

UNIVERZITET DŽEMAL BIJEDIĆ U MOSTARU
FAKULTET INFORMACIJSKIH TEHNOLOGIJA MOSTAR

SEMINAR

Poslovna inteligencija bazirana na "open source" softveru

Student: *Ernad Husremović, DL 2792*

Mentor: *prof.dr Vanja Bevanda*

ver: 2.0.1

Mostar, februar 2012.

SADRŽAJ

1. Uvod	1
1.1. Poslovna inteligencija (BI)	1
1.1.1. Operativni podaci, skladišta podataka	1
1.2. Poslovni motivi za implementaciju BI rješenja	2
1.2.1. OLAP kocka	3
1.2.2. Alati za integraciju podataka (DI)	3
1.2.3. Rudarenje podataka (DM)	4
2. "Open source" (OSS) BI software	5
2.1. "Open core" model	5
2.2. Pentaho BI	5
2.3. Komponente Pentaho BI-a	6
2.3.1. Mondrian OLAP	6
2.3.2. Mondrian OLAP shema	7
2.3.3. Pentaho ETL dizajner 'Spoon'	8
2.3.4. Analiziranje OLAP kocki	8
3. Case study: Kreiranje i analiza 'Mondrian' OLAP kocke	10
3.1. F18 knowhow ERP	10
3.2. Poslovni cilj	10
3.3. Model skladišta podataka (DMart)	11
3.4. Transformacija operativnih podataka: klasifikacija i 'cleansing'	12
3.4.1. Šifarnik artikala	13
3.4.2. Šifarnik klijenata	14
3.5. Kreiranje vremenske dimenzije	16
3.6. Kreiranje 'facts' tabele prodaje	17
3.7. Kettle 'jobs'	18
3.8. OLAP reporting	20

3.8.1. Pregled prodaje po grupama artikala	21
3.8.2. Pregled prodaje po regionima	22
4. Zaključak	24
4.1. Rezultati ‘case study’-ja	24
4.2. Pogled sa aspekta menadžera	24
4.3. Pogled sa aspekta informatičara	25
4.4. Rezime	25
5. Literatura	26
A. Izvorni kod, dostupni resursi	28
B. Bilješke	29

1. Uvod

1.1. Poslovna inteligencija (BI)

Poslovna inteligencija (nadalje BI¹) se primarno odnosi na računarski bazirane tehnike identificiranja, ekstrakcije i analize poslovnih podataka. BI tehnologije obezbeđuju pregled ranijih i tekućih poslovnih operacija, kao i predviđanje budućih trendova u poslovanju² Wikipedia (2012a).

Standardne funkcije BI-a su:

1. izvještavanje (reporting)
2. analitičko procesiranje (analytical processing)
3. rudaranje podataka (data mining)
4. prediktivne analize (predictive analytics)

Sve ove funkcije pomažu poslovnom odlučivanju i planiranju - kako strateškom³ tako i operativnom planiranju⁴.

1.1.1. Operativni podaci, skladišta podataka

BI kao izvor podataka koristi skladišta podataka.

Skladišta podataka treba razlikovati od operativnih podataka. Operativni podaci - podaci o tekućem poslovanju nalaze se unutar poslovnih aplikacija (ERP)⁵.

ERP sistemi pohranjuju poslovne transakcije (poslovne dokumente) u realnom vremenu.

¹BI Business intelligence

²prediktivna analiza

³Odlučivanje 'top' managera

⁴'department' manageri

⁵ERP - Enterprise Resource Planning software, software za podršku tekućem poslovanju

ERP sistemi sadrže sisteme izvještavanja⁶ koji su prevashodno usmjereni na davanje (detaljnih) podataka o tekućem poslovanju⁷.

Operativni podaci su glavni izvor za gradnju skladišta podataka. Međutim, skladišta podataka se najčešće grade iz heterogenih izvora.

Iako slični po načinu konstrukcije, u BI terminologije se pravi distinkcija između 'data mart' i 'data warehouse' skladišta podataka⁸:

Data mart (DMart) sadrži informacije o jednom dijelu organizacije (npr. prodaja, ljudski resursi),

Data warehouse (DW) sadrži informacije iz više područja - analizira organizaciju globalno.

DW je stoga usmjeren na podršku 'top' menadžmenta, dok DMart obezbjeđuje informacije za upravljanje i operativno planiranje pojedinih dijelova organizacije (Roland, 2010, str. 391).

1.2. Poslovni motivi za implementaciju BI rješenja

Analiza poslovnih podataka u cilju podrške kvalitetnom poslovnom odlučivanju postojala je informatizacije poslovanja.

Slijedeća poslovna pitanja su univerzalna (nisu aktuelna tek u informatičkoj eri):

- Kako se kretala prodaja određene grupe artikala u predhodnom periodu ?
- Koja grupa artikala se najviše zadržava na lageru ?
- Kakva je struktura prodaje po regijama ?
- Unutar koje grupacije artikala / proizvoda je najbolja marža ?

Sva ova pitanja praktično vrše analizu efekata poslovanja usljed različitih uticaja (multidimenzionalna analiza).

Iz svih gornjih pitanja mogu se uočiti dva tipa informacija:

- mjere - poslovni indikatori (prodaja, vrijednost zalihe, visina marže)
- dimenzije - atributi poslovanja (geografske dimenzije - grad, region, vemenske dimenzije, grupe artikala, rang cijena ...)

⁶ Tradicionalni reporting sistem. Podjelu na tradicionalni i BI reporting treba ipak uslovno postaviti. Mnogi savremeni ERP sistemi imaju u svom sastavu segmente BI reporting-a

⁷ period tekuće poslovne godine do nivoa pojedinačne poslovne transakcije

⁸ U domaćoj literaturi se najčešće za oba pojma koristi termin 'skladište podataka'

Ovaj koncept je polazište za konstrukciju multidimenzionalnih skladišta podataka (OLAP cube) kao osnovnog gradivnog elementa BI rješenja.

1.2.1. OLAP kocka

OLAP kocka (OLAP cube⁹) je set podatka organizovanih na taj način da omogućavaju *nedeterminirane* upite nad agregiranim podacima, odnosno online analitičko procesiranje podataka Wikipedia (2012b).

Ovakva organizacija podataka omogućava OLAP klijentima pregled podataka u različitim varijantama.

Ono što RDBMS¹⁰ predstavlja za ERP sistem, OLAP kocka predstavlja za BI sistem.

Analogija postoji i u dijelu pretrage podataka:

- RDBMS <-> SQL structured query language
- OLAP cube <-> MDX - multidimensional query language

Današnje implementacije OLAP kocki su slijedeće:

- ROLAP - podaci smješteni u relacijske baze podataka
- MOLAP - podaci su u proprietary formatu prilagođenom procesiranju multidimenzionalnih struktura podataka

Pored gornjih postoje i hibridne implementacije OLAP kocki koje kombinuju gornje tehnologije.

1.2.2. Alati za integraciju podataka (DI)

DI¹¹ alati omogućavaju vezu BI sistema sa "vanjskim" svijetom (vidi 1.1.1 operativni podaci)

Glavni dio DI 'toolset' -a su je ETL softver¹².

ETL softver obavlja sljedeće funkcije:

1. *Extract*: uzimanje podataka iz vanjskih izvora
2. *Transform*: transformacija operativnih podataka u format koji je pogodan za pohranu u skladište podataka (vidi 1.1.1)
3. *Load*: snimanje 'prečišćenih' podataka u skladište podataka (DW/DMart)

⁹OLAP - online analytical processing

¹⁰RDBMS relational database management system

¹¹DI - data integration

¹²ETL - Extract/Transform/Load

1.2.3. Rudarenje podataka (DM)

Pojam rudarenje podataka¹³ se može definisati kao pronalaženje zakonitosti među podacima.

