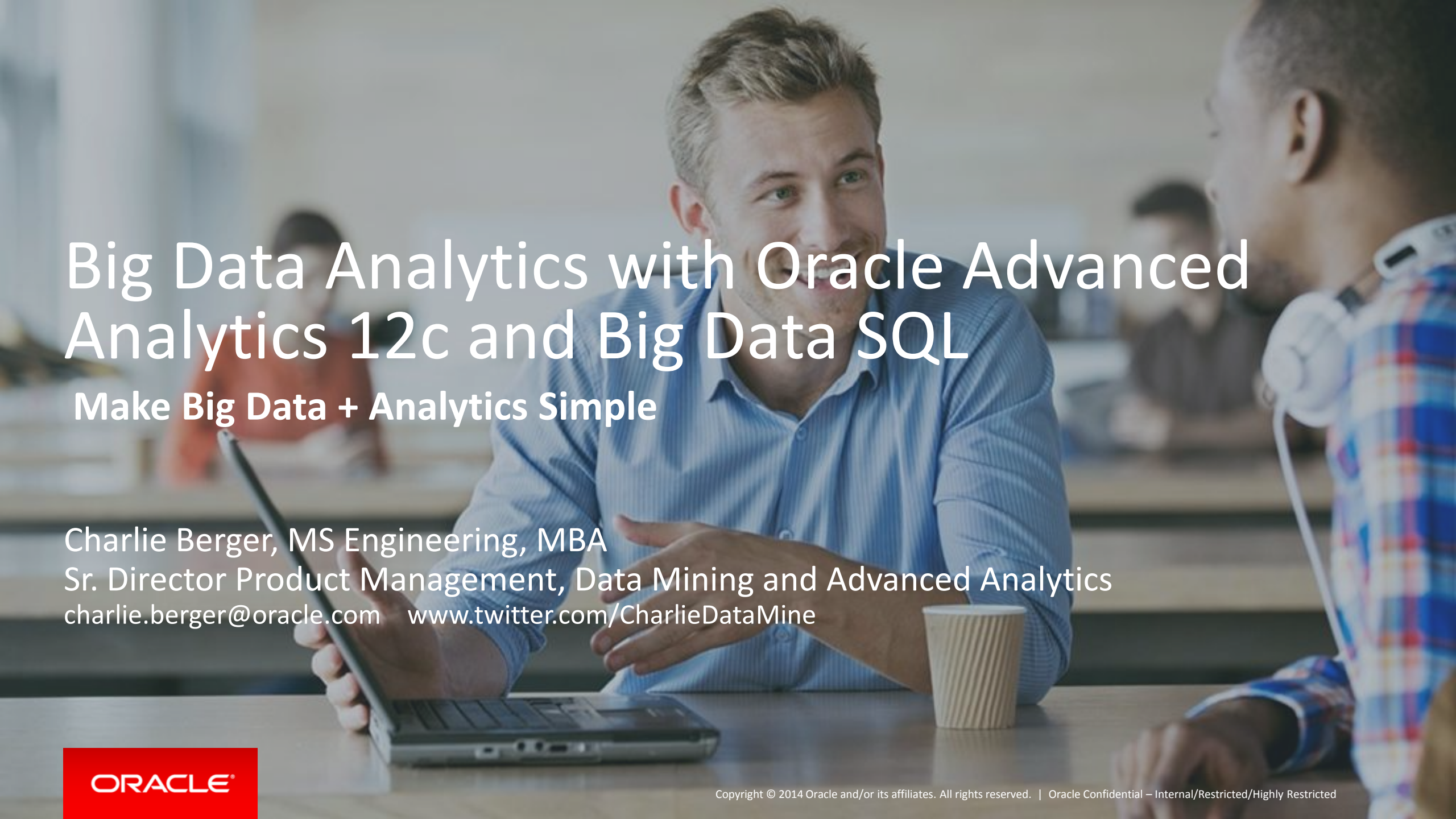


Big Data Analytics with Oracle Advanced Analytics 12c and Big Data SQL

Make Big Data + Analytics Simple

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

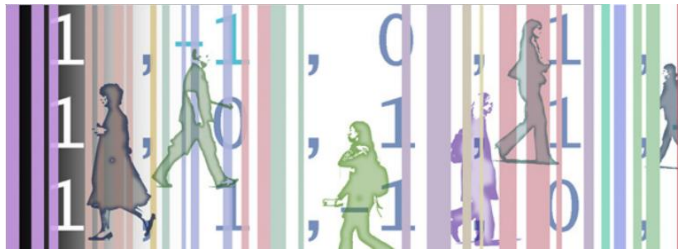
ORACLE

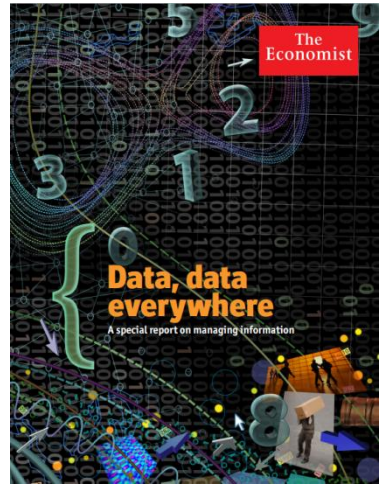
Safe Harbor Statement

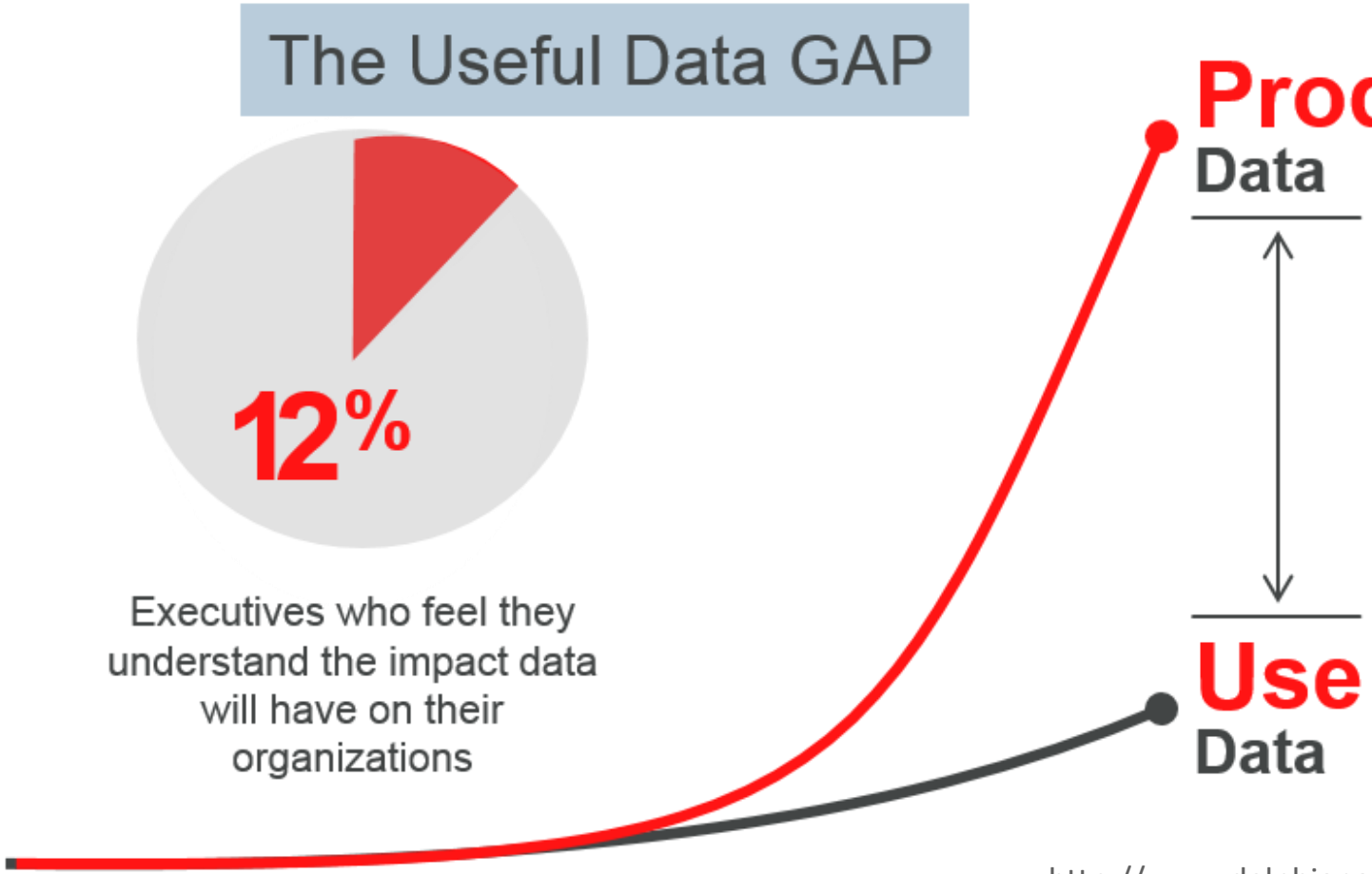
The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development,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

Data, data everywhere

Growth of Data Exponentially Greater than Growth of Data Analysts!







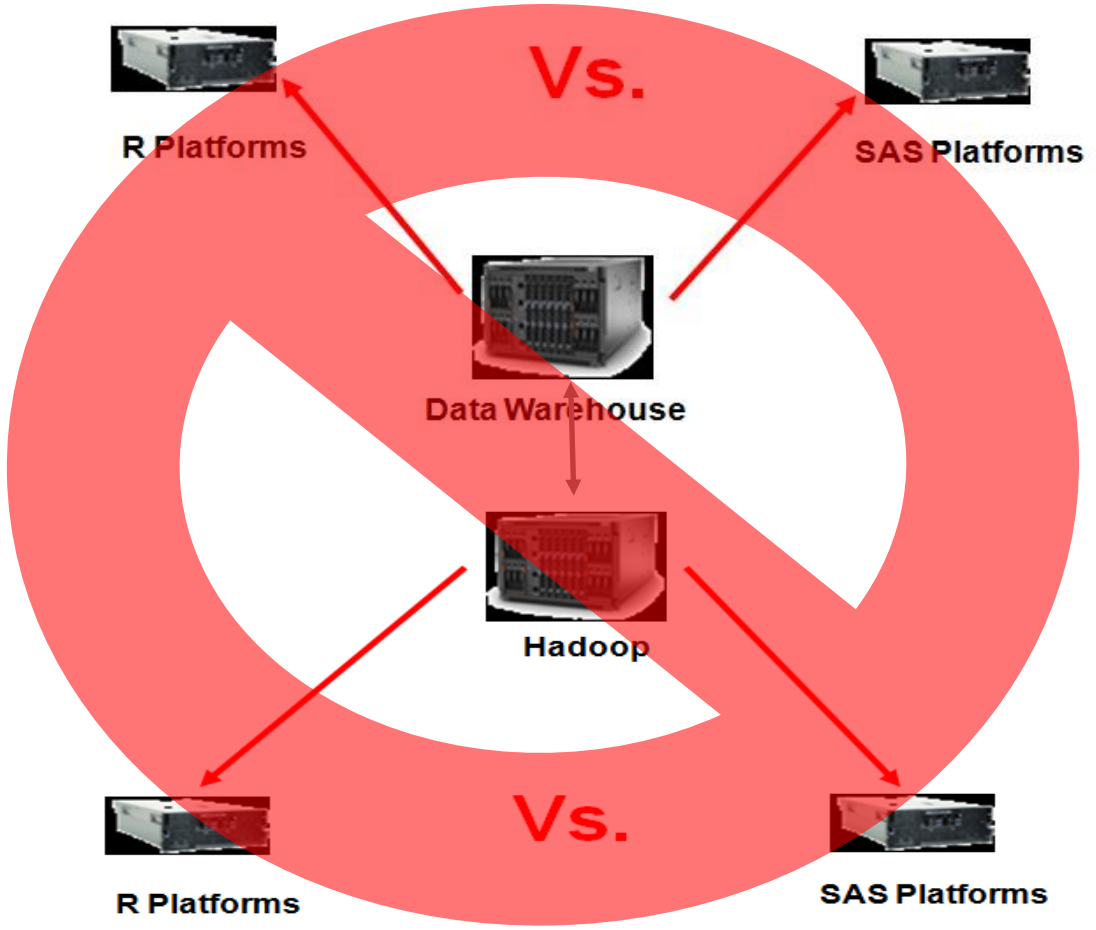
Data Analysis platforms requirements:

- 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/
http://uk.emc.com/collateral/analyst-reports/ar-the-economist-data-data-everywhere.pdf

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 + Analytic Platform for the Era of Big Data and Cloud
  - **Make Big Data + Analytics *Simple***
    - Any data size, on any computer infrastructure
    - Any variety of data, in any combination
  - **Make Big Data + Analytics Deployment *Simple***
    - As a service, as a platform, as an application

Oracle Advanced Analytics Database Evolution

1998
- *7 Data Mining "Partners"*

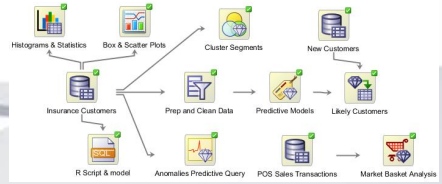
1999
- *Oracle acquires Thinking Machine Corp's dev. team + "Darwin" data mining software*

2002
- *Oracle Data Mining 9.2i launched – 2 algorithms (NB and AR) via Java API*

2004
- *Oracle Data Mining 10g & 10gR2 introduces SQL dm functions, 7 new SQL dm algorithms and new Oracle Data Miner "Classic" wizards driven GUI*

2005
- *ODM 11g & 11gR2 adds AutoDataPrep (ADP), text mining, perf. improvements*
- *SQLDEV/Oracle Data Miner 3.2 "work flow" GUI launched*
- *Integration with "R" and introduction/addition of Oracle R Enterprise*
- *Product renamed "Oracle Advanced Analytics (ODM + ORE)*

2008



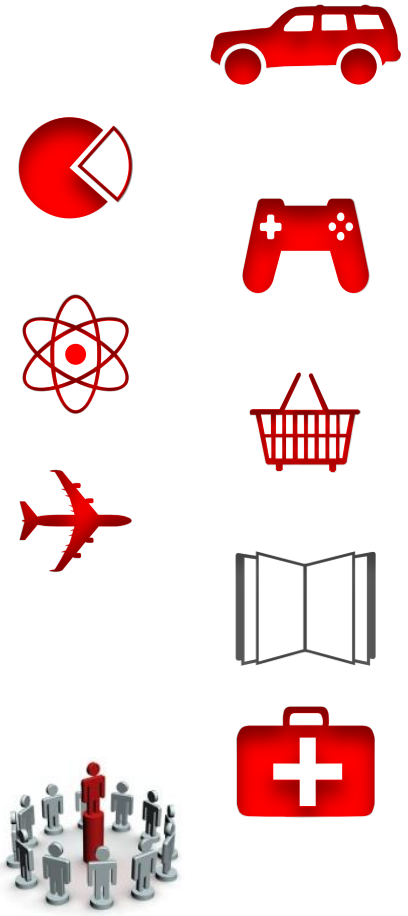
2011
- *New algorithms (EM, PCA, SVD)*
- *Predictive Queries*
- *SQLDEV/Oracle Data Miner 4.0 SQL script generation and SQL Query node (R integration)*
- *OAA/ORE 1.3 + 1.4 adds NN, Stepwise, scalable R algorithms*
- *Oracle Adv. Analytics for Hadoop Connector launched with scalable BDA algorithms*

2014

Predictive Analytics & Data Mining

Typical Use Cases

- Targeting the right customer with the right offer
- How is a customer likely to respond to an offer?
- Finding the most profitable growth opportunities
- Finding and preventing customer churn
- Maximizing cross-business impact
- Security and suspicious activity detection
- Understanding sentiments in customer conversations
- Reducing medical errors & improving quality of health
- Understanding influencers in social networks

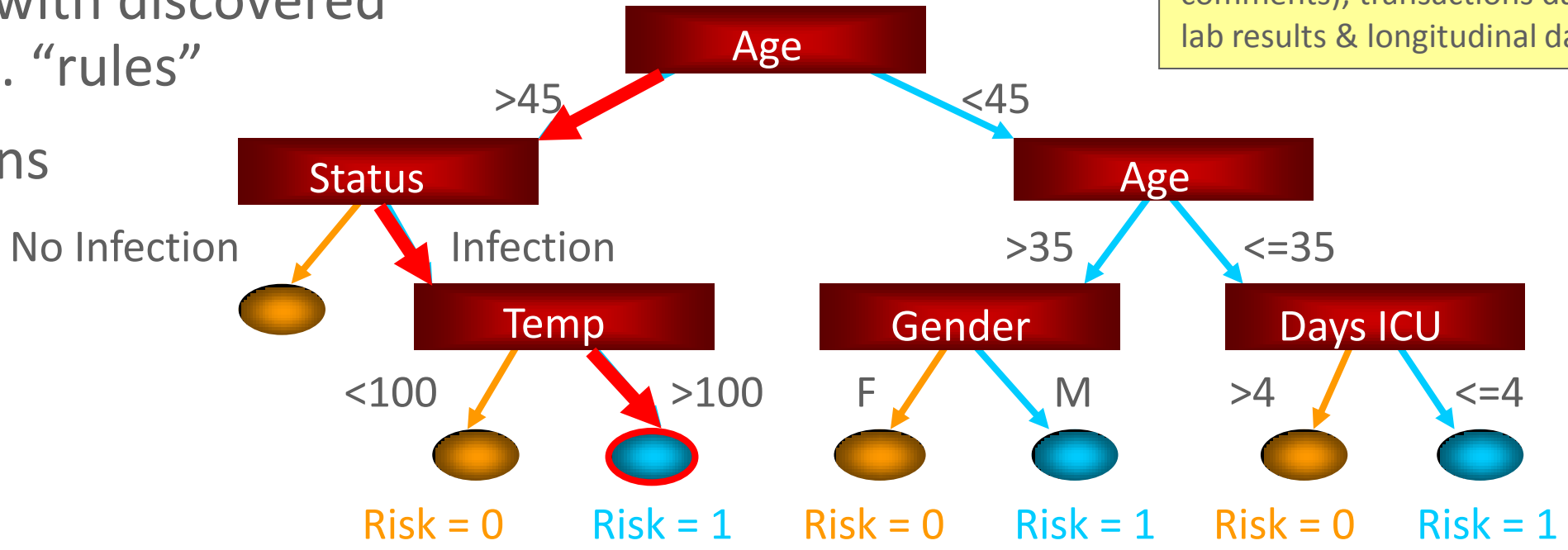


Oracle Advanced Analytics: Supervised Learning

"Classification" Decision Tree Algorithm

- Profiling with discovered If…Then… "rules"
- Predictions

Simple model: Could include unstructured data (e.g. physician's comments), transactions data (e.g. lab results & longitudinal data), etc.



