

UNIVERZITET DŽEMAL BIJEDIĆ U MOSTARU  
FAKULTET INFORMACIJSKIH TEHNOLOGIJA MOSTAR

SEMINAR

# **Analiza poslovnih podataka sa "open source" software-om**

Student: *Ernad Husremović, DL 2792*

Mentor: *prof.dr Vanja Bevanda*

ver: 1.9.7

Mostar, februar 2012.

# SADRŽAJ

<b>1. Uvod</b>	<b>1</b>
1.1. BI Pojmovi . . . . .	1
1.1.1. ETL . . . . .	1
1.1.2. Ulazni podaci - ERP podaci . . . . .	1
<b>2. OLAP Case study: Analiza podataka ‘F18 knowhow’ ERP-a</b>	<b>2</b>
<b>3. Iza case study-ja ?</b>	<b>11</b>
3.1. Ne znam . . . . .	11
3.1.1. Mondrian . . . . .	11
3.1.2. datamart vs datawarehouse . . . . .	11
3.1.3. Konstrukcija OLAP kocke . . . . .	11
3.2. Pentaho . . . . .	12
3.3. Spoon . . . . .	12
3.4. Data mining . . . . .	12
3.4.1. dimension table . . . . .	13
3.4.2. facts table . . . . .	13
3.4.3. ETL (Extract Transform Load) . . . . .	13
3.5. Poslovna pitanja (Business questions) . . . . .	13
3.6. Analiza podataka . . . . .	13
3.6.1. Redovi, Kolone, Filteri . . . . .	15
3.6.2. Ekspert . . . . .	15
<b>4. Zaključak</b>	<b>18</b>
<b>5. Literatura</b>	<b>19</b>
<b>6. Rezime</b>	<b>20</b>

<b>A. Korišteni alati</b>	<b>21</b>
<b>B. Izvorni kod, dostupni resursi</b>	<b>22</b>
<b>C. Bilješke autora</b>	<b>23</b>

# **1. Uvod**

## **1.1. BI Pojmovi**

### **1.1.1. ETL**

### **1.1.2. Ulazni podaci - ERP podaci**

## 2. OLAP Case study: Analiza podataka ‘F18 knowhow’ ERP-a

The screenshot displays the F18 ERP application interface. The top window shows a list of transactions with columns: RJ, VD, Brdok, VP, Datum, Partner, Ukupno-Rab, Rabat, Ukupno, and Datum placanja. The bottom window shows a detailed invoice (RAČUN/OTPREMNICA br. 00123) with columns: kolicina, C.b.PDV, and Uk.bez.PDV. The invoice details include a list of items, their quantities, and prices, followed by a summary of the total amount and tax.

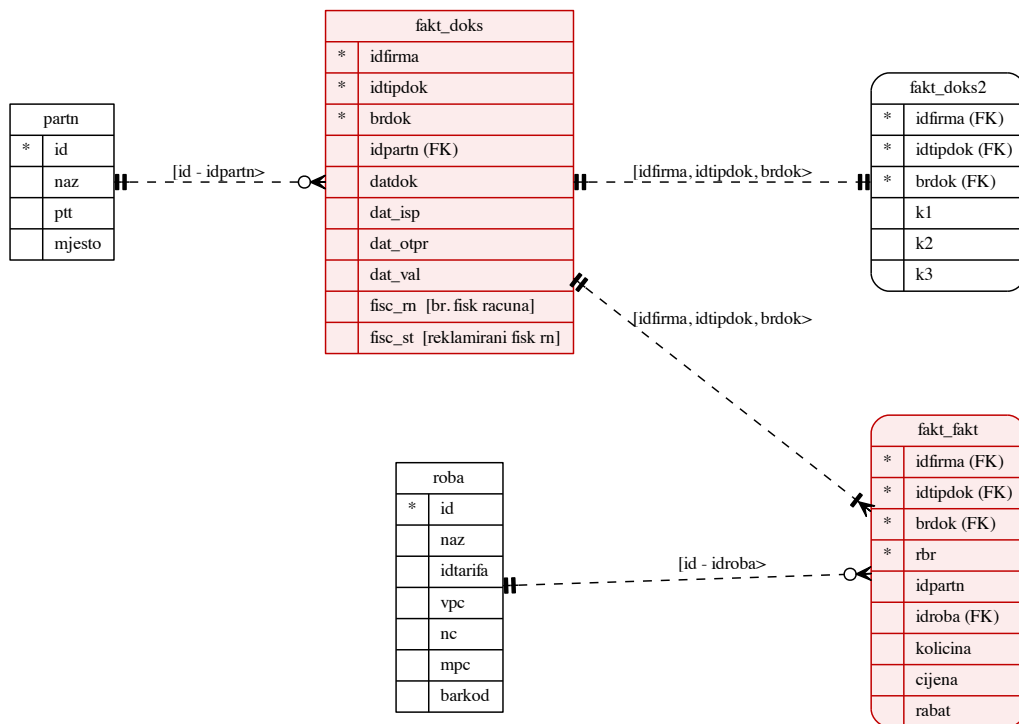
RJ	VD	Brdok	VP	Datum	Partner	Ukupno-Rab	Rabat	Ukupno	Datum placanja
10	10	00101		06.03.07	SVJE				
10	10	00102		06.03.07	FINK				
10	10	00103		06.03.07	DA-4				
10	10	00104		06.03.07	DIKI	13			
10	10	00105		06.03.07	VIND	14			
10	10	00106		06.03.07	ENER	15			
10	10	00107		06.03.07	RAHA	16			
10	10	00108		06.03.07	CET	17			
10	10	00109		06.03.07	MAKS	18			
10	10	00110		06.03.07	AUTO	19			
10	10	00111		06.03.07	OPCI	20			
10	10	00112		06.03.07	MBH	21			
10	10	00113		06.03.07	BETR	22			
10	10	00114		06.03.07	FUEL	23			
10	10	00115		06.03.07	ZAVO	24			
10	10	00116		06.03.07	HORI	25			
10	10	00117		06.03.07	STAN	26			
10	10	00118		06.03.07	ZEDE	27			
10	10	00119		06.03.07	PRIV	28			
10	10	00120		14.03.07	RAHA	29			
10	10	00121		14.03.07	STAN	30			
10	10	00122		14.03.07	CENT	31			
10	10	00123		14.03.07	CENT	32			
10	10	00124		15.03.07	OPCI	33			
10	10	00125		15.03.07	OPCI	34			
10	10	00126		16.03.07	OPCI	35			
10	10	00127		16.03.07	OPCI	36			

kolicina	C.b.PDV	Uk.bez.PDV
1.000	25.00	25.00
1.000	30.00	30.00
1.000	50.00	50.00
Ukupno bez PDV (KM) :		105.00
PDV 17% :		17.85
** SVEUKUPNO SA PDV (KM) :		122.85

Slika 2.1: ERP aplikacija, F18 klijent

Operativni podaci ‘F18 knowhow’ smješteni su u ovaj relacijski model:



**Slika 2.2:** F18 transakcijski db model (relevantni dio)

Podaci svake poslovne godine nalaze se u posebnoj PostgreSQL bazi podataka.

