirajo na podlagi dogodkov, ki so se zgodila v preteklosti – spremembe malo verjetne.
- Primarni ključ tabele dejstev je sestavljen iz dveh ali več tujih ključev posameznih dimenzijskih tabel.
- Tabela dejstev skupaj z dimenzijskimi tabelami tvori ti. zvezdno shemo.

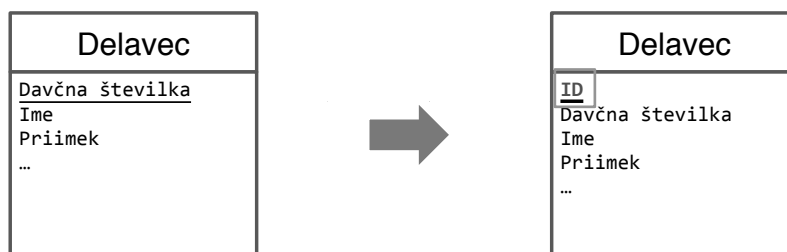


## Zvezdna shema



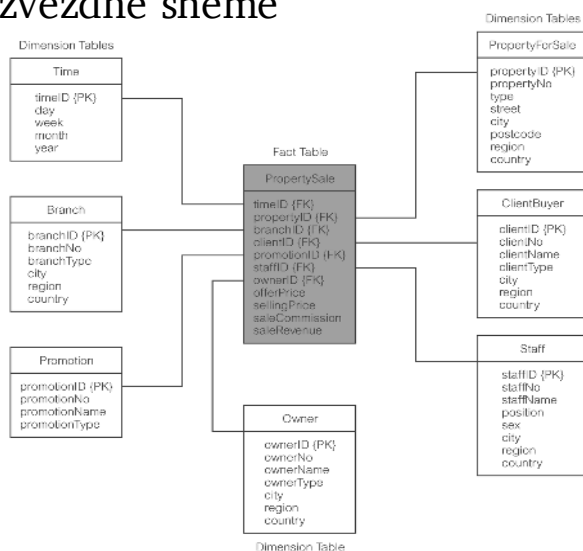
## Uporaba umetnih ključev

- Vse naravne primarne ključe se nadomesti s surogati – umetni ključi.
- Surogati omogočajo neodvisnost med podatki podatkovnega skladišča in posameznimi viri.





## Primer zvezdne sheme



Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

37



## Podatki v tabelah dimenzij in dejstev

- Podatki v tabelah dejstev
  - referenčni podatki – namenjeni branju,
  - skozi čas se ne spreminjajo (le dodajajo),
  - numerične vrednosti.
  - Večina podatkov v podatkovnem skladišču se hrani v tabeli dejstev → lahko ekstremno obsežne tabele.

Numerične vrednosti (dogodki)

