

UNIVERZITET „DŽEMAL BIJEDIĆ“ U MOSTARU
Fakultet informacijskih tehnologija

Predmet: Menadžment informacioni sistemi

Tema: **“knowhow ERP” Business intelligence platforma**

Seminarski rad - neuspješno, blog style

ver 0.9.2

Mentor:
asistent Mohamed El-Zayat

Student:
Ernad Husremović, 2792

Mostar, januar 2012 godine

Tabela sadržaja

1) Uvod.....	3
2) Pripremne radnje.....	3
2.1) Upoznavanje sa konceptima BI-a, OLAP-a.....	3
2.2) Traženje tehnološke osnove za “knowhow ERP”	3
3) Odabir tehnološke osnove: trojac “pentaho / mondrian / saiku”	4
3.1 Šta je Saiku ?.....	4
4) prvi OLAP Workflow by Ernad.....	6
4.1) Definisanje Inputa - operativni podaci iz ERP sistema.....	6
4.2) Kettle ETL dizajner - Spoon.....	7
4.3) Mondrian OLAP Schema.....	8
4.4) Instalacija OLAP-a na server.....	9
4.5) I konačno ... “Saiku” OLAP Analyzer.....	9
4.6) Bilješke tokom kreiranja OLAP cube-a	10
4.6.1) Moj prvi OLAP cube se flat fakt tabelom.....	10
4.6.2) Finalni cube - star db schema sa više dimenzija, sample data identičan.....	10
5) Artifakti projekta.....	11
6) Zaključci.....	11
7) Uočeni problemi.....	11
8) Neriješene stavke.....	11
9) budućnost “knowhow ERP” BI-a.....	11
10) Način izrade seminarskog rada.....	12
10.1) “knowhow ERP” development.....	12
11) Popis software-a i korištenih tehnologija.....	12
11.1) Developer's toolset.....	12
11.2) Background tehnologije:.....	13
11.3) Sistemska infrastruktura (fit.bring.out.ba deployment):.....	13
11.4) Aplikativno rješenje.....	13
12) Zahvala.....	14
R) Revizije:.....	14

1) Uvod

O BI¹-u sam imao samo površna znanja. Stepen mog znanja se može ocijeniti “Čuo sam o čemu se tu radi, ali da nešto konkretno znam - ne znam”.

Prvobitno sam mislio da ovaj seminarski bude reporting sistem baziran na Pentaho BI-u, ali reporting sistem i nije “neki fol”. Analiza podataka sa OLAP kockama je suština BI-a. Svaki dobar ERP ima razvijen reporting sistem.

Traganjem za rješenjima došao sam do “Saiku” projekta koji me je oduševio.

Na kraju se moj rad sveo na formiranje OLAP kocke u Mondrian ROLAP/PostgreSQL Data warehouse-u i instalacija te kocke na “Saiku” analytics server.

OLAP je sjajna stvar i sjajan koncept. OLAP je stvarno nešto sasvim drugačije od klasičnog reporting sistema.

Ovaj rad je “hello world” OLAP rješenje.

2) Pripremne radnje

2.1) Upoznavanje sa konceptima BI-a, OLAP-a

BI teorija: <http://redmine.bring.out.ba/issues/26337>

Pentaho BI glossary: <http://redmine.bring.out.ba/issues/26320>

Internet resursi o Pentaho-u: <http://redmine.bring.out.ba/issues/26324>

Šta su MDX queries ? <http://redmine.bring.out.ba/issues/26329>

Multidimenzionalni upiti na OLAP kocku.

2.2) Traženje tehnološke osnove za “knowhow ERP”

Prvi pokušaj BI-a je bio podesiti ErpBI: <http://redmine.bring.out.ba/issues/26319>, ali neuspješno.

Kada to nisam uspio nastavio sam dalje tragati. Traganje završio sa saiku. Ne želim što ErpBI nije proradio.

Međutim, ima se smisla vratiti ponovo na ErpBI nekom drugom prilikom jer on ima podešene kocke i

¹ BI - Business intelligence

reporte za xtuple (knowhowERP) transakcijske - operativne podatke.

3) Odabir tehnološke osnove: trojac "pentaho / mondrian / saiku"

Pentaho BI (opensource CE edicija) je već u početku kao glavni kandidat.

Pentaho BI u komercijalnom dijelu funkcionira kao "open core"² projekat.

Upravo je područje OLAP kocki koje su kreatori Pentaho-a uzeli kao područje koje pripada "enterprise" (plaćenj verziji).

