
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.

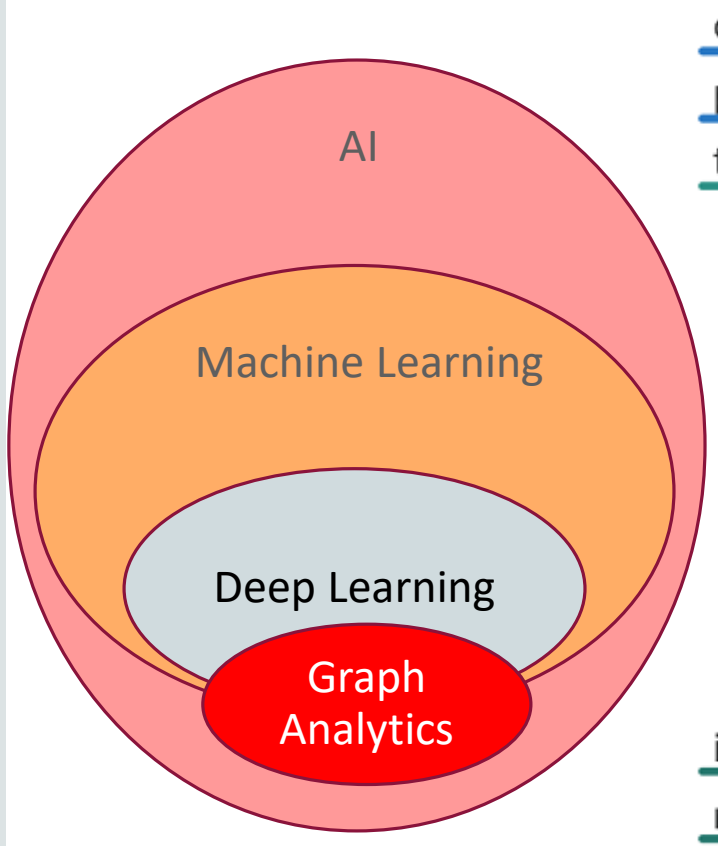
Oracle's Machine Learning

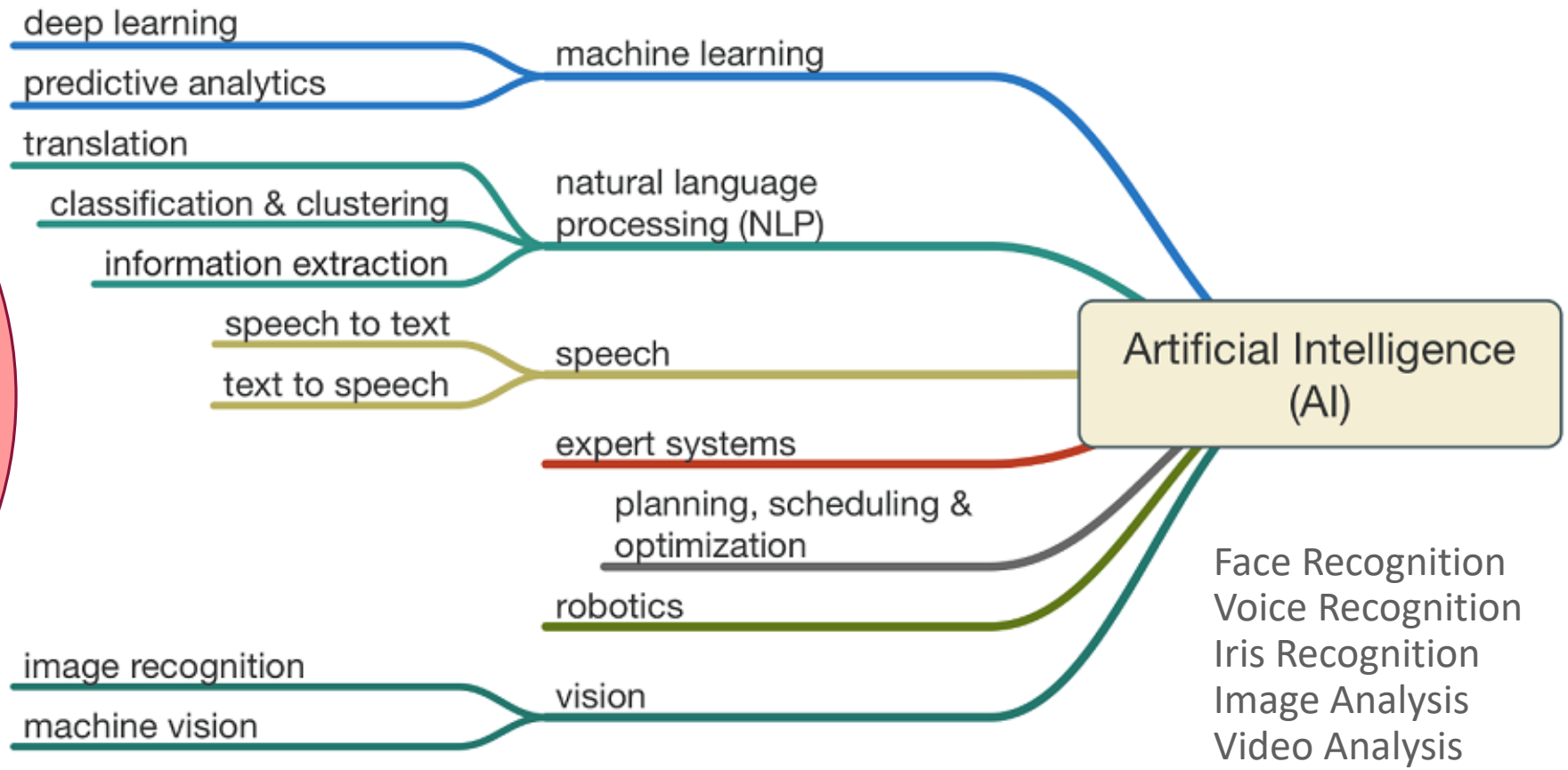
Oracle Database 19c Overview

Olivier Perard
Olivier.perard@oracle.com

Iberia Technology Sales Consulting
January, 2020

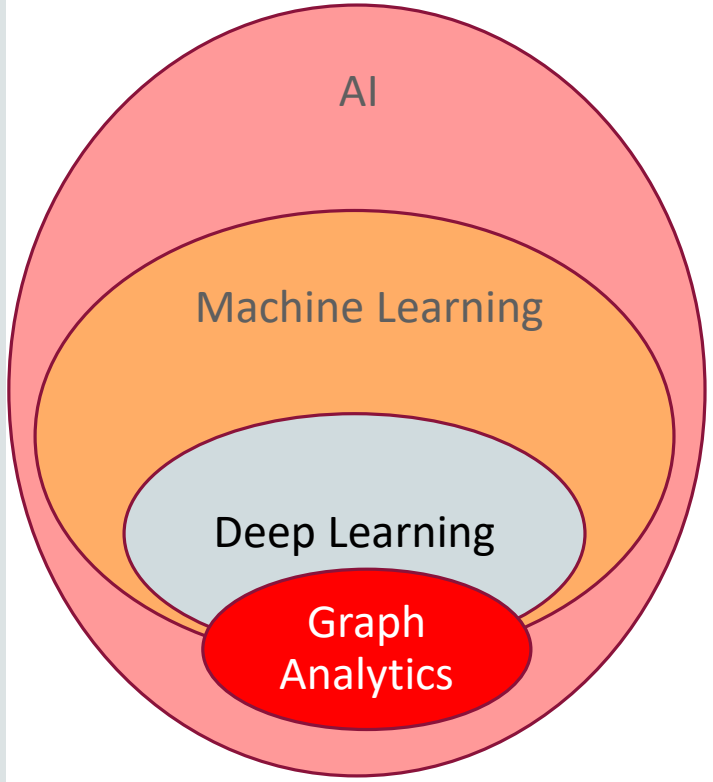
Artificial Intelligence Concepts





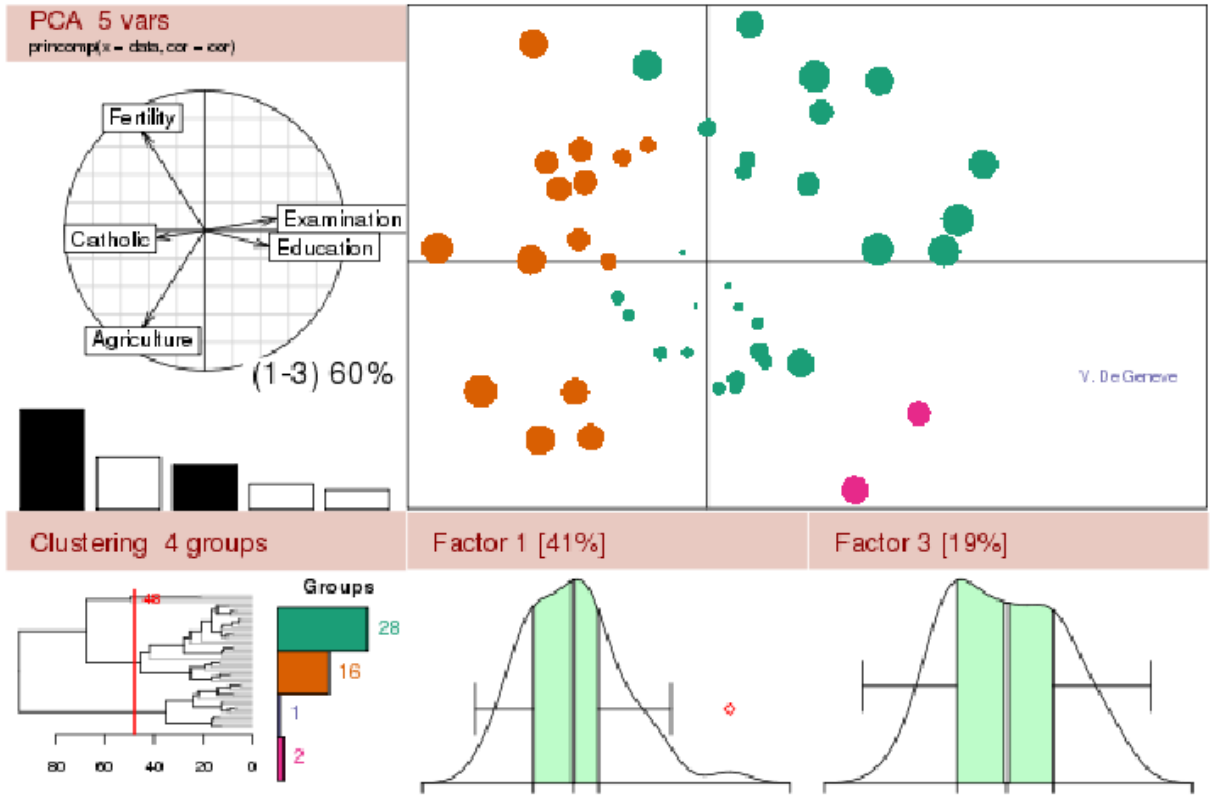
Face Recognition
Voice Recognition
Iris Recognition
Image Analysis
Video Analysis
...etc....

Machine Learning Concepts



- Neural Network - NN (feed)
- Recurrent Neural Networks -(RNN)