Jedna od definicija rudarenje glasi: rudarenje podataka je sistematičan, interaktivan i iterativan proces izvođenja i prikazivanja korisnika, implicitnog i inovativnog *znanja* iz podataka (Mršić, 2004, str. 40).

Unutar ovog rada nećemo se baviti ovim dijelom BI-a.

Dodatno proučavanje

Open source software koji pokriva ove oblasti:

- Data mining ‘Weka’ projekat: University of Waikato (2012), Pentaho Community (2012)
- ‘R’ statistički paket foundation (2012) - prediktivne, trend analize

Treba uočiti da je ‘Weka’ jedan od podprojekata ‘Pentaho BI suite’-a (vidi 2.3).

¹³DM - data mining

2. "Open source" (OSS) BI software

"Open source" (nadalje OSS¹) rješenja dostupne su za sve glavne funkcije BI-a.

Najveći OSS BI projekti vođeni su od strane komercijalnih opensource vendora. Najpoznatiji OSS BI vendori su:

- Jaspersoft BI ²
- Pentaho BI ³
- Talend⁴

Treba naglasiti da su mnoge komponente pojedinih BI rješenja bazirane na istim tehnologijama ili su čak identične.

Tako vendori koji su na komercijalnom polju konkurencija, zajednički rade na razvoju baznih komponenti.

Ova sinergija u razvoju, što je uostalom generalna odlika OSS-a, predstavlja snažan poticaj brzom razvoju OSS BI rješenja.

2.1. "Open core" model

Većina opensource BI vendora bazira se na 'open core' modelu. Ovaj model organizacije opensource softvera predviđa dvije verzije:

- "community" verzija (OSS),
- "enterprise" verzija ('closed source', 'proprietary addon'-ovi vendora)

2.2. Pentaho BI

Pentaho BI obuhvata praktično funkcije BI-a (Roldan, 2010, str. 7):

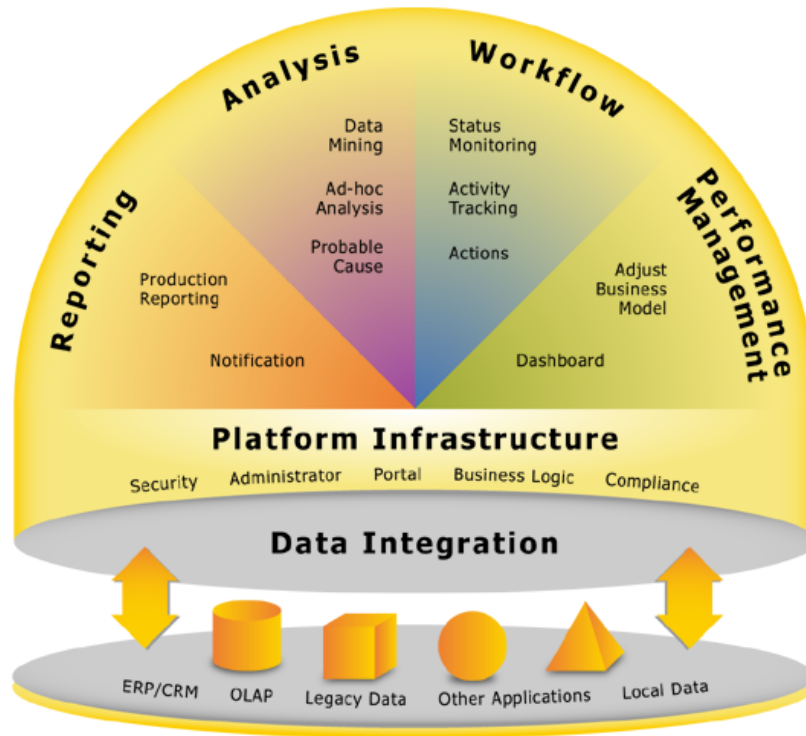
¹OSS - open source software

²<http://www.jaspersoft.com/>

³<http://www.pentaho.com/>

⁴<http://www.talend.com/>

1. multidimenzionalna analiza
2. izvještavanje
3. BI "dashboards"⁵



Slika 2.1: Pentaho arhitektura (Bimonte i Wehrle (2007))

2.3. Komponente Pentaho BI-a

2.3.1. Mondrian OLAP

Pentaho implementacija OLAP kocke naziva se 'Mondrian'. Mondrian je kocka RO-LAP tipa.

Treba pomenuti da je 'Mondrian' XMLA kompatibilan provajder⁶

⁵prikaz glavnih poslovnih indikatora u putem sažetog grafičkog prikaza

⁶XML za Analizu (XMLA) je industrijski standard za pristup podacima u analitičkim sistemima kao što su OLAP i 'data mining'. Baziran je na drugim industrijskim standardima kao što su XML, SOAP i HTML Wikipedia (2012c).

2.3.2. Mondrian OLAP shema

Kao ROLAP implementacija, podaci se nalaze u relacijskog bazi podataka⁷

Kod konstrukcije sheme koriste se sljedeći pojmovi:

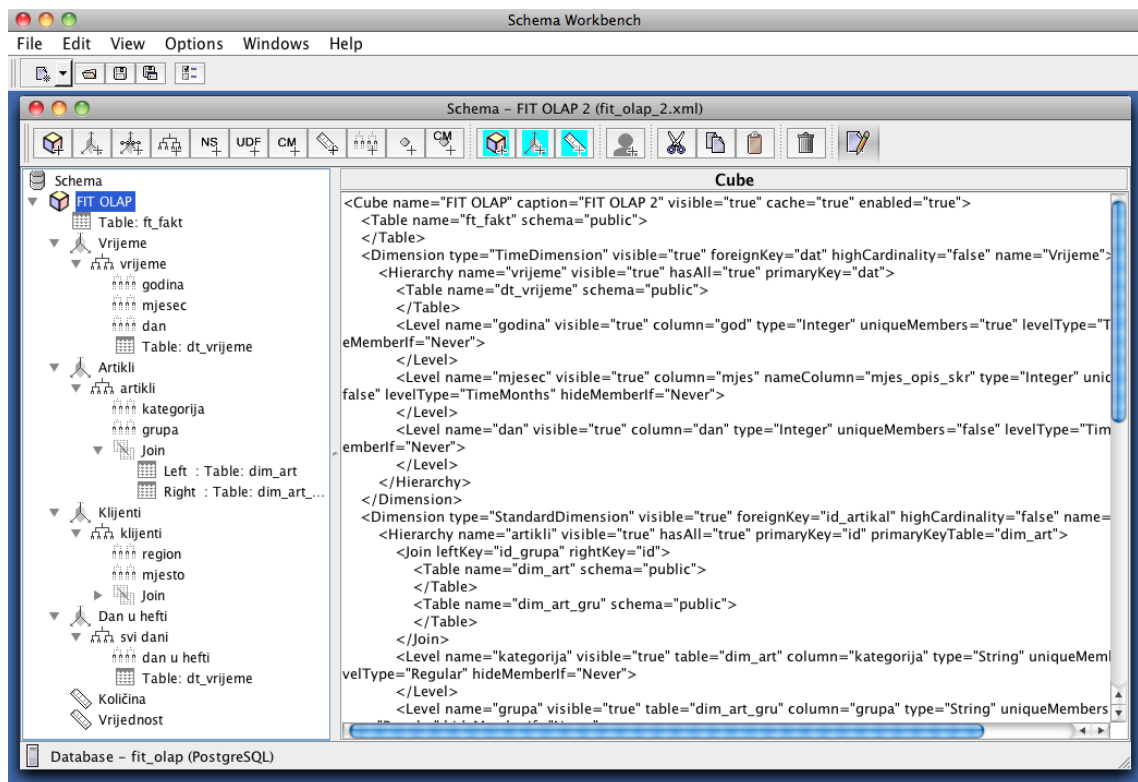
- ‘dimension table’ - tabela u kojoj su pohranjene dimenzije
- SCD - slow changing dimension (Roldan, 2010, poglavlje 12)
- SCD Type I - čuva se samo jedna vrijednost dimenzije
- SCD Type II - čuva se istorija vrijednosti dimenzije kroz vremenski period ⁸
- ‘facts table’ - tabela u kojoj su mjere - poslovni indikatori koje analiziramo
- ‘business key’ (bk) - ključ koji koriste aplikacije za rukovanje operativnim podacima (ERP software)
- ‘surogat key’ (id) - ključ u bazi podataka koji koristi OLAP storage
- ‘snowflake’ shema - šema u kojoj su dimenzije u sopstvenim tabelama, dimenzije su visokog stepena normalizacije podataka⁹ Pentaho (2012)
- Type II

