



Oracle Machine Learning Overview

Mark Hornick

Oracle Machine Learning Product Management



Mark Hornick

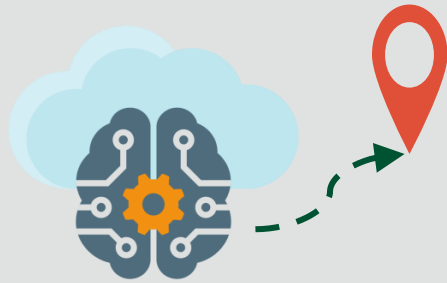
Senor Director
Data Science and Machine Learning
Product Management
Oracle



Safe Harbor

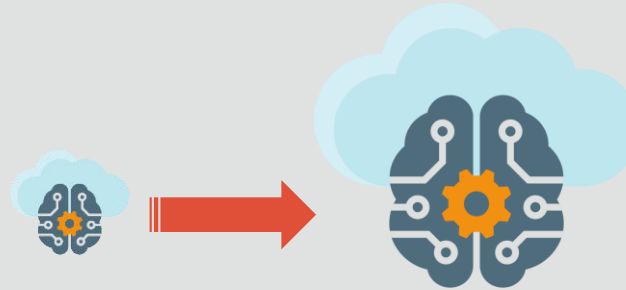
The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development,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

Oracle Machine Learning **Key Attributes**



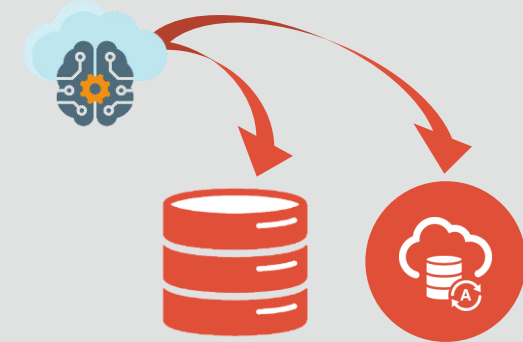
Automated

Get better results faster with less effort – even non-expert users



Scalable

Handle big data volumes using parallelized, distributed algorithms – no data movement

