Oracle’s Advanced Analytics 12.2c
New Features & Road Map
Bigger, Better, Faster, More!

Charlie Berger, MS Engineering, MBA
Sr. Director Product Management,
Advanced Analytics and Machine Learning
charlie.berger@oracle.com
www.twitter.com/CharlieDataMine
Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle’s products remains at the sole discretion of Oracle.
Predictive Analytics 101

• Data, data everywhere – explosive growth
• Growth of data exponentially greater than growth of data analysts!

The Useful Data GAP

12%

Executives who feel they understand the impact data will have on their organizations

Produce Data

Use Data

Machine Learning/Data Analysis platforms requirements:

• Be extremely powerful and handle large data volumes
• Be easy to learn
• Be highly automated & enable deployment

http://www.delphianalytics.net/more-data-than-analysts-the-real-big-data-problem/
Machine Learning/Analytics + Data Warehouse + Hadoop

- Platform Sprawl
  - More Duplicated Data
  - More Data Movement Latency
  - More Security challenges
  - More Duplicated Storage
  - More Duplicated Backups
  - More Duplicated Systems
  - More Space and Power
Vision

• Big Data + Machine Learning/Analytics Platform for the Era of Big Data and Cloud
  
  – Make Big Data + **ML/Analytics** Model Discovery *Simple*
    • Any data size, on any computer infrastructure—on-premise and/or cloud
    • Any variety of data (structured, unstructured, transactional, geospatial), in any combination

  – Make Big Data + **ML/Analytics** Model Deployment *Simple*
    • As a service, as a platform, as an application
    • On-premise and/or cloud
What is Machine Learning, Data Mining & Predictive Analytics?

Automatically sifting through large amounts of data to create models that find previously hidden patterns, discover valuable new insights and make predictions

- Identify most important factor (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)
- Find fraudulent or “rare events” (Anomaly Detection)
- Determine co-occurring items in a “baskets” (Associations)
**Oracle Advanced Analytics DB Option**

**In-Database Machine Learning Algorithms** — SQL & GUI Access

### Classification
- Decision Tree
- Logistic Regression (GLM)
- Naïve Bayes
- Support Vector Machine (SVM)
- Random Forest

### Regression
- Multiple Regression (GLM)
- Support Vector Machine (SVM)
- Stepwise Linear Regression
- Linear Model
- Generalized Linear Model
- Multi-Layer Neural Networks

### Anomaly Detection
- 1-Class Support Vector Machine

### Clustering
- Hierarchical k-Means
- Orthogonal Partitioning Clustering
- Expectation-Maximization

### Attribute Importance
- Minimum Description Length
- Unsupervised pair-wise KL div.

### Market Basket Analysis
- Apriori – Association Rules

### Text Mining
- All OAA/ODM SQL ML support
- Explicit Semantic Analysis

### Predictive Queries
- Clustering
- Regression
- Anomaly Detection
- Feature Extraction

### Feature Extraction & Creation
- Nonnegative Matrix Factorization
- Principal Component Analysis
- Singular Value Decomposition

### Time Series
- Single & Double Exp. Smoothing

### Open Source R Algorithms
- Ability to run any R package (9,000+) via Embedded R mode

### Additional Features
- Ability to Mine Unstructured, Structured & Transactional data
- Partitioned Models

---

*Oracle Advanced Analytics DB Option is a comprehensive solution that includes a wide range of machine learning algorithms and data mining techniques, designed to enhance the analytical capabilities of Oracle Database. This includes both SQL and GUI access to a variety of algorithms and features, making it a powerful tool for data analysis and machine learning tasks.*
Oracle’s Advanced Analytics
Fastest Way to Deliver Scalable Enterprise-wide Predictive Analytics

Key Features

- Parallel, scalable data mining algorithms and R integration
- In-Database + Hadoop—Don’t move the data
- Data analysts, data scientists & developers
- Drag and drop workflow, R and SQL APIs
- Extends data management into powerful advanced/predictive analytics platform
- Enables enterprise predictive analytics deployment + applications
Google “Oracle Advanced Analytics”

Oracle Advanced Analytics

Scalable enterprise-wide predictive analytics

Architecture Overview

Oracle Advanced Analytics 12c delivers parallel in-database implementations of data mining algorithms and integration with open source R data analysis tool. Data analysts use Oracle Data Miner GUI and R to build and evaluate predictive models and leverage these models to predict events and market trends. Application developers deploy Oracle Advanced Analytics models using SQL data mining functions and R. With the Oracle Advanced Analytics option, Oracle provides a set of tools that form more data analysis and data types, eliminates data movement, and reduces storage security to anticipate customer behavior, detect patterns, and deliver actionable insights. Oracle Big Data SQL adds new big data sources to Oracle Advanced Analytics for Hadoop provides algorithms that run on Hadoop.

Oracle Advanced Analytics, a combination of Oracle Data Mining and Oracle R Enterprise, delivers predictive analytics, data mining, text mining, statistical analysis, interactive graphics inside the database. If there's a problem in your company, improvements in your sales, data analysts, data processing tools, and BI tools can develop methodologies for the database and gain

Oracle R Advanced Analytics for Hadoop

NEW ORACLE RELEASE 2.7.0 Introducing the fastest OLM and LM algorithms on Spark with full summary, enhanced Deep Neural Networks and support for Spark MLlib Gaussian Mixture Models.