Kako "knowhow ERP" za svoju tehnološku osnovu koristi isključivo otvoreni software, Pentahovo rješenje OLAP-a je isključeno kao opcija.

Tako sam došao do "Saiku" projekta.

3.1 Šta je Saiku ?

Na [naslovnoj stranici](#) proizvođača³ piše:

Modular open-source analysis suite offering lightweight OLAP which remains easily embeddable, extendable and configurable.

Saiku je opensource projekat baziran na Pentaho BI-u. Saiku je OLAP Analyzer sa cool HTML5 user interface-om⁴.

Bilješke: <http://redmine.bring.out.ba/issues/26327>

Saiku na fit.bring.out.ba:

The screenshot shows the Saiku web application interface. The browser address bar displays fit.bring.out.ba/olap/. The interface includes a sidebar with a tree view of dimensions (Artikal, Grad, Kategorija Artikala, Partner, Vrijeme) and measures (Kolicina, Prodaja, Prosječna kolicina). The main area shows a query result table for the 'prodaja-po-kategorijama' repository. The table has columns for 'MeasuresLevel', 'Kategorija', and years from 2010 to 2011. The data is filtered by 'Prodaja' and 'Kategorija'.

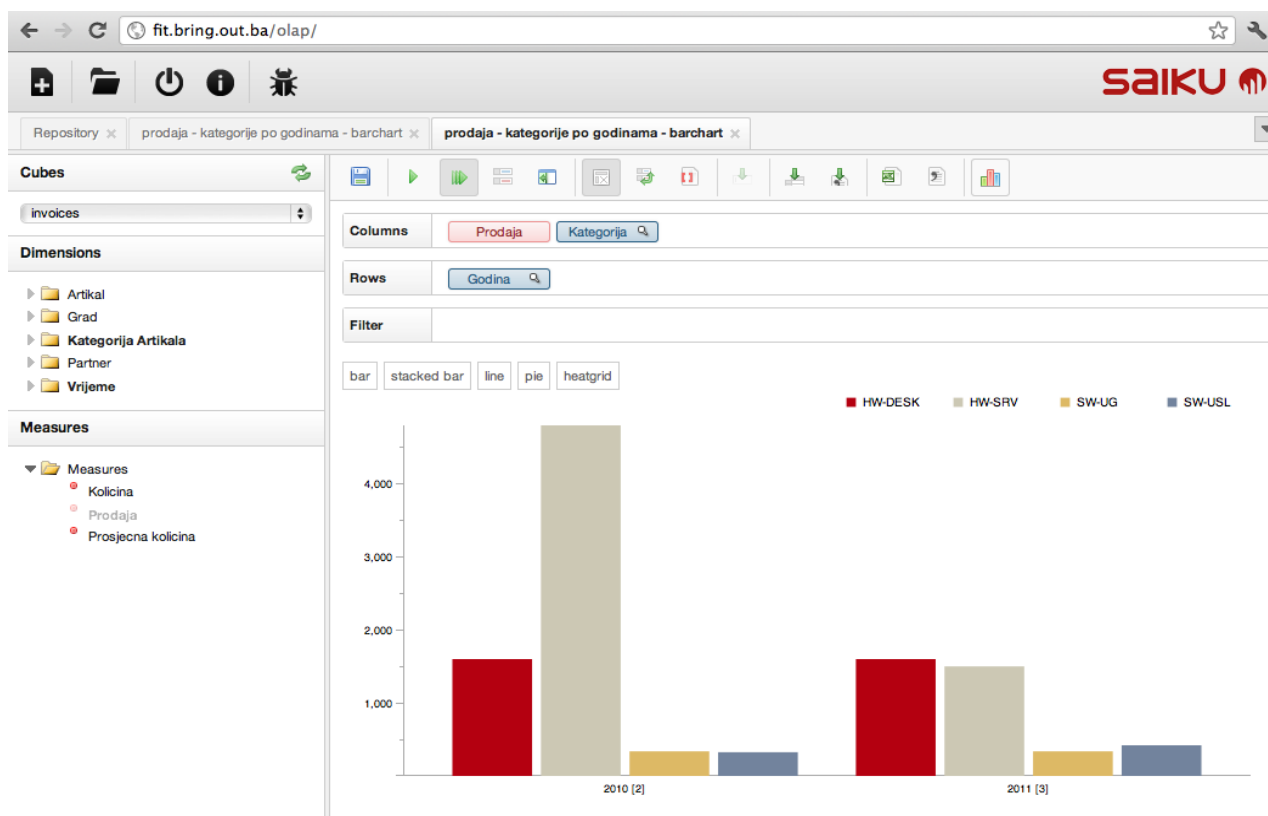
MeasuresLevel	Kategorija	2010	2010	2010	2010	2010	2010	2010	2010	2011	2011	2011	2011	2011
		1	2	3	4	5	6	12	1	2	3	4	10	
Prodaja	HW-DESK			1,600							1,600			
	HW-SRV			1,700				3,100			1,500			
	SW-UG		297				40			297			40	
	SW-USL	120			65	140			55	160		65	140	

