

ORACLE

Oracle Machine Learning

Move the Algorithms – Not the Data

Umut Nazlıca

Principal Technology Architect , Technology Solution and Cloud Engineering

May 2022

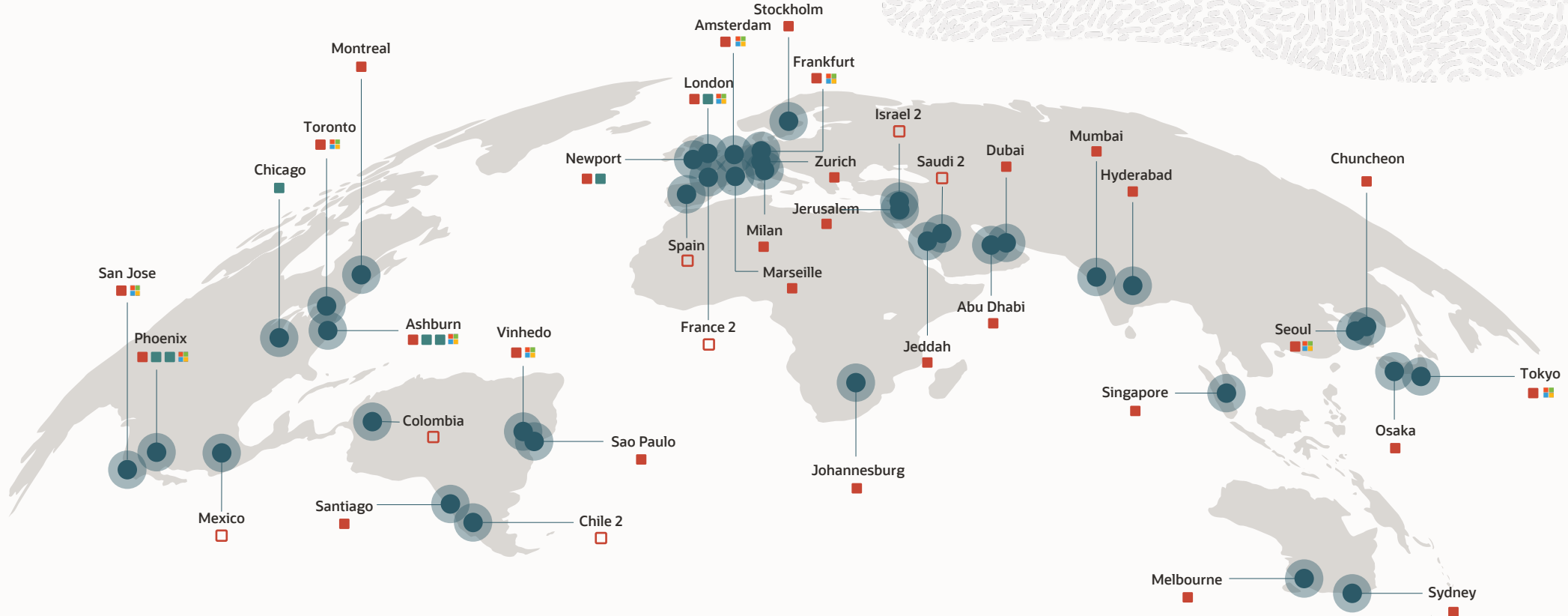


Safe harbor statement

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, 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 Cloud Infrastructure