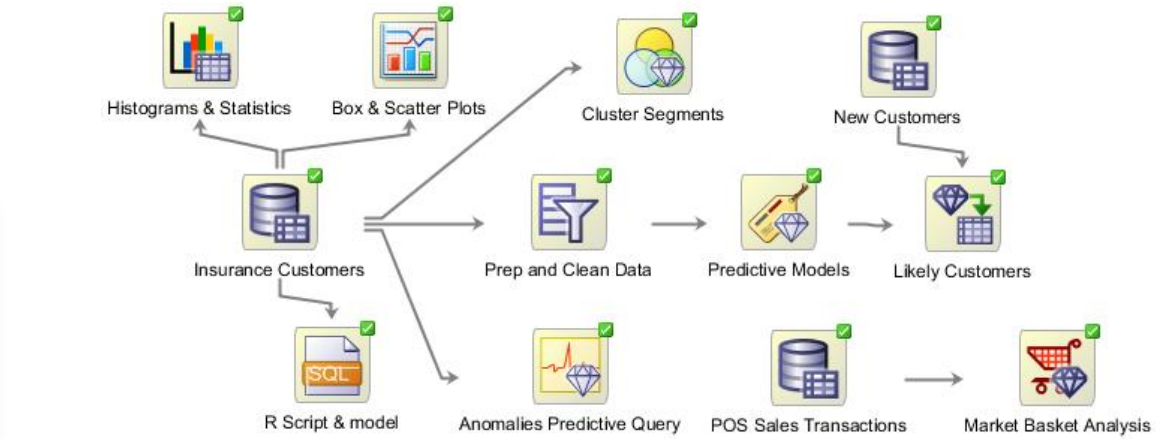
IF *(Age > 45 AND Status = Infection AND Temp = >100)*
THEN *Probability(High Risk=1) = .77 and Support = 250 cases*

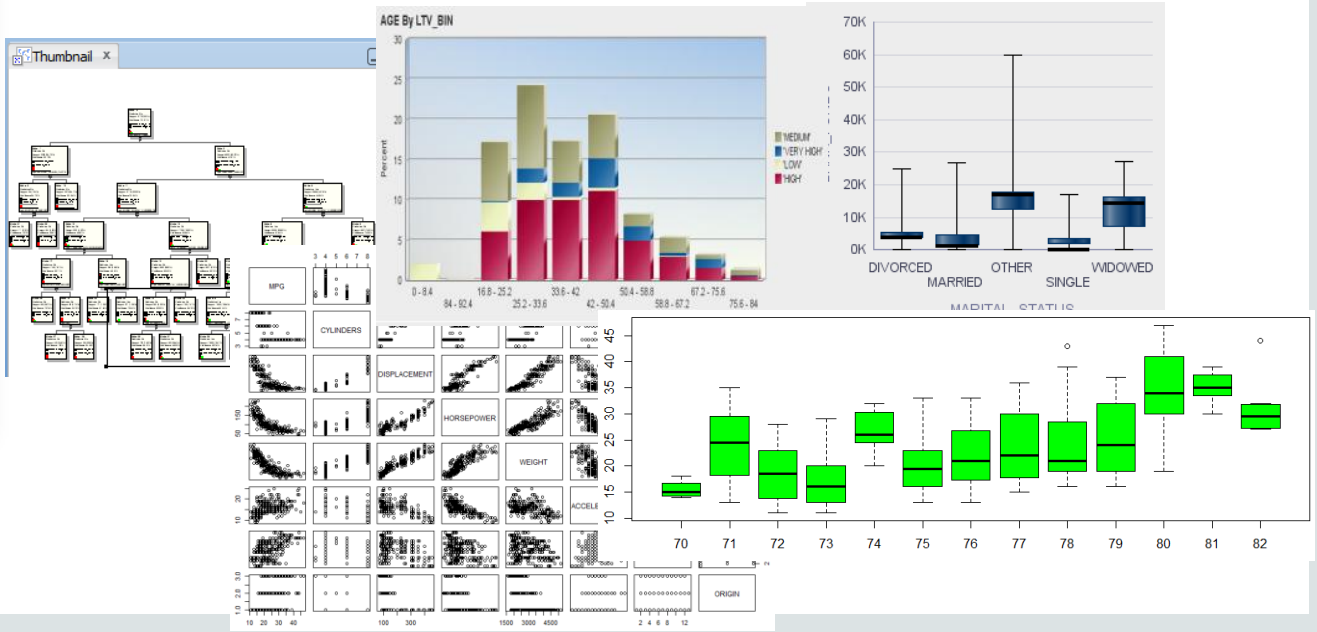
Oracle Advanced Analytics Database Option

Fastest Way to Deliver Scalable Enterprise-wide Predictive Analytics

Key Features

- In-database data mining algorithms and open source R algorithms
- Trilingual component of Oracle Database—SQL, SQLDev/ODMr GUI, R
- Scalable, parallel in-database execution
- Workflow GUI and IDEs
- Integrated component of Database
- Enables enterprise analytical applications





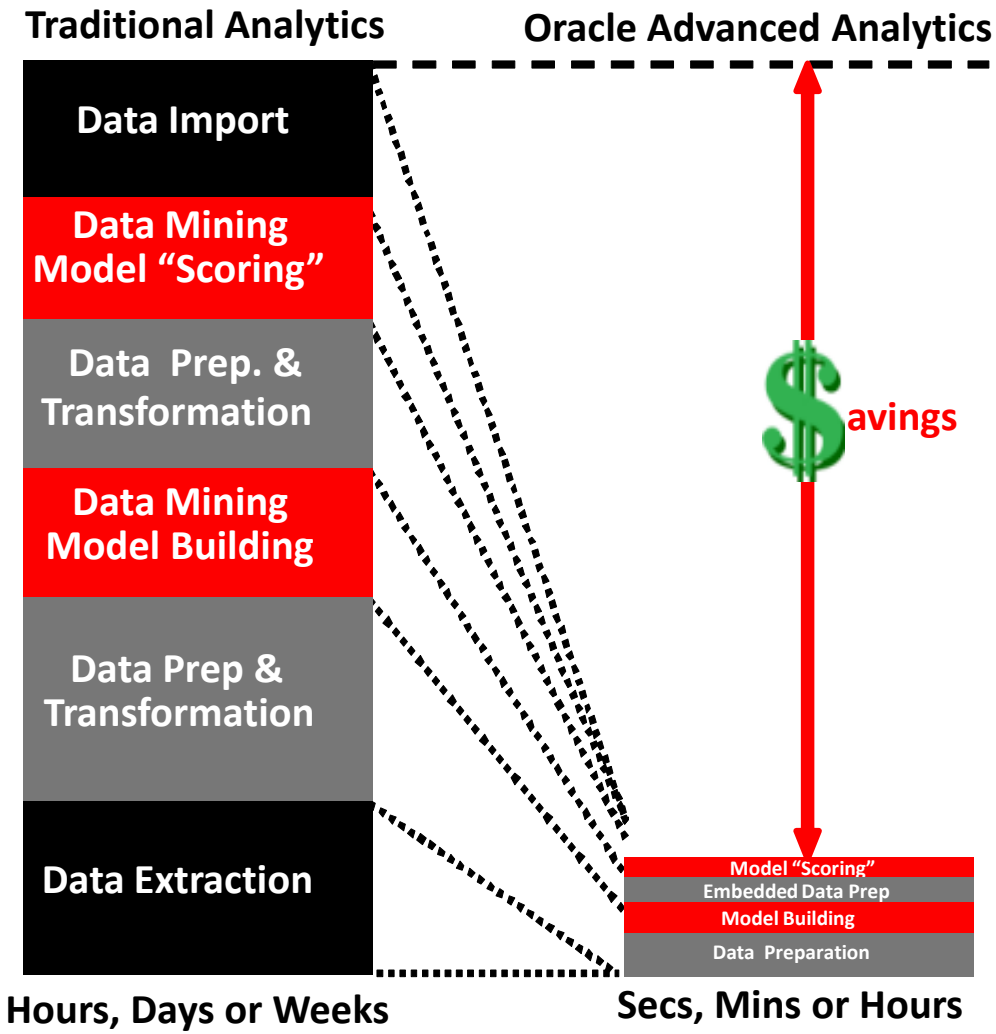
Oracle Advanced Analytics Database Option

Fastest way to deliver enterprise-wide predictive analytics

Key Features

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



Oracle Advanced Analytics Database Option

Fastest way to deliver enterprise-wide predictive analytics

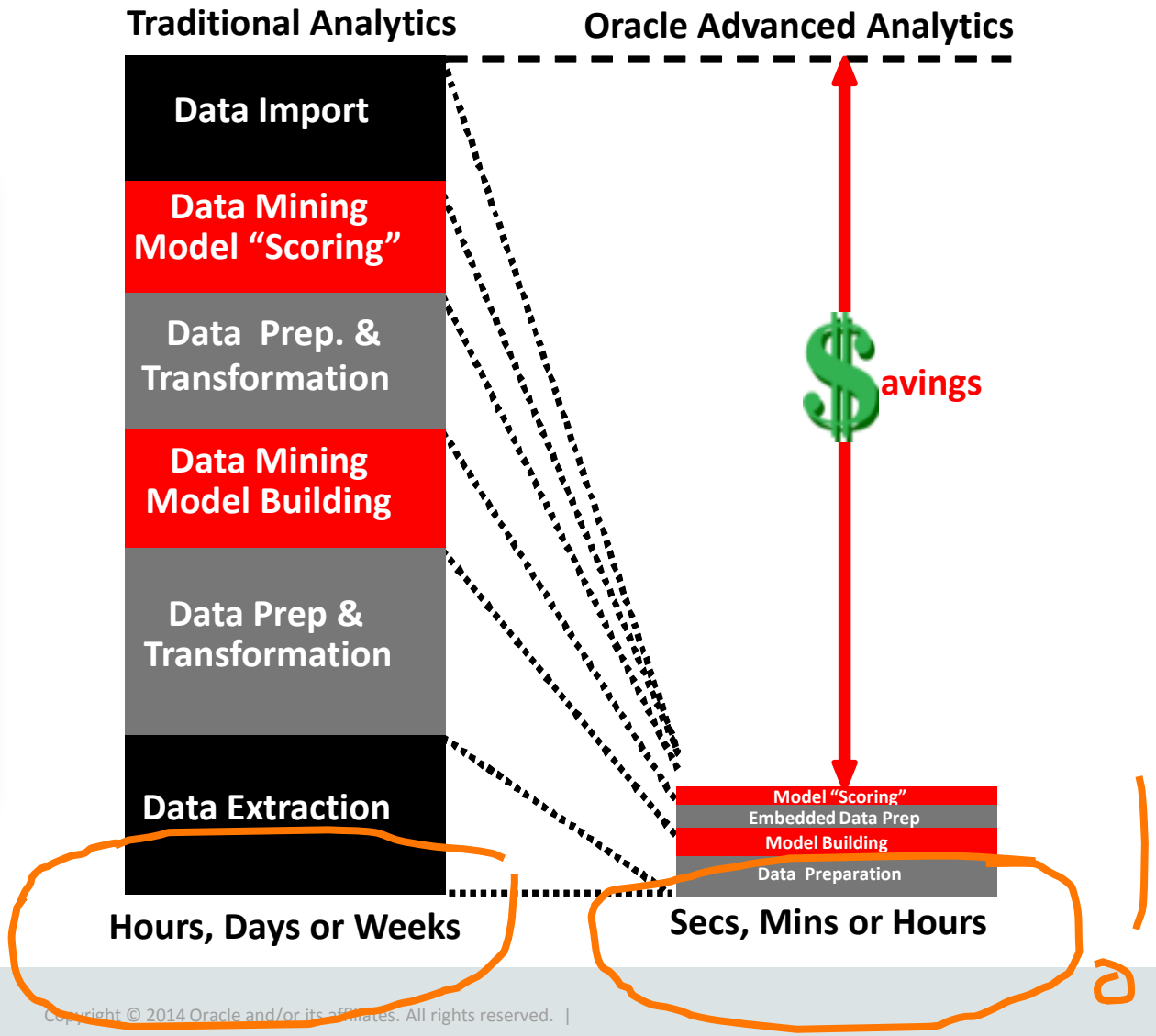
Key Features

Lowest Total Cost of Ownership

- Eliminate data duplication
- Eliminate separate analytical servers
- Leverage investment in Oracle IT

Fastest way to deliver *enterprise-wide* predictive analytics

- Integrated GUI for Predictive Analytics
- Database scoring engine



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`

+

Oracle Advanced Analytics (SQL & R)

- 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

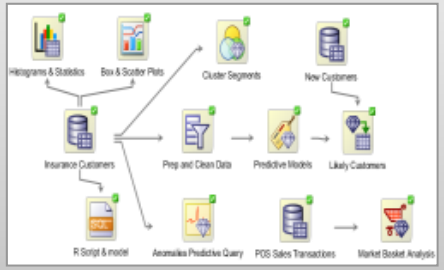
Oracle Advanced Analytics Database Architecture

Trilingual Component of Oracle Database—SQL, SQLDev/ODMr GUI, R

Users

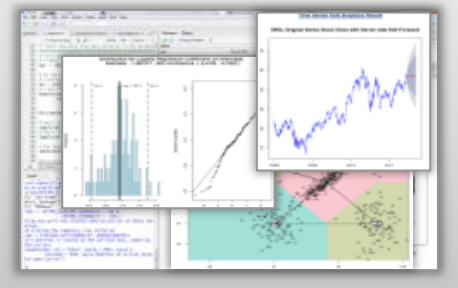
Data & Business Analysts

SQL Developer



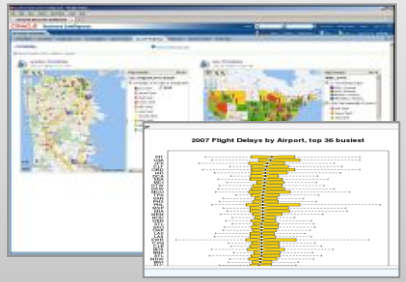
R programmers

R Client



Business Analysts/Mgrs

OBIEE



Domain End Users

Applications

Platform

Oracle Database Enterprise Edition

Oracle Advanced Analytics

Native SQL Data Mining/Analytic Functions + High-performance

R Integration for Scalable, Distributed, Parallel Execution

ORACLE

Turkcell

Combating Communications Fraud

Objectives

- Prepaid card fraud—millions of dollars/year
- Extremely fast sifting through huge data volumes; with fraud, time is money

Solution

- Monitor 10 billion daily call-data records
- Leveraged SQL for the preparation—1 PB
- Due to the slow process of moving data, Turkcell IT builds and deploys models in-DB
- Oracle Advanced Analytics on Exadata for extreme speed. Analysts can detect fraud patterns almost immediately

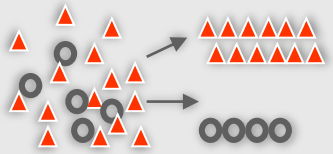
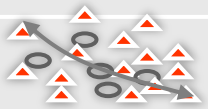
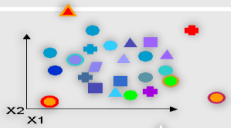
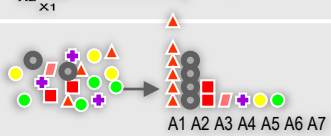
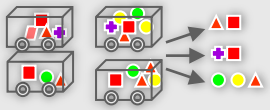
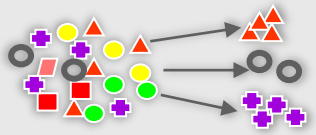
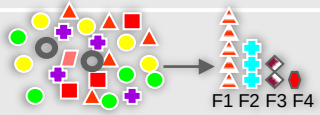
- "Turkcell manages 100 terabytes of compressed data—or one petabyte of uncompressed raw data—on Oracle Exadata. With Oracle Data Mining, a component of the Oracle Advanced Analytics Option, we can analyze large volumes of customer data and call-data records easier and faster than with any other tool and rapidly detect and combat fraudulent phone use."
  – Hasan Tonguç Yılmaz, Manager, Turkcell İletişim Hizmetleri A.Ş.

Oracle Advanced Analytics In-Database Fraud Models

Exadata

Oracle Advanced Analytics

In-Database Data Mining Algorithms—SQL & R & GUI Access

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

Oracle Advanced Analytics Database Option

Wide Range of In-Database Data Mining and Statistical Functions

- **Data Understanding & Visualization**
  - Summary & Descriptive Statistics
  - Histograms, scatter plots, box plots, bar charts
  - R graphics: 3-D plots, link plots, special R graph types
  - Cross tabulations
  - Tests for Correlations (t-test, Pearson's, ANOVA)
  - Selected Base SAS equivalents
- **Data Selection, Preparation and Transformations**
  - Joins, Tables, Views, Data Selection, Data Filter, SQL time windows, Multiple schemas
  - Sampling techniques
  - Re-coding, Missing values
  - Aggregations
  - Spatial data
  - SQL Patterns
  - R to SQL transparency and push down
- **Classification Models**
  - Logistic Regression (GLM)
  - Naive Bayes
  - Decision Trees
  - Support Vector Machines (SVM)
  - Neural Networks (NNs)
- **Regression Models**
  - Multiple Regression (GLM)
  - Support Vector Machines
- **Clustering**
  - Hierarchical K-means
  - Orthogonal Partitioning
  - Expectation Maximization
- **Anomaly Detection**
  - Special case Support Vector Machine (1-Class SVM)
- **Associations / Market Basket Analysis**
  - A Priori algorithm
- **Feature Selection and Reduction**
  - Attribute Importance (Minimum Description Length)
  - Principal Components Analysis (PCA)
  - Non-negative Matrix Factorization
  - Singular Vector Decomposition
- **Text Mining**
  - Most OAA algorithms support unstructured data (i.e. customer comments, email, abstracts, etc.)
- **Transactional & Spatial Data**
  - All OAA algorithms support transactional data (i.e. purchase transactions, repeated measures over time, distances from location, time spent in area A, B, C, etc.)
- **R packages—ability to run open source**
  - Broad range of R CRAN packages can be run as part of database process via R to SQL transparency and/or via Embedded R mode

ORACLE

* included free in every Oracle Database

Why Netflix Never Implemented The Algorithm | That Won The Netflix $1 Million Challenge

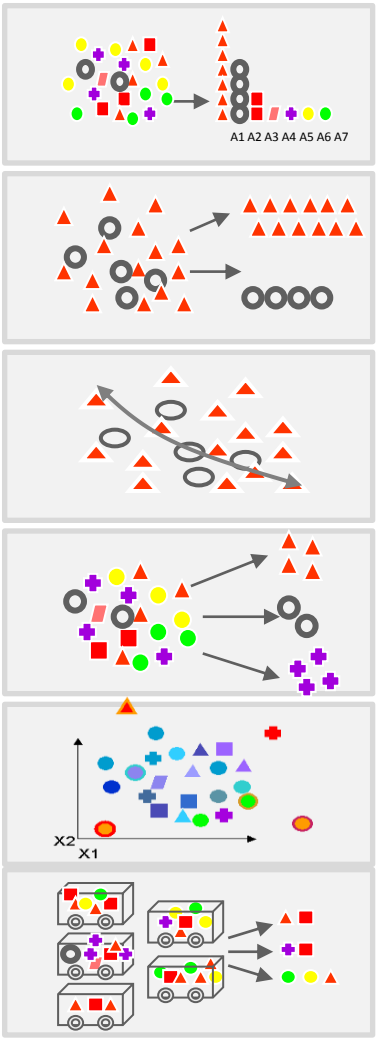
"We evaluated some of the new methods offline but the additional accuracy gains that we measured did not seem to justify the engineering effort needed to bring them into a production environment."

https://www.techdirt.com/blog/innovation/articles/20120409/03412518422/why-netflix-never-implemented-algorithm-that-won-netflix-1-million-challenge.shtml

What is Data Mining?

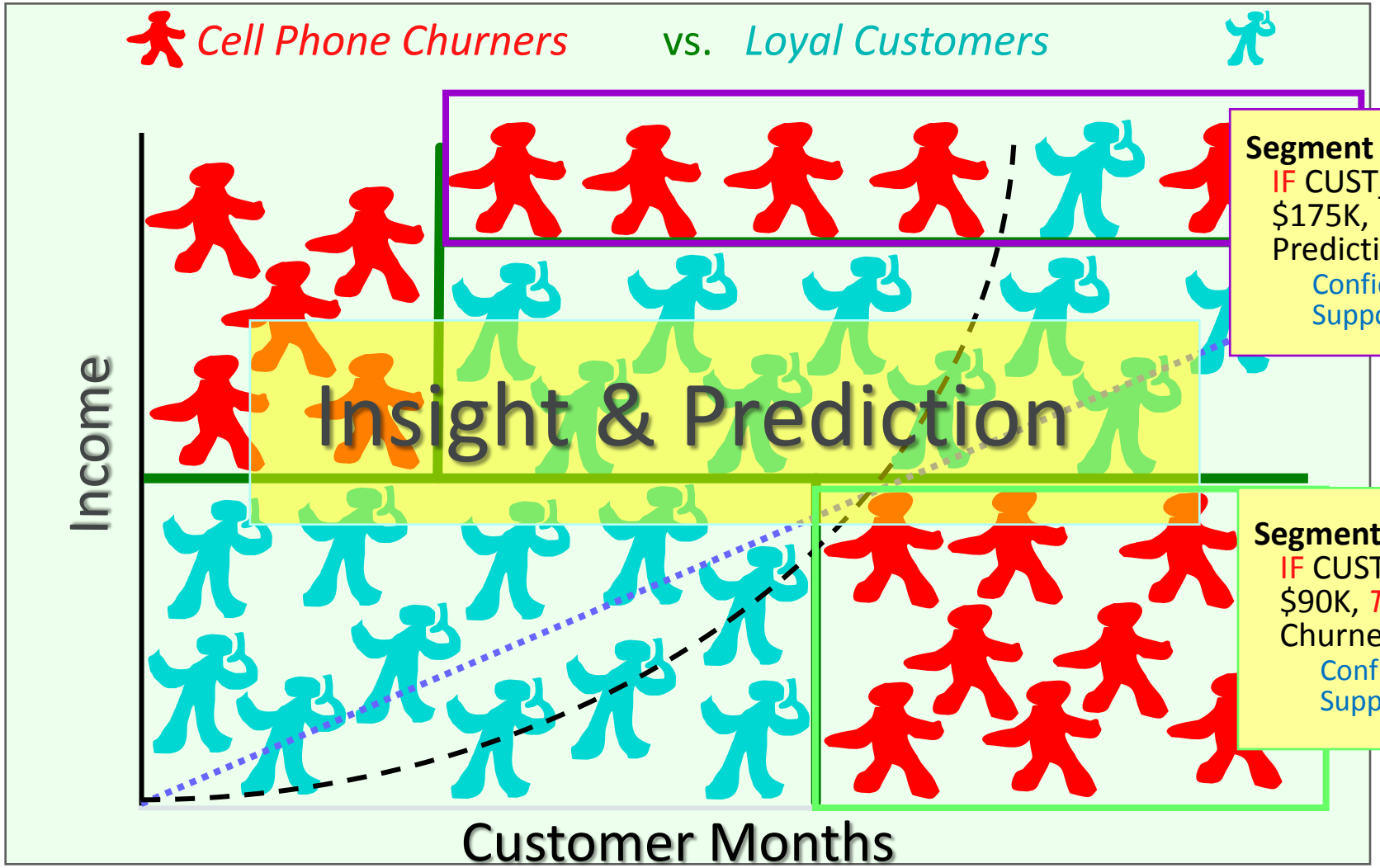
Automatically sifting through large amounts of data to find previously hidden patterns, discover valuable new insights and make predictions

- 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)*



Data Mining Provides

Better Information, Valuable Insights and Predictions



Source: Inspired from *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* by Michael J. A. Berry, Gordon S. Linoff

Typical Data Mining Initial Questions

Start with a Business Problem Statement

- Predict employees that voluntarily churn
- Predict customers that are likely to churn
- Target “best” customers
- Find items that will help me sell more most profitable items
- What is a specific customer most likely to purchase next?
- Who are my “best customers”?
- How can I combat fraud?
- I’ve got all this data; can you “mine” it and find useful insights?