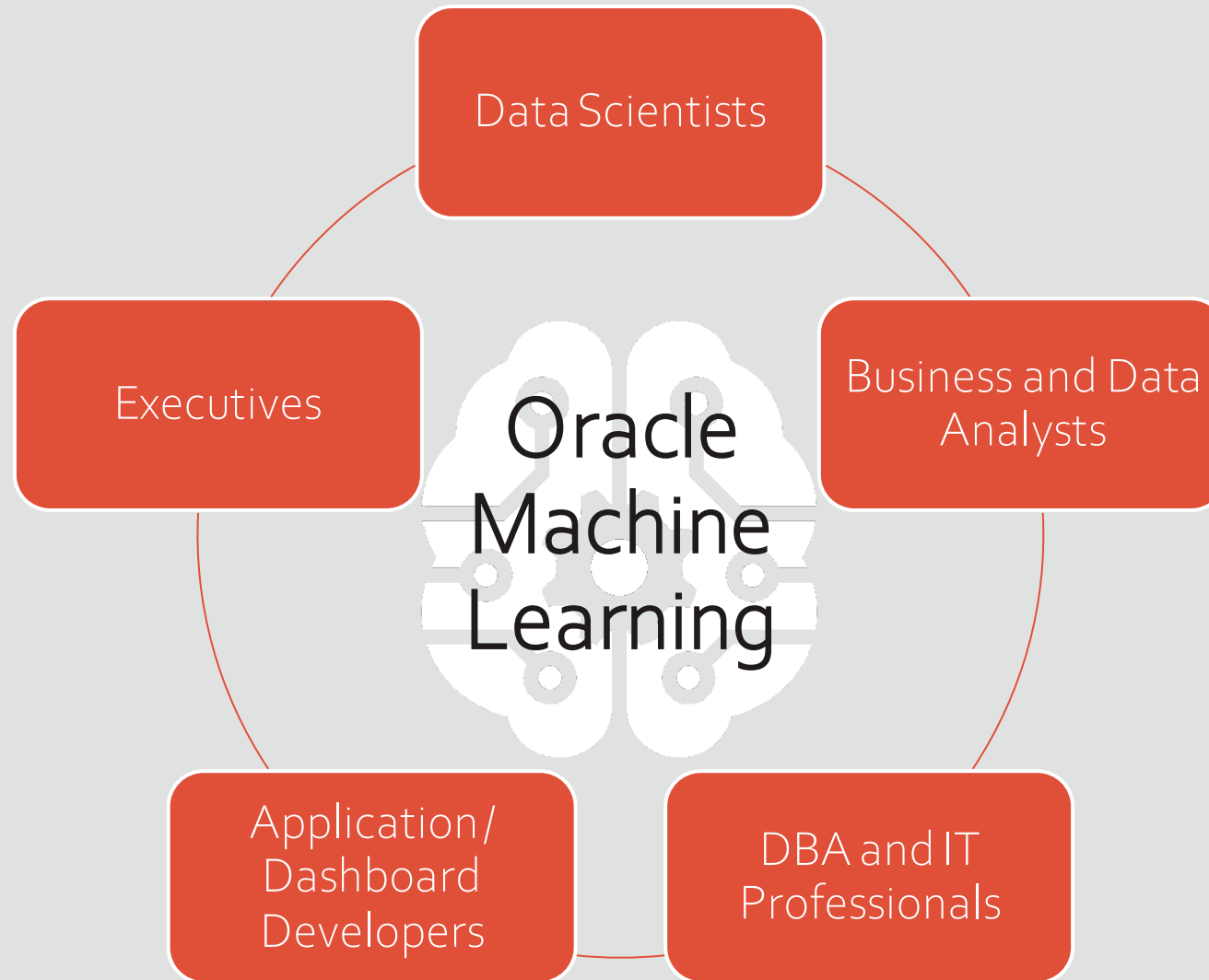


Production-ready

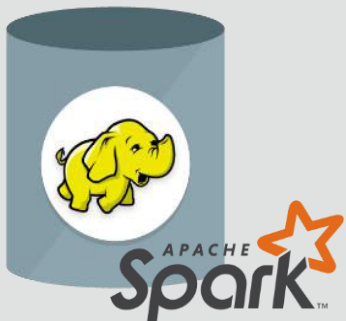
Deploy and update data science solutions faster with integrated ML platform

Increase productivity | Achieve enterprise goals | Innovate more

Empowering Enterprise Users



ORACLE
APPLICATIONS



Oracle Machine Learning

OML4SQL
SQL API

OML Notebooks
with Apache Zeppelin on
Autonomous Database

OML4R
R API

Oracle Data Miner
Oracle SQL Developer extension

OML4Py*
Python API

OML4Spark
R API on Big Data

OML AutoML UI*
Code-free AutoML interface
on Autonomous Database

OML Services*
Model Deployment and Management,
Cognitive Text

ORACLE
APPLICATIONS



* Coming soon

Oracle Machine Learning Algorithms and Analytics

CLASSIFICATION

Naïve Bayes
Logistic Regression (GLM)
Decision Tree
Random Forest
Neural Network
Support Vector Machine (SVM)
Explicit Semantic Analysis
*XGBoost**

ANOMALY DETECTION

One-Class SVM
*MSET-SPRT**

CLUSTERING

Hierarchical K-Means
Hierarchical O-Cluster
Expectation Maximization (EM)

TIME SERIES

Forecasting - Exponential Smoothing
Includes popular models
e.g. Holt-Winters with trends, seasonality, irregularity, missing data

REGRESSION

Linear Model
Generalized Linear Model (GLM)
Support Vector Machine (SVM)
Stepwise Linear regression
Neural Network
LASSO
*XGBoost**

ATTRIBUTE IMPORTANCE

Minimum Description Length
Principal Component Analysis (PCA)
Unsupervised Pair-wise KL Div
CUR decomposition for row & AI

ASSOCIATION RULES

A priori/ market basket

PREDICTIVE QUERIES

Predict, cluster, detect, features

SQL ANALYTICS

SQL Windows
SQL Patterns
SQL Aggregates

FEATURE EXTRACTION

Principal Comp Analysis (PCA)
Non-negative Matrix Factorization
Singular Value Decomposition (SVD)
Explicit Semantic Analysis (ESA)

ROW IMPORTANCE

CUR Decomposition

RANKING

*XGBoost**

TEXT MINING SUPPORT

Algorithms support text columns
Tokenization and theme extraction
Explicit Semantic Analysis (ESA)

STATISTICAL FUNCTIONS

min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.

R AND PYTHON PACKAGES

Third-party R and Python Packages through Embedded Execution
Spark MLlib algorithm integration





Oracle Machine Learning Notebooks

Autonomous Database as a Data Science Platform

Collaborative UI

Based on Apache Zeppelin

Supports data scientists, data analysts, application developers, DBAs with SQL and Python

