Oracle Machine Learning Platform Overview

Move the algorithms, not the data!

Mark Hornick, Senior Director, Product Management
Data Science and Machine Learning

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.

Agenda

- Oracle Machine Learning family of products
- Supporting multiple personas
- OML component details
- Enabling Oracle Applications use of ML

ORACLE APPLICATIONS

Oracle Machine Learning

OML4SQL
Oracle Advanced Analytics SQL API

OML Notebooks
with Apache Zeppelin on Autonomous Database

OML4R
Oracle R Enterprise R API

Oracle Data Miner
Oracle SQL Developer extension

OML4Py*
Python API

OML4Spark
Oracle R Advanced Analytics for Hadoop

OML Microservices**
Supporting Oracle Applications
Image, Text, Scoring, Deployment, Model Management

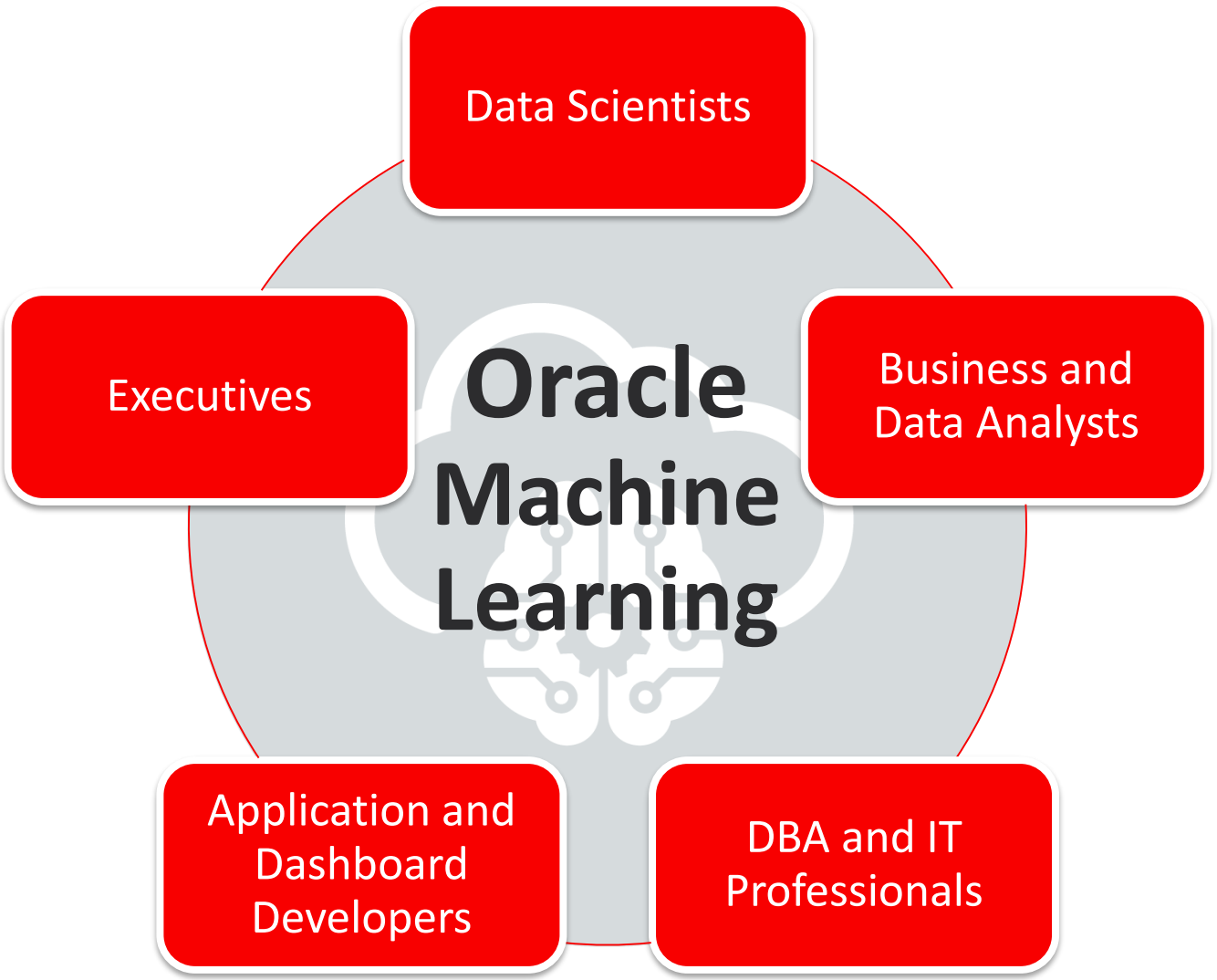
ORACLE FUSION APPLICATIONS

* Coming soon ** For use through Oracle Application only

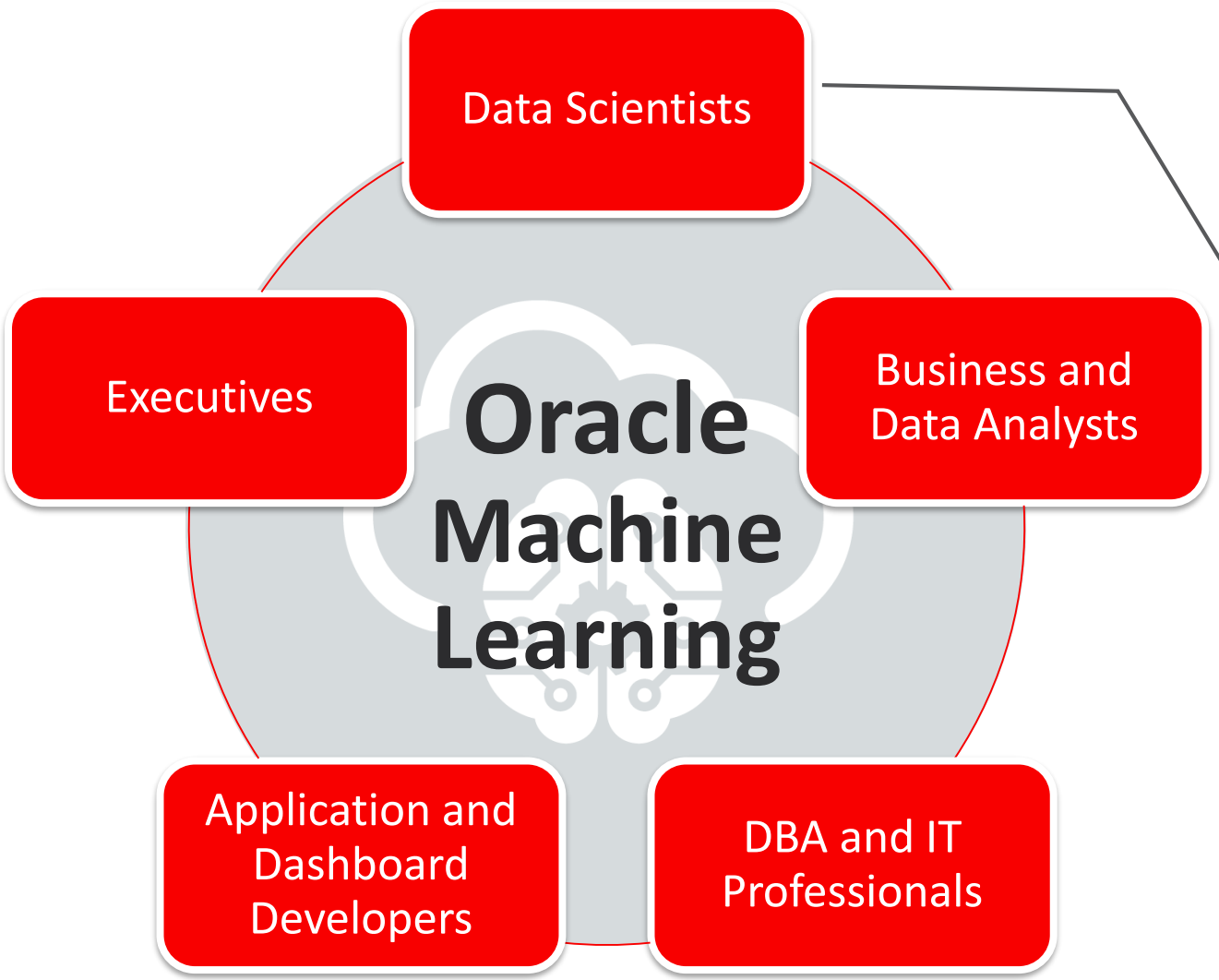
New Branding	Corresponding Products and Components
Oracle Machine Learning (OML)	Oracle Advanced Analytics (OAA) and Oracle R Advanced Analytics for Hadoop (ORAAH)
Oracle Machine Learning for SQL (OML4SQL)	Oracle Advanced Analytics (OAA) / Oracle Data Mining (ODM)
Oracle Machine Learning for R (OML4R)	Oracle R Enterprise (ORE)
Oracle Machine Learning for Python (OML4Py)	Oracle Machine Learning for Python (OML4Py)
Oracle Machine Learning for Spark (OML4Spark)	Oracle R Advanced Analytics for Hadoop (ORAAH)
Oracle Machine Learning Notebooks	*previously known as* Oracle Machine Learning
	Oracle Data Miner (ODMr)

Note: The official price list product names have not changed.