² <http://hernad.bring.out.ba/open-source-money>

³ <http://analytical-labs.com>

⁴ <http://documentcloud.github.com/backbone/>

Podrška i grafički prikaz⁵:



⁵ Ova opcija je poprilično bug-ovita, tabelarni prikaz je, očekivano, u većini slučajeva u ovom trenutku razvoja Saiku jedina opcija

4) prvi OLAP Workflow by Ernad

Sve je ovo lijepo, ali nije konkretno :).

Nadam se da ste oduševljeni gornjim “screen shot”-ovima kao i ja kada sam prvi put vidio Saiku u akciji⁶.

Međutim, da bih ovladao osnovnim znanjima i došao do sopstvenih konkretnih rezultata, trebalo mi je 30-40 sahata.

Sada ću demonstrirati svoje svoje rezultate:

4.1) Definisanje Inputa - operativni podaci iz ERP sistema

Nakon što sam počeo čitati (vidi “Pripremne radnje”) brzo sam uočio sam da je “osvajanje” i usvajanje OLAP -a “veliki zalogaj”. O pojmovima kao što su OLAP-u⁷, ETL⁸, Data Mining, Data mart, Data warehouse sam čuo kako rekoah ranije, ali de-factno nisam ništa znao.

Zato sam se odmah odlučio da ulazni podaci budu što jednostavniji, da ne bih na tome izgubio vrijeme.

Radi jednostavnosti sam napravio jednostavnu spreadsheet tabelu (vidi dole “Aplikativno rješenje”) koju sem tretirao kao ulazne operativne podatke za moj OLAP Data warehouse.

Inače ti ulazni podaci se u OLAP terminologiji označavaju terminima “Data mart”, “operational data”.

Zamislite da je naš spreadsheet neka tabela prodaje iz našeg ERP sistema koja je dobijena joinom par tabela i da je to input u naš OLAP sistem⁹.

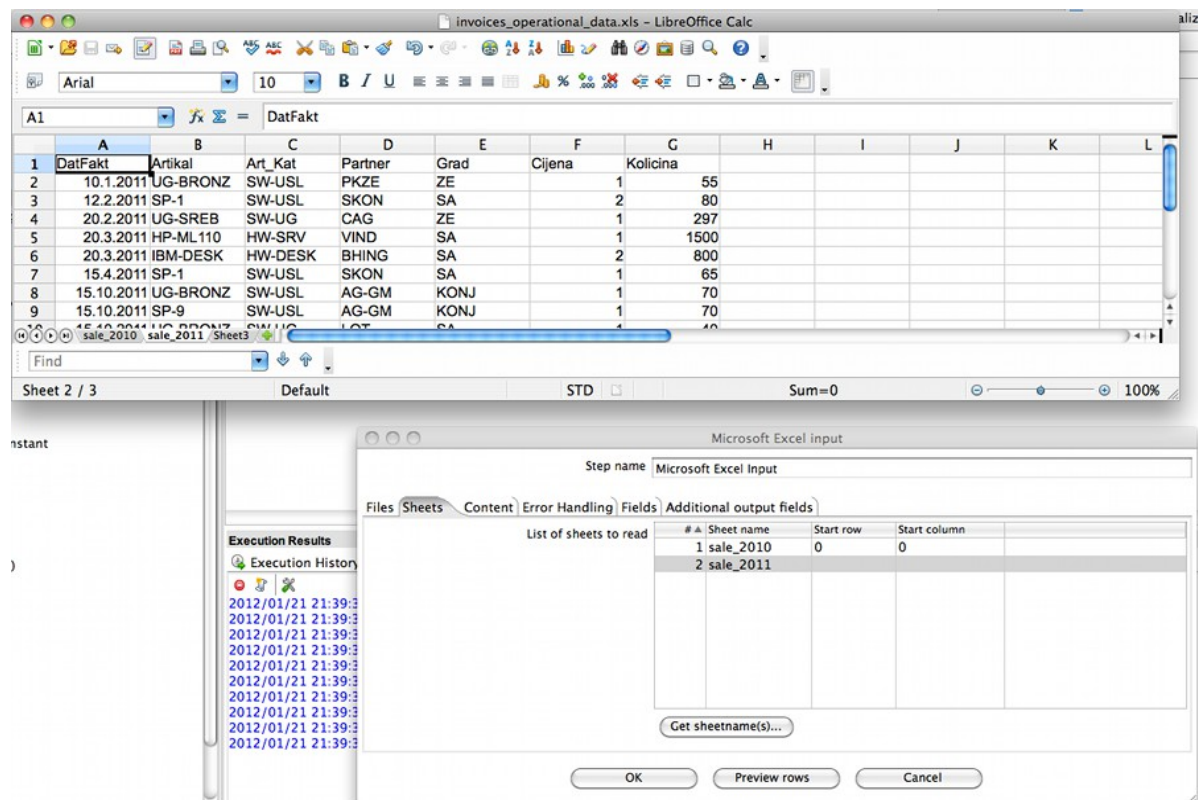
6 Na njihovoj naslovnoj stranici nalazi se odličan promotivni video <http://analytical-labs.com/>