Easy sharing of notebooks and templates

Permissions, versioning, and execution scheduling

Included with Autonomous Database

Automatically provisioned, managed, backed up

In-database algorithms and analytics functions

Explore and prepare, build and evaluate models, score data, deploy solutions

Soon to be augmented with R

The screenshot shows the Oracle Machine Learning interface. At the top, it says "ORACLE Machine Learning". Below that is a "Back" button. The main title is "Credit Score Predictions". Underneath, there's a section titled "Review Data by Occupation" with a "FINISHED" status. The notebook content includes a SQL query and a pie chart visualization. The SQL query is:

```
%sql
-- This shows an alternative presentation style - a pie chart. Note that Zeppelin
visualizations are limited. In lab 400 we will use Oracle Data Visualization to
create more more interesting perspectives.

select customer_id, age, income, tenure, loan_type, loan_amount, occupation,
marital_status
from credit_scoring_100k_v where rownum < 1000
```

 The pie chart shows the distribution of occupations: Professional (blue), Clerical (light blue), Farmer (orange), Manager (light orange), Worker (green), NaN (light green), Army (red), and Technician (pink).



Oracle Machine Learning for SQL

Empower SQL users with immediate access to ML included with Oracle Database and Oracle Autonomous Database

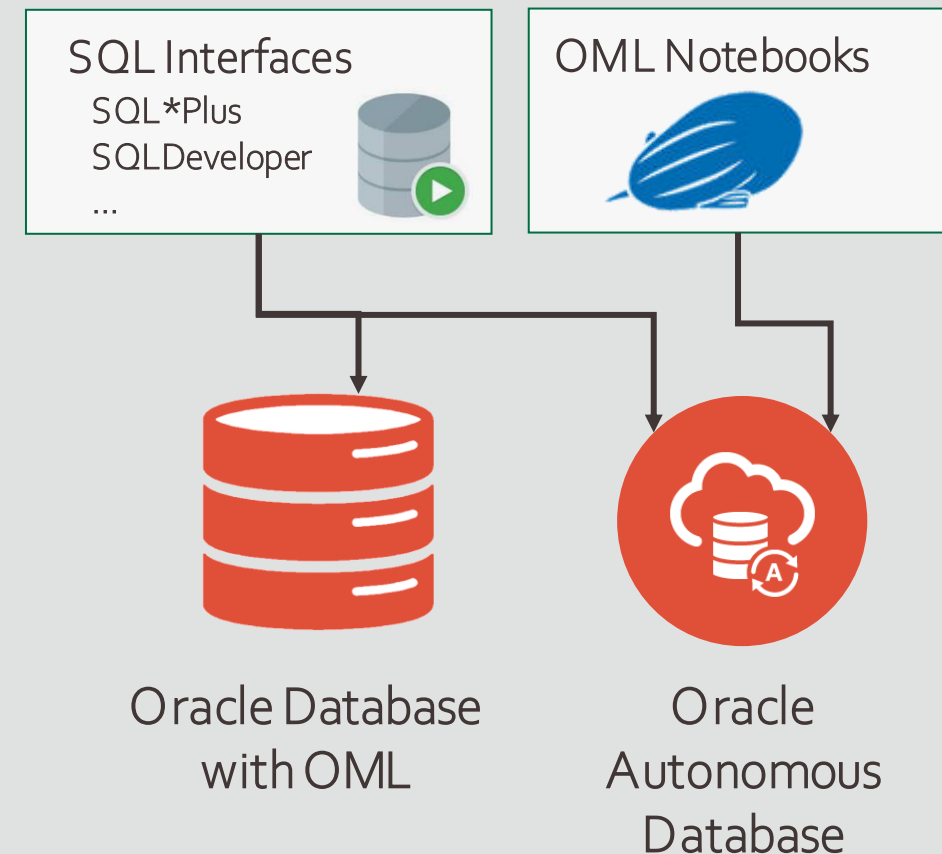
In-database, parallelized, distributed algorithms

- No extracting data to separate ML engine
- Fast and scalable
- Batch and real-time scoring
- Explanatory prediction details

ML models as first class database objects

- Access control via permissions
- Audit user actions
- Export / import models across databases

Leverage ML across Oracle stack



Oracle Data Miner User Interface

Create analytical workflows – productivity tool for data scientists – enables citizen data scientists



- SQL Developer Extension for Oracle Database on-premise and DBCS
- Automates typical data science steps
- Easy to use drag-and-drop interface
- Analytical workflows quickly defined and shared
- Wide range of algorithms and data transformations

The screenshot displays the Oracle SQL Developer interface with the Oracle Data Miner extension. The main workspace shows a workflow diagram with the following steps: Clustering Segmentation 1 (outputting CUST_INSUR_LTV1), Filter Columns (outputting Filter Columns Details 1), Multiple Classification Models (outputting Most Likely Customers), and Most Likely Customers (outputting LIKELY_BUY_INSURANCE_CUSTMRS 1). A 'Query Builder' window is open, showing a query to identify suspicious fraud policy holder claims.

```
begin
dms_data_mining.create_model('CLAIMSMODEL', 'CLASSIFICATION',
'CLAIMS', 'POLICYNUMBER', null, 'CLAIMS_SET');
end;
```