- Recursive Neural Networks
- Convolutional Neural Network -(CNN)
- Long Short Term Memory (LSTM)

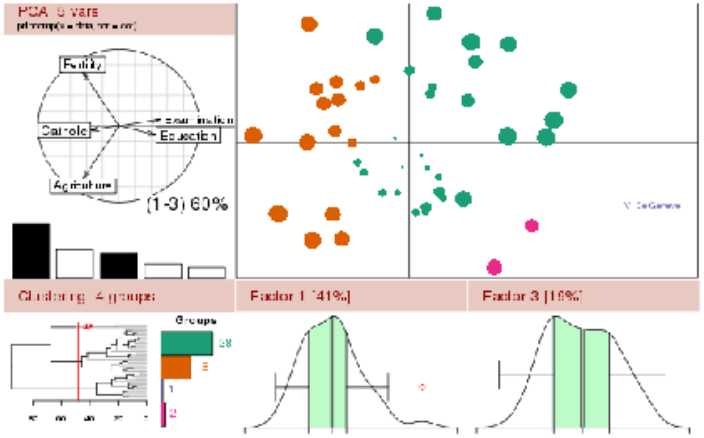
R Statistical Programming Language

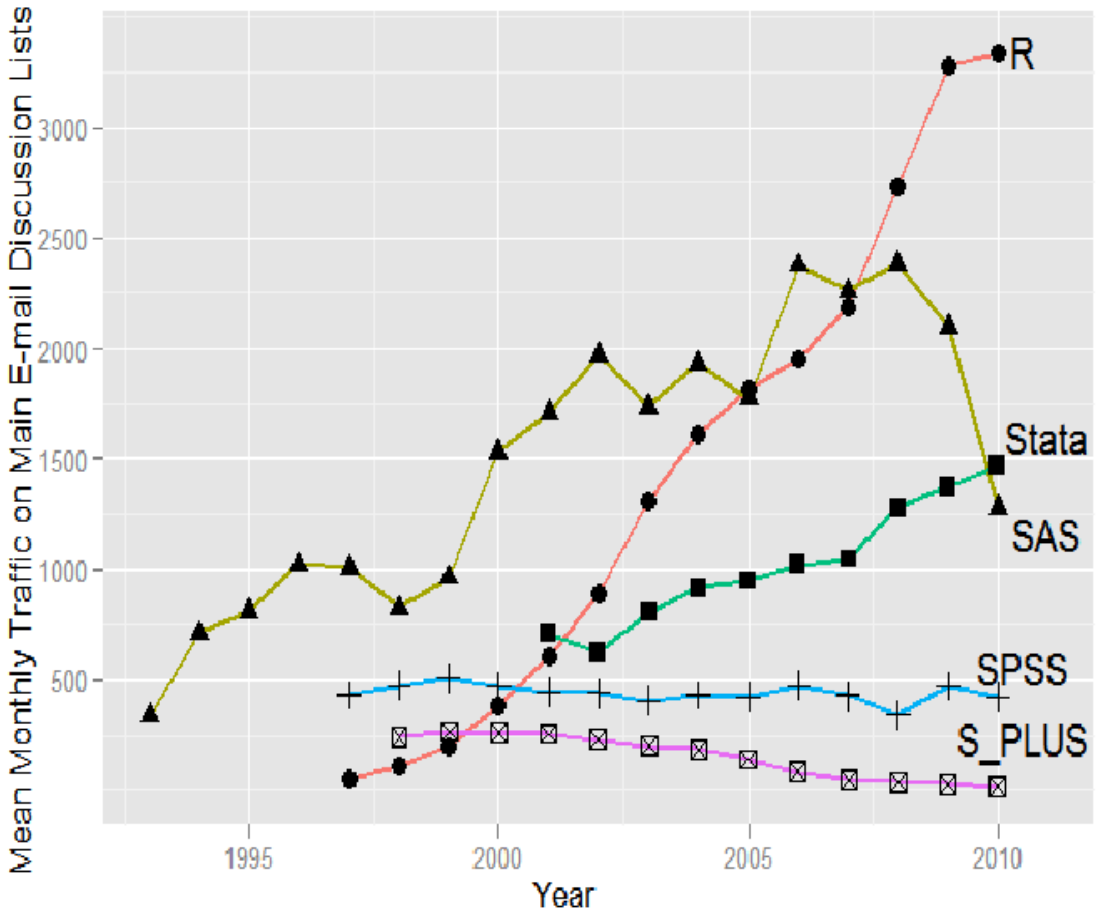


- Open source language and environment
- Used for statistical computing and graphics
- Strength in easily producing publication-quality plots
- Highly extensible with open source community R packages

Growing Popularity

The R Project for Statistical Computing





- R's rapid adoption over several years has earned its reputation as a new statistical software standard
  - Rival to SAS and SPSS