7 OLAP - Online analytical processing

8 ETL - Extract - transform - load

9 Prvo sam napravio sheet koji se odnosio na prodaju jedne godine, pa sam dodao i drugu da bi testni podaci bili što smisleniji za analizu.

Evo kako to izgleda:

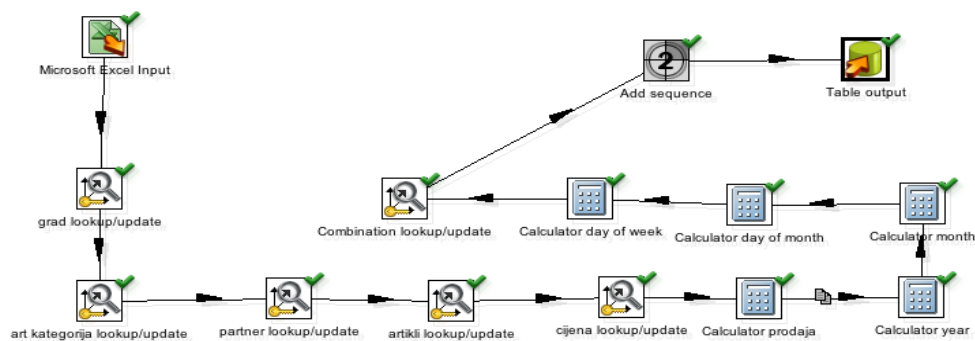


4.2) Kettle ETL dizajner - Spoon

Gornja tabela pored tabele u Libreoffice-u pokazuje i screen “Spoon” aplikacije (vidi dole Developer's toolset).

Naime “Spoon” je mjesto gdje se definiše **extract** podataka (naša spreadsheet excel tabela), **transformation** (predprocesiranje, proračun vrijednosti koje naš ulazni stream podataka ne sadrži) te **load podataka** (OLAP PostgreSQL data warehouse).

Evo kako izgleda na desktopu naš ETL:



Alat je jednostavan i intuitivan¹⁰. Koncept dizajna mi se pravo svidio. Jednostavno pratiš svoj cilj i kettle/Spoon će te dovesti do njega¹¹.

U gornjoj tabeli uočavamo sljedeće korake transformacije:

1. ulazni podaci - MS Excel
2. lookup/update koraci koji kreiraju “dimension” tabele
3. calculator koraci koji proračunavaju podatke koje ulazni stream ne nudi (prodaja = količina * cijena, year = YEAR(Datum fakture))
4. Generator id-ova. U našem inputu nemamo definisane id-ove za pojedine zapise, što nam je potrebno za našu ROLAP PostgreSQL “facts” tabelu. Ovaj korak generiše ID po algoritmu jednostavnog inkrementa
5. izlazni strim - snimanje “facts” tabele u naš OLAP store