Cleansing

F18 'cleansing' podaci (Dodatak B, olap\_cleansing 'spreadsheet' dokument)

Klasificiranje izvornih podataka - šifarnik artikala

olap\_cleansing.xls - LibreOffice Calc

Find

	A	B	C	D	E	F	G	H
1	idroba4	kategorija	grupa					
2	7ADV	SW	3-rd party software					
3	7LXV	SW	3-rd party software					
4	7LXW	SW	3-rd party software					
5	7MSO	SW	3-rd party software					
6	9ADV	SW	3-rd party software					
7	9AV0	SW	3-rd party software					
8	9DOD	SW	3-rd party software					
9	9MSO	SW	3-rd party software					
10	9XBA	SW	3-rd party software					
11	MSWI	SW	3-rd party software					
12	WSER	HW	3-rd party software					
13	WXPH	SW	3-rd party software					
14	WXPP	SW	3-rd party software					
15	NSC-	HW	fiskalni uređaji					
16	6OBR	SW	fmk software					
17	9FMK	SW	fmk software					
18	9MOD	SW	fmk software					

Sheet 3 / 3 PageStyle\_artikli\_kod STD Sum=0 100%

Slika 2.3: F18 klasificiranje - šifarski sistem artikala

olap\_cleansing.xls - LibreOffice Calc

Find

	A	B	C	D	E	F	G	H
1	Mjestoln	Mjesto						
2	BEGOV HAN	Begov Han						
3	ŽELJEZNO POLJE	Begov Han						
4	BIHAĆ	Bihać						
5	BUGOJNO	Bugojno						
6	BUSOVAČA	Busovača						
7	DONJI VAKUF	Donji Vakuf						
8	GORA@DE	Goražde						
9	GORAŽDE	Goražde						
10	GORAŽDE	Goražde						
11	U.S.A	inostranstvo						
12	U.S.A.	inostranstvo						
13	DEUTSCHLAND	inostranstvo						
14	USA	inostranstvo						
15	KAKANJ	Kakanj						
16	KISELJAK	Kiseljak						
17	KONJIC	Konjic						
18	MAGLAJ	Maglaj						

Sheet 1 / 3 PageStyle\_mjesto STD Sum=0 100%

Slika 2.4: 'cleansing' F18 podataka - klijenti - mjesta/gradovi

olap\_cleansing.xls – LibreOffice Calc

Arial 10 B I U

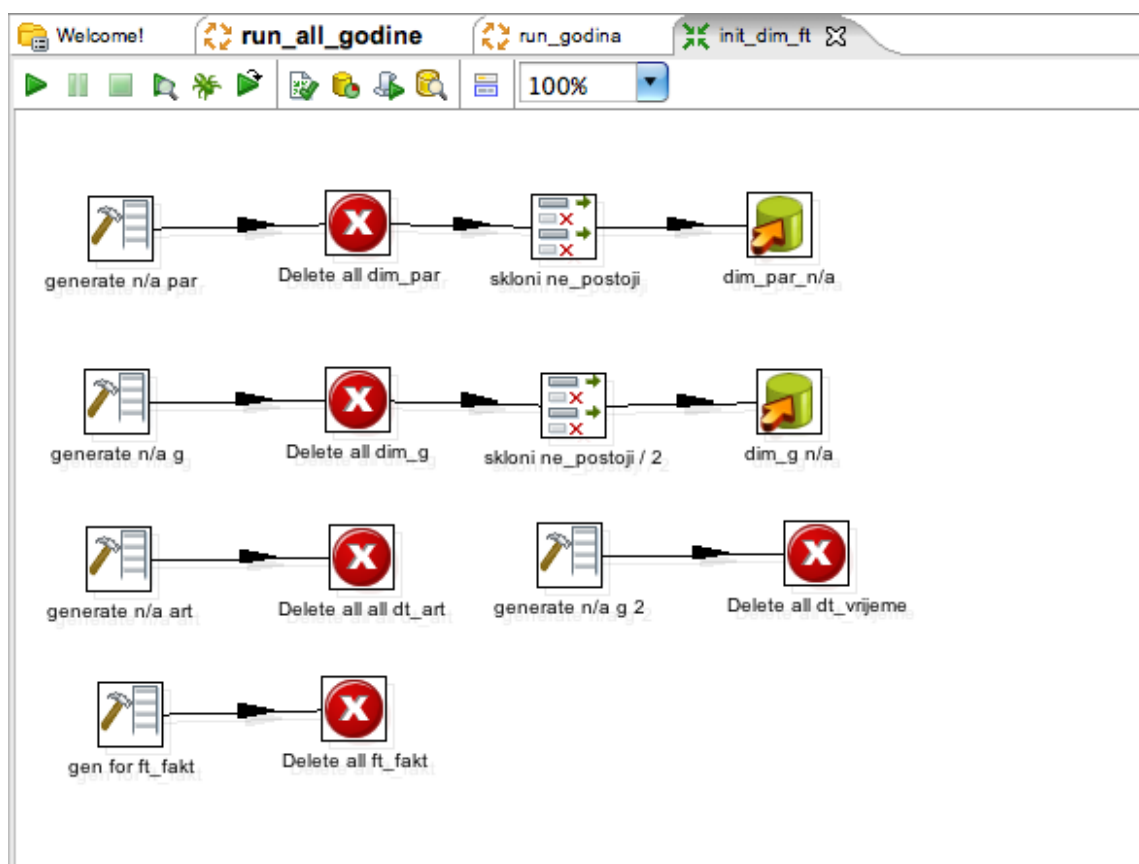
A1 Mjesto

	A	B	C	D	E	F	G	H
1	Mjesto	Oznaka						
2	Begov Han	ŽEP						
3	Bihać	BI						
4	Bugojno	BUG						
5	Busovača	VIT						
6	Donji Vakuf	BUG						
7	Goražde	GOR						
8	inostranstvo	INO						
9	Kakanj	ZE						
10	Kiseljak	KIS						
11	Konjic	KO						
12	Maglaj	ZAV						
13	n/a	XX						
14	Olovo	OLO						
15	Sanski Most	BI						
16	Sarajevo	SA						
17	Tešanj	TEŠ						
18	Travnik	TRA						

