Oracle Advanced Analytics 12c & SQLDEV/Oracle Data Miner 4.0
New Features

Charlie Berger, MS Eng, MBA
Sr. Director Product Management, Data Mining and Advanced Analytics
charlie.berger@oracle.com    www.twitter.com/CharlieDataMine
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 Evolution

1998

- 7 Data Mining “Partners”
- Oracle acquires Thinking Machine Corp’s dev. team + “Darwin” data mining software

1999

- Oracle Data Mining 9.2i launched – 2 algorithms (NB and AR) via Java API

2002

- Oracle Data Mining 10g & 10gR2 introduces SQL dm functions, 7 new SQL dm algorithms and new Oracle Data Miner “Classic” wizards driven GUI

2004

- Oracle Data Mining 11g & 11gR2 adds AutoDataPrep (ADP), text mining, perf. improvements

2005

- ODM 11g & 11gR2 adds AutoDataPrep (ADP), text mining, perf. improvements

2008

- SQLDEV/Oracle Data Miner 4.0 “work flow” GUI launched with SQL script generation and SQL Query node (R integration)
- OAA/ORE 1.3 + 1.4 launched adding several new scalable R algorithms
- Oracle Adv. Analytics for Hadoop Connector launched with scalable BDA algorithms

2011

- Product renamed “Oracle Advanced Analytics (ODM + ORE)
- Integration with “R” and introduction/addition of Oracle R Enterprise
- Integration with “R” and introduction/addition of Oracle R Enterprise

2015

- New algorithms (EM, PCA, SVD)
- SQLDEV/Oracle Data Miner 4.0 “work flow” GUI launched with SQL script generation and SQL Query node (R integration)
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
- Fast parallelized native SQL data mining functions, SQL data preparation and efficient execution of R open-source packages
- High-performance parallel scoring of SQL data mining functions and R open-source models

Fastest way to deliver enterprise-wide predictive analytics
- Integrated GUI for Predictive Analytics
- Database scoring engine

Lowest TCO
- Eliminate data duplication
- Eliminate separate analytical servers
- Leverage investment in Oracle IT
More Data Variety—Better Predictive Models

- Increasing sources of relevant data can boost model accuracy

Model with “Big Data” and hundreds -- thousands of input variables including:
- Demographic data
- Purchase POS transactional data
- “Unstructured data”, text & comments
- Spatial location data
- Long term vs. recent historical behavior
- Web visits
- Sensor data
- etc.

0% 100% 100% 100%
Population Size

Responders

Naïve Guess or Random
Model with 20 variables
Model with 250 variables
Model with 75 variables
Oracle Advanced Analytics Architecture

Oracle Database Enterprise Edition

Oracle Advanced Analytics
Native SQL Data Mining/Analytic Functions + High-performance R Integration for Scalable, Distributed, Parallel Execution
Fusion HCM Predictive Workforce
Predictive Analytics Applications

Fusion Human Capital Management Powered by OAA

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

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 (Database) Option

Oracle Data Miner 4.0 Summary New Features

- **Oracle Data Miner/SQLDEV 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 Advanced Analytics 12c features exposed in Oracle Data Miner**
  - New SQL data mining algorithms/enhancements
    - Expectation Maximization clustering algorithm
    - PCA & Singular Vector Decomposition algorithms
    - Improved/automated Text Mining, Prediction Details and other algorithm improvements
  - Predictive SQL Queries—automatic build, apply within SQL query
SQL Developer/Oracle Data Miner 4.0

New Features

- **Graph node**
  - Scatter, line, bar, box plots, histograms
  - `Group_by` supported
SQL Developer/Oracle Data Miner 4.0

New Features

- **SQL Query node**
  - Allows any form of query/transformation/statistics within an ODM'r work flow
  - Use SQL anywhere to handle special/unique data manipulation use cases
    - Recency, Frequency, Monetary (RFM)
    - SQL Window functions for e.g. moving average of $$ checks written past 3 months vs. past 3 days
  - Allows integration of R Scripts
SQL Developer/Oracle Data Miner 4.0

New Features

- **SQL Script Generation**
  - Deploy entire methodology as a SQL script
  - Immediate deployment of data analyst’s methodologies
SQL Developer/Oracle Data Miner 4.0

New Features

- **SQL Query node**
  - Allows integration of R Scripts
SQL Developer/Oracle Data Miner 4.0

New Features

- **SQL Query node**
  - Allows integration of R Scripts
SQL Developer/Oracle Data Miner 4.0

New Features

- **Database/Data Mining Parallelism On/Off Control**
  - Allows users to take full advantage of Oracle parallelism/scalability on an Oracle Data Miner node by node basis
    - Default is “Off”
  - Important for large Oracle Database & Oracle Exadata shops
3 New Oracle Data Mining SQL functions algorithms

- **Expectation Maximization (EM) Clustering**
  - New Clustering Technique
    - Probabilistic clustering algorithm that creates a density model of the data
    - Improved approach for data originating in different domains (for example, sales transactions and customer demographics, or structured data and text or other unstructured data)
    - Automatically determines the optimal number of clusters needed to model the data.