Query Result

POLICYNUMBER	PERCENT_FRAUD	RNK
1	654	61.87
2	11068	57.37
3	7435	55.47

Generate SQL code for immediate deployment

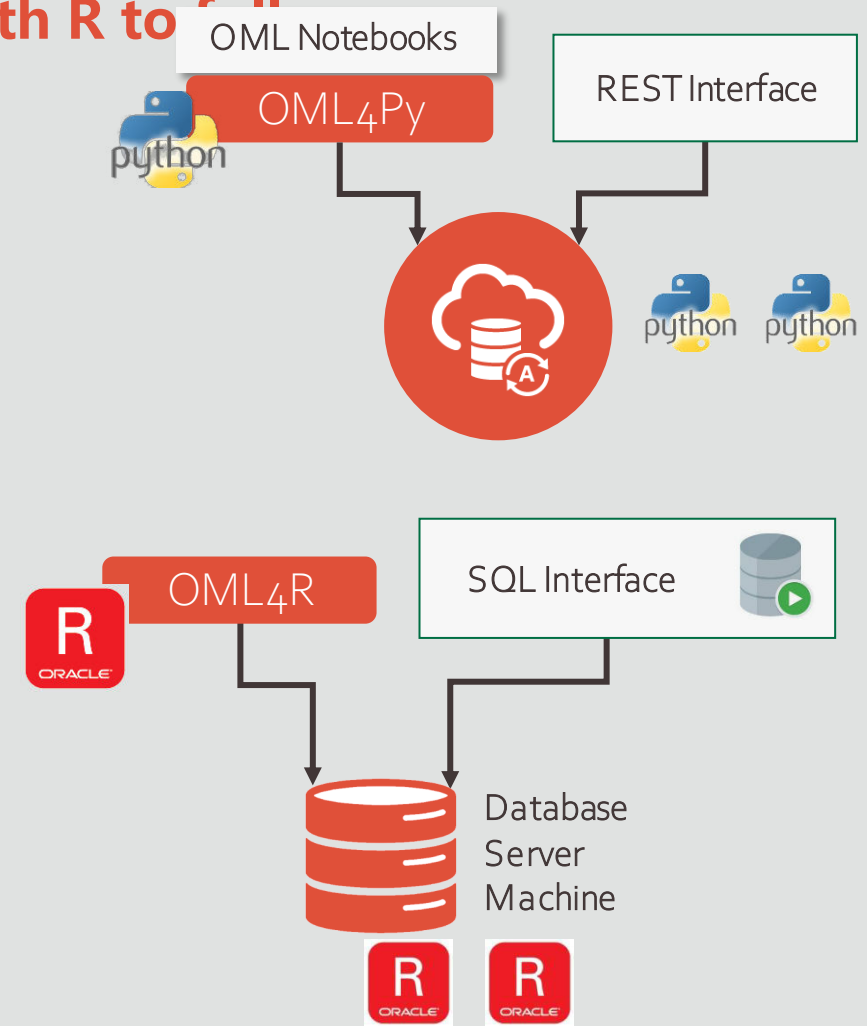


Oracle Machine Learning for R and Python

Oracle Database – R (Python coming soon)

Oracle Autonomous Database – Python coming soon, with R to follow

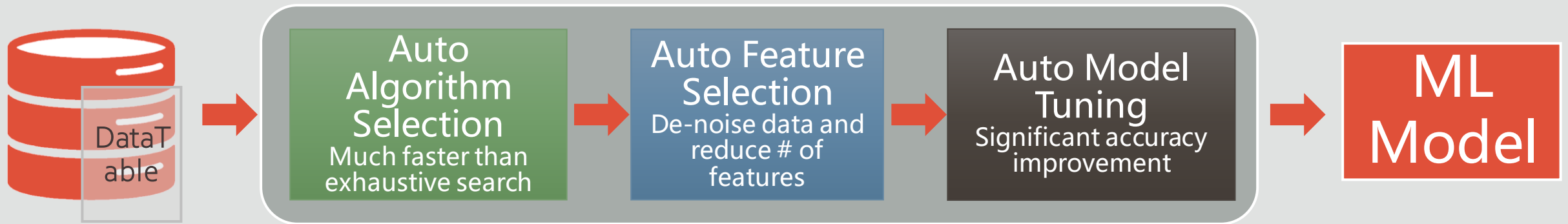
- Oracle Database as HPC environment
- In-database parallelized and distributed machine learning algorithms
- Manage scripts and objects in Oracle Database
- Integrate results into applications and dashboards via SQL or REST
- OML4Py automated machine learning
- Empower data scientists with open source environments***