Ocak 2022

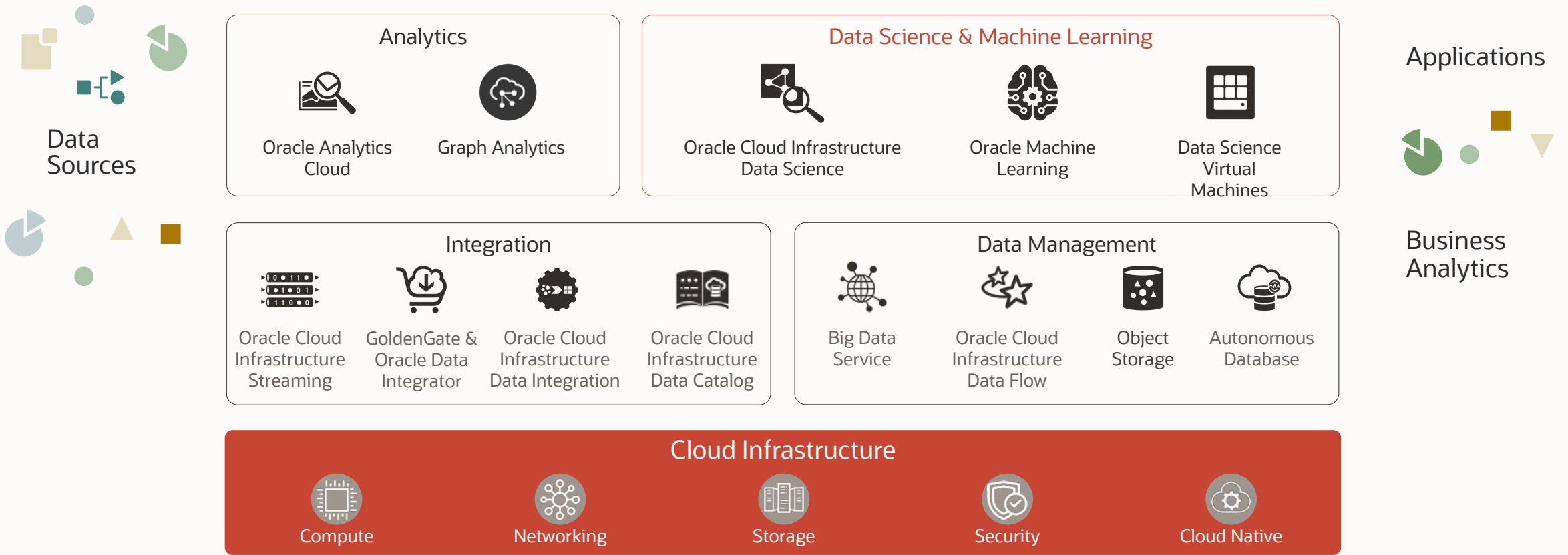
37 bölge; 2022 sonuna kadar planlanan bölge 7

10 Azure Interconnect Bölgesi



Oracle Data Science Platform

Machine learning supported by data sources, ingest, management and analytics



Oracle Cloud Infrastructure Data Science

Support for Python and open source

Accelerated

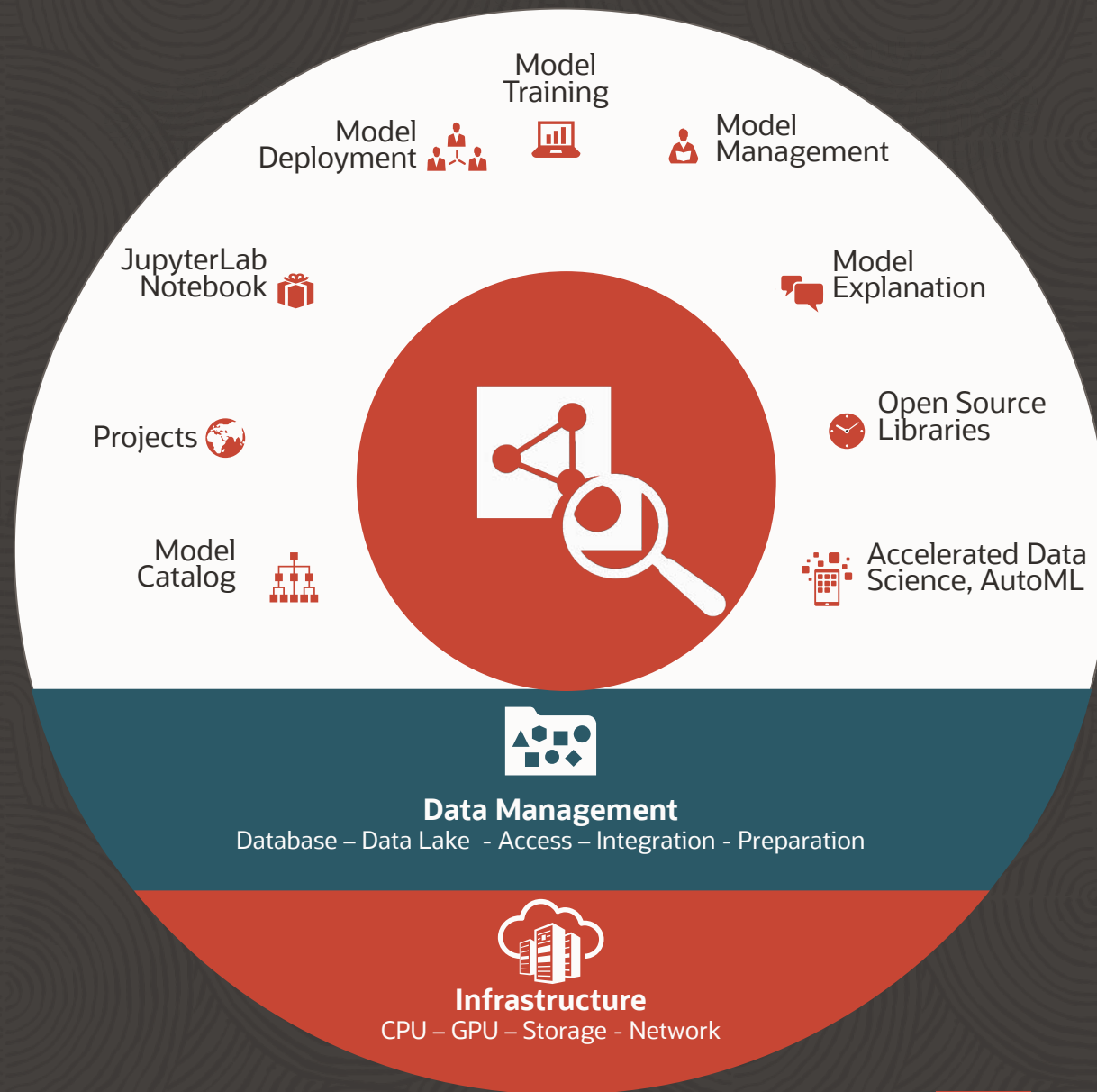
Allow data scientists to work the way they want to, and provide access to automated workflows, the best of open source, and a streamlined approach to building models.