Oracle Advanced Analytics—*Best Practices*

Nothing is Different; Everything is Different

1. Start with a Business Problem Statement
2. Don't Move the Data
3. Assemble the "Right Data" for the Problem
4. Create New Derived Variables
5. Be Creative in Analytical Methodologies
6. Quickly Transform "Data" to "Actionable Insights"
7. Automate and Deploy Enterprise-wide

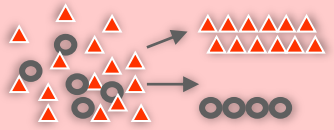
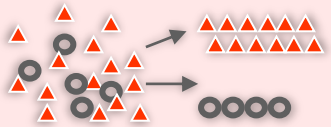
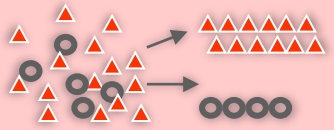
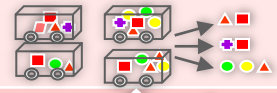
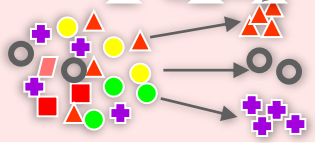
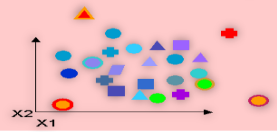
Be Specific in Problem Statement

Poorly Defined
Predict employees that leave
Predict customers that churn
Target “best” customers
How can I make more $$?
Which customers are likely to buy?
Who are my “best customers”?
How can I combat fraud?

Be Specific in Problem Statement

Poorly Defined	Better
Predict employees that leave	• Based on past employees that voluntarily left: • Create New Attribute `EmplTurnover` → O/1
Predict customers that churn	• Based on past customers that have churned: • Create New Attribute `Churn` → `YES/NO`
Target "best" customers	• Recency, Frequency Monetary (RFM) Analysis • Specific Dollar Amount over Time Window: • Who has spent $500+ in most recent 18 months
How can I make more $$?	• What helps me sell soft drinks & coffee?
Which customers are likely to buy?	• How much is each customer likely to spend?
Who are my "best customers"?	• What descriptive "rules" describe "best customers"?
How can I combat fraud?	• Which transactions are the most anomalous? • Then roll-up to physician, claimant, employee, etc.

Be Specific in Problem Statement

Poorly Defined	Better	Data Mining Technique
Predict employees that leave	• Based on past employees that voluntarily left: • Create New Attribute EmplTurnover → O/1	
Predict customers that churn	• Based on past customers that have churned: • Create New Attribute Churn → YES/NO	
Target "best" customers	• Recency, Frequency Monetary (RFM) Analysis • Specific Dollar Amount over Time Window: • Who has spent $500+ in most recent 18 months	
How can I make more $$?	• What helps me sell soft drinks & coffee?	
Which customers are likely to buy?	• How much is each customer likely to spend?	
Who are my "best customers"?	• What descriptive "rules" describe "best customers"?	
How can I combat fraud?	• Which transactions are the most anomalous? • Then roll-up to physician, claimant, employee, etc.	

Oracle Advanced Analytics

Brief Demos

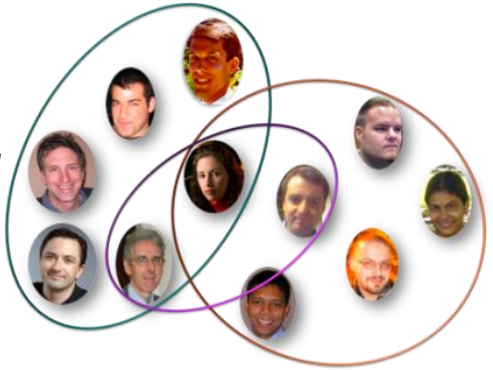
ORACLE

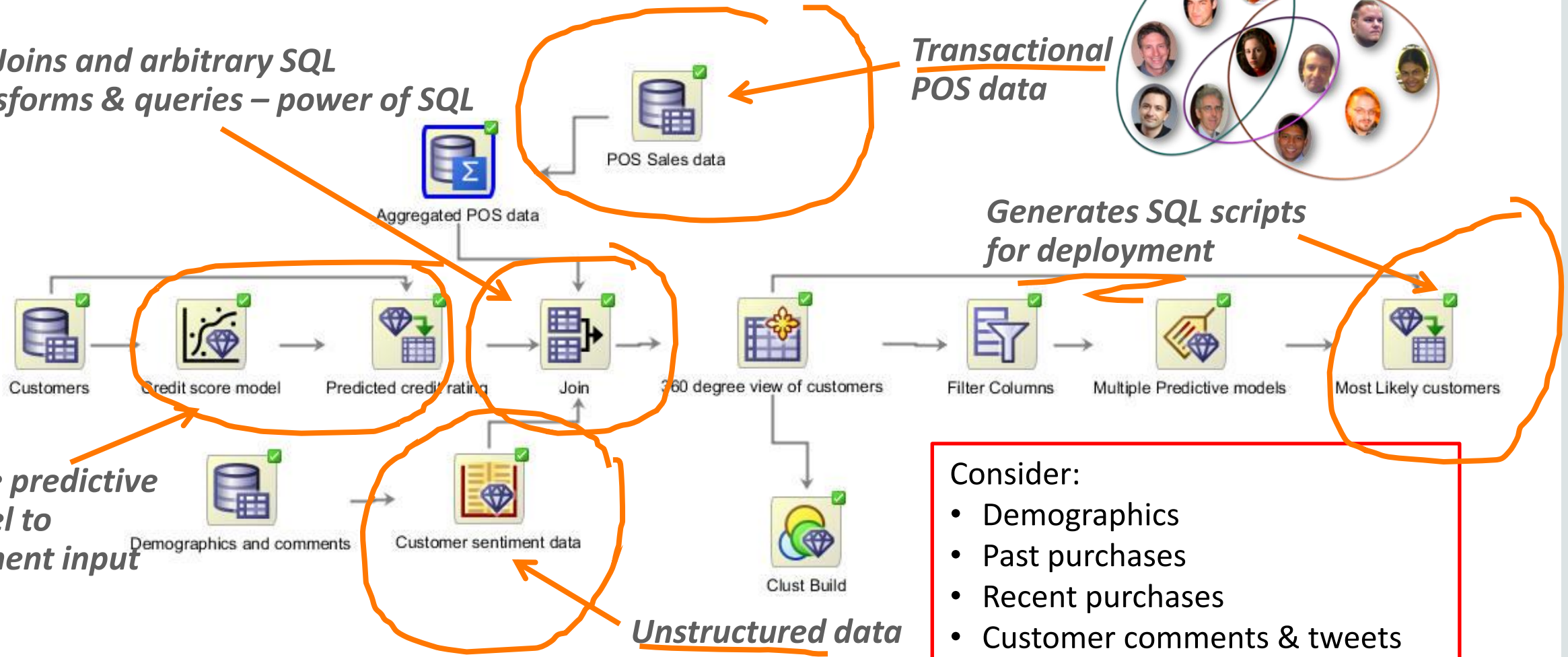
Predicting Behavior

Identify "Likely Behavior" and their Profiles

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

Transactional POS data





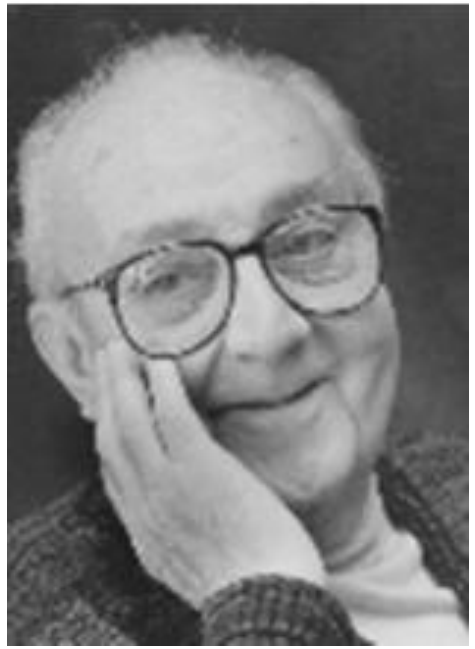
Generates SQL scripts for deployment

Inline predictive model to augment input data

Unstructured data also mined by algorithms

Consider:

- Demographics
- Past purchases
- Recent purchases
- Customer comments & tweets



“Essentially, all models are wrong, …but some are useful.”

– George Box
(One of the most influential statisticians of the 20th century and a pioneer in the areas of quality control, time series analysis, design of experiments and Bayesian inference.)

dunnhumby

Accelerates Complex Segmentation Queries from Weeks to Minutes—Gains Competitive Advantage

Objectives

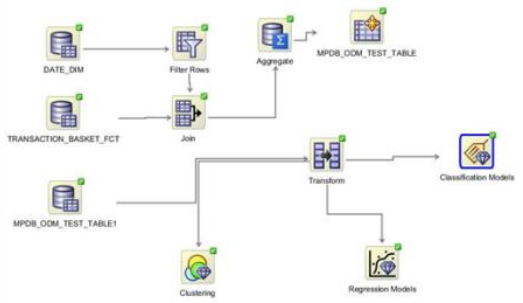
- World's leading customer-science company
- Accelerate analytic capabilities to near real time using Oracle Advanced Analytics and third-party tools, enabling analysis of unstructured big data from emerging sources, like smart phones

Solution

- Accelerated segmentation and customer-loyalty analysis from one week to just four hours—enabling the company to deliver more timely information & finer-grained analysis
- Generated more accurate business insights and marketing recommendations with the ability to analyze 100% of data—including years of historical data—instead of just a small sample

- "Improved analysts' productivity and focus as they can now run queries and complete analysis without having to wait hours or days for a query to process"
- "Improved accuracy of marketing recommendations by analyzing larger sample sizes and predicting the market's reception to new product ideas and strategies"

– dunnhumby Oracle Customer Snapshot)



ORACLE

Customer Snapshot: http://www.oracle.com/us/corporate/customers/customersearch/dunnhumby-1-exadata-ss-2137635.html

Market Basket & Advanced Analytics at Dunkin Brands

Objectives

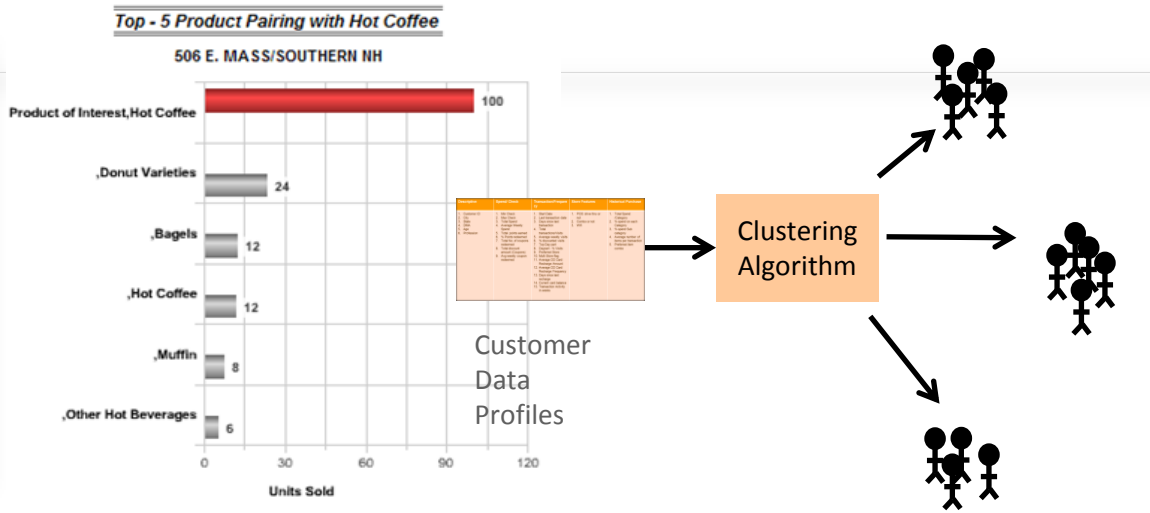
- Store development dashboards to identify opportunities
- 8 M daily transactions, ~25M transaction detail lines
- 20 TB data warehouse size, sales data about 10 TB
- Market basket analysis and customer loyalty & segmentation

Solution

- Exadata Engineered Systesm
- Oracle Advanced Analytics Option
- Market Basket Analysis, Clustering, Classification, Segmentation, Loyalty Analysis

- "Exponential growth in combinations with each hierarchy. 2 years of pre-computed Market Baskets and associated sales measures for reporting. Nightly compute within ETL window data with 1 day latency."
  - Dunkin Brands, Mahesh Jagannath, Senior Manager, Business Intelligence

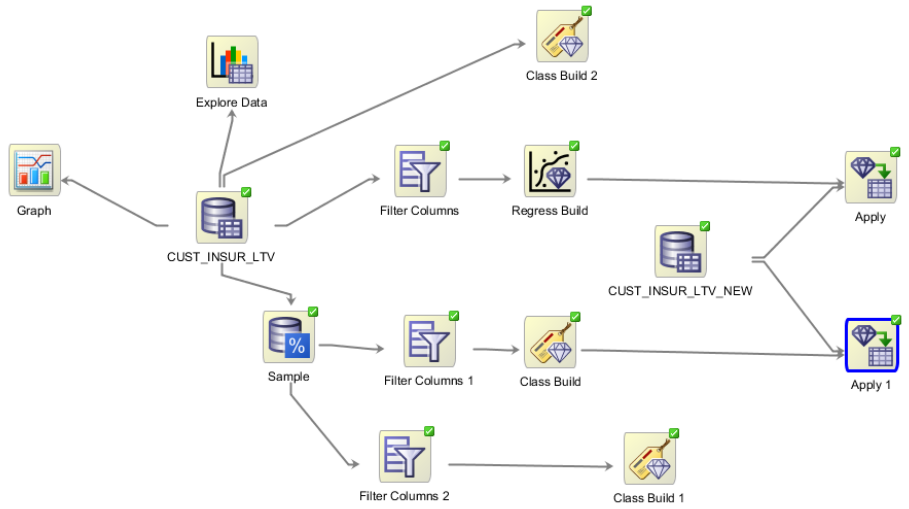
  (Excerpts from Dunkin Brands presentation at Oracle Open World 2014)

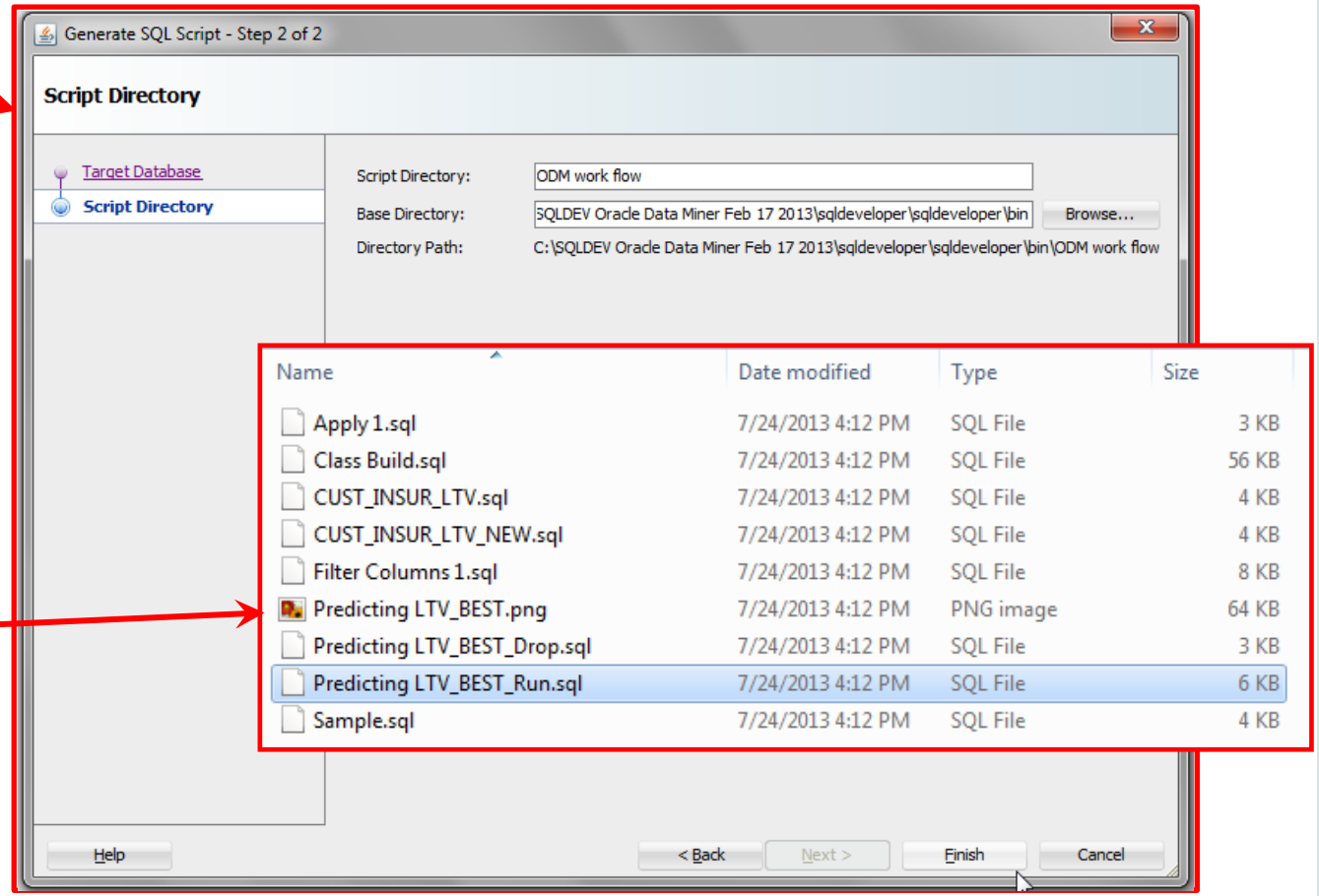


SQL Developer/Oracle Data Miner 4.0

New Features

- **SQL Script Generation**
  - Deploy entire methodology as a SQL script
  - Immediate deployment of data analyst's methodologies





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;
```

POLICYNUMBER	PERCENT_FRAUD	RNK
6532	64.78	1
2749	64.17	2
3440	63.22	3
654	63.1	4
12650	62.36	5

Automated Monthly "Application"! *Just add:*

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

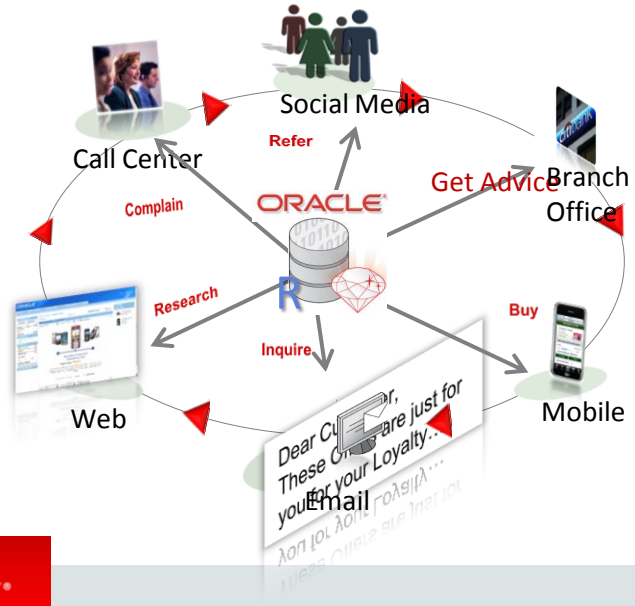
Time measure: set timing on;

Oracle Advanced Analytics

More Details

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