Find

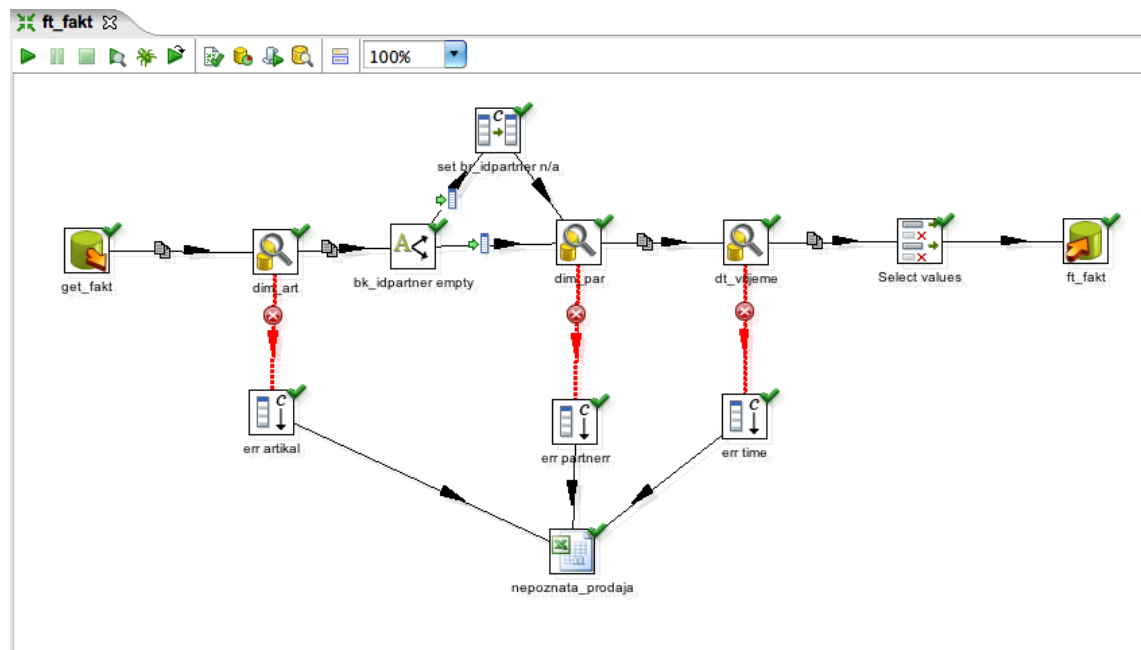
Sheet 2 / 3 PageStyle\_mjesto\_kod STD Sum=0 100%