Collaborative

Enable data science teams to work together with ways to share and reproduce models in a structured, secure way for enterprise-grade results.

Enterprise-Grade

Provide a fully managed platform built to meet the needs of the modern enterprise



Why Oracle Machine Learning for the database environment?



Data
in the database



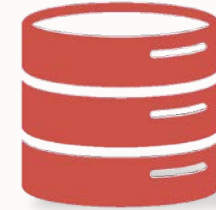
Algorithms
in the database



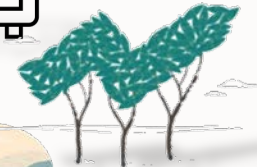
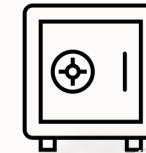
Models
in the database



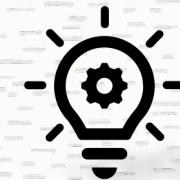
No data
movement



Security
Backup & Recovery
Scalability
Performance



Fewer moving parts
Reduced complexity



Oracle Machine Learning

OML4SQL

OML Notebooks

OML4R

Oracle Data Miner

OML4Py

OML4Spark

OML AutoML UI

OML Services

Interfaces for 3 popular data science languages: SQL, R, and Python

Collaborative notebook environment based on Apache Zeppelin with Autonomous Database

SQL Developer extension to create, schedule, and deploy ML solutions through a drag-and-drop interface

ML for the big data environment from R with scalable algorithms

No-code AutoML interface on Autonomous Database

Model Deployment and Management, Cognitive Text on Autonomous Database

Oracle Machine Learning Algorithms and Analytics in Oracle Database

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, irregular time series

REGRESSION

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

ATTRIBUTE IMPORTANCE

- Minimum Description Length
- Principal Component Analysis (PCA)
- Unsupervised Pairwise KL Divergence
- CUR decomposition for row & AI

ASSOCIATION RULES

- A priori

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.

Includes support for Automatic Data Preparation, Partitioned Models, Transactional data and aggregations

