

A faint background graphic of a network diagram with several circular nodes connected by lines. One node in the center-left is highlighted in dark blue.

O stvarima koje bi svaki DW/BI sustav trebao imati

Marino Ljubić
marino.ljubic@multicom.hr

Multicom

- **Glavna područja ekspertize:**

- Data Mining
- Obračun i naplata (**Billing**)
- Upravljanje matičnim podacima (**MDM**)
- Skladišta podataka (**DWH**) i Poslovna Inteligencija (**BI**)
- **B2B**
- Upravljanje korisničkim procesima (**CRM**)



Reference

Telco

...

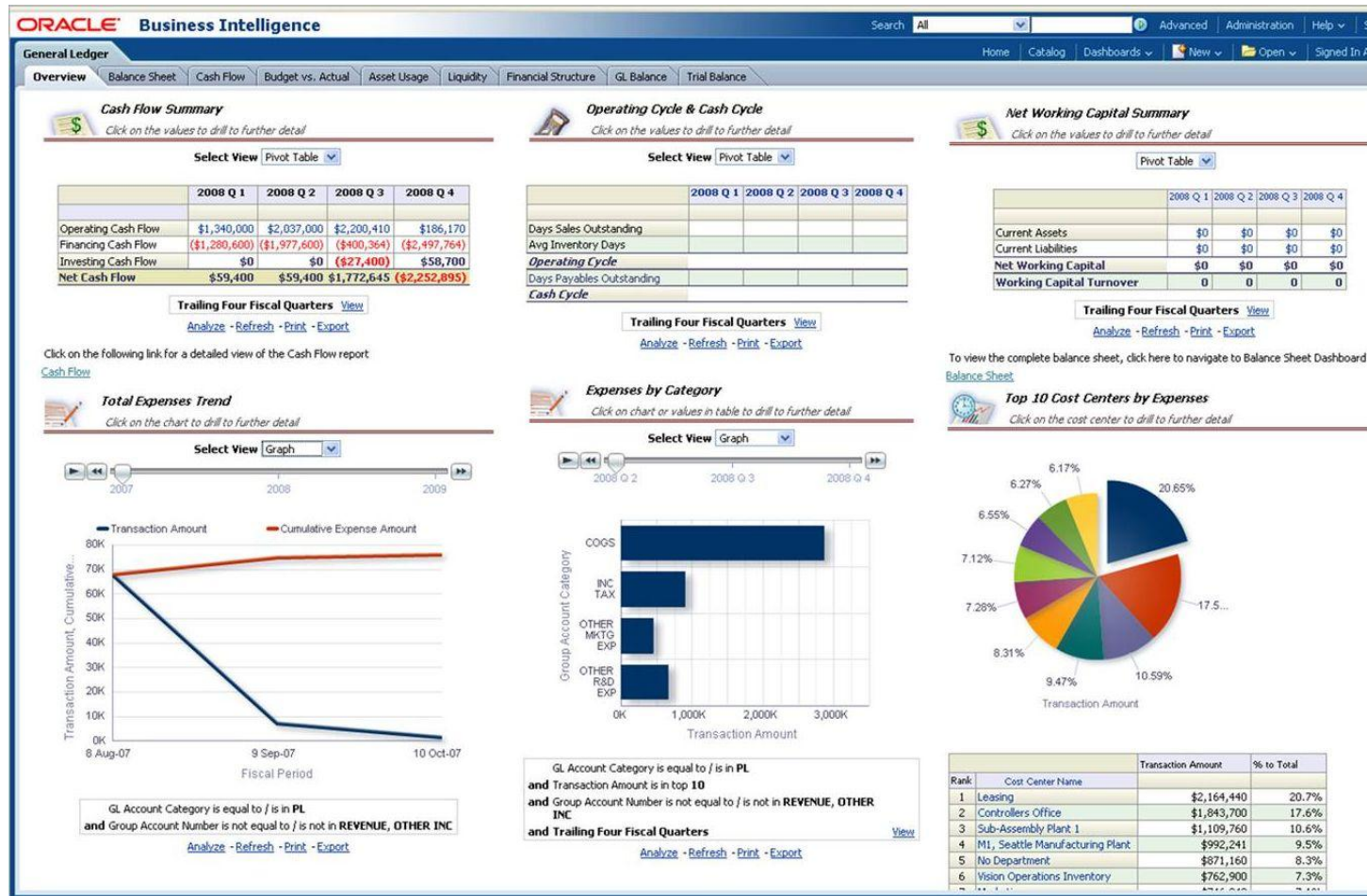
Finance

...

Public/Utility



Postoji li netko tko ovo nije vidio?