OML empowers Enterprise Users

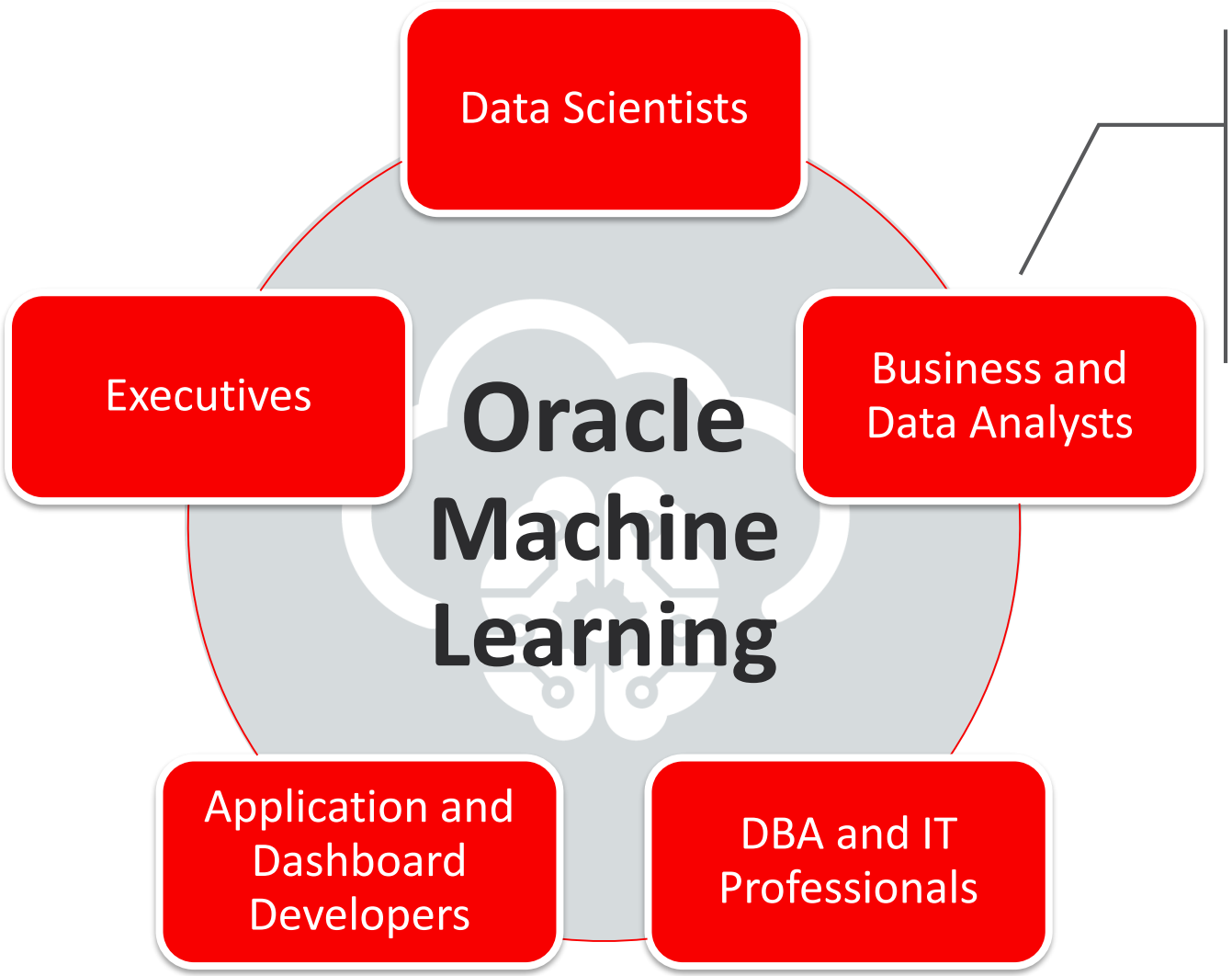


Data Scientists



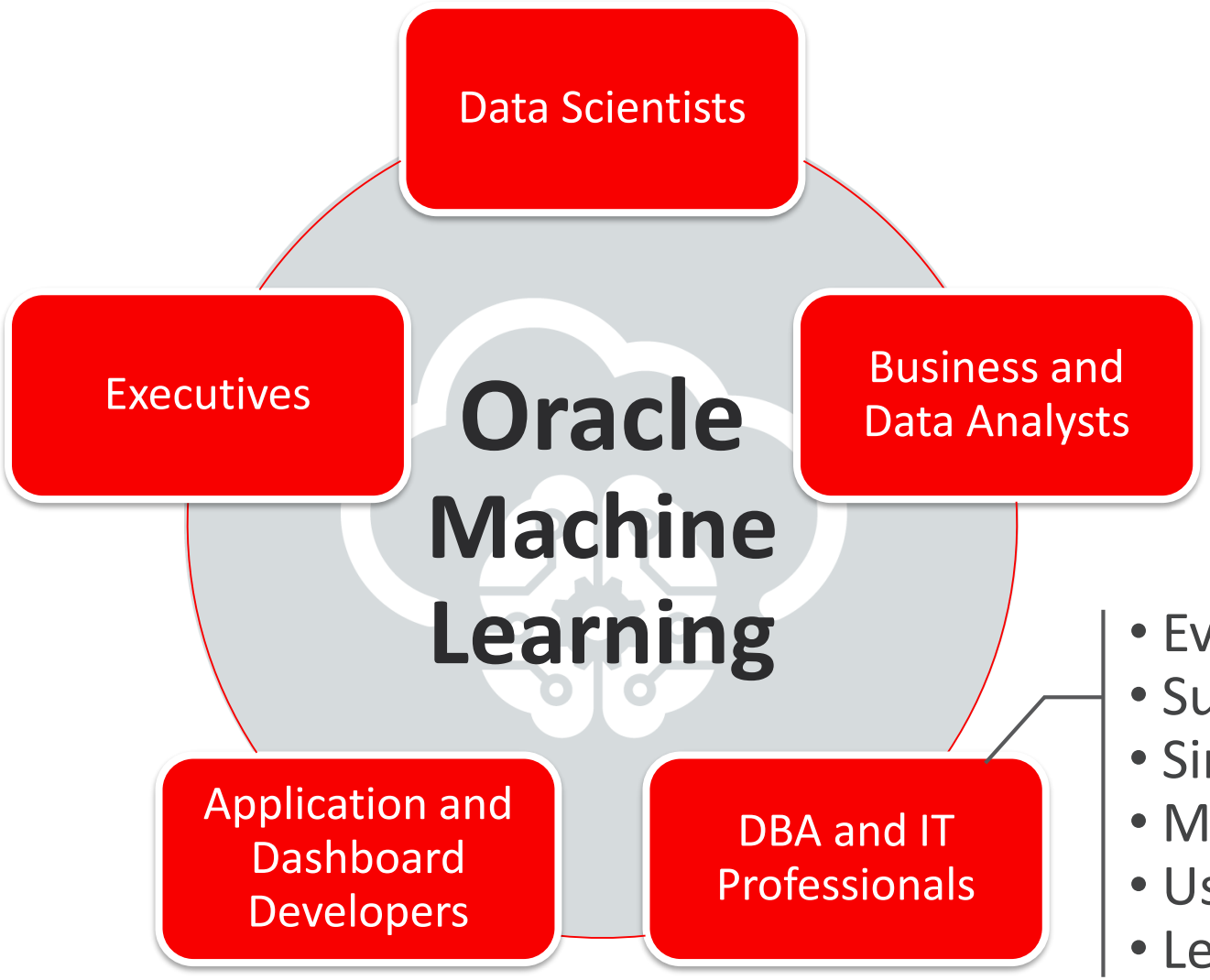
- Popular data science languages: R, Python, SQL
- Augment with 3rd party packages
- Scalability and performance
- Automation-enhanced productivity
- Greater enterprise collaboration
- Integrate and analyze data across the enterprise

Business and Data Analysts



- Expand analytical tool set with ML
- Enable non-ML experts with AutoML
- Leverage domain knowledge for better results
- Collaborate with Data Scientists and IT

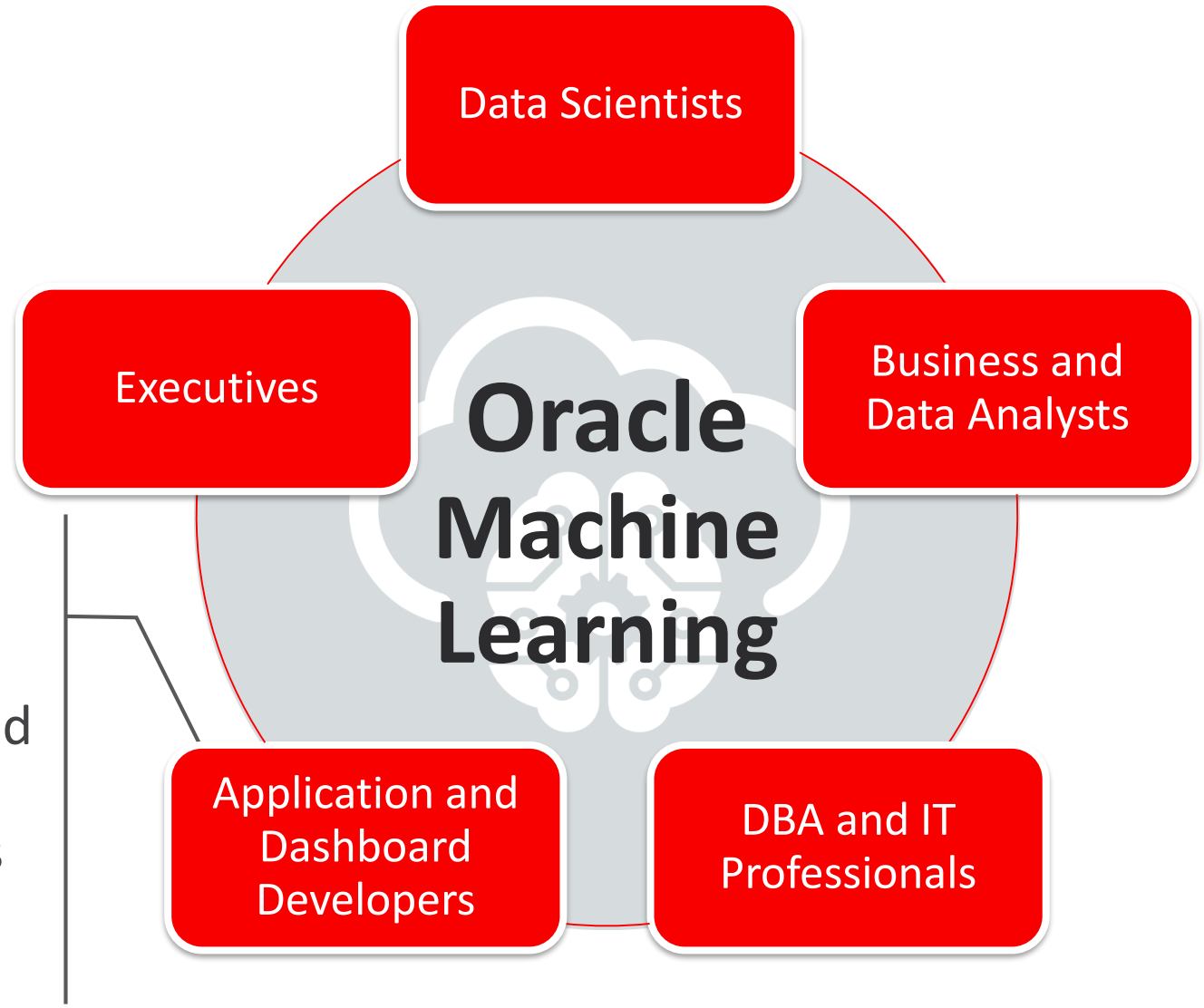
DBA and IT Professionals



- Even greater value from Oracle investment
- Support scalability and performance
- Simpler, streamlined infrastructure
- Maintain data security, backup, recovery
- Use SQL, expand to Python and R
- Leverage Database and Big Data sources

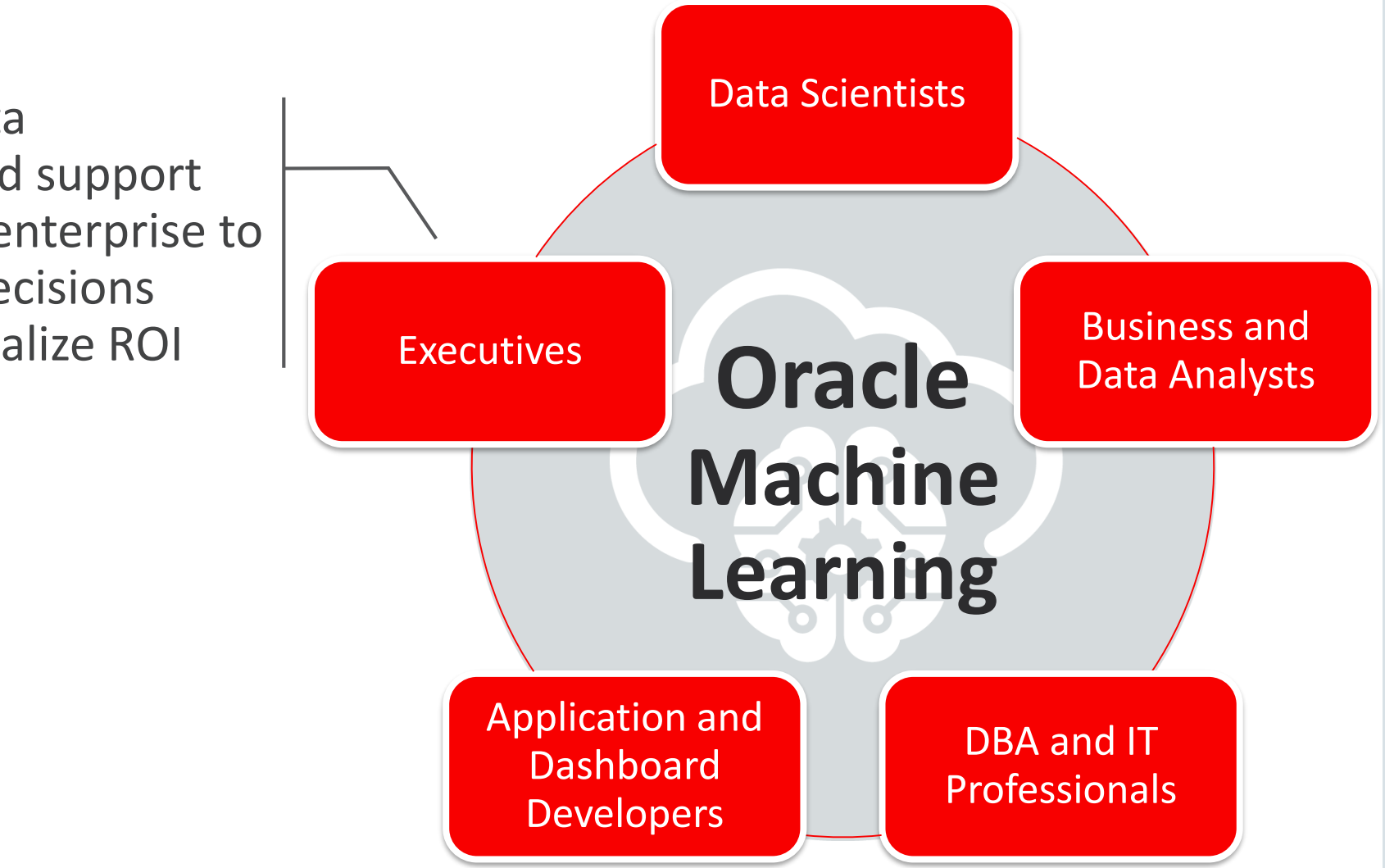
Application and Dashboard Developers

- Realize intelligent solutions faster through Oracle stack integration
- Easily uptake data scientists' R, Python, and SQL scripts and rapidly deploy solutions
- Embed ML in applications and dashboards using SQL, REST, and SODA APIs