The latest release of Oracle R Advanced Analytics for Hadoop (ORAAN), version 2.7.0, is one of the components of the Oracle Big Data Connectors software suite, an option to the Oracle Big Data Appliance. At its core, ORAAN provides an R interface for manipulating data stored in HDFS, using both R and transparency capabilities and mapping to R objects into R-like dataframes. The algorithms that can run are MapReduce jobs inside an Apache Spark container.
Oracle Data Mining/ Machine Learning/Predictive Analytics

Data Preparation & Adv. Analytical Process Runs In-Database

Additional relevant data and “engineered features”

Sensor data, Text, unstructured data, transactional data, spatial data, etc.

Historical data

Assembled historical data

Oracle Database 12c

Build Predictive Model

Historical or Current Data to be “scored” for predictions

Make Predictions

Predictions & Insights
Oracle’s Advanced Analytics
Fastest Way to Deliver Scalable Enterprise-wide ML/Predictive Analytics

Major Benefits

- Data remains in Database & Hadoop
  - Model building and scoring occur in-database
  - Use R packages with data-parallel invocations
- Leverage investment in Oracle IT
  - Eliminate data duplication
  - Eliminate separate analytical servers
- Deliver enterprise-wide applications
  - GUI for ML/Predictive Analytics & code gen
  - R interface leverages database as HPC engine

Oracle’s Advanced Analytics

Traditional Analytics

- Data Import
- Data Mining
  - Model “Scoring”
- Data Prep. & Transformation
- Data Mining
  - Model Building
- Data Prep. & Transformation
- Data Import

Secs, Mins or Hours

Oracle Advanced Analytics

- Model “Scoring”
- Embedded Data Prep
- Model Building
- Data Preparation

Hours, Days or Weeks

Savings
## Oracle Advanced Analytics 12.2
### Model Build Time Performance

<table>
<thead>
<tr>
<th>OAA 12.2 Algorithms</th>
<th>Rows (Ms)</th>
<th>T7-4 (Sparc &amp; Solaris) Model Build Time (Secs / Degree of Parallelism)</th>
<th>X5-4 (Intel and Linux)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributes Importance</td>
<td>640</td>
<td>28s / 512</td>
<td>44s / 72</td>
</tr>
<tr>
<td>K Means Clustering</td>
<td>640</td>
<td>161s / 256</td>
<td>268s / 144</td>
</tr>
<tr>
<td>Expectation Maximization</td>
<td>159</td>
<td>455s / 512</td>
<td>588s / 144</td>
</tr>
<tr>
<td>Naive Bayes Classification</td>
<td>320</td>
<td>17s / 256</td>
<td>23s / 72</td>
</tr>
<tr>
<td>GLM Classification</td>
<td>640</td>
<td>154s / 512</td>
<td>363s / 144</td>
</tr>
<tr>
<td>GLM Regression</td>
<td>640</td>
<td>55s / 512</td>
<td>93s / 144</td>
</tr>
<tr>
<td>Support Vector Machine (IPM solver)</td>
<td>640</td>
<td>404s / 512</td>
<td>1411s / 144</td>
</tr>
<tr>
<td>Support Vector Machine (SGD solver)</td>
<td>640</td>
<td>84s / 256</td>
<td>188s / 72</td>
</tr>
</tbody>
</table>

The way to read their results is that they compare 2 chips: X5 (Intel and Linux) and T7 (Sparc and Solaris). They are measuring scalability (time in seconds) with increase degree of parallelism (dop). The data also has high cardinality categorical columns which translates in high mining attributes (when algorithms require explosion). There are no comparisons to 12.1 and it is fair to say that the 12.1 algorithms could not run on data of this size.
Oracle’s Advanced Analytics and Machine Learning Platform

Multiple interfaces across platforms — SQL, R, GUI, Dashboards, Apps

Information Producers
R Client
R programmers
Data & Business Analysts
SQLDEV/Oracle Data Miner

Information Consumers
Business Analysts/Mgrs
OBIEE/DV
Domain End Users (HCM, CRM)
Applications

Users

Platform

Oracle Database Enterprise Edition

Oracle Advanced Analytics - Database Option
SQL Data Mining, ML & Analytic Functions + R Integration
for Scalable, Distributed, Parallel in-DB ML Execution

Oracle Cloud

Copyright © 2017, Oracle and/or its affiliates. All rights reserved.
You Can Think of Oracle Advanced Analytics Like This...

