

OpenWorld 2017 Machine Learning 101

Aali Masood
Peter Jeffcock

11:30 pm, Oct 3rd, 2017
Moscone West 3002

ORACLE

ORACLE
OPEN
WORLD

October 1-5, 2017
SAN FRANCISCO, CA



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.

Program Agenda

- 1 Introduction
- 2 Machine Learning Techniques
- 3 Oracle Solution
- 4 Customer Stories
- 5 Questions

Have You Ever...

Had a credit card transaction unexpectedly (and incorrectly) declined



Received a personalized email, direct mail or web ad



Been influenced by a recommendation from your music player or an online shopping site



Had an unexpected bonus / incentive to stay with a company





Machine Learning is based on algorithms that can learn from data without relying on rules-based programming.

McKinsey

CLASSIFYING CUSTOMERS

Basic Query

Features

Age / Gender

Marketing Preferences






Basic Analytics

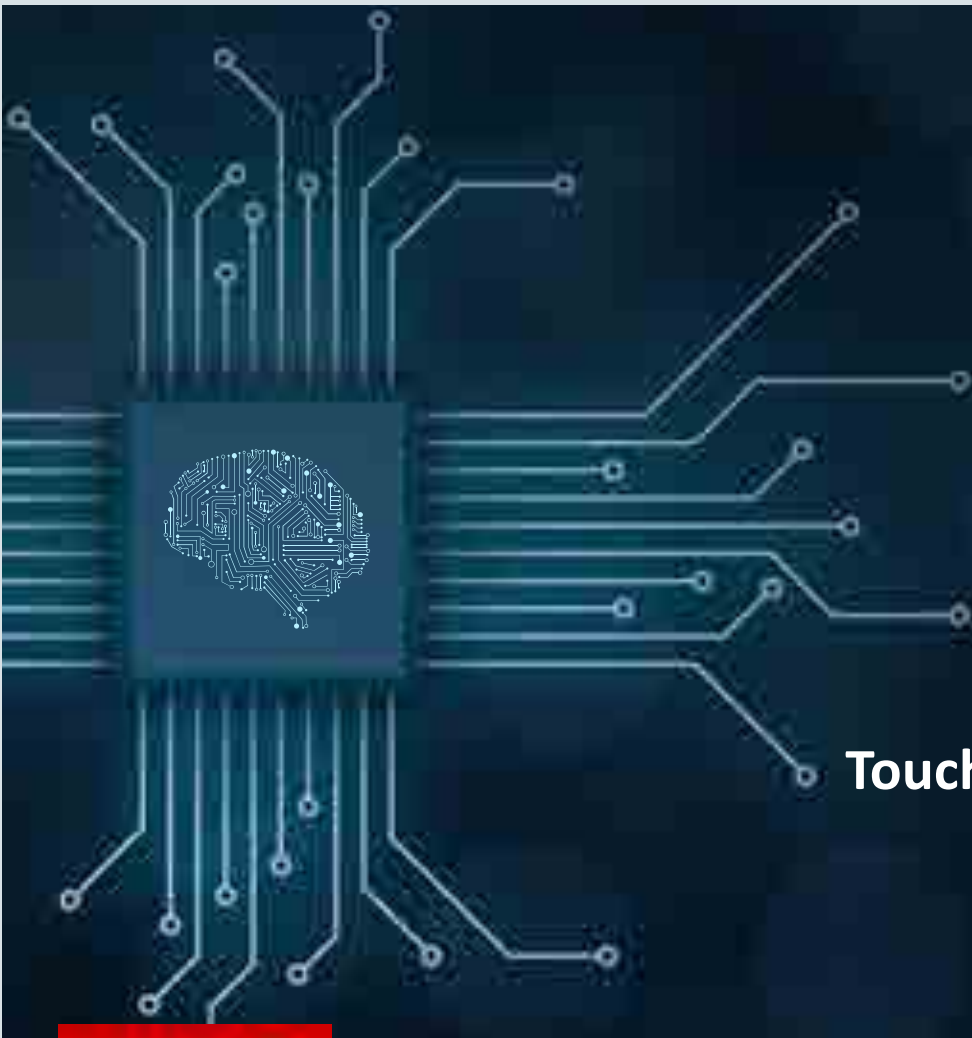
RFM (Recency, Frequency and Monetary Value): Purchases in the Last 3/6/12 mo.