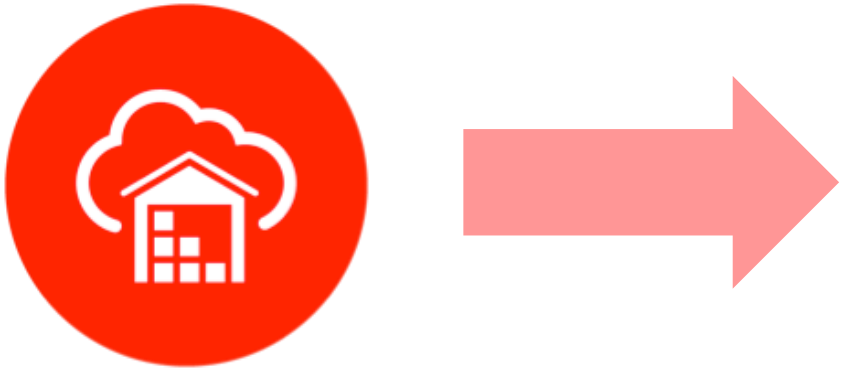


Executives

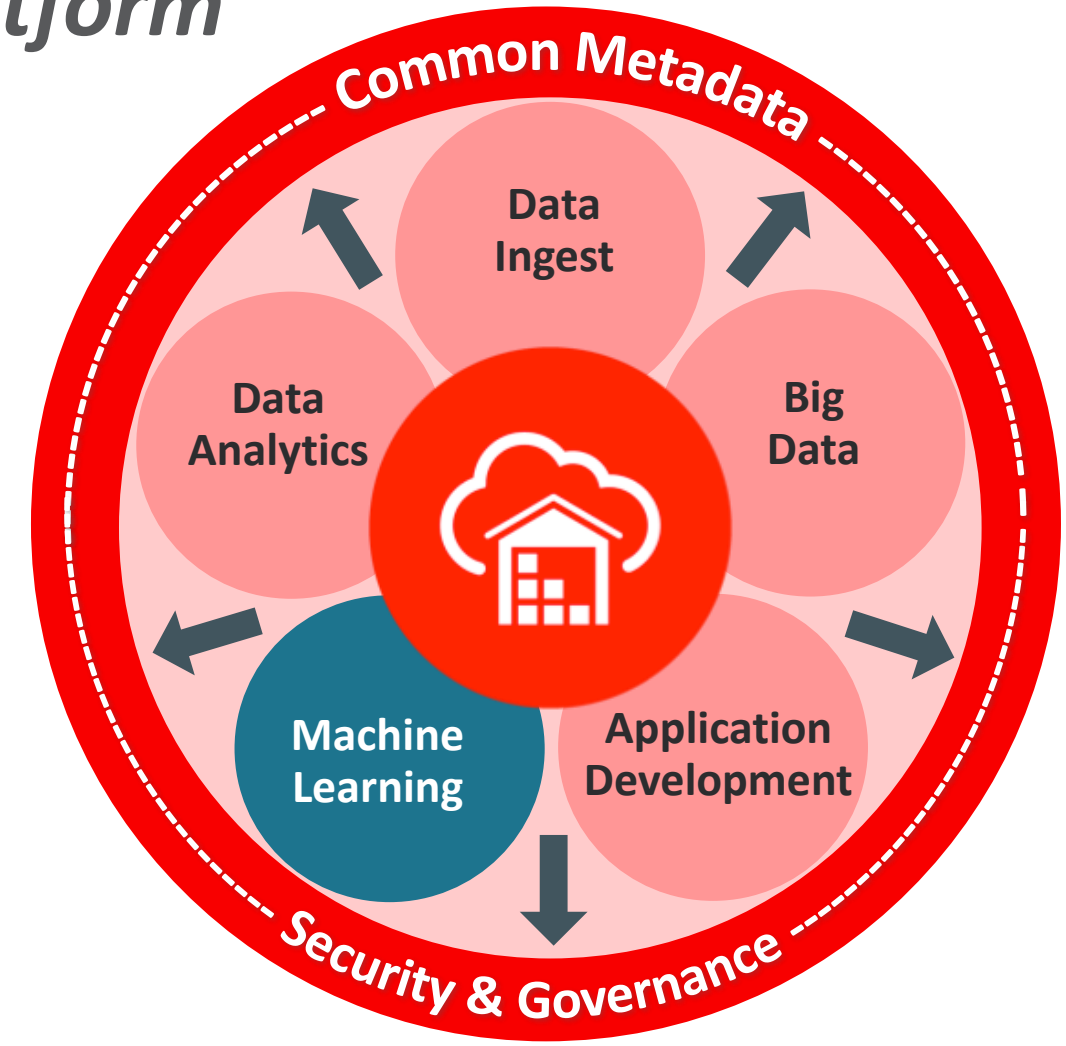
- Benefit from world-class data management technology and support
- Democratize ML across the enterprise to enable better data-driven decisions
- Deploy solutions faster to realize ROI



Vision: Autonomous Database evolves into ***Autonomous Data Platform***

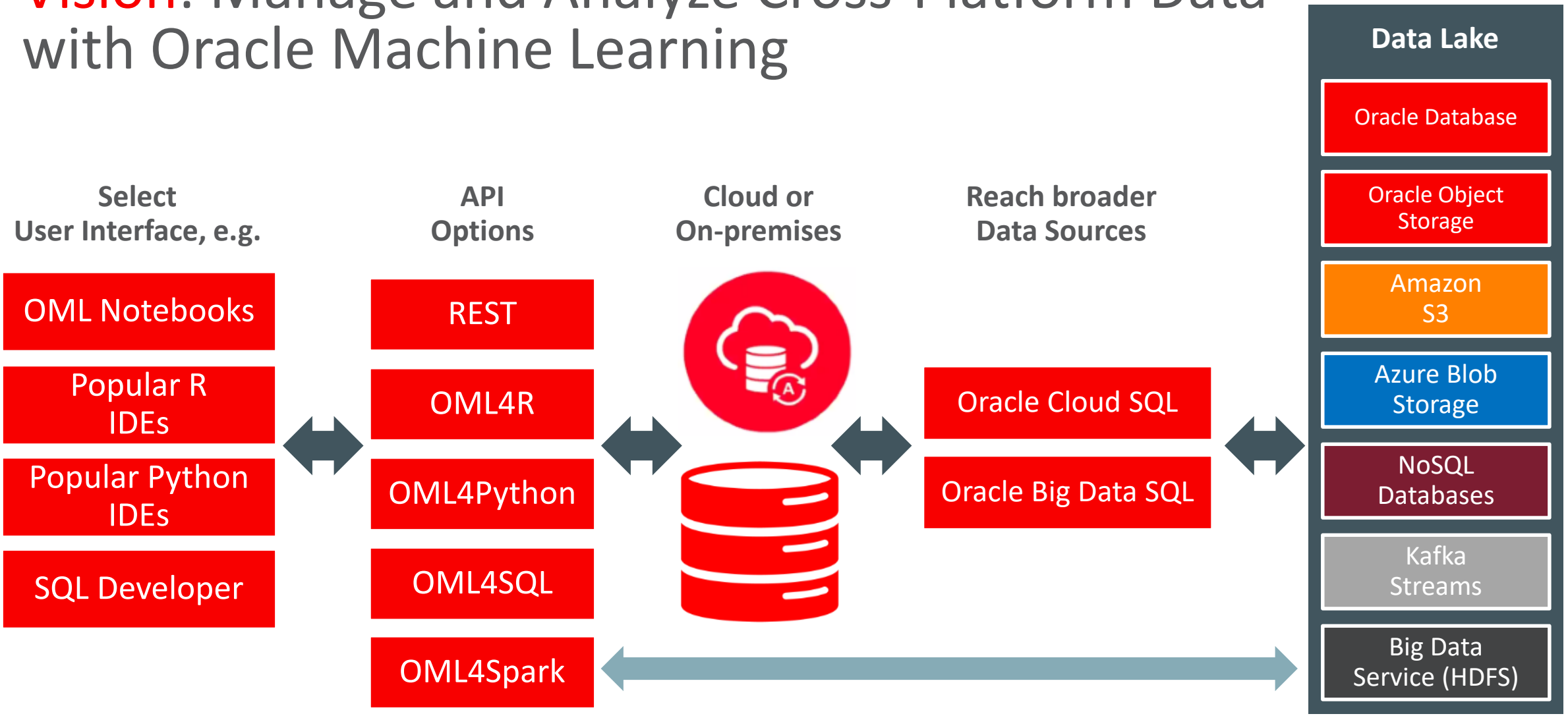


From essentially a database cloud service…



…to a tightly integrated data platform

Vision: Manage and Analyze Cross-Platform Data with Oracle Machine Learning



Oracle Machine Learning Algorithms and Analytics

Classification
- Naïve Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine
- Explicit Semantic Analysis

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

Anomaly Detection
- One-Class SVM

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

Regression
- Linear Model
- Generalized Linear Model
- Support Vector Machine (SVM)
- Stepwise Linear regression
- Neural Network
- LASSO

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)

Text Mining Support
- Algorithms support text columns
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA) for document similarity

Statistical Functions
- Basic statistics: 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

OML algorithms differentiators

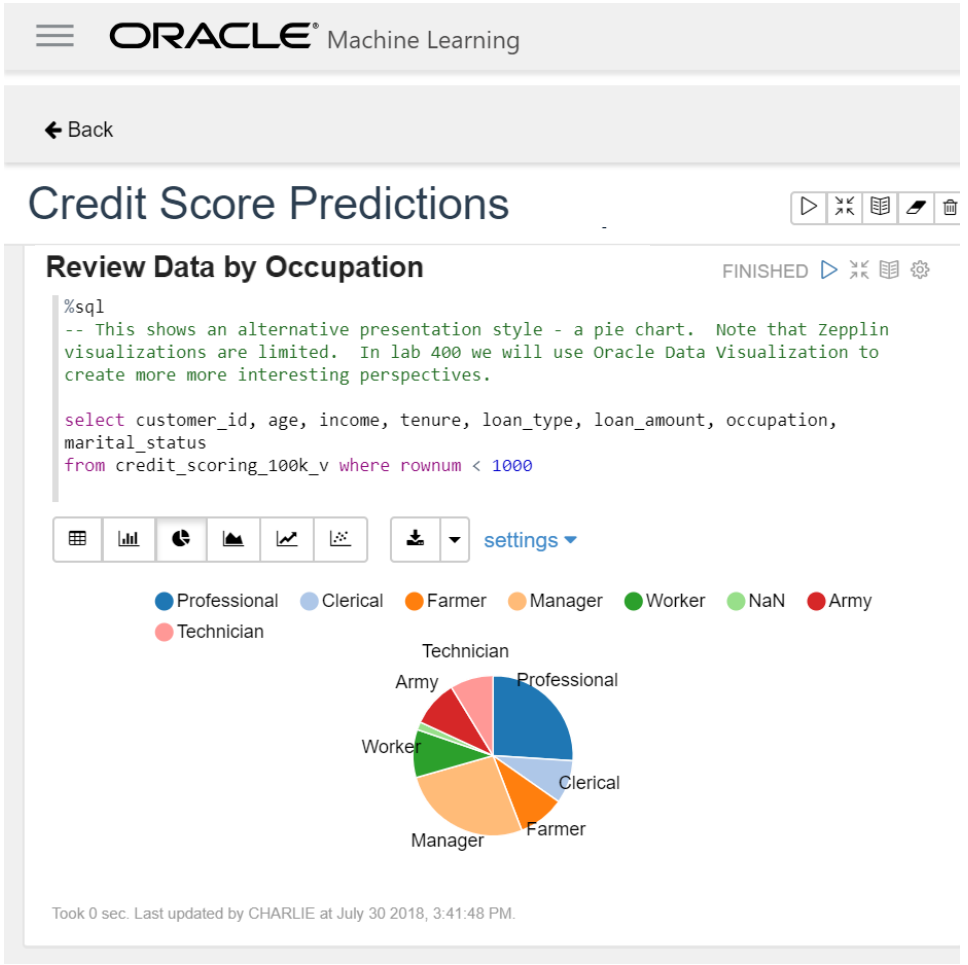
Feature	In-Database	Spark
No data movement to separate analytical engines	✓	✓
Wide range of ML techniques supported	✓ Native	✓ Native and Spark MLlib
High performance from parallel, distributed execution	✓	✓ Spark 2-based. Use all nodes from Hadoop cluster
Greater scalability from improved memory utilization	✓	✓
Handle narrow, wide, and sparse data	✓ Plus star schema, nested data	✓
Automation	✓ Data preparation, text mining, partitioned models, AutoML	
Multiple language interfaces	✓ SQL, R, Python	✓ R, Java
High performance R Formula for R Language	✓	✓