Slika 2.5: F18 kodiranje regiona - klasifikacija mjesta/gradova

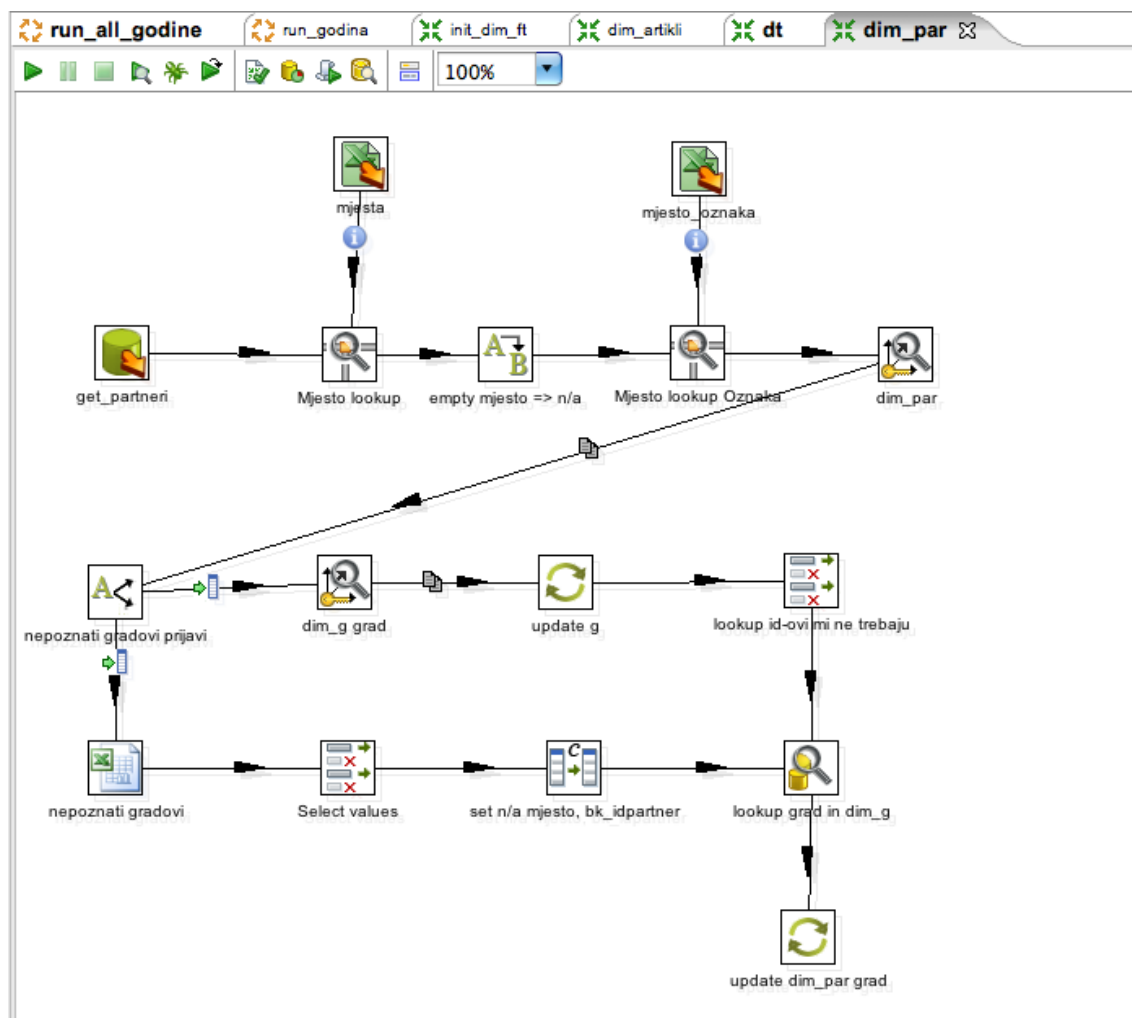


Slika 2.6: Inicijalizacija 'dimension' i 'facts' tabela

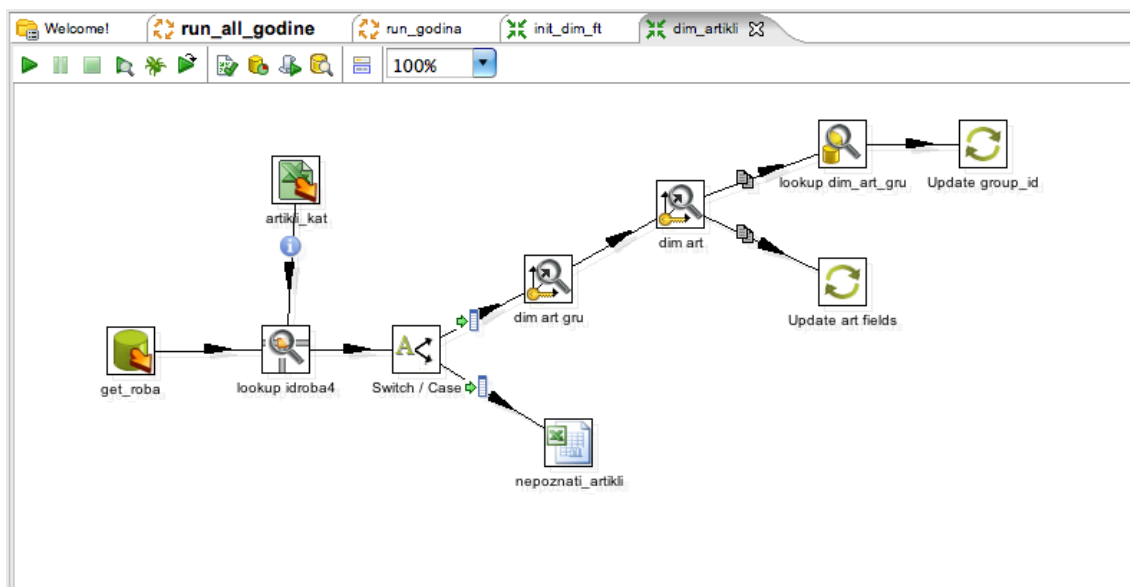




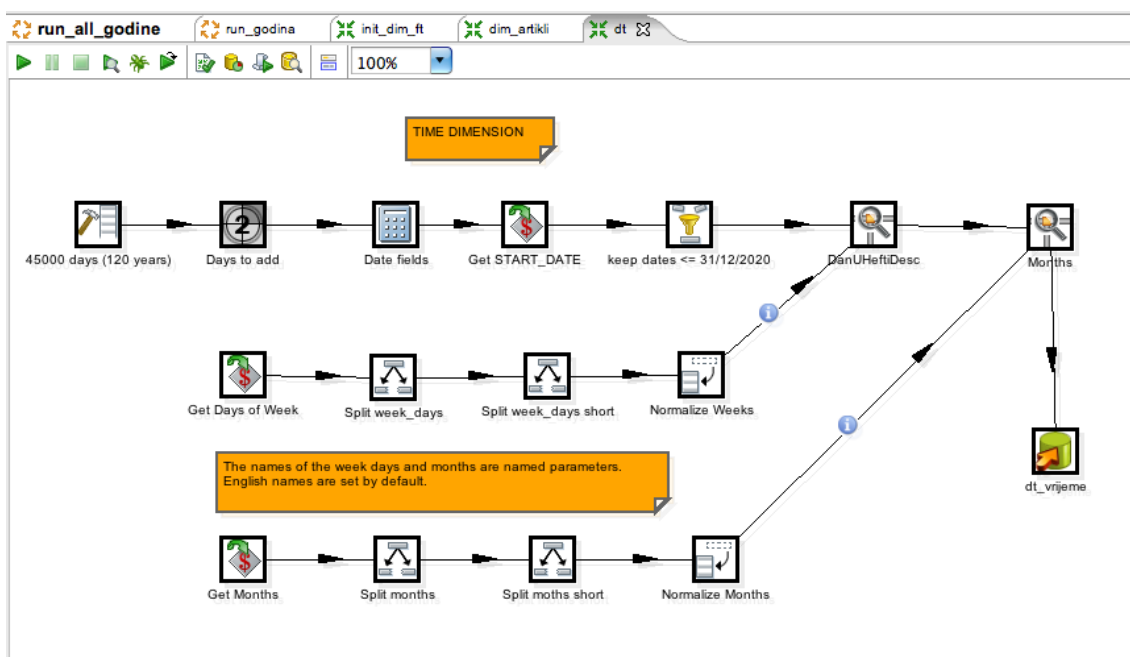
Slika 2.7: Generacija "ft\_fakt" 'facts' tabele za određenu poslovnu godinu



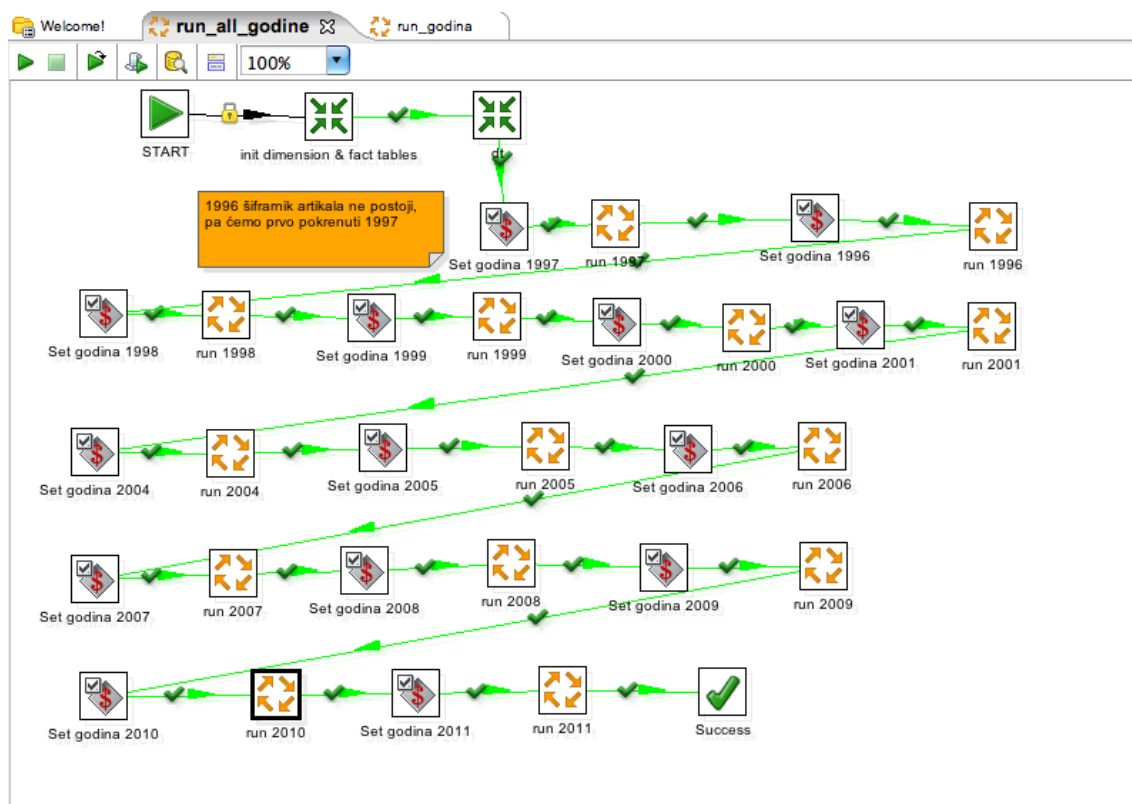
**Slika 2.8:** Kettle transformacija: Generacija "dim\_par" i "dim\_g" 'dimension' tabela za određenu poslovnu godinu