** New in 21c*

Oracle Machine Learning for SQL

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

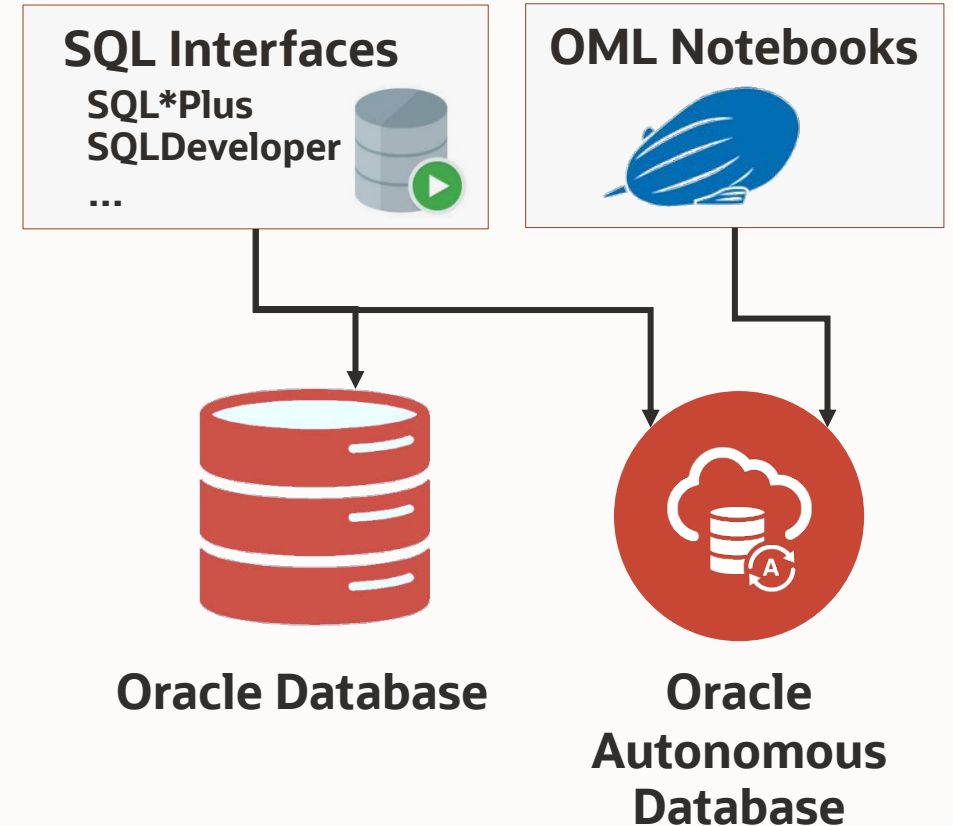
In-database, parallelized, distributed algorithms

- No extracting data to separate ML engine
- Fast and scalable
- Batch and real-time scoring at scale that leverages Exadata storage-tier function pushdown
- Algorithm-specific automatic data preparation
- Explanatory prediction details

ML models as first-class database objects

- Access control per model
- Audit user actions
- Export / import models across databases
- Ease of backup, recovery, and security

Faster time-to-market through immediate solution deployment



Oracle Machine Learning for R and Python

Empower data scientists with open source environments

Oracle Database as HPC environment

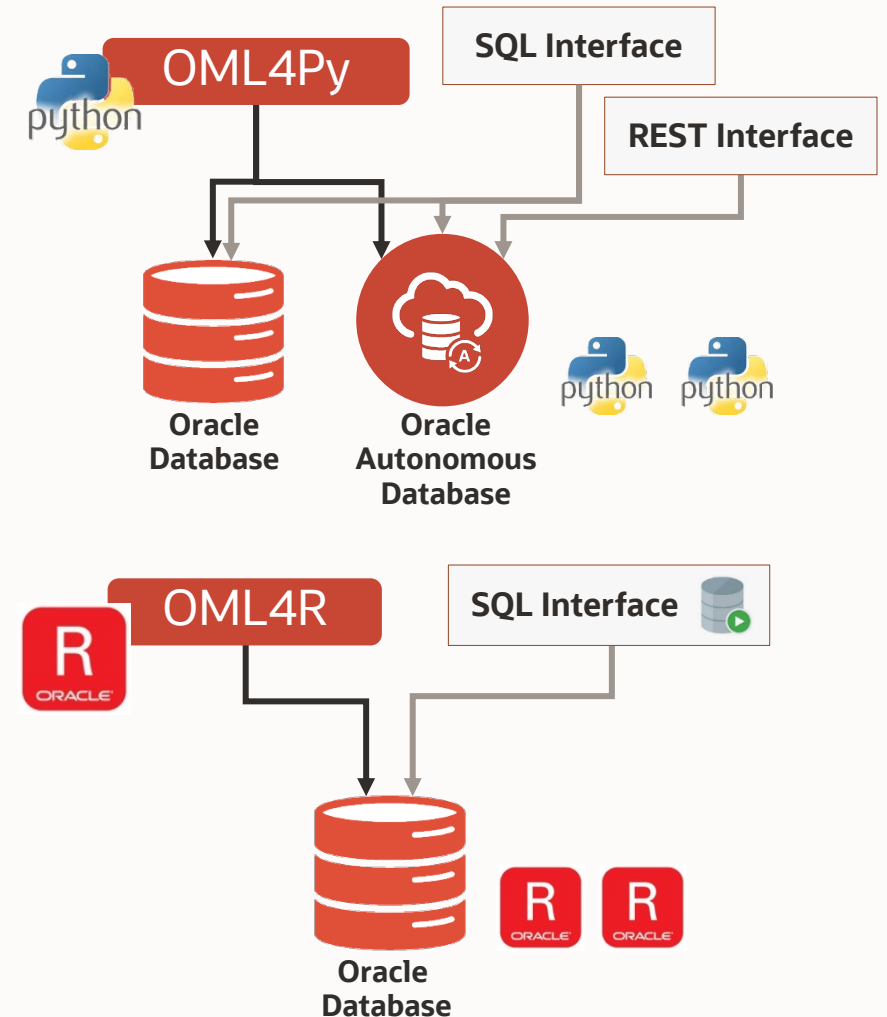
In-database parallelized and distributed machine learning algorithms