**Traditional SQL**
- “Human-driven” queries
- Domain expertise
- Any “rules” must be defined and managed

**SQL Queries**
- SELECT
- DISTINCT
- AGGREGATE
- WHERE
- AND OR
- GROUP BY
- ORDER BY
- RANK

**SQL Statistical Functions - SQL & R**
- Automated knowledge discovery, model building and deployment
- Domain expertise to assemble the “right” data to mine/analyze

**Statistical SQL “Verbs”**
- MEAN, STDEV
- MEDIAN
- SUMMARY
- CORRELATE
- FIT
- COMPARE
- ANOVA

FREE!
You Can Think of Oracle’s Advanced Analytics Like This…

**Traditional SQL**
- “Human-driven” queries
- Domain expertise
- Any “rules” must be defined and managed

**SQL Queries**
- SELECT
- DISTINCT
- AGGREGATE
- WHERE
- AND OR
- GROUP BY
- ORDER BY
- RANK

**Oracle Advanced Analytics - SQL & Analytical SQL “Verbs”**
- Automated knowledge discovery, model building and deployment
- Domain expertise to assemble the “right” data to mine/analyze

**Analytical SQL “Verbs”**
- PREDICT
- DETECT
- CLUSTER
- CLASSIFY
- REGRESS
- PROFILE
- IDENTIFY FACTORS
- ASSOCIATE
Oracle Advanced Analytics
How Oracle R Enterprise Compute Engines Work

1. R-> SQL Transparency “Push-Down”
   - R language for interaction with the database
   - R-SQL Transparency Framework overloads R functions for scalable in-database execution
   - Function overload for data selection, manipulation and transforms
   - Interactive display of graphical results and flow control as in standard R
   - Submit user-defined R functions for execution at database server under control of Oracle Database

2. In-Database Adv Analytical SQL Functions
   - 15+ Powerful data mining algorithms (regression, clustering, AR, DT, etc.)
   - Run Oracle Data Mining SQL data mining functioning (ORE.odmSVM, ORE.odmDT, etc.)
   - Speak “R” but executes as proprietary in-database SQL functions—machine learning algorithms and statistical functions
   - Leverage database strengths: SQL parallelism, scale to large datasets, security
   - Access big data in Database and Hadoop via SQL, R, and Big Data SQL

3. Embedded R Package Callouts
   - R Engine(s) spawned by Oracle DB for database-managed parallelism
   - ore.groupApply high performance scoring
   - Efficient data transfer to spawned R engines
   - Emulate map-reduce style algorithms and applications
   - Enables production deployment and automated execution of R scripts

Copyright © 2016, Oracle and/or its affiliates. All rights reserved.
Oracle Text Native Capability of every Oracle Database

• Oracle Text uses standard SQL to index, search, and analyze text and documents stored in the Oracle database, in files, and on the web.

• Oracle Text supports multiple languages and uses advanced relevance-ranking technology to improve search quality.

• Oracle Advanced Analytics leverages Oracle Text to pre-process (“tokenize”) unstructured data for the OAA SQL ML/data mining functions.
Rapidly Build, Evaluate & Deploy Analytical Methodologies
Leveraging a Variety of Data Sources and Types

SQL Joins and arbitrary SQL transforms & queries – power of SQL

Transactional POS data

Modeling Approaches

Consider:
- Demographics
- Past purchases
- Recent purchases
- Comments & tweets

Unstructured data also mined by algorithms

Generates SQL scripts and workflow API for deployment

Advanced Analytics

ORACLE
Oracle Advanced Analytics—On Premise or Cloud

100% Compatibility Enables Easy Coexistence and Migration

Transparency move workloads and ML/analytical methodologies between On-premise and public cloud

Same Architecture
Same ML/Analytics
Same Standards

On-Premise

Oracle Cloud

Advanced Analytics
Manage and **Analyze** All Data—SQL & Oracle Big Data SQL

**Big Data SQL + Advanced Analytics**

**Structured and Unstructured Data Reservoir**
- JSON data
- HDFS / Hive
- NoSQL
- Spatial and Graph data
- Image and Video data
- Social Media

**Store business-critical data in Oracle**
- Customer data
- Transactional data
- Unstructured documents, comments
- Spatial and Graph data
- Image and Video data
- Social Media

**Data analyzed via SQL / R / GUI**
- R Clients
- SQL Clients
- Oracle Data Miner

**Oracle Big Data Appliance**

**Oracle Database 12c**

**SQL / R**
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.

**Engineered Features** – Derived attributes/variable that reflect domain knowledge—key to best models

- Naïve Guess or Random
- Model with 20 variables
- Model with 75 variables
- Model with 250 variables
- Model with “Big Data” and hundreds

Copyright © 2016, Oracle and/or its affiliates. All rights reserved.
Oracle Advanced Analytics
Brief Demos
Oracle Data Miner GUI
Easy to Use for “Citizen Data Scientist”

- Easy to use to define analytical methodologies that can be shared
- SQL Developer Extension
- Workflow API and generates SQL code for immediate deployment
Explore Data

