Transform Big Data into Bigger Insight with Oracle Exadata and Oracle Advanced Analytics

Charlie Berger, Senior Director, Product Mgt. OAA
Marcos Arancibia, Product Manager, OAA
Michael Bramley, Science Director R&D, dunnhumby
Oracle Big Data Solution Architecture

Stream
- Oracle Event Processing
- Apache Flume
- Oracle GoldenGate

Acquire – Organize – Analyze
- Cloudera Hadoop
  - Oracle NoSQL Database
  - Oracle R Distribution
- Oracle Big Data Connectors
- Oracle Data Integrator
- Oracle Database
- Oracle Advanced Analytics
- Oracle Spatial & Graph

Decide
- Oracle Real-Time Decisions
- Endeca Information Discovery
- Oracle BI Foundation Suite
Oracle In-Database Analytics

- Statistical Functions
- Data Mining & Predictive Analytics
- Text Mining
- Text Search
- Graph Analysis
- Spatial Analysis
- Semantic Analysis
- In-Database MapReduce
Oracle Advanced Analytics
Fastest Way to Deliver Scalable Enterprise-wide Predictive Analytics

Key Features

- In-database data mining algorithms and open source R algorithms
- SQL, PL/SQL, R languages
- Scalable, parallel in-database execution
- Workflow GUI and IDEs
- Integrated component of Database
- Enables enterprise analytical applications
Oracle Advanced Analytics
Performance and Scalability with Low Total Cost of Ownership

Data remains in the Database
- Scalable, parallel Data Mining algorithms in SQL kernel
- Efficient execution of R open-source packages with in-database data preparation
- High-performance parallel scoring of Data Mining and R open-source models

Fastest path from data to insights
- Integrated GUI for Predictive Analytics
- Database scoring engine

Lowest TCO
- Eliminate data duplication
- Eliminate separate analytical servers
Oracle Advanced Analytics Architecture

Oracle Database Enterprise Edition

Oracle Advanced Analytics
Native SQL-PL/SQL Analytic Libraries plus high-performance R interface
Scalable, Distributed, Parallel Execution

Oracle R Distribution
Oracle Advanced Analytics Architecture

Oracle Database Enterprise Edition

Oracle Advanced Analytics
Native SQL-PL/SQL Analytic Libraries plus high-performance R interface
Scalable, Distributed, Parallel Execution

Oracle R Distribution
## Oracle Advanced Analytics

### In-Database Data Mining Algorithms

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classification</strong></td>
<td></td>
</tr>
<tr>
<td>Logistic Regression (GLM)</td>
<td>Classical statistical technique</td>
</tr>
<tr>
<td>Decision Trees</td>
<td>Popular / Rules / transparency</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>Embedded app</td>
</tr>
<tr>
<td>Support Vector Machines (SVM)</td>
<td>Wide / narrow data / text</td>
</tr>
<tr>
<td><strong>Regression</strong></td>
<td></td>
</tr>
<tr>
<td>Linear Regression (GLM)</td>
<td>Classical statistical technique</td>
</tr>
<tr>
<td>Support Vector Machine (SVM)</td>
<td>Wide / narrow data / text</td>
</tr>
<tr>
<td><strong>Anomaly Detection</strong></td>
<td></td>
</tr>
<tr>
<td>One Class SVM</td>
<td>Unknown fraud cases or anomalies</td>
</tr>
<tr>
<td><strong>Attribute Importance</strong></td>
<td></td>
</tr>
<tr>
<td>Minimum Description Length (MDL)</td>
<td>Attribute reduction, Reduce data noise</td>
</tr>
<tr>
<td>Principal Components Analysis (PCA)</td>
<td></td>
</tr>
<tr>
<td><strong>Association Rules</strong></td>
<td></td>
</tr>
<tr>
<td>Apriori</td>
<td>Market basket analysis / Next Best Offer</td>
</tr>
<tr>
<td><strong>Clustering</strong></td>
<td></td>
</tr>
<tr>
<td>Hierarchical k-Means</td>
<td>Product grouping / Text mining</td>
</tr>
<tr>
<td>Hierarchical O-Cluster</td>
<td>Gene and protein analysis</td>
</tr>
<tr>
<td>Expectation-Maximization Clustering (EM)</td>
<td></td>
</tr>
<tr>
<td><strong>Feature Extraction</strong></td>
<td></td>
</tr>
<tr>
<td>Nonnegative Matrix Factorization (NMF)</td>
<td>Text analysis / Feature reduction</td>
</tr>
<tr>
<td>Singular Value Decomposition (SVD)</td>
<td></td>
</tr>
</tbody>
</table>
Oracle Advanced Analytics

