



Optimizacija skladišta podataka korištenjem Oracle OLAP 11g

Josip Pojatina

Agrokor ICT (mStart telekomunikacije d.o.o.)

josip.pojatina@agrokor.hr



Sadržaj

- O tvrtki mStart telekomunikacije
- OLAP – što je to?
- Povijest Oracle OLAP-a
- OLAP – tipična područja primjene
- Optimizacija skladišta podataka korištenjem Oracle OLAP 11g
- OLAP DML programski jezik
- Demo
- Pitanja i odgovori

O tvrtki mStart telekomunikacije

- Agrokor ICT 1.7.2010. promijenio naziv u mStart telekomunikacije
- Djeluje kao samostalni subjekt unutar Agrokor koncerna s ciljem pružanja podrške za svih 36 kompanija unutar Agrokor grupacije

Ulje i margarini						
Sladoledi i zamrznuta hrana						
Vode i pića						
Meso i mesne prerađevine						
Poljoprivreda						
Maloprodaja, veleprodaja i distribucija						
Ostale djelatnosti						



OLAP – što je to?

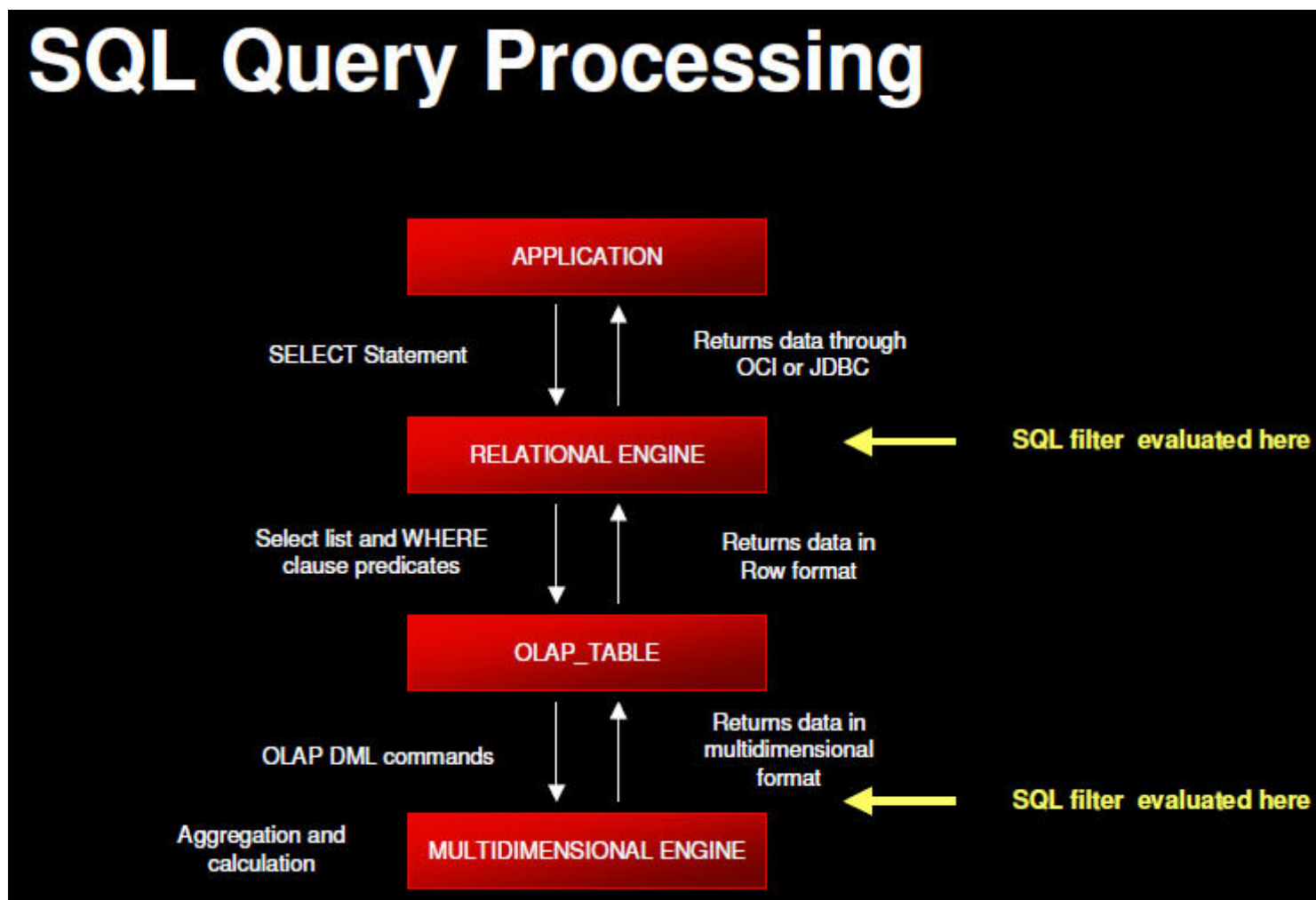
- OLAP je akronim za Online Analytical Processing (Dr. E.F. Codd, 1993)
- Točnija definicija bi bila FASMI (Fast Analysis of Shared Multidimensional information)
- Multidimensionalnost je osnovno obilježje OLAP-a
- Podaci su nisu spremljeni u tablicama već u složenim multidimenzijskim strukturama (BLOB)
- Cilj OLAP-a je brza analiza složenih multidimenzijskih podataka, izrada složenih kalkulacija te jednostavna interpretacija rezultata
- OLAP obuhvaća najbolje osobine klasičnih DW izvještaja kao i statističke analize i data mining-a