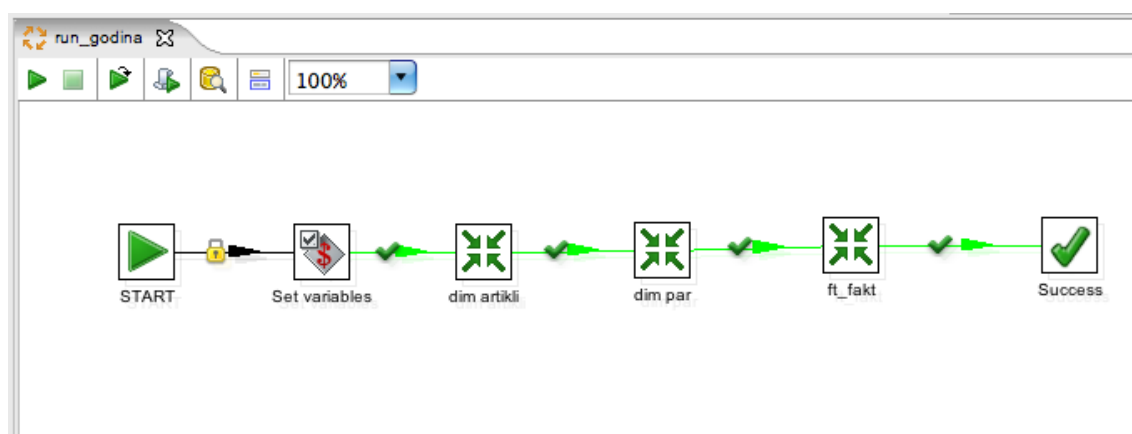
**Slika 2.9:** Kettle transformacija: Generacija "dim\_art" i "dim\_art\_gru" 'dimension' tabela za određenu poslovnu godinu



**Slika 2.10:** Kettle transformacija: Generacija "dim\_dt" 'dimension' tabele - vremenska dimenzija



**Slika 2.11:** Kettle job: inicijalizacija OLAP tabela, te generacija OLAP pdoataka za sve poslovne godine 1996-2011



**Slika 2.12:** Kettle job: Generacija OLAP podataka iz F18 ERP izvora za jednu poslovnu godinu

	A	B	C	D	E	F	G	H	I
8	2NAP001	2NAP	NAPOJNA JEDINICA L MODEL *		40				
9	3	3							
10	3 .....	3 ..	ŠTAMPAČI, KOPIR APARATI						
11	4 .....	4 ..	OSTALI PERIFERALI/DODATNA OPREMA						
12	5 .....	5 ..	MREŽNA OPREMA						
13	6 .....	6 ..	POTROŠNI MATERIJAL						
14	7 .....	7 ..	SOFTVER DRUGIH PROIZVOĐAČA						
15	8 .....	8 ..	OSTALO						
16	9 .....	9 ..	SOFTVER SIGMA-COMA						
17	9OST.....	9OST	OSTALI PROGRAMI						
18	9OST0001	9OST	SCC EXPLORER - MANAGER PROJEKATA		1465				
19	S.....	S...	NOMENKLATURA ZA SERVIS						
20	SEMON	SEMO	SERVIS MONITOR						
21	SEOST	SEOS	OSTALI DIJELOVI						
22	SERAC	SERA	RAČUNARI SERVIS						
23	SESTA	SEST	ŠTAMPAČI SERVIS						
24									

**Slika 2.13:** Error reporting putem 'spreadsheet' dokumenata - artikli za koje nisu definisani kodovi u olap\_cleansing tabelama

	A	B	C	D	E	F	G	H	I
1	ERROR	idtpdok	brdok	dat	kolicina	cijena	rabat	vrijednost	bk_idroba
2	ERR_ART	10	00004	2008-01-04	1,00	166,00		166	OS-DI
3	ERR_ART	10	00005	2008-01-04	1,00	50,00		50	PA-1
4	ERR_ART	10	00022	2008-02-08	2,00	50,00		100	PA-1
5	ERR_ART	10	00023	2008-02-08	3,00	40,00	20,00	120	PA-1
6	ERR_ART	10	00031	2008-02-08	2,00	167,40	7,00	334,8	KALK-DI
7	ERR_ART	10	00066	2008-02-25	1,00	318,50	30,00	318,5	VIRM
8	ERR_ART	10	00067	2008-02-27	12,00	85,00		1020	UPS600
9	ERR_ART	10	00135	2008-03-10	1,00	513,00	10,00	513	EPDV
10	ERR_ART	10	00144	2008-03-25	1,00	110,70	10,00	110,7	KADEV-MI
11	ERR_ART	10	00152	2008-03-25	1,00	454,00		454	DGSIGASW
12	ERR_ART	10	00152	2008-03-25	20,00	20,00		400	RJ45UTDUP
13	ERR_ART	10	00159	2008-04-08	1,00	50,00		50	PA-1
14	ERR_ART	10	00160	2008-04-11	50,00	1,00		50	KABUTP05

**Slika 2.14:** Dokumenti prodaje u kojima su neispravni podaci potrebni za popunjavanje dimension tabela (datum, klijent, roba)

## 3. Iza case study-ja ?

### 3.1. Ne znam

#### 3.1.1. Mondrian

Snowflake mondrian - join

Pentaho (2012)

#### 3.1.2. datamart vs datawarehouse

'Data mart' sadrži informacije o jednom dijelu organizacije (npr. prodaja, ljudski resursi), dok 'datawarehouse' sadrži informacije iz više područja - obrađuje organizaciju globalno.

'Data warehouse' je stoga usmjeren na podršku 'top' menadžmenta, dok 'datamart' obezbjeđuje informacije za upravljanje i operativno planiranje pojedinih dijelova organizacije (Roldan, 2010, str. 391).

#### 3.1.3. Konstrukcija OLAP kocke

surogat key (id)

business key (bk)