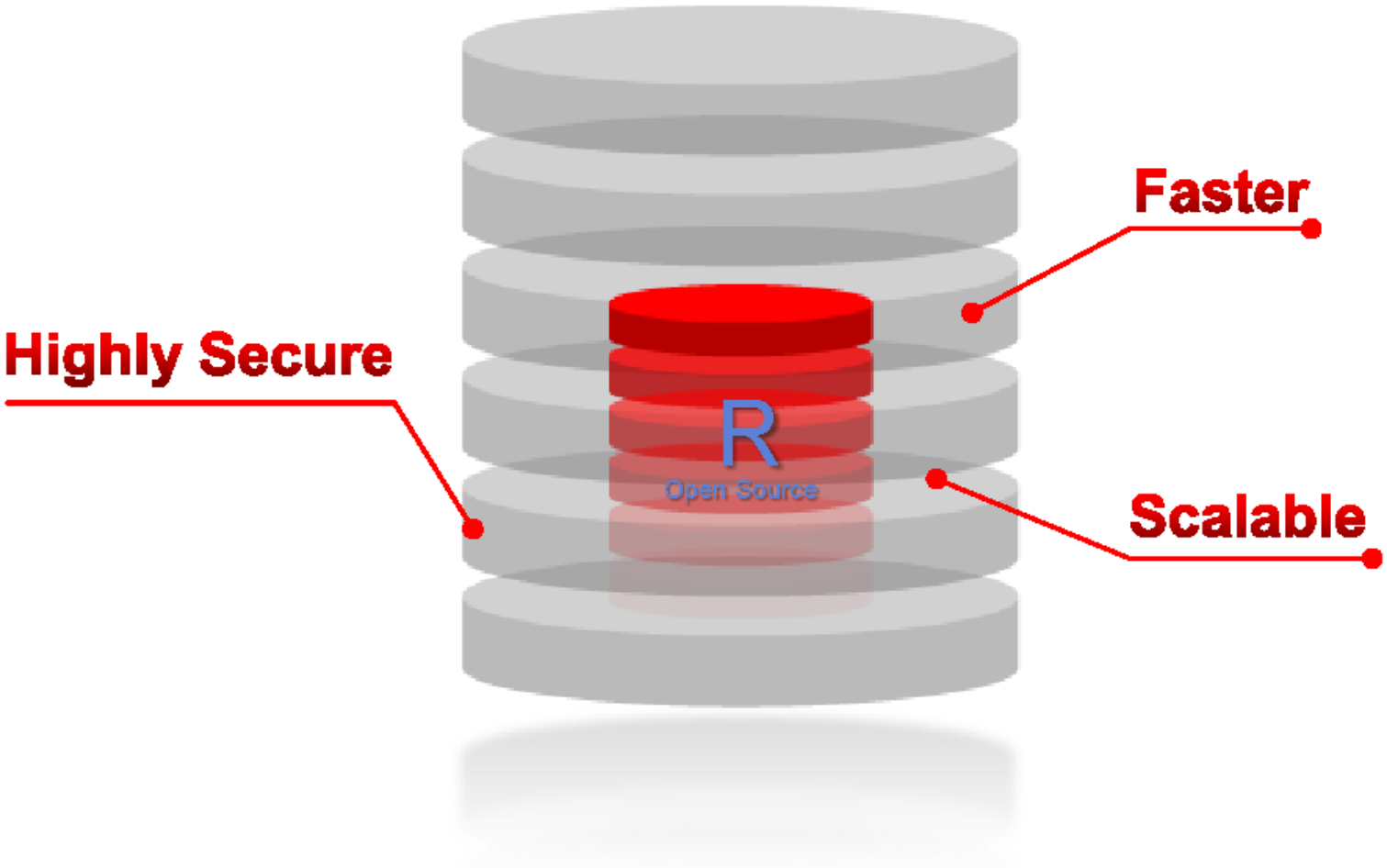
While it is difficult to calculate exactly how many people use R, those most familiar with the software estimate that close to 250,000 people work with it regularly.
"Data Analysts Captivated by R's Power", New York Times, Jan 6, 2009

http://www.r-project.org/

What Are R 's Challenges?

1. R is memory constrained
   - R processing is single threaded - does not exploit available compute infrastructure
   - R lacks industrial strength for enterprise use cases
2. R has lacked mindshare in Enterprise market
   - R is still met with caution by the long established SAS and IBM/SPSS statistical community
     - However, major university (e.g. Yale) Statistics courses now taught in R
     - The FDA has recently shown indications for approval of new drugs for which the submission's data analysis was performed using R

Oracle R Enterprise Approach



Data and statistical analysis are stored and run in-database

Same R user experience & same R clients

Embed in operational systems

Complements Oracle Data Mining

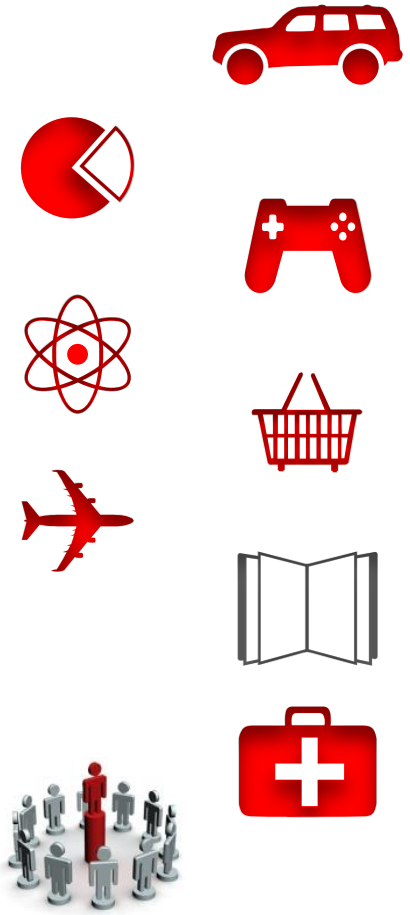
Vision

- Big Data + Data Science Platform for the Era of Big Data and Cloud
  - Make Big Data + Machine Learning Model Building *Simple*
  - Make Big Data + Machine Learning Model Deployment *Simple*
  - *Key Differentiators:*
    - Fully integrated into Oracle and Hadoop platforms
    - Scalable and Distributed algorithms run where the data is, in-Database or in-Cluster.
    - Easy to use using familiar interfaces like SQL and R.
    - Support for open-source R packages running in the Database Server or in the Cluster
    - Integrated with Oracle solutions like Graph, OBIEE, BDD, IoT, OSA, RTD
    - Easy GUI provided for SQL Developer; Compatible with 3rd party GUIs like RStudio
    - Low TCO, included in Oracle Cloud Services like EXADATA, DBaaS (Extreme and High Performance editions) and BDCS