Povijest Oracle OLAP-a

- Početak 1970. pod nazivom Express kao prva multidimenzijaska baza na tržištu
- 1995. Oracle kupuje Express s ciljem integracije u Oracle bazu
- Prva integracija Express-a 2002.-2003. s Oracle 9i
- 2004.-2006. S pojavom 10g baze, Oracle OLAP dobiva nove funkcionalnosti (kompresija podataka, OLAP_TABLE funkcija – integracija SQL i OLAP engine-a...)

Povijest Oracle OLAP-a



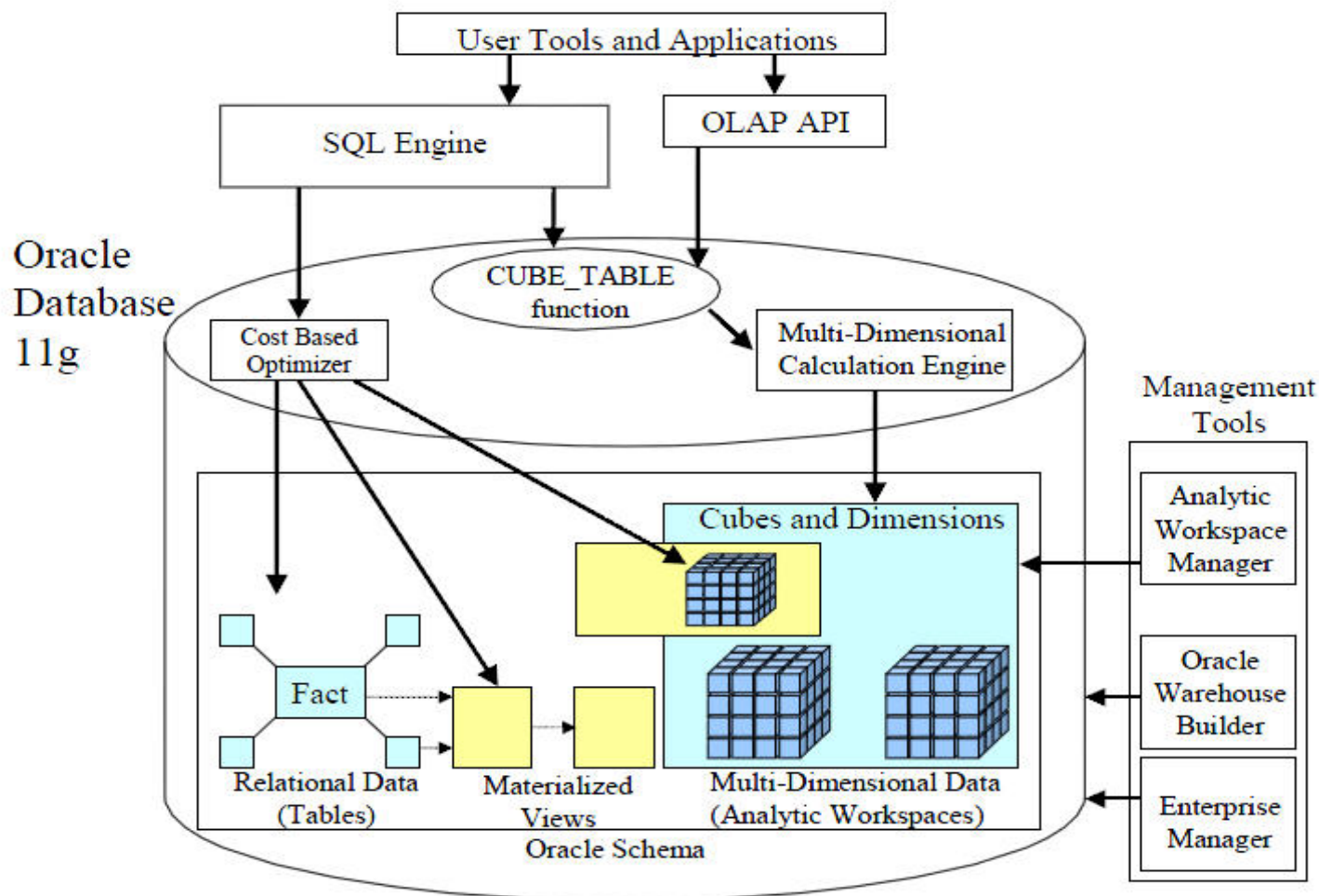


Povijest Oracle OLAP-a

- Dolaskom 11g baze (2007-2009) Oracle je nakon 15 godina završio proces integracije Express-a u Oracle bazu
- CBO je sposoban napraviti query rewrite ne samo nad klasičnim MV-ima, već i nad OLAP kockama (Cube Organized Materialized Views)