dimension table

facts table

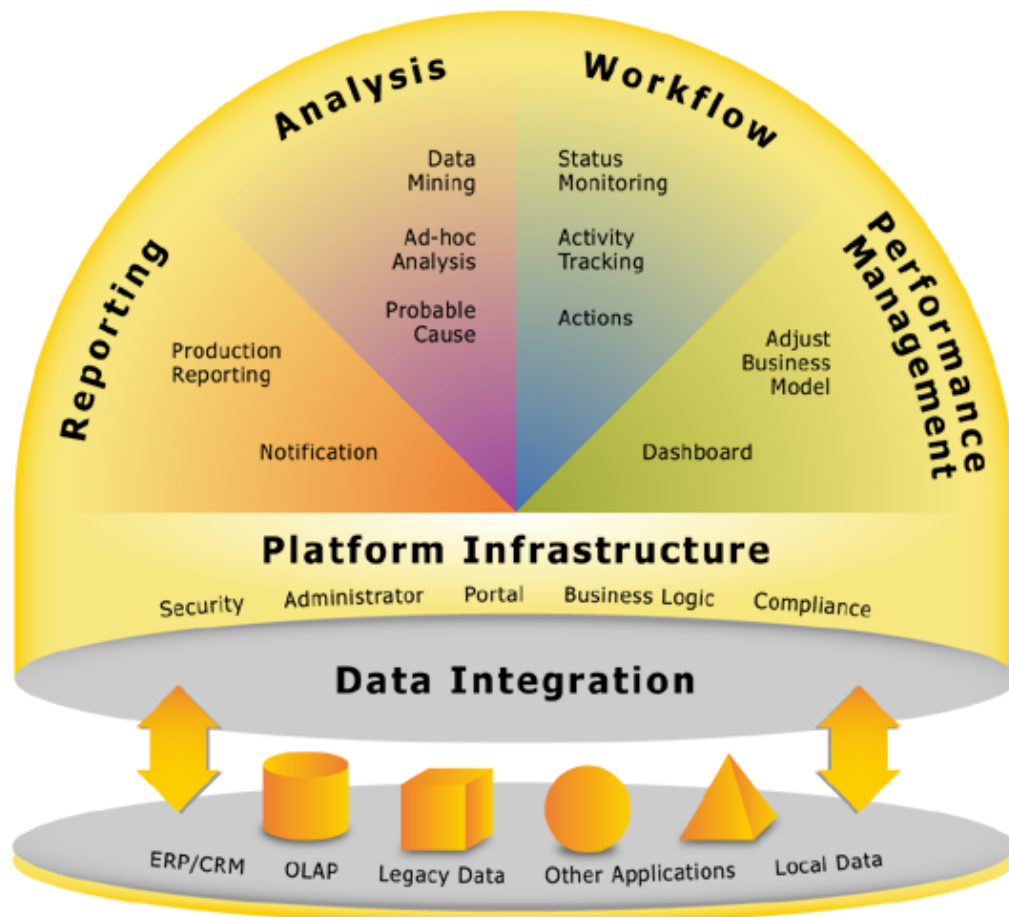
SCD slow changing dimension

– Type I

– Type II

## 3.2. Pentaho

Pentaho: analysis multidimensional, reporting, dashboards (key performance indicators) (Roldan, 2010, str. 7).



Slika 3.1: Pentaho arhitektura (Bimonte i Wehrle (2007))

## 3.3. Spoon

## 3.4. Data mining

Data mining Weka projekat: University of Waikato (2012), Pentaho Community (2012)  
R statistički paket foundation (2012)

### 3.4.1. dimension table

Mondrian schema:

### 3.4.2. facts table

### 3.4.3. ETL (Extract Transform Load)

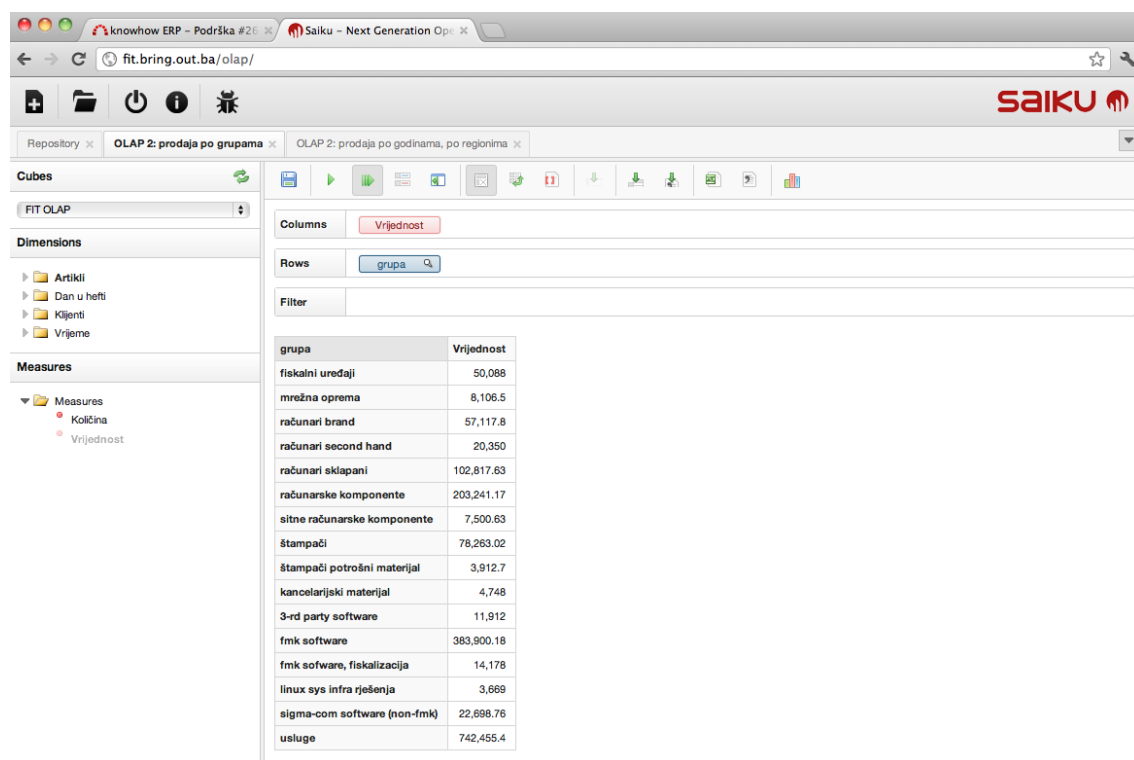
## 3.5. Poslovna pitanja (Business questions)

Kolika je prodaja u određenom vremenskom periodu ?

Kakav je odnos prodaje prodaje za određeni period tekuće godine u odnosu na predhodne ?

Koji su efekti zapošljavanja radnika po pitanju ostvarenih prihoda ?

## 3.6. Analiza podataka



The screenshot shows the Saiku web interface with the following components:

- Repository:** OLAP 2: prodaja po grupama
- Cubes:** FIT OLAP
- Dimensions:** Artikli, Dan u tjedni, Klijenti, Vrijeme
- Measures:** Measures, Količina, Vrijednost
- Columns:** Vrijednost
- Rows:** grupa
- Filter:**

grupa	Vrijednost
fiskalni uređaji	50,088
mrežna oprema	8,106.5
računari brand	57,117.8
računari second hand	20,350
računari sklopiani	102,817.63
računarske komponente	203,241.17
sitne računarske komponente	7,500.63
šampači	78,263.02
šampači potrošni materijal	3,912.7
kancelarijski materijal	4,748
3-rd party software	11,912
fmk software	383,900.18
fmk software, fiskalizacija	14,178
linux sys infra rješenja	3,669
sigma-com software (non-fmk)	22,698.76
usluge	742,455.4