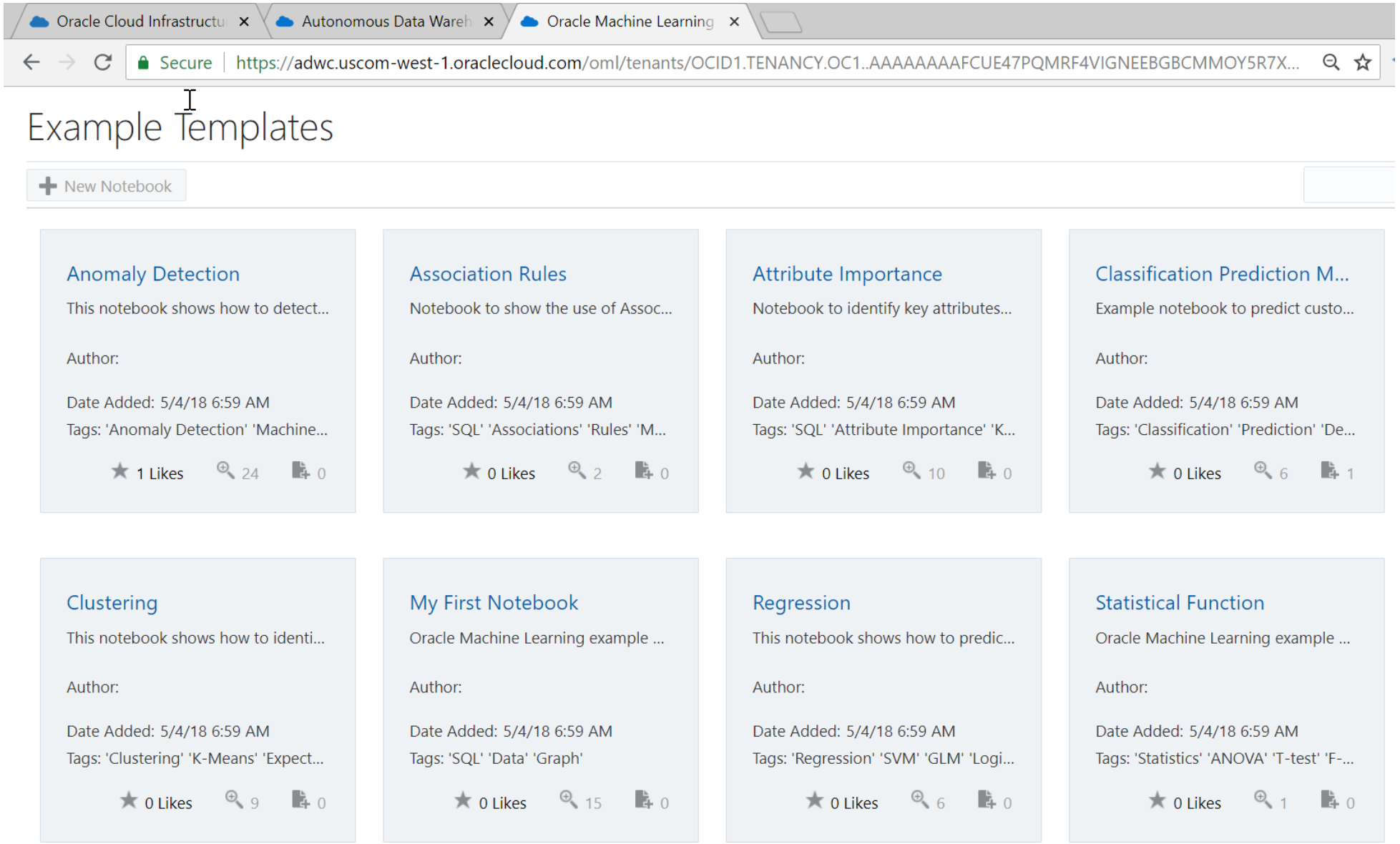
Oracle Machine Learning Notebooks
for Autonomous Database

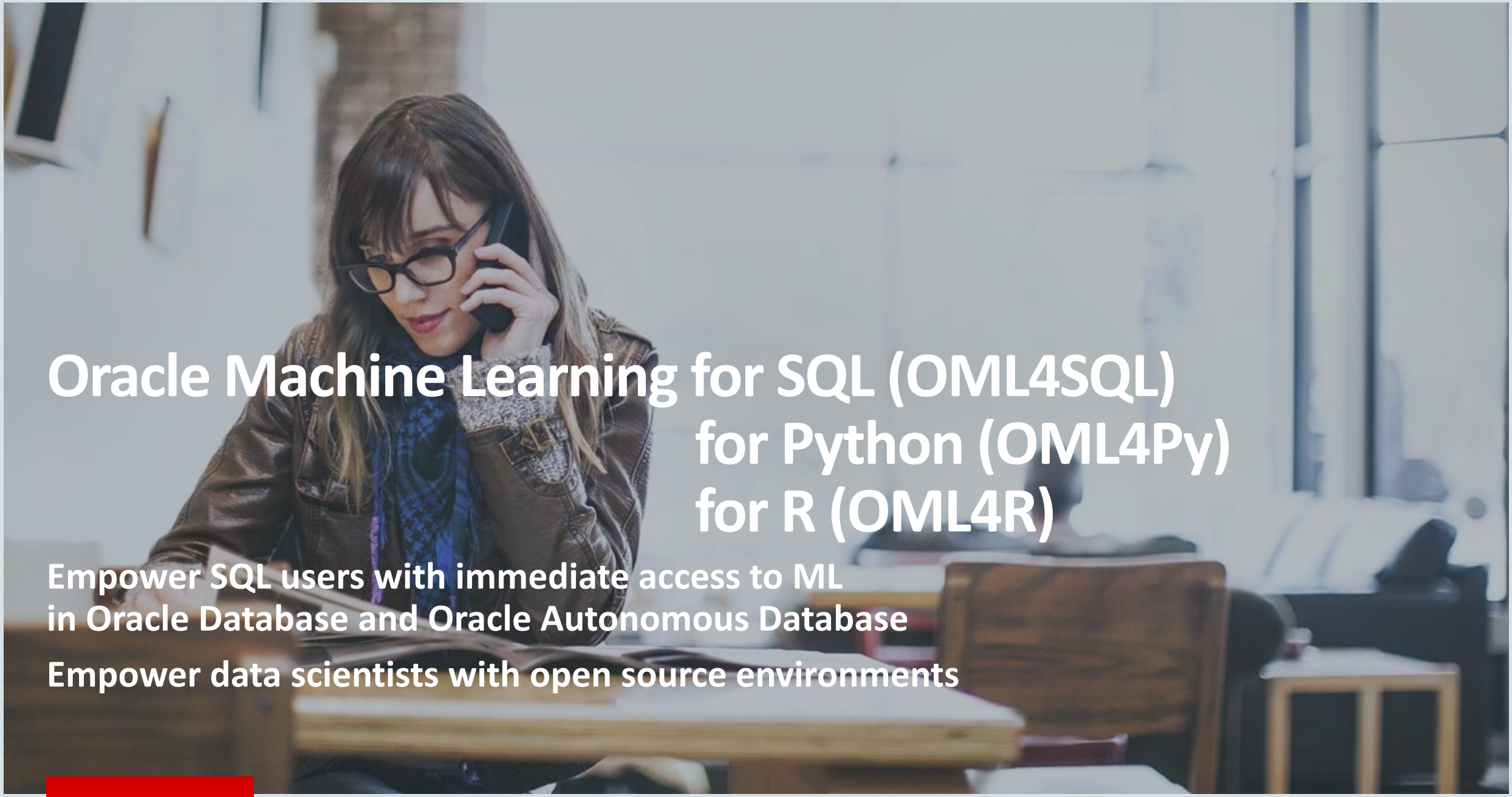
Oracle Machine Learning Notebooks

Autonomous Database as a Data Science Platform

- Collaborative UI
  - Based on Apache Zeppelin
  - Supports data scientists, data analysts, application developers, DBAs
  - Easy sharing of notebooks and templates with permissions, versioning, and execution scheduling
- Included with Autonomous Database
  - Automatically provisioned, managed, backed up
  - In-database machine learning algorithms and analytics functions via OML4SQL, and soon to be augmented with Python and R







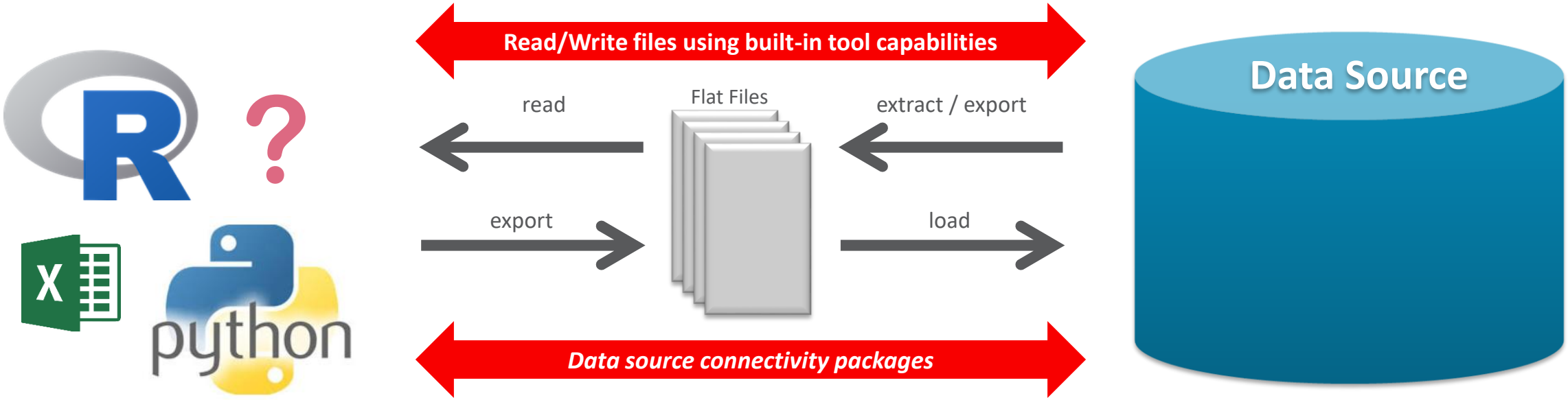
Oracle Machine Learning for SQL (OML4SQL) for Python (OML4Py) for R (OML4R)