Machine Learning

Behavioral Customer Segment

Probability to Buy New Product X

				
Known	Known	Known	Unknown	Unknown
Mail and e-mail	e-mail	e-mail and Facebook	e-mail and Google+	Mail, e-mail and Twitter
1 item / \$35 in the last 3 mo	2 items / \$150 in the last 6 mo	3 items / \$75 in the last 3 mo	3 items / \$225 in the last 12 mo	9 items / \$250 in the last 6 mo
"Retired Cosmopolitan"	"Affluent Executive"	"New Home Mom"	"Young Successful startup"	"Executive product collector"
31%	45%	55%	21%	72%



Machine Learning Techniques

- Focus on**
- ▶ Regression
 - ▶ Classification
 - ▶ Clustering
 - ▶ Anomaly Detection

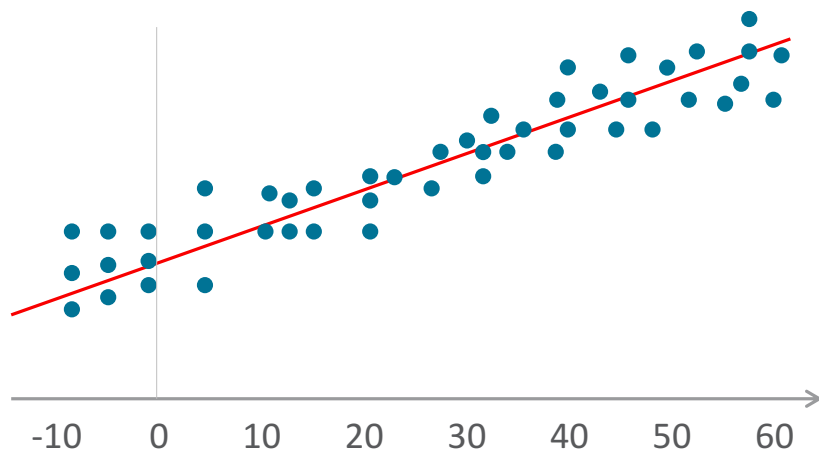
- Touch on**
- ▶ Association Rules
 - ▶ Time Series
 - ▶ Neural Networks