BI trendovi

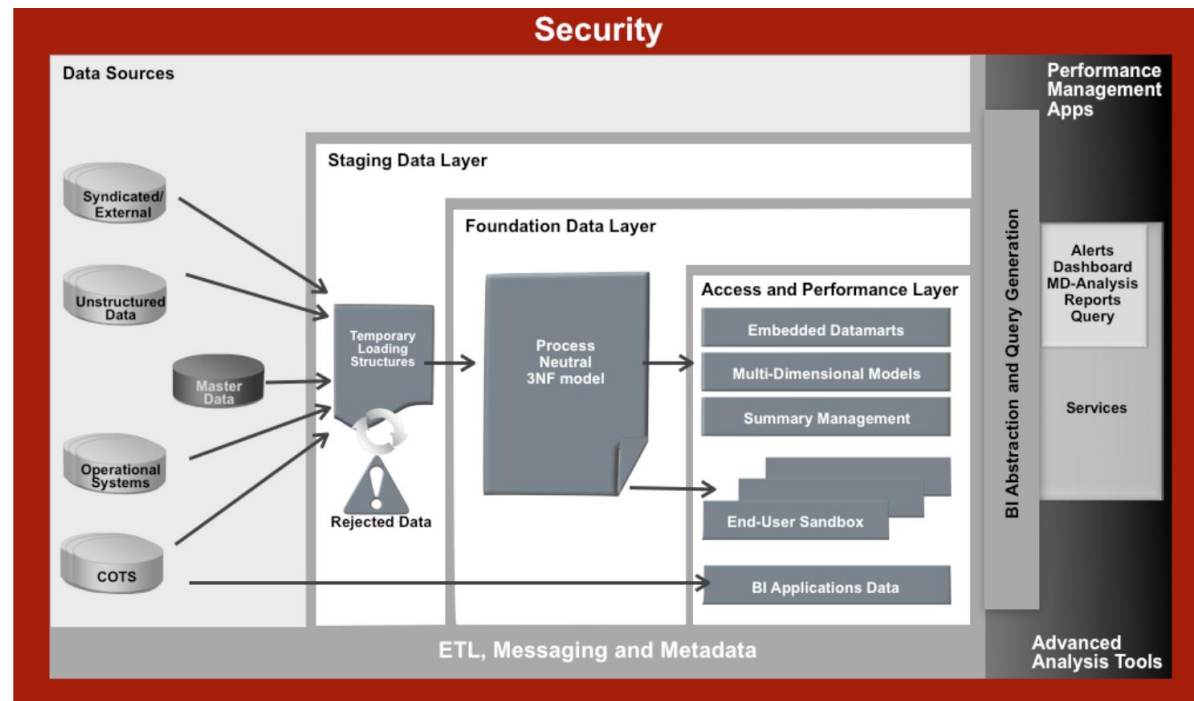
- In-Memory rješenja
 - Izgradnja podatkovne kocke u radnoj memoriji računala
- Nema zasebne analitičke baze
 - OLTP je ujedno i OLAP
 - Paralelni dvostruki storage (Row/Column)
- “Agile BI”
- Brže, bolje, jednostavnije
- Hmm...

BI bez DW-a: Mane

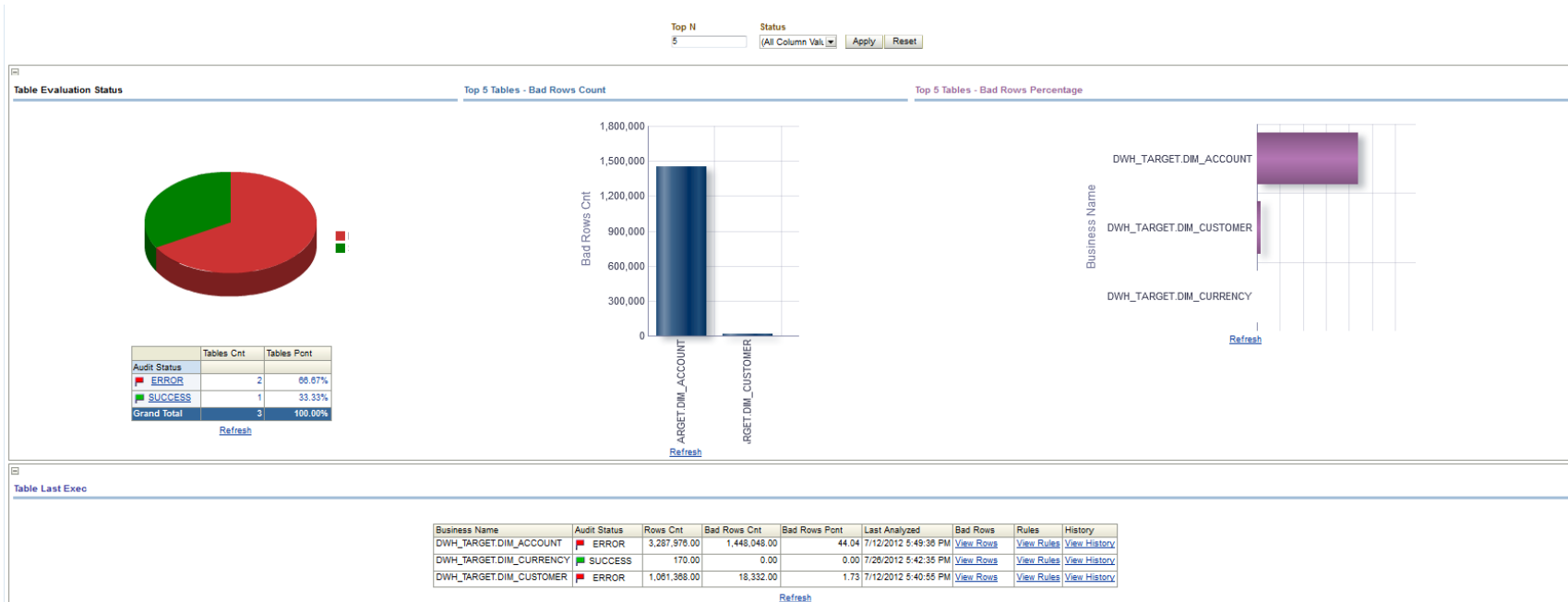
- Konsolidacija podataka
- Data Quality
- Praćenje povijesti
- Performanse
 - Skupi join-ovi (heterogeni izvori, transakcijske tablice)
 - Kompleksne “on-the-fly” mjere
 - Masovne agregacije, trendovi
- Opterećenje operativne baze