Wide Range of In-Database Data Mining and Statistical Functions

- Data Understanding & Visualization
  - Summary & Descriptive Statistics
  - Cross tabulations
  - Tests for Correlations (t-test, Pearson’s, ANOVA)
  - Histograms, scatter plots, box plots, bar charts
  - R graphics: 3-D plots, link plots, special R graph types
  - Selected Base SAS equivalents

- Data Selection, Preparation & Transformations
  - Joins, Tables, Views, Data Selection, Data Filter,
  - Join multiple databases
  - Select, Filter, Rank,
  - SQL time windows,
  - Sample
  - Re-coding, Missing values
  - Aggregations
  - Spatial data
  - R to SQL transparency and push down

- In-Database Algorithms
  - Classification Models
  - Regression Models
  - Clustering
  - Anomaly Detection
  - Associations / Market Basket Analysis
  - Text Mining
    - Most OAA algorithms support unstructured data (i.e. customer comments, email, abstracts, etc.)

- R Integration:
  - Additional custom Oracle R packages with algorithms that run against Database and Hadoop (like Neural Networks and Stepwise Regression)
  - Open-source R packages—ability to run open source R CRAN packages

* included in every Oracle Database
OAA SQL DM Fraud Example

begin
  dbms_data_mining.create_model('CLAIMSMODEL', 'CLASSIFICATION', 'CLAIMS', 'POLICYNUMBER', null, 'CLAIMS_SET');
end;
/

-- Top 5 most suspicious fraud policy holder claims
select
  POLICYNUMBER,
  round(prediction_probability('CLAIMSMODEL', '0' using *)*100,2) prob_fraud
from
  CLAIMS
where
  PASTNUMBEROFCLAIMS in ('2to4', 'morethan4')
order by
  prob_fraud desc
fetch first 5 rows only;

<table>
<thead>
<tr>
<th>POLICYNUMBER</th>
<th>PERCENT_FRAUD</th>
<th>RNK</th>
</tr>
</thead>
<tbody>
<tr>
<td>6532</td>
<td>64.78</td>
<td>1</td>
</tr>
<tr>
<td>2749</td>
<td>64.17</td>
<td>2</td>
</tr>
<tr>
<td>3440</td>
<td>63.22</td>
<td>3</td>
</tr>
<tr>
<td>654</td>
<td>63.1</td>
<td>4</td>
</tr>
<tr>
<td>12650</td>
<td>62.36</td>
<td>5</td>
</tr>
</tbody>
</table>

For Automated Monthly “Application”! Just add:
Create
View CLAIMS2_30
As
Select * from CLAIMS2
Where mydate > SYSDATE – 30
Oracle Advanced Analytics Architecture

Oracle Database Enterprise Edition

Oracle Advanced Analytics
Native SQL-PL/SQL Analytic Libraries plus high-performance R interface
Scalable, Distributed, Parallel Execution

Oracle R Distribution
Oracle Data Miner GUI

SQL Developer 4.0 Extension
Free OTN Download

- **Easy to Use**
  - Oracle Data Miner GUI for data analysts
  - "Work flow" paradigm
- **Powerful**
  - Multiple algorithms & data transformations
  - Runs 100% in-DB
  - Build, evaluate and apply models
- **Automate and Deploy**
  - Save and share analytical workflows
  - Generate SQL scripts for deployment
Oracle Advanced Analytics Architecture

Oracle Database Enterprise Edition

Oracle Advanced Analytics
Native SQL-PL/SQL Analytic Libraries plus high-performance R interface
Scalable, Distributed, Parallel Execution

