Oracle Machine Learning Overview

Mark Hornick, Senior Director
Oracle Machine Learning Product Management
mark.hornick@oracle.com
@MarkHornick

Sample of common enterprise machine learning pain points

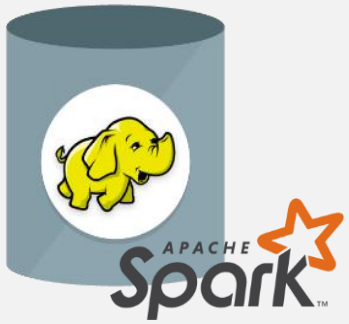
"It takes too long to get my data or to get the 'right' data"

"I can't analyze or mine all of my data – it has to be sampled"

"Putting open source models and results into production takes too long and is ad hoc and complex"

"Our company is concerned about data security, backup and recovery"

"We need to build and score with 100s or 1000s of models fast to meet business objectives"



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

* 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
*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

** New in 20c*

Oracle Machine Learning Notebooks

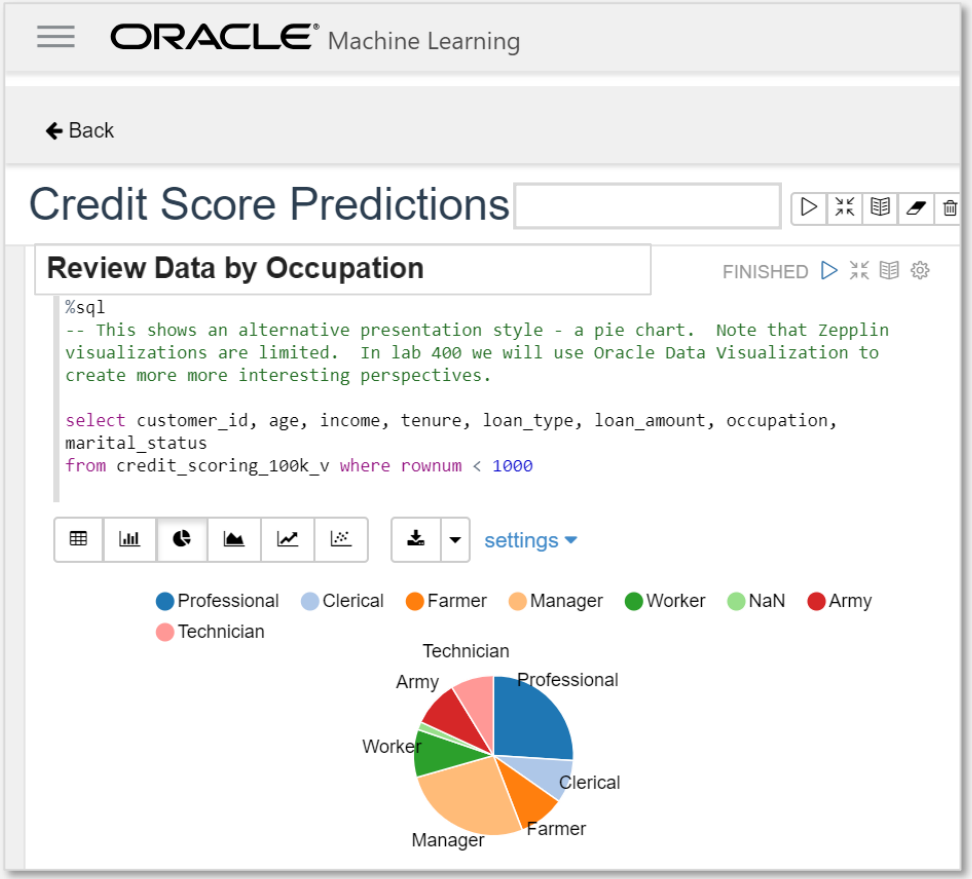
Autonomous Database as a Data Science Platform