```
Select prediction_probability(CLAS_DT_1_2, 'Yes'
  USING 7800 as bank_funds, 125 as checking_amount, 20 as
  credit_balance, 55 as age, 'Married' as marital_status,
  250 as MONEY_MONTLY_OVERDRAWN, 1 as house_ownership)
from dual;
```



Likelihood to respond:

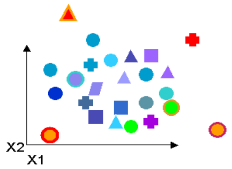
Query Result

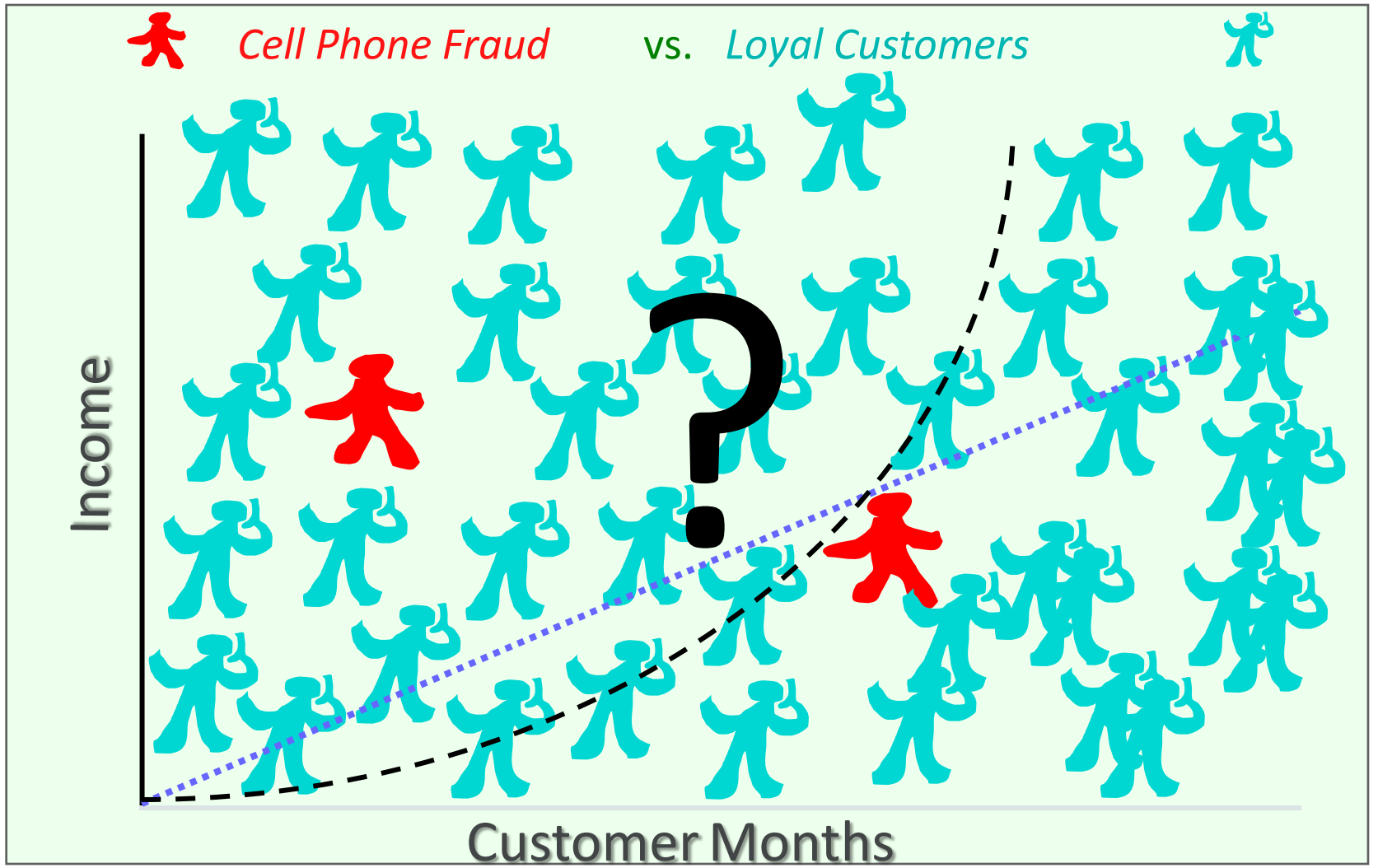
SQL | All Rows Fetched: 1 in 0 seconds

PREDICTION_PROB...
0.8382936507936...

Data Mining When Lack Examples

Better Information, Valuable Insights and Predictions

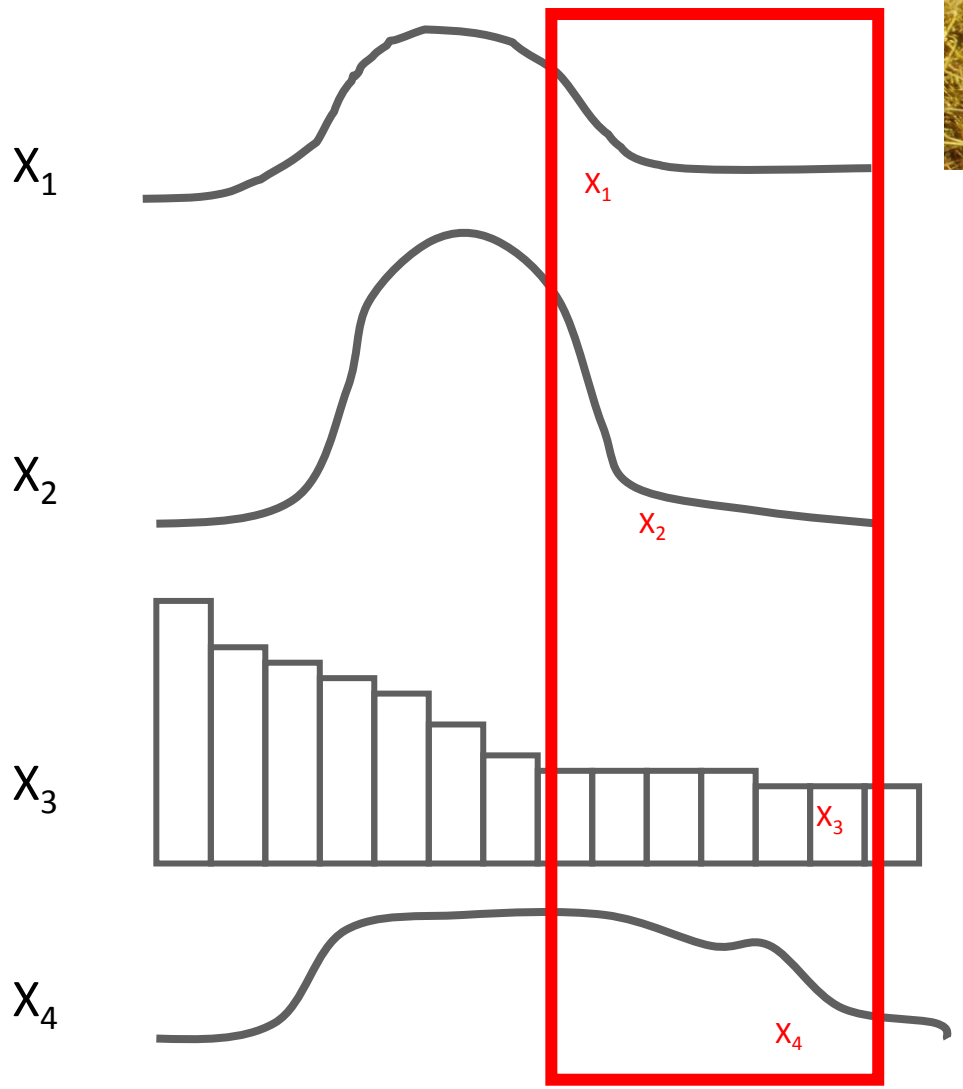


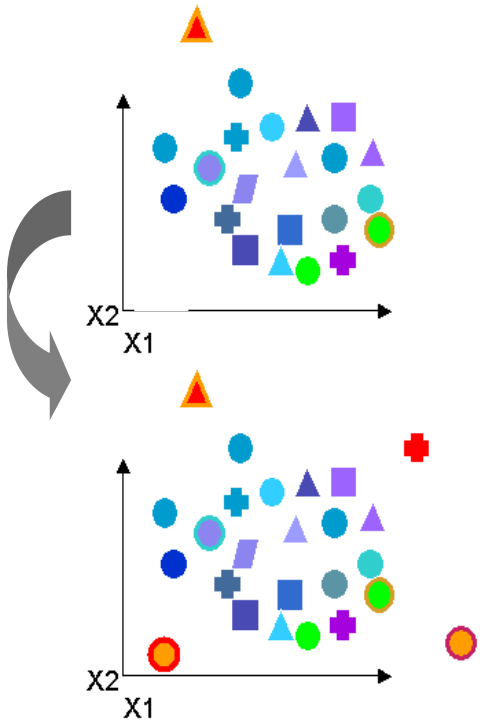


Source: Inspired from *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* by Michael J. A. Berry, Gordon S. Linoff

Challenge: Finding Anomalies

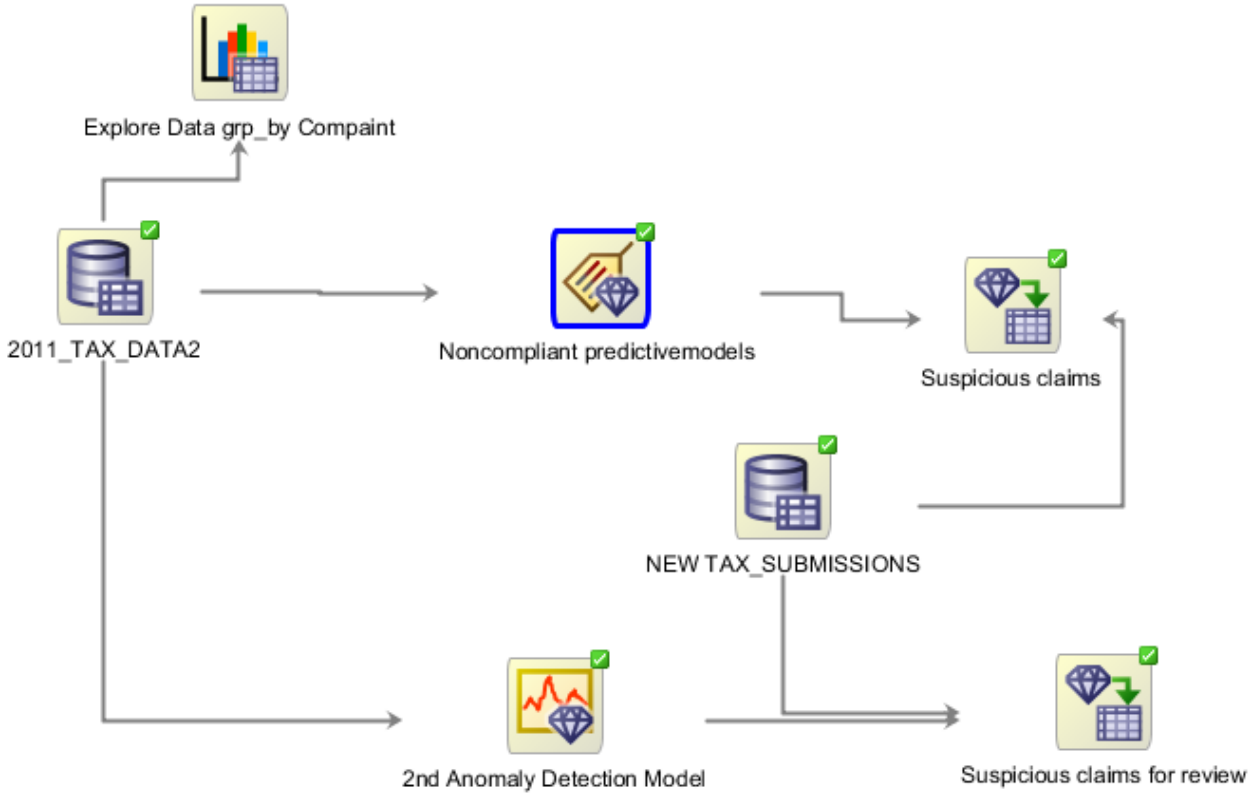
- Considering multiple attributes
- Taken alone, may seem “normal”
- Taken collectively, a record may appear to be anomalous
- Look for what is “*different*”





Tax Noncomplaince Audit Selection

- Simple Oracle Data Mining predictive model
  - Uses Decision Tree for classification of Noncompliant tax submissions (yes/no) based on historical 2011 data



Oracle Advanced Analytics

OAA/Oracle R Enterprise (R integration)

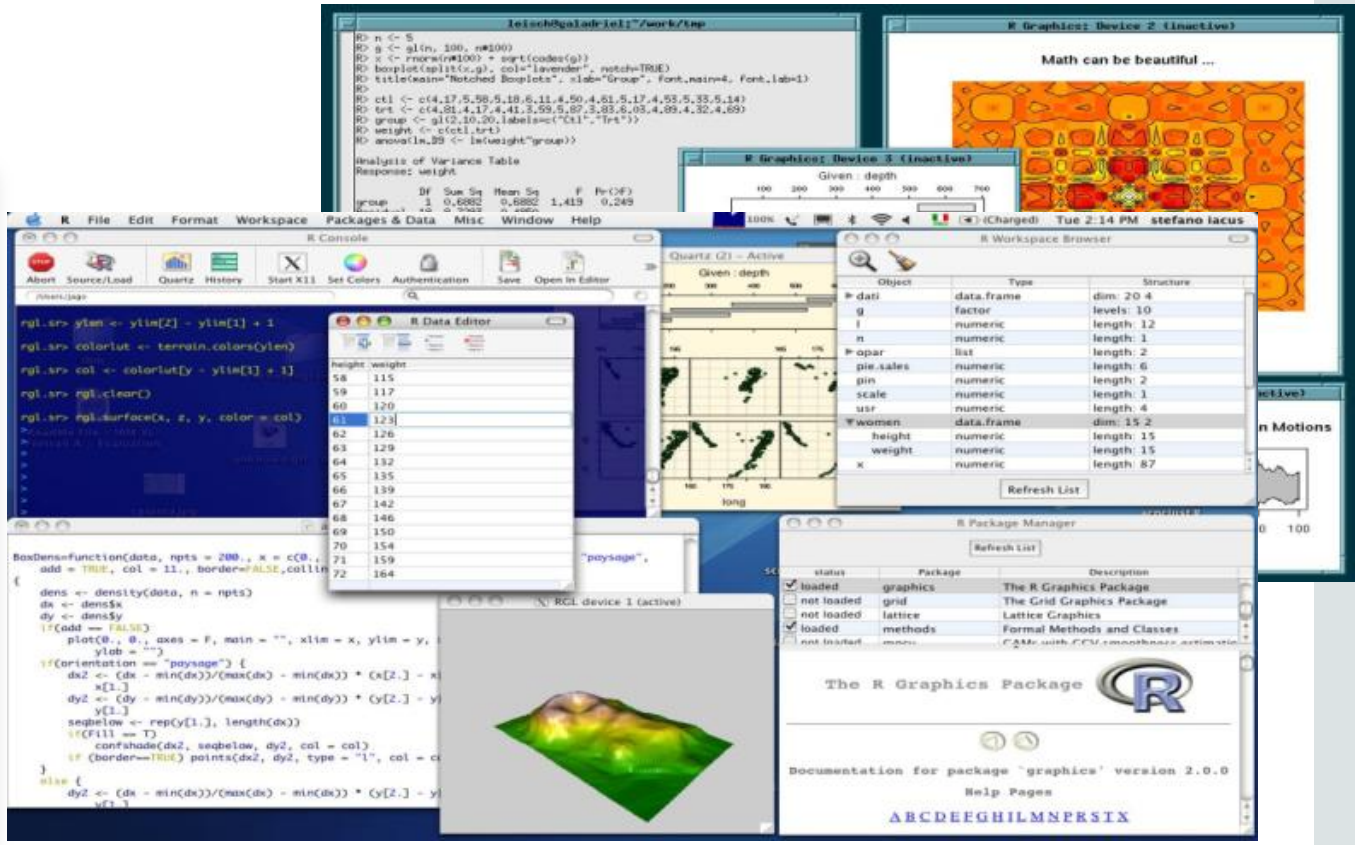
ORACLE

R—Widely Popular

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

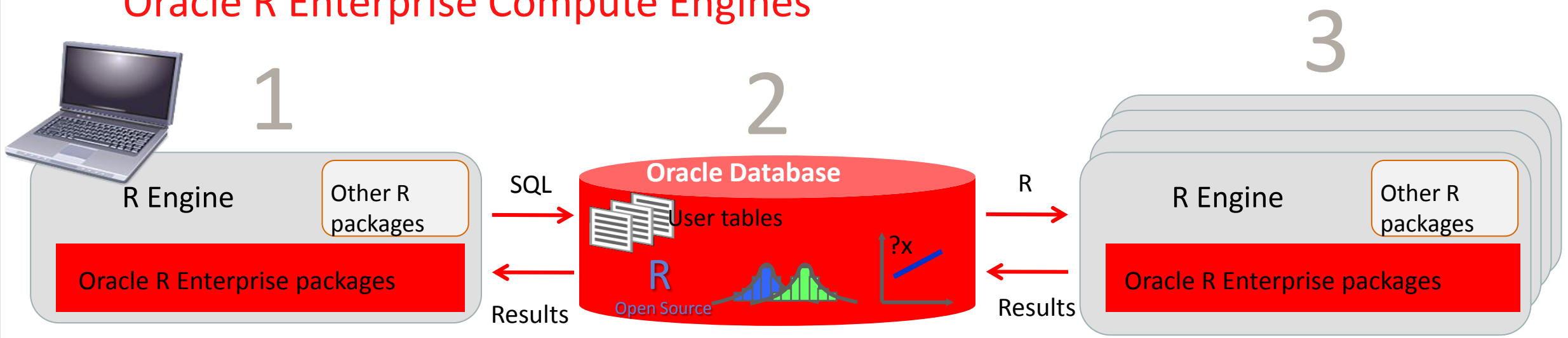
R environment

- Strengths
  - Powerful & Extensible
  - Graphical & Extensive statistics
  - Free—open source
- Challenges
  - Memory constrained
  - Single threaded
  - Outer loop—slows down process
  - Not industrial strength



Oracle Advanced Analytics

Oracle R Enterprise Compute Engines



User R Engine on desktop

- R-SQL Transparency Framework overloads R functions for scalable in-database execution
- Function overload for data transforms, statistical functions and advanced analytics
- 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

Database Compute Engine

- Scale to large datasets
- Access tables, views, and external tables, as well as data through DB LINKS
- Leverage database SQL parallelism
- Leverage new and existing in-database statistical and data mining capabilities

R Engine(s) spawned by Oracle DB

- Database can spawn multiple R engines for database-managed parallelism
- Efficient data transfer to spawned R engines
- Emulate map-reduce style algorithms and applications
- Enables production deployment and automated execution of R scripts

ORACLE

R: Transparency through function overloading

Invoke in-database aggregation function

R Console

```
Oracle Distribution of R version 3.0.1 (2012-06-22) -- "Good Sport"

> 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
```

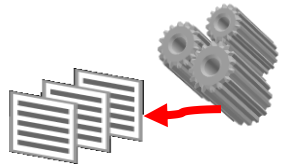
Oracle Advanced Analytics
ORE Client Packages

Transparency Layer

Oracle SQL

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

Oracle Database



In-db Stats

ONTIME_S

Database Server

R: Transparency through function overloading

Invoke in-database Data Mining model (Support Vector Machine)

R Console

```
Oracle Distribution of R version 3.0.1 (2012-06-22) -- "Good Sport"
> 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:
                        value
prep.auto                  on
active.learning    al.enable
complexity.factor 46.044899
conv.tolerance         1e-04
kernel.function       linear
```

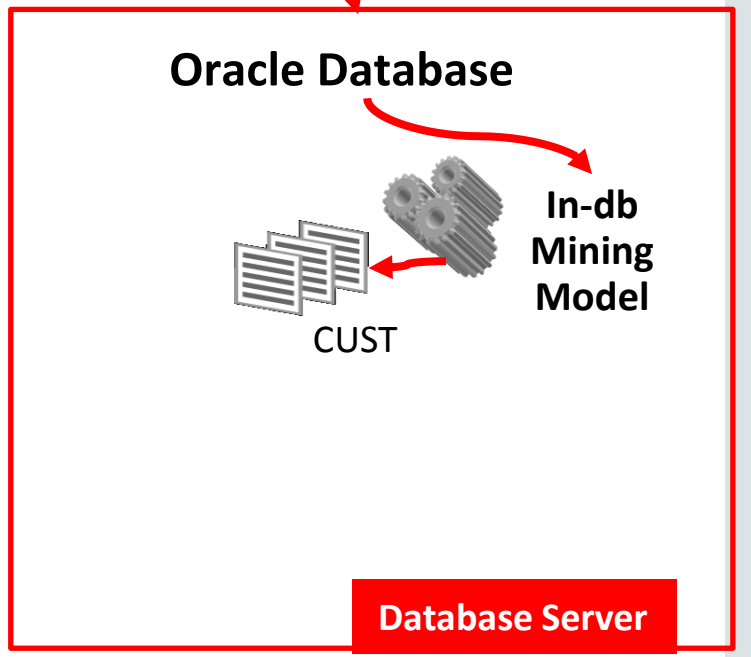
```
Coefficients:
   class       variable value      estimate
1      0          INCOME         5.204561e-05
2      0 MARITAL_STATUS      M -4.531359e-05
3      0 MARITAL_STATUS      S  4.531359e-05
4      0       YRS_CUST         1.264948e-04
5      0    (Intercept)         9.999269e-01
6      1          INCOME         2.032340e-05
7      1 MARITAL_STATUS      M  2.636552e-06
8      1 MARITAL_STATUS      S -2.636555e-06
9      1       YRS_CUST        -1.588211e-04
10     1    (Intercept)        -9.999324e-01
```

Oracle Advanced Analytics
ORE Client Packages

Transparency Layer

Oracle PL/SQL