Povijest Oracle OLAP-a

Figure 2 Relational and Multidimensional Data in Oracle Database 11g





OLAP – tipična područja primjene

- Svugdje gdje se traži napredna analiza (time-series, non-additive calculations, financial calculations, statistical forecasting, modelling...)
- Primjeri primjene OLAP-a:
 - Prodaja i marketing – proračun profitabilnosti proizvoda, profiliranje kupaca, analiza distribucije, performance/efikasnost prodaje, praćenje prodaje, analiza promocija, predviđanje prodaje...
 - Odjel financija - financijskom modeliranje, upravljanje i praćenje novčanih tokova, buđetiranje, konsolidacija...



OLAP – tipična područja primjene

- Primjeri primjene OLAP-a:
 - Odjel plana i analize – analiza investicija/akvizicija/reorganizacija, dugoročno planiranje, alokacija resursa, planiranje kapaciteta, planiranje ljudskih resursa...
 - Rukovodstvo kompanije – analiza trendova, praćenje ključnih pokazatelja, analiza konkurentnosti
- OLAP nudi mogućnost napredne analize podataka korištenjem value-setova, dimenzija, hijerarhija, relacija i naprednih kalkulacija



Optimizacija DW-a korištenjem OLAP-a 11g

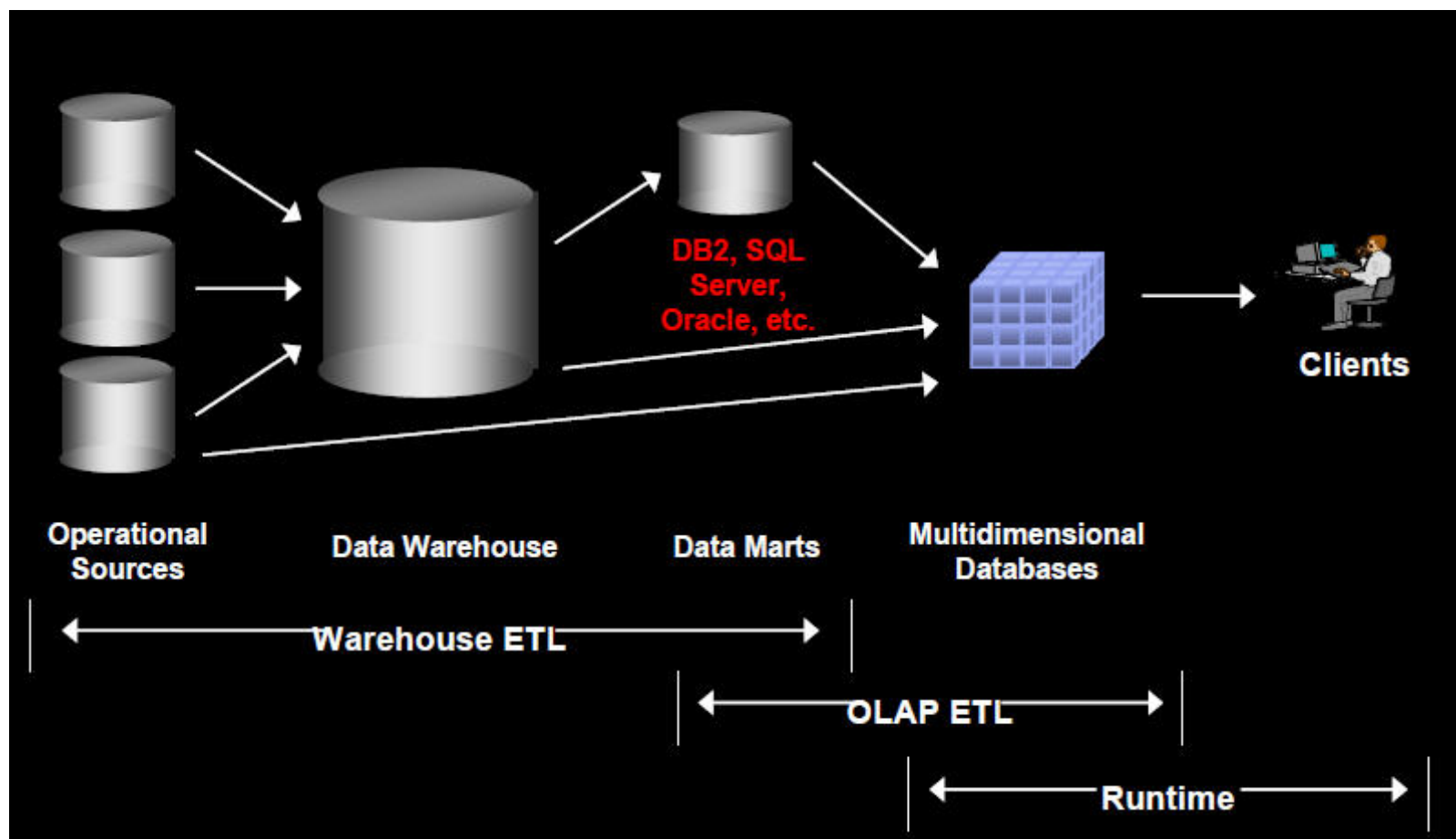
- Izrada sigurnosnih kopija multidimenzijskih podataka (backup)
 - Backup OLAP-a se ne razlikuje od standardnog backup-a uz korištenje istih alata (rman). Najčešće je backup OLAP-a uključen u standardni backup baze.
- Zaštita OLAP podataka
 - identična postavljanju zaštite nad relacijskim podacima, čime je izbjegnuto učenje postavljanja zaštite na OLAP serveru.
 - Zaštita se postavlja na jednom (DW baza) a ne dva mjesta (DW baza i OLAP server). Posljedica je jednostavnost upravljanja i smanjeno vrijeme administriranja sigurnosnim postavkama