Oracle R Distribution
Integration through SQL and R

- All predictions, insights and models are in the Database—any BI tool can access and query using SQL
- OBIEE’s integrated spatial mapping can be used to Map predictions
- OBIEE dashboards can launch parameterized R calculations that can return data or visualizations
- Any BI tool or application that supports SQL can take advantage

Customer “most likely” to be HIGH and VERY HIGH value customer in the future

Advanced R Statistical graphic output directly in the Dashboard
Integration through SQL and R

- All predictions, insights and models are in the Database—any BI tool can access and query using SQL
- OBIEE’s integrated spatial mapping can be used to Map predictions
- OBIEE dashboards can launch parameterized R calculations that can return data or visualizations
- Any BI tool or application that supports SQL can take advantage
Oracle Advanced Analytics Architecture

Oracle Database Enterprise Edition

Oracle Advanced Analytics
Native SQL-PL/SQL Analytic Libraries plus high-performance R interface
Scalable, Distributed, Parallel Execution

Oracle R Distribution
Enabling Predictive Applications
Example Oracle Applications Using Oracle Advanced Analytics

- **HCM Fusion**
  - **Predictive Workforce**—employee turnover and performance prediction and “What if?” analysis

- **CRM Fusion**
  - **Sales Prediction Engine**—prediction of sales opportunities, what to sell, amount, timing, etc.

- **Supply Chain Management**
  - **Spend Classification**—real-time flagging of noncompliance and anomalies in expense submissions

- **Identity Management**
  - **Oracle Adaptive Access Manager**—real-time security and fraud analytics

- **Industry Data Models**
  - **Communications Data Model** implements churn prediction, segmentation, profiling, etc.
  - **Retail Data Model** implements loyalty and market basket analysis
  - **Airline Data Model** implements analysis frequent flyers, loyalty, etc.

- **Oracle Fin. Services Analytic Applications**
  - **Customer Insight, Enterprise Risk Management**
  - **Enterprise Performance, Financial Crime and Compliance**

- **OFSAA CI Retail Customer Analytics**
  - **Attrition Analysis**—Mortgage Prepay, Savings Account Attrition, Term Deposit, Cards…
  - **Survival analysis**
  - **Customer Lifetime value**
  - **Propensity Models**—Credit Cards <-> Auto loans, Savings <-> Cards

- **Retail Analytics**
  - **Oracle Retail Customer Analytics**—“shopping cart analysis” and next best offers

- **Customer Support**
  - **Predictive Incident Monitoring (PIM)** Customer Service offering for Database customers
Oracle Communications Industry Data Model

Pre-Built Predictive Models

- Fastest Way to Deliver Scalable Enterprise-wide Predictive Analytics
- OAA’s clustering and predictions available in-DB for OBIEE
- Automatic Customer Segmentation, Churn Predictions, and Sentiment Analysis
OCDM Telco Churn Enhanced by SNA Analysis

Social Network Analysis of Large Volumes of CDR Data

- Integrated with OCDM, OBIEE, and leverages Oracle Data Mining with specialized SNA code
- Identification of social network communities
- Predictive scores for churn and influence at a node level, as well as potential revenue/value at risk
- User interface targeted at business users and flexible ad-hoc reporting
Fusion HCM Predictive Workforce

Fusion Human Capital Management Powered by OAA

- Oracle Advanced Analytics factory-installed predictive analytics
- Employees likely to leave,
- Top reasons, expected performance
- Real-time "What if?" analysis
Oracle Advanced Analytics Architecture

Oracle Database Enterprise Edition

Oracle Advanced Analytics
Native SQL-PL/SQL Analytic Libraries plus high-performance R interface
Scalable, Distributed, Parallel Execution
Why statisticians/data analysts use R
R is a statistics language similar to Base SAS or SPSS statistics

The R environment is ..