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

Empower data scientists with open source environments

ORACLE

Traditional Analytics and Data Source Interaction



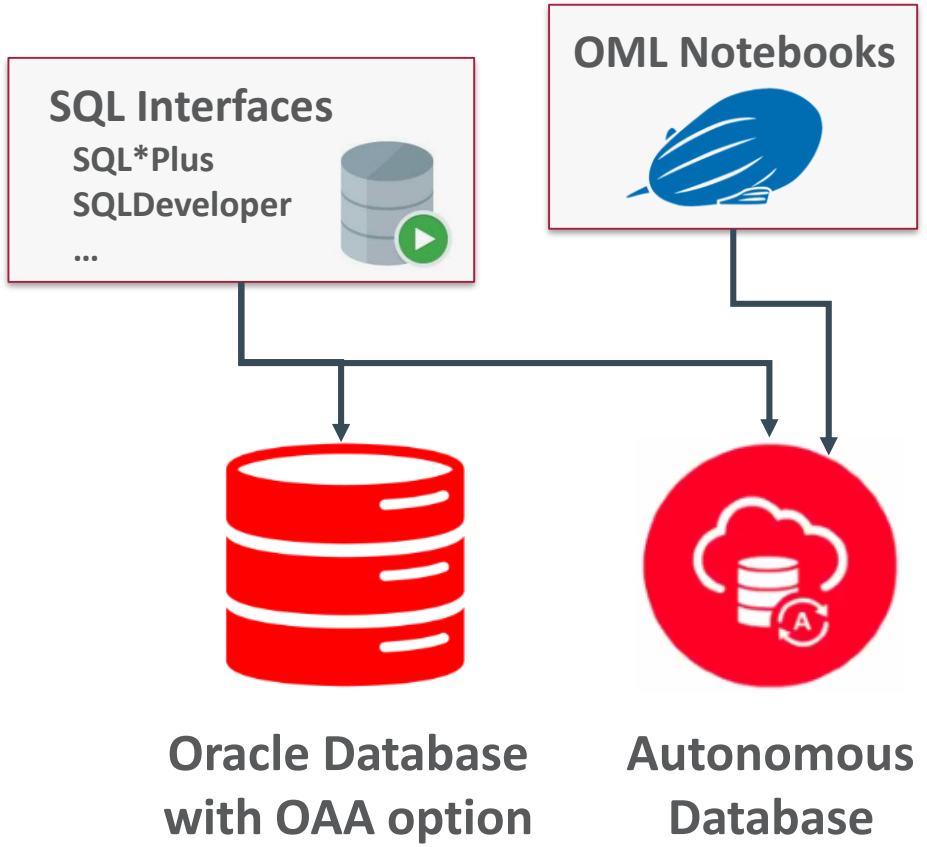
Deployment
Ad hoc
cron job

- Access latency
- Paradigm shift: e.g., R/Python → *Data Access Language* → R/Python
- Memory limitation – data size, in-memory processing
- Single threaded
- Issues for backup, recovery, security
- Ad hoc production deployment

Oracle Machine Learning for SQL

Component of ADB and Oracle Advanced Analytics option to Oracle Database

- Use in-database parallel and distributed machine learning algorithms from SQL and PL/SQL
- ML models as first class database objects
- Export / import models across databases
- Batch and real-time scoring with explanatory predictive details
- Leverage machine learning across SQL-enabled Oracle stack



OML4SQL: Model Build and Real-time Prediction

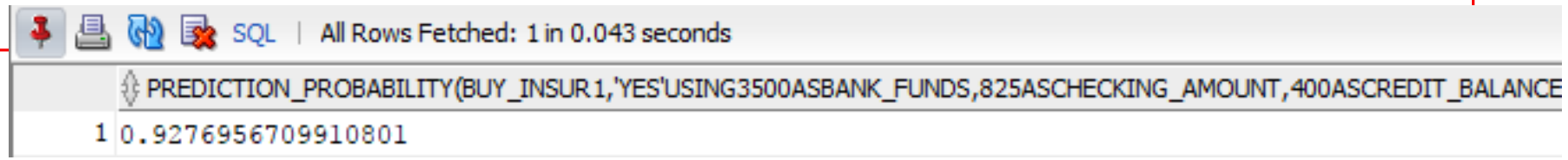
Simple SQL Syntax—Classification Model

Model build (PL/SQL)

```
BEGIN
   DBMS_DATA_MINING.CREATE_MODEL(
     model_name           => 'BUY_INSUR1',
     mining_function      => dbms_data_mining.classification,
     data_table_name      => 'CUST_INSUR_LTV',
     case_id_column_name  => 'CUST_ID',
     target_column_name   => 'BUY_INSURANCE',
     settings_table_name  => 'CUST_INSUR_LTV_SET');
END;
```

Real-time scoring (SQL query)

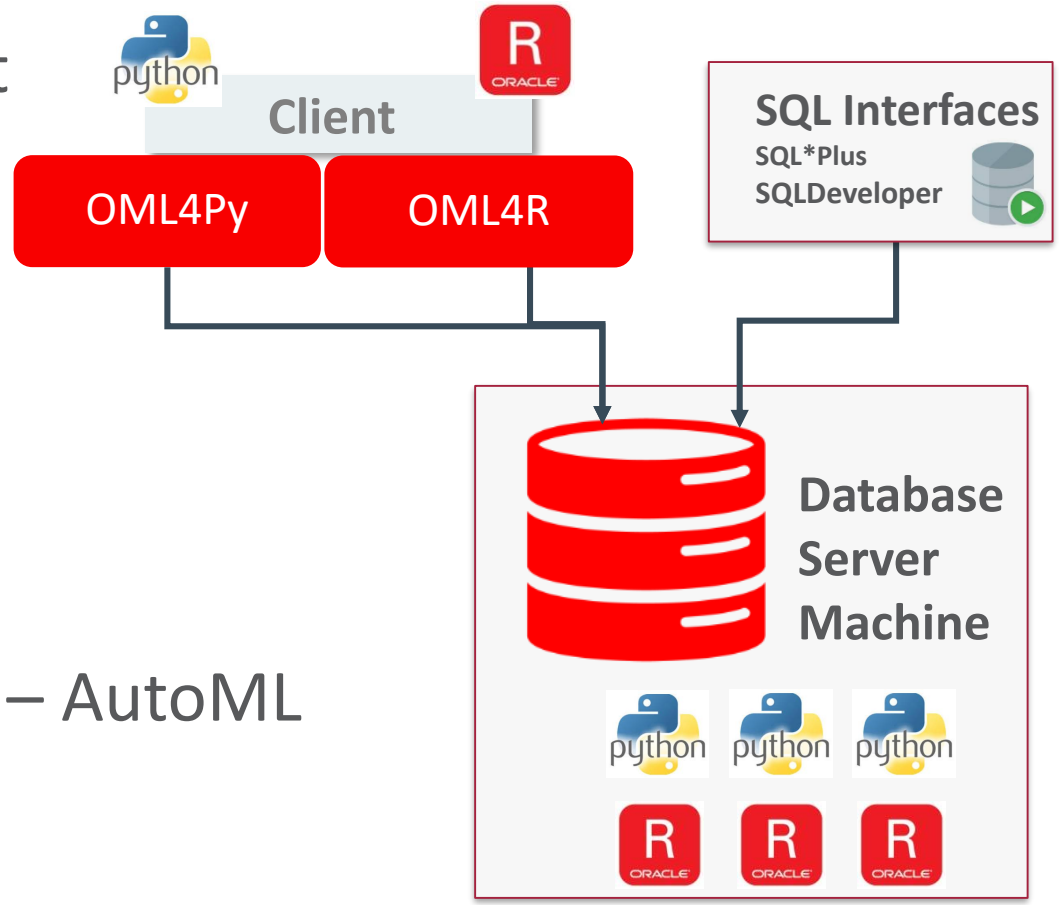
```
SELECT prediction_probability(BUY_INSUR1, 'Yes'
  USING 3500 as bank_funds, 825 as checking_amount, 400 as credit_balance, 22 as age,
  'Married' as marital_status, 93 as MONEY_MONTLY_OVERDRAWN, 1 as house_ownership)
FROM dual;
```



Oracle Machine Learning for R and Python

Components of Oracle Advanced Analytics option to Oracle Database

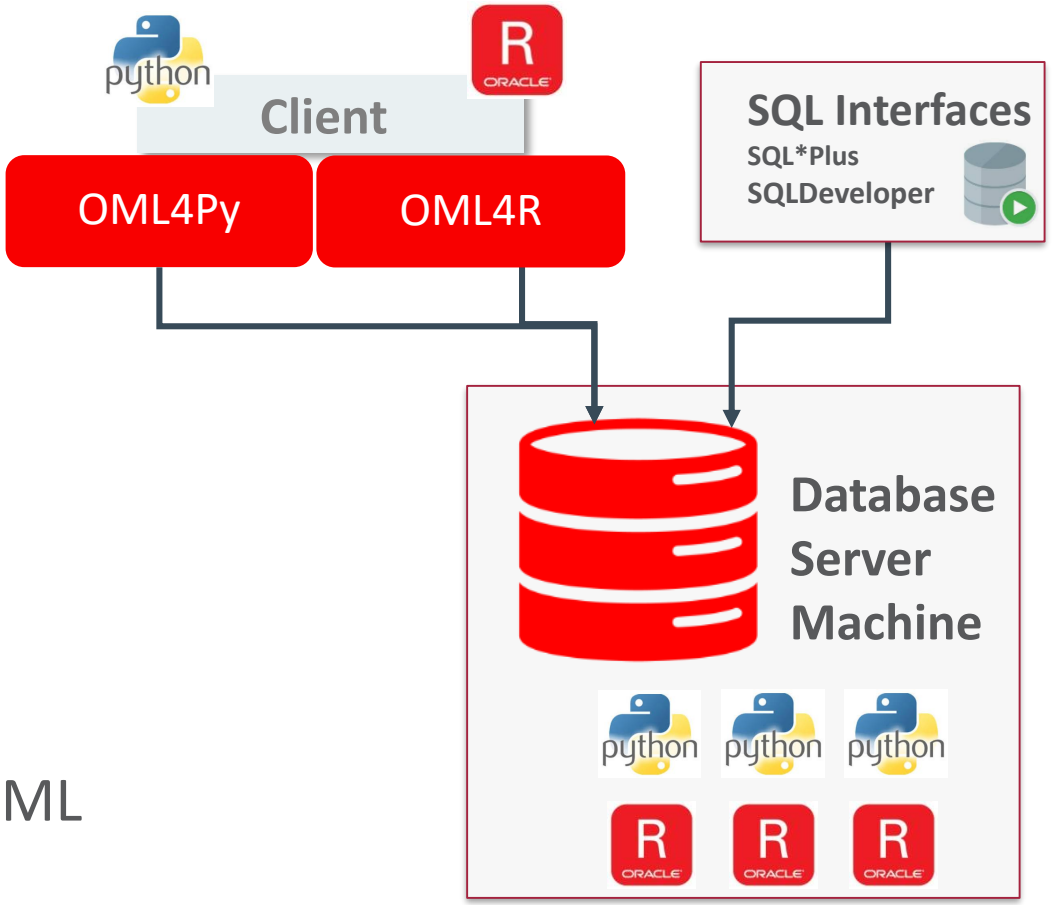
- Use Oracle Database as HPC environment
- Use in-database parallel and distributed machine learning algorithms
- Manage R and Python scripts and objects in Oracle Database
- Integrate open source results into applications and dashboards via SQL
- In OML4Py, automated machine learning – AutoML