Optimizacija DW-a korištenjem OLAP-a 11g

- Pojednostavnjivanje ETL-a i punjenja OLAP kocki.
 - Izbjegnuta je izrada OLAP ETL procedure punjenja OLAP kocki.
Za osvježavanje OLAP-a dovoljno je dodati jednu naredbu u PL/SQL-u (`dbms_cube.refresh_mview`).
 - Ubrzan je cijeli ETL jer nema seljenja podataka između dva servera

Optimizacija DW-a korištenjem OLAP-a 11g



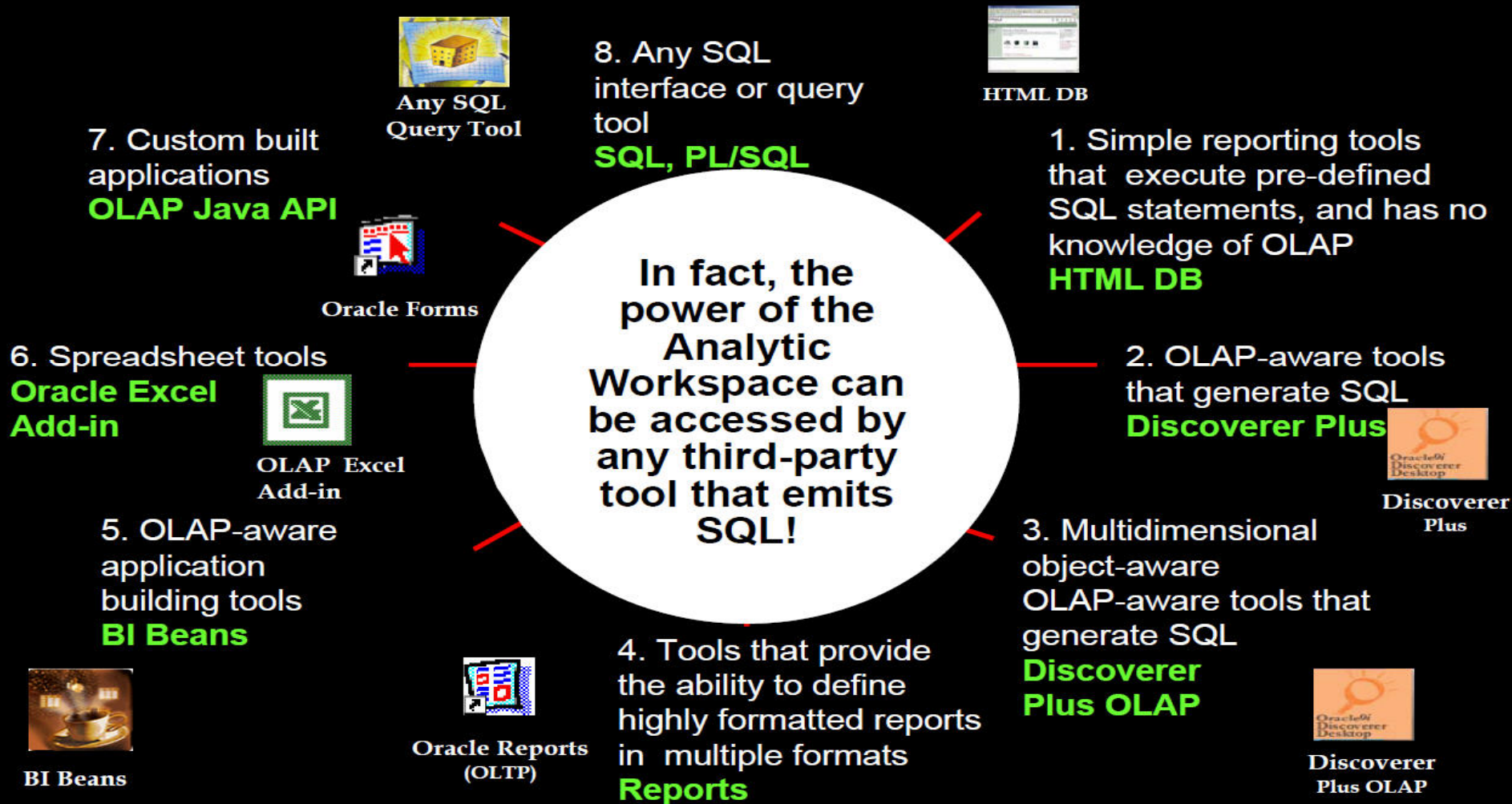
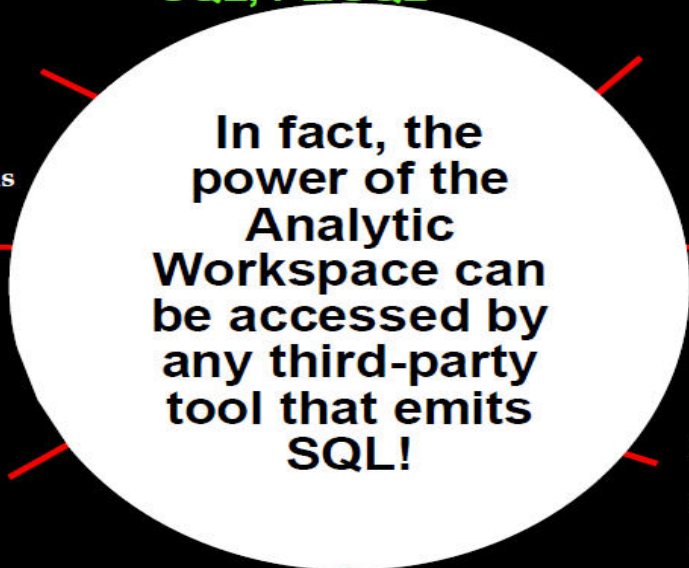