AutoML – *new* with OML4Py

Increase data scientist productivity – reduce overall compute time



Auto Algorithm Selection

- Identify in-database algorithm that achieves highest model quality
- Find best algorithm faster than with exhaustive search

Auto Feature Selection

- Reduce # of features by identifying most predictive
- Improve performance and accuracy

Auto Model Tuning

- Automatic tuning of algorithm hyperparameters
- Significantly improve model accuracy
- Avoid manual or exhaustive search techniques

Enables non-expert users to leverage Machine Learning

Oracle Machine Learning for Spark



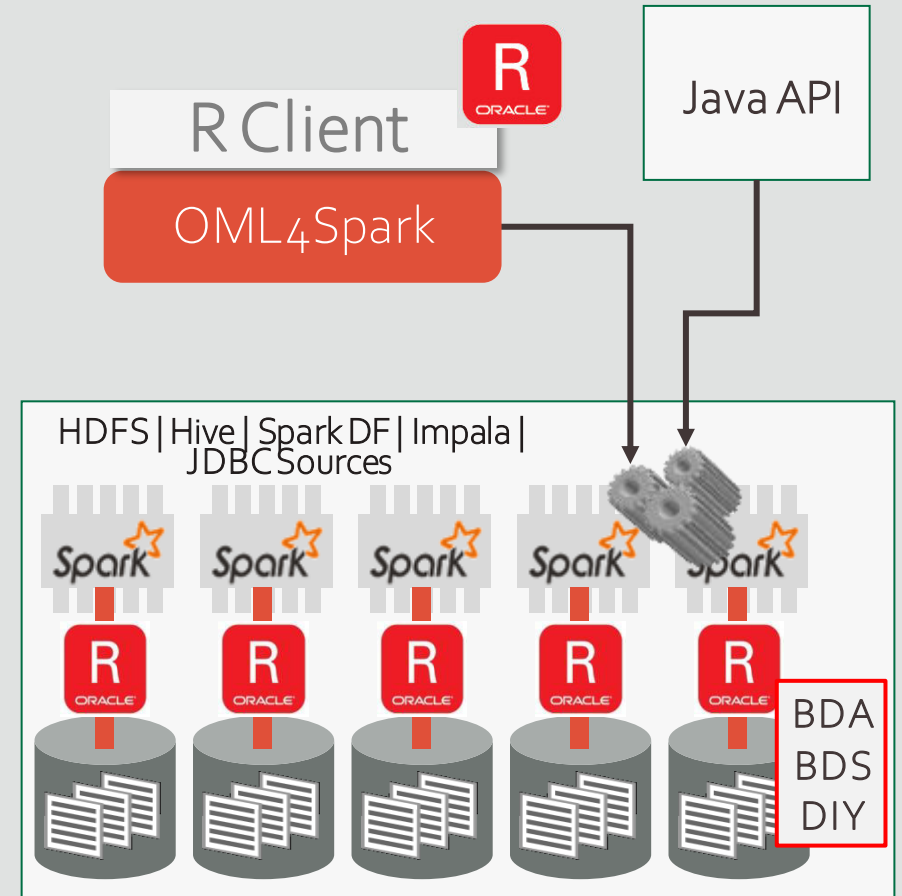
R Language API Component to Oracle Big Data Connectors

Leverage Spark 2 environment for powerful data preparation and machine learning

Use data across range of Data Lake sources

Achieve scalability and performance using full Hadoop cluster

Parallelized and distributed ML algorithms from native and Spark MLlib implementations



Oracle Applications integrating OML

HCM Cloud
Workforce Predictions

CRM Sales Cloud
Sales Prediction

Retail GBU
Customer Insights,
Customer Segmentation

Adaptive Intelligent Applications
for Manufacturing

Configure, Price, Quote Cloud

Content and Experience
Unstructured Data Analytics

Integration Cloud
Digital Process Automation

Copyright © 2020 Oracle and/or its affiliates.

ORACLE
HUMAN CAPITAL
MANAGEMENT
CLOUD

ORACLE
SALES CLOUD

ORACLE
RETAIL CLOUD



ORACLE
CONFIGURE, PRICE,
AND QUOTE
CLOUD

ORACLE
CLOUD PLATFORM
CONTENT AND
EXPERIENCE



Industry Data Models
Communications, SNA, Utilities, Airlines, Retail, ...



EBS Spend Classification
Organize spend into logical categories

EBS Depot Repair
Optimize speed, cost, quality of
product repair, reuse, recycling

Identity Management
Adaptive Access Management

FSGBU
Analytical Applications
Infrastructure

ORACLE
E-BUSINESS SUITE

ORACLE
IDENTITY MANAGEMENT

ORACLE
FINANCIAL SERVICES



Why Oracle for Machine Learning?