BI s DW-om: Mane

- Spora implementacija
- ETL vještine
- Cijena



Data Quality



Rules Last Exec

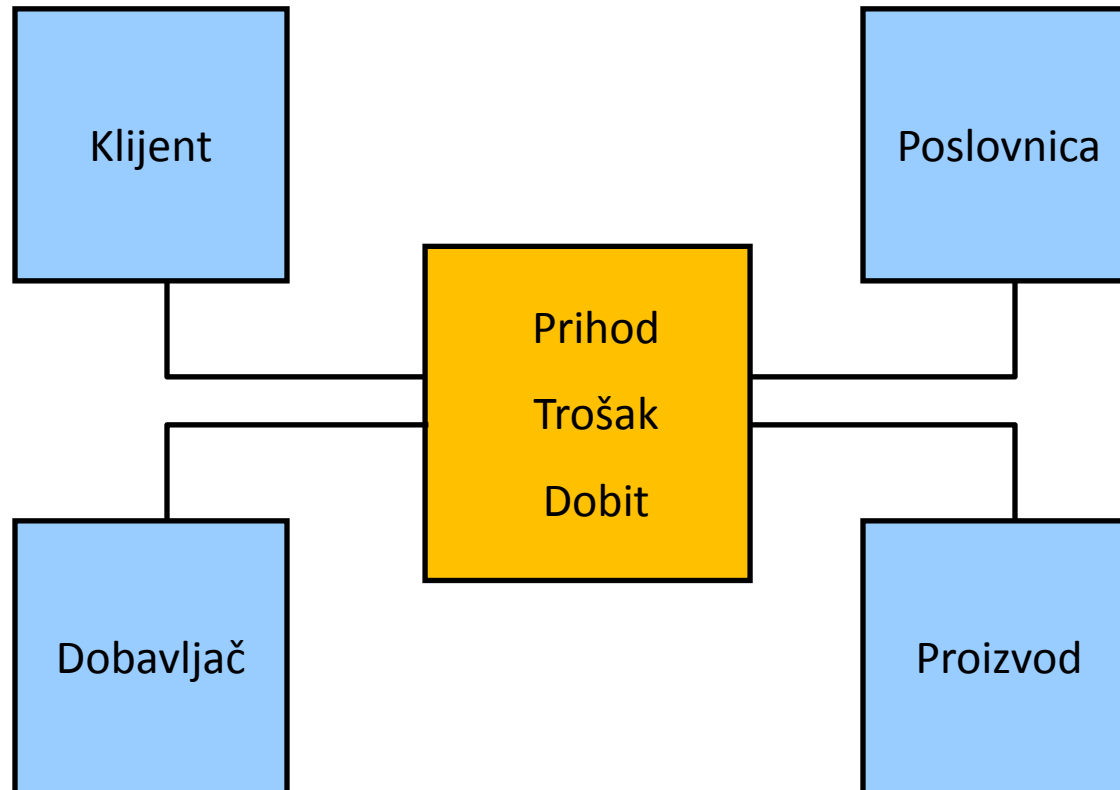
Rule Long Name	Rule Type	Audit Status	Bad Rows Cnt	Bad Rows Error Pont	Bad Rows Table Pont	Last Analyzed	Bad Rows	Definition	History
DIM_ACCOUNT: Dimenzijaki klijuš ACCNT_DIM_ACCOUNT_MANAGER ima nepostojeću vrijednost.	Custom	ERROR	969.00	0.07	0.03	12.07.2012 17:49:35	View Rows	View Definition	View History
DIM_ACCOUNT: Dimenzijaki klijuš ACCNT_DIM_ACCOUNT_TYPE ima nepostojeću vrijednost.	Custom	ERROR	31,331.00	2.18	0.95	12.07.2012 17:49:35	View Rows	View Definition	View History
DIM_ACCOUNT: Dimenzijaki klijuš ACCNT_DIM_CLOSE_DATE ima nepostojeću vrijednost.	Custom	ERROR	14.00	0.00	0.00	12.07.2012 17:49:35	View Rows	View Definition	View History
DIM_ACCOUNT: Dimenzijaki klijuš ACCNT_DIM_CURRENCY ima nepostojeću vrijednost.	Custom	SUCCESS	0.00	0.00	0.00	12.07.2012 17:49:35	View Rows	View Definition	View History
DIM_ACCOUNT: Dimenzijaki klijuš ACCNT_DIM_DATA_SOURCE ima nepostojeću vrijednost.	Custom	SUCCESS	0.00	0.00	0.00	12.07.2012 17:49:35	View Rows	View Definition	View History
DIM_ACCOUNT: Dimenzijaki klijuš ACCNT_DIM_DISBURSEMENT_DATE ima nepostojeću vrijednost.	Custom	SUCCESS	0.00	0.00	0.00	12.07.2012 17:49:35	View Rows	View Definition	View History
DIM_ACCOUNT: Dimenzijaki klijuš ACCNT_DIM_EXCHANGE_RATE_SOURCE ima nepostojeću vrijednost.	Custom	SUCCESS	0.00	0.00	0.00	12.07.2012 17:49:35	View Rows	View Definition	View History
DIM_ACCOUNT: Dimenzijaki klijuš ACCNT_DIM_EXCHANGE_RATE_TYPE ima nepostojeću vrijednost.	Custom	SUCCESS	0.00	0.00	0.00	12.07.2012 17:49:35	View Rows	View Definition	View History
DIM_ACCOUNT: Dimenzijaki klijuš ACCNT_DIM_FIRST_TRANSAC_DATE ima nepostojeću vrijednost.	Custom	SUCCESS	0.00	0.00	0.00	12.07.2012 17:49:35	View Rows	View Definition	View History
DIM_ACCOUNT: Dimenzijaki klijuš ACCNT_DIM_GL_ACCOUNT ima nepostojeću vrijednost.	Custom	SUCCESS	0.00	0.00	0.00	12.07.2012 17:49:35	View Rows	View Definition	View History
DIM_ACCOUNT: Dimenzijaki klijuš ACCNT_DIM_LAST_TRANSAC_DATE ima nepostojeću vrijednost.	Custom	SUCCESS	0.00	0.00	0.00	12.07.2012 17:49:35	View Rows	View Definition	View History
DIM_ACCOUNT: Dimenzijaki klijuš ACCNT_DIM_MATURITY ima nepostojeću vrijednost.	Custom	SUCCESS	0.00	0.00	0.00	12.07.2012 17:49:35	View Rows	View Definition	View History