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

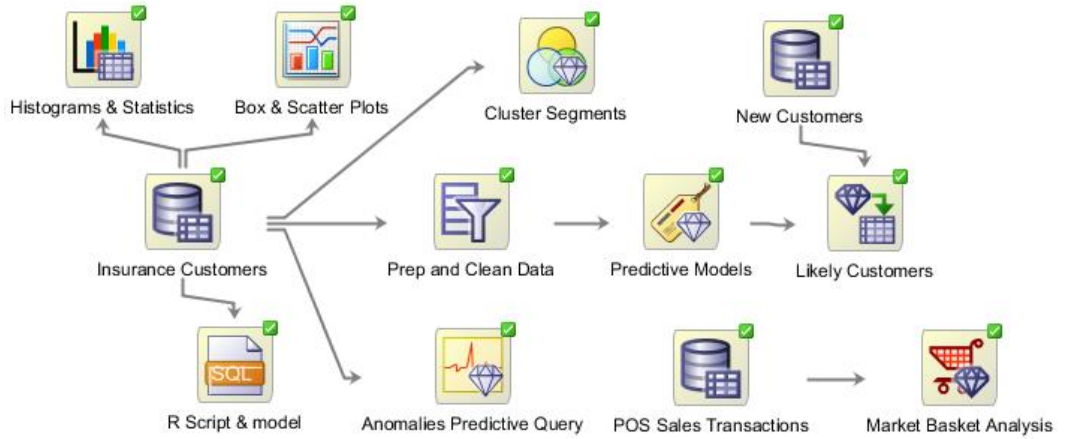


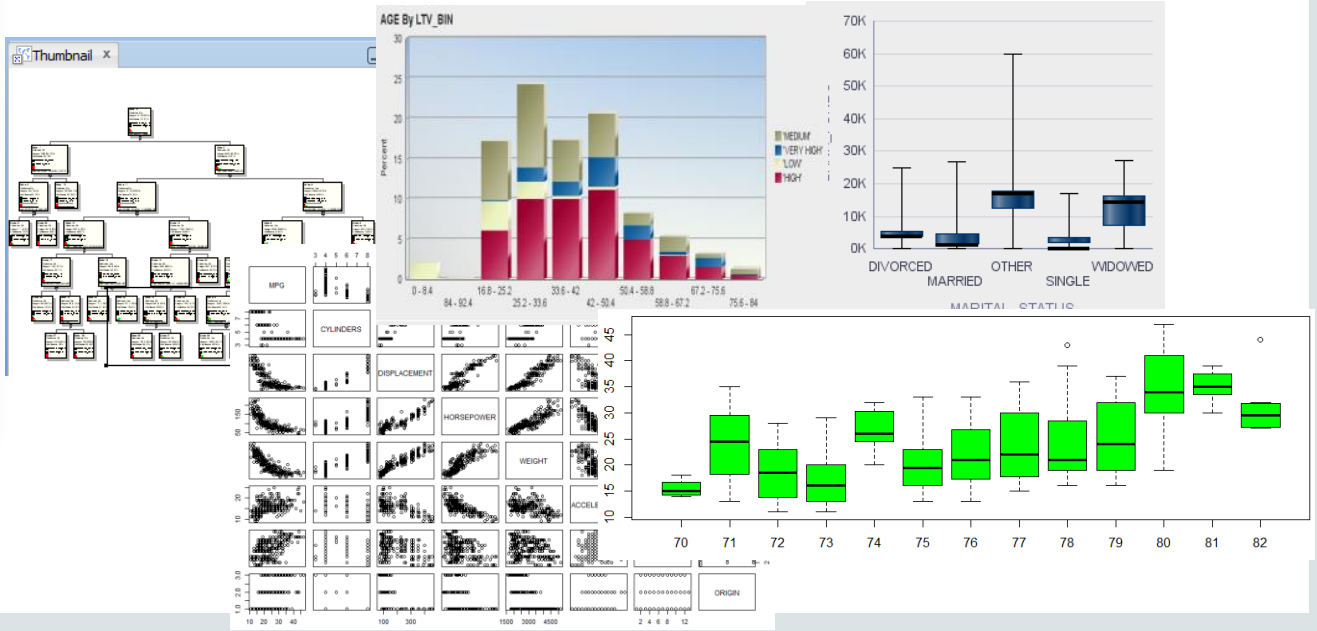
Oracle Advanced Analytics Database Option

Fastest Way to Deliver Scalable Enterprise-wide Predictive Analytics

Key Features

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





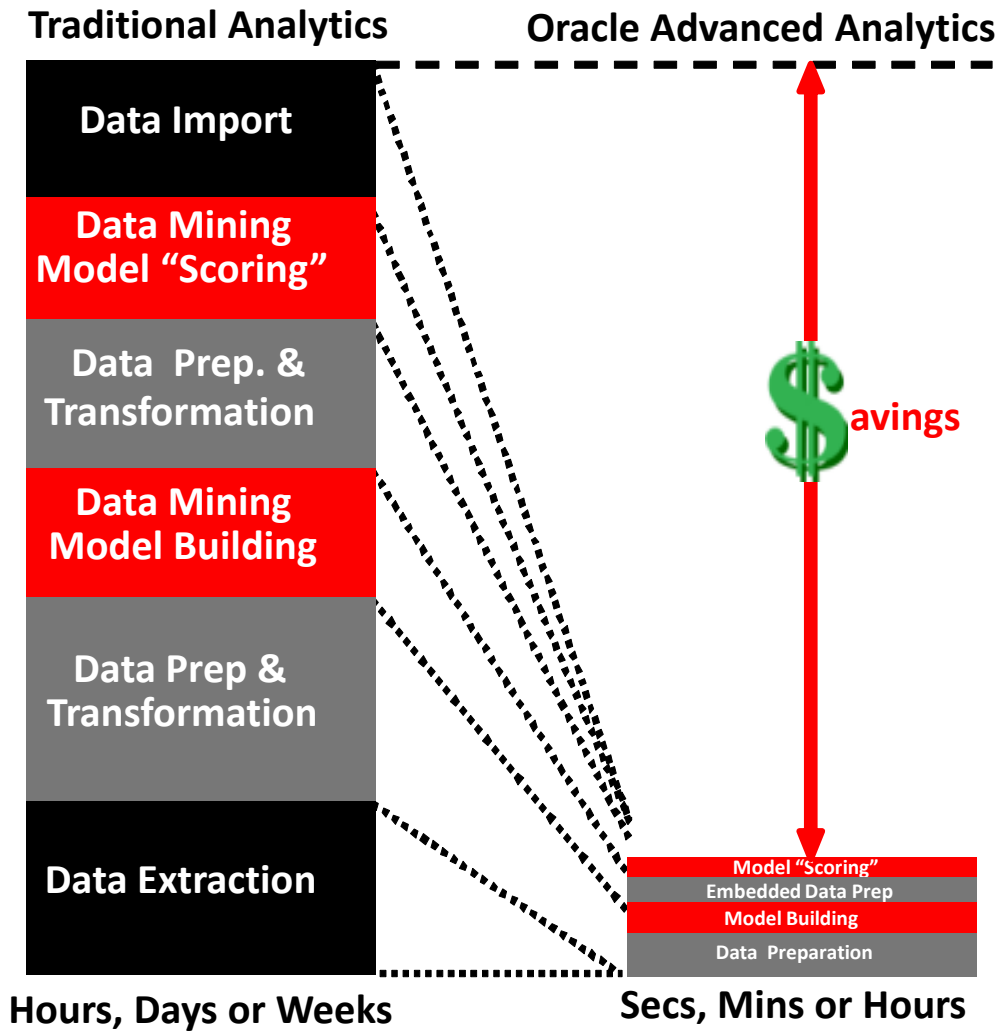
Oracle Advanced Analytics Database Option

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

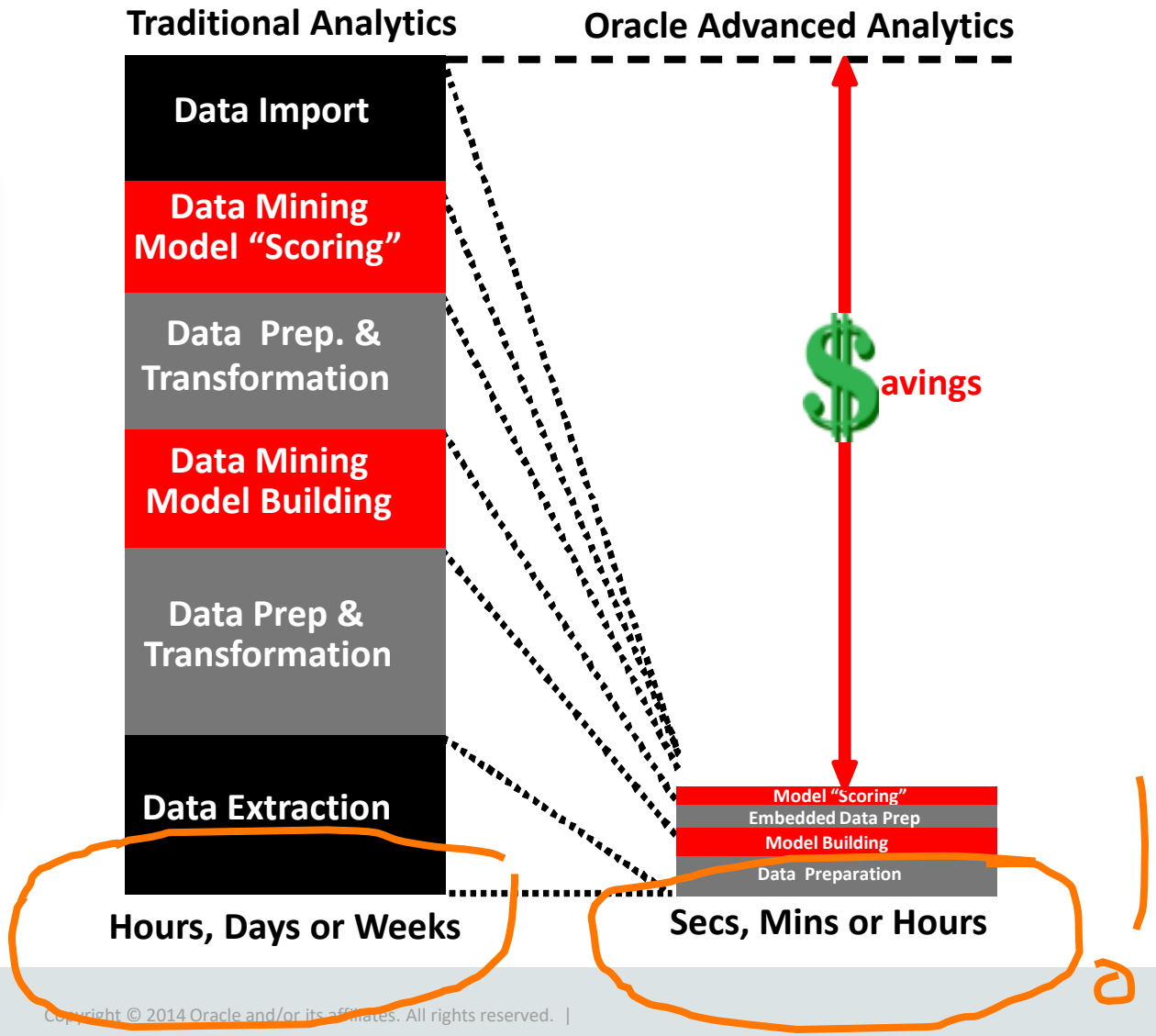
Key Features

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

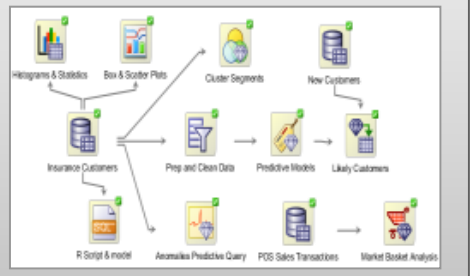
Oracle Advanced Analytics Database Architecture