- **Principal Components Analysis (PCA)**
  - Data Reduction & improved modeling capability
    - Based on SVD, powerful feature extraction method use orthogonal linear projections to capture the underlying variance of the data

- **Singular Value Decomposition (SVD)**
  - Big data “workhorse” technique for matrix operations
    - Scales well to very large data sizes (both rows and attributes) for very large numerical data sets (e.g. sensor data, text, etc.)
Text Mining Support Enhancements

- This enhancement greatly simplifies the data mining process (model build, deployment and scoring) when text data is present in the input:
  - Manual pre-processing of text data is no longer needed.
  - No text index needs to be created
  - Additional data types are supported: CLOB, BLOB, BFILE
  - Character data can be specified as either categorical values or text
12c New Features

New Server Functionality

- **Predictive Queries**
  - Immediate build/apply of ODM models in SQL query
  - Classification & regression
    - Multi-target *(nested)* problems
  - Clustering query
  - Anomaly query
  - Feature extraction query

OAA automatically creates multiple anomaly detection models “Grouped_By” and “scores” by partition via powerful SQL query

```
Select
cust_income_level, cust_id,
round(probanom,2) probanom, round(pctrank,3)*100 pctrank from ( 
  select 
cust_id, cust_income_level, probanom,
  percent_rank()
  over (partition by cust_income_level order by probanom desc)
pctrank
from ( 
  select 
cust_id, cust_income_level,
prediction_probability(of anomaly, 0 using *)
  over (partition by cust_income_level) probanom
  from customers
) 
) 
where pctrank <= .05 
order by cust_income_level, probanom desc;
```
12c New Features
New Server Functionality

- Predictive Queries
  - **Immediate** build/apply of ODM models in SQL query
    - Classification & regression
      - Multi-target (nested) problems
    - Clustering query
    - Anomaly query
    - Feature extraction query

OAA automatically creates multiple anomaly detection models “Grouped_By” and “scores” by partition via powerful SQL query
Oracle Advanced Analytics Previews
Oracle Data Miner 4.1 & OAA 12.2 Preview

- **Oracle Data Miner/SQLDEV 4.0**
  - New JSON Query node to work with Data Source node

- **Oracle Advanced Analytics 12.2c**
  - Re-engineering all OAA/ODM algorithms for maximum scalability & parallelism
  - Analytical components—available for faster/tighter R integration for selected R algorithms
    - Enables direct uptake into Oracle Data Miner GUI
      - Model viewers, advanced options settings, etc.
  - Partition-By modeling
Oracle Advanced Analytics Option

Links and Resources

- Oracle Advanced Analytics Overview:
  - Link to presentation—*Big Data Analytics using Oracle Advanced Analytics In-Database Option*
  - [OAA data sheet](#) on OTN
  - [Oracle Internal OAA Product Management Wiki and Workspace](#)

- YouTube recorded OAA Presentations and Demos:
  - [Oracle Advanced Analytics and Data Mining at the YouTube Movies](#) (6 + OAA Demos on Retail, Fraud, Loyalty, Overview, etc.)

- Getting Started:
  - Link to [Getting Started w/ ODM blog entry](#)
  - Link to [New OAA/Oracle Data Mining 2-Day Instructor Led Oracle University course](#).
  - Link to [OAA/Oracle Data Mining 4.0 Oracle by Examples (free) Tutorials](#) on OTN
  - Take a [Free Test Drive of Oracle Advanced Analytics (Oracle Data Miner GUI) on the Amazon Cloud](#) (Vlamis Partner)
  - Link to [SQL Developer Days Virtual Event w/ downloadable Virtual Machine (VM) images of Oracle Database + ODM/ODMr and e-training for Hands on Labs](#)
  - Link to [OAA/Oracle R Enterprise (free) Tutorial Series](#) on OTN

- Additional Resources:
  - [Oracle Advanced Analytics Option on OTN](#) page
  - [OAA/Oracle Data Mining on OTN](#) page, [ODM Documentation & ODM Blog](#)
  - [OAA/Oracle R Enterprise page on OTN](#) page, [ORE Documentation & ORE Blog](#)
  - [Oracle SQL based Basic Statistical functions](#) on OTN
  - Business Intelligence, Warehousing & Analytics—[BIWA Summit 2014](#), Jan 14-16 at Oracle HQ Conference Center