CUST_INSUR_LTV1
<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Output</th>
<th>Rank</th>
<th>Importance</th>
<th>% Null</th>
<th>% Unique</th>
<th>% Constant</th>
<th>Hints</th>
</tr>
</thead>
<tbody>
<tr>
<td>BANK_FUNDS</td>
<td>NUMBER</td>
<td>1</td>
<td>1</td>
<td>0.2039</td>
<td>0</td>
<td>21.3511</td>
<td>35.8016</td>
<td></td>
</tr>
<tr>
<td>N_TRANS_ATM</td>
<td>NUMBER</td>
<td>2</td>
<td>3</td>
<td>0.1217</td>
<td>0</td>
<td>21.499</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N_TRANS_TELLER</td>
<td>NUMBER</td>
<td>3</td>
<td>3</td>
<td>0.4931</td>
<td>0</td>
<td>32.5471</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MONEY_MONTHLY_OVERDRAWN</td>
<td>NUMBER</td>
<td>4</td>
<td>3</td>
<td>0.1207</td>
<td>0</td>
<td>18.787</td>
<td>16.3215</td>
<td></td>
</tr>
<tr>
<td>T_AMOUNT_AUTOM_PAYMENTS</td>
<td>NUMBER</td>
<td>5</td>
<td>3</td>
<td>0.1054</td>
<td>0</td>
<td>59.856</td>
<td>21.5483</td>
<td></td>
</tr>
<tr>
<td>MONTHLY_CHECKS_WRITTEN</td>
<td>NUMBER</td>
<td>6</td>
<td>3</td>
<td>0.0219</td>
<td>0</td>
<td>19.428</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N_OF_DEPENDENTS</td>
<td>NUMBER</td>
<td>7</td>
<td>3</td>
<td>0.0219</td>
<td>0</td>
<td>19.428</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIME_AS_CUSTOMER</td>
<td>NUMBER</td>
<td>8</td>
<td>3</td>
<td>0.0219</td>
<td>0</td>
<td>19.428</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHECKING_AMT</td>
<td>NUMBER</td>
<td>9</td>
<td>3</td>
<td>0.0219</td>
<td>0</td>
<td>19.428</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CREDIT_BALANCE</td>
<td>NUMBER</td>
<td>10</td>
<td>3</td>
<td>0.0219</td>
<td>0</td>
<td>19.428</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N_TRANS_KIOSK</td>
<td>NUMBER</td>
<td>11</td>
<td>3</td>
<td>0.0219</td>
<td>0</td>
<td>19.428</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MORTGAGE_AMT</td>
<td>NUMBER</td>
<td>12</td>
<td>3</td>
<td>0.0219</td>
<td>0</td>
<td>19.428</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEX</td>
<td>VARCHAR2</td>
<td>13</td>
<td>3</td>
<td>0.0219</td>
<td>0</td>
<td>19.428</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MARITAL_STATUS</td>
<td>VARCHAR2</td>
<td>14</td>
<td>3</td>
<td>0.0219</td>
<td>0</td>
<td>19.428</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOUSE_HO贰ERSHIP</td>
<td>NUMBER</td>
<td>15</td>
<td>3</td>
<td>0.0219</td>
<td>0</td>
<td>19.428</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
<Details algorithm="Naive Bayes" class="Yes">
  <Attribute name="LTV_BII" actualValue="MEDIUM" operator="in" range="LOW,MEDIUM,VERY HIGH" weight=".587" rank="1"/>
  <Attribute name="N_TRANS_WF_BANK" actualValue="2300" operator="greaterThan" value="1419.5" weight=".963" rank="2"/>
  <Attribute name="LTV" actualValue="21389.75" operator="greaterThan" value="6861.625" weight=".848" rank="3"/>
  <Attribute name="CAR_OWNERSHIP" actualValue="1" operator="between" range="(17.5:55.5)" weight=".8" rank="5"/>
</Details>
PREDICTION_COUNT 781.00
FULL_SIMPLE_RULE (BANK_FUNDS = 248) AND (CHECKING_AMOUNT = 282) AND (CREDIT_BALANCE = 2145) AND (MONEY_MONTHLY_OVERDRAWN = 54.095) AND (T_AMOUNT_AUTOM_PAYMENTS = 14990)
Sharing, Automation and Deployment
Immediately Go to “Productionization” of Analytical Methodologies

• Share ODMr workflows
• Workflow API for 100% automation
  • Immediate deployment of data analyst’s methodologies
• SQL Script Generation
  • Deploy methodology as SQL scripts
Fraud Prediction Demo

Automated In-DB Analytical Methodology

```
drop table CLAIMS_SET;
exec dbms_data_mining.drop_model('CLAIMSMODEL');
create table CLAIMS_SET (setting_name varchar2(30), setting_value varchar2(4000));
insert into CLAIMS_SET values ('ALGO_NAME', 'ALGO_SUPPORT_VECTOR_MACHINES');
insert into CLAIMS_SET values ('PREP_AUTO', 'ON');
commit;
begin
    dbms_data_mining.create_model('CLAIMSMODEL', 'CLASSIFICATION', 'CLAIMS', 'POLICYNUMBER', null, 'CLAIMS_SET');
end;
/

-- Top 5 most suspicious fraud policy holder claims
select * from
(select POLICYNUMBER, round(prob_fraud*100,2) percent_fraud,
    rank() over (order by prob_fraud desc) rnk from
(select POLICYNUMBER, prediction_probability CLAIMSMODEL, '0' using *) prob_fraud
from CLAIMS
where PASTNUMBEROFCLAIMS in ('2to4', 'morethan4'))
where rnk <= 5
order by percent_fraud desc;
```