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

Users

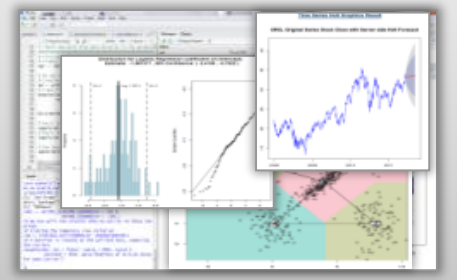
Data & Business Analysts

SQL Developer



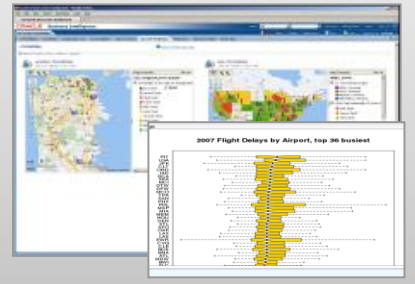
R,Python programmers

R, Python Client



Business Analysts/Mgrs

OBIEE



Domain End Users

Applications

Platform

Oracle Database Enterprise Edition

Oracle Advanced Analytics

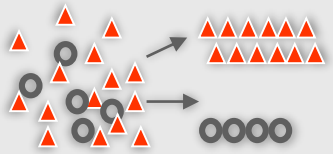
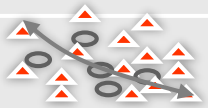
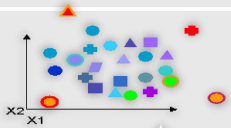
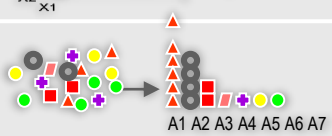
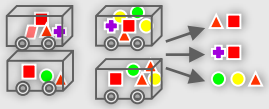
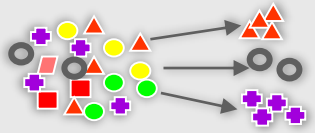
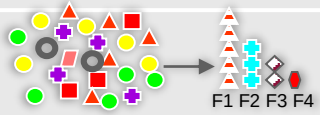
Native SQL Data Mining/Analytic Functions + High-performance

R & Python Integration for Scalable, Distributed, Parallel Execution

ORACLE

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 In-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

Oracle Advanced Analytics

Brief Demos

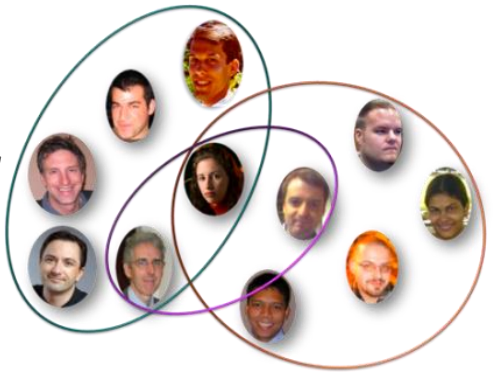
ORACLE

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

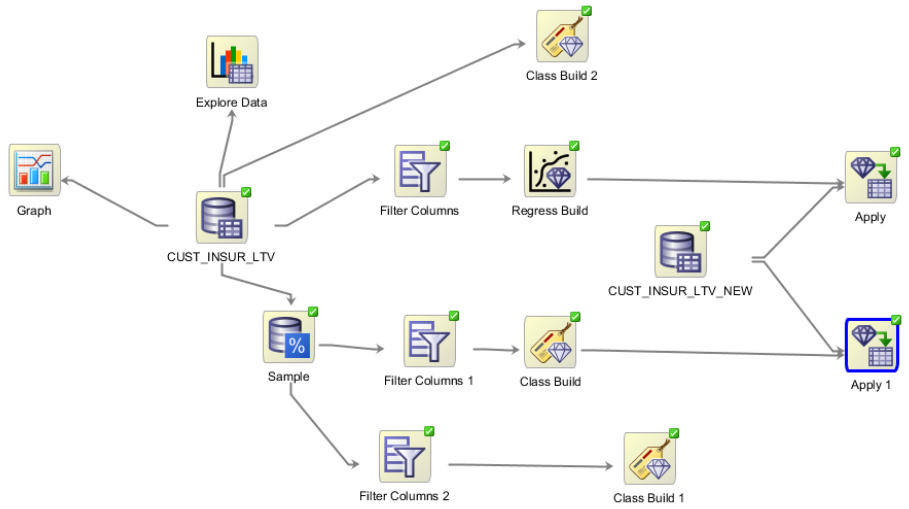
Consider:

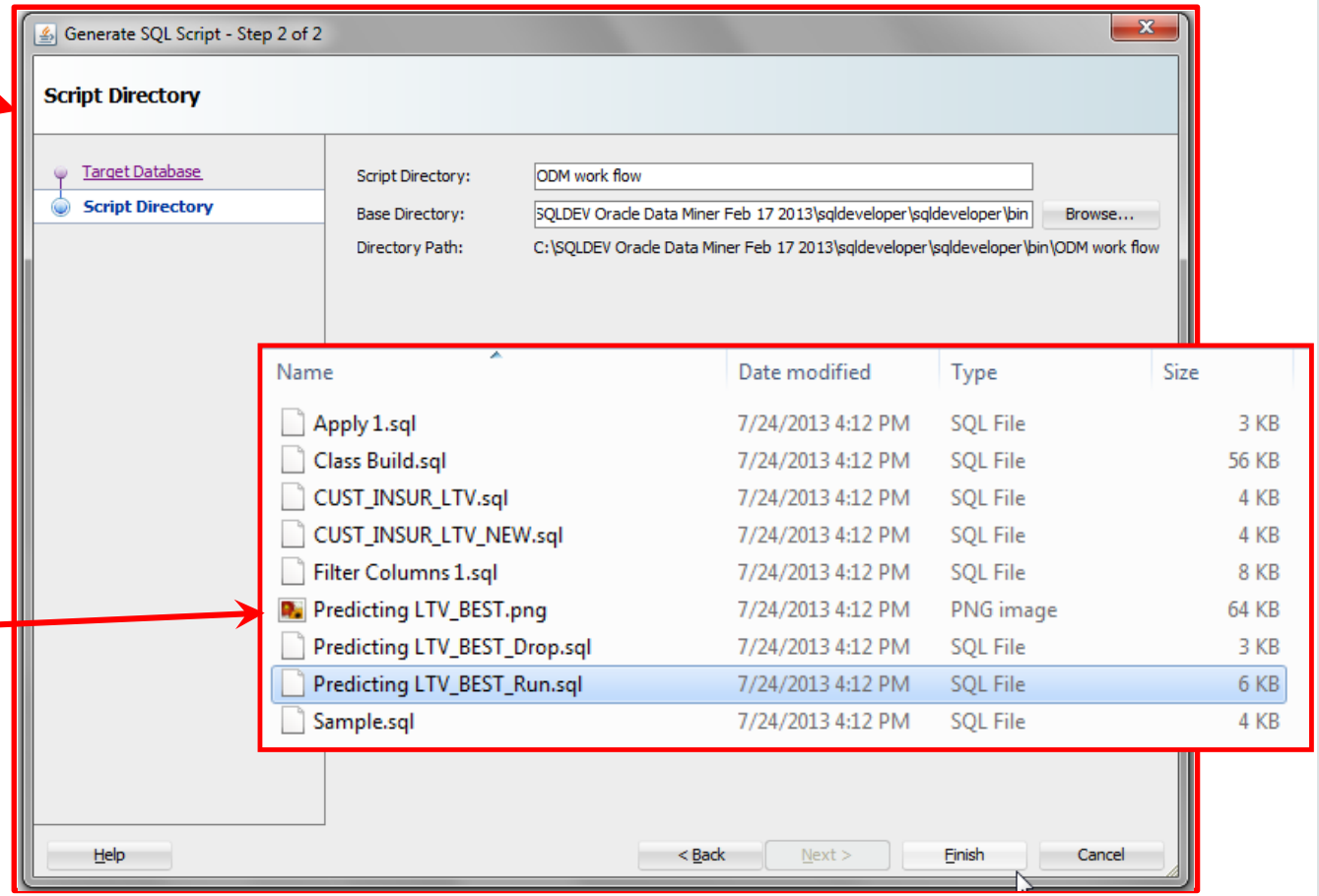
- Demographics
- Past purchases
- Recent purchases
- Customer comments & tweets