Oracle Machine Learning for R and Python

Oracle Advanced Analytics option to Oracle Database

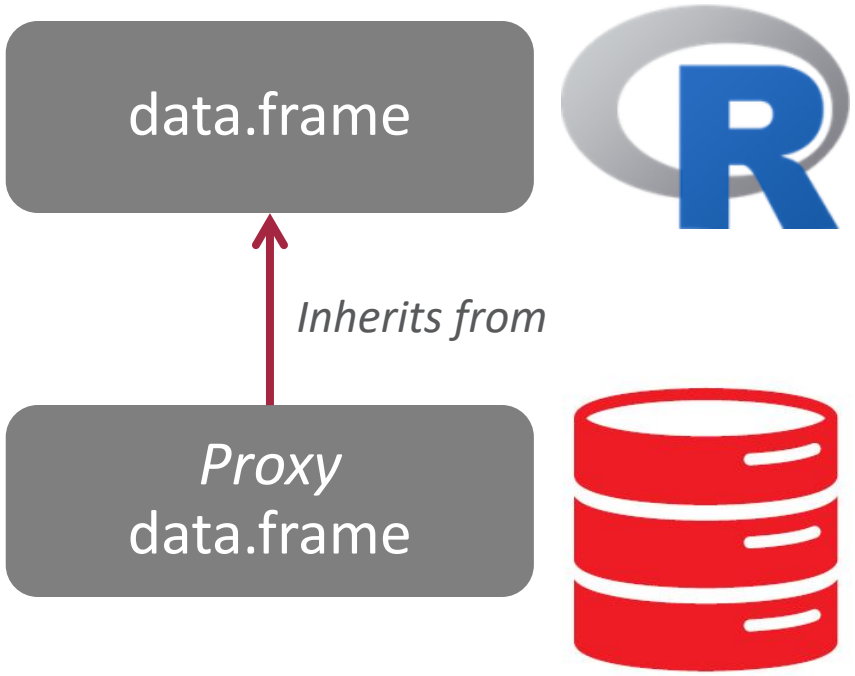
- Transparency layer
  - Leverage proxy objects so data remain in database
  - Overload native functions translating functionality to SQL
  - Use familiar R / Python syntax to manipulate database data
- Parallel, distributed algorithms
  - Scalability and performance
  - Exposes in-database algorithms 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
- In OML4Py, Automated Machine Learning - AutoML
  - Feature selection, model selection, hyper-parameter tuning



Proxy objects for scalability

Example using OML4R interface

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa




```
> str(iris)
'data.frame':   150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
> str(IRIS)
'data.frame':   150 obs. of  5 variables:
Formal class 'ore.frame' [package "OREbase"] with 12 slots
  ..@ .Data    : list()
  ..@ dataQry  : Named chr "( select /*+ no_merge(t) */  \"Sepal.Length\" VAL001,\"Sepal.Wid
th\" VAL002,\"Petal.Length\" VAL003,\"Petal.Width\" VAL004,\"Sp"| __truncated__
  ..- attr(*, "names")= chr "2539 1"
```

```
"( select /*+ no_merge(t) */  \"Sepal.Length\" VAL001,\"Sepal.Width\" VAL
002,\"Petal.Length\" VAL003,\"Petal.Width\" VAL004,\"Species\" VAL005 fro
m \"RQUSER\".\"IRIS\" t  )"
```

```
  ..@ sqlName  : chr
  ..@ sqlValue : chr  "\"Sepal.Length\"" "\"Sepal.Width\"" "\"Petal.Length\"" "\"Petal.Width
\"" ...
  ..@ sqlTable : chr "\"RQUSER\".\"IRIS\""
  ..@ sqlPred  : chr ""
  ..@ extRef   : list()
  ..@ names    : chr
  ..@ row.names: int
  ..@ .S3Class : chr "data.frame"
```

Transparency Layer

In-database performance from indexes, query optimization, parallelism, partitioning

- Leverages proxy objects for database data: *oml.DataFrame*

```
# Create table from Pandas DataFrame data
DATA = oml.create(data, table = 'BOSTON')

# Get proxy object to DB table boston
DATA = oml.sync(table = 'BOSTON')
```

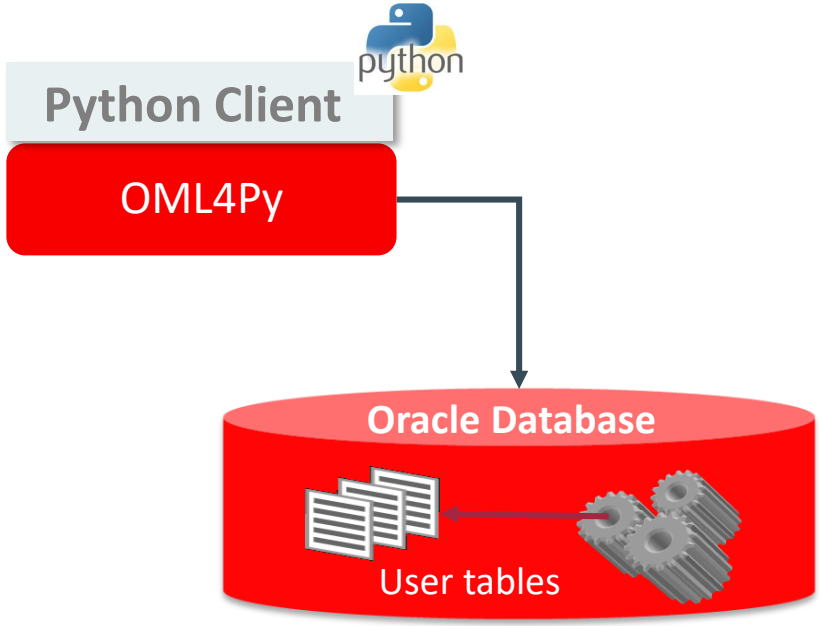
- Uses familiar Python syntax to manipulate database data
- Overloads Python functions translating functionality to SQL

```
DATA.shape

DATA.head()

DATA.describe()

DATA.std()

DATA.skew()


TRAIN, TEST =
   DATA.split()

TRAIN.shape

TEST.shape
```

Parallel, Distributed Algorithms

In-database modeling using Support Vector Machine

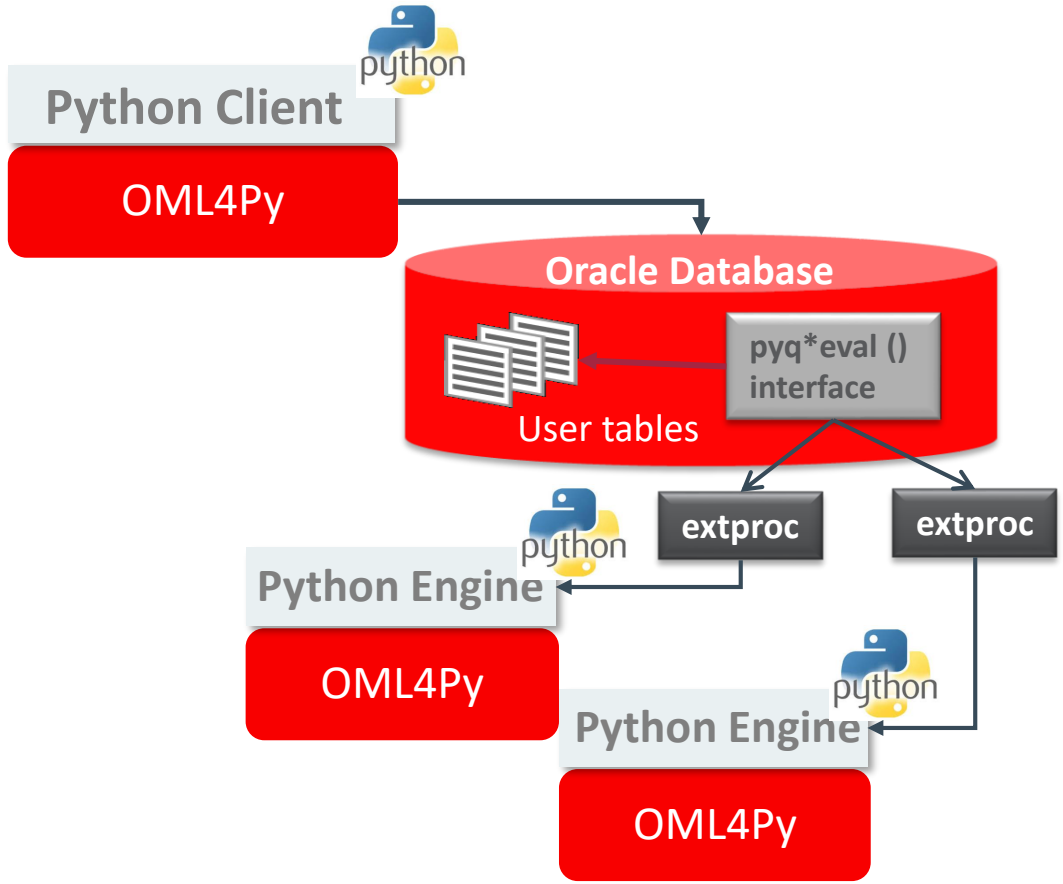
```
from oml import svm

# create proxy object
ONTIME_S = oml.sync(table='ONTIME_S')

# define model object
settings = {'svms_outlier_rate' : 0.01}
svm_mod = svm('anomaly_detection',
              svms_kernel_function =
                 'dbms_data_mining.svms_linear',
              **settings)
# build anomaly detection model
svm_mod = svm_mod.fit(x=ONTIME_S, y=None)
# view model object
svm_mod
```



Embedded Python Execution

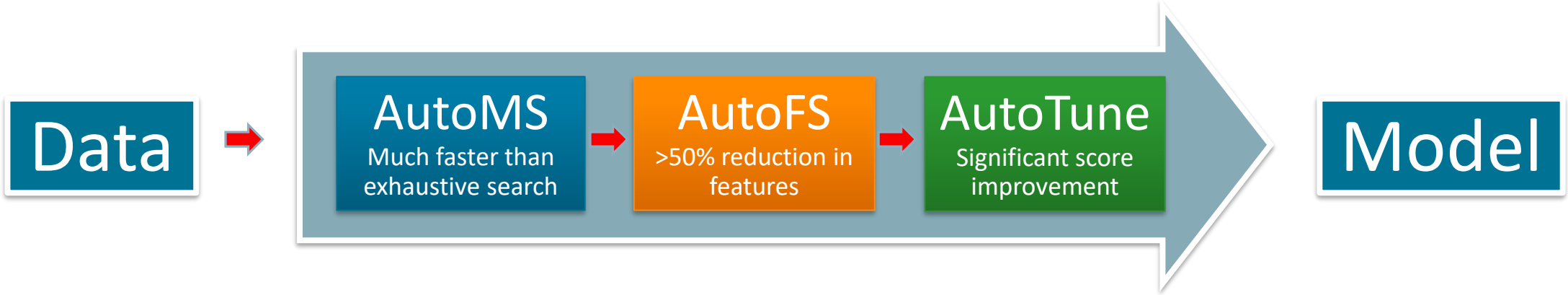
Example of parallel execution for partitioned data flow using third party package

```
# user-defined function using sklearn
def build_lm(dat):
  from sklearn import linear_model
  lm = linear_model.LinearRegression()
  X = dat[['PETAL_WIDTH']]
  y = dat[['PETAL_LENGTH']]
  lm.fit(X, y)
  return lm
# select column(s) for partitioning data
index = oml.DataFrame(IRIS['SPECIES'])
# invoke function in parallel on IRIS table
mods = oml.group_apply(IRIS, index,
                       func=build_lm,
                       parallel=2)
mods.pull().items()
```



AutoML – *new* with OML4Py

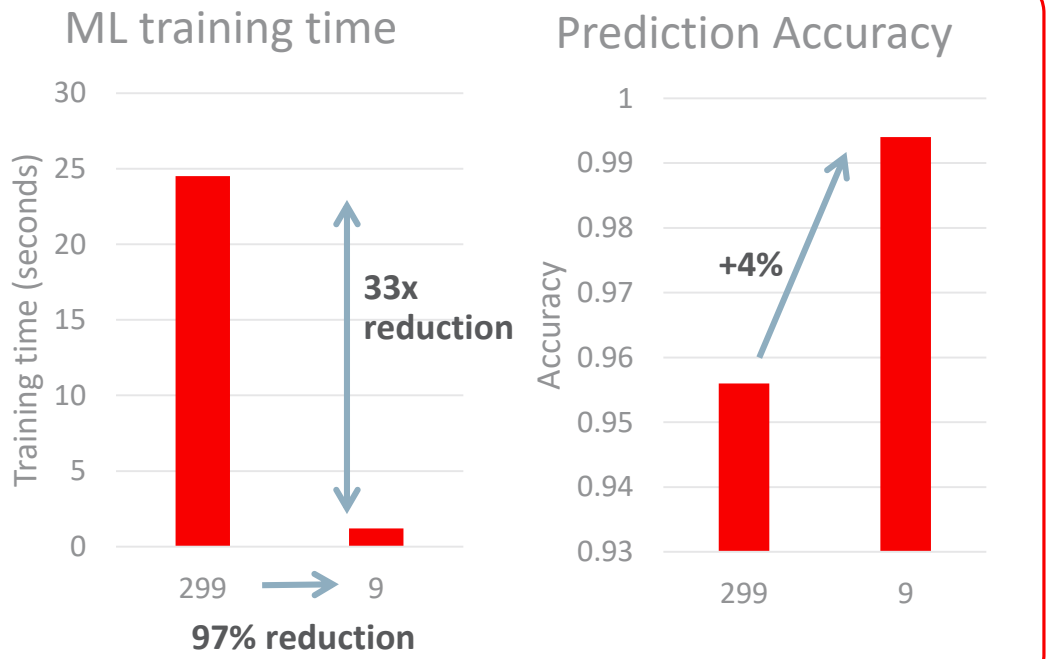
Increase data scientist productivity – reduce overall compute time
Uses in-database algorithms