Regression



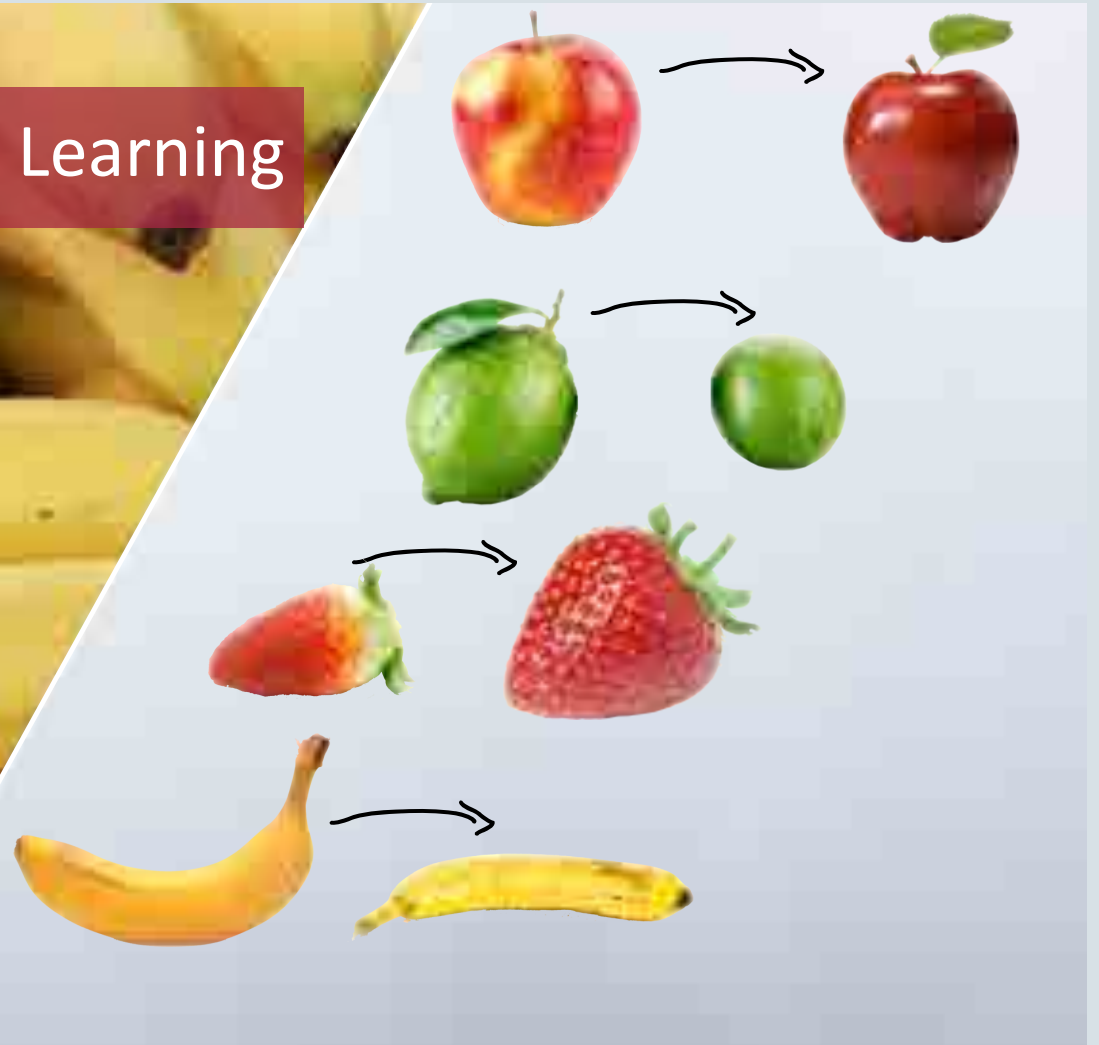
ORACLE

○ REGRESSION

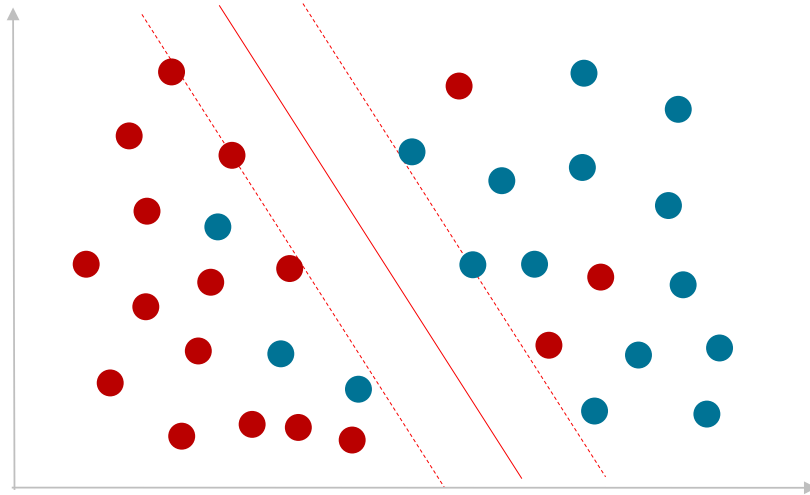


- **Predicting numbers**
- Customer lifetime value
- Estimate optimal pricing
- House price estimates

Classification – Supervised Learning



○ CLASSIFICATION

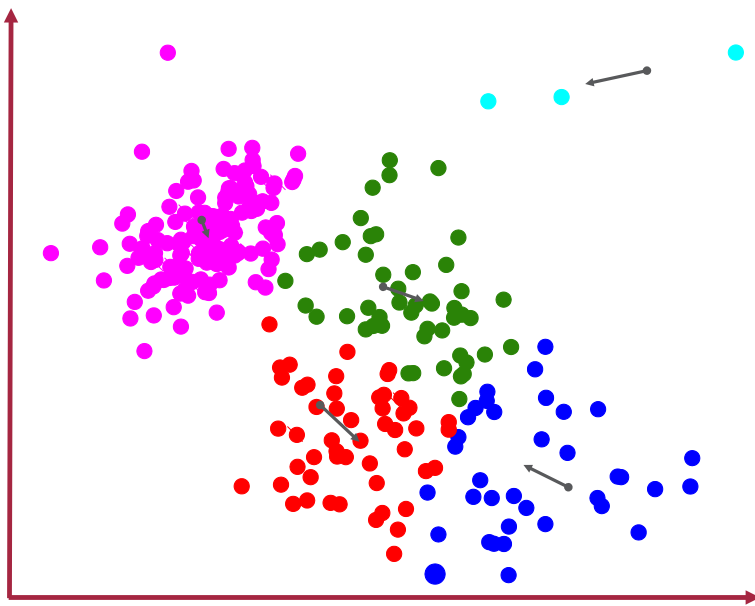


- **Membership of a known class**
- Identify likely high value customers
- Find customer likely to churn
- Fraud detection