Data Quality



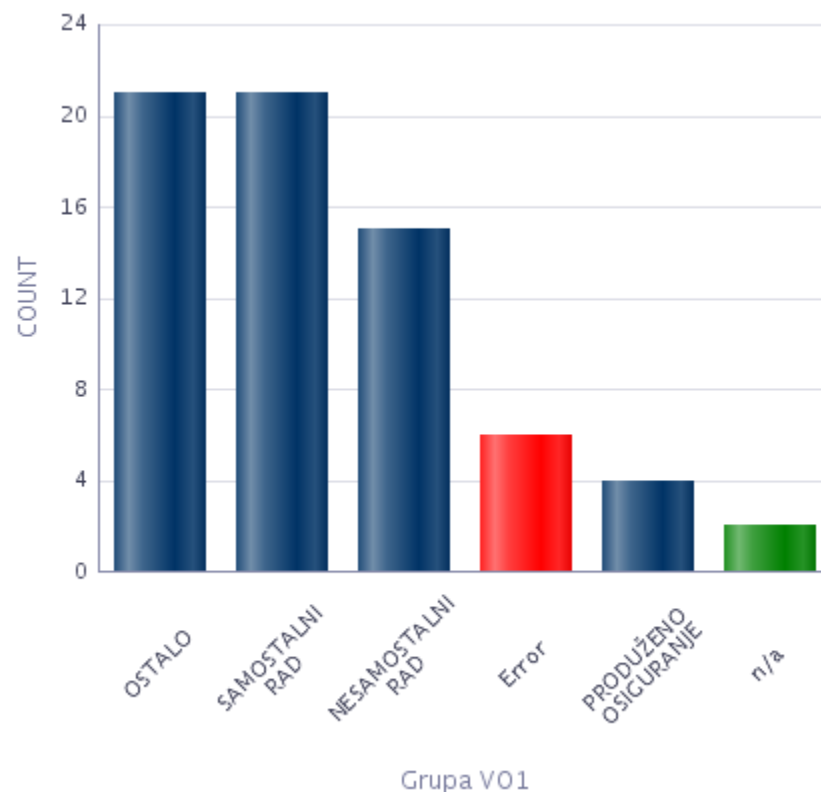
- Statistike grešaka, trend grešaka
- Putanje svrdlanja (Drill Paths):
 - Neispravni retci -> Podatkovna pravila
 - Podatkovna pravila -> Neispravni retci

Nepotpuni šifrarnici



Nepotpuni šifrarnici

- Dvije vrste “dummy” redaka:
 - **N/A**: Nema vrijednosti (null)
 - **Error**: Ima vrijednost koja ne postoji u referentnom šifrarniku
- Potpuna informacija o vezi prema šifrarnicima



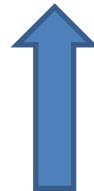
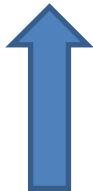
Praćenje povijesti