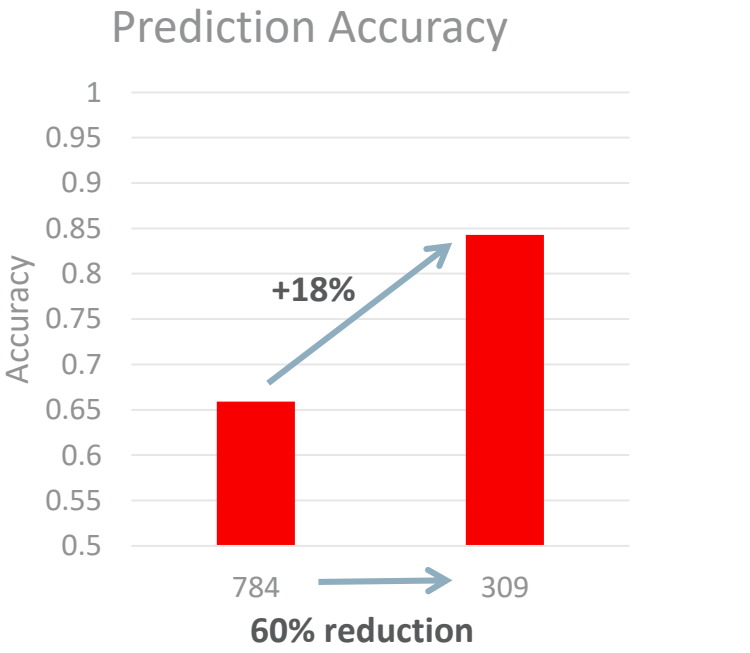
- Auto Model Selection
  - Identify algorithm that achieves highest model quality
  - Find best model faster than with exhaustive search
- Auto Feature Selection
  - Reduce # of features by identifying most predictive
  - Improve performance and accuracy
- Auto Tune Hyperparameters
  - Significantly improve model accuracy
  - Avoid manual or exhaustive search techniques

OML4Py Auto Feature Selection: examples

Reduce # features by identifying most relevant
Improve performance and accuracy



OpenML dataset 312 with 1925 rows, 299 columns



OpenML dataset 40996 56K rows, 784 columns

Core APIs Feature Summary

SQL plus two most popular open source languages for machine learning

Feature	OML4SQL	OML4Py	OML4R
Transparency Layer	n/a	✓	✓
Parallel, Distributed Algorithms	✓	✓	✓
Embedded Execution	n/a	✓	✓
Automated Data Preparation	✓	✓	✓
Automated Text Processing	✓	✓	✓
Partitioned Models	✓	✓	✓
Automated Machine Learning (AutoML)		✓	
PGX Integration for Graph Analytics	implicit		✓
DML table package transparency	n/a		✓
Extensible Algorithm Models	✓		✓

n/a – not applicable

Oracle Machine Learning for Spark (OML4Spark)

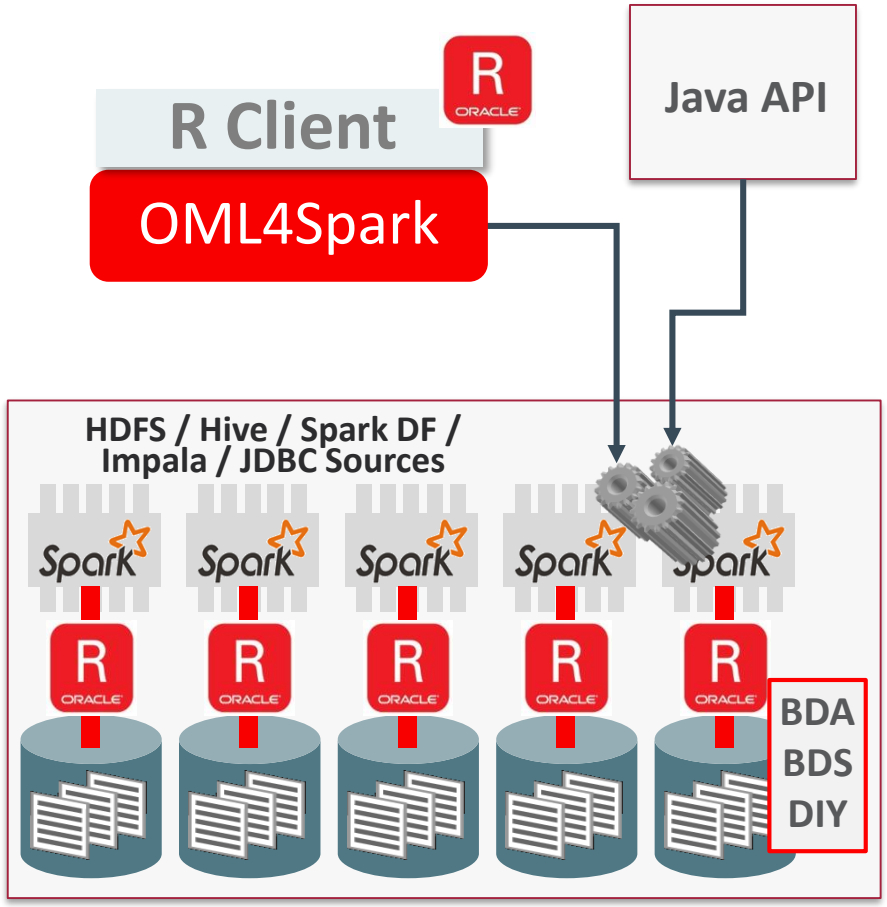
supported by Oracle R Advanced Analytics for Hadoop

ORACLE

Oracle Machine Learning for Spark

R Language API Component to Oracle Big Data Connectors

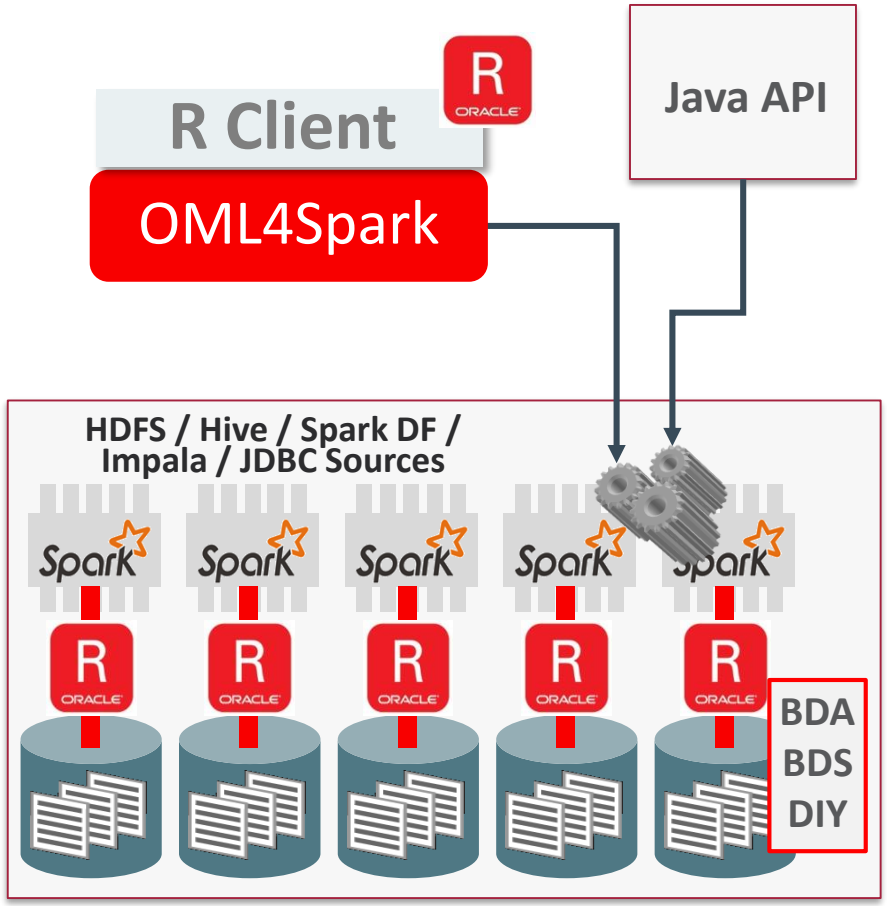
- Leverage Spark 2 environment for powerful data preparation and machine learning
- Use data across range of Data Lake sources
- Achieve scalability and performance using full Hadoop cluster
- Parallel and distributed machine learning algorithms from native and Spark MLlib implementations
- Use expressive R Formula specification



Oracle Machine Learning for Spark

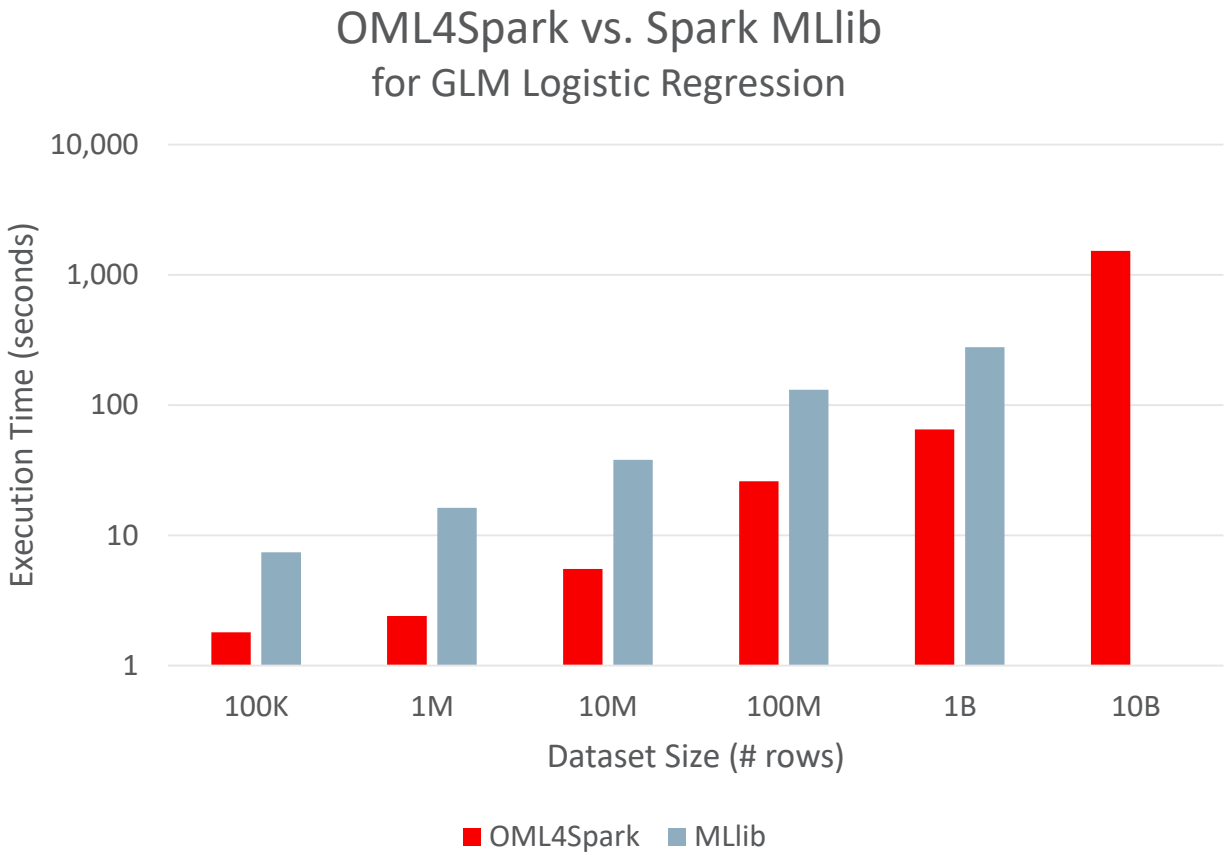
R Language API Component to Oracle Big Data Connectors

