

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: 1.9.6

Mostar, februar 2012.

SADRŽAJ

1. Uvod	1
1.1. BI Pojmovi	1
1.1.1. ETL	1
1.1.2. Mondrian	3
1.1.3. datamart vs datawarehouse	3
1.1.4. Konstrukcija OLAP kocke	3
1.2. Pentaho	4
1.3. Spoon	5
1.4. Data mining	5
1.4.1. dimension table	5
1.4.2. facts table	6
1.4.3. ETL (Extract Transform Load)	6
1.5. Poslovna pitanja (Business questions)	6
1.6. Analiza podataka	6
1.6.1. Redovi, Kolone, Filteri	8
1.6.2. Ekspert	8
2. Zaključak	11
3. Literatura	12
4. Rezime	13
A. Korišteni alati	14
B. Izvorni kod, dostupni resursi	15
C. Bilješke autora	16

1. Uvod

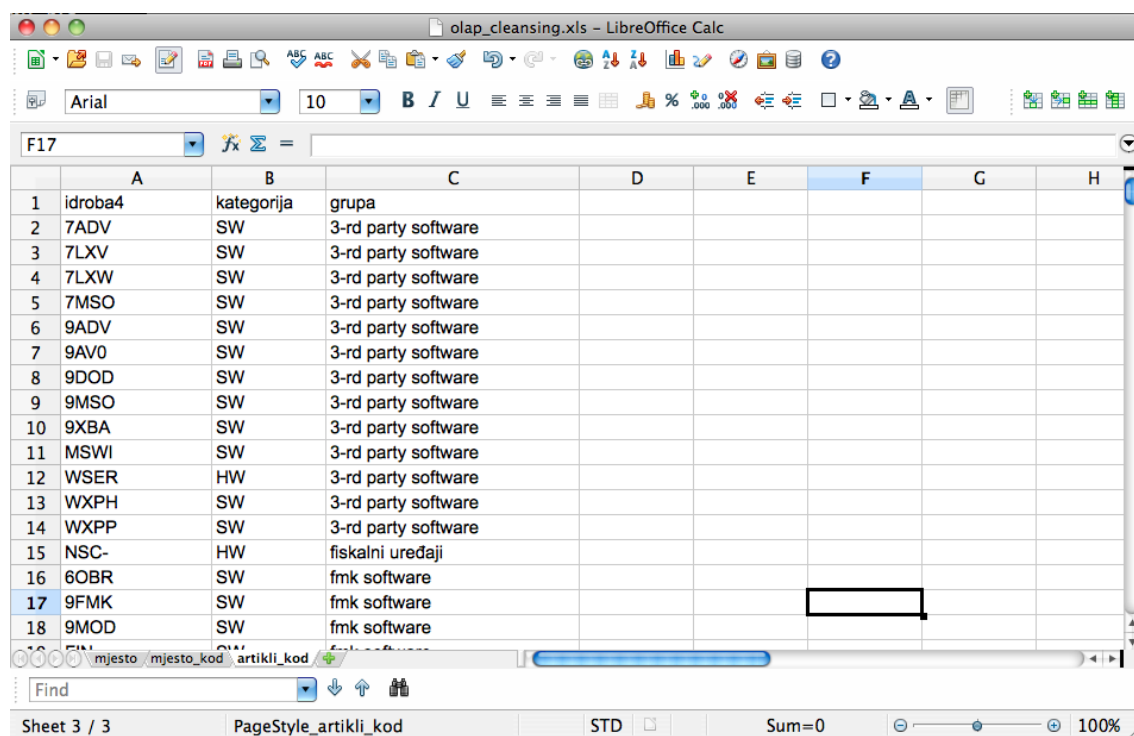
1.1. BI Pojmovi

1.1.1. ETL

Cleansing

F18 'cleansing' podaci (Dodatak B, olap_cleansing 'spreadsheet' dokument)

Klasificiranje izvornih podataka - šifarnik artikala



	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					

Slika 1.1: F18 klasificiranje šifarski sistem artikala

	A	B	C	D	E	F	G	H
1	MjestoIn	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	GORAZDE	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						

Slika 1.2: F18 "cleansing" klijenti - mjesta/gradovi

	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						

Slika 1.3: F18 kodiranje regiona - klasifikacija mjesta/gradova

1.1.2. Mondrian

Snowflake mondrian - join

Pentaho (2012)

1.1.3. 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).