- Povijesno točni podaci
- SCD2

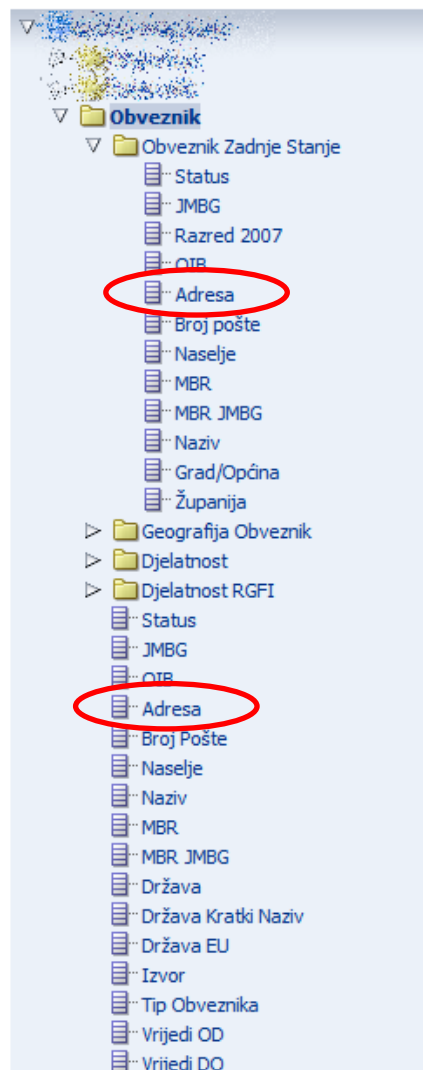
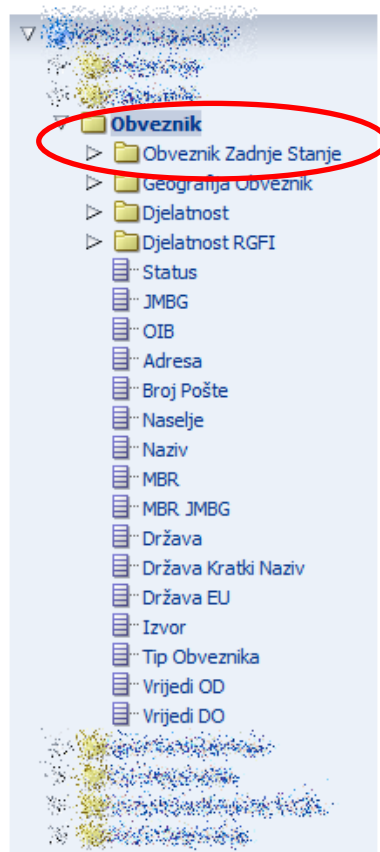
PROD_CODE	PROD_NAME	PROD_PRICE	PROD_CATEGORY	PROD_VALID_FROM	PROD_VALID_TO	CATG_CODE	CATG_NAME
1689	Mouse Pad	50	ACC	01.01.2013	31.03.2014	ACC	Accessories
1689	Mouse Pad	45	ACC	01.04.2014		ACC	Accessories
5009	32 GB Memory Card	160	CAM	01.06.2013	12.09.2013	Error	Error
5009	32 GB Memory Card	160	MMA	13.09.2013		Error	Error

Kad praćenje povijesti smeta...