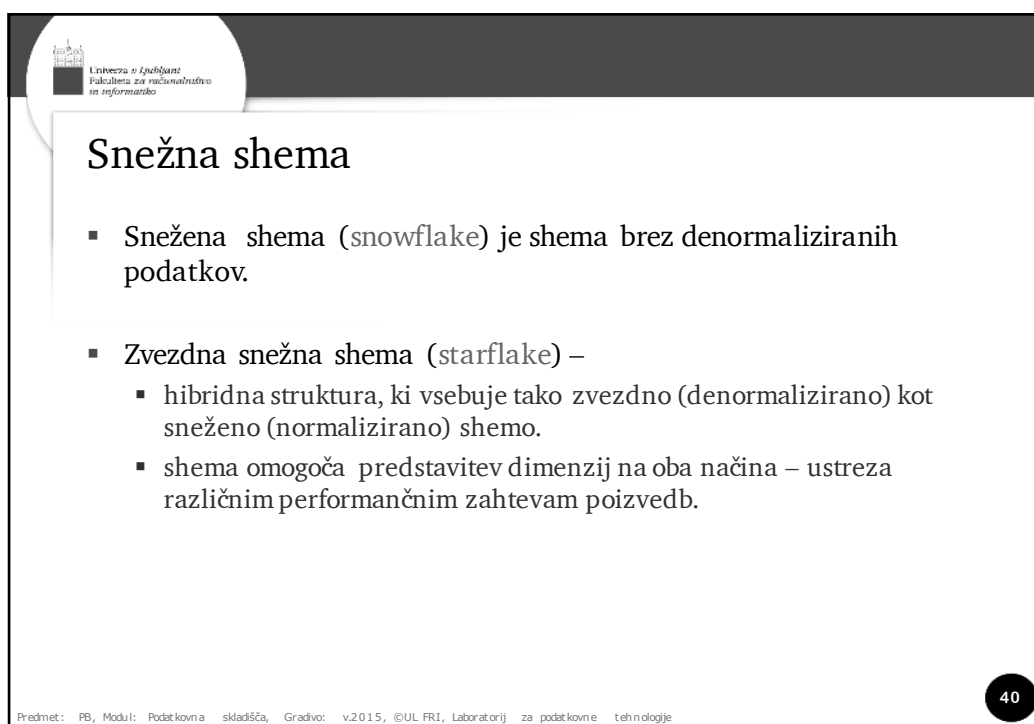
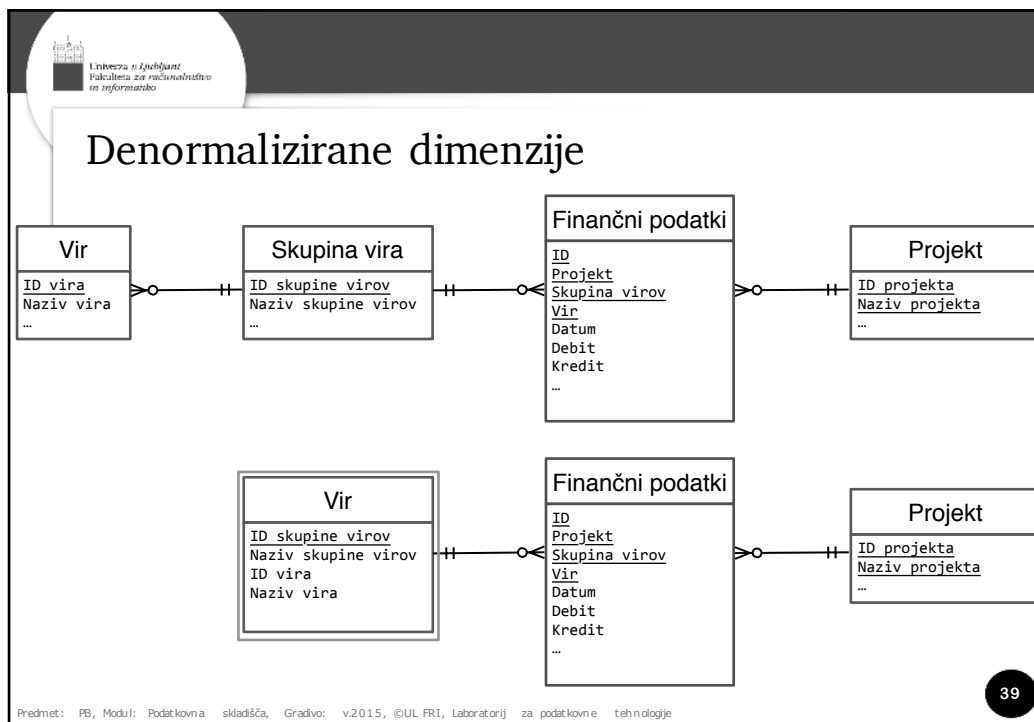
Tabela dejstev