Slika 3.4: Pregled prodaje po grupama artikala

1 SELECT

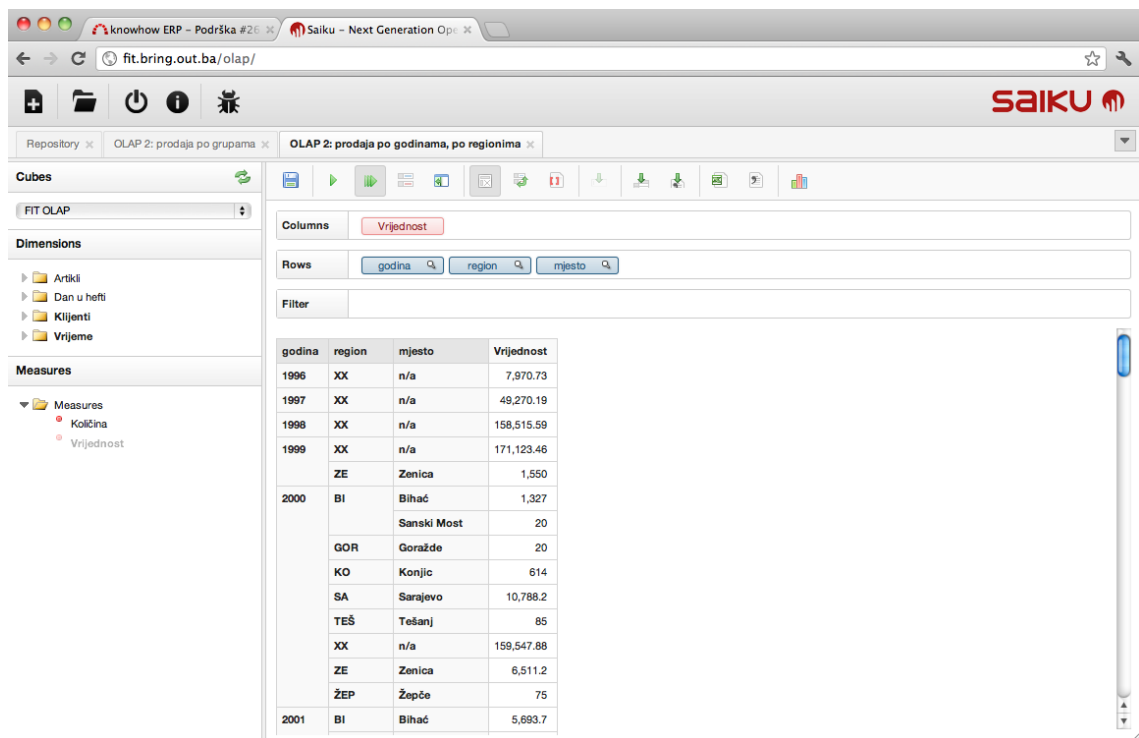


```

2 NON EMPTY { Hierarchize ( { [ Measures ]. [ Vrijednost ] } ) }
3 ON COLUMNS,
4   NON EMPTY { Hierarchize ( { [ Artikli . artikli ]. [ grupa ]. Members } ) }
5 ON ROWS
6 FROM [ FIT OLAP ]
7 WHERE { Hierarchize ( { [ Vrijeme . vrijeme ]. [ All Vrijeme . vrijeme ] } ) }

```

**Listing 3.1:** Pregled prodaje po grupama artikala



The screenshot shows the Saiku web interface with a query result table. The table has four columns: godina, region, mjesto, and Vrijednost. The data is grouped by year (godina) and region (region), with specific locations (mjesto) listed for each region. The values represent the sales amount (Vrijednost).

godina	region	mjesto	Vrijednost
1996	XX	n/a	7,970.73
1997	XX	n/a	49,270.19
1998	XX	n/a	158,515.59
1999	XX	n/a	171,123.46
2000	ZE	Zenica	1,550
		Bihac	1,327
	GOR	Sanski Most	20
		Gorazde	20
	KO	Konjic	614
	SA	Sarajevo	10,788.2
	TEŠ	Tešanj	85
	XX	n/a	158,547.88
2001	ZE	Zenica	6,511.2
		Žepče	75
	BI	Bihac	5,693.7

**Slika 3.5:** Pregled prodaje po regionima, po godinama

```

1 SELECT
2   NON EMPTY { Hierarchize ( { [ Measures ]. [ Vrijednost ] } ) }
3 ON COLUMNS,
4   NON EMPTY
5     Hierarchize (
6       Union ( CrossJoin ( [ Vrijeme . vrijeme ]. [ godina ]. Members ,
7         [ Klijenti . klijenti ]. [ region ]. Members ) , CrossJoin ( [
8           Vrijeme . vrijeme ]. [ godina ]. Members ,
9         [ Klijenti . klijenti ]. [ mjesto ]. Members ) )

```

```
10 ON ROWS
11 FROM [FIT OLAP]
```

**Listing 3.2:** Pregled prodaje po regionima

### **3.6.1. Redovi, Kolone, Filteri**

objasniti OLAP analysis ...