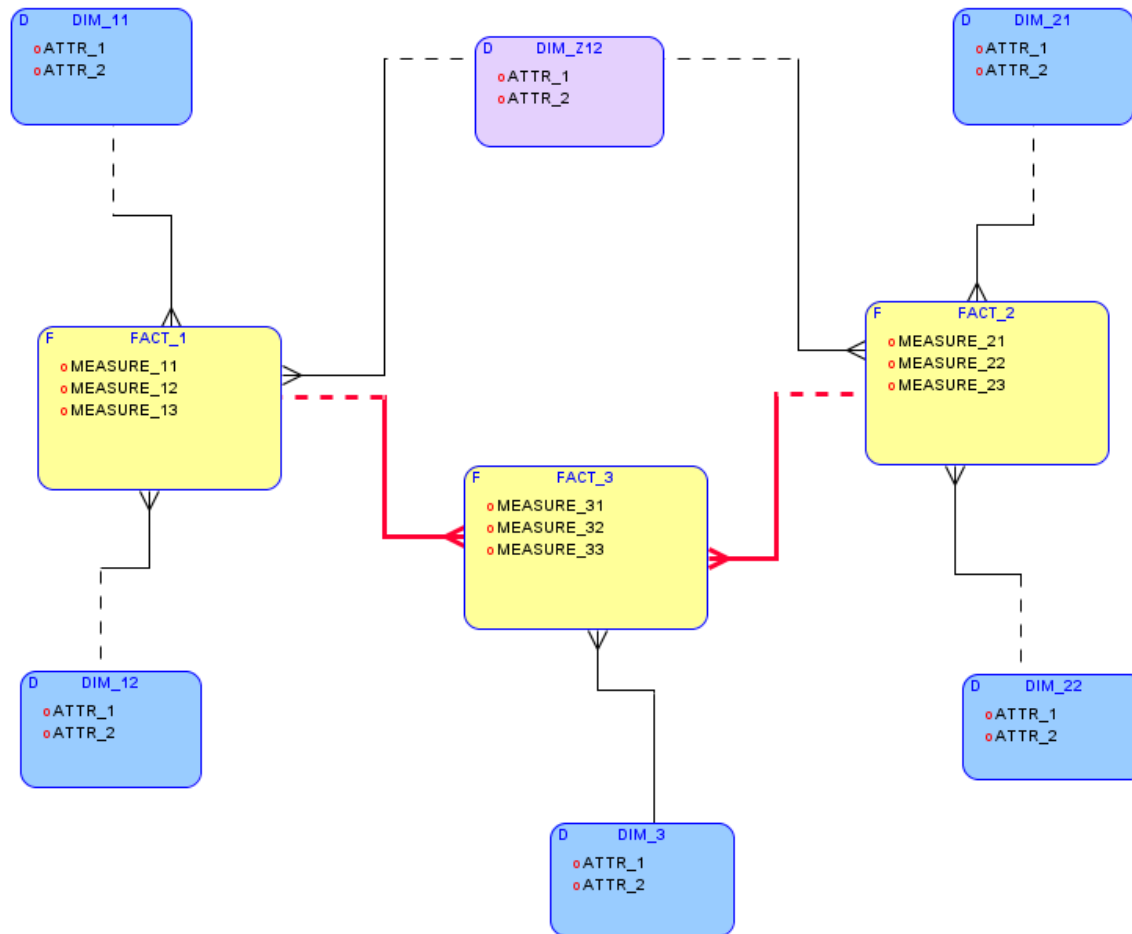
OIB	Naziv	Adresa	Naziv Zadnje Stanje	Adresa Zadnje Stanje	Naselje	Vrijedi OD	Vrijedi DO
81793146560	Hrvatski Telekom d.d.	Savska cesta 32	HT d.d.	Roberta Frangeša Mihanovića 9	ZAGREB	01.01.1900 00:00:00	23.02.2014 23:59:59
81793146560	HT d.d.	Savska cesta 32	HT d.d.	Roberta Frangeša Mihanovića 9	ZAGREB	24.02.2014 00:00:00	02.03.2014 23:59:59
81793146560	HT d.d.	Roberta Frangeša Mihanovića 9	HT d.d.	Roberta Frangeša Mihanovića 9	ZAGREB	03.03.2014 00:00:00	11.05.2014 23:59:59
81793146560	Hrvatski Telekom d.d.	Roberta Frangeša Mihanovića 9	HT d.d.	Roberta Frangeša Mihanovića 9	ZAGREB	12.05.2014 00:00:00	01.06.2014 23:59:59
81793146560	HT d.d.	Roberta Frangeša Mihanovića 9	HT d.d.	Roberta Frangeša Mihanovića 9	ZAGREB	02.06.2014 00:00:00	



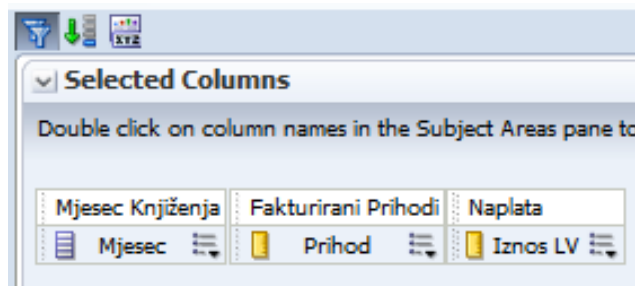
Kad praćenje povijesti smeta...