```
BEGIN
DBMS_DATA_MINING.CREATE_MODEL(
 model_name => ' SVM_MOD' ,
 mining_function =>
dbms_data_mining.classification
...
```

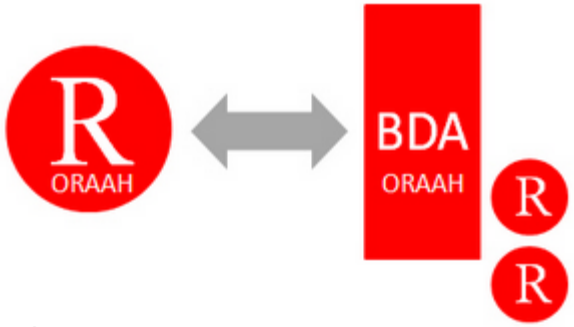


Oracle Advanced Analytics for Hadoop

Predictive algorithms that execute in a parallel/distributed manner on Hadoop with data in HDFS

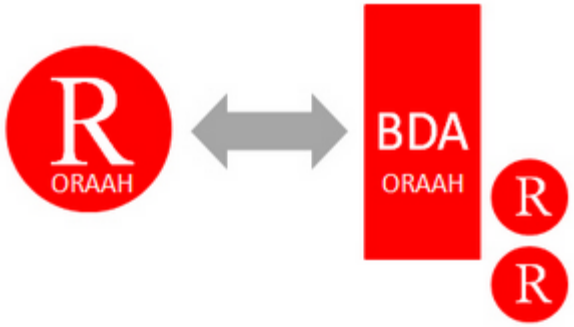
ORACLE

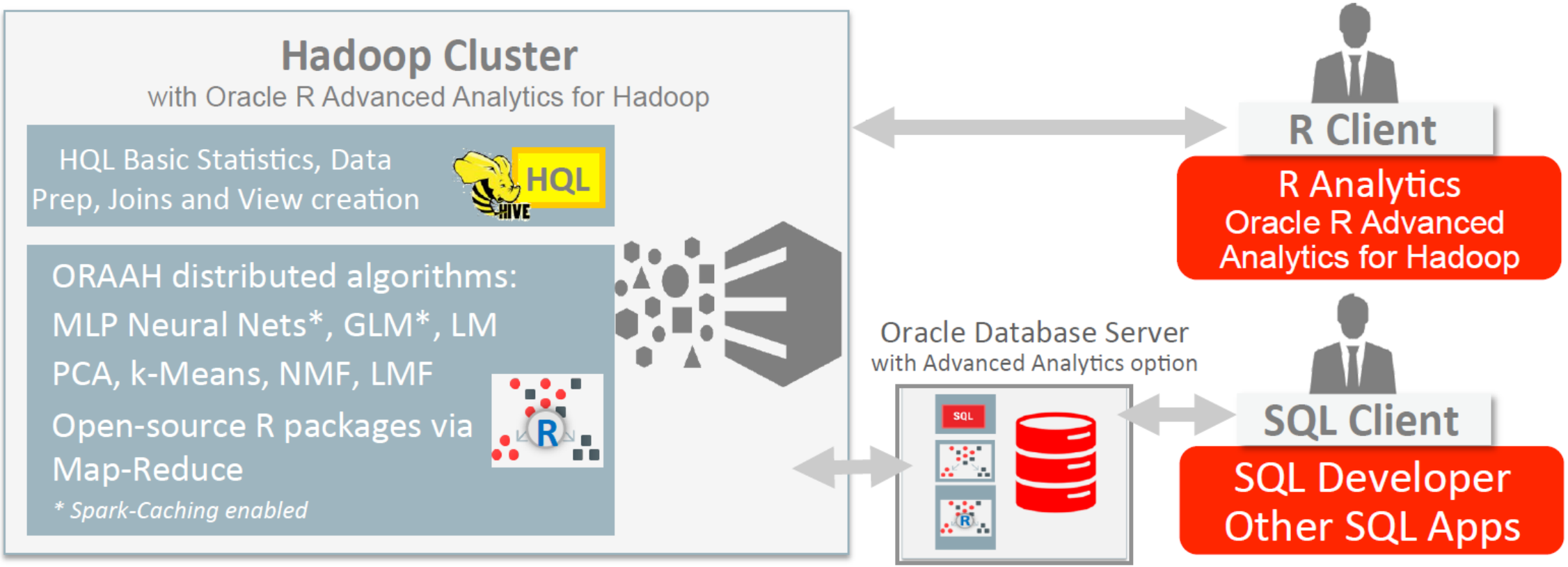
Oracle R Advanced Analytics for Hadoop



- ORAAH = Oracle R Advanced Analytics for Hadoop, part of Big Data Software Connectors Suite (Oracle Big Data Appliance Option)
- ORAAH transparency layer enables certain overloaded R functions to operate on Hive tables using R syntax and behavior (transparently translating R to HiveQL)
- R interface for manipulating HDFS data and writing mapper and reducer functions in R – where you can leverage open source CRAN packages – and invoke those Hadoop jobs from R
- Provides a range of predictive algorithms that execute on the Hadoop cluster with data in HDFS in a parallel/distributed manner.

Oracle R Advanced Analytics for Hadoop



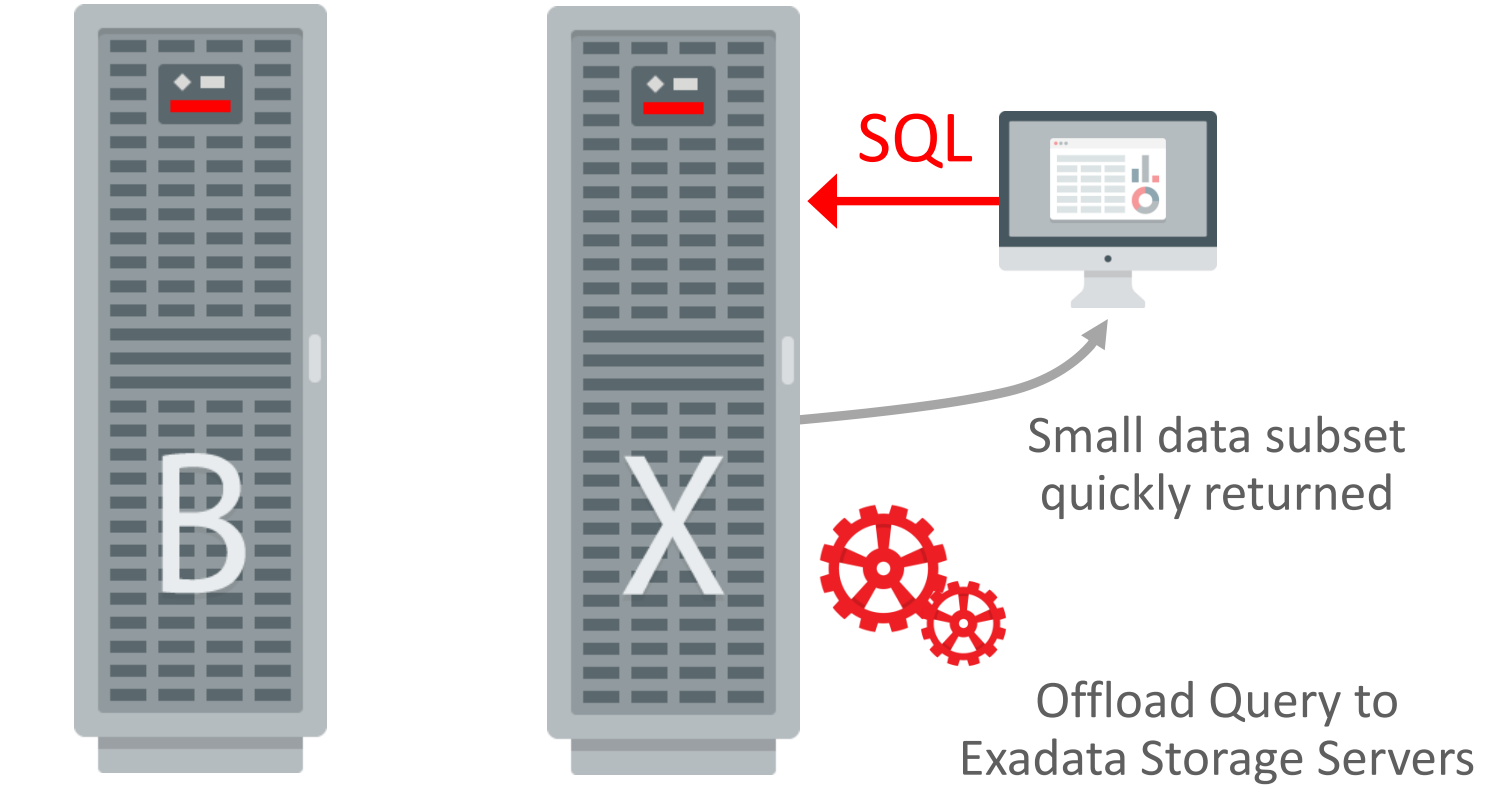


Big Data SQL

Push down SQL predicts to storage layers

ORACLE

What gives Exadata extreme performance?

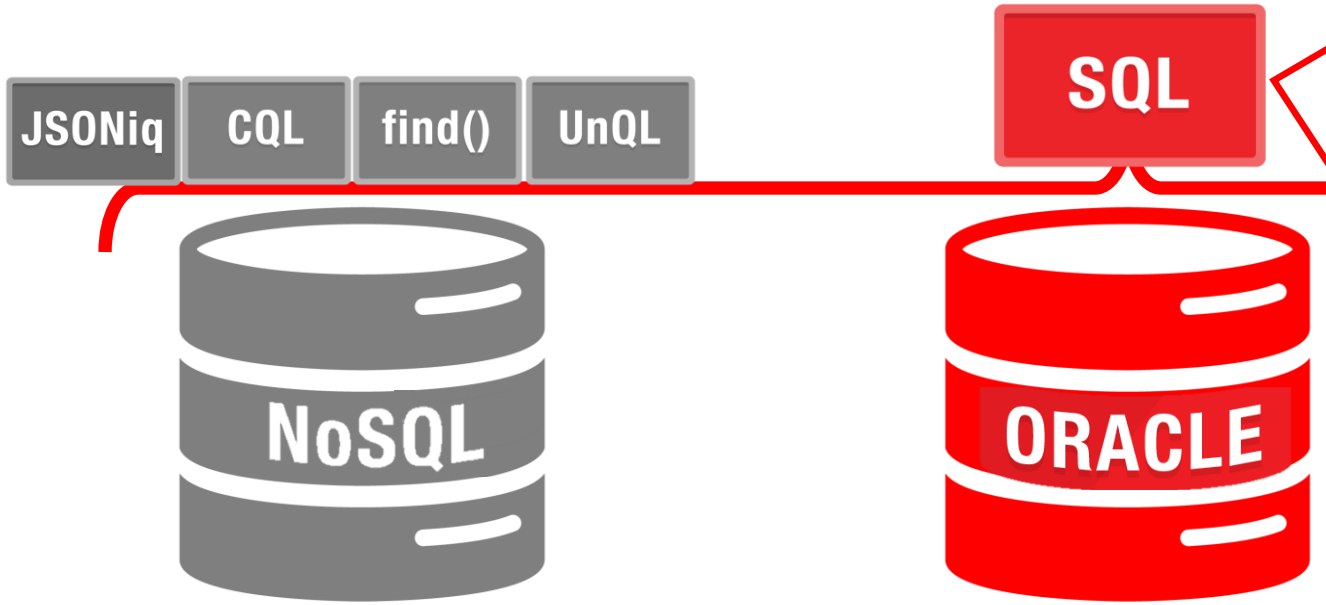


Data Analytics Challenge

Separate silos with separate data access interfaces

What customers want: Oracle Big Data SQL

Rich, comprehensive SQL access to all enterprise data



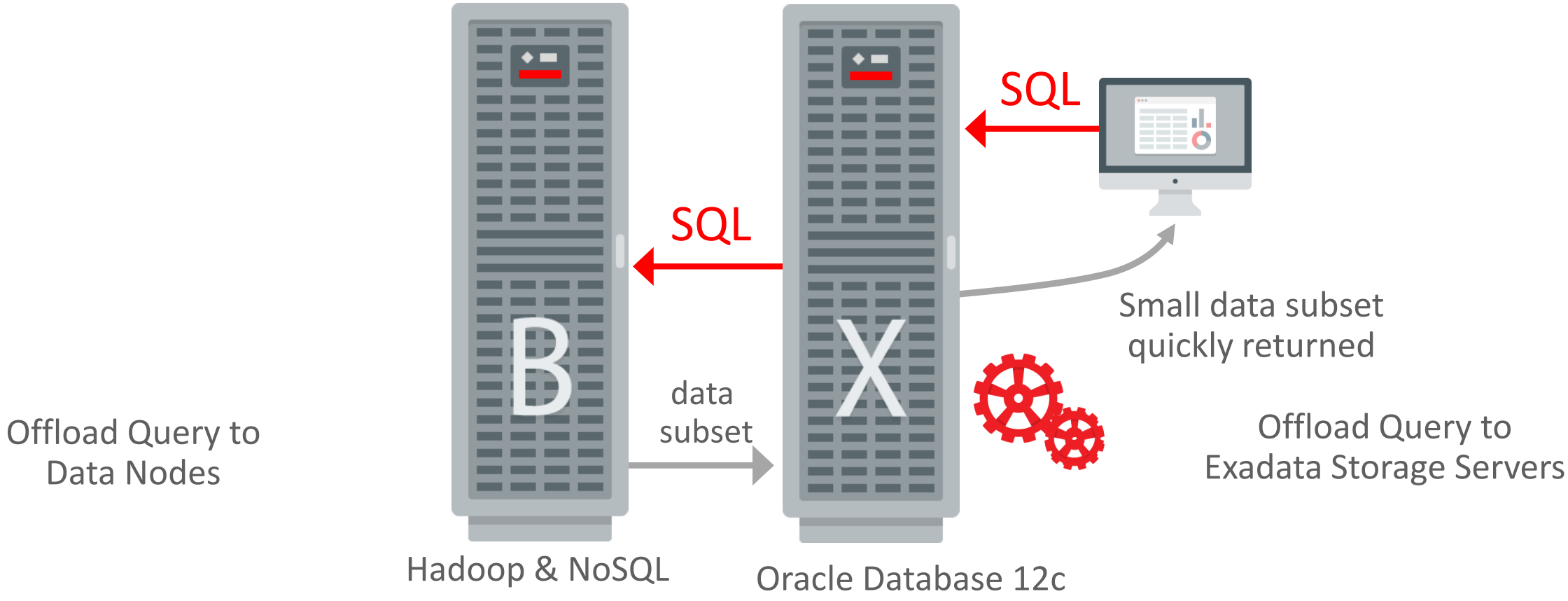
The Power of Oracle SQL

- Wide variety of 'Big Data' types
  - Structured data
    - Numeric, string, date, …
  - Unstructured data
    - LOBs, Text, XML, JSON, Spatial, Graph, Multimedia

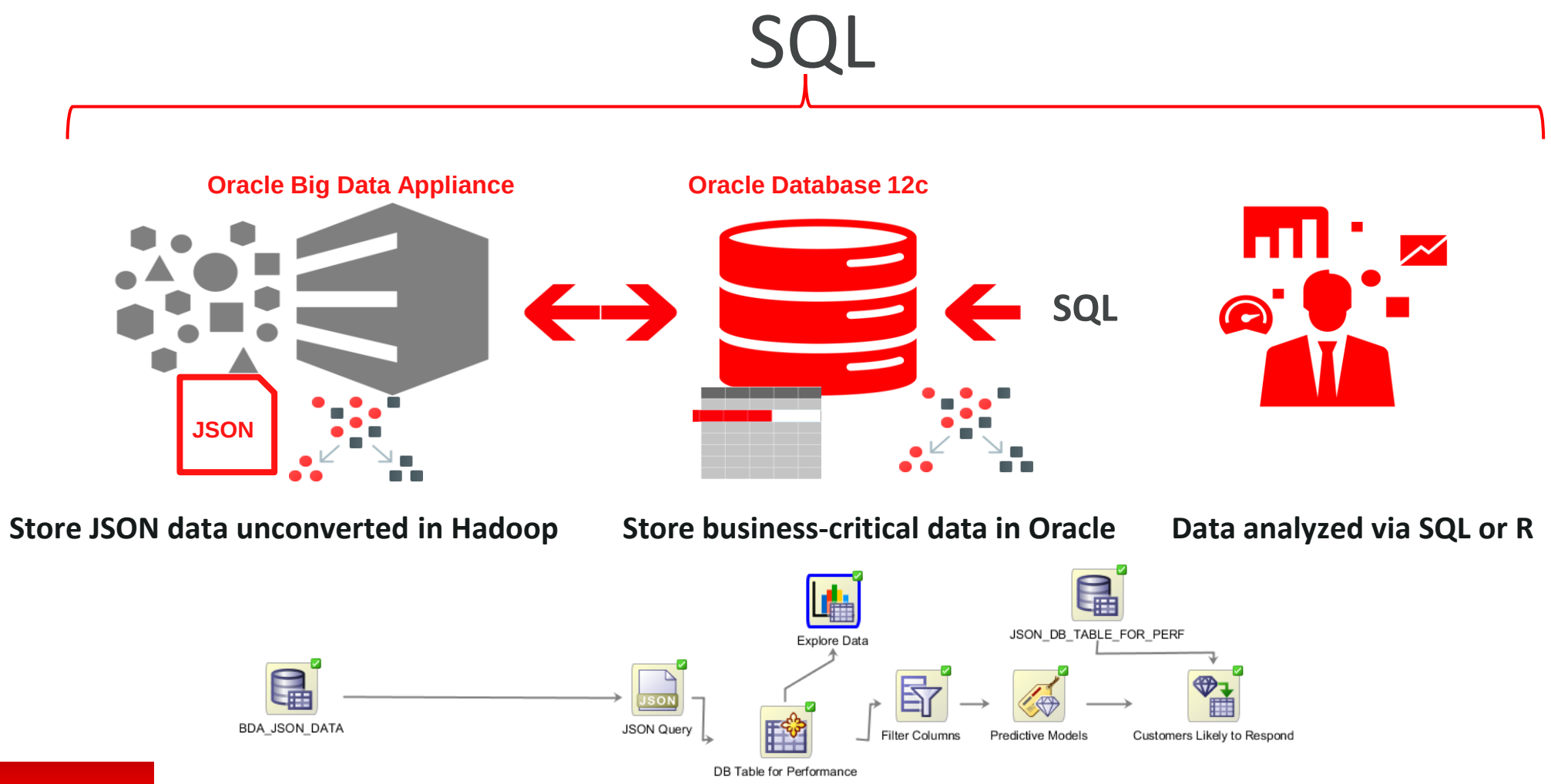
- Rich SQL Analytic Functions
  - Ranking, Windowing, LAG/LEAD, Aggregate, Pattern Matching, Cross Tabs, Statistical, Linear Regression, Correlations, Hypothesis Testing, Distribution Fitting, …

Introducing Oracle Big Data SQL

Massively Parallel SQL Query across Oracle, Hadoop and NoSQL



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

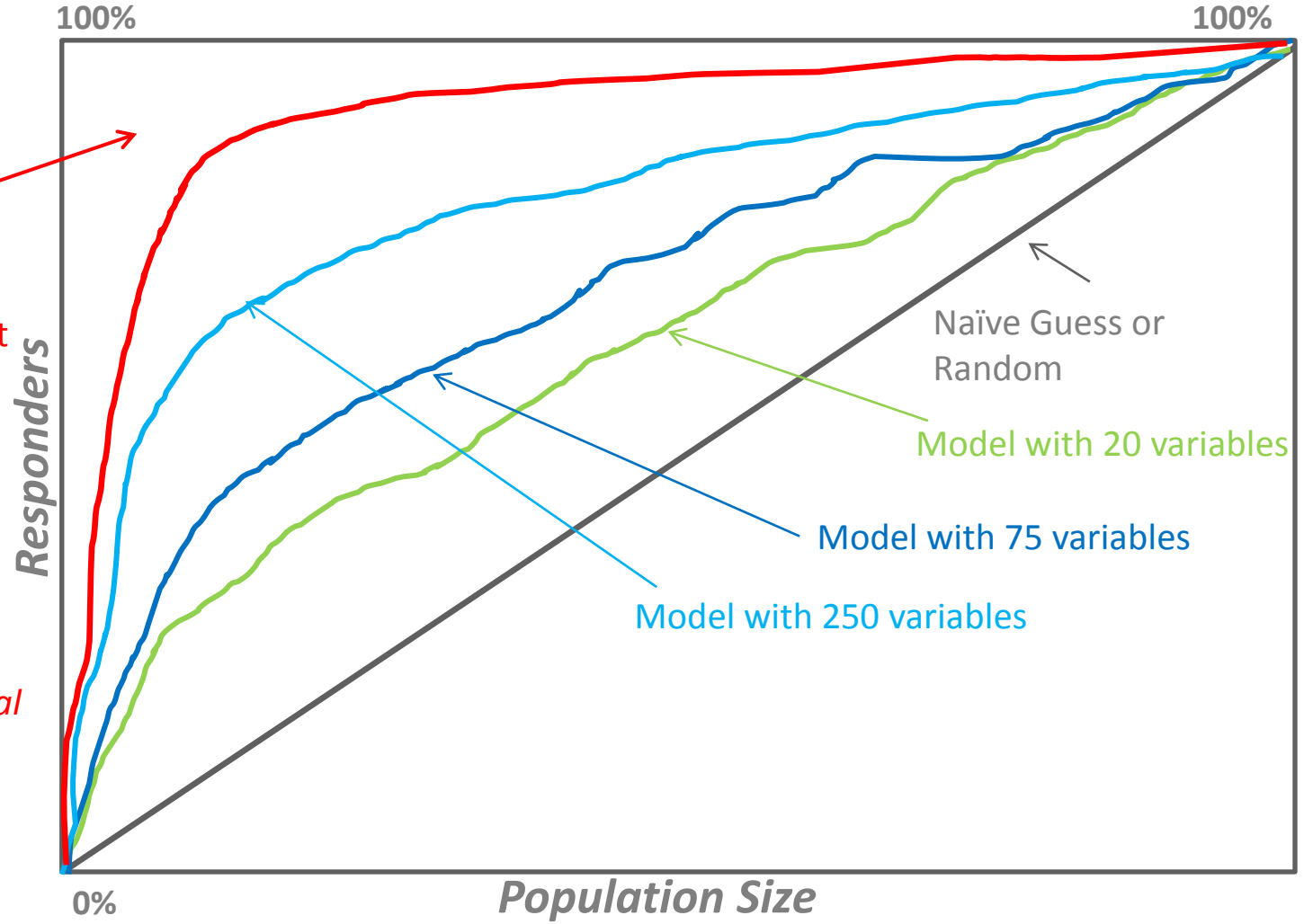


More Data Variety—Better Predictive Models

- 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.*



ORACLE

Wargaming Creates Complex Analytical Models in Minutes, Ensures Superior Gaming Experience

Objectives

- Online game developer, publisher, and leader in free-to-play massively multiplayer online game market
- Implement a scalable and flexible data warehouse
- Ensured optimal game play for more than 110 million players
- Expand insight into game health, market health, and business

Solution

- Ran predicative analytics model in just three minutes instead of more than six hours
- Enabled developers to adapt game, based on where players experience most exciting play
- Expanded insight with increased intelligence on customer acquisition and loyalty as well as identifying potential monetization opportunities