Clustering – Unsupervised Learning



○ CLUSTERING



- **Membership of an inferred class**
- Customer segmentation
- Credit risk evaluation
- Document similarity

Anomaly Detection



ORACLE

ANOMALY DETECTION

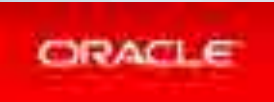
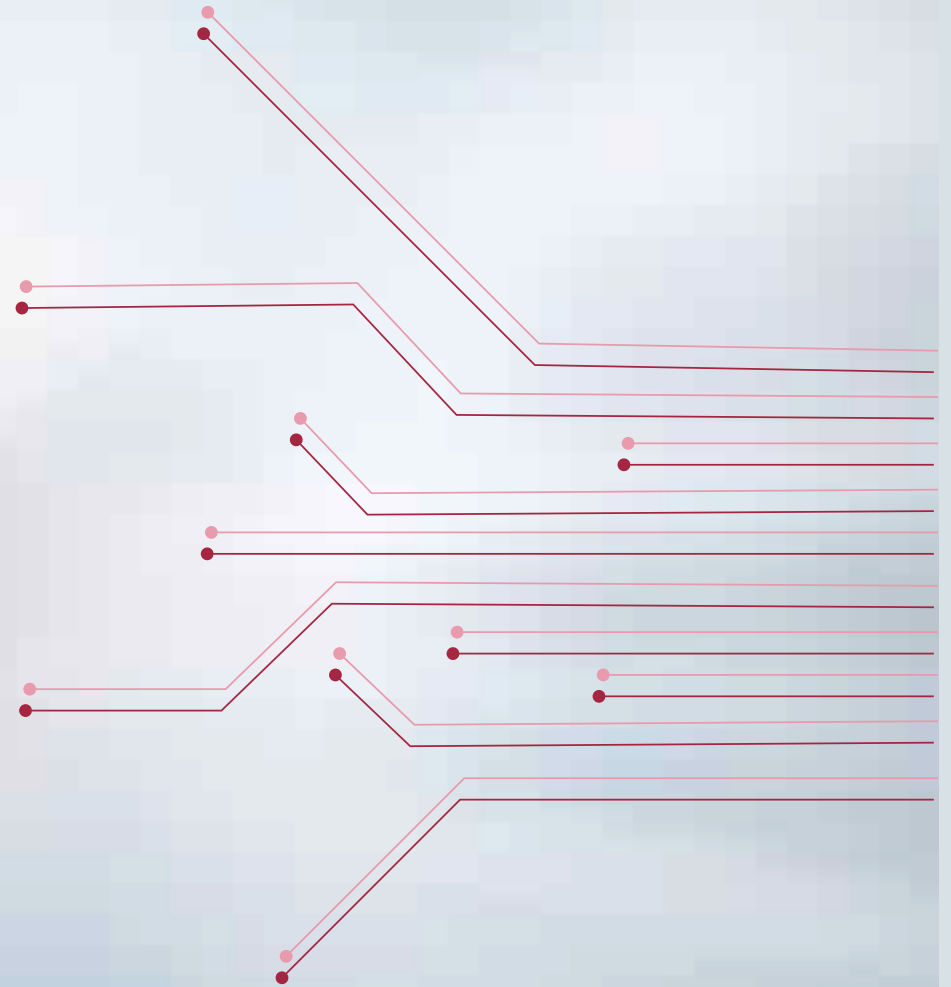


- **Outliers**
- Dentist billing
85 fillings / hour
- Employees with high claims
/ grade
- One variable moving out
of sync

▶ Association Rules

▶ Time Series

▶ Neural Networks



ASSOCIATION RULES



- **Finding like-minded people**
- You might be interested in...
- Root cause analysis
- Identify “harbingers of failure”