Manage scripts and objects in Oracle Database

Integrate results into applications

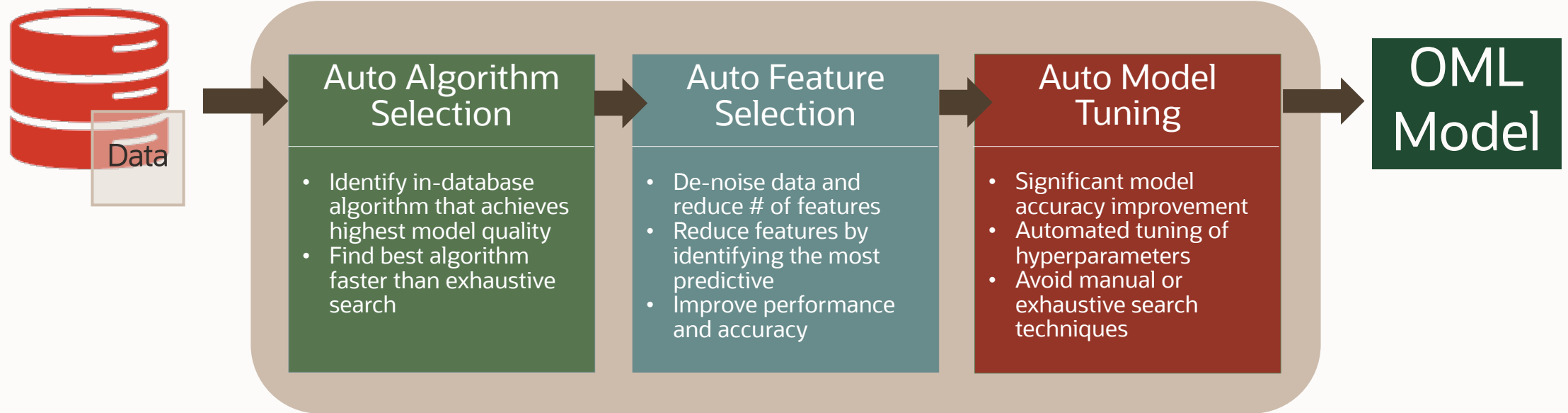
and dashboards via SQL or REST

OML4Py automated machine learning



AutoML with OML4Py