1.1.4. Konstrukcija OLAP kocke

surogat key (id)

business key (bk)

dimension table

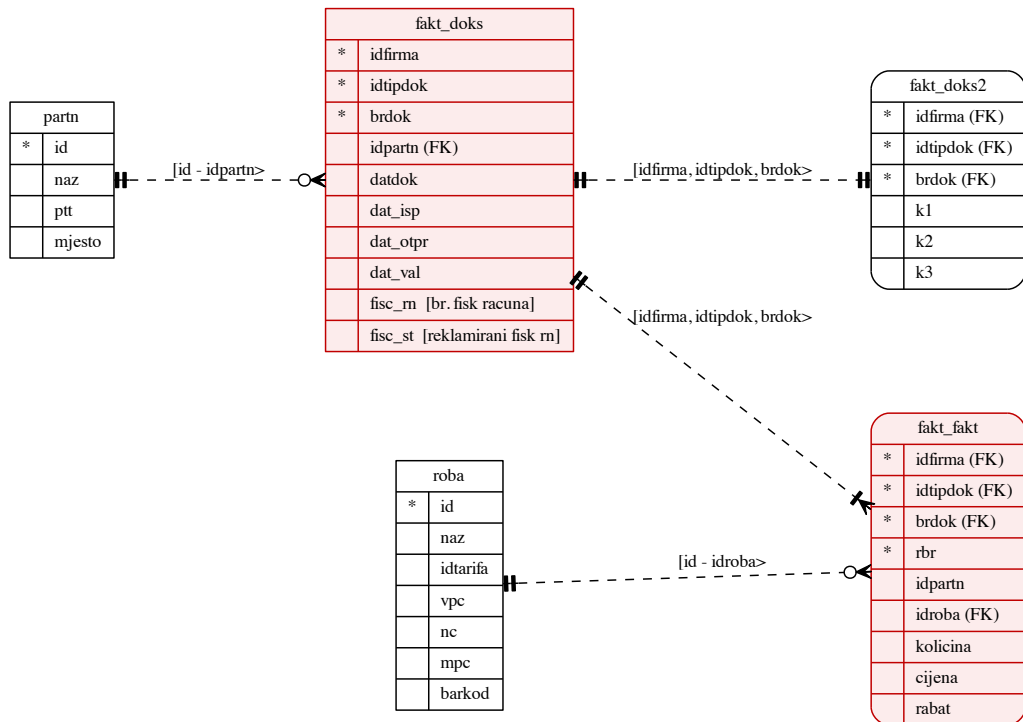
facts table

SCD slow changing dimension

– Type I

– Type II

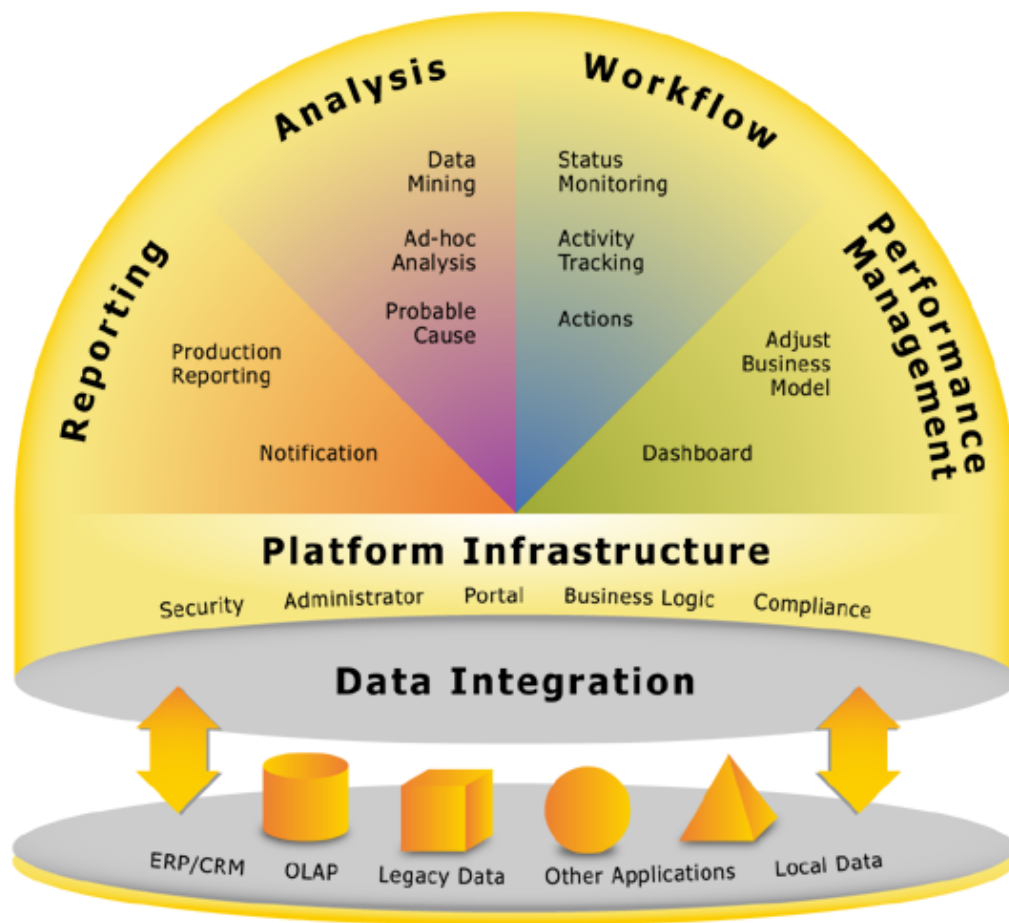
Operativni podacii smješteni su u sljedeći relacijski model:



Slika 1.4: F18 transakcijski db model (relevantni dio)

1.2. Pentaho

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



Slika 1.5: Pentaho arhitektura (Bimonte i Wehrle (2007))

1.3. Spoon

1.4. Data mining

Data mining Weka projekat: University of Waikato (2012), Pentaho Community (2012)

R statistički paket foundation (2012)

1.4.1. dimension table

Mondrian schema:

1.4.2. facts table

1.4.3. ETL (Extract Transform Load)

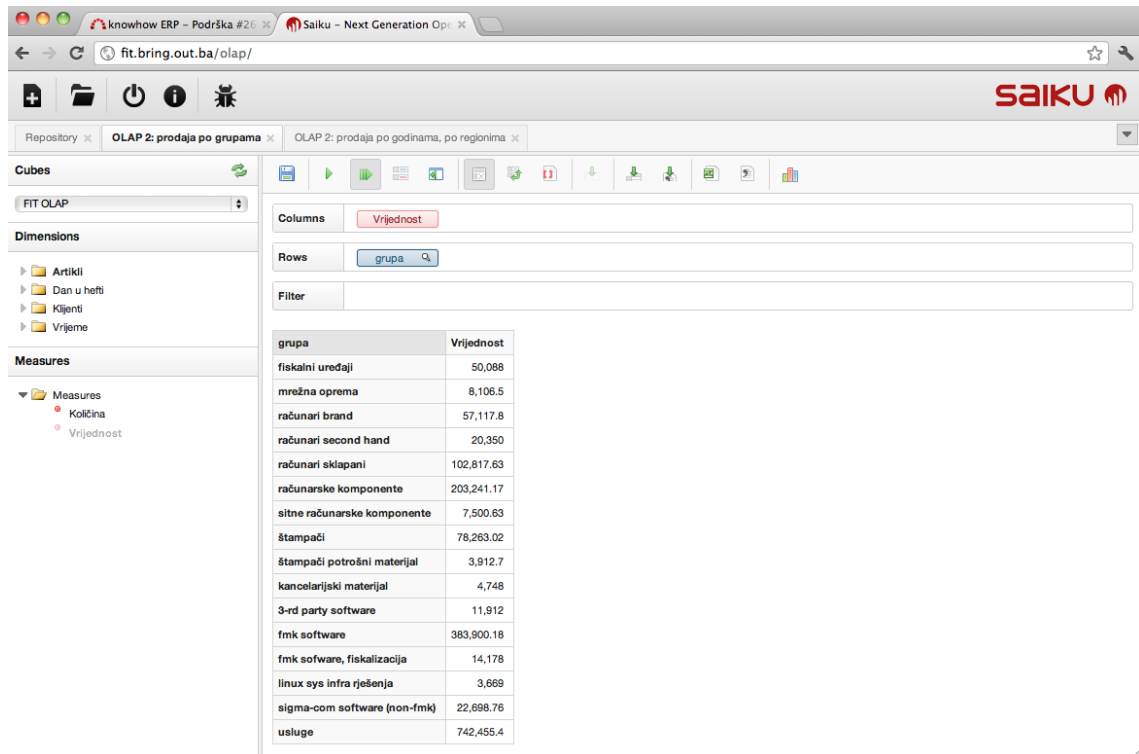
1.5. Poslovna pitanja (Business questions)

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 ?