○ TIME SERIES

○ **Temporal Aspect**

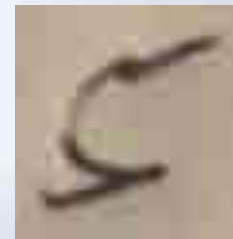
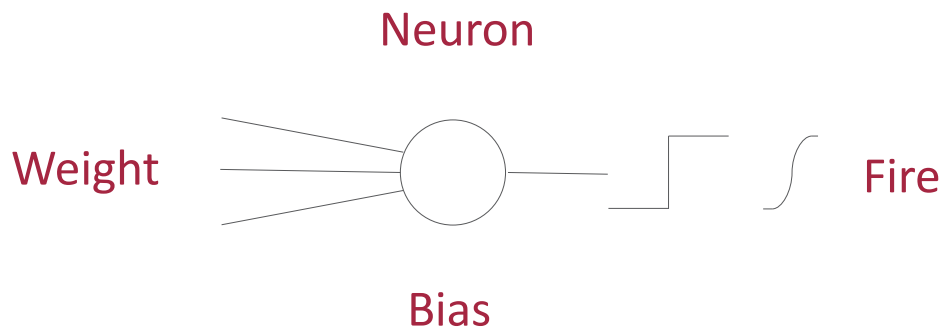
○ Hitting a threshold

○ Forecasting energy use

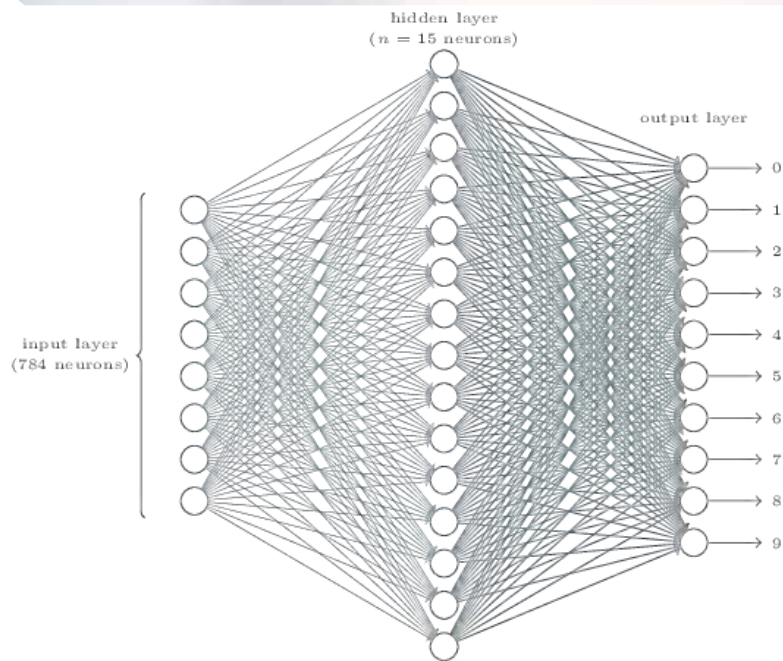
○ Seasonality of data



NEURAL NETWORKS



NEURAL NETWORKS



- Learn (More) Like A Human
- Classification
- Regression
- Deep learning

<http://neuralnetworksanddeeplearning.com/chap1.html>

ORACLE

Oracle Machine Learning Algorithms

CLASSIFICATION

- Logistic Regression
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine
- Naïve Bayes
- Explicit Semantic Analysis
- Gaussian Mixture Models

CLUSTERING

- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization

ANOMALY DETECTION

- One-Class Support Vector Machine

REGRESSION

- Generalized Linear Model
- Support Vector Machine
- Random Forest
- Linear Model
- Stepwise Linear regression
- LASSO

ASSOCIATION RULES

- A priori

ATTRIBUTE IMPORTANCE

- Minimum Description Length
- Principal Component Analysis
- Unsupervised Pairwise KL Divergence

SQL PREDICTIVE QUERIES

ALGORITHM TEXT SUPPORT

- Algorithms support text type
- Tokenization and theme extraction
- Document similarity

FEATURE EXTRACTION

- Principal Component Analysis
- Non-negative Matrix Factorization
- Singular Value Decomposition