Collaborative UI

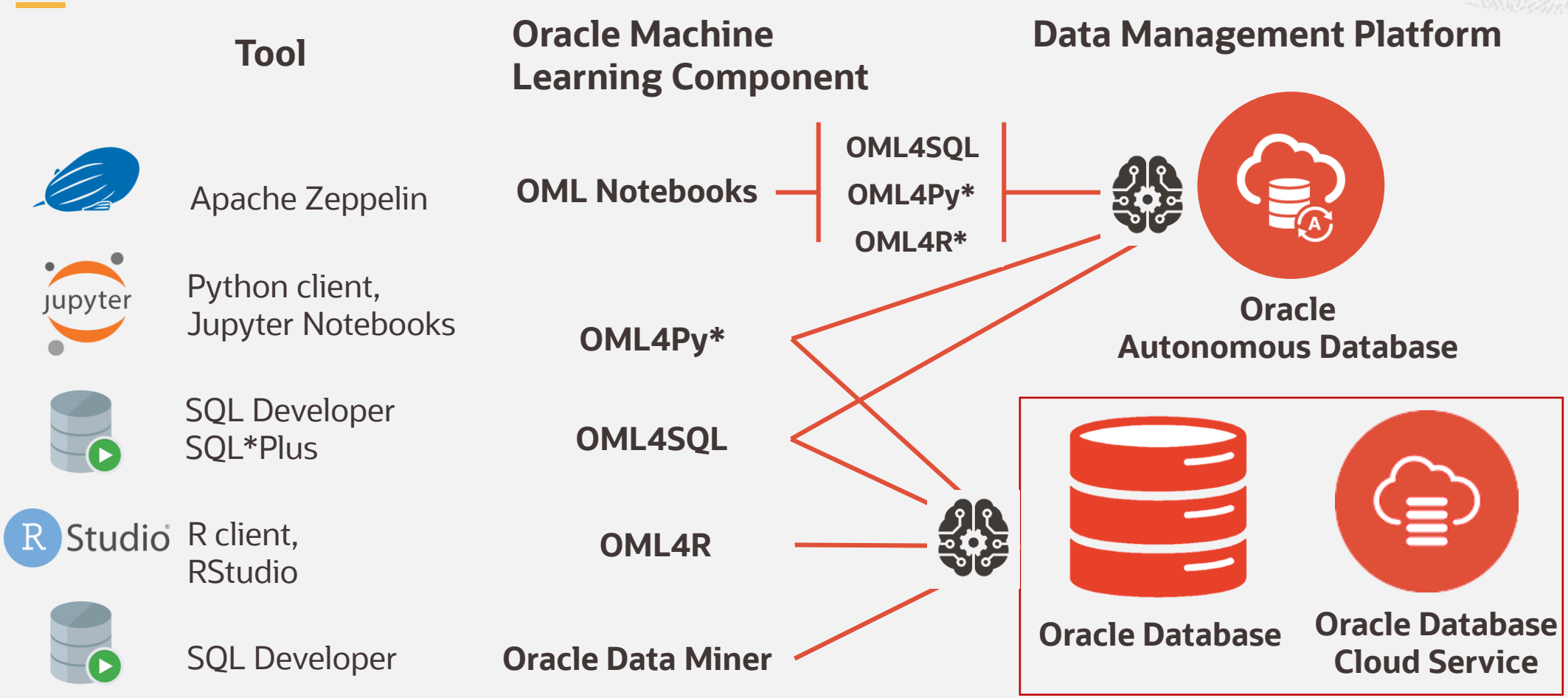
- Based on Apache Zeppelin
- Supports data scientists, data analysts, application developers, DBAs with SQL and Python*
- Easy sharing of notebooks and templates
- Permissions, versioning, and execution scheduling