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





Oracle Advanced Analytics

OAA/Oracle R Enterprise (R integration)

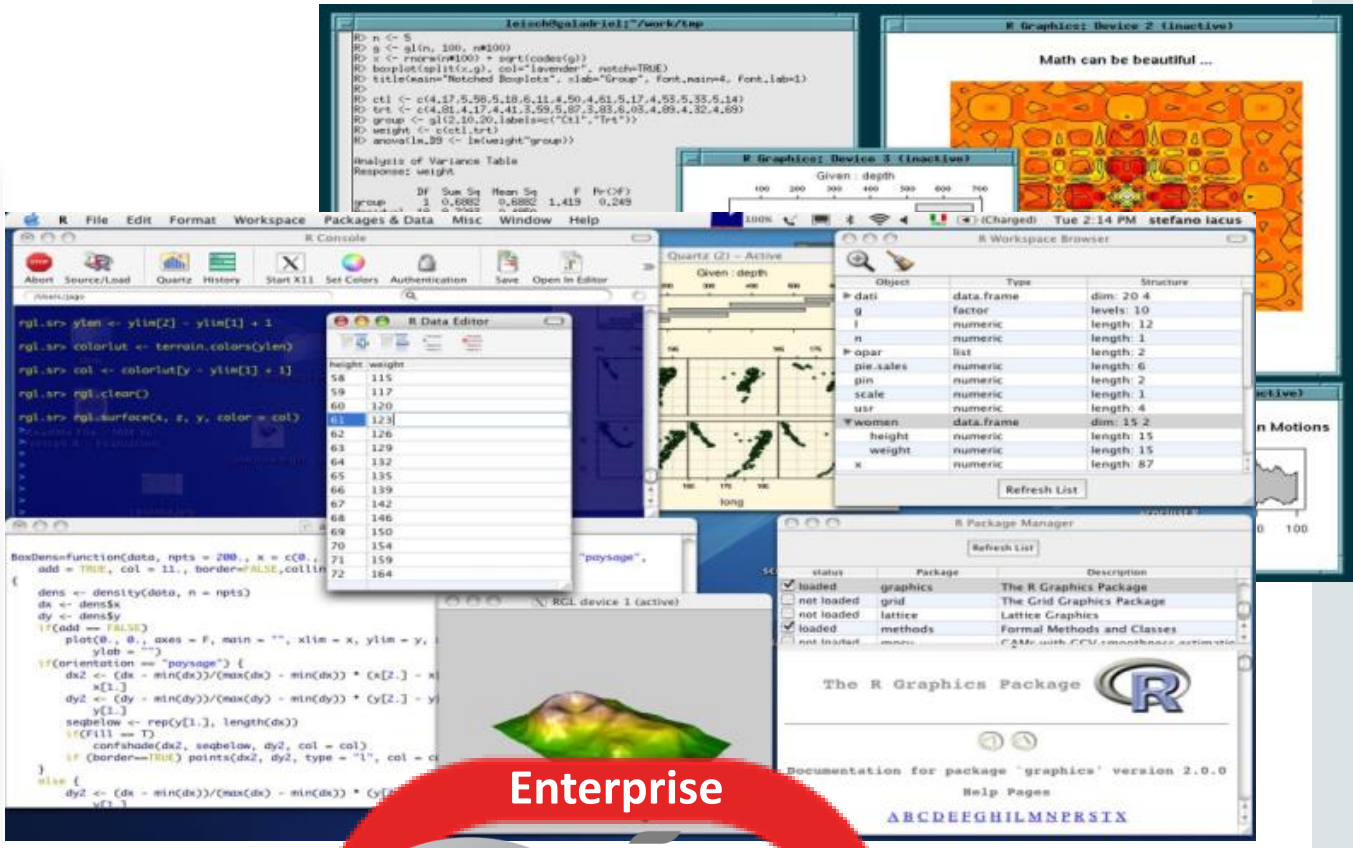
ORACLE

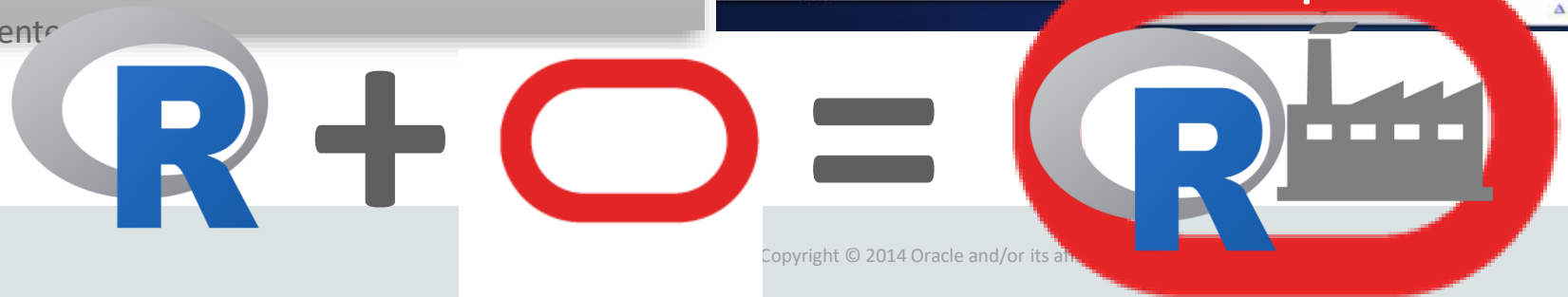
R—Widely Popular

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

R environment

- Strengths
  - Powerful & Extensible
  - Graphical & Extensive statistics
  - Free—open source (CRAN + 9000 components)
  - Standard for Data Scientist
- Challenges
  - Memory constrained
  - Single threaded
  - Outer loop—slows down process
  - Not Enterprise Orient

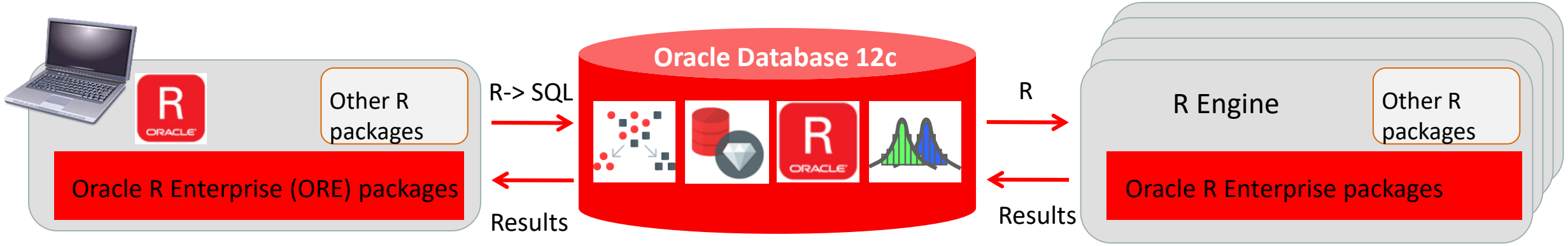




Oracle Advanced Analytics

How Oracle R Enterprise Compute Engines Work

Advanced Analytics



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

ORACLE

08. Advanced Analytics

8.0 Oracle BI EE 12c Advanced Analytics

8.00 Advanced Analytics:
Overview, Binning, Trendline, Forecast, Outlier, Cluster, Regression

8.01 Visualizing using R:
Interactive Boxplot, Interactive 3D Scatter, Bubble Chart Grid, Variable Width Bar, Random Dots

8.02 Functional Examples (Evaluate Script):
Text Sentiment, Text Term Frequency, Timeseries Decomposition, Market Basket Analysis, Collaborative Filtering, Delay Prediction (Precomputed Model)