Pogledajmo jednu gotovu mondrian shemu dobijenu u *case study*-ju (pogl. 3) koji slijedi:

⁷Podržane su sve JDBC podržane relacijske baze: PostgreSQL, MySQL, MSSQL, Oracle

⁸npr. cijena artikla se mijenja tokom vremena, SCD type II dimenzija za svaku novu cijenu pravi novi zapis u dimension tabeli

⁹normalizacija podataka u relacijskog bazi podataka



Slika 2.2: Mondrian schema OLAP 2 cube

2.3.3. Pentaho ETL dizajner ‘Spoon’

‘Spoon’ je GUI aplikacija u kojoj se vrši definicija i testiranje ETL transformacija i ‘job’-ova.

Kod ETL operacija bitno je poznavati sljedeće pojmove:

- ‘cleansing’ - “čišćenje” podataka - ispravka (ili izbacivanje) netačnih podataka

2.3.4. Analiziranje OLAP kocki

Pentaho sadrži dva ‘OLAP analyzer’¹⁰ rješenja:

- JPilot - starije rješenje, napušta se njegov razvoj¹¹
- novi ‘OLAP Analyzer’ koji se nalazi samo u ‘Enterprise’ (plaćenj verziji)¹²

Kao ‘OLAP Analysis’ sofver izabran je Saiku (vidi B).

Saiku je modularni ‘open source’ analitički sotver koji nudi jednostavnu OLAP analizu podataka analytical labs (2012)

¹⁰često se za ove aplikacije kaže ‘OLAP cube client’ ili ‘OLAP data consumer’ aplikacije

¹¹označeno od Pentaho razvojnog tima kao ‘deprecated’

¹²kako jeova komponenta je ‘closed source’ software, ona se u ovom radu neće razmatrati.

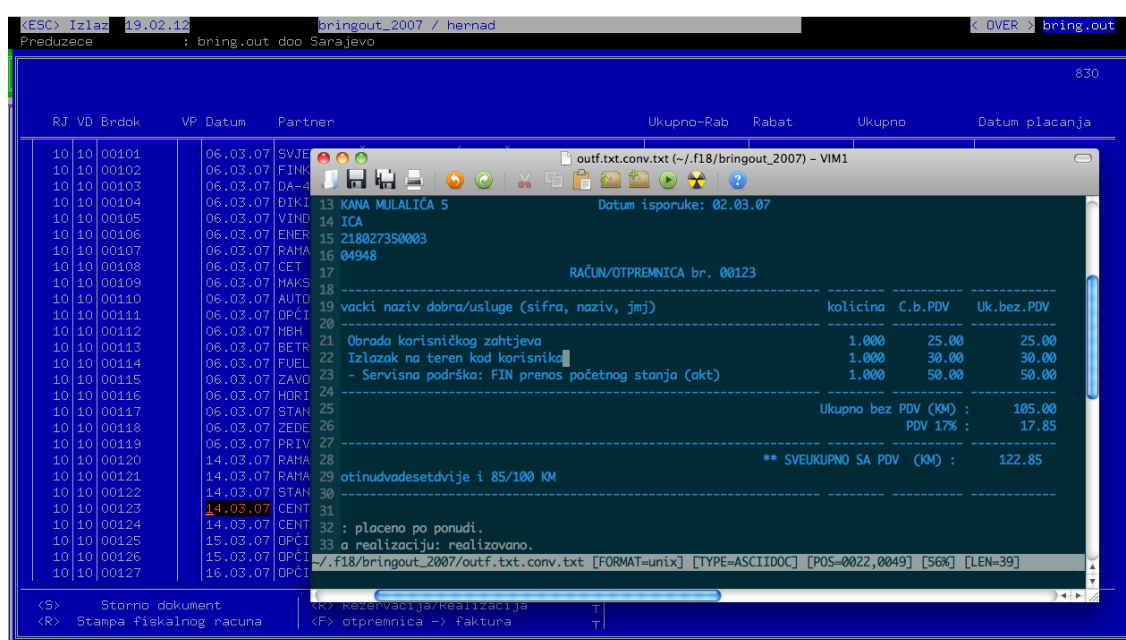
Kod korištenja OLAP analitičkog softvera koriste se sljedeći pojmovi:

- redovi - sadrži jednu ili više mjera i/ili dimenzija koje se prikazuju u redovima kod prikaza podataka
- kolona - sadrži jednu ili više mjera i/ili dimenzija koji se prikazuju u kolonama prikaza podatka
- filteri - ograničenje podataka po određenim vrijednostima dimenzija

3. Case study: Kreiranje i analiza 'Mondrian' OLAP kocke

U ovom 'case study'-ju ćemo izvršiti formiranje OLAP kocke za operativne podatke ERP aplikacije 'F18 knowhow' radi analize podataka o prodaji.

3.1. F18 knowhow ERP



Slika 3.1: ERP aplikacija, F18 klijent

3.2. Poslovni cilj

Cilj je analizirati prodaju firme firme "bring.out"¹ po godinama pri čemu nas interesuje struktura klijenata po gradovima i regionima, te prodaja artikala po određenim katego-

¹<http://www.bring.out.ba>

rijama i grupama.