Optimizacija DW-a korištenjem OLAP-a 11g

- Smanjenje hardverskih zahtjeva.
 - Kako je Oracle OLAP dio Oracle baze, nalazi se na istom serveru. Time je izbjegnuta kupovina posebnog servera za potrebe samostojećeg (stand alone) OLAP servera.
- Mogućnost pristupa OLAP podacima direktno iz SQL-a
 - Prednost OLAP tehnologije će osjetiti sve aplikacije koje mogu pristupiti Oracle bazi (BI Publisher, Oracle Reports Builder, Cognos, APEX, TOAD, SQL*Plus...)

Optimizacija DW-a korištenjem OLAP-a 11g

Architecture – Open access





Optimizacija DW-a korištenjem OLAP-a 11g

- Korištenje kompleksnih kalkulacija u SQL-u kao posljedica direktnog pristupa OLAP podacima

Kalkulacije koje bi bilo nemoguće i/ili neefikasno napraviti u okvirima SQL-a, sada postaju dostupne korištenjem izračunatih mjera (calculated measures) OLAP kocki.

- Pojednostavnjivanje SQL upita kao posljedica direktnog pristupa OLAP podacima iz SQL-a



Optimizacija DW-a korištenjem OLAP-a 11g

```
SELECT [breakout columns],
       sales,
       sales_prior_year
       sales_ytd,
       sales_ytd_prior_year
FROM sales_cube_view
WHERE [star join]
```

```
WITH sales_dense AS
(SELECT [breakout columns]
  sales,
  SUM(sales) over(PARTITION BY [breakout columns]
ORDER BY [time column] ASC range BETWEEN unbounded
           preceding AND CURRENT ROW) AS sales_ytd
FROM
  (SELECT [breakout columns]
   a.sales
  FROM
   (SELECT [breakout columns]
    SUM(f.sales) sales
   FROM [table list]
   WHERE [star join and other filters]
   GROUP BY [breakout columns])
  a PARTITION BY(breakout columns)
  RIGHT OUTER JOIN
   (-- need list of all time periods
   SELECT DISTINCT [time columns]
   FROM time_dim
   b ON([join on relevant time level]))
  ) ...
Continued...
```