Automated Monthly “Application”!

```
Just add:
Create View CLAIMS2_30 As
Select * from CLAIMS2
Where mydate > SYSDATE – 30

Time measure: set timing on;
```
Oracle Advanced Analytics
Real-Time Scoring, Predictions and Recommendations

• On-the-fly, single record apply with new data (e.g. from call center)

```sql
SELECT prediction_probability(CLAS_DT_1_64, 'Yes'
    USING 7800 as bank_funds, 125 as checking_amount, 20 as credit_balance, 55 as age, 'Married' as marital_status,
    250 as MONEY_MONLY_OVERDRAWN, 1 as house_ownership)
FROM dual;
```

Likelihood to respond:

<table>
<thead>
<tr>
<th>Query Result</th>
<th>All Rows Fetched: 1 in 0 seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREDICTION_PROB...</td>
<td></td>
</tr>
<tr>
<td>0.8382936507936...</td>
<td></td>
</tr>
</tbody>
</table>
Oracle’s Advanced Analytics

Example Customer References
Objectives

- Prevent $200M in losses every year using data to monitor, understand and anticipate fraud

Solution

- We installed OAA analytics for model development during 2014
- When choosing the tools for fraud management, speed is a critical factor
- OAA provided a fast and flexible solution for model building, visualization and integration with production processes

“When choosing the tools for fraud management, speed is a critical factor. Oracle Advance Analytics provided a fast and flexible solution for model building, visualization and integration with production processes.”

– Miguel Barrera, Director of Risk Analytics, Fiserv Inc.
– Julia Minkowski, Risk Analytics Manager, Fiserv Inc.
UK National Health Service

Combating Healthcare Fraud

Objectives

- Use new insight to help identify cost savings and meet goals
- Identify and prevent healthcare fraud and benefit eligibility errors to save costs
- Leverage existing data to transform business and productivity

Solution

- Identified up to GBP100 million (US$156 million) potentially saved through benefit fraud and error reduction
- Used anomaly detection to uncover fraudulent activity where some dentists split a single course of treatment into multiple parts and presented claims for multiple treatments
- Analyzed billions of records at one time to measure longer-term patient journeys and to analyze drug prescribing patterns to improve patient care

“Oracle Advanced Analytics’ data mining capabilities and Oracle Exalytics' performance really impressed us. The overall solution is very fast, and our investment very quickly provided value. We can now do so much more with our data, resulting in significant savings for the NHS as a whole”

– Nina Monckton, Head of Information Services, NHS Business Services Authority

Oracle Exadata Database Machine
Oracle Advanced Analytics
Oracle Exalytics In-Memory Machine
Oracle Endeca Information Discovery
Oracle Business Intelligence EE
Objectives

- Cloud-based solution
- Increase revenue
- Reduce time-to-market

Solution

The company considered only two solution vendors --SAS and Oracle to host its consumer data. SAS offered to help build the IT infrastructure from scratch and helped develop a one-year plan. But when they looked at the number of personnel needed to manage the infrastructure including administrators, security specialists and analysts as well as Security & HIPPA compliance needed, Oracle’s DBCS solution looked far more attractive. Hence, they decided to go with Oracle. Oracle’s solution offered:

- Scalability
- Built in analytical tools including data mining.
- Built in HIPPA compliance and security features.
- Required fewer resources --only two analysts –Data Engineer and an expert in Predictive Analytics who now manage the entire eco system.

“Time to market has significantly improved from 4-6 weeks to less than a week with the result the company can bring new clients on board faster. This has helped boost revenues by 25% in the six months since using Oracle’s DBCS..”

– DX Marketing

 DX Marketing Expands Customer Acquisition with Oracle Cloud – YouTube video
Zagrebačka Bank (biggest bank in Croatia)

Increases Cash Loans by 15% Within 18 Months of Deployment

**Objectives**

- Needed to speed up entire advanced analytics process; data prep was taking 3 days; model building 24 hours
- Faster time to “actionable analytics” for Credit Risk Modeling and Targeted Customer Campaigns

**Solution**

- Zaba migrated from SAS to the Oracle Advanced Analytics platform for statistical modeling and predictive analytics
- Increased prediction performance by leveraging the security, reliability, performance, and scalability of Oracle Database and Oracle Advanced Analytics for predictive analytics—running data preparation, transformation, model building, and model scoring within the database

“With Oracle Advanced Analytics we execute computations on thousands of attributes in parallel—impossible with open-source R. Analyzing in Oracle Database without moving data increases our agility. Oracle Advanced Analytics enables us to make quality decisions on time, increasing our cash loans business 15%.”