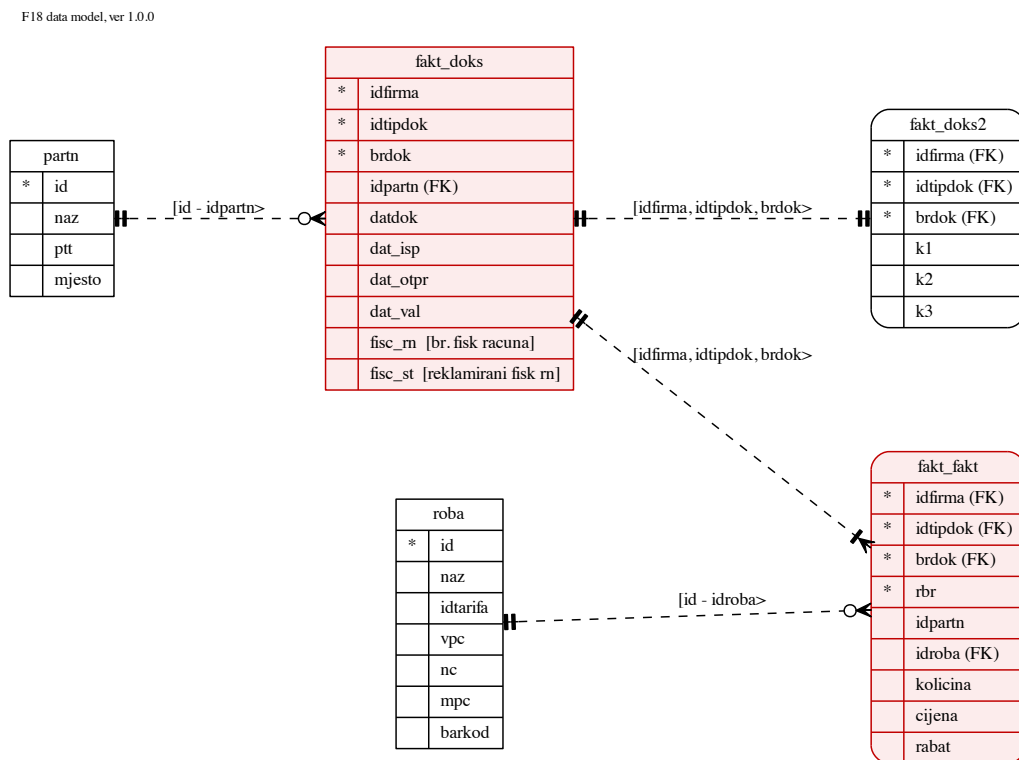
Znači, potrebno je kreirati DMart prodaje za kompletan period poslovanja (1996-2011).

Glavni operativni podaci nalaze se u ERP aplikaciji, u PostgreSQL relacijskoj bazi.

Dio potrebnih dimenzija (regioni, kategorije i grupe artikala) nisu implementirani unutar operativnih podataka, tako da analitičar mora unutar ETL procesa ove informacije dodati.

Podaci svake poslovne godine nalaze se u posebnoj bazi podataka.

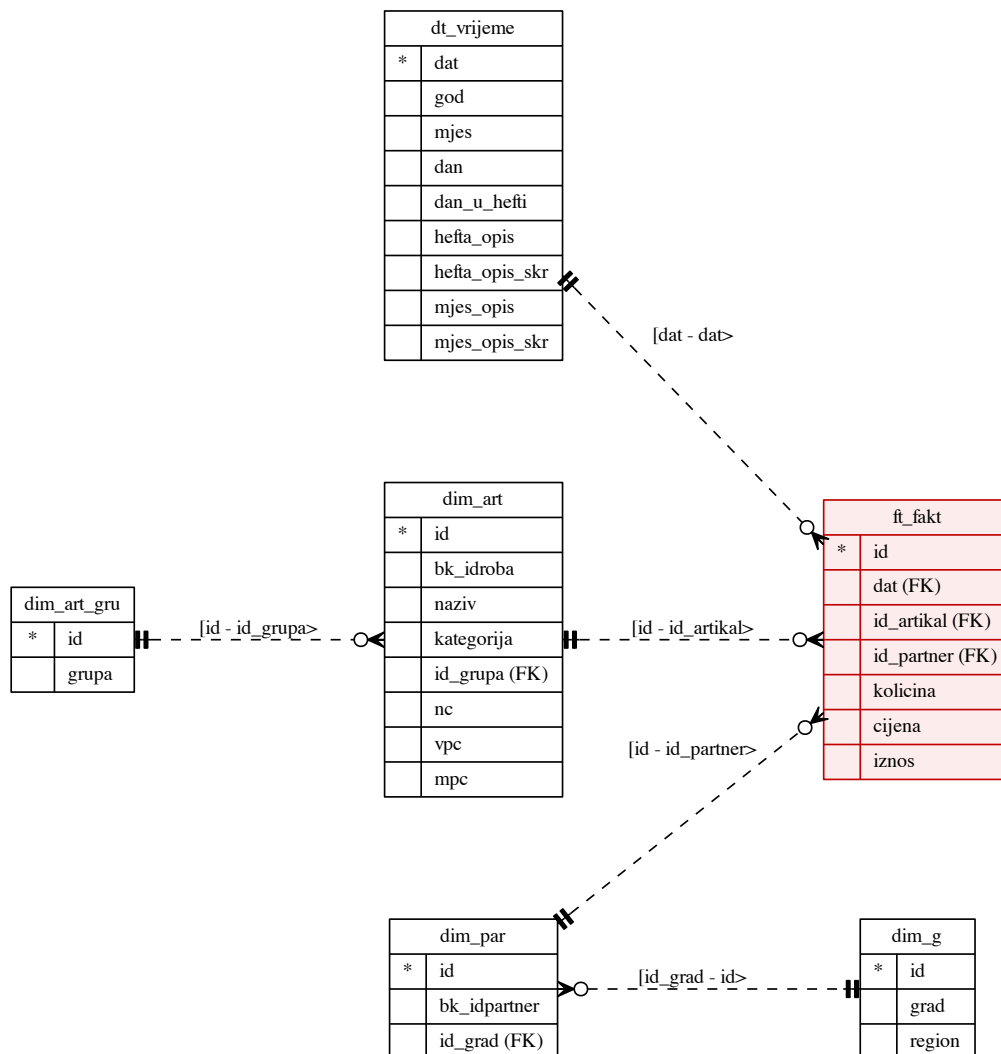
Operativni podaci 'F18 knowhow' smješteni su u sljedeći relacijski model:



Slika 3.2: F18 transakcijski db model (relevantni dio)

3.3. Model skladišta podataka (DMart)

Na osnovu gornjeg zahtjeva utvrđujemo model skladišta podataka koji ćemo koristiti za izgradnju OLAP kocke:



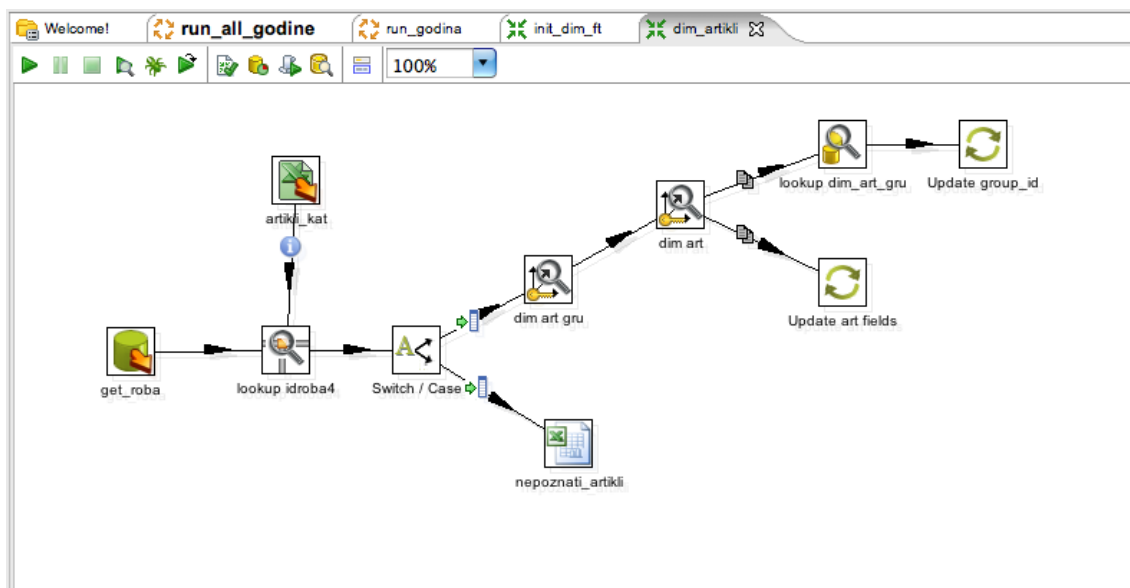
Slika 3.3: OLAP schema

3.4. Transformacija operativnih podataka: klasifikacija i ‘cleansing’

F18 ‘cleansing’ podataka vrši se na šifarskom sistemu (matičnim podacima) artikala i klijenata.