8.1 Descriptive Analytics

8.10 Binning and Tiling:
Comparative Dist, Distr. II, Distribution, Large Data, Ntiling, Percentiles, Width Bucket

8.11 Comparative Analysis:
Age Pyramid, Benchmark, Index To Avg, Indexing, Lift, Tiering, TopN, TopN History

8.12 Descriptive Stats:
8020, Control Chart, Correlation, Data Density, Deviants, Scatter, See Also, StdDev, Variability

8.13 History and Trend:
History, Seasonality, Trend Lines, Trending

8.2 Oracle Database Analytics

8.20 Text Analytics:
Cost Analysis per Token, Cost Per Token Frequency, Text Aggregation, Text Classes, Text Classification, Text Filtering, Words Distribution

8.21 Temporal and Time:
Months Between, Temporal Query (12c Session), Temporal Query (12c), TimeZone with DST, Timezone Conversion

8.22 Analytic Clauses:
Frequent Itemset, Model Projection, Pattern Detection (12c), Projection Interactive

8.23 Other DB Analytics:
CLOB Datatype, Column Statistics, DB Web Services, Text Aggregation, JSON Parsing, Approximate Count Distinct

8.3 Oracle Data Mining

8.30 ODM Classification:
Classification Tree, Geo LTV Prediction, LTV Details, LTV Prediction, LTV Probabilities, LTV What If Scoring, Dynamic Classification (12c)

8.31 ODM Regression:
Regression, Regression Variance, Variance Heatmap, Dyn Predictive Regression (12c)

8.32 ODM Clustering, Association and Attribute Importance:
Clustering, Market Basket Analysis, Attribute Importance

8.33 ODM Mining On-the-fly:
Anomaly Influencers (12c), Dyn Anomaly Detection (12c), Dyn Predictive Regression (12c), Dynamic Classification (12c), Dyn Prediction Delay Grp (12c In-mem), Dyn Anomaly Analysis (12c In-mem)

8.34 ODM Data Miner Workflows:
List of Examples, Overview

8.4 Oracle R Enterprise

8.40 ORE Integration:
R Integration, R End-User Interaction, R Workbench, R Results Object in RPD, BIP Sourcing from R, Quality Control Chart (BIP)

8.41 ORE Time Series:
T. Series Decomposition, T. Series Forecasting, T. Series Moving Average, T. Series Auto ARIMA, T. Series Holt, T. Series SES, T. Series ACF PACF

8.42 ORE Datamining:
Multivariate Adaptive Regression Splines, Support Vector Machines, Association Rules, Variable Importance, Clustering with k-Means++, In-Database Associations, ORE GroupApply, ORE IndexApply

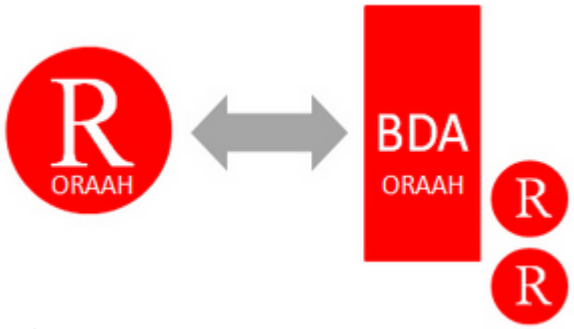
8.43 ORE Visualizations:
Quality Control Chart, Boxplot, Cond. Histogram, Corr. Matrix Circles, Corr. Matrix Ellipses, Heatmap, Multipanel Geo Lattice, Volcano, sinc Perspective

Oracle Advanced Analytics for Hadoop

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

ORACLE

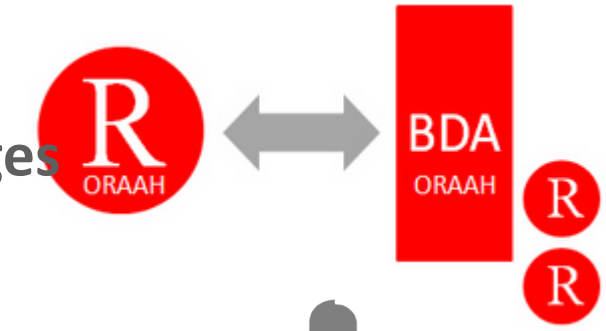
Oracle R Advanced Analytics for Hadoop



- 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:

Using Hadoop and HIVE, plus R Engine and Open-Source R Packages



Hadoop Cluster

with Oracle R Advanced Analytics for Hadoop

HQL Basic Statistics, Data Prep, Joins and View creation

HQL

HIVE

HDFS Access, Store, Load, Data Prep and Transform.

- ORAAH Spark algorithms: Deep Neural, GLM, LM
- Spark MLlib algorithms: LM, GLM, LASSO, Ridge Regression, Decision Trees, Random Forests, SVM, k-Means, PCA
- Open-source R packages distributed via Map-Red function in R

Spark + Spark MLlib

Rⁿ ORACLE

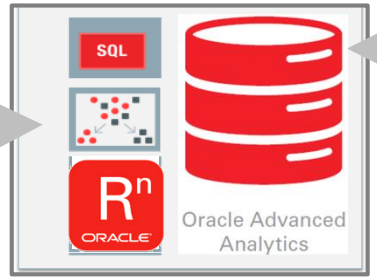
B

R Client

R Analytics
Oracle R Advanced Analytics for Hadoop

Oracle Database Server

with Advanced Analytics option



SQL Client

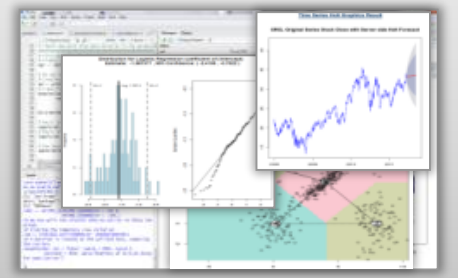
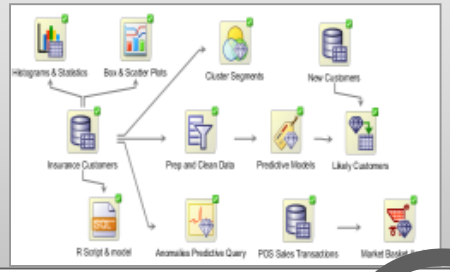
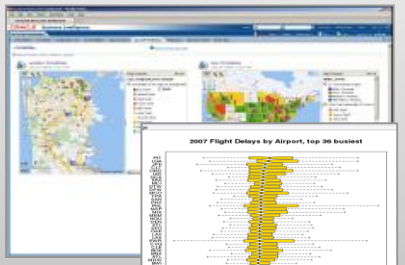
SQL Developer
Other SQL Apps

ORACLE

Oracle's Advanced Analytics

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

	Information Producers		*Information Consumers*	
Users	R programmers	Data & Business Analysts	Business Analysts/Mgrs	Domain End Users