Povezani poslovni procesi



Povezani poslovni procesi

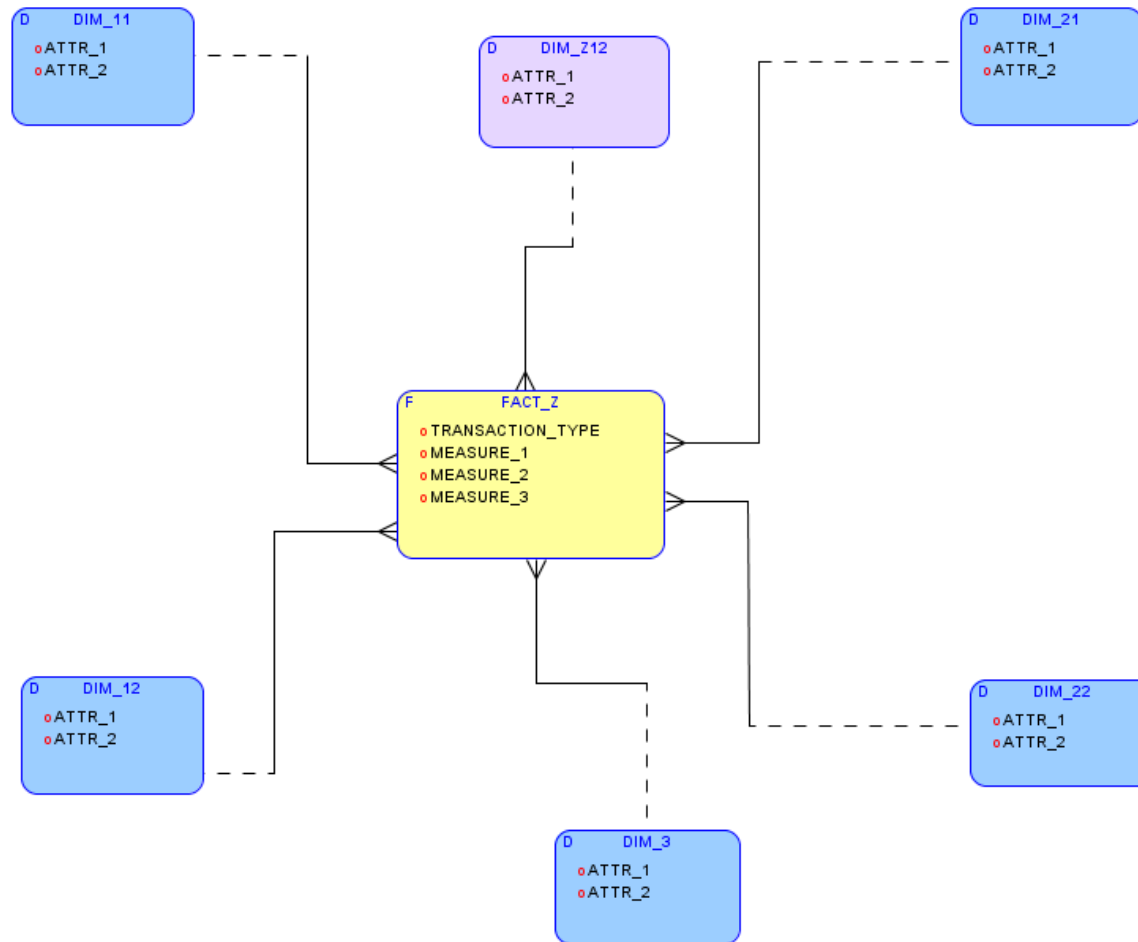


- Heterogene mjere na istom izvješću
- OBI 10g: UNION (ALL)
- OBI 11g: Multiple Subject Areas (Full Outer Join)
- OBI 10g/11g: Sve u istom SA
 - Nije Best Practice
 - Zaštita od multipliciranja iznosa (SQL Chasm & Fan Trap Detection)

Povezani poslovni procesi

- Više odvojenih upita čiji rezultati se spajaju
- Problem: Indirektne veze prema dimenzijama
 - Skupi join između velikih činjeničnih tablica
 - Loše performanse
- Nije uvijek intuitivno za BI korisnika

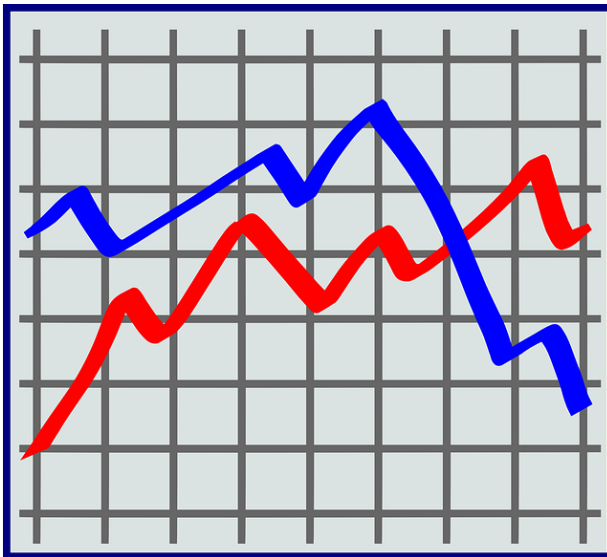
Objedinjene činjenice



Objedinjene činjenice

- Samo jedan upit
- Implicitni filter po tipu transakcije
 - OBI funkcija FILTER () ... USING ()
- Samo direktne veze prema dimenzijama
 - Materijaliziran join između velikih činjeničnih tablica
 - Nema opasnosti za SQL Chasm & Fan Trap
 - Dobre performanse
- Sve u istom SA
- Najjednostavnije za BI korisnika

Pametne mjere



- Ne postoje izvorno kao kolone
- Jednostavnost korištenja za BI korisnika
- Dva tipa:
 - BI-Only pametne mjere
 - DW/BI pametne mjere

BI-Only pametne mjere

- Kad SQL izraz:
 - Sadrži sve attribute iz iste tablice
 - Nema string operacija (REGEXP, LIKE, SUBSTR...)
 - Nema PL/SQL funkcija
- Bez dodatnih redaka u činjeničnoj tablici

AMOUNT	VERIFIED	BI mjera (User)	BI mjera (Admin)
123	N	Neverificirani Iznos	Iznos
456	Y	Verificirani Iznos	

DW/BI pametne mjere

- Kad SQL izraz sadrži:
 - Attribute iz više tablica
 - String operacije
 - PL/SQL funkcije (!)
- Materijalizacija u obliku dodatnih redaka, a ne kolona
- Jednostavna agregacija (MV)
- Brzina