- Powerful
- Extensible
- Graphical
- Extensive statistics
- OOTB functionality with many ‘knobs’ but smart defaults
- Ease to install and use
- Free
Oracle Strategy for R

Provide high-performance, scalable R environment tightly integrated with Oracle RDBMS and Hadoop

For R users

- Full access to Database and HDFS objects
- High performance and scalability for all R operations
- Scalable, Natively integrated machine learning algorithms
- Deploy R scripts and store R calculation results in Database or Hadoop

For Database & Big Data developers

- Execute embedded R scripts containing any R algorithm or calculation
- Access stored R results in Database or Hadoop
- Retrieve R computation results in graphical formats like XML or PNG
- Integrate R results into BI Applications
Oracle Advanced Analytics: Database Integration
Using the in-Database Integration and Open-Source R Packages

Client Interfaces

R Client Interface
Oracle R Enterprise packages:
- Transparency
- Embedded R

SQL Interfaces
- Any SQL & PL/SQL
- New “SQL Query Node” in ODM GUI

Oracle Database Server

Oracle Database
Advanced Analytics Option
SQL Basic Statistics
Data Mining algorithms
Registered R Scripts called via SQL

Oracle Database Server
Parallel ExtProc Interconnect

Oracle R Distribution
- Enhanced linear algebra performance
- Parallel distributed analytic techniques that leverage R language constructs
- Custom R algorithms: Neural/Stepwise
- Access to open-source R packages
Oracle Advanced Analytics: Hadoop Integration

Using the Hadoop-HDFS Integration, Custom and Open-Source R Packages

Client Interfaces

R Client Interface
- Oracle R Connector for Hadoop packages:
  - Hadoop
  - MapReduce
  - HIVE Transparency Layer
- Oracle R Enterprise packages:
  - Transparency
  - Embedded R

Oracle Database
- Advanced Analytics Option
- Oracle R Distribution

Hadoop Cluster

Oracle R Connector for Hadoop
- Translation of R requests to Hadoop:
  - $\Sigma(x)$: HDFS Utilities: Data Movement and Statistics, pushing data to R, Data Sampling
  - $f(x)$: ORCH Utilities: Connect/Disconnect R Sessions
  - $\partial(x)$: HIVE Interfaces: Load table metadata and interface
  - ORCH Custom R algorithms: Neural, GLM, kMeans, NMF, LMF
- Custom R Analytics are written once for a Mapper & Reducer framework, and are reused as is. I/O is then built for both the Database and Hadoop

Oracle Database

Parallel MapReduce Calls

HDFS engine

Hadoop Cluster

Oracle R Connector for Hadoop packages:
- Hadoop
- MapReduce
- HIVE Transparency Layer

Oracle R Enterprise packages:
- Transparency
- Embedded R

R Client Interface

R, Java
Oracle Advanced Analytics

Summary New Features

- **Oracle Advanced Analytics 12c**
  - New SQL data mining algorithms (Expectation Maximization, PCA, Singular Vector Decomposition, Text Mining and other algorithm improvements)
  - Predictive SQL Queries—automatic build, apply within SQL query

- **Oracle Data Miner/SQL Developer 4.0** *(for Oracle Database 11g and 12c)*
  - New Graph node (box, scatter, bar, histograms)
  - SQL Query node + integration of R scripts
  - Automatic SQL script generation for deployment

- **Oracle R Enterprise 1.4** *(for Oracle Database 11g and 12c)*
  - Parallelized Neural Networks with ore.neural() against Database data
  - Scoring Database tables with open-source R Models; in-Database Sampling
  - Support for Date and Time data types for Time Series Analysis
  - Persist and Manage R Objects in-Database; Improved integration with OBIEE
More Information on OAA

▪ **Google**: “Oracle Advanced Analytics”

▪ **Oracle Demo Campgrounds Demo Pod**
  - OOW Exhibit Hall Hours (Mon-Wed) Moscone South, Left,
    ▪ Workstation ID: SL-063, Database, Data Warehousing

▪ **OAA Hands on Labs:**
  - Big Data, Bigger Insights with Oracle Advanced Analytics and Oracle SQL Developer [HOL10074]
    ▪ Monday, Sep 23, 3:15 PM - 4:15 PM - Marriott Marquis - Salon 3/4
  - Make the Right Offers to Customers Using Oracle Advanced Analytics [HOL10075]
    ▪ Tuesday, Sep 24, 10:30 AM - 11:30 AM - Marriott Marquis - Salon 3/4
Hardware and Software
Engineered to Work Together