1.6. 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 hefti, 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 sklapani	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 1.8: 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 1.1: Pregled prodaje po grupama artikala

The screenshot shows the Saiku web application interface. The browser address bar displays 'fit.bring.out.ba/olap/'. The application header includes the 'SAIKU' logo. The main content area shows a table with the following data:

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
	BI	Bihać	1,327
		Sanski Most	20
	GOR	Goražde	20
	KO	Konjic	614
	SA	Sarajevo	10,788.2
	TEŠ	Tešanj	85
	XX	n/a	159,547.88
	ZE	Zenica	6,511.2
	ŽEP	Žepče	75
2001	BI	Bihać	5,693.7

Slika 1.9: 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    )
11 FROM [FIT OLAP]

```

Listing 1.2: Pregled prodaje po regionima

1.6.1. Redovi, Kolone, Filteri

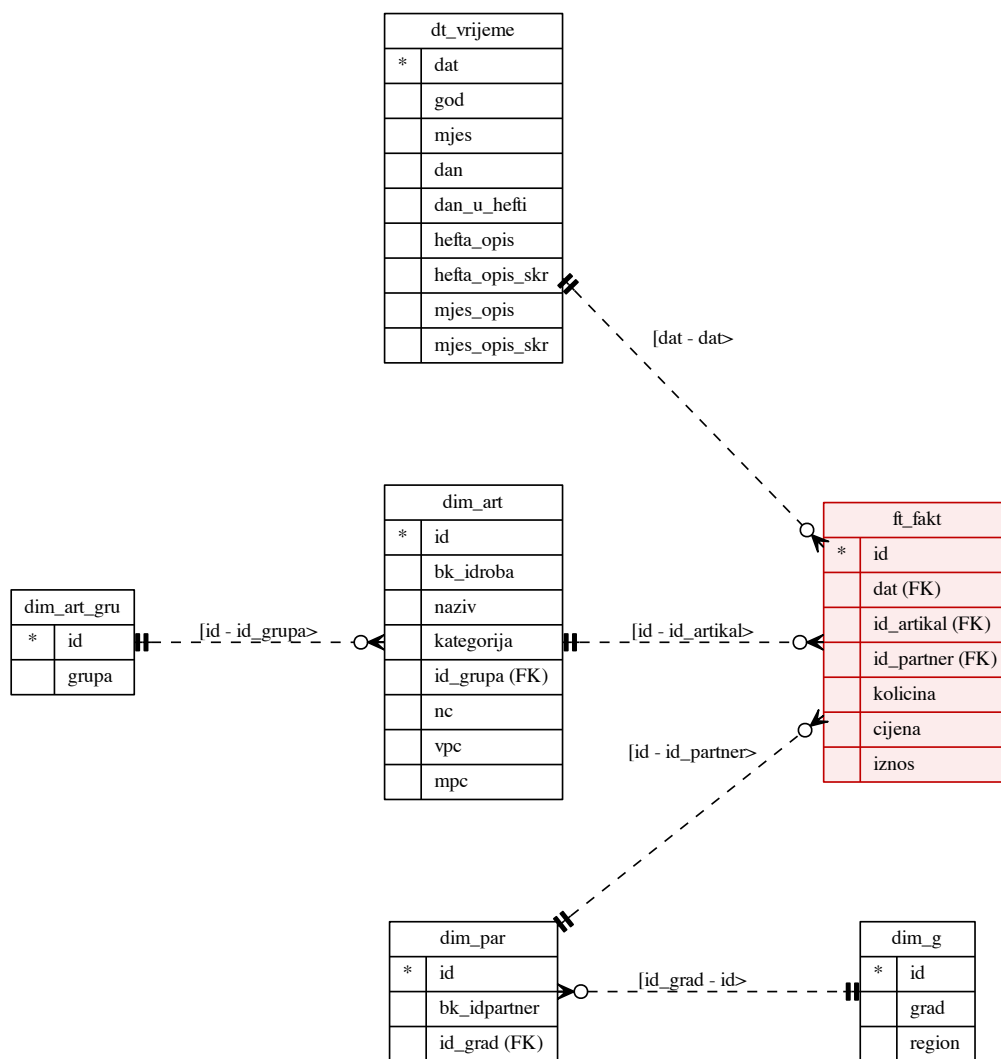
1.6.2. Ekspert

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 1.6: OLAP schema

2. Zaključak

Zaključak.

3. 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.

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.

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.

Wikipedia. Xml for analysis, Februar 2012b. URL http://en.wikipedia.org/wiki/XML_for_Analysis.

4. 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
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
4. Saiku demo server online: <http://fit.bring.out.ba/olap/#>

Dodatak C

Bilješke autora

1. FIT OLAP 2 cube: <http://redmine.bring.out.ba/issues/26711>