- "Oracle Big Data Appliance gives us unprecedented insight into game, business, and market health. The possibilities are endless with this highly extensible solution that enables us to gather, analyze, and use data, including social-media data, in ways that were simply not possible before."

– Craig Fryar, Head of Wargaming Business Intelligence

Oracle Big Data Appliance, Oracle Database Appliance, Oracle Advanced Analytics, Oracle Big Data Connectors

Customer Snapshot: http://www.oracle.com/us/corporate/customers/customersearch/wargaming-1-bda-ss-2408474.html

Oracle Advanced Analytics

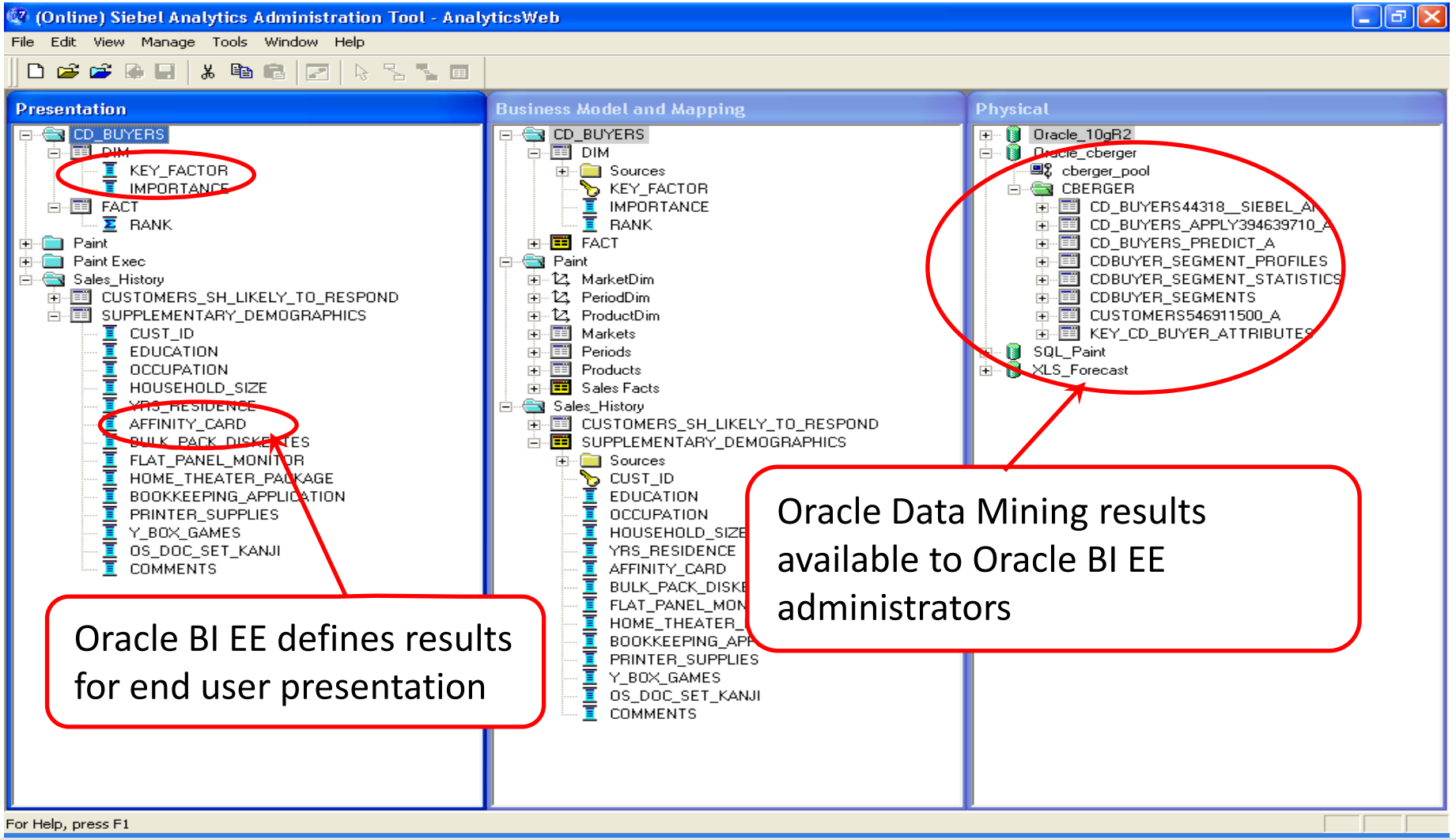
Applications Integration + OBIEE Integration

ORACLE

Integrated Business Intelligence

Enhance Dashboards with Predictions and Data Mining Insights

- In-database predictive models “mine” customer data and predict their behavior
- OBIEE’s integrated spatial mapping shows location
- All OAA results and predictions available in Database via OBIEE Admin to enhance dashboards

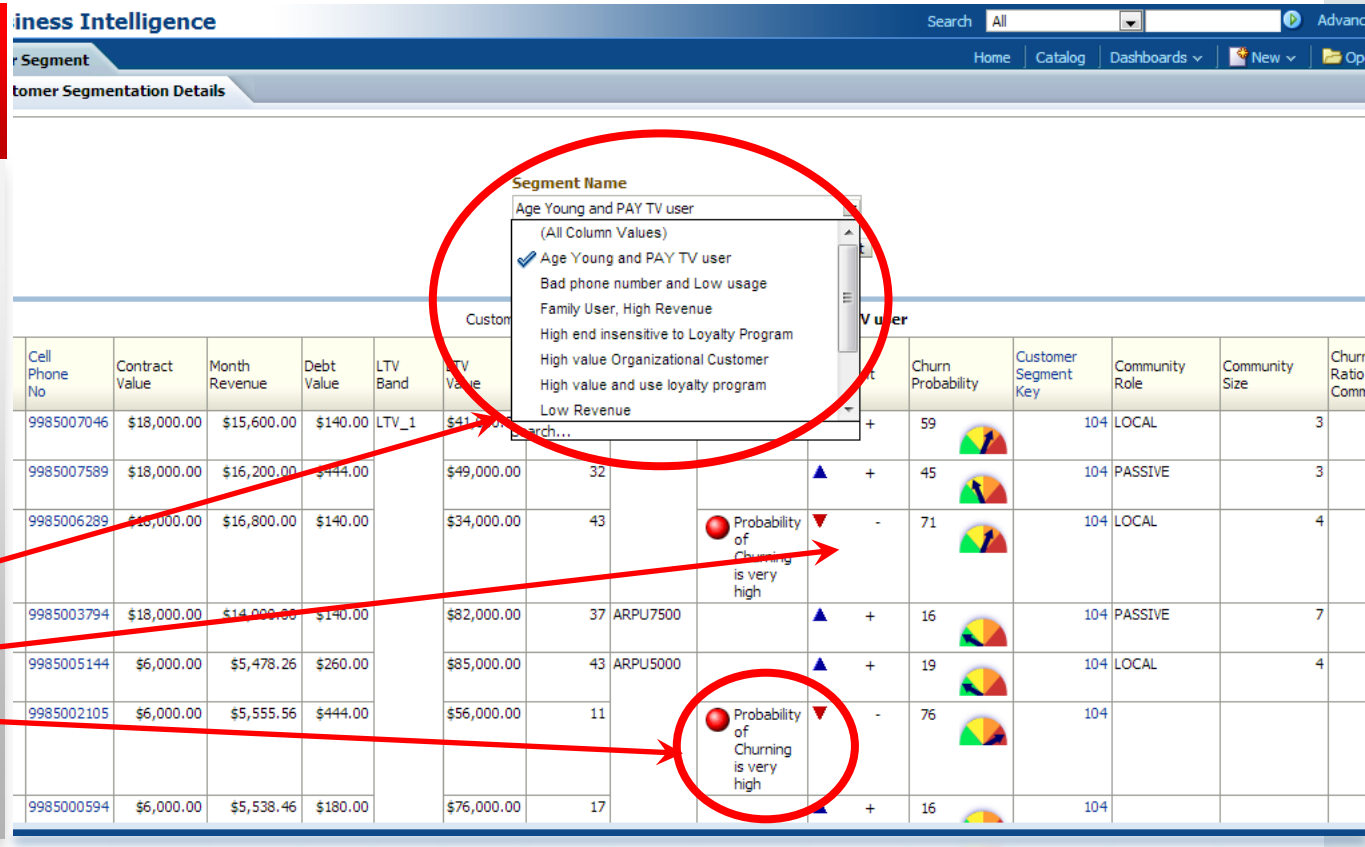


Oracle Communications Industry Data Model

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



Cell Phone No	Contract Value	Month Revenue	Debt Value	LTV Band	Churn Probability	Customer Segment Key	Community Role	Community Size
9985007046	$18,000.00	$15,600.00	$140.00	LTV_1	59	104	LOCAL	3
9985007589	$18,000.00	$16,200.00	$444.00		45	104	PASSIVE	3
9985006289	$18,000.00	$16,800.00	$140.00		71	104	LOCAL	4
9985003794	$18,000.00	$14,000.00	$140.00		16	104	PASSIVE	7
9985005144	$6,000.00	$5,478.26	$260.00		19	104	LOCAL	4
9985002105	$6,000.00	$5,555.56	$444.00		76	104		
9985000594	$6,000.00	$5,538.46	$180.00		16	104		

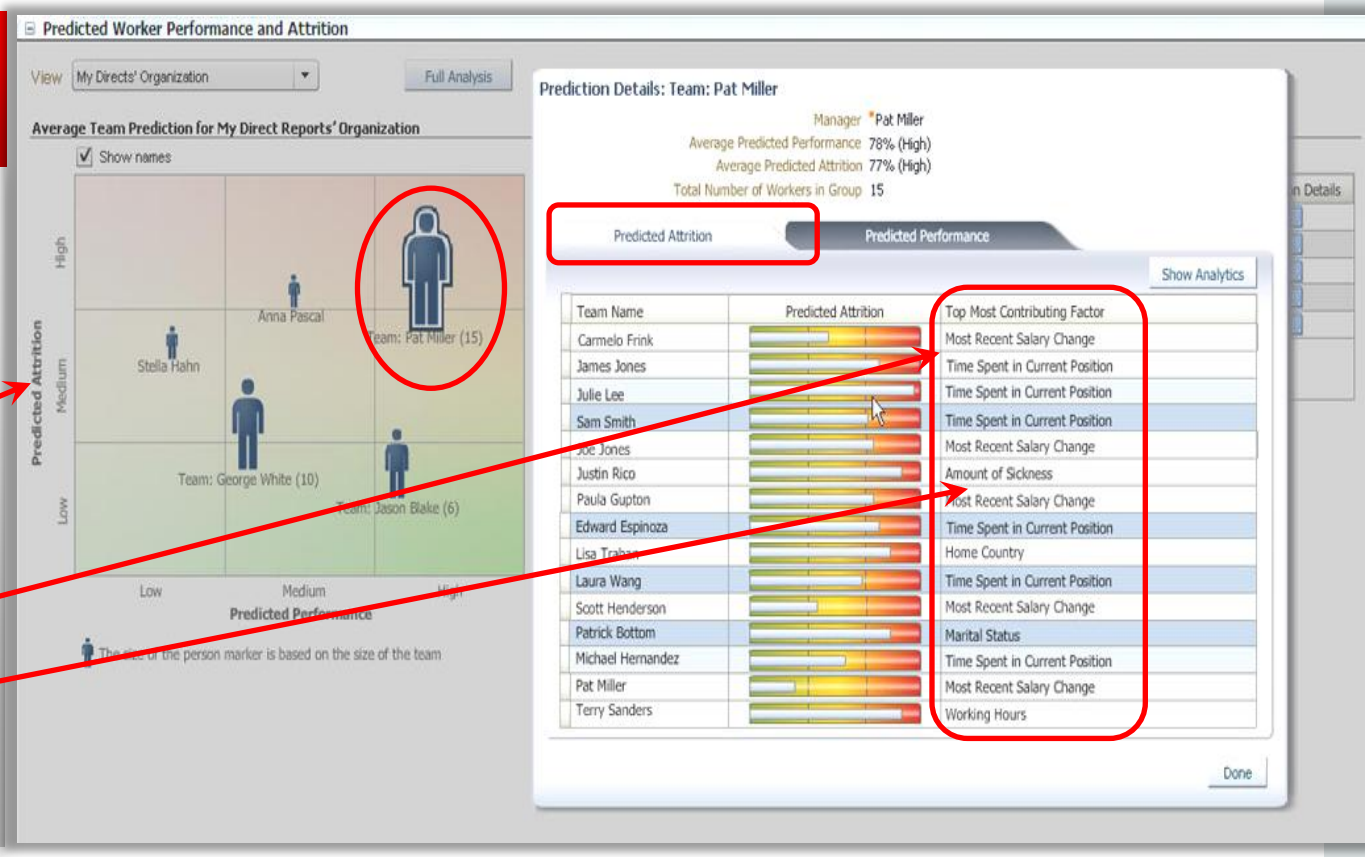
LTV Value		
$49,000.00	32	
$34,000.00	43	
$82,000.00	37	ARPU7500
$85,000.00	43	ARPU5000
$56,000.00	11	
$76,000.00	17	

Fusion HCM Predictive Workforce

Predictive Analytics Applications

Fusion Human Capital Management Powered by OAA

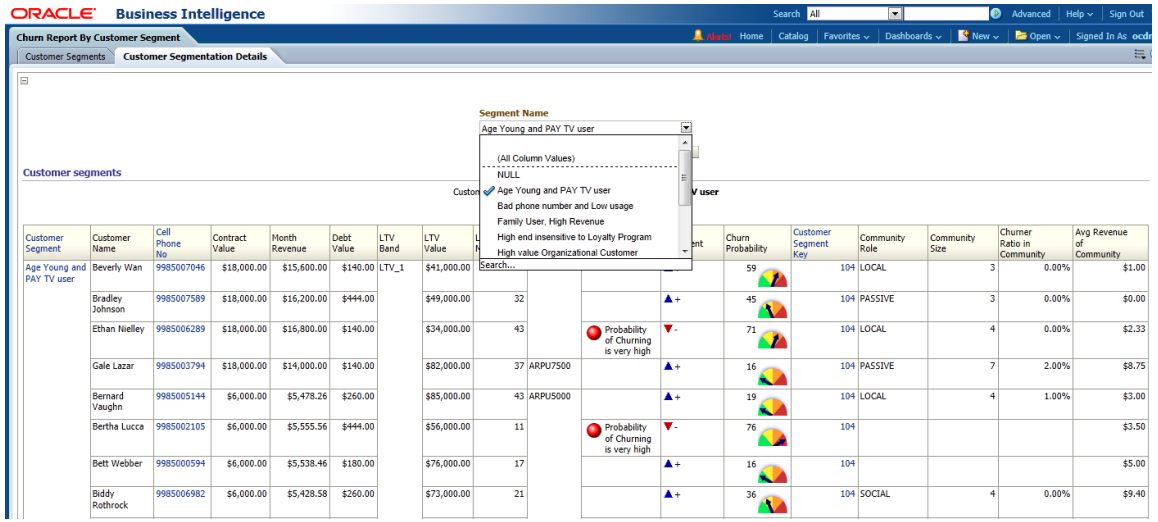
- Oracle Advanced Analytics factory-installed predictive analytics
- Employees likely to leave and predicted performance
- Top reasons, expected behavior
- Real-time "What if?" analysis

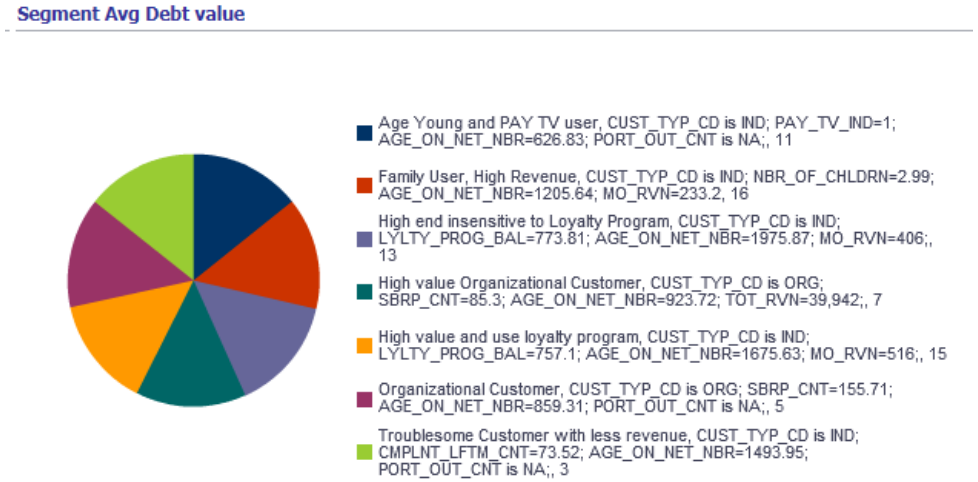


Oracle Communications Data Model

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



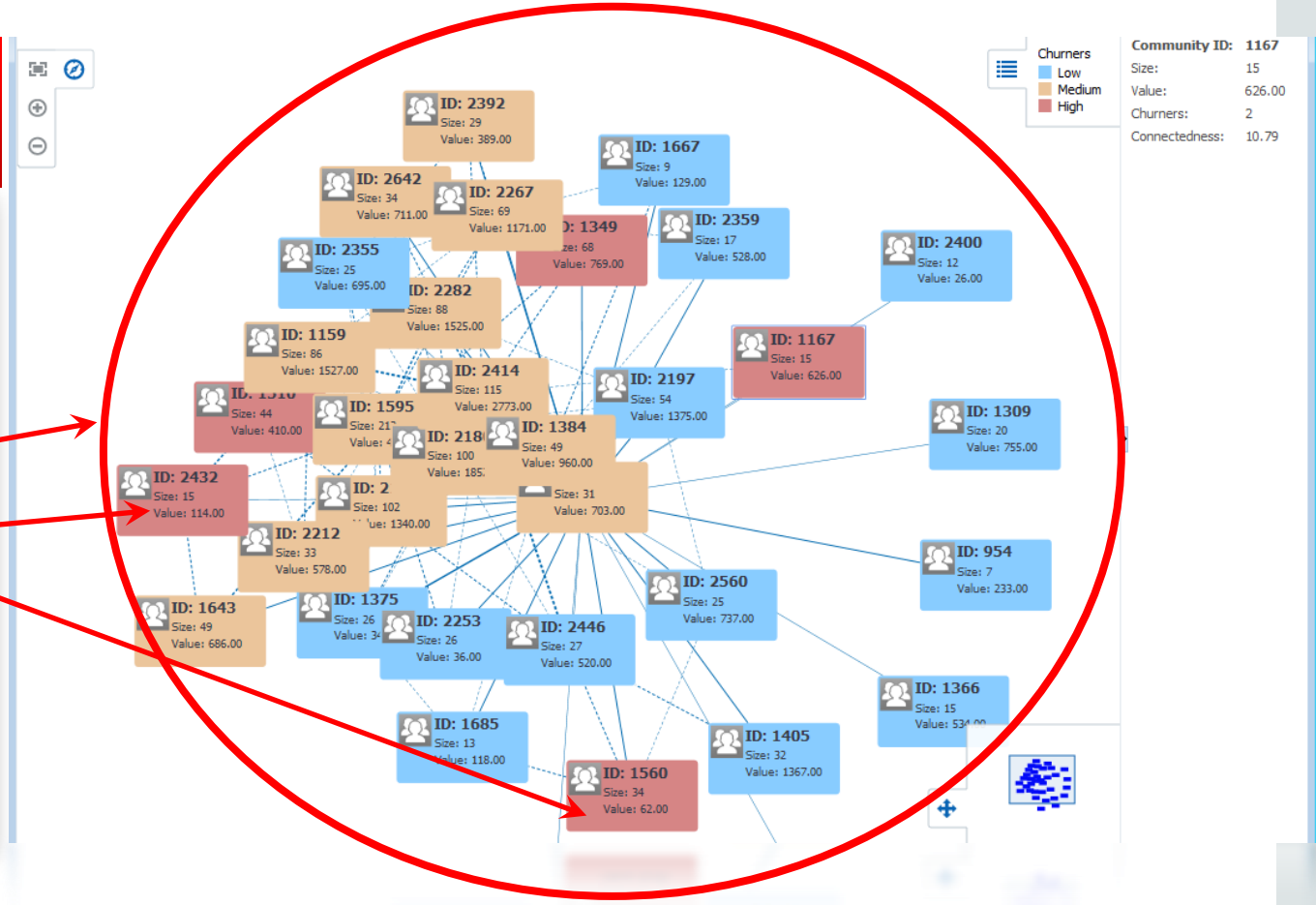


Oracle Communications Industry Data Model

Predictive Analytics Applications

OCDM Telco Churn Enhanced by SNA Analysis

- Integrated with OCDM, OBIEE, and leverages Oracle Data Mining with specialized SNA code
- Identification of social network communities from CDR data
- 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



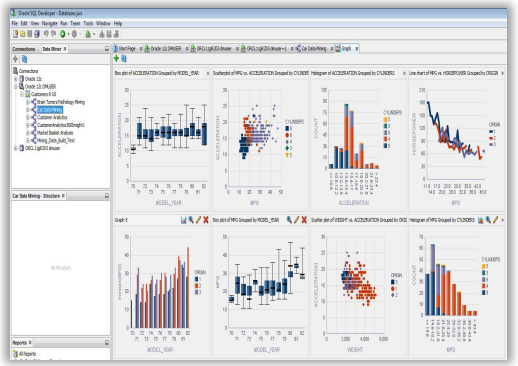
12c New Features

ORACLE

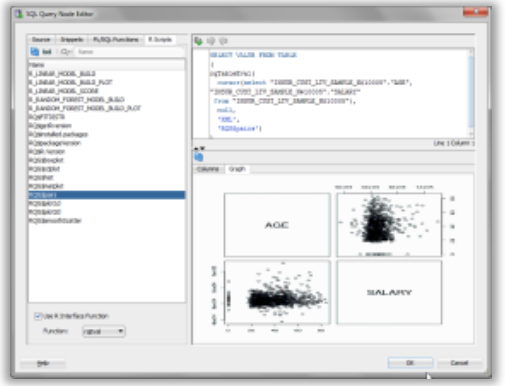
Oracle Advanced Analytics Database Option

Oracle Data Miner 4.X Summary New Features

- Oracle Data Miner/SQLDEV 4.1 EA2 (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