Za te potrebe se u kettle transformacijama prilikom formiranja odgovarajućih dimenzija vrši ‘lookup’ na osnovu posebno pripremljenih tabelarnih podataka (vidi A, olap_cleansing xls dokument).

3.4.1. Šifarnik artikala



Slika 3.4: Kettle transformacija: Generacija "dim_art" i "dim_art_gru" 'dimension' tabela za određenu poslovnu godinu

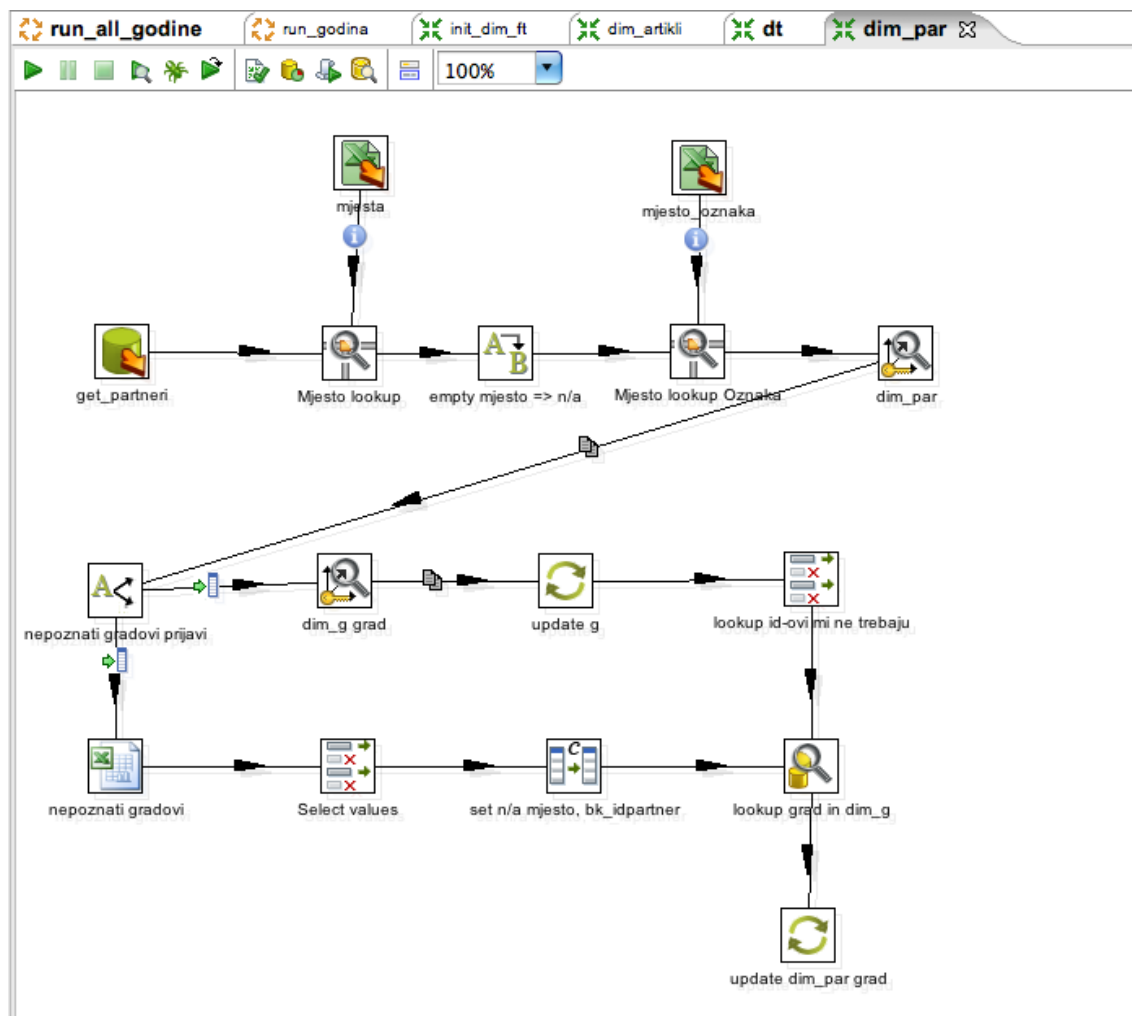
olap_cleansing.xls - LibreOffice Calc

	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					
19	FIN	SW	fmk software					

Sheet 3 / 3 PageStyle_artikli_kod STD Sum=0 100%

Slika 3.5: F18 klasificiranje - šifarski sistem artikala

3.4.2. Šifrarnik klijenata



Slika 3.6: Kettle transformacija: Generacija "dim_par" i "dim_g" 'dimension' tabela za određenu poslovnu godinu

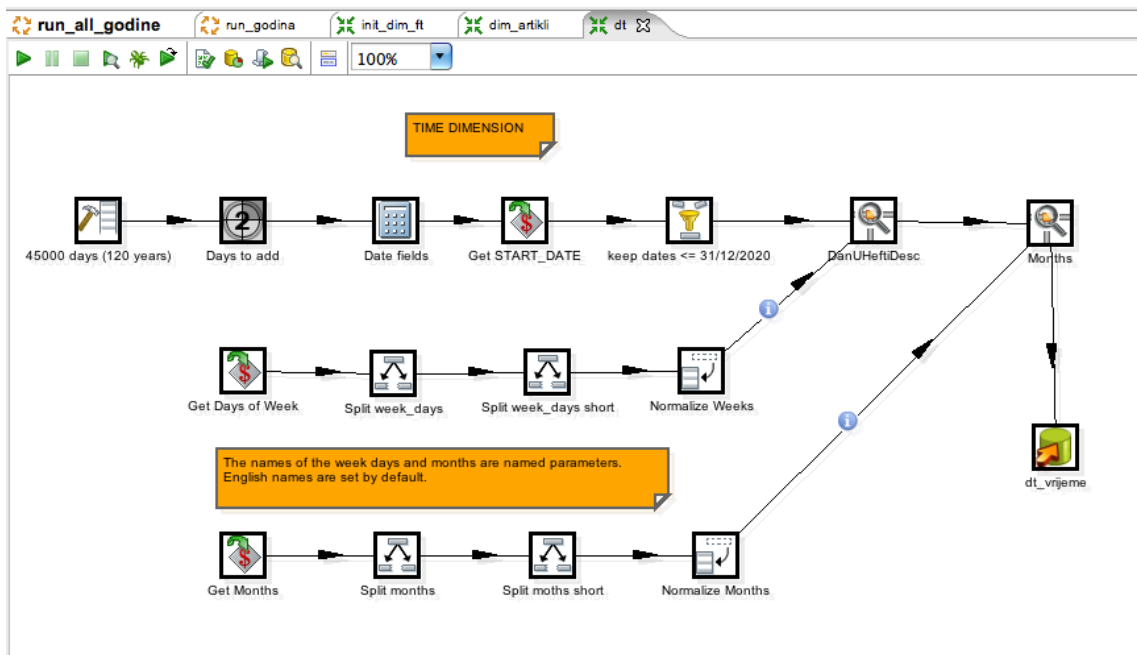
	A	B	C	D	E	F	G	H
1	MjestoIn	Mjesto						
2	BEGOV HAN	Begov Han						
3	ŽELJEZNO POLJE	Begov Han						
4	BIHAĆ	Bihać						
5	BUGOJNO	Bugojno						
6	BUSOVAČA	Busovača						
7	DONJI VAKUF	Donji Vakuf						
8	GORA@DE	Goražde						
9	GORAŽDE	Goražde						
10	GORAZDE	Goražde						
11	U.S.A	inostranstvo			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