DIM_MEASURE	
P * MEASURE_CODE	VARCHAR2 (30)
MEASURE_NAME	VARCHAR2 (200)
MEASURE_EXPRESSION	VARCHAR2 (2000)
MEASURE_DESCRIPTION	VARCHAR2 (2000)
DIM_MEASURE_PK (MEASURE_CODE)	

MEASURE_CODE	MEASURE_NAME
A000	Duguje/Potražuje
A001	Naplaćena Izravno
A002	Nedospjela Kamata
A003	Obračunata redovna Kamata
A004	Obračunata zatezna Kamata
A005	Obračunata naknada Kredita
A006	Prihodovni diskont Zatezna
A007	Obračunata Kamata Kamatni Prihodi
A008	Obračunata Naknada po osnovu Eskonta Mjenica
A009	Obračunata Garancija
A010	Obračunata Kamata Nekamatni Prihodi
A011	Broj Transakcija
A012	Iznos Priljeva Sredstava
A013	Iznos Priljeva Sredstava dom.
A014	Iznos Odljeva Sredstava
A015	Iznos Odljeva Sredstava dom.
A016	Iznos Plasmana
A017	Iznos Plasmana dom.
A018	Iznos Naplaćenih Naknada
A019	Iznos Naplaćenih Naknada dom.
A020	Broj Opomena
A021	Broj Storniranih Opomena
A022	Naplaćeni Iznos Opomena

OBI implementacija

- Za oba tipa pametnih mjera:
 - Pivotiranje (obavezno) -> CASE WHEN ...
 - Implicitno filtriranje (opcionalno) -> WHERE ...
- Bez implicitnog filtera
 - Upit vraća retke bez obzira da li je vrijednost mjere (izraza) NULL
 - Loše performanse
- S implicitnim filterom
 - Upit vraća samo one retke u kojima barem jedna od uključenih mjera (izraza) nije NULL
 - Dobre performanse

Trendovi

- Prema periodu:
 - Godišnji
 - Polugodišnji
 - Kvartalni
 - Mjesečni
 - Tjedni
 - Dnevni
- Prema vrsti iznosa:
 - Iznos u prethodnom periodu
 - Apsolutni iznos promjene
 - Relativni iznos promjene (%)

Prethodni period

Period/Mjera	MAGO (MoM)	QAGO (QoQ)	YAGO (YoY)
2014-M09	2014-M08	2014-M06	2013-M09
2014-Q3	-	2014-Q2	2013-Q3
2014	-	-	2013

- Svaka mjera ima svoje AGO članove
- SUM, COUNT, AVG, MIN, MAX
- Podržano:
 - Dodavanje novih dimenzija u izvješće
 - Svrđlanje po hijerarhiji

Prethodni period



Prethodni period

Objekt	AVG Radna Snaga (MW)	AVG Radna Snaga (MW) MAGO	% Prom AVG Radna Snaga (MW) MAGO	
CS Buško Blato			-85,63	↓
CSBB-Agr_1			-4,71	↓
CSBB-Agr_2			-135,14	↑
CSBB-Agr_3			-94,88	↓
HE Đale			-13,09	↓
HE Kraljevac			1,36	↑
HE Orlovac			27,64	↑
HE Peruća			-41,71	↓
HE Zakučac			-20,99	↓
HEĐ-Agr_1			-23,55	↑
HEĐ-Agr_2			-12,59	↓
HEK-ABM			0,56	↑
HEK-Agr_1			0,00	▬
HEK-Agr_2			-100,00	↓
HEO-Agr_1			0,03	↑
HEO-Agr_2			1,02	↑
HEO-Agr_3			13,81	↑
HEP-Agr_1			-2,69	↓
HEP-Agr_2			-1,52	↓
HEZ-Agr_1			-30,59	↓
HEZ-Agr_2			-6,98	↓
HEZ-Agr_3			-18,54	↓
HEZ-Agr_4			-4,76	↓
Sliv rijeke Cetine			-8,94	↓

To Date trendovi

Mjera/Mjera	MAGO	QAGO	YAGO
MTD (01.09.2014 – 15.09.2014)	01.08.2014 – 15.08.2014	01.06.2014 – 15.06.2014	01.09.2013 – 15.09.2013
QTD (01.07.2014 – 15.09.2014)	-	01.04.2014 – 15.06.2014	01.07.2013 – 15.09.2013
YTD (1.1.2014 – 15.09.2014)	-	-	01.01.2013 – 15.09.2013

- Primjer za datum: 15.09.2014
- Svaka mjera ima svoje xTD članove
- Svaka xTD mjera ima svoje AGO članove

OBI implementacija

- OBI funkcije Ago () i ToDate ()
 - Brzina implementacije
 - Loše performanse
- Višestruki LTS (Logical Table Sources)
 - Veći napor implementacije
 - Dobre performanse
- OBI Calculation Wizard

Zaključak

- BI korisnik želi:
 - Velike analitičke mogućnosti
 - Da bude jednostavno
 - Da bude brzo
 - Vjerovati dobivenim informacijama

Zaključak

- Da bismo mu to omogućili
 - Pojednostavnjena slika stvarnosti
 - Puno truda i znanja oko izgradnje DW/BI modela i ETL-a
- Ne tjerajte poslovnog korisnika da razmišlja kao informatičar!

...Hvala!



Q&A