Oracle integrates ML across the Oracle Stack and the Enterprise

Empower data scientists and analysts, developers, and DBAs/IT with ML

Eliminate costly data movement and latency

Fast and scalable data exploration, data preparation, and ML algorithms

Over 30 in-database algorithms supporting: regression, classification, time series,

clustering, feature extraction, anomaly detection

Automate key ML process steps

R and Python integration supports data scientists

Ease of ML model and R/Python script deployment

Leverage existing backup, recovery, and security mechanisms and protocols

That's where most enterprise data lives – bring the algorithms to the data!



Roadmap: Expanding Oracle's investment in machine learning

Key focus areas for OML

- **Extend Oracle data management platform**
Database as a platform for machine learning/data science
- **Support data science teams with multiple personas using multiple languages**
Data scientists, business/data analysts, application/dashboard developers
SQL, Python, R
- **Provide a platform for application integration**
SQL and REST
- **Enable machine learning through multiple interfaces**
Apache Zeppelin, No-code AutoML UI
Oracle Analytics Cloud, OCI Data Science

Coming soon...

Roadmap: Expand Autonomous Database with Python

Autonomous Database as a Data Science Platform

OML Notebooks add support for Python

In addition to SQL, PL/SQL, and Markdown

Scalable Python execution (OML4Py)

Transparency layer-enabled database functionality

In-database machine learning algorithms

Automatic Machine Learning (AutoML)

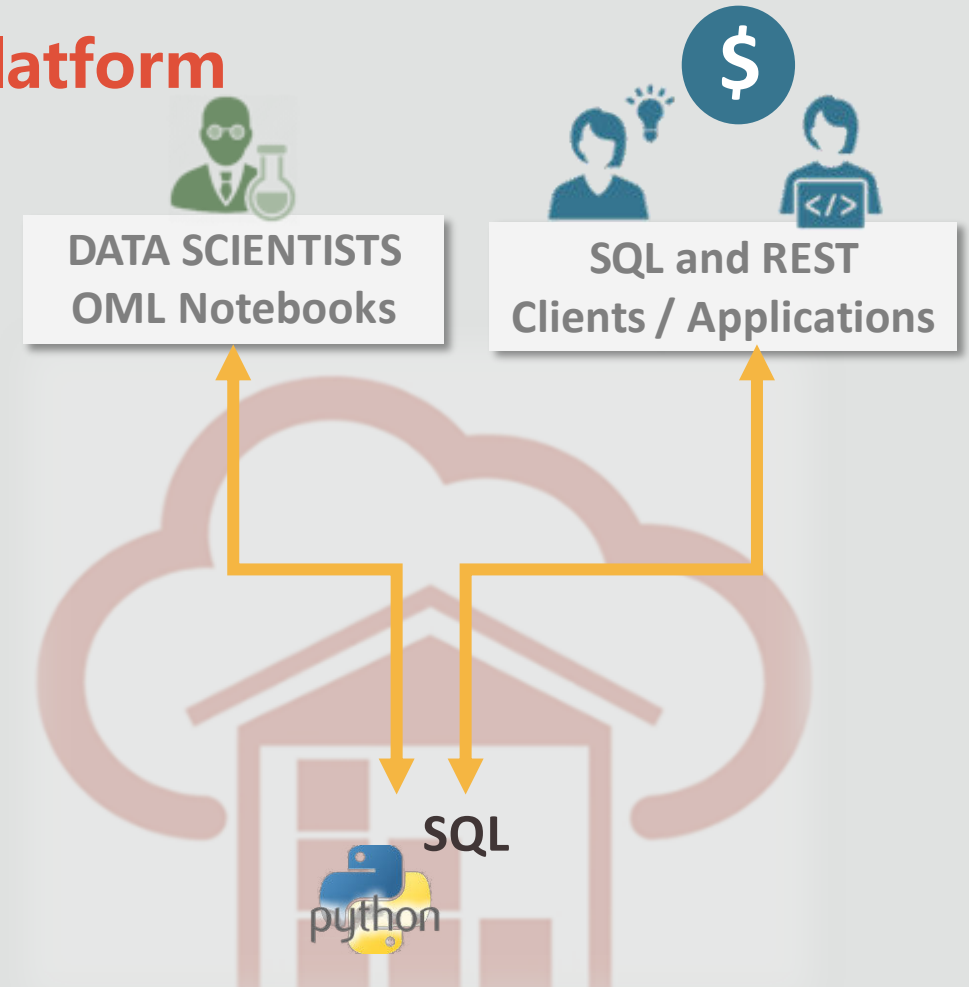
Algorithm and feature selection

Model tuning

Python scripts managed in-database

Invoke from OML Notebooks and REST APIs

Deploy easily into Web applications



Roadmap: OML Services

OML-specific REST APIs – develop and deploy models outside the database

Model Management and Deployment Services

Build and deploy OML models

Monitor models for accuracy and prediction/predictor drift

Models in OML format and ONNX format

Import ONNX for Tensorflow, PyTorch, MXNet, scikitlearn, etc.