Slika 3.7: 'cleansing' F18 podataka - 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						

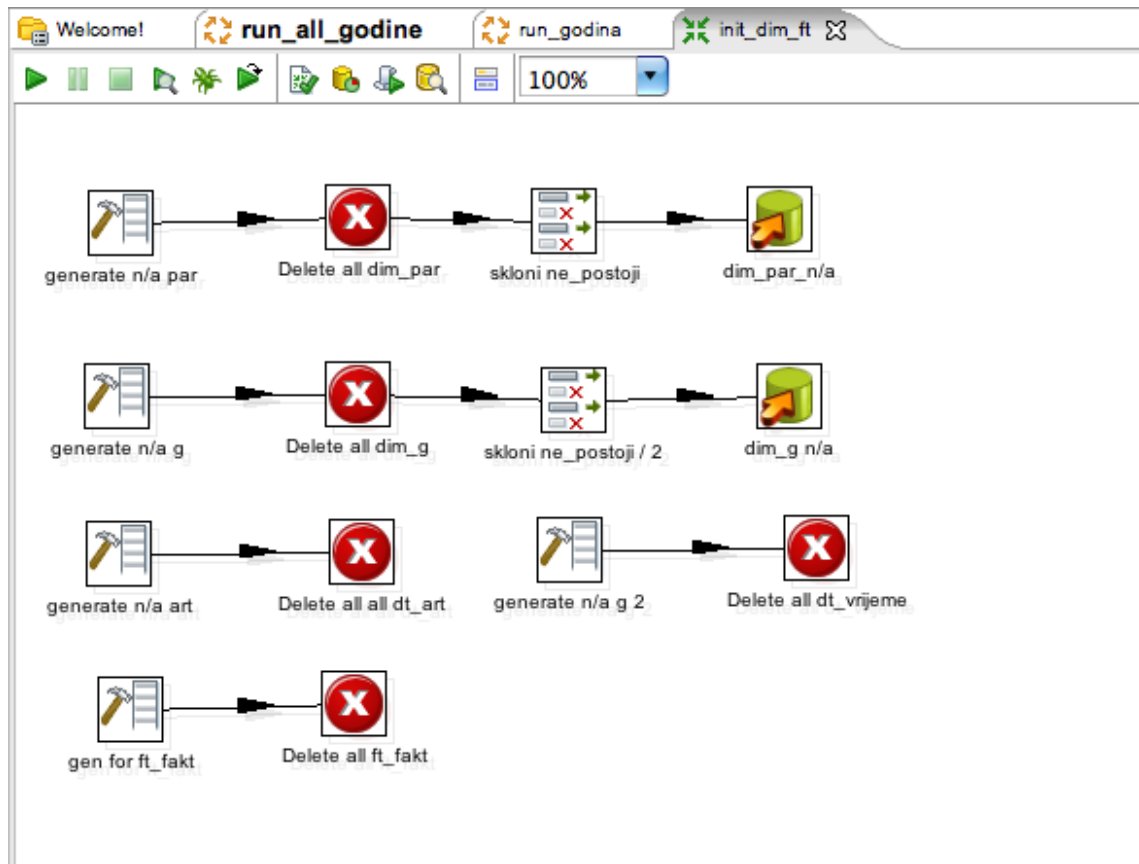
Slika 3.8: F18 kodiranje regiona - klasifikacija mjesta/gradova

3.5. Kreiranje vremenske dimenzije

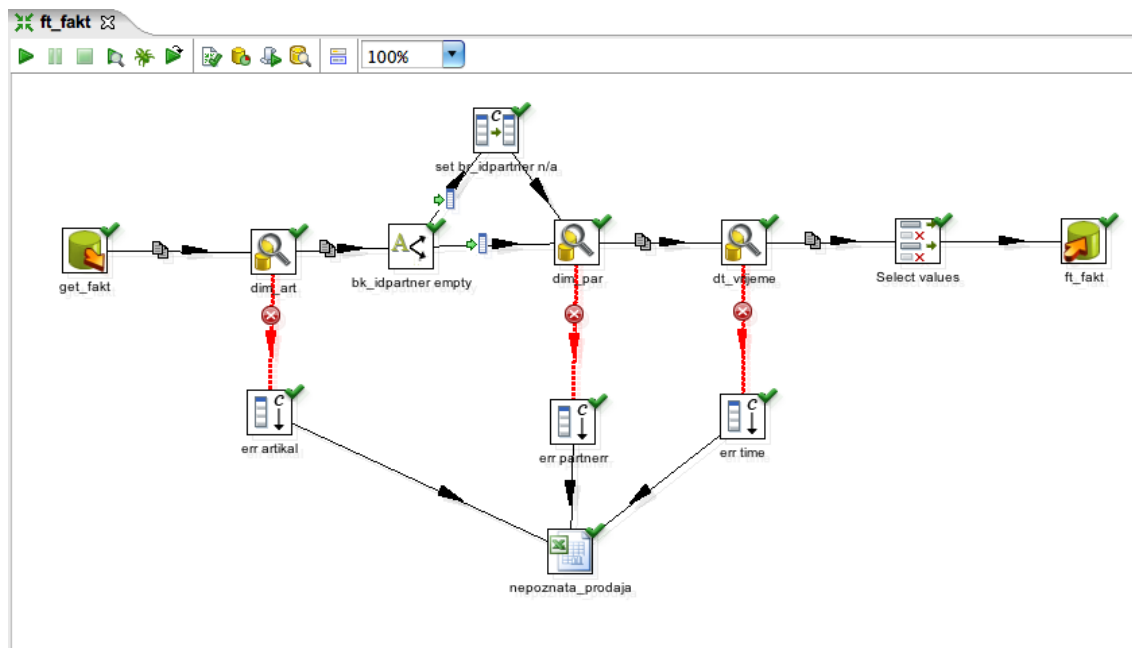


Slika 3.9: Kettle transformacija: Generacija "dim_dt" 'dimension' tabele - vremenska dimenzija

3.6. Kreiranje 'facts' tabele prodaje



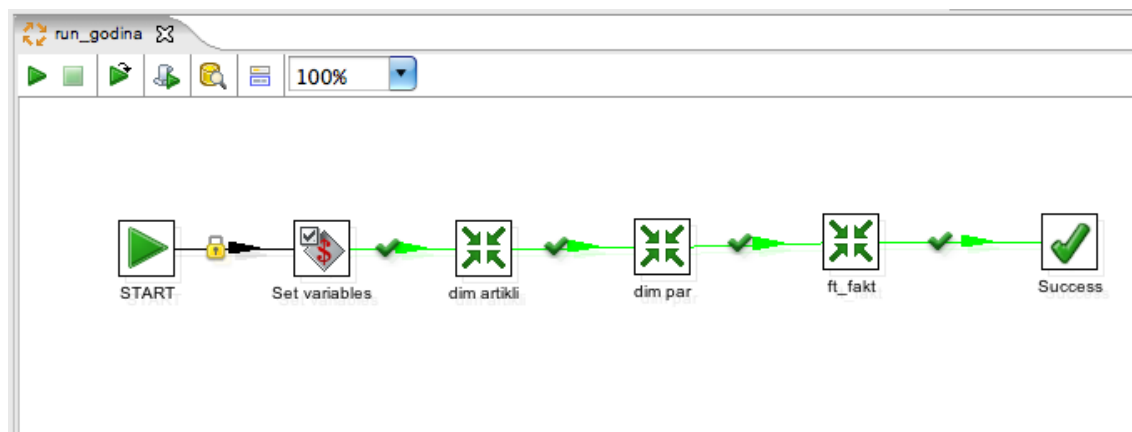
Slika 3.10: Inicijalizacija 'dimension' i 'facts' tabela