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



Oracle Machine Learning interfaces to Oracle Database



*Coming soon

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

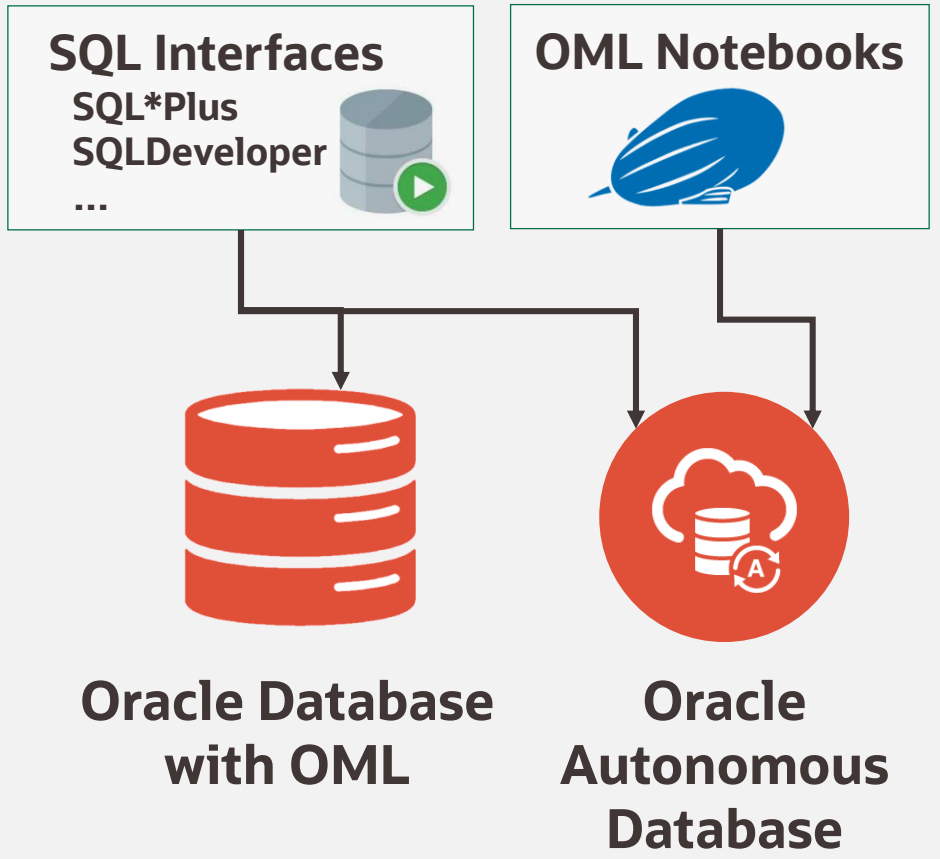
- No extracting data to separate ML engine
- Fast and scalable
- Batch and real-time scoring
- Explanatory prediction details