- Transparency layer
  - Proxy objects reference data from file system, HDFS, Hive, Impala, Spark DataFrame and JDBC sources
  - Overloaded R functions translate functionality to native language, e.g., HiveQL for HIVE and Impala
  - Users manipulate data via standard R syntax
- Parallel, distributed machine learning algorithms
  - Scalability and performance leveraging full Hadoop cluster
  - Spark-based custom LM, GLM, NN, K-Means plus Spark MLlib algorithms
  - Use expressive R Formula specification
- Compute framework with custom R mappers/reducers
  - Data-parallel and task-parallel execution
  - Allows for open source CRAN packages run on the Cluster Nodes



OML4Spark Performance

- Logistic Regression (GLM)
- Data fits in memory
  - Up to 7x faster than Spark MLlib
- Data cannot fit memory
  - Able to solve a 10B row model
- Benchmark environment
  - ORAAH 2.8.0
  - Big Data Appliance X7-2
  - 6 Nodes, 256GB of RAM per Node



Formula: cancelled ~ distance + origin + dest + as.factor(month) + as.factor(year) + as.factor(dayofmonth) + as.factor(dayofweek) + as.factor(flightnum)

Enabling Oracle Applications with OML Models and Microservices

Applications integrating Oracle Machine Learning

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

Oracle Integration Cloud
Digital Process Automation

ORACLE HUMAN CAPITAL MANAGEMENT CLOUD

ORACLE SALES CLOUD

ORACLE RETAIL CLOUD

ORACLE CONFIGURE, PRICE, AND QUOTE CLOUD

ORACLE CLOUD PLATFORM CONTENT AND EXPERIENCE

Integration
ORACLE CLOUD

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

ORACLE E-BUSINESS SUITE

Oracle Identity Management
Adaptive Access Management

ORACLE IDENTITY MANAGEMENT

FSGBU
Analytical Applications Infrastructure

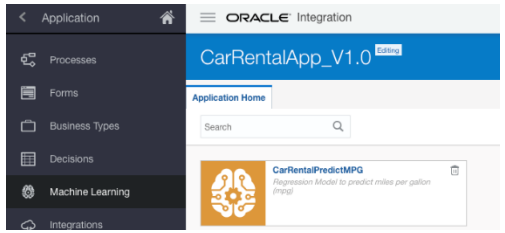
ORACLE FINANCIAL SERVICES

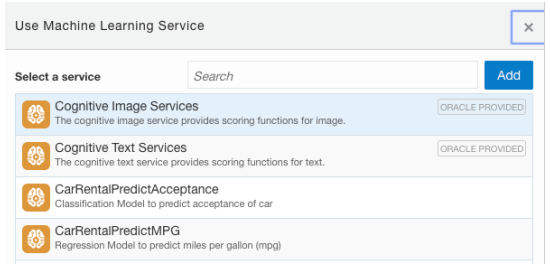
ORACLE

Application Integration Spotlight – Platforms *using* OML Platform

Oracle Integration Cloud (OIC)

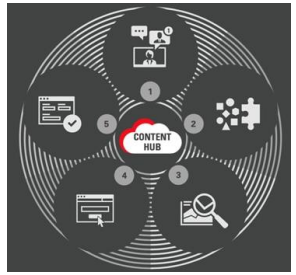
- Digital Process Automation
  - Help business users make better decisions by using recommendations from ML models
  - Increase automation of human-centric approval workflows
- Used by Oracle SaaS process-centric apps
- PaaS service that exposes OML features
  - Build models in ADB, deploy via OML Microservices





Oracle Content and Experience (OCE)

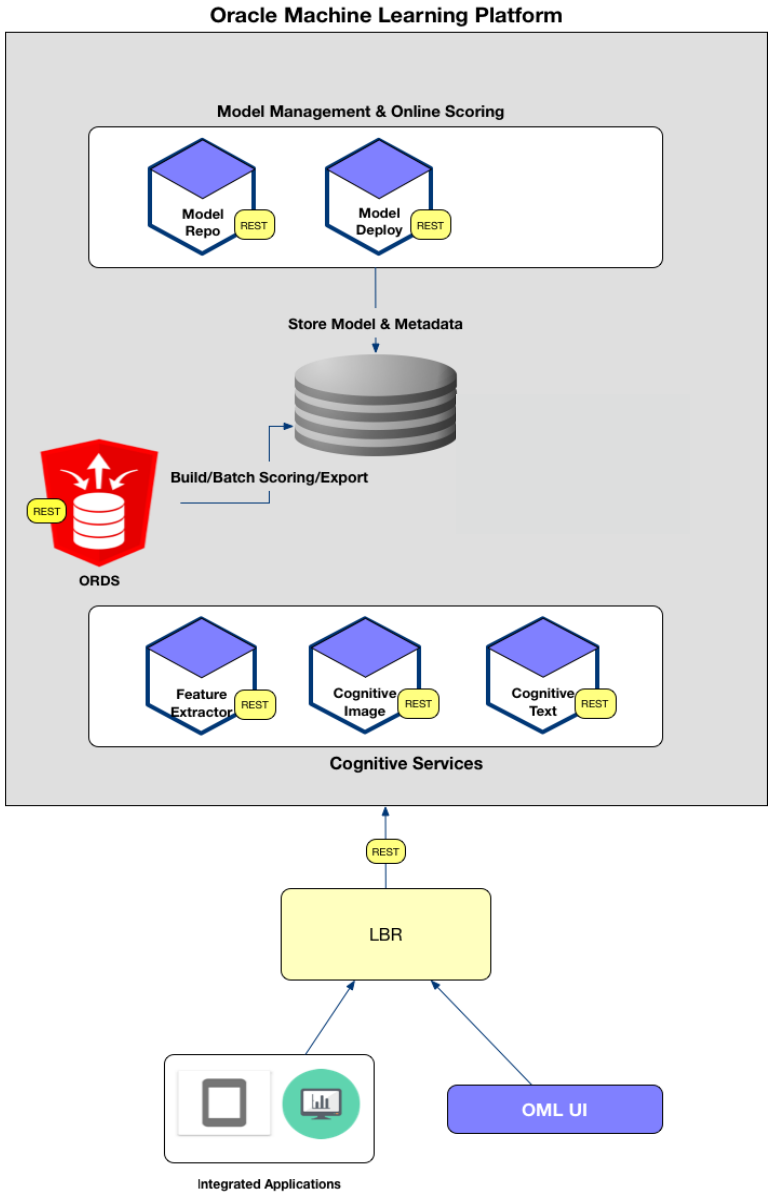
- Improve Content Discoverability
  - Search, organize content, reduce duplication
  - Find relevant images/docs during content creation
  - Automatic tagging and classification of images, videos, text
  - Visual search
- Cloud-based content hub to drive omni-channel content management and accelerate experience delivery
- Leverages OML cognitive microservices



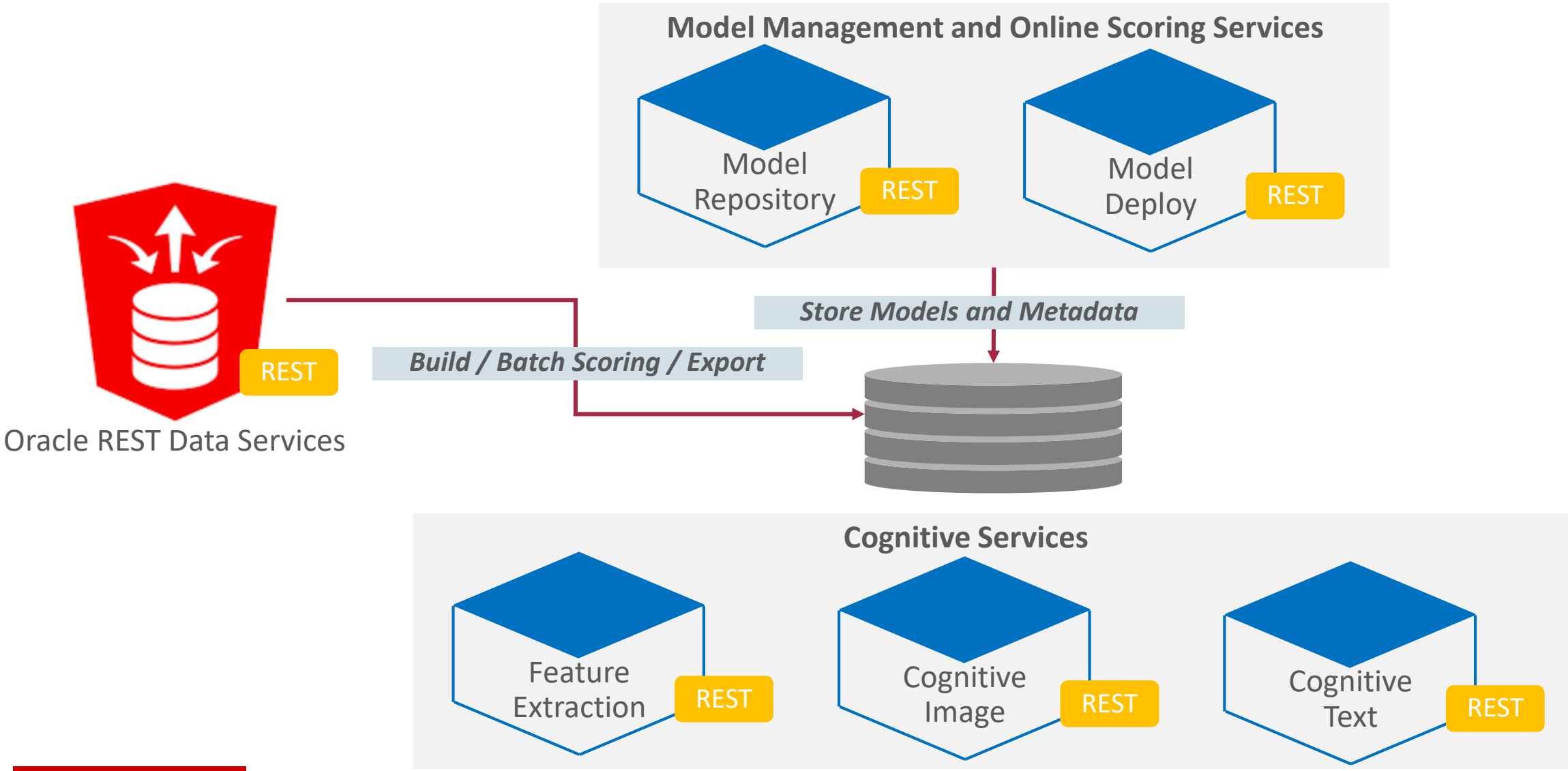
Oracle Machine Learning Microservices

Available now *internally* to Oracle Applications teams

- Model Management Services
  - Building and deploying OML models
- Cognitive Services
  - Feature Extraction, Image and Text
- Model repository
  - Store, version, compare ML models
- REST APIs for application integration
- Docker Containers for portability
- Kubernetes support for container management



Oracle Machine Learning Platform REST APIs



Oracle Machine Learning

- Automation, scalability, and performance
- Machine learning model deployment for applications
- Integrated with Oracle Database, Big Data, and Cloud environments
- APIs for REST, SQL, R, and Python

For more information…

https://www.oracle.com/database/technologies/datawarehouse-bigdata/machine-learning.html


ORACLE®