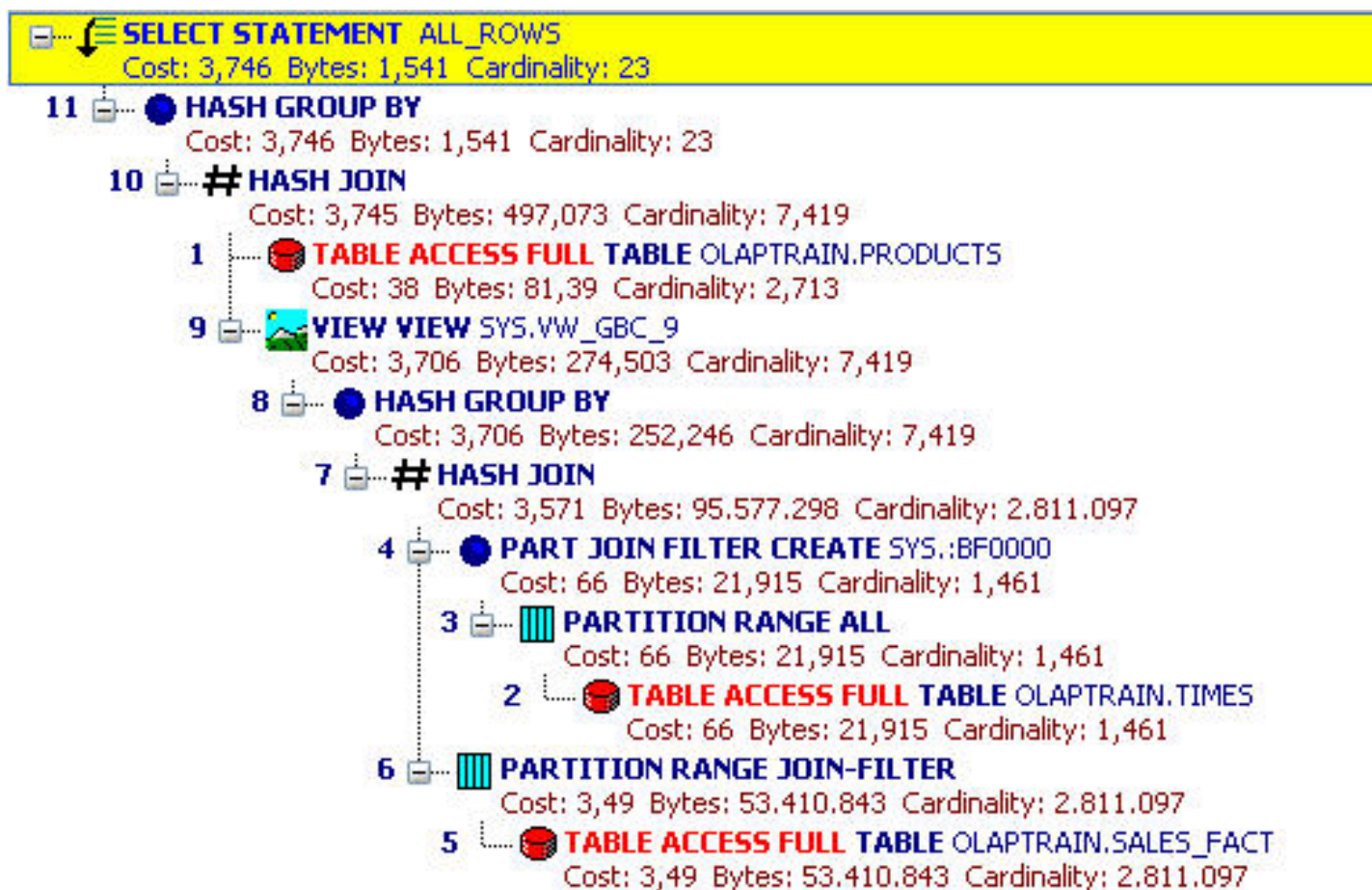


Optimizacija DW-a korištenjem OLAP-a 11g

- Cube Organized Materialized Views
 - Radi se o izlaganju OLAP kocke CBO-u kao da se radi o klasičnoj sumarnoj tablici (MV – Materialized View).
 - Na slijedeća dva slide-a je prikazan klasičan plan izvršavanje upita, kao i plan izvršavanja kad se koristi cube query rewrite.