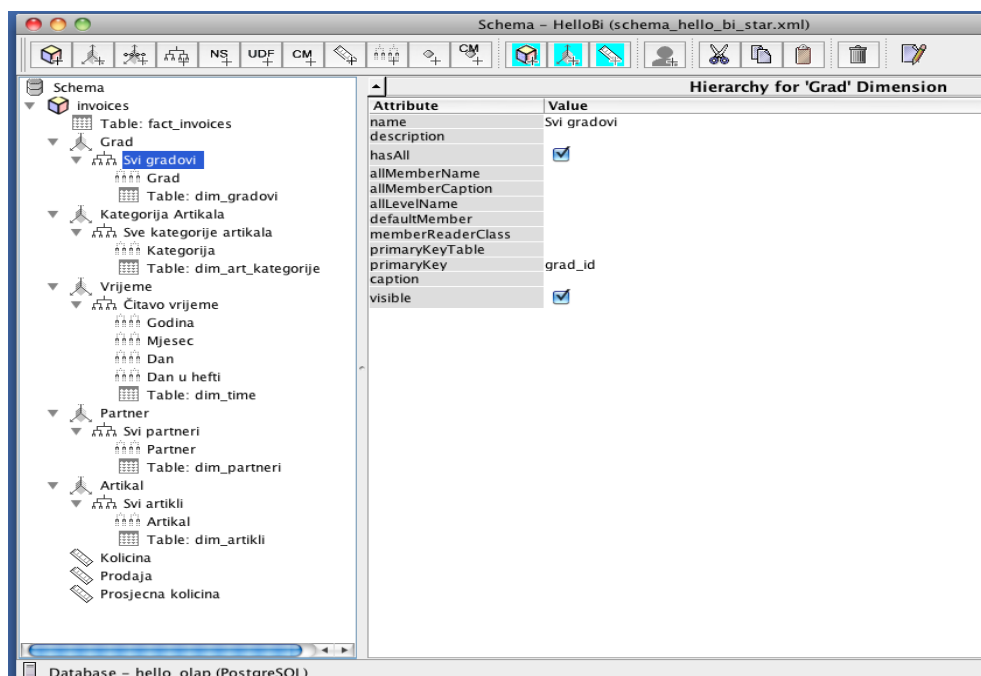
Eto to je to. Sada mi čak i jednostavno izgleda :).

A vama, dragi moj čitaocē ? Razumijete li vi o čemu pričam ?

4.3) Mondrian OLAP Schema

Ova schema je u pojavnom smislu obični xml fajl koji Mondrian OLAP server koristi da relacione podatke (facts tabele i dimension tabele) poveže u multidimenzionu OLAP kocku.

Pustimo se priče. Evo kako to izgleda u schema designer-u:



¹⁰ Spoon je klasična Java SWING aplikacija. Kao i sve SWING aplikacije usability baš nije “boli glava”, ali za naše potrebe sasvim dovoljno. Ne smije se zaboraviti benefit koji nam Java nudi - multiplatformska aplikacija.

¹¹ Na redmine ticketima postoji dosta bilješki o tome.

Međutim, ima smisla pogledati kako to izgleda “ispod haube”: [fit_olap_star_schema.xml fajl](#)¹².

XML fajl je prilično čitljiv i logičan. Čak sam se u prvim koracima učenja bolje snalazio u xml-u nego li sa grafičkim dizajnerom.

4.4) Instalacija OLAP-a na server

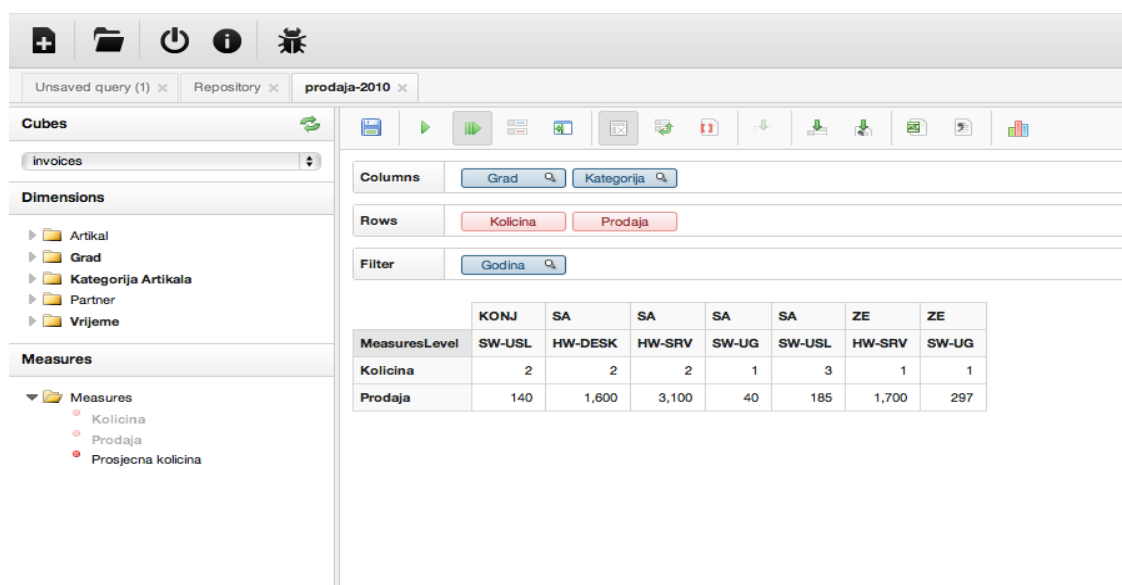
Kako god bilo, sa pravilno definisanom shemom, naša OLAP kocka je definisana, i možemo je deploy-ati na naš “Saiku” OLAP analytics server.