ML models as first class database objects

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

Supports R and Python interfaces

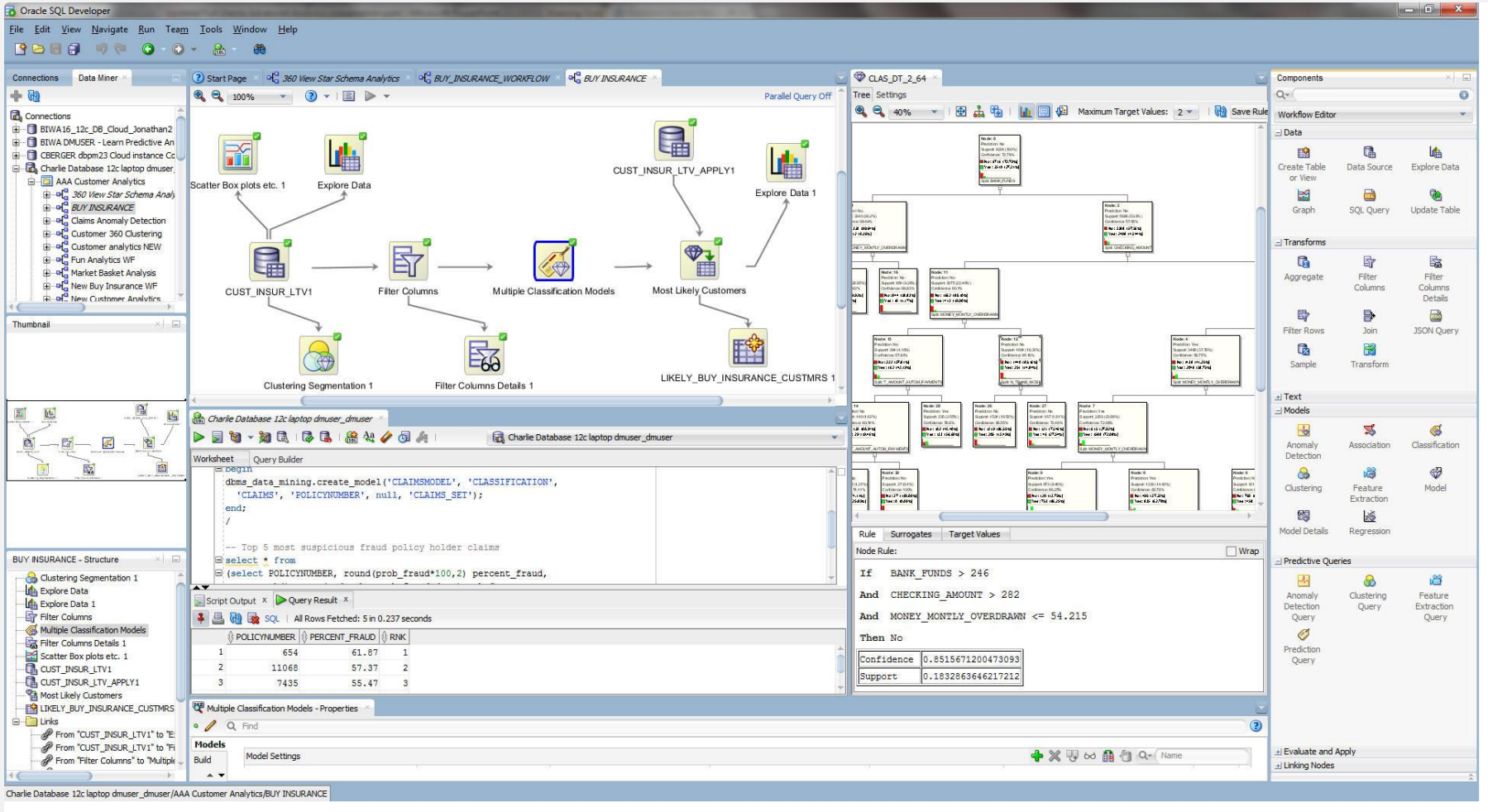
Leverage ML across Oracle stack



Oracle Data Miner User Interface

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

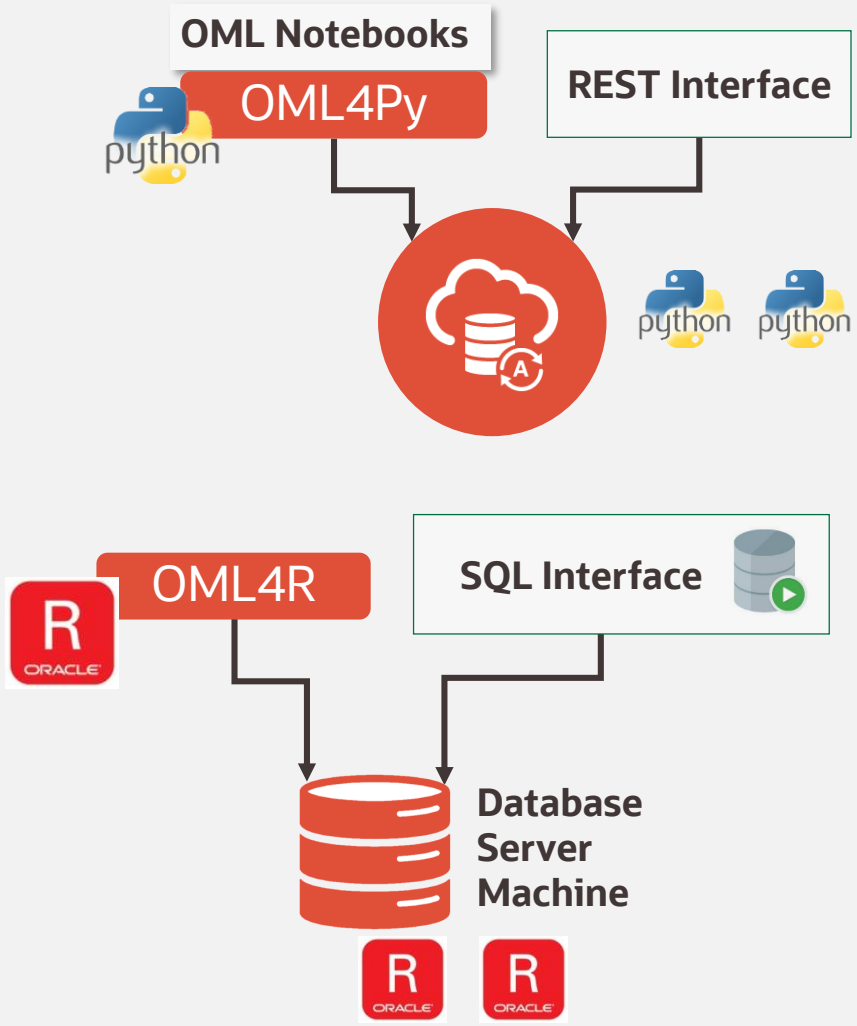
- SQL Developer Extension for Oracle Database on premises and DBCS