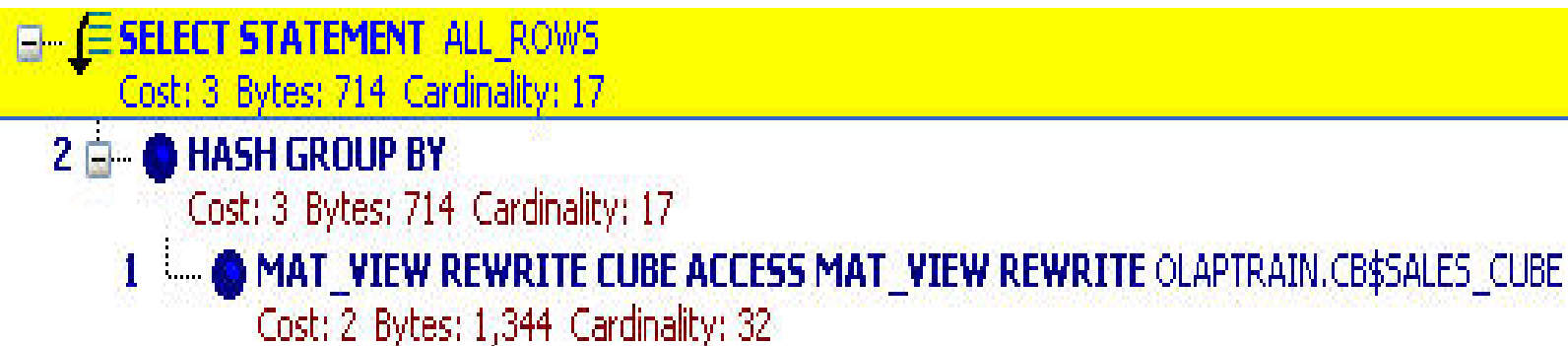


Optimizacija DW-a korištenjem OLAP-a 11g





Optimizacija DW-a korištenjem OLAP-a 11g





Optimizacija DW-a korištenjem OLAP-a 11g

- Veliki broj kombinacija u kojima dolazi do cube query rewrite-a
 - Mana klasičnih MV-ova je limitiran broj SQL upita u kojima dolazi do query rewrite-a. Kako OLAP kocke imaju veći broj dimenzija (5 - 7), dok su podaci agregirani na različitim nivoima, broj kombinacija u kojima dolazi do query rewrite-a je neusporedivo veći.
 - Jedna OLAP kocka zamjenjuje desetine MV-ova, zbog čega se ubrzavaju većina drugih upita, a ne samo oni najkritičniji.



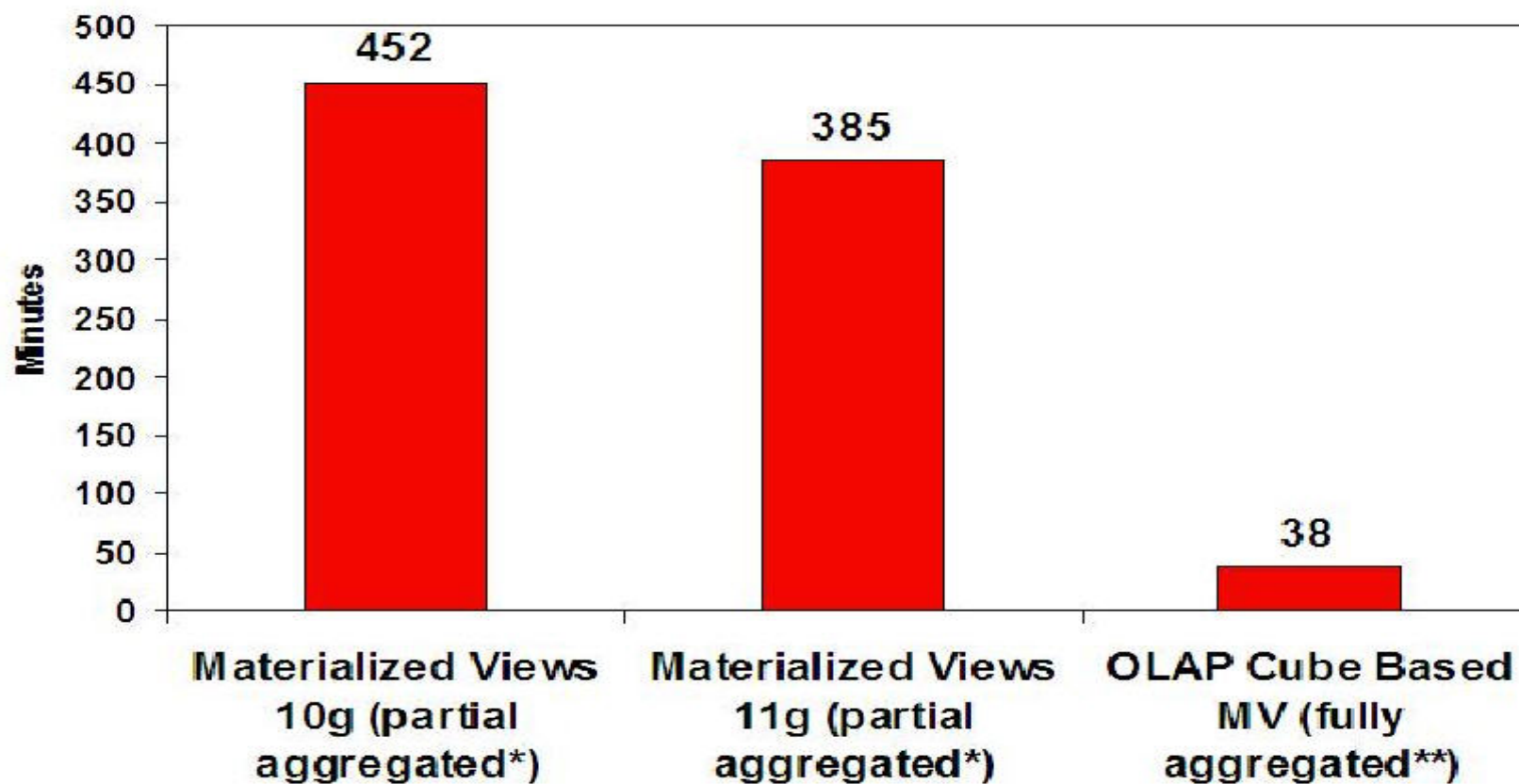
Optimizacija DW-a korištenjem OLAP-a 11g