Na ticketu [“fit.bring.out.ba - saiku server OLAP analyzer”](#) dokumentovana je instalacija servera na produkcijski server¹³

<http://fit.bring.out.ba/olap/>

4.5) I konačno ... “Saiku” OLAP Analyzer

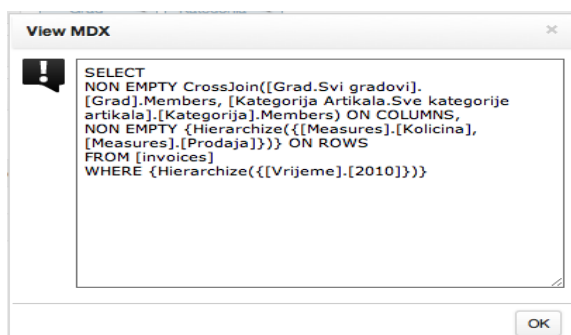
Tabelarni pregled podataka:



The screenshot shows the Saiku OLAP Analyzer interface. On the left, there's a sidebar with 'Cubes' (invoices), 'Dimensions' (Artikal, Grad, Kategorija Artikala, Partner, Vrijeme), and 'Measures' (Measures, Kolicina, Prodaja, Prosjecna kolicina). The main area shows a configuration for a cube named 'prodaja-2010'. The 'Columns' section has 'Grad' and 'Kategorija'. The 'Rows' section has 'Kolicina' and 'Prodaja'. The 'Filter' section has 'Godina'. Below this, a data table is displayed with columns for 'MeasuresLevel', 'KONJ', 'SA', 'SA', 'SA', 'SA', 'ZE', and 'ZE'. The table contains data for 'Kolicina' and 'Prodaja' across different levels and categories.

	KONJ	SA	SA	SA	SA	ZE	ZE
MeasuresLevel	SW-USL	HW-DESK	HW-SRV	SW-UG	SW-USL	HW-SRV	SW-UG
Kolicina	2	2	2	1	3	1	1
Prodaja	140	1,600	3,100	40	185	1,700	297

Iza svakog pregleda “krije” se MDX¹⁴ query:



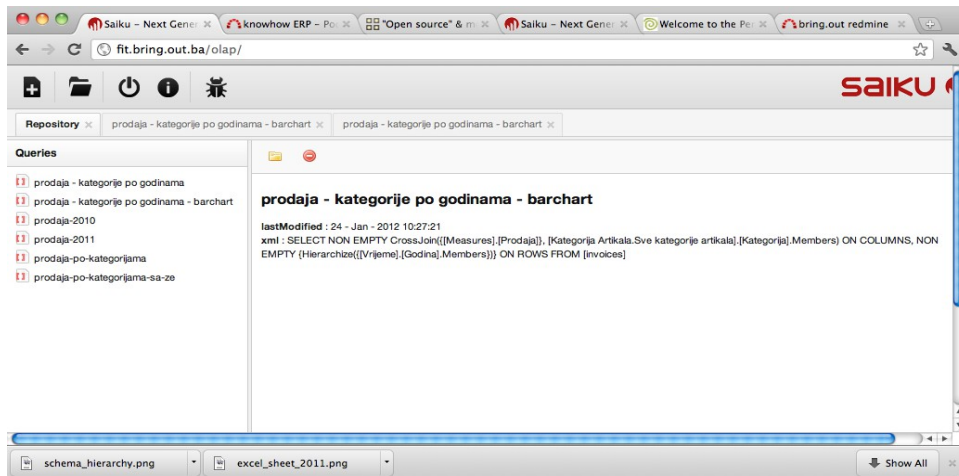
```
SELECT
NON EMPTY CrossJoin([Grad.Svi gradovi],
[Grad].Members, [Kategorija Artikala.Sve kategorije
artikala].[Kategorija].Members) ON COLUMNS,
NON EMPTY {Hierarchize({[Measures].[Kolicina],
[Measures].[Prodaja]})} ON ROWS
FROM [invoices]
WHERE {Hierarchize({[Vrijeme].[2010]})}
```