TIME SERIES

- Single Exponential Smoothing
- Double Exponential Smoothing

OPEN SOURCE ML ALGORITHMS

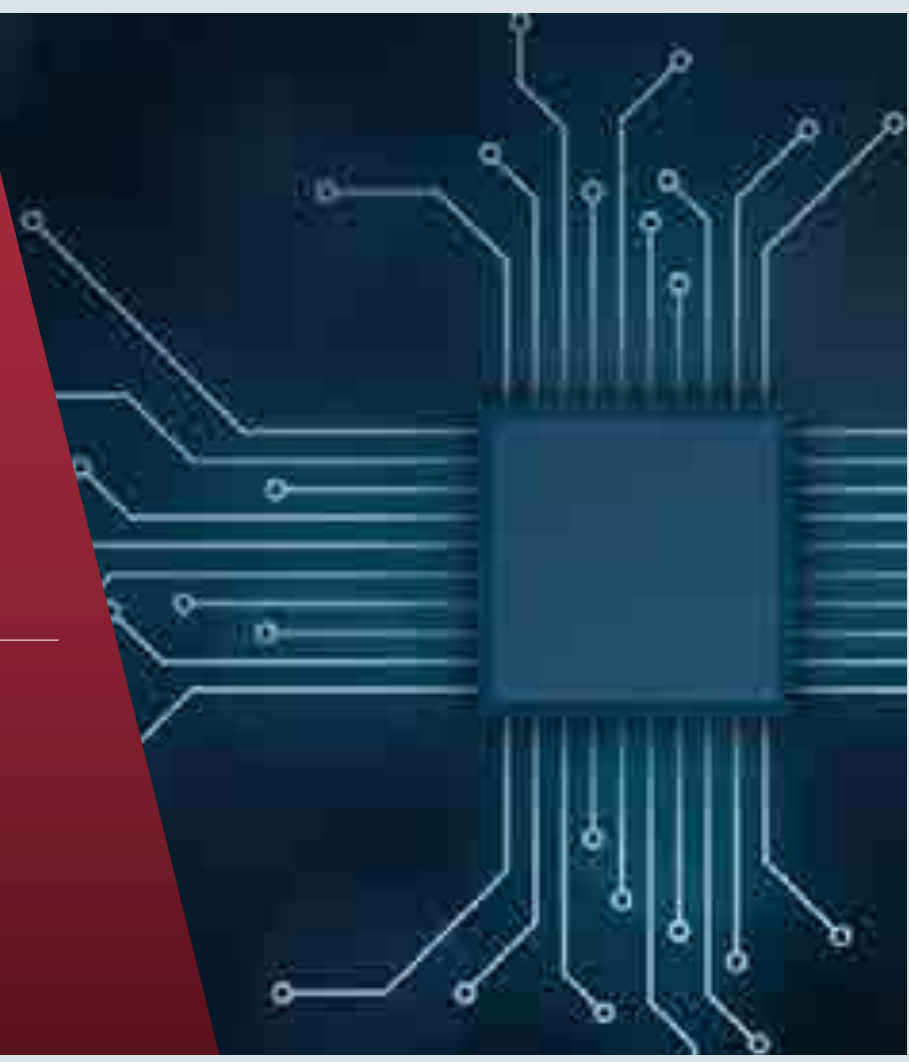
- CRAN R Algorithm Packages through Embedded R Execution
- Spark MLlib algorithm integration

Build Your Own Models

Database
Cloud

Big Data
Cloud

Analytics
Cloud



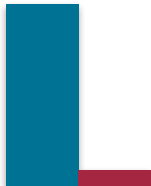
The Leading Choice for Machine Learning



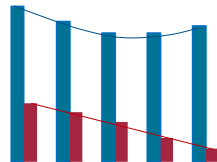
FAST

Optimized in-memory algorithms

Up to 32X
Faster



Scales above
8-16 nodes



OPEN

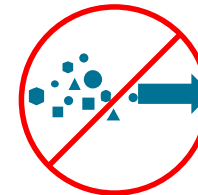
Based on open source



INTEGRATED

Bringing analytics to the data

Don't move
the data



ORACLE

National Health Service



- Finding savings in healthcare budget

DX Marketing



- Predicting propensity to purchase

StubHub



- Real-time prediction of fraudulent ecommerce transactions

Zagreb Bank



- Customer 360 and credit risk analysis

National Tax Authority

- Detecting tax fraud
- Classification and anomaly detection

Financial Payments

- Prevent, monitor, understand and anticipate fraud
- Classification (generalized linear models and decision trees)

Retail

- Customer loyalty and segmentation
- Association rules, clustering, classification

Bank

- Accelerate product launch, evaluate credit risk
- Neural networks

Integrated Cloud

Applications & Platform Services

ORACLE®