Store, version, compare ML models

Shared authentication with OML4Py REST API

Cognitive Text Services

Extract topics and keywords

Sentiment analysis

Text summary and similarity

Model Management

GET /models
GET /{model name}
GET /{model name}/{version}
POST /{model name}
POST /{model name}/{version}
DELETE /{model name}/{version}

Model Deployment

GET /models
GET /{uri}
GET /{uri}/api
POST /{uri}
POST /{uri}/score
DELETE /{uri}

Cognitive Text

POST /topics
POST /keywords
POST /sentiment
POST /summary
POST /similarity

Roadmap: OML AutoML UI

“Code-free” AutoML-based user interface supporting automated end-to-end ML

Powerful, easy to use UI

Enable non-expert users to use ML

Automate model build and deployment

Enhance data scientist productivity

Support model management

Features

Minimal user input: data, target

Model leaderboard

Model deployment via REST endpoints

The screenshot displays the Oracle Machine Learning AutoML UI interface. The top section shows the experiment name 'AutoML Experiment Demo' and its status 'Running'. A 'Metric Chart' displays a line graph showing performance over time. Below the chart is a 'Leader Board' table listing various models and their accuracy scores.

Name	Algorithm	Accuracy (default)
Random Forest 1	Random Forest	89
Neural Network 1	Neural Network	87
GLMR 1	Generalized Linear Model (Ridge Regression)	86
GLM 1	Generalized Linear Model	
Decision Tree 1	Decision Tree	

The bottom section shows a 'Features' table with columns for Name, Type, Percent NULLs, Distinct Values, Min, Max, Mean, and Std Dev.

Name	Type	Percent NULLs	Distinct Values	Min	Max	Mean	Std Dev
PROD_CATEGORY	VARCHAR2	0	5				
PROD_CATEGORY_DESC	VARCHAR2	0	5				

Additional features shown include 'Insight Options' for 'SVM Linear 1' and 'Neural Network', and a 'Confusion Matrix' for 'SVM Linear 1'.



FY2021...