Increase data scientist productivity – reduce overall compute time



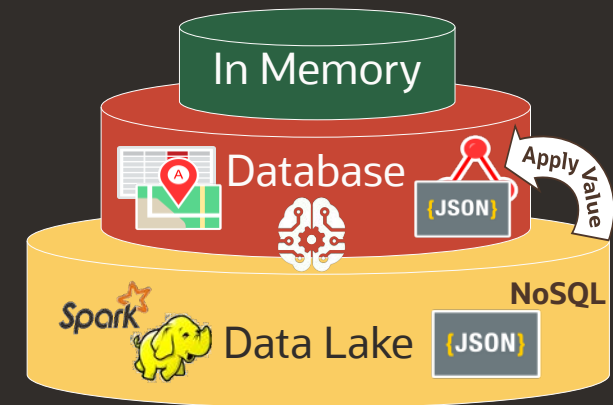
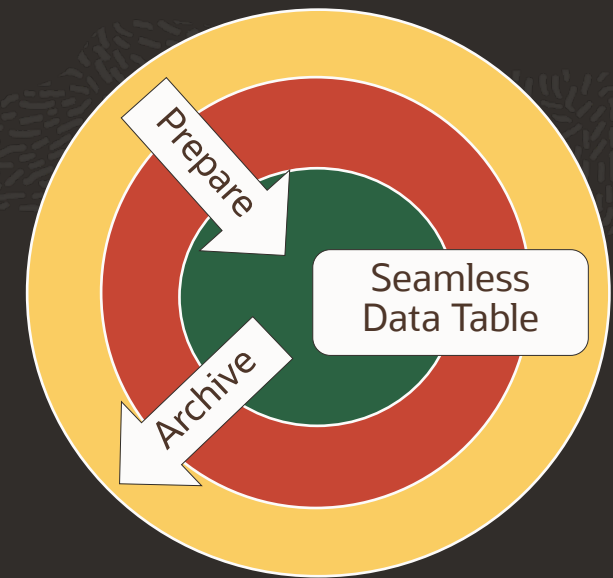
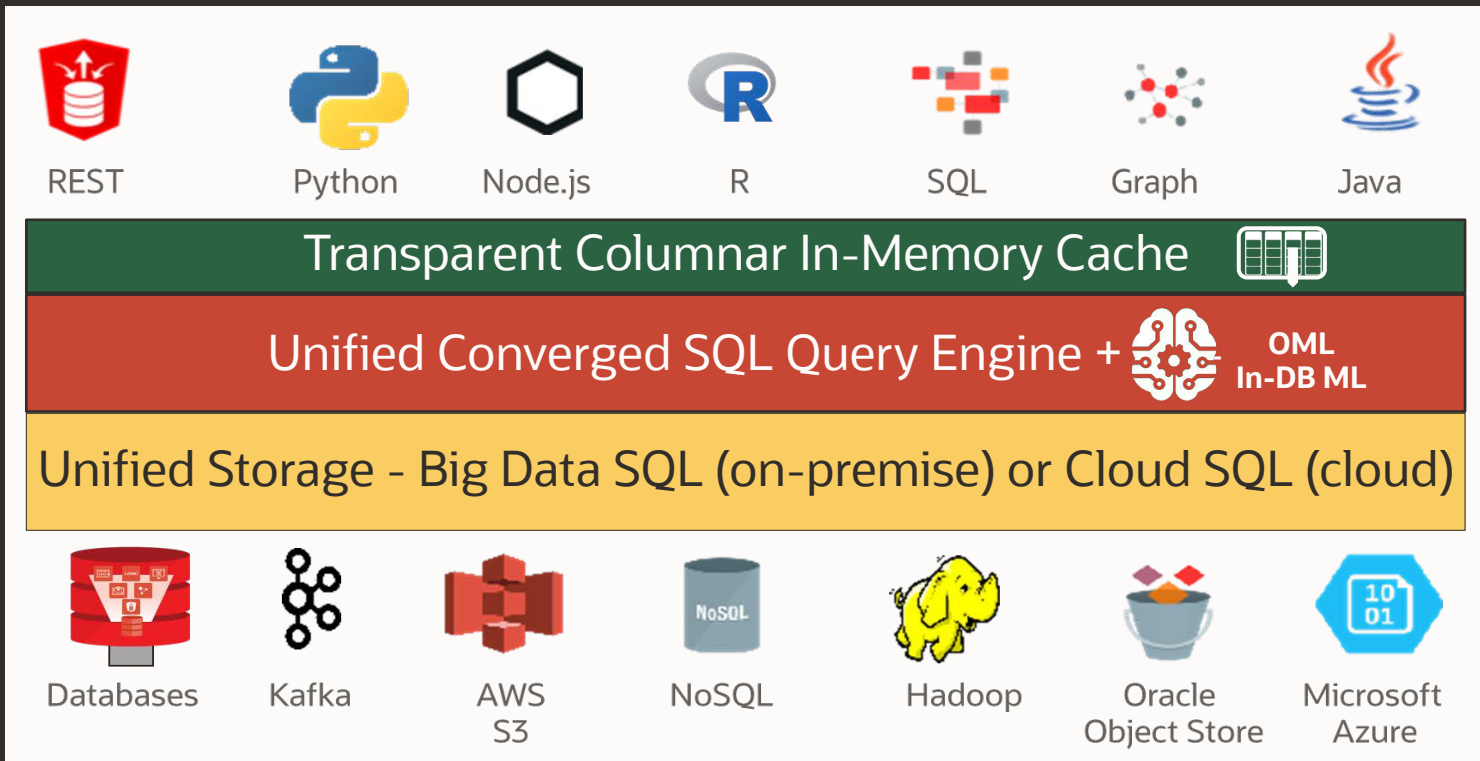
Enables non-expert users to leverage Machine Learning