M1	M2	M3	M4	M5

- Podatki v tabelah dimenzij
  - navadno tekstovne vrednosti
  - filter

Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

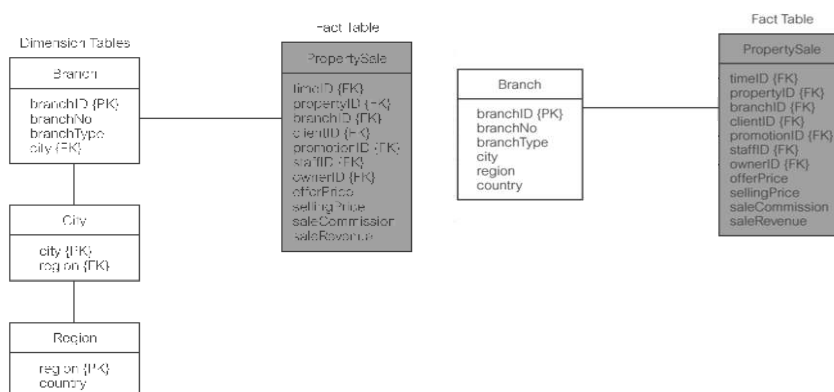
38



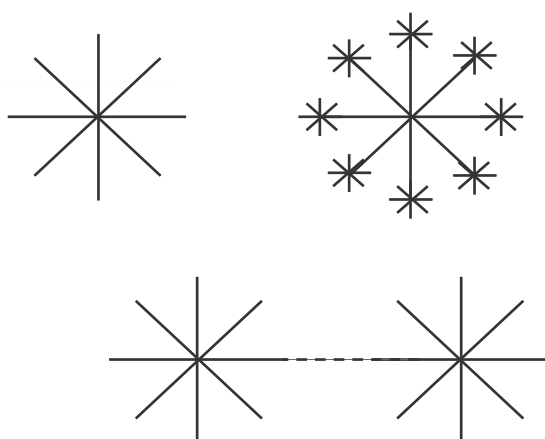


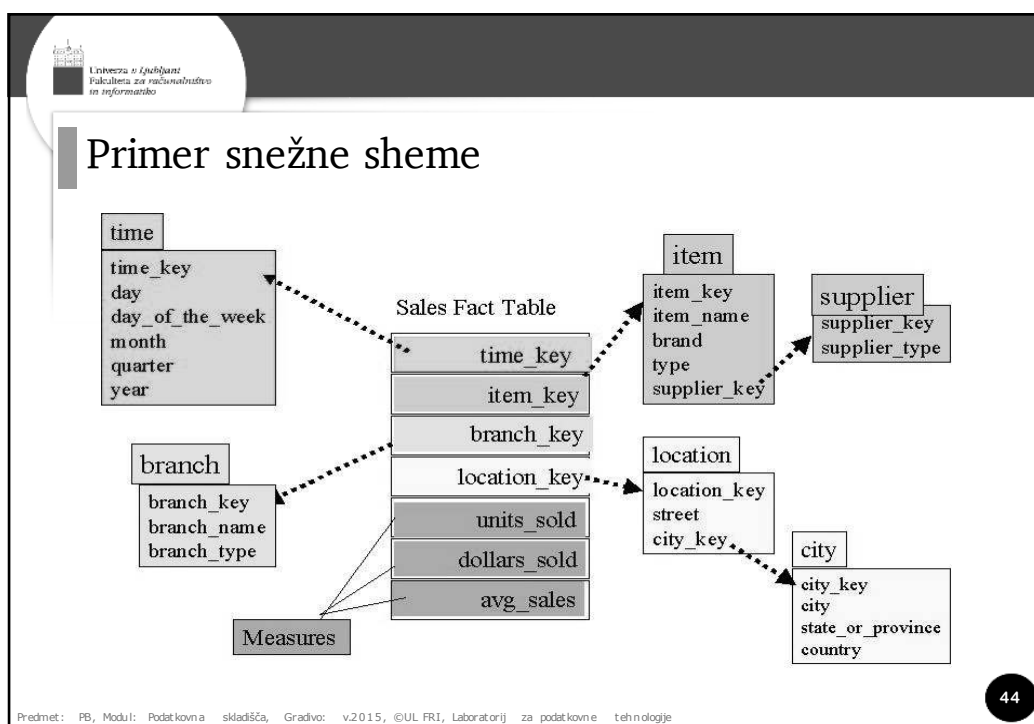
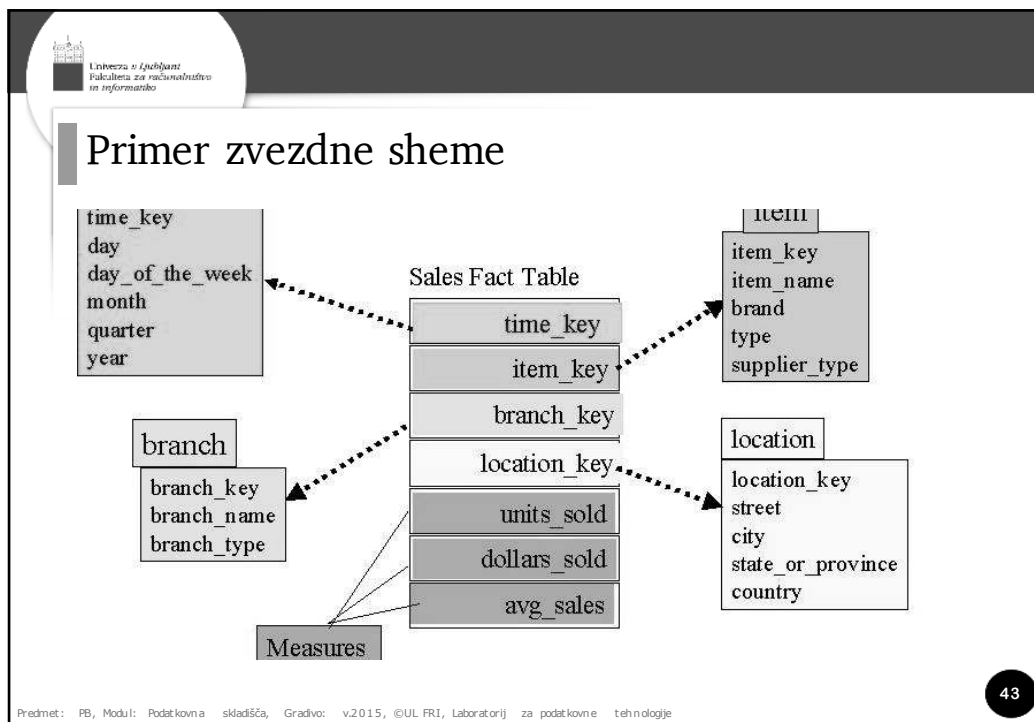
## Modeliranje dimenzij...