12 Pored ovog schema fajla u github reposu postoji i prva varijanta koja je kao OLAP store imala jednu jedinu facts tabelu bez dimenzija https://github.com/hernad/hello_bi/blob/master/schema_hello_bi.xml

13 Naravno u kontekstu ovog seminarskog rada

14 MDX - Multidimension OLAP cube query

Nakon definicije, MDX query se može snimiti za kasniju “one-click” upotrebu:



Svaki pregled podataka može se exportovati u spreadsheet aplikaciju (saiku-export):

The screenshot shows the LibreOffice Calc spreadsheet titled 'saiku-export (6).xls'. The spreadsheet has columns A through J and rows 1 through 7. The data is organized into a table with headers and values. The table structure is as follows:

	A	B	C	D	E	F	G	H	I	J
1										
2			KONJ	SA	SA	SA	SA	ZE	ZE	
3		MeasuresLevel	SW-USL	HW-DESK	HW-SRV	SW-UG	SW-USL	HW-SRV	SW-UG	
4		Kolicina	2	2	2	1	3	1	1	
5		Prodaja	140	1600	3100	40	185	1700	297	
6										
7										

The spreadsheet also shows a 'Find' bar at the bottom and a status bar indicating 'Sheet 1 / 1', 'PageStyle_Sheet', 'STD', 'Sum=0', and '100%'.

4.6) Bilješke tokom kreiranja OLAP cube-a

4.6.1) Moj prvi OLAP cube se flat fakt tabelom

hello Pentaho BI, hello Mondrian OLAP cube: <http://redmine.bring.out.ba/issues/26338>

4.6.2) Finalni cube - star db schema sa više dimenzija, sample data identičan

hello_bi OLAP star cube: <http://redmine.bring.out.ba/issues/26339>

5) Artifakti projekta

Svi artifakti nalaze se ovdje: https://github.com/hernad/hello_bi

Source code (moj “fork”) Saiku projekta: <https://github.com/hernad/saiku>

6) Zaključci

TBD

7) Uočeni problemi

“Saiku” je projekat u razvoju. Uočio sam dosta bug-ova kod podešavanja izvještaja.

Bug-ove na koje sam do sada naišao sam uglavnom riješio poznatim informatičkim principom “izađi/uđi” :).

U većini slučajeva bilo je dovoljno napraviti izvještaj od početka, a ako to ne ide, restart java servera¹⁵

To je u ovom slučaju bilo - počni praviti izvještaj od početka.

Sama sesija fit.bring.out.ba je za memorijski zahtjevne java procese poprilično slaba¹⁶, i često prijavljuje “java heap probleme” što uzrokuje potrebu za restartom saiku servera.

8) Neriješene stavke

Glavne neriješene stavke

1. ETL proces je napravljen manuelno. On se naravno treba autmatizirati u jedan “batch proces”
2. Input iz ERP sistema knowhow ERP database-a

9) budućnost “knowhow ERP” BI-a

Smatram da je ovaj seminarski rad dobro polazište za daljnji razvoj BI funkcija “knowhow ERP”.

Procjenjujem da je potreno sljedeće vrijeme za razvoj verzija koje se mogu nuditi krajnjim kupcima:

1. knowhow ERP BI ver 1.0.0
 - procjena vremena: 8 hefti, od toga 2 hefte za dodatno upoznavanje sa alatima i tehnologijama
 - Sadržaj: OLAP cubes prodaje.
 - Prvi koraci već napravljeni u ovom radu
2. knowhow ERP BI ver 2.0.0
 1. procjena vremena: 6 hefti
 2. Sadržaj: OLAP cubes koje kao pomoć analizi proizvodnje

¹⁵ stop_saiku.sh / start_saiku.sh

¹⁶ Naime “openvz” container - light virtualizacijska tehnologija, čak i kada sesiji daš 4GB tih 4GB nije garantovano. bring.out ima u planu portiranje virtualnih rješenja na linux KVM bazirana virtualizacijska rješenja (openstack, ganeti)

10) Način izrade seminarskog rada

Ovaj seminarski rad urađen je drugačije. Tema se razlikuje od onoga što su radili drugi studenti, ali nije u tome “kvaka”.