  - **JSON Query node** to mine Big Data external tables



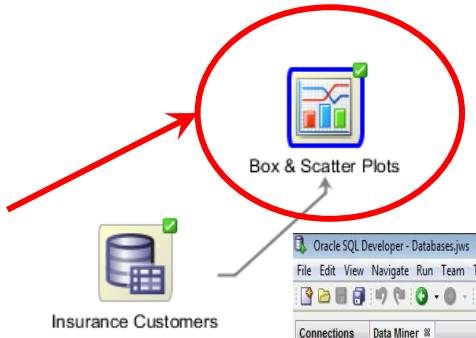
- Oracle Advanced Analytics 12*c* features exposed in Oracle Data Miner
  - New SQL data mining algorithms/enhancements
    - Expectation Maximization clustering algorithm
    - PCA & Singular Vector Decomposition algorithms
    - Improved/automated Text Mining, Prediction Details and other algorithm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

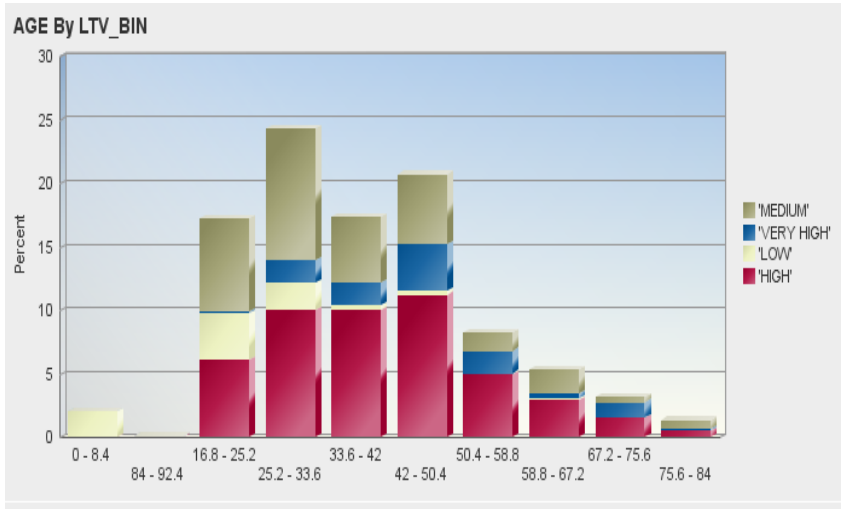


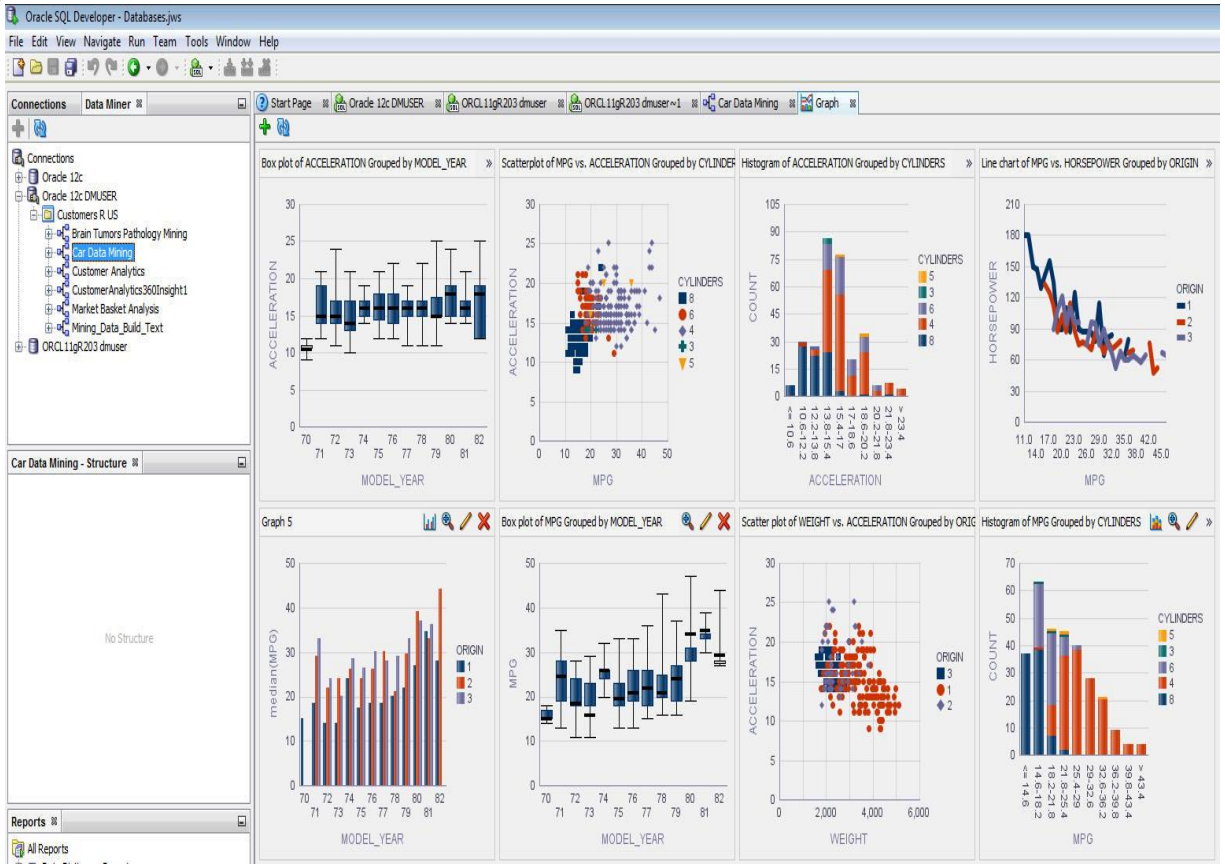
SQL Developer/Oracle Data Miner 4.0

New Features

- **Graph node**
  - Scatter, line, bar, box plots, histograms
  - Group_by supported



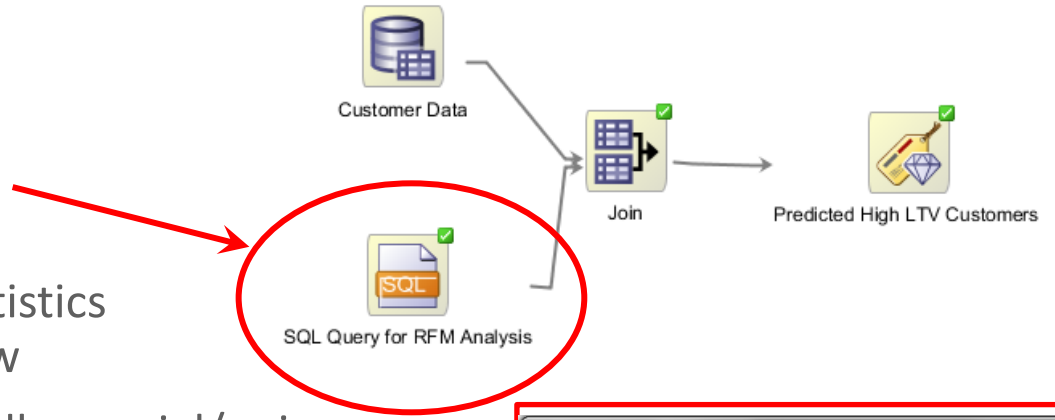


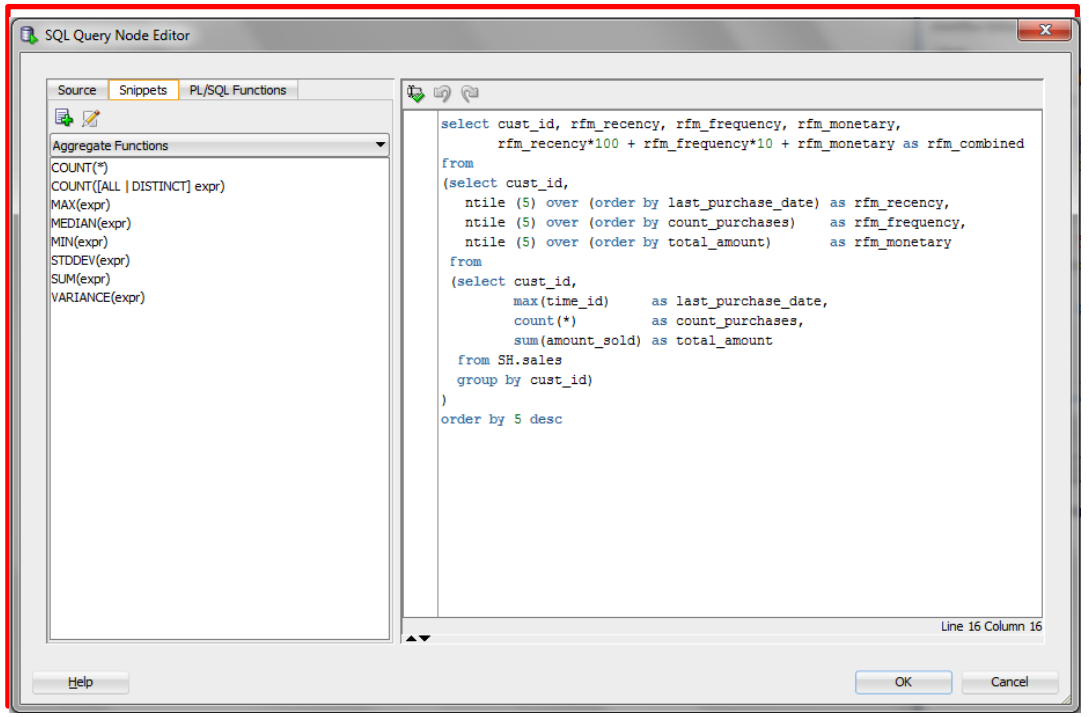


SQL Developer/Oracle Data Miner 4.0

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

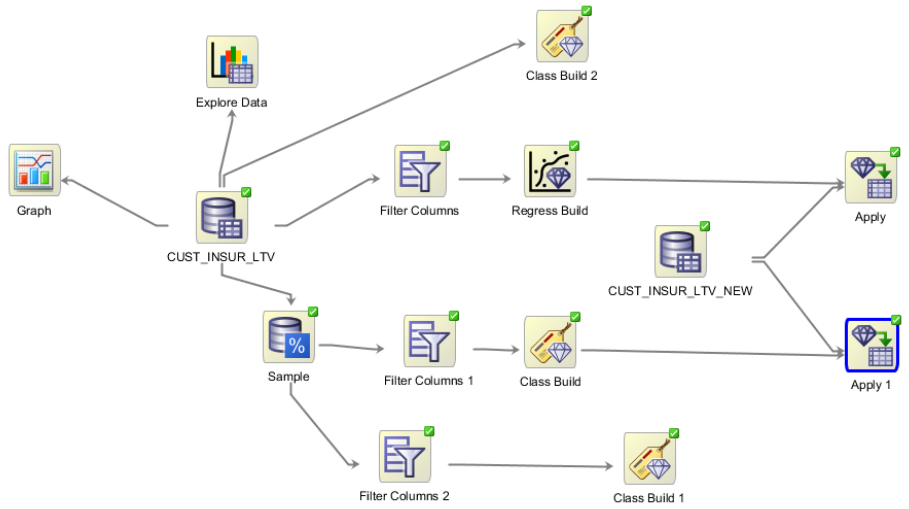


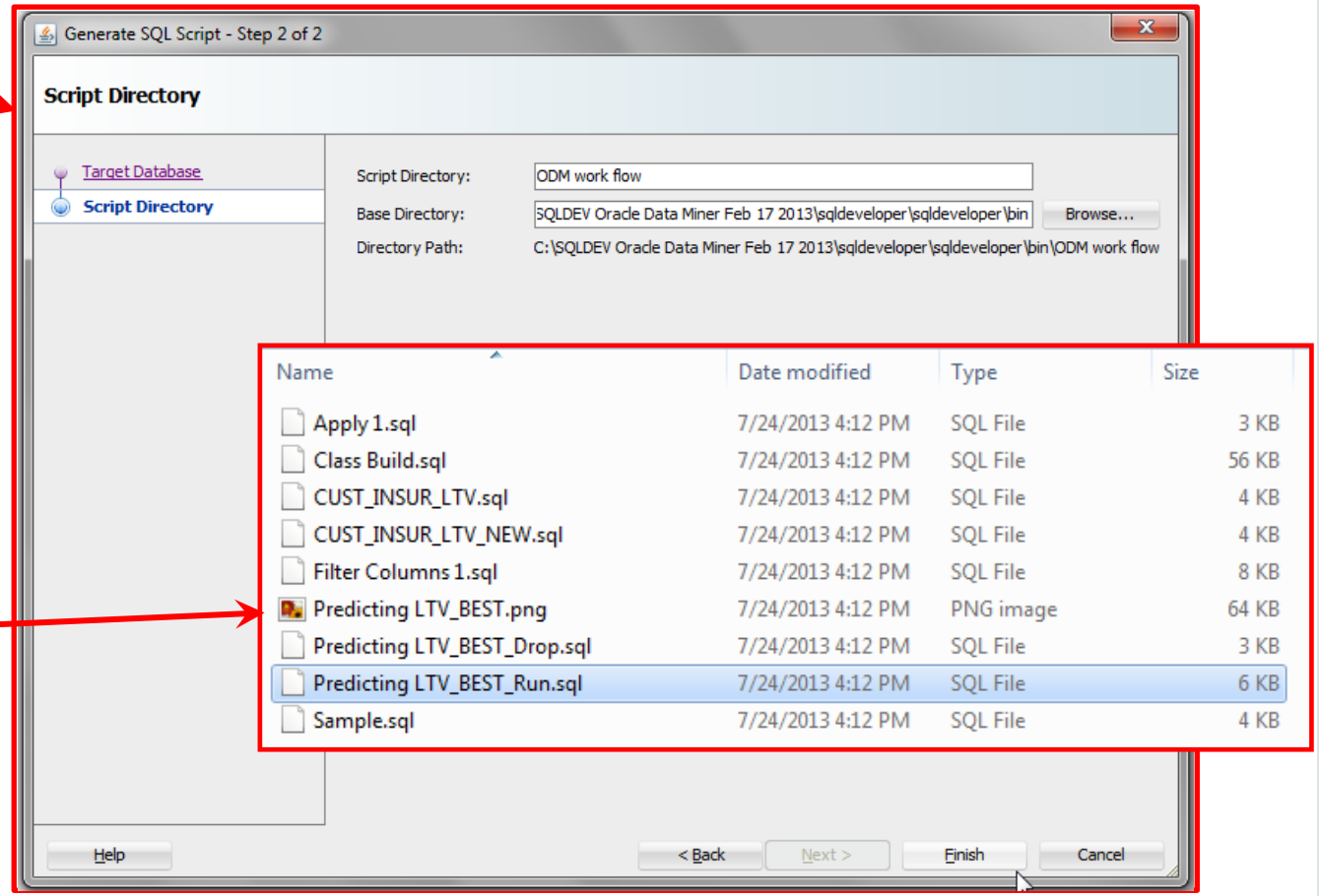


SQL Developer/Oracle Data Miner 4.0

New Features

- **SQL Script Generation**
  - Deploy entire methodology as a SQL script
  - Immediate deployment of data analyst's methodologies

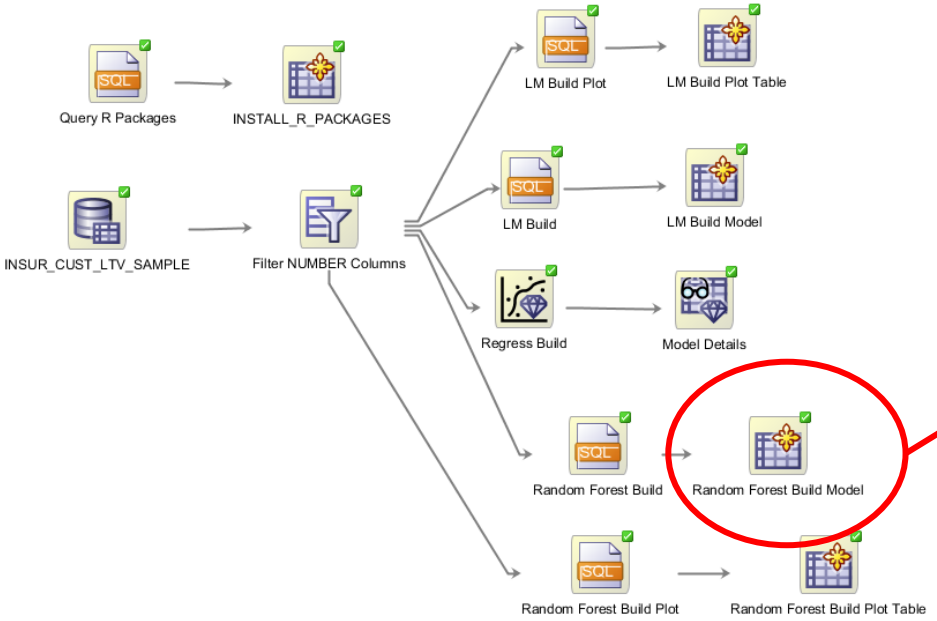


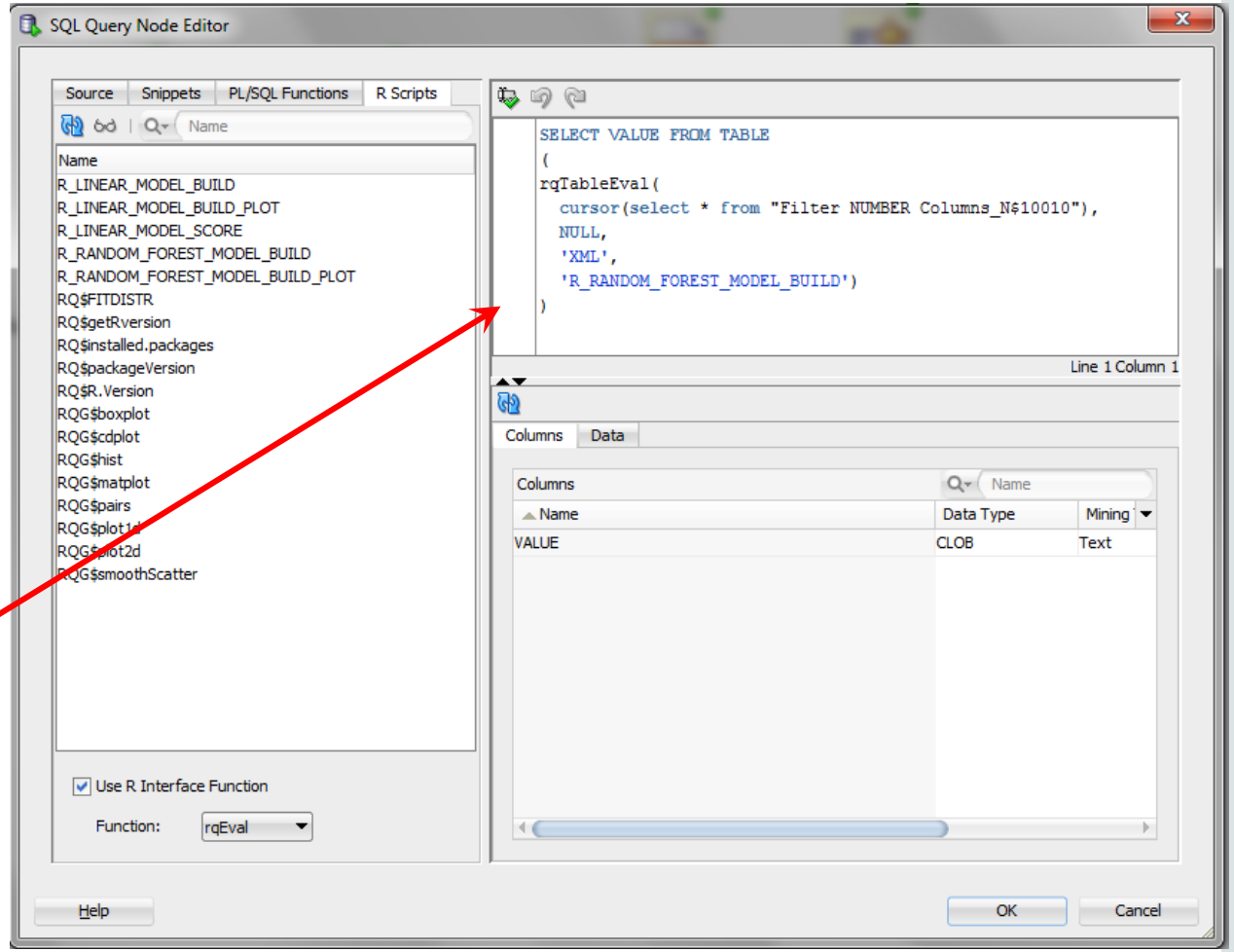


SQL Developer/Oracle Data Miner 4.0

New Features

- **SQL Query** node
  - Allows integration of R Scripts





12c *New Features*

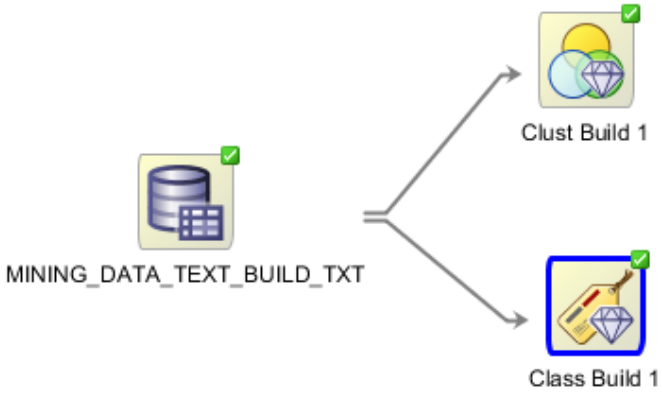
New Server Functionality

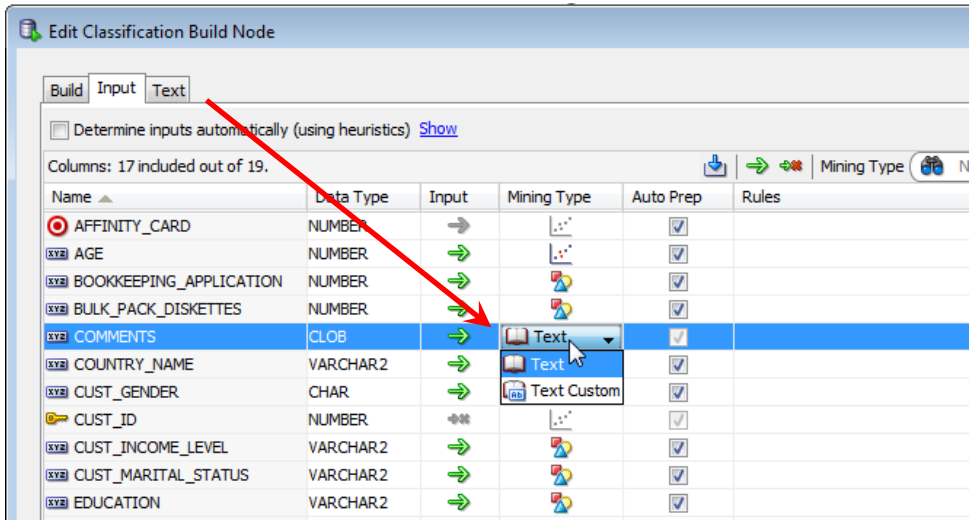
- **3 New Oracle Data Mining SQL functions algorithms**
  - **Expectation Maximization (EM) Clustering**
    - New Clustering Technique
      - Probabilistic clustering algorithm that creates a density model of the data
      - Improved approach for data originating in different domains (for example, 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.)

12c *New Features*

New Server Functionality

- **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



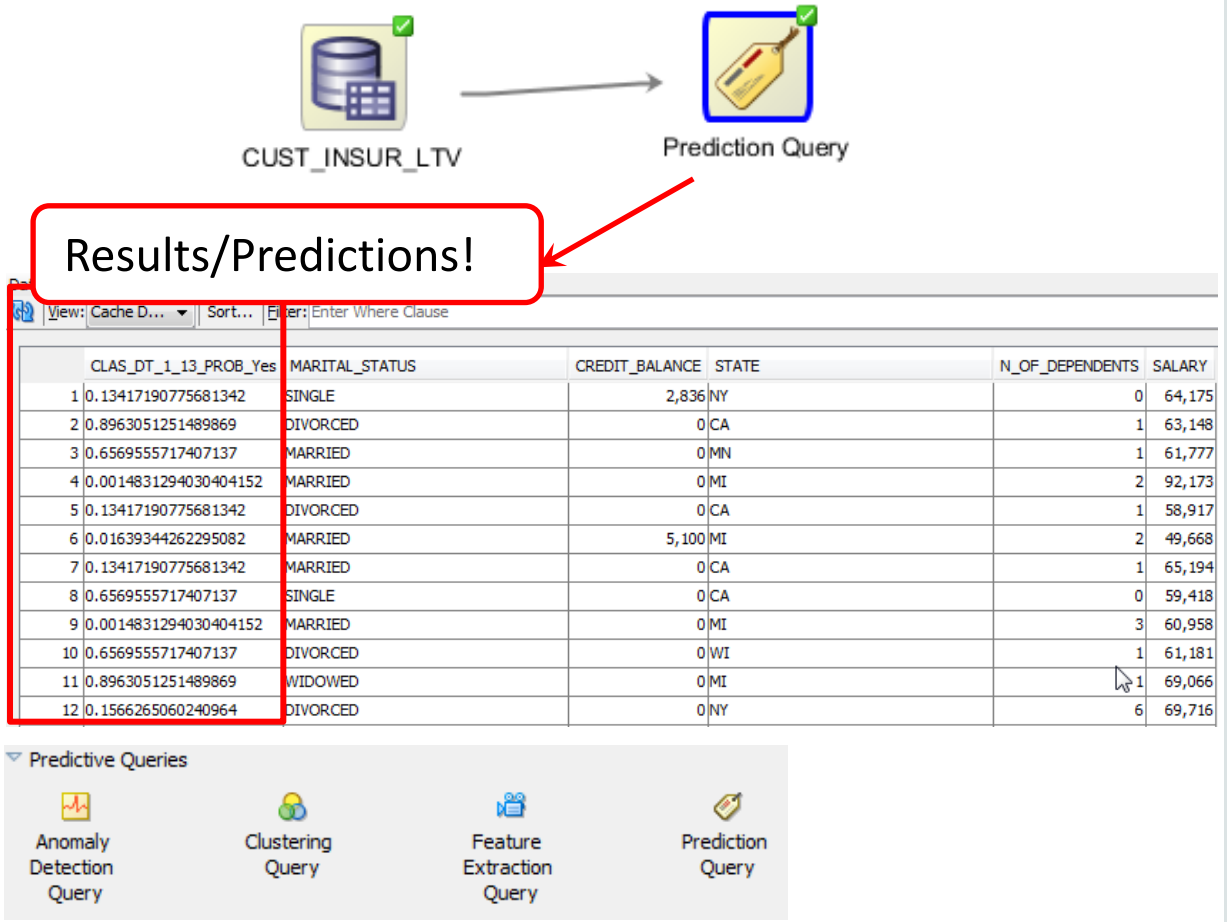


12c *New Features*

New Server Functionality

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

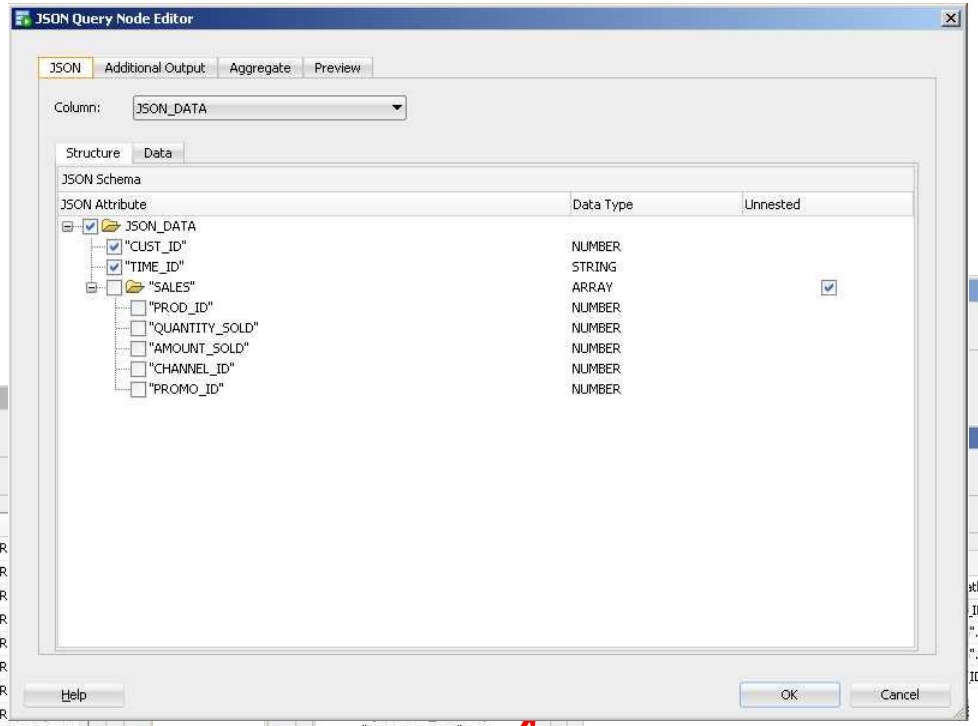
OAA automatically creates multiple anomaly detection models “Grouped_By” and “scores” by partition via powerful SQL query

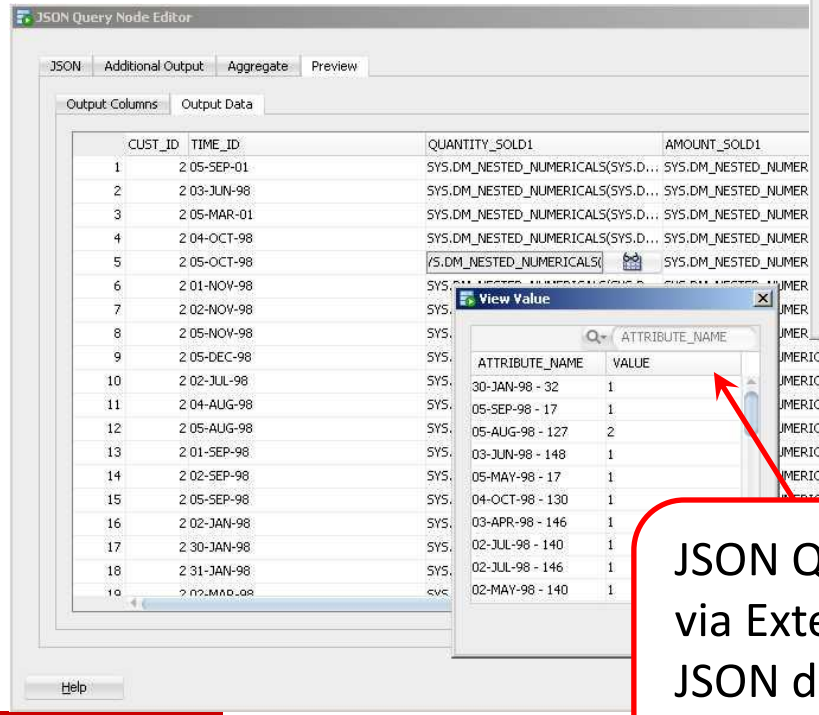


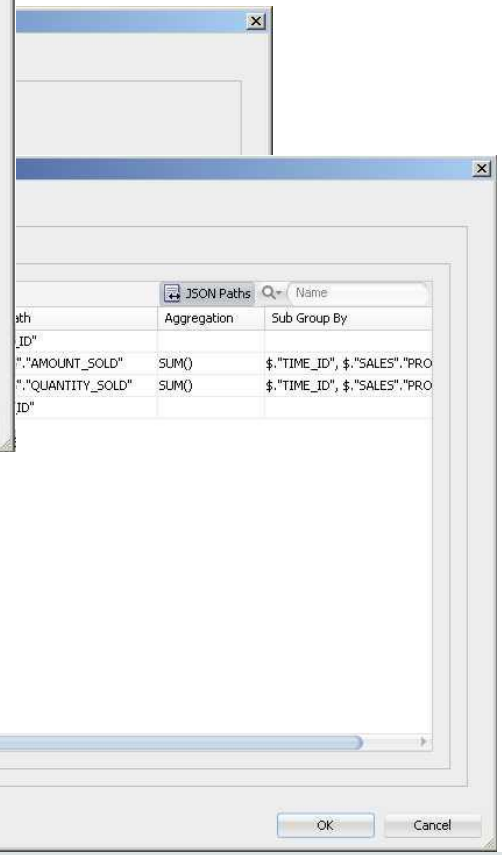
Oracle Data Miner 4.1

New Features

- **JSON Query node**







JSON Query node extracts BDA data via External Tables and parses out JSON data type and assembles data for data mining

Oracle Data Miner 4.1

New Features

- **Oracle Data Miner Workflow API to Manage, Schedule and Run Workflows**
  - PL/SQL APIs to enable applications to execute workflows immediately or schedule them
  - Oracle Scheduler for scheduling functionality
  - ODMr repository views can be queried for project and workflow information
  - Applications can monitor workflow execution and query generated results