- Normalizirana vs denormalizirana varianta dimenzije Branch:



## Zvezdna, snežna, sozvezdna shema...

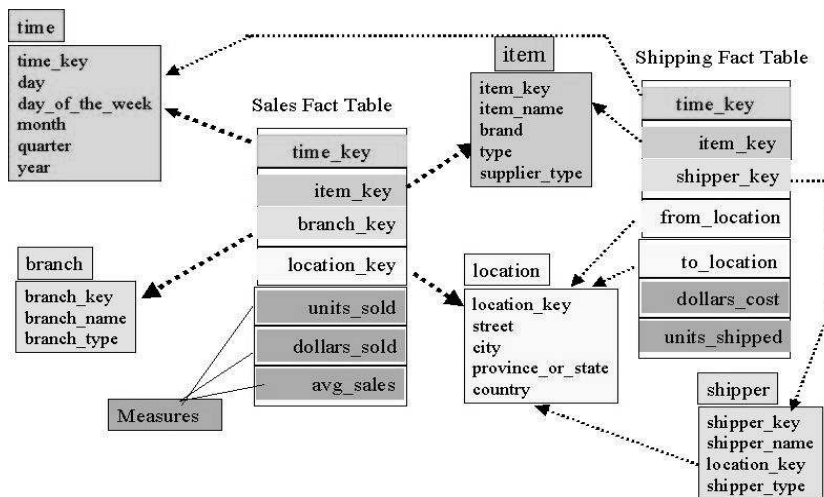






Univerza na Ljubljani  
Fakulteta za humanistične  
in informacijske

## Primer sozvezdna shema



Predmet: PB, Modul: Podatkovna skladišča, Gradivo: v.2015, ©UL FRI, Laboratorij za podatkovne tehnologije

45