– Jadranka Novoselovic, Head of BI Dev., Zagrebačka Bank

“We chose Oracle because our entire data modeling process runs on the same machine with the highest performance and level of integration. With Oracle Database we simply switched on the Oracle Advanced Analytics option and needed no new tools,”

– Sinisa Behin, ICT coordinator at BI Dev. Zagrebačka Bank
An Post Boosts Retail & Postal Services with Big Data & Analytics Platform

Objectives
- Provide a scalable big data and analytics platform to manage millions of daily transactions, facilitate new postal services, and align with market needs and the growth in post office retail services

Solution
- Deployed a big data and analytics platform using Oracle Exadata Database Machine, Oracle Advanced Analytics, ... to ..., combat fraud, and enable readily-available enterprise wide business ... within core areas such as mails, parcels, and retail
- Enabled An Post to rapidly analyze and respond ... using Oracle Advanced Analytics...
- Vertice (partner) provided expert services including ... in-depth product and data mining expertise services

“With our Oracle solution for Big Data and Analytics we now have an analytics platform that has completely transformed our service delivery model. It provides An Post with a single source of truth while automatically consolidating data from all post offices nationwide.”

– John Cronin, Group Chief Information Officer, An Post Limited

“Big data is a vital part of our future,” says John Cronin, CIO of Ireland’s postal network.

https://www.siliconrepublic.com/enterprise/2016/02/05/an-post-oracle-john-cronin-five-minute-cio
Oracle Advanced Analytics
OAA/Oracle R Enterprise (R integration)
R—Widely Popular

R is a statistics language similar to Base SAS or SPSS statistics

R environment

• Strengths
  – Powerful & Extensible
  – Graphical & Extensive statistics
  – Free—open source

• Challenges
  – Memory constrained
  – Single threaded
  – Outer loop—slows down process
  – Not industrial strength
R: Transparency via function overloading

Invoke in-database aggregation function

```r
aggdata <- aggregate(ONTIME_S$DEST,
+ by = list(ONTIME_S$DEST),
+ FUN = length)

> class(aggdata)
[1] "ore.frame"
attr("package")
[1] "OREbase"

> head(aggdata)
   Group.1 x
1 ABE 237
2 ABI  34
3 ABQ 1357
4 ABY  10
5 ACK   3
6 ACT  33
```

Oracle SQL

```
select DEST, count(*)
from ONTIME_S
group by DEST
```

Oracle Database

In-db Stats

ONTIME_S

Database Server

Oracle Advanced Analytics

ORE Client Packages

Transparency Layer
R: Transparency via function overloading
Invoke in-database Data Mining model (Support Vector Machine)

```
> svm_mod <- ore.odmSVM(BUY~INCOME+YRS_CUST+MARITAL_STATUS, data=CUST, 
                        "classification", kernel="linear")

> summary(svm_mod)

Call:
ore.odmSVM(formula = BUY ~ INCOME + YRS_CUST + MARITAL_STATUS, data = CUST, 
type = "classification", kernel.function = "linear")

Settings:

<table>
<thead>
<tr>
<th></th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>prep.auto</td>
<td>on</td>
</tr>
<tr>
<td>active.learning</td>
<td>al.enable</td>
</tr>
<tr>
<td>complexity.factor</td>
<td>46.044899</td>
</tr>
<tr>
<td>conv.tolerance</td>
<td>1e-04</td>
</tr>
<tr>
<td>kernel.function</td>
<td>linear</td>
</tr>
</tbody>
</table>

Coefficients:

<table>
<thead>
<tr>
<th>class</th>
<th>variable</th>
<th>value</th>
<th>estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>INCOME</td>
<td>5.204561e-05</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>MARITAL_STATUS M</td>
<td>-4.531359e-05</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>MARITAL_STATUS S</td>
<td>4.531359e-05</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>YRS_CUST</td>
<td>1.264948e-04</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>(Intercept)</td>
<td>9.999269e-01</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>INCOME</td>
<td>2.032340e-05</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>MARITAL_STATUS M</td>
<td>2.636552e-06</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>MARITAL_STATUS S</td>
<td>-2.636555e-06</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>YRS_CUST</td>
<td>-1.588211e-04</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(Intercept)</td>
<td>-9.999324e-01</td>
<td></td>
</tr>
</tbody>
</table>
```
```r
# Load required packages
library(ORE)
library(OREmodels)
library(OREData)
library(lattice)
library(ORe)
library(OReData)
library(ORePredict)
library(OReX)

# Connect to database
ore.connect("dmuser", "oral2c", "localhost", "dmuser", all=TRUE)

data(pistormps)

# Perform operations
occasion <- occ.of(pistormps$diameter, pistormps$sample)
occ.of(125, type="xbar")

hist(CARSTATSS$MPG, col="red", breaks=25)
plot(CARSTATSS$MPG, col="red")

hist(CARSTATSS$MPG, col="red", breaks=25)
plot(CARSTATSS, col="red")

hist(CARSTATSS$MPG, col="red", breaks=25)
library(ORE)

ore.connect("dmuser", "oral2c", "localhost", "dmuser", all=TRUE)

hist(CARSTATSS$MPG, col="red", breaks=25)
ore.ts()
```
Oracle Advanced Analytics for Hadoop

Predictive algorithms that execute in a parallel/distributed manner on Hadoop with data in HDFS
Oracle R Advanced Analytics for Hadoop
Using Hadoop and HIVE Integration, plus R Engine and Open-Source R Packages

Hadoop Cluster
with Oracle R Advanced Analytics for Hadoop (ORAAH)

- R interface to HQL Basic Statistics, HQL
- Data Prep, Joins and View creation

Parallel, distributed algorithms:
- MLP Neural Nets*, GLM*, LM, PCA,
- k-Means, NMF, LMF