Holistic View of All Data breaking Silos

Seamless Access to All Data where it resides

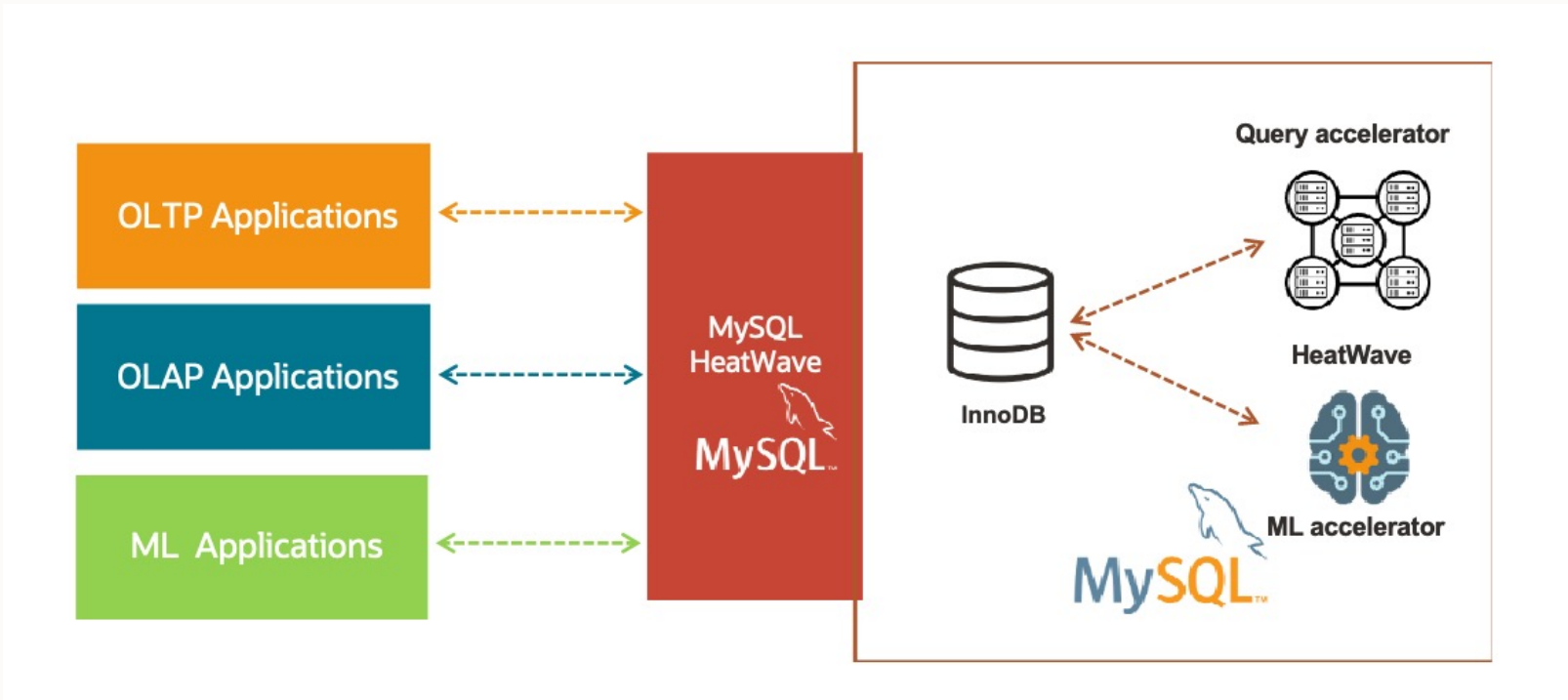
In-Database ML and Auto-ML

Much more than just Database Federation or Virtualization



MySQL HeatWave

The only MySQL service with a native, massively-scalable query accelerator



Unified solution for OLTP & OLAP

No ETL

Real-time analytics