Ovaj seminarski je bio studentsko istraživanje. Student je tražio određeno (ne precizno definisano) rješenje za odgovarajući poslovni i tehnički problem.

Mislim da je to najveći kvalitet ovog rada.

10.1) “knowhow ERP” development

Drugi segment koji želim istaći je “knowhow ERP” način rada.

Seminarski radovi koje sam gledao su uglavnom ... dosadni i sterilni.

Oni naime uopšte ne pokazuju kako je student došao do rješenja, šta je radio, na koje je probleme nailazio. Ukratko, izgledaju mi kao priče za malu djecu.

Informatički zadaci i problemi su kompleksni. U njima se krije hiljadu problema i zamki. Izvršilac treba isprocesirati ogromnu količinu informacija, često sa nepoznatim ishodom.

Svaka ta informacija, svaki taj problem, posebno u timskom radu može biti presudni faktor kvaliteta i produktivnosti projekta.

Zato “knowhow ERP” projekat zahtjeva od svakog izvršioca da svoje operacije evidentira, bez obzira da li su uspješne ili ne.

Zato je seminarski rad je prepun linkova na moje radne bilješke (redmine tickete, u žargonu “knowhow ERP”, odnosno bring.out).

Te bilješke mogu stvoriti značajnu distrakciju čitaocu, ali kako je ovo studentski rad, rad koji treba demonstrirati postignuća studenta, namjerno sve te linkove ostavio.

Rijetko koji rad navodi probleme na koje student naišao, šta je pokušao, zašto je “zaglavio”.

Tako sam primjera radi naveo ErpBI koji nisam uspio konfigurisati kako treba. Iako to izgleda kao neuspjeh, pokazalo se kao dobra strategija. Nisam potrošio dva dana na pokušaj da podesim ErpBI nego sam nastavio traganje za drugim rješenjima i, što je najbitnije, upoznavanje OLAP koncepata.

11) Popis software-a i korištenih tehnologija

11.1) Developer's toolset

1. Mac OSX Desktop 10.6
2. java (OS builtin)
3. vim/mvim programmer's editor 7.3.x
4. libreoffice, 3.4.3
 - korišten za izradu transaction sistem sample data spreadsheet
 - izrada ovog dokumenta

5. redmine project management & ticketing sistem (<http://redmine.bring.out.ba>)
6. gimp (2.6.11)
 - resize screenshot-ova
7. Pentaho developer toolset:
 1. Pentaho data integration 4.2.1 CE
 - Kettle Spoon ETL transformation designer
 2. Mondrian schema workbench ver 3.3.0.14703 (EPL license)

11.2) Background tehnologije:

- javascript backbone web rich client HTML5 library (used by saiku ui)
- java server tomcat 6.0.x (used by pentaho, saiku server)
- Mondrian (pentaho Analysis 3.3 OLAP cube)

11.3) Sistemska infrastruktura (fit.bring.out.ba deployment):

1. VM Host: Proxmox/debian/openvz environment
2. openvz sesija fit.bring.out.ba:
 1. 4GB RAM (shared)
 2. linux/ubuntu lucid 10.04
 3. openjdk java 6b20
 4. PostgreSQL 9.1
 - java JDBC drajver: postgresql-8.4-702.jdbc3.jar

11.4) Aplikativno rješenje

1. Saiku 2.2-SNAPSHOT¹⁷
2. kettle ETL transformation¹⁸
 - input - operational sample data: spreadsheet¹⁹
 - output - ROLAP mondrian store PostgreSQL

¹⁷ <https://github.com/hernad/saiku>

¹⁸ https://github.com/hernad/hello_bi/raw/master/hello_bi_star.ktr

¹⁹ https://github.com/hernad/hello_bi/raw/master/invoices_operational_data.xls

12) Zahvala

- Pentaho što su napravili sjajan BI paket kao opensource software
- Analytical-labs što su napravili odličan i cool OLAP Analyzer.
- SLAWOMIR CHODNICKI čiji mi je blog bio od ogromne pomoći
- Asistentu Mohamedu i profesorici Bevanda što su mi omogućili da radim Seminarski rad po sopstvenom izboru

R) Revizije:

- draft, 23.01.2012, u neko doba noći
- ver 0.9.0, 24.01.2012
- ver 0.9.1, asistent se zove: Mohamed El-Zayat
- ver 0.9.2, typos