Slika 3.11: Generacija "ft_fakt" 'facts' tabele za određenu poslovnu godinu

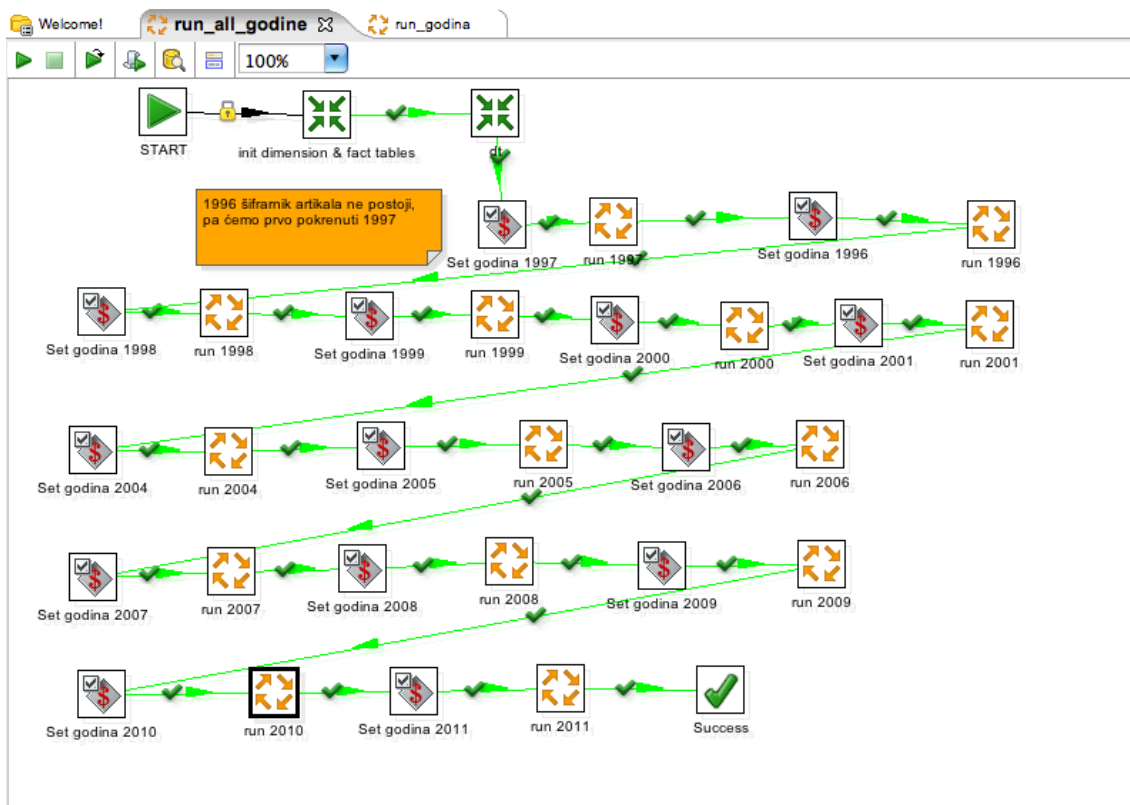
3.7. Kettle 'jobs'

Kettle 'job' omogućava nam da se pojedinačne transformacije i jednostavniji 'job'-ovi integriraju u jedinstven ETL proces koji se može u automatizirati²



Slika 3.12: Kettle job: Generacija OLAP podataka iz F18 ERP izvora za jednu poslovnu godinu

²kreiranje serverskih 'batch' procesa



Slika 3.13: Kettle job: inicijalizacija OLAP tabela, te generacija OLAP pdoataka za sve poslovne godine 1996-2011

The screenshot shows a LibreOffice Calc spreadsheet titled 'artikli_nepoznato.xls'. The spreadsheet has columns A through I. The data is organized into rows, with the first row (row 8) containing the following values: A: 2NAP001, B: 2NAP, C: NAPOJNA JEDINICA L MODEL *, D: 40, E: 40, F: 40, G: 40, H: 40, I: 40. The subsequent rows (rows 9-24) contain various article codes and descriptions, such as 'ŠTAMPAČI, KOPIR APARATI', 'OSTALI PERIFERALI/DODATNA OPREMA', 'MREŽNA OPREMA', 'POTROŠNI MATERIJAL', 'SOFTVER DRUGIH PROIZVOĐAČA', 'OSTALO', 'SOFTVER SIGMA-COMA', 'OSTALI PROGRAMI', 'SCC EXPLORER - MANAGER PROJEKATA', 'NOMENKLATURA ZA SERVIS', 'SERVIS MONITOR', 'OSTALI DIJELOVI', 'RAČUNARI SERVIS', and 'ŠTAMPAČI SERVIS'. The value 1465 is visible in cell E18.

Slika 3.14: Error reporting putem 'spreadsheet' dokumenata - artikli za koje nisu definisani kodovi u olap_cleansing tabelama

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

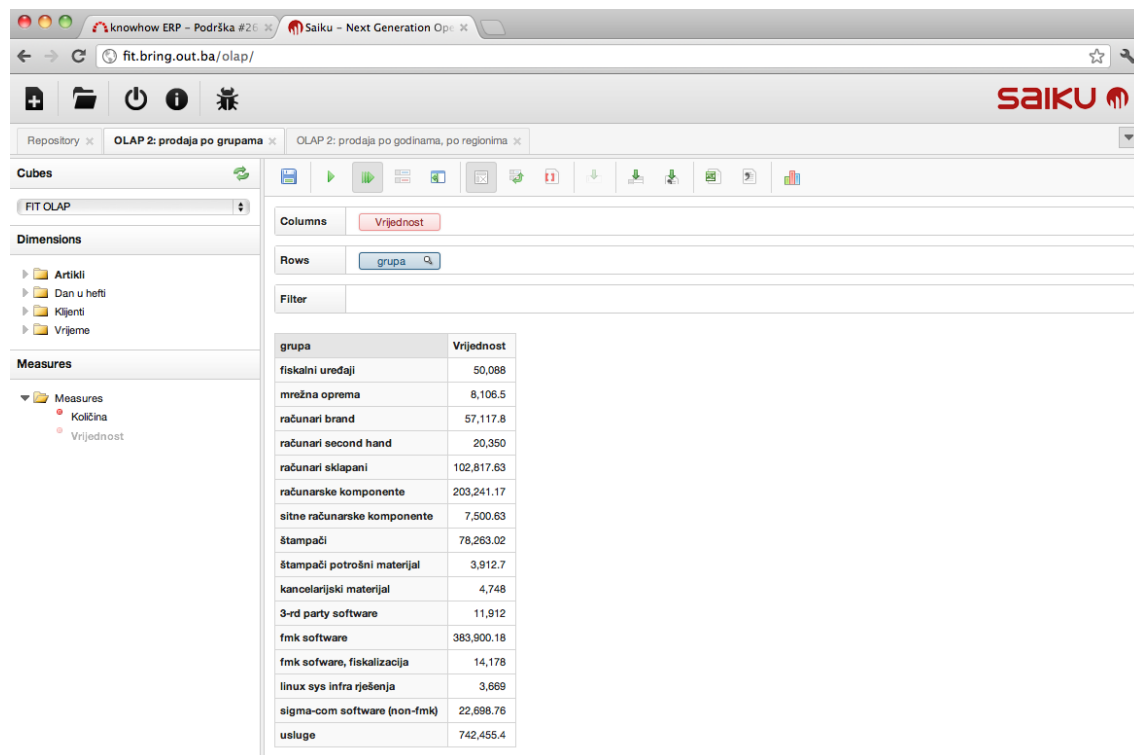
Slika 3.15: Dokumenti prodaje u kojima su neispravni podaci potrebni za popunjavanje dimension tabela (datum, klijent, roba)