- Smanjeno zauzeće mjesta na diskovima
 - Sumarne tablice (MV) zauzimaju diskovni prostor. Kako jedna OLAP kocka zamjenjuje desetine klasičnih MV-ova, zahtjevi za diskovnim prostorom su neusporedivo manji.
- Smanjeno vrijeme i jednostavnost osvježavanja
 - Dok je održavanje klasičnih sumarnih tablica (pogotovo kad postoji međuovisnost) zahtjevno (vrijeme + resursi), osvježavanje OLAP kocki je bitno brže.



Optimizacija DW-a korištenjem OLAP-a 11g

The graph below shows the results for loading and aggregating relational views and cube based materialized views for the 10g and 11g database benchmarks:



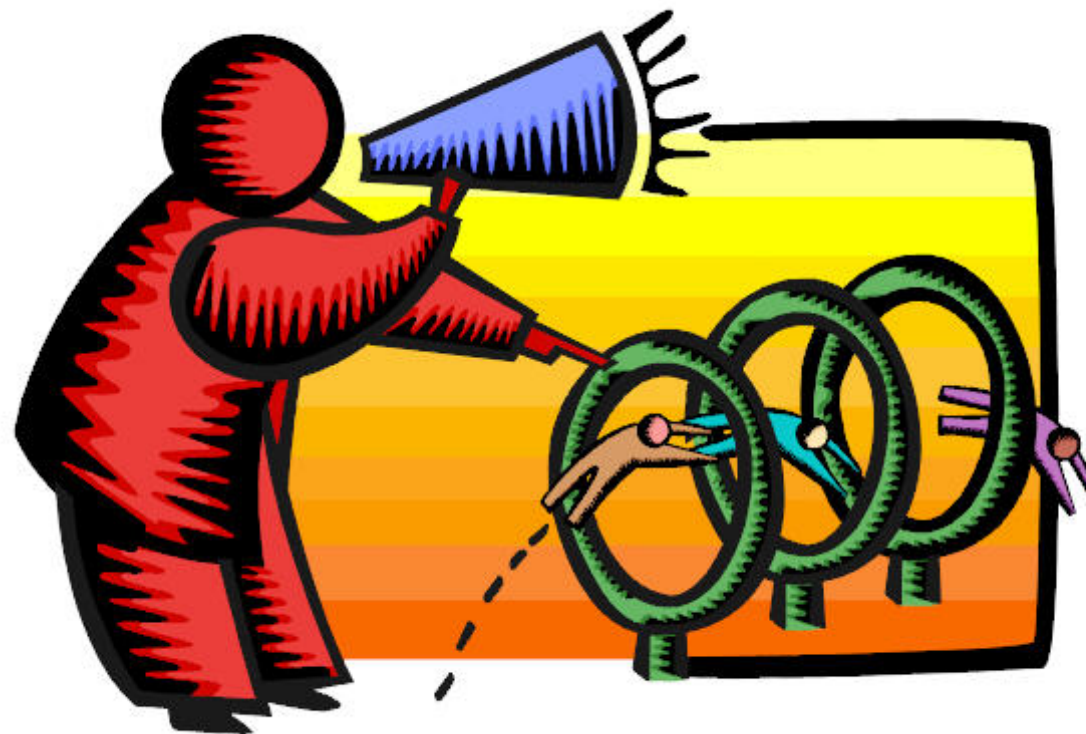
- As shown above, build times were significantly faster when using OLAP 11g. Compared to 10g and improved 11g relational materialized views, a single CPU cube built in 38 minutes (without using available parallel processing options). Other parallel tests performed the same task in only 10 to 15 minutes. By comparison, it took at least 11x longer to build partially aggregated relational materialized views on the same database server instance.



Optimizacija DW-a korištenjem OLAP-a 11g

- OLAP DML programski jezik
 - Od samih početaka poznat po performansama i mogućnostima
 - Reference Manual za OLAP DML veći i od SQL Reference Manual-a
 - DML u nazivu nema veze s pojmom DML u SQL jeziku
 - Dizajniran za izvođenje analitika nad multidimenzionalnim podacima
 - Dostupan iz OLAP Worksheet-a, PL/SQL-a i Jave
 - Proširuje mogućnostima AWM (Analytic Workspace Manager)

Demo





Optimizacija DW-a korištenjem OLAP-a 11g