- Automates typical data science steps
- Easy to use drag-and-drop interface
- Analytical workflows quickly defined and shared
- Wide range of algorithms and data transformations
- Generate SQL code for immediate deployment



Oracle Machine Learning for R and Python

Empower data scientists with open source environments

- 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 automatic machine learning



Oracle Machine Learning for R and Python

Empower data scientists with open source environments

Transparency layer

- Leverage proxy objects so data remain in database
- Overload native functions translating functionality to SQL
- Use familiar R / Python syntax on database data

Parallel, distributed algorithms

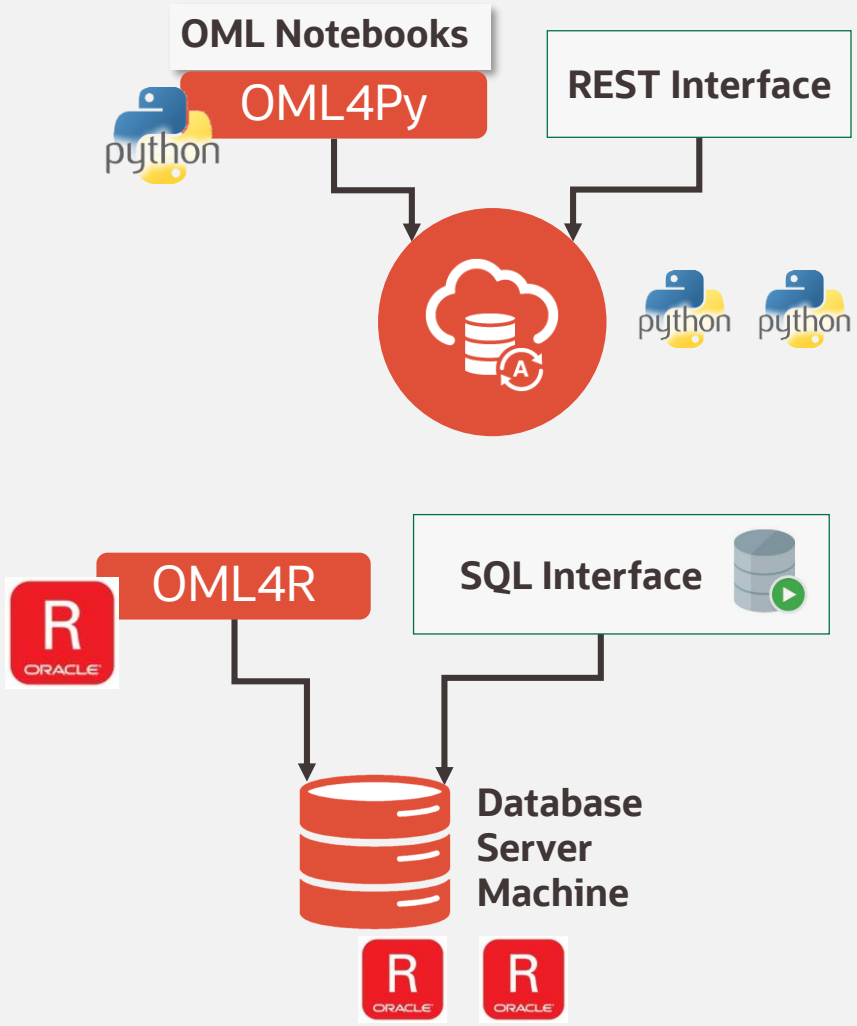
- Scalability and performance
- Exposes in-database algorithms available from OML4SQL