Roadmap: Expand Autonomous Database with Python and R

Autonomous Database as a Data Science Platform

OML Notebooks add support for R

R scripts managed in-database

Invoke from OML Notebooks and REST APIs

Deploy into Web applications easily

Scalable R execution

Transparency layer-enabled database functionality

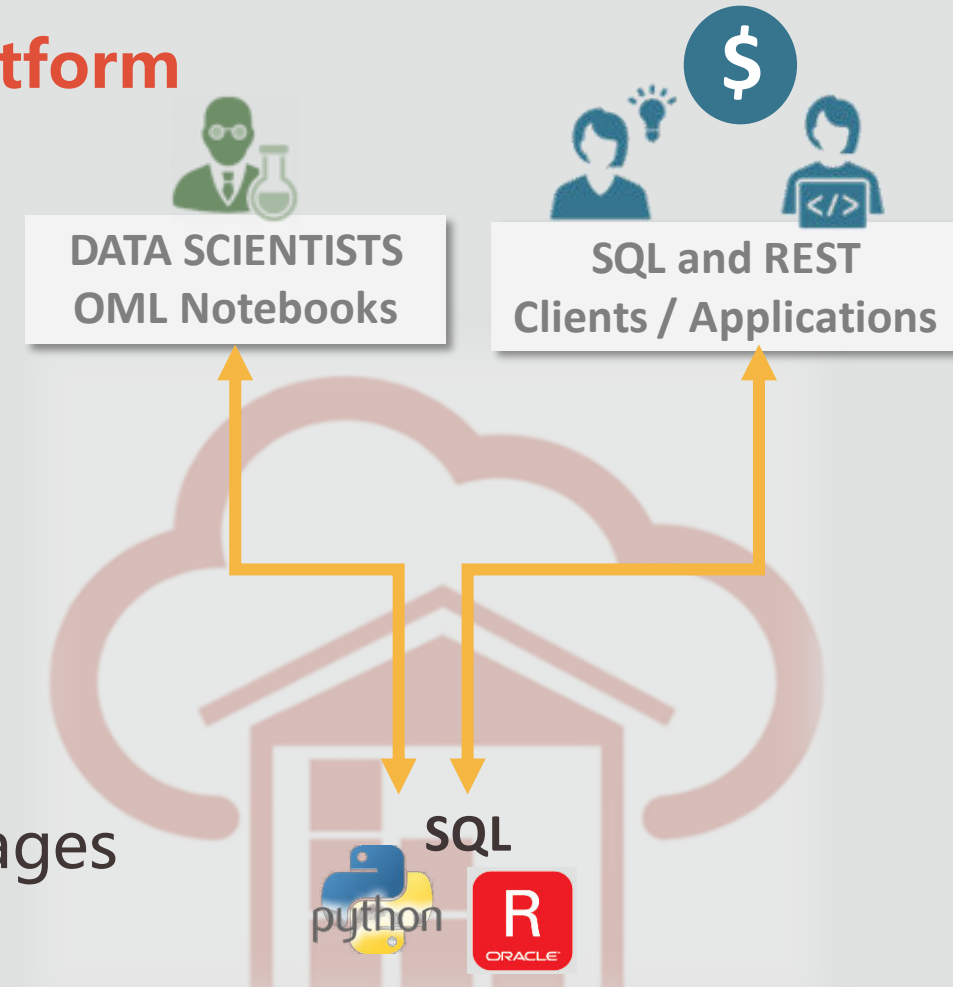
In-database machine learning algorithms

Use external OML4Py and OML4R clients

Python and R scripts invoked from SQL

Extend use of open source Python and R packages

OML4Py integrated with OCI Data Science



Roadmap: OML4R and OML4Py

Expand support for open source languages and ecosystems

Expose additional OML4SQL algorithms to Python and R

Support for recent R and Python releases

Enable Oracle Database standard integrated installation, patching, upgrade/downgrade

OML4Py AutoML introduces *pipeline* function

OML4Py available on premises and DBCS



Roadmap: OML4Spark

New cloud-based architecture with powerful Spark analytics

Enable OML4Py integration

- Add support for OML4Spark algorithms

- Add support for Hive and Impala via transparency layer

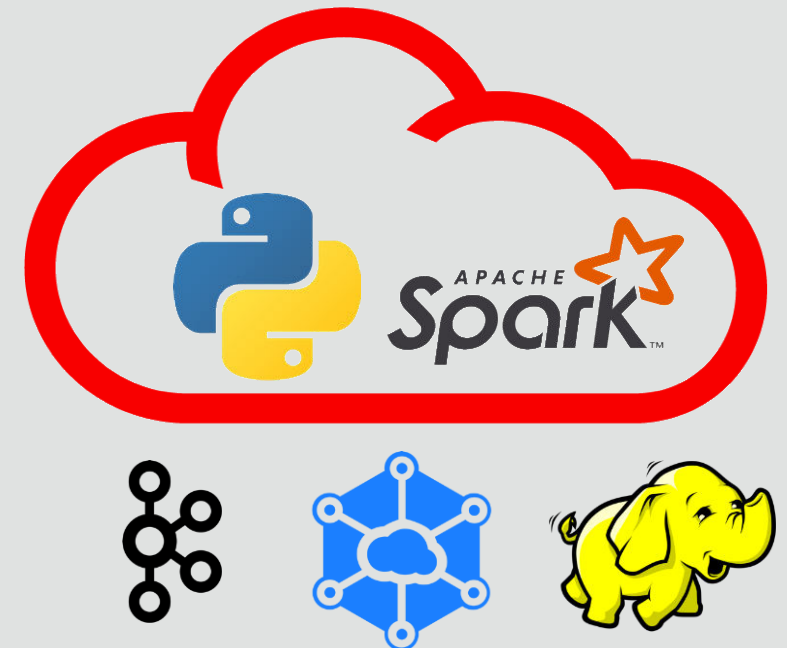
Expand set of natively supported data formats and sources

- Oracle Object Storage

- Spark streaming data

- Parquet, AVRO, RC, ORC, and other Hadoop formats

- SparkSQL via transparency layer



Roadmap: OML Services

OML-specific REST APIs – develop and deploy models outside the database

Extend Model Management and Deployment Services

Enable monitoring for classification and regression models

Roadmap: OML AutoML UI

“Code-free” AutoML-based user interface supporting automated end-to-end ML

Enable model monitoring with model management

Cognitive features for processing text

For more information...

oracle.com/machine-learning

Database / Technical Details /
Machine Learning



Oracle Machine Learning

The Oracle Machine Learning product family enables scalable data science projects. Data scientists, analysts, developers, and IT can achieve data science project goals faster while taking full advantage of the Oracle platform.

Oracle Machine Learning consists of complementary components supporting scalable machine learning algorithms for in-database and big data environments, notebook technology, SQL and R APIs, and Hadoop/Spark environments.

See also [AskTOM OML Office Hours](#)

Thank You



MarkHornick
Oracle Machine Learning Product Management