R Client	SQL Developer/ Oracle Data Miner	OBIEE	Applications
			

Platform

Hadoop

Oracle R Advanced Analytics for Hadoop

Parallel, distributed algorithms

Oracle Database Enterprise Edition

Oracle R enterprise in Database

SQL Data Mining & Analytic Functions

+ R Integration for Scalable,

Distributed, Parallel in-Database ML Execution

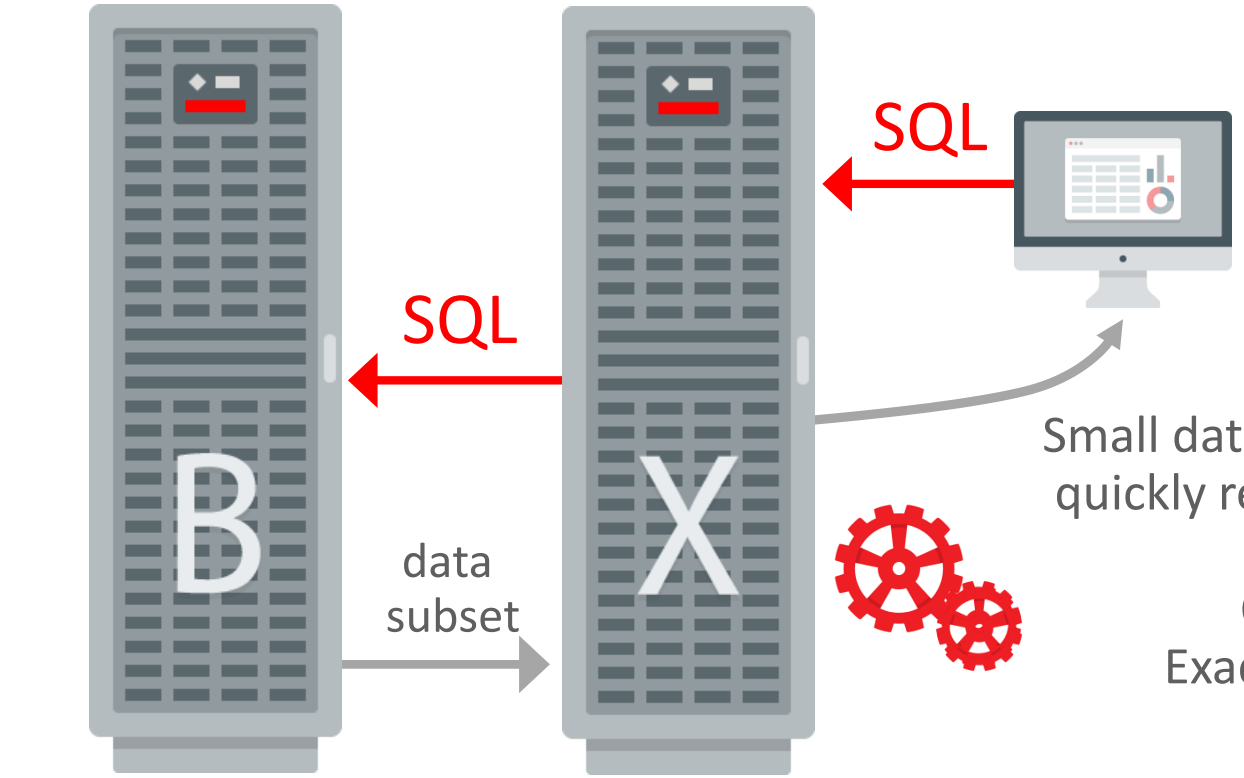
Big Data SQL

Push down SQL predicts to storage layers

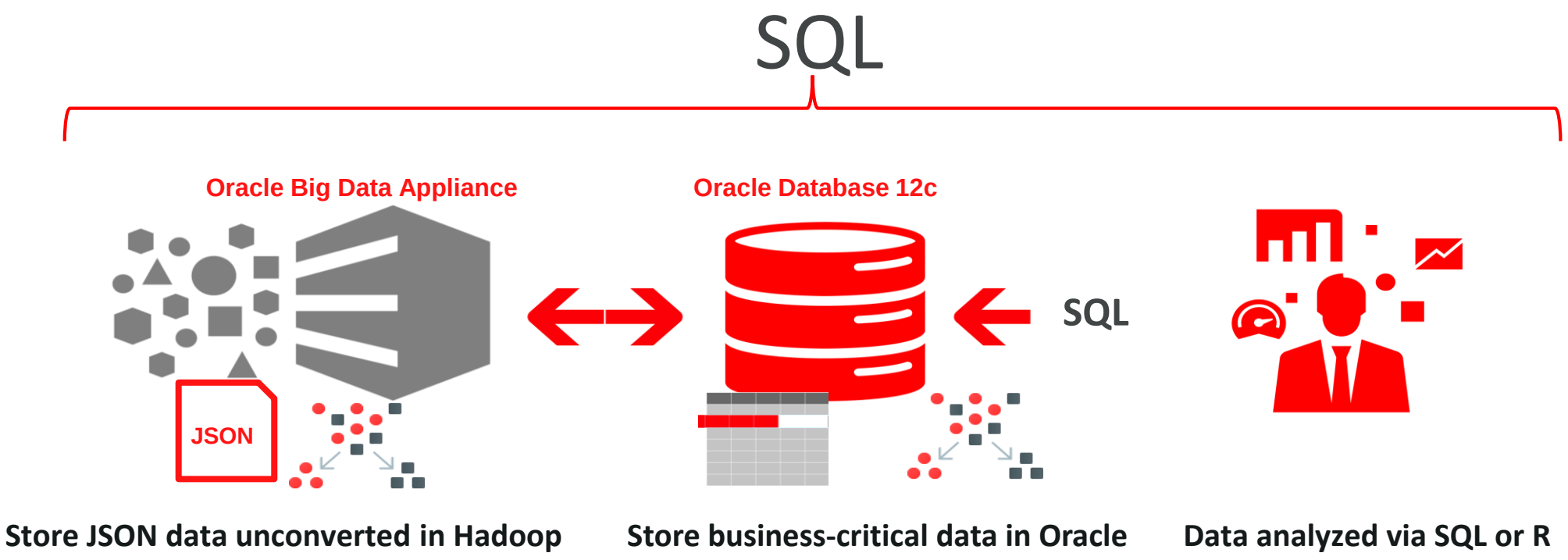
ORACLE

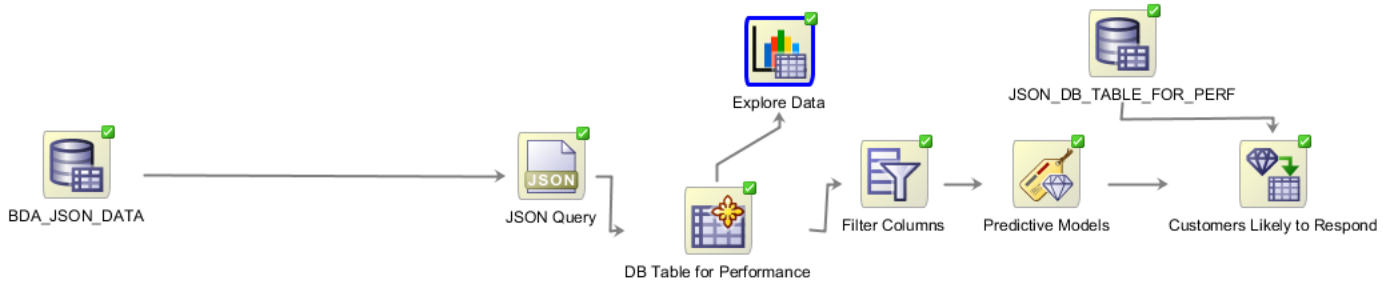
Introducing Oracle Big Data SQL

Massively Parallel SQL Query across Oracle, Hadoop and NoSQL



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





Oracle’s Advanced Analytics

Predictive Applications + OBIEE Integration

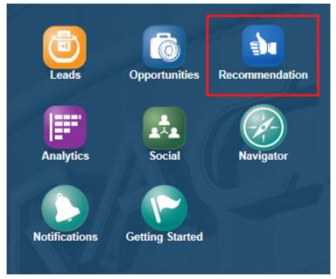
ORACLE

Enabling “Predictive” Enterprise Applications

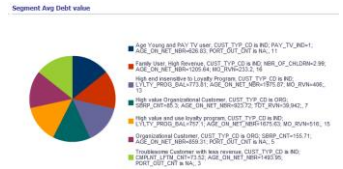
Oracle Applications Using Oracle Advanced Analytics—Partial List

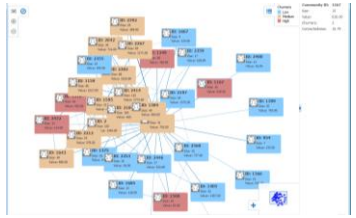
- **Oracle HCM Fusion**
  - Employee turnover and performance prediction and *“What if?”* analysis

- **Oracle CRM Fusion**
  - Prediction of sales opportunities, what to sell, amount, timing, etc.

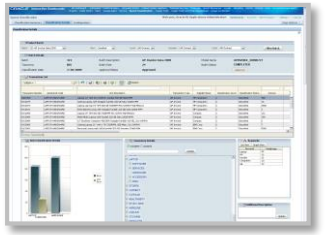


- **Oracle Industry Data Models**
  - **Communications Data Model** churn prediction, segmentation, profiling, etc.
  - **Retail Data Model** loyalty and market basket analysis
  - **Airline Data Model** analysis frequent flyers, loyalty, etc.
  - **Utilities Data Model** customer churn, cross-sell, loyalty, etc.





- **Oracle Retail Customer Analytics**
  - ”Shopping cart analysis” and next best offers
- **Oracle Customer Support**
  - Predictive Incident Monitoring (PIM)
- **Oracle Spend Classification**
  - Real-time and batch flagging of noncompliance and anomalies in expense submissions



- **Oracle FinServ Analytic Applications**
  - Customer Insight, Enterprise Risk Management, Enterprise Performance, Financial Crime and Compliance
- **Oracle Adaptive Access Manager**
  - Real-time security and fraud analytics

Hardware and Software
Engineered to Work Together


ORACLE®