```
CONNECT DMUSER/DMUSER
SET SERVEROUTPUT ON DECLARE
      v_jobId                    VARCHAR2(30) := NULL;
      v_status                   VARCHAR2(30) := NULL;
      v_projectName              VARCHAR2(30) := 'Project';
      v_workflow_name            VARCHAR2(30) := 'build_workflow';
      v_node                     VARCHAR2(30) := 'MODEL_COEFFCIENTS';
      v_run_mode                  VARCHAR2(30) := ODMRSYS.ODMR_WORKFLOW.RERUN_NODE_PARENTS;
      v_failure                  NUMBER := 0;
      v_nodes                    ODMRSYS.ODMR_OBJECT_NAMES := ODMRSYS.ODMR_OBJECT_NAMES();
BEGIN
      v_nodes.extend();
      v_nodes(v_nodes.count) := v_node;
      v_jobId := ODMRSYS.ODMR_WORKFLOW.WF_RUN(p_project_name => v_projectName,
                                    p_workflow_name => v_workflow_name,
                                    p_node_names => v_nodes,
                                    p_run_mode => v_run_mode,
                                    p_start_date => '31-DEC-14 12.00.00 AM AMERICA/NEW_YORK',
                                    p_repeat_interval => 'FREQ=MONTHLY;BYMONTHDAY=-1',
                                    p_end_date => '31-DEC-15 12.00.00 AM AMERICA/NEW_YORK');
DBMS_OUTPUT.PUT_LINE('Job: '||v_jobId);
```

Getting started

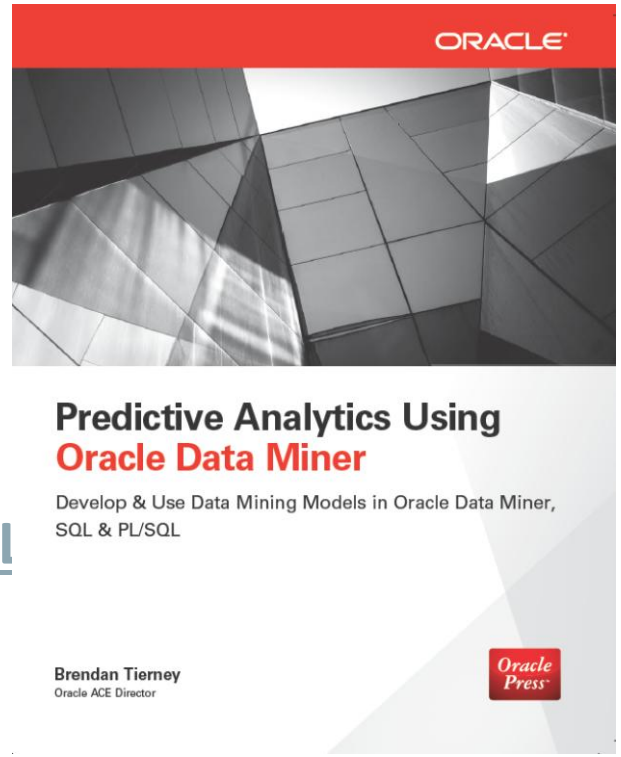
OAA Links and Resources

- **Oracle Advanced Analytics Overview:**
  - Link to **OAA presentation**— Big Data Analytics in Oracle Database 12c With **Oracle Advanced Analytics** & Big Data SQL
  - **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
- ***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, Ret
- **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
  - 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'16, Jan 26-28, 2016 at Oracle HQ Conference Center

Books on Oracle Advanced Analytics

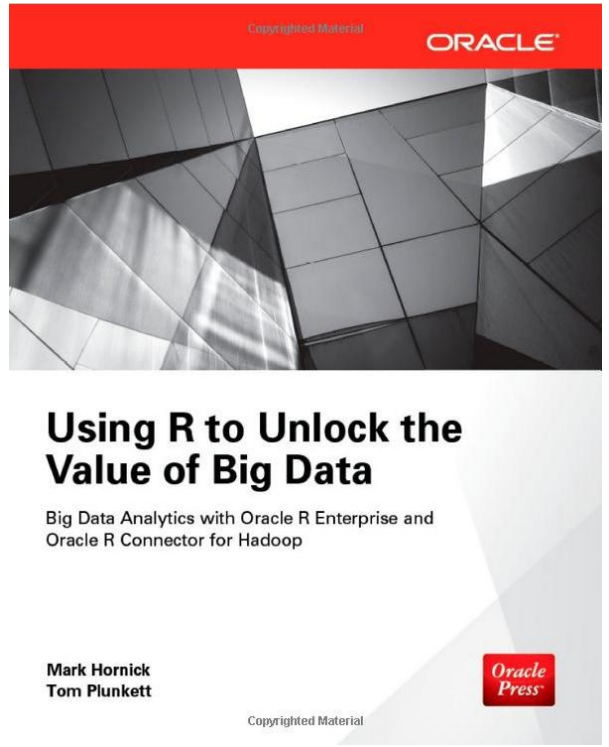
Book available on Amazon

[Predictive Analytics Using Oracle Data Miner: Develop for ODM in SQL & PL/SQL]



Book available on Amazon

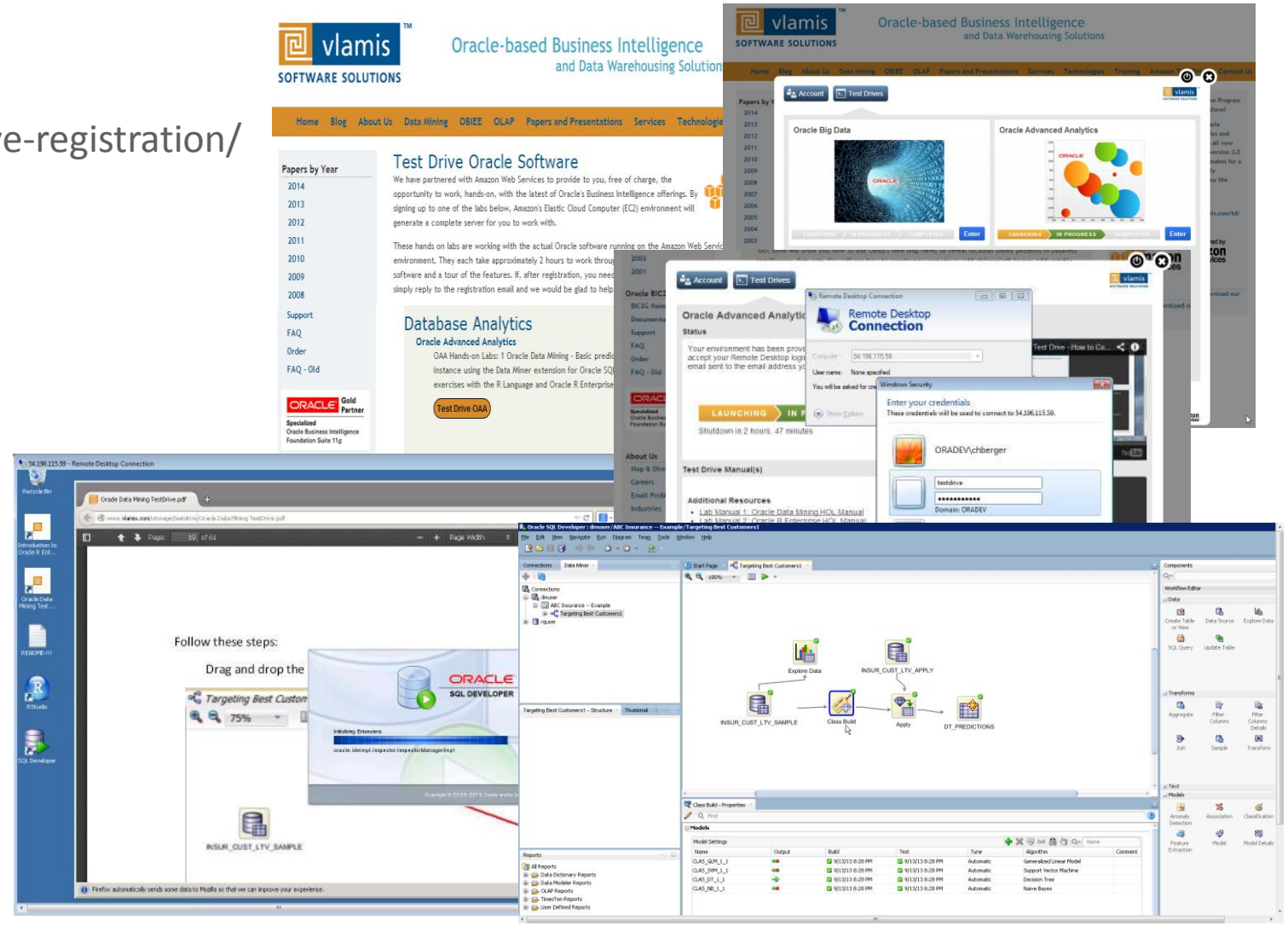
[Using R to Unlock the Value of Big Data]



Take a Test Drive!

Vlamis Software, Oracle Partner Offers FREE Test Drives on the Amazon Cloud

- Step 1—Fill out request
  - Go to http://www.vlamis.com/testdrive-registration/
- Step 2—Connect
  - Connect with Remote Desktop
- Step 3—Start Test Drive!
  - Oracle Database +
  - Oracle Advanced Analytics Option
  - SQL Developer/Oracle Data Miner GUI
  - Demo data for learning
  - Follow Tutorials



Login

IOUG independent oracle users group

SUPER

For the Complete Technology & Database Professional

Join #IOUGenius Resource Center Publications Communities Events Partners Marketplace About IOUG

Home Page Hotel and Travel Agenda Sessions Sponsorship Registration Abstract Submission

Click for BIWA Summit 2015 info

Thank you to all that made BIWA Summit 2015 which took place January 27-29, 2015 such a fantastic success! BIWA Summit 2015 presentations will be available soon to attendees.

Publicity

- Oracle Business Analytics Newsletter
- DB Insider Dec 2014
- Oracle Magazine
- Latest BIWA SIG Blog Entry
- Jeff Shauer Blog Entry

Oracle BIWA Summit 2016

Including Oracle Spatial Summit 2016

In partnership with NoCOUG

January 26-28, 2016
Oracle Conference Center at Oracle HQ Campus, Redwood Shores, CA

January 26, 27, 28, 2016
at Oracle HQ Campus

BIWA Summit 2016 is being planned now. Send an email to Dan Vlamis at dvlamis@vlamis.com or Shyam Varan Nath at shyamvaran@gmail.com to get involved.

See the BIWA Summit 2015 home page for more information on BIWA Summit 2015, including presentations, agenda, etc.

ORACLE


ORACLE®