* Spark-Caching enabled

Use of Open-source R packages via custom R Mappers / Reducers

R Client
- R Analytics
- Oracle R Advanced Analytics for Hadoop

SQL Client
- SQL Developer
- Other SQL Apps

Oracle Database
with Advanced Analytics option

HQL
Basic Statistics,
Data Prep, Joins and View creation

Parallel, distributed algorithms:
- MLP Neural Nets*, GLM*, LM, PCA,
- k-Means, NMF, LMF

* Spark-Caching enabled

Use of Open-source R packages via custom R Mappers / Reducers
## Oracle R Advanced Analytics for Hadoop

### AA Algorithms in a Hadoop Cluster: Map-Reduce and Spark (2.7)

<table>
<thead>
<tr>
<th>Classification</th>
<th>Regression</th>
<th>Feature Extraction</th>
<th>Attribute Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM ORAAH</td>
<td>MLP Neural Networks ORAAH</td>
<td>Non-negative Matrix Factorization</td>
<td>Principal Components Analysis</td>
</tr>
<tr>
<td>Logistic Regression ORAAH</td>
<td>LASSO</td>
<td>Collaborative Filtering (LMF)</td>
<td>Principal Components Analysis</td>
</tr>
<tr>
<td>Random Forests</td>
<td>Ridge Regression</td>
<td></td>
<td>LASSO</td>
</tr>
<tr>
<td>Decision Trees</td>
<td>Support Vector Machines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Support Vector Machines</td>
<td>Linear Regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaussian Mixture Models</td>
<td>Random Forest</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Clustering</strong></td>
<td><strong>Basic Statistics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-Means</td>
<td>Correlation/Covariance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Copyright © 2016, Oracle and/or its affiliates. All rights reserved.
Oracle’s Advanced Analytics
Predictive Applications + OBIEE Integration
• Sales Predictor helps sales reps answer critical sales questions:
  • Which products should be offered to a customer?
  • Who are the customers buying products?
  • What are the reasons a product is being bought?
• Sales Predictor offers product recommendations that have a higher likelihood of being converted to a win.
Oracle Communications Industry Data Model

Example Predictive Analytics Application

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

Link to OCDM on OTN
Oracle Communications Data Model

Pre-Built Data Mining Models

1. Churn Prediction
2. Customer Profiling
3. Customer Churn Factor
4. Cross-Sell Opportunity
5. Customer Life Time Value
6. Customer Sentiment
7. Customer Life Time Value

Link to OCDM on OTN
Oracle Advanced Analytics 12.2, Oracle Data Miner 4.2 and ORAAH 2.6

New Features + Road Map
New Oracle Database Features

• **Major Performance Improvements for Algorithms**
  – New parallel model build / apply infrastructure to enable faster algorithm introduction
  – Scale to larger data volumes found in big data and cloud use cases

• **Unsupervised Feature Selection**
  – Uses unsupervised pair-wise Kullback-Leibler Divergence (KLD - correlations analysis - numeric and categorical attributes) to find highest “information containing” attributes

• **Association Rules Enhancements**
  – Adds calculation of values associated with AR rules such as sales amount to indicate the value of co-occurring items in baskets

• **Partitioned Models**
  – Instead of building, naming and referencing 10s or 1000s of models, a partitioned model will organize and represent these multiple models as partitions in a single model entity
Explicit Semantic Analysis (ESA) algorithm

- Useful technique for extracting meaningful, interpretable features, better than LDA
- Use ESA for document similarity and topic identification
- Wikipedia provides a large corpus of existing documents to provide sensible features and topics

Document 1

'Senior members of the Saudi royal family paid at least $560 million to Osama bin Laden terror group and the Taliban for an agreement his forces would not attack targets in Saudi Arabia, according to court documents. The papers, filed in a $US3000 billion ($5500 billion) lawsuit in the US, allege the deal was made after two secret meetings between Saudi royals and leaders of al-Qa ida, including bin Laden. The money enabled al-Qa ida to fund training camps in Afghanistan later attended by the September 11 hijackers. The disclosures will increase tensions between the US and Saudi Arabia.'

Document 2

'The Saudi Interior Ministry on Sunday confirmed it is holding a 21-year-old Saudi man the FBI is seeking for alleged links to the Sept. 11 hijackers. Authorities are interrogating Saud Abdulaziz Saud al-Rasheed "and if it is proven that he was connected to terrorism, he will be referred to the sharia (Islamic) court," the official Saudi Press Agency quoted an unidentified ministry official as saying.'

