Business Intelligence

- Comprehensive Platform
- OLAP
- Data Mining
## Oracle 10g for Data Warehousing and Business Intelligence

*A comprehensive platform*

| Industry-leading scalability and performance | Oracle Partitioning  
|                                            | Oracle Real Application Clusters |
| Deep integrated analytics                   | Oracle OLAP  
|                                            | Oracle Data Mining                  |
| Embedded data quality and data integration  | Oracle Warehouse Builder  
|                                            | Enterprise ETL  
|                                            | Oracle Warehouse Builder Data Quality  
|                                            | Oracle Warehouse Builder Connectors |
Why Online Analytic Processing?

• “Does your organization use SQL-based business intelligence applications such as BusinessObjects, MicroStrategy, Cognos ReportNet or Oracle Business Intelligence Enterprise Edition?”

• Would business users benefit from:
  • **significantly improved query performance?**
  • **ability to explore data sets with rich calculations, rather than being confined to predefined reports?**

• Would IT benefit from:
  • **fast, efficient updates of data sets?**
  • **fewer servers to manager?**
What is Oracle OLAP Option?

- A multidimensional calculation and aggregation ‘engine’ embedded in the database
  - multidimensional data types (Cubes and Dimensions)
  - OLAP API for cube definition and multidimensional query

and coming in Oracle Database 11g ....
- transparent SQL interface to OLAP Cubes and dimensions
- cube-organized materialized views
- OLAP metadata in the Oracle Data Dictionary
Oracle OLAP 11g

New Features in Oracle Database 11g

• Data accessibility: Enhanced SQL access to OLAP cubes
  • SQL Optimizer awareness of OLAP cubes
  • Cube Organized Materialized Views
  • System-maintained generation of cube and dimension views

• Feature accessibility: Shift the focus from the skilled OLAP specialist to the database generalist
  • Cube and dimensions in the Oracle Data Dictionary
  • Cost-based aggregation
  • Dimensionally aware SQL-like calculation syntax
Cube-Organized Materialized Views

• Materialized view with data stored in an OLAP cube
  • Automatic query write of summary queries into the cube for improving performance of summary queries
  • Materialized view managed refresh of cubes and dimensions
  • Similar to materialized view on pre-built table; data is stored in cube and MV is a meta data object
Cube-Organized Materialized Views

- Detail data is stored in relational tables
- Application queries relational tables with SQL
- Summary data is managed by OLAP cube
- Access to summary data occurs via automatic query rewrite to the cube
- Application is unchanged, but updates and queries are faster
select a13.FAMILY_ID FAMILY_ID,
a13.FAMILY_DSC FAMILY_DSC,
a12.CALENDAR_YEAR_ID CALENDAR_YEAR_ID,
a12.CALENDAR_YEAR_DSC CALENDAR_YEAR_DSC,
sum(a11.SALES) WJXBFS1
from UNITS_FACT a11
join TIME_DIM a12
  on (a11.MONTH_ID = a12.MONTH_ID)
join PRODUCT_DIM a13
  on (a11.ITEM_ID = a13.ITEM_ID)
group by a13.FAMILY_ID,
a13.FAMILY_DSC,
a12.CALENDAR_YEAR_ID,
a12.CALENDAR_YEAR_DSC;
Query Rewrite to Cube

OPERATION
---------------------------------------
SELECT STATEMENT
HASH
HASH JOIN
HASH JOIN
   CUBE SCAN CB$UNITS_CUBE
   VIEW
      HASH
      TABLE ACCESS PRODUCT_DIM
   VIEW
      HASH
      TABLE ACCESS TIME_DIM
Why Oracle OLAP?

• Compared to Relational Aggregates
  • faster to create
  • efficient storage for large numbers of dimensions
  • can benefit unknown aggregates
  • faster to maintain
• Object and data security for OLAP cubes
  • managed using Oracle database
• Fully compatible with:
  • Real Application Clusters and Grid Computing
• Cubes are easily queried using SQL
Why Data Mining?

“Do you wish to ..

Identify factors most associated with a business problem?  
(Attribute Importance)

Predict customer behavior?  
(Classification)

Predict or estimate a value?  
(Regression)

Find profiles of targeted people or items?  
(Decision Trees)

Segment a population?  
(Clustering)

Determine important relationships/“market baskets” within the population?  
(Associations)

Find fraudulent or “rare events”?  
(Anomaly Detection)
What is Oracle Data Mining?

- In-Database Mining Engine
  - wide range of algorithms
    - anomaly detection
    - attribute importance
    - association rules
    - clustering
    - classification & regression
    - nonnegative matrix factorization
    - structured & unstructured data (text mining)
    - BLAST (life sciences similarity search algorithm)
- Oracle Data Miner (GUI)
  - simplified, guided data mining
- Spreadsheet Add-In for Predictive Analytics
  - “1-click data mining” from a spreadsheet
- PL/SQL API & Java (JDM) API
  - develop advanced analytical applications
Oracle Data Mining 11g
New Features in Oracle Database 11g

• Simplified development and deployment of models
  • supermodels: data preparation combined with mining model
  • tighter database integration – DMSYS schema eliminated; improved sharing of objects, security, manageability

• New algorithms: “General Linear Models”
  • multivariate linear regression; logistic regression
    • basic tool for microeconomics analysis and business decision-making. Indicate the impact of different attributes on a key performance indicator.
      “What is the impact on SUVs sales if the price of gas goes up by a $1.00?”
Why In-Database Analytics?

- **Fewer moving parts**: data remains in the database at all times … with appropriate access security control mechanisms
- **Straightforward inclusion within interesting and arbitrarily complex queries**
- **Real-world scalability and performance**
  - Fast scoring:
    - 2.5 million records scored in 6 seconds on a single CPU system
  - Real-time scoring:
    - 100 models on a single CPU: 0.085 seconds
Integration with Oracle BI EE

ODM provides likelihood of fraud and other important questions.
Key Differentiators for Database Options

- **Best-of-breed** capabilities for OLAP, Data Mining, and ETL
- **Integrated** into the Oracle Database
  - A true BI consolidation platform: eliminates the costs and inefficiencies of multiple data stores
  - All analytics available via SQL
  - Enterprise-level scalability
  - Enterprise-level security
  - Enterprise-level availability
  - Optimized for the grid