Embedded execution

- Manage and invoke R or Python scripts in Oracle Database
- Data-parallel, task-parallel, and non-parallel execution
- Use open source packages to augment functionality

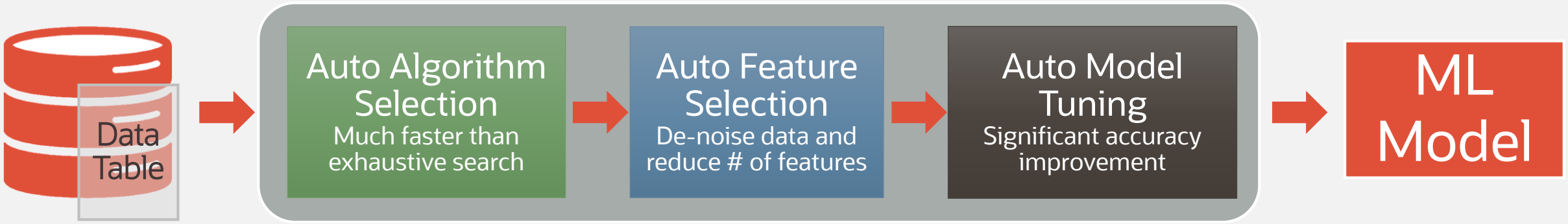
OML4Py AutoML

- Algorithm selection, feature selection, model tuning



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

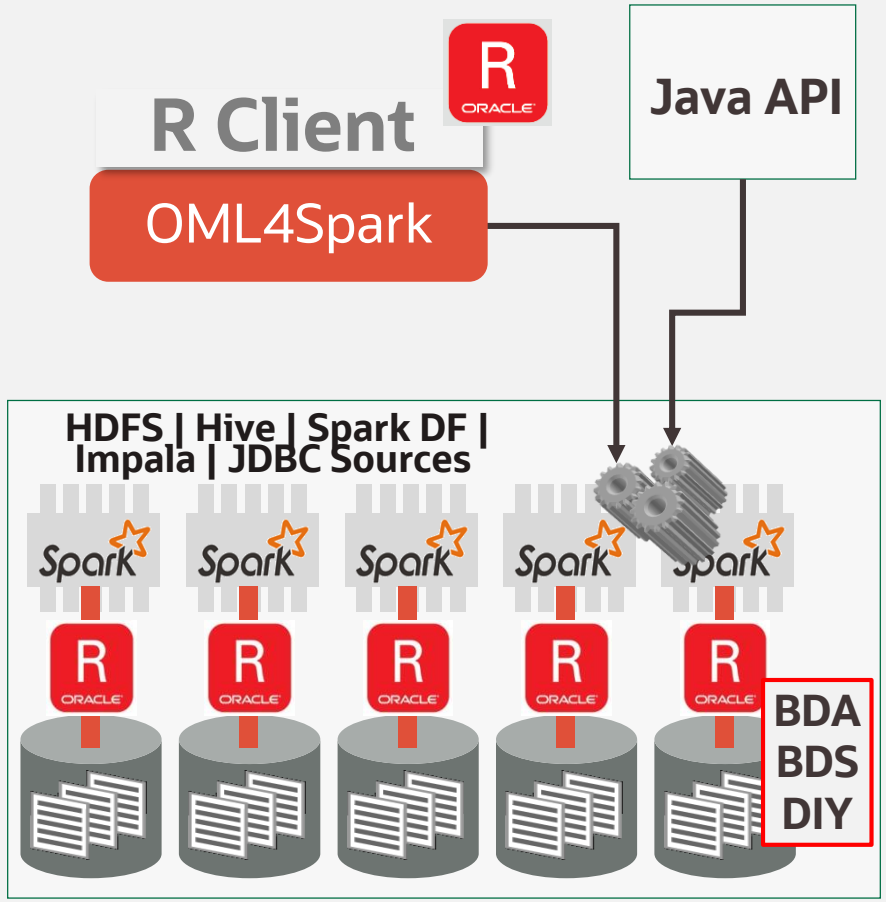
Enables non-expert users to leverage Machine Learning

Demo

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



Coming soon...

Roadmap: OML Services

REST API – deploy models outside the database

Model Management and Deployment Services

- Deploy models in OML format and ONNX format
- Import ONNX for Tensorflow, PyTorch, MXNet, scikitlearn, etc.
- Store, version, compare ML models