ESA Similarity 0.62
Oracle Advanced Analytics 12.2
New Oracle Database Features

• Extensibility for R Models
  – Register R models as in-database models for build, apply, settings, and viewing
  – Supports data with “nested” attributes, handling text and aggregated transactional data for open source R packages
  – Extends ease of advanced analytics development from R to Oracle Database
  – Enables R users to roll out new analytics and more rapidly take advantage of existing R packages
Oracle Data Miner

4.2 New Features
Oracle Data Miner 4.2

New Features for OAA

• Add/Expose all 12.2 features in Oracle Data Miner UI
NEW IN 4.2

Workflow Scheduler

NEW IN 4.2
Oracle R Advanced Analytics for Hadoop
New Features in ORAAH 2.7

• Updated ORAAH GLM and LM algorithms which are much faster, stable and light on memory than comparable GLM and LM methods from Spark Mllib
• Both methods also bring a new summary feature that makes them comparable to solutions from open-source R glm and lm, but capable of handling Big Data at enterprise scale
• The Neural Networks algorithm has been enhanced to support the full formula processing and a full build and scoring in Spark
• The new Gaussian Mixture Models is an addition to the set of algorithms supported in Spark Mllib
• ORAAH's Spark-based LM with full formula support and summary - orch.lm2()
• ORAAH's Spark-based GLM with full formula support and summary - orch.glm2()
Getting started
Getting started: OAA Links and Resources

Oracle Advanced Analytics Overview:
- OAA presentation — [Big Data Analytics with Oracle Advanced Analytics](#) or just watch [YouTube video presentation and demo(s)](#)
- [Big Data Analytics with Oracle Advanced Analytics: Making Big Data and Analytics Simple](#) white paper on OTN
- Oracle Internal [OAA Product Management Wiki and Workspace](#)
- [Oracle Advanced Analytics Customer Successes](#)

YouTube recorded OAA Presentations and Demos:
- [Oracle Advanced Analytics and Data Mining at the YouTube Movies](#) (6 + OAA “live” Demos on ODM’r 4.0 New Features, Retail, Fraud, Loyalty, Overview, etc.)

Getting Started:
- [Link to OAA/Oracle Data Miner Workflow GUI Online (free) Tutorial Series on OTN](#)
- [Link to OAA/Oracle R Enterprise (free) Tutorial Series on OTN](#)
- [Link to Free Oracle Advanced Analytics "Test Drives" on Oracle Cloud](#) via Vlamis Partner
- [Link to Getting Started w/ ODM blog entry](#)
- [Link to New OAA/Oracle Data Mining 2-Day Instructor Led Oracle University course](#)
- [Oracle Data Mining Sample Code Examples](#)

Additional Resources:
- [Oracle Advanced Analytics Option on OTN](#)
- [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](#)
- [Oracle R Advanced Analytics for Hadoop (ORAAH) on OTN](#)
- [Business Intelligence, Warehousing & Analytics—BIWA Summit’17, Jan 31, Feb 1 & 2, 2017 at Oracle HQ Conference Center (w/ links to customer presentations)](#)
One more thing...
Oracle Advanced Analytics Strategy & Road Map

- One server side product, with a single analytic library, supporting multiple data platforms, analytical engines, UIs and deployment strategies.
Oracle Machine Learning
Multi-Platform, Multi-Engine Oracle Machine Learning for the Cloud

Introduction

• A collaborative unified notebook user interface and cloud platform for Oracle’s machine learning capabilities

• Enables teams to...
  – Leverage the languages, platforms and engines of their choice
  – Explore and prepare data
  – Build, access, and use machine learning models
  – Deploy machine learning solutions
Oracle Machine Learning
Multi-Platform, Multi-Engine Oracle Machine Learning for the Cloud

Features

• Zeppelin-based Unifying Notebook UI
  – Multiple backend servers, analytical engines and languages
  – Oracle and open source algorithms, platforms, tools, and data sources (SQL, R, Spark, Python)
  – Provides a flexible toolbox for the data scientist

• Supports collaborative development
  – Shared notebooks and templates, with access permissions

• Supports deployment of predictive analytics solutions
  – Enables publishing libraries, templates, examples of common use cases

• Dynamic and expandable platform
Oracle’s Advanced Analytics and Machine Learning Platform

Multiple interfaces across platforms — SQL, R, GUI, Dashboards, Apps

**Information Producers**

- **Oracle**
- **Oracle Machine Learning**
- **Oracle Database Enterprise Edition**

**Oracle’s Advanced Analytics and Machine Learning Platform**

- **Multiple interfaces across platforms**
  - SQL, R, GUI, Dashboards, Apps

**Users**

- R programmers
- Data & Business Analysts
- Data Scientists (R, SQL, Python, etc.)

**Platform**

- **Hadoop**
  - ORAAH
  - Parallel, distributed algorithms
- **Oracle Database Enterprise Edition**
  - **Oracle Advanced Analytics - Database Option**
    - SQL Data Mining, ML & Analytic Functions + R Integration
    - for Scalable, Distributed, Parallel in-DB ML Execution

**Oracle Cloud**

- Open Source MLLib, RServ, Spark, etc.
Oracle Machine Learning

Beta Program
- Seeking knowledgeable, articulate data scientists/data analysts users for early product access, beta testing and providing valuable customer feedback
- Oracle Machine Learning will be hosting as “i-Betas” in the Oracle Cloud for easier customer access and testing
- OML Beta Program 2017

OML Contacts in Product Management
- Charlie Berger – charlie.berger@oracle.com
- Mark Hornick – mark.hornick@oracle.com
- Marcos Arancibia – marcos.arancibia@oracle.com