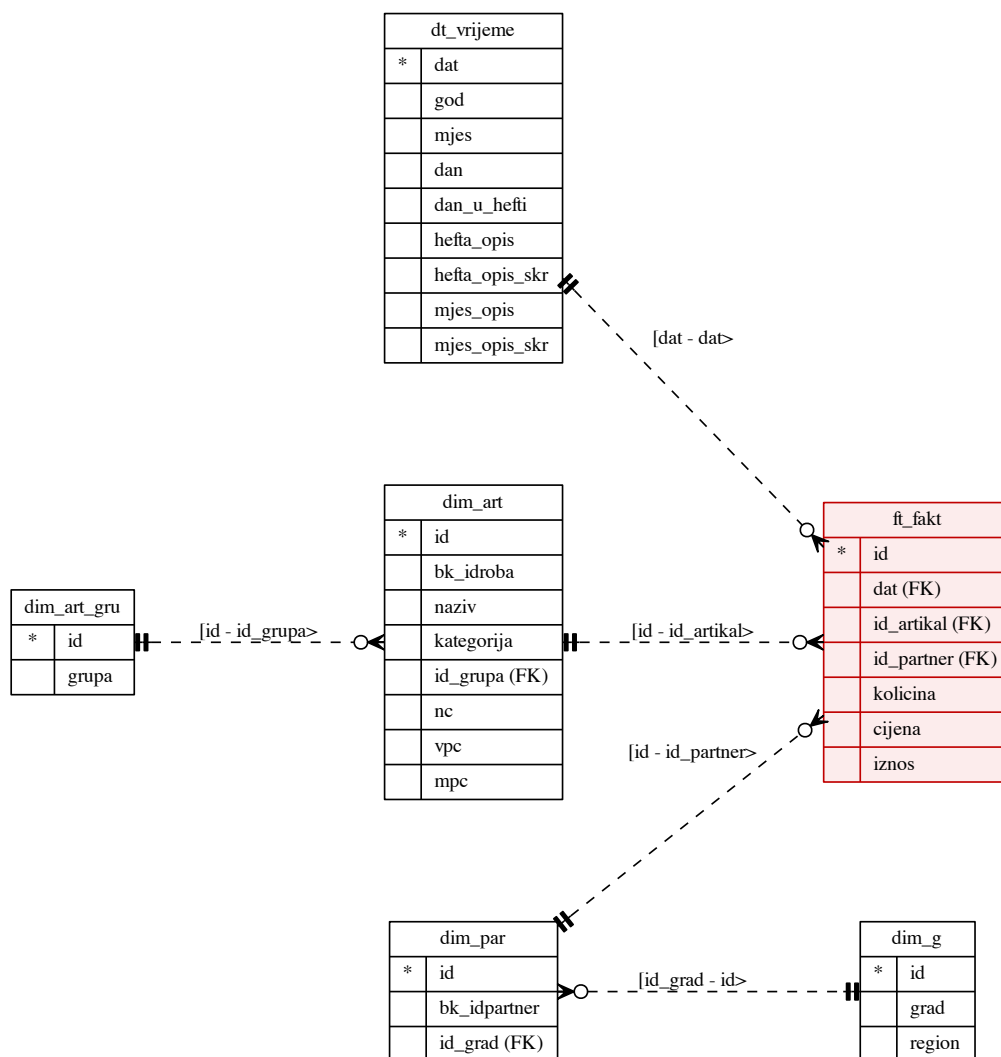
### **3.6.2. Ekspert**

Poznavanje sadržaja i postojećih struktura podataka.

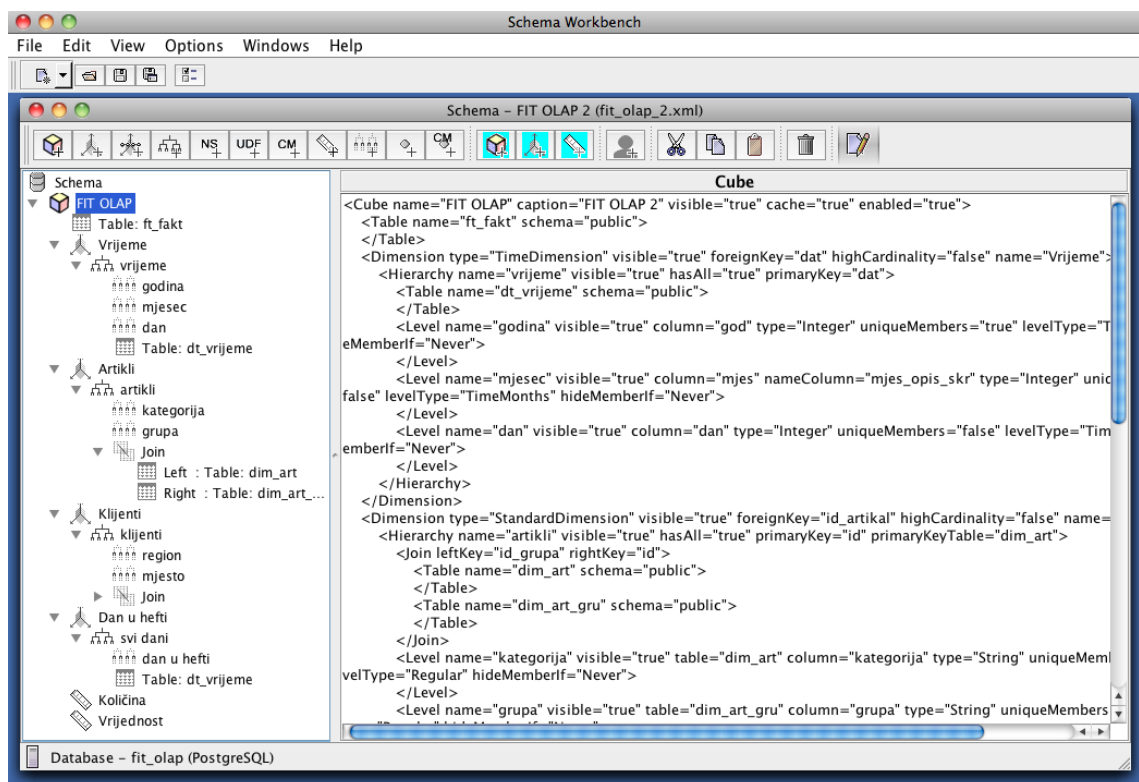
navodim pentaho: Roldan (2010)

navodim stranu 215: (Roldan, 2010, str. 215)

wikipedia olap cube: Wikipedia (2012a) wikipedia xmla: Wikipedia (2012b)



Slika 3.2: OLAP schema



Slika 3.3: Mondrian schema OLAP 2 cube

## **4. Zaključak**

Zaključak.

## 5. Literatura

Sandro Bimonte i Pascal Wehrle. An olap solution using mondrian and jpivot, 2007. URL [http://eric.univ-lyon2.fr/~sbimonte/doc/presentation\\_2007-02.pps](http://eric.univ-lyon2.fr/~sbimonte/doc/presentation_2007-02.pps).

R foundation. The r project for statistical computing, Februar 2012. URL <http://www.r-project.org>.

Pentaho. Mondrian snowflake schema, Februar 2012. URL [http://mondrian.pentaho.com/documentation/schema.php#Star\\_schemas](http://mondrian.pentaho.com/documentation/schema.php#Star_schemas).

Pentaho Community. Pentaho weka project, Februar 2012. URL <http://weka.pentaho.com/>.

Maria Carina Roldan. *Pentaho 3.2 Data Integration: Beginner's Guide*. Packt Publishing, 2010. URL <http://www.packtpub.com/pentaho-32-data-integration-beginners-guide/book>.

Machine Learning Group University of Waikato. The weka data mining software: An update, Februar 2012. URL <http://www.cs.waikato.ac.nz/ml/weka>.

Wikipedia. Olap cube, Februar 2012a. URL [http://en.wikipedia.org/wiki/OLAP\\_cube](http://en.wikipedia.org/wiki/OLAP_cube).

Wikipedia. Xml for analysis, Februar 2012b. URL [http://en.wikipedia.org/wiki/XML\\_for\\_Analysis](http://en.wikipedia.org/wiki/XML_for_Analysis).

## 6. Rezime

Rezime.

## **Dodatak A**

### **Korišteni alati**



## **Dodatak B**

### **Izvorni kod, dostupni resursi**

1. OLAP mondrian, kettle transformacije i job-ovi, erviz modeli: [https://github.com/hernad/hello\\_bi](https://github.com/hernad/hello_bi)
2. Latex kod ovog dokumenta <https://github.com/hernad/MIS/tree/master/latex>
3. olap\_cleansing 'spreadsheet' dokument [https://github.com/hernad/hello\\_bi/raw/master/olap\\_cleansing.xls](https://github.com/hernad/hello_bi/raw/master/olap_cleansing.xls)
4. Saiku demo server online: <http://fit.bring.out.ba/olap/#>

# **Dodatak C**

## **Bilješke autora**

1. Prva verzija ovog seminarskog rada, neuspješno [https://github.com/hernad/MIS/raw/master/knowhowERP\\_OLAP\\_blog\\_style.pdf](https://github.com/hernad/MIS/raw/master/knowhowERP_OLAP_blog_style.pdf)
2. FIT OLAP 2 cube: <http://redmine.bring.out.ba/issues/26711>