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}

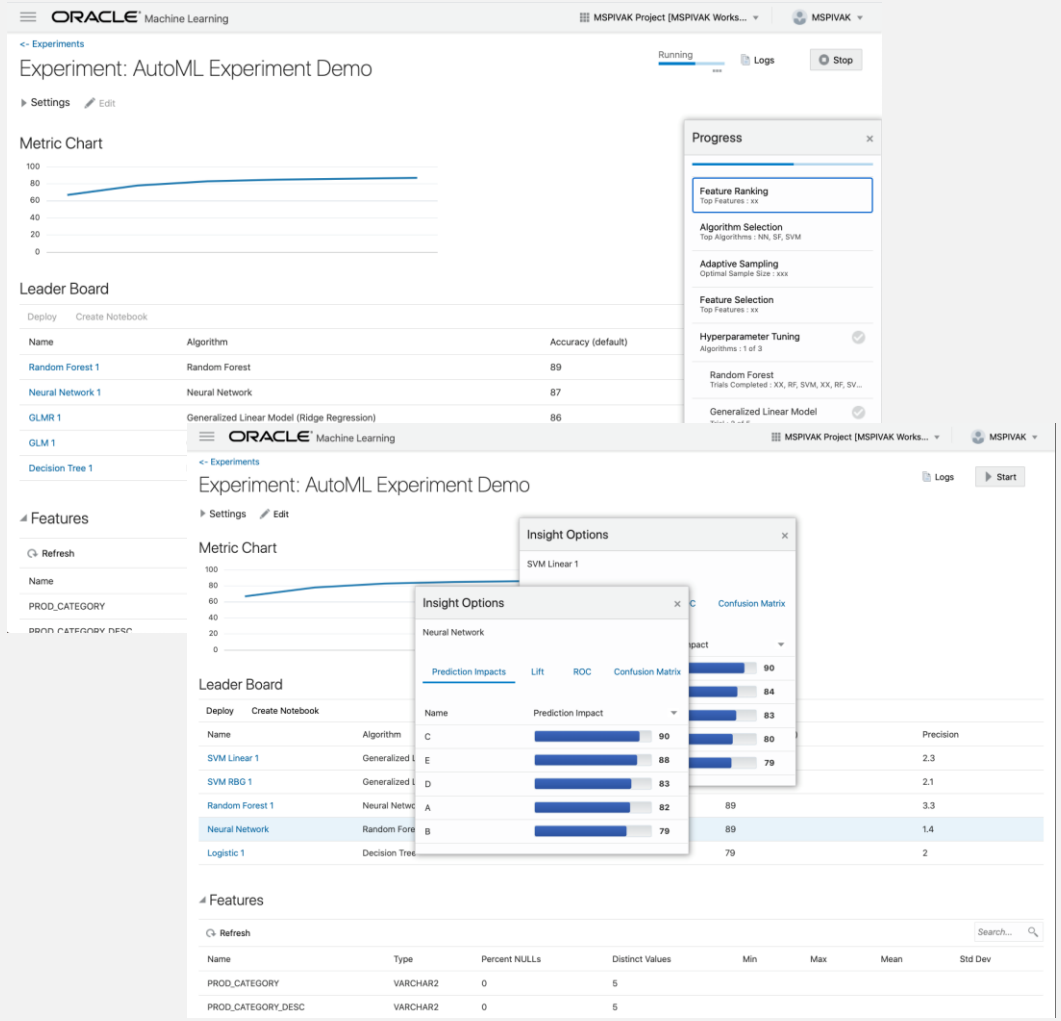
Cognitive Text

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

Roadmap: OML AutoML UI

No-code AutoML-based user interface supporting automatic machine learning

- Powerful, easy to use UI
  - Enable non-expert users to use ML
- Automates model building, tuning, and deployment
  - Enhance data scientist productivity
  - Support model management
  - Empower non-expert users
- Featuring
  - Minimal user input: data, target
  - Model leaderboard
  - Model deployment via REST endpoints



CY2021…

Roadmap: OML4R and OML4Py

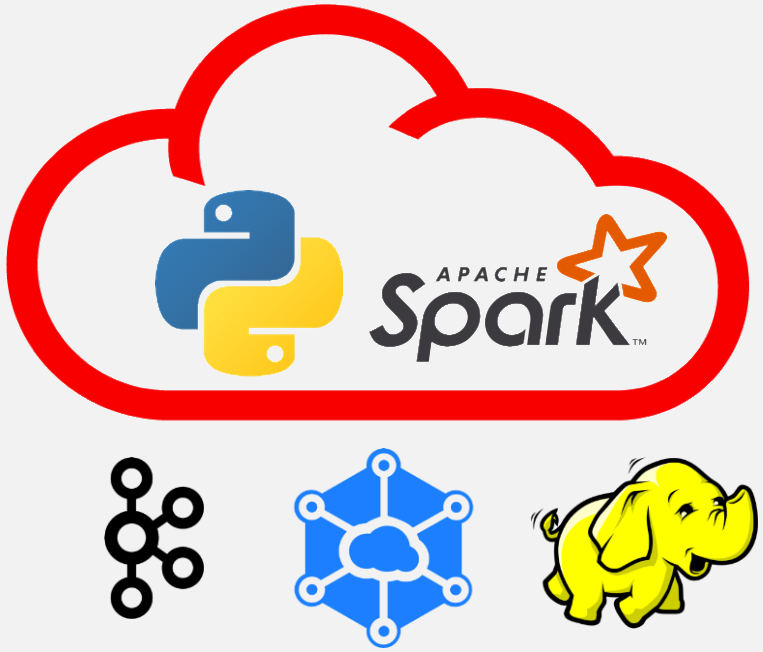
Expand support for open source languages and ecosystems

- 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



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!

Oracle Database and Oracle Autonomous Database

For more information…

oracle.com/machine-learning

See also AskTOM OML Office Hours

Thank You

Mark Hornick
Oracle Machine Learning Product Management