3.8. OLAP reporting

‘Data mart’ dobijen ETL procesom koristimo za podešavanje Mondrian sheme OLAP kocke (vidi sliku 2.2).

Dobijena shema se instalira na "Saiku" server. Saiku putem svog web interfejsa omogućava jednostavnu definiciju redova, kolona i filtera putem ‘drag & drop’ operacija.

3.8.1. Pregled prodaje po grupama artikala



grupa	Vrijednost
fiskalni uredaji	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	3,668
sigma-com software (non-fmk)	22,698.76
usluge	742,455.4

Slika 3.16: Pregled prodaje po grupama artikala

‘Saiku’ automatski kreira ‘MDX query’:

```
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

3.8.2. Pregled prodaje po regionima

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
	ZE	Zenica	1,550
2000	BI	Bihac	1,327
		Sanski Most	20
	GOR	Gorazde	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	Bihac	5,693.7

Slika 3.17: 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         [ Klijeti . klijeti ] . [ region ] . Members ) , CrossJoin ( [
8           Vrijeme . vrijeme ] . [ godina ] . Members ,
9           [ Klijeti . klijeti ] . [ mjesto ] . Members ) )
10    )
11 FROM [ FIT OLAP ]

```

Listing 3.2: Pregled prodaje po regionima

Na gornjem izvještaju se mogu očititi mjesta sa oznakom 'n/a'. Naime, radi se o tome da se 'cleansing' procesom nisu identificirali klijenti kod dijela poslovnih transakcija. Ti klijenti su označeni sa "n/a"³.

³n/a not available

Visoki iznosi za određene godine (1996-1999) indiciraju da postoje određeni propusti u ETL procesima za te podatke.

Uzroke treba tražiti u ETL transformacijama vezanim za klijente.

Detalji o neuspješnim transformacijama nalaze se u "nepoznato_" spreadsheet dokumentima⁴.

Vjerovatno je dovoljno kodirati nedostajuće šifre klijenata iz tih godina, te ponoviti kettle job proces "run_all_godine".

⁴primjer takvog dokumenta: dodatak A, artikli_nepoznato xls dokument

4. Zaključak

4.1. Rezultati ‘case study’-ja

Realizacijom ‘case study’-ja dobili smo ‘data mart’ osnovnih parametara prodaje sa kojim smo konstruisali jednostavnu OLAP kocku.

Očekivano, najveći dio posla desio se u fazi pripreme podataka (ETL transformacije i ‘job’-ovi).

Analitičar mora dobro poznavati kako poslovnu domenu tako i tehnološke principe konstrukcije OLAP-a.

Ovakve poslove redovno rade timovi koji se sastoje od biznis analitičara i IT analitičara.

Da bi se došlo do iskoristivih podataka u skladištu podataka potrebna je velika uključenost klijenata (korisnika i vlasnika poslovnih procesa i podataka).

Realni podaci sadrže puno netačnih ili nedostajućih podataka. U procesu sklapanja DW/DMart-a mnogi nedostaci mogu se otkloniti samo na osnovu dodatnih tumačenja podataka od strane klijenata.

4.2. Pogled sa aspekta menadžera

S obzirom da sam kao kreator OLAP kocke ujedno i direktor firme čiji su podaci analizirani, izniću i par direktnih dojmova kao klijent dobijene OLAP kocke:

- Podaci o prodaji potvrđuju niz zaključaka do kojih sam kao nosilac poslovnih procesa dolazio "ad-hoc" zaključivanjem
- Čak i ova jednostavna OLAP kocka omogućava mi da se fokusiram na određene aspekte poslovanja koje tradicionalnim reportingom nisam mogao postići
- Postojeća instalacija ukazala mi je na potrebu i mogućnosti koje bi dobio ovdje dodatnih indikatora poslovanja:

- uvođenje mjere broj_radnika, te na osnovu toga omjer broj_radnika / vrijednost_realizacije po mjesecima, radi analize dosadašnjih efekata zapošljavanja
- finansijski troškovi po mjesecima
- trgovačka marža po mjesecima

Gornja zapažanja upućuju na naredne korake izgradnje skladišta podataka. Očigledno je da naš "data mart" prodaje treba dopuniti podacima iz drugih segmenata poslovanja (finansije, ljudski resursi).

U ishodu, evolucija našeg BI rješenja vodi ka izgradnji "data warehouse"-a firme.

4.3. Pogled sa aspekta informatičara

Pentaho DI (kettle ETL) aplikacije imaju primjenljivost i van užeg konteksta izgradnje BI rješenja.

Oni su moćan alati u svim poslovima transformacije podataka. Mogu se koristiti za jednostavni "ad-hoc" reporting određenih podataka koje postojeći ERP sistem ne obezbjeđuje.

Njihov stepen integracije sa 'spreadsheet' xls dokumentima daje mnoge mogućnosti koje su do sada isključivo tretirani oslovima koje samo software developeri mogu rješavati.

Ukratko, ovi alati značajno pomjeraju granice rukovanja i manipulacije poslovnih podataka.

4.4. Rezime

Izgradnja BI rješenja, čak i najjednostavnijeg koje je u 'case study'-ju demonstrirano, traži značajne tehničke i poslovne resurse.

Izgradnja BI rješenja je dugotrajan i skup proces koji sebi ne može priuštiti svaka organizacija.

OSS BI rješenja tu granicu ipak mogu značajno pomjeriti u korist organizacija sa manjim budžetima.

5. Literatura

The analytical labs. Saiku analytics software, Februar 2012. URL <http://analytical-labs.com>.

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>.

Leo Mršić. *Primjena metoda rudarenja podataka u trgovini tekstilnim i srodnim proizvodima*. Magistarski rad, Ekonomski fakultet u Zagrebu, November 2004.

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. Business intelligence, Februar 2012a. URL http://en.wikipedia.org/wiki/Business_intelligence.

Wikipedia. Olap cube, Februar 2012b. URL http://en.wikipedia.org/wiki/OLAP_cube.

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

Dodatak A

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. artikli_nepoznato 'spreadsheet' dokument https://github.com/hernad/hello_bi/raw/master/artikli_nepoznato.xls
5. Saiku demo server online: <http://fit.bring.out.ba/olap/#>

Dodatak B

Bilješke

1. Prva verzija ovog seminarskog rada, neuspješno https://github.com/hernad/MIS/raw/master/knowhowERP_OLAP_blog_style.pdf
2. FIT OLAP 2 cube: <http://redmine.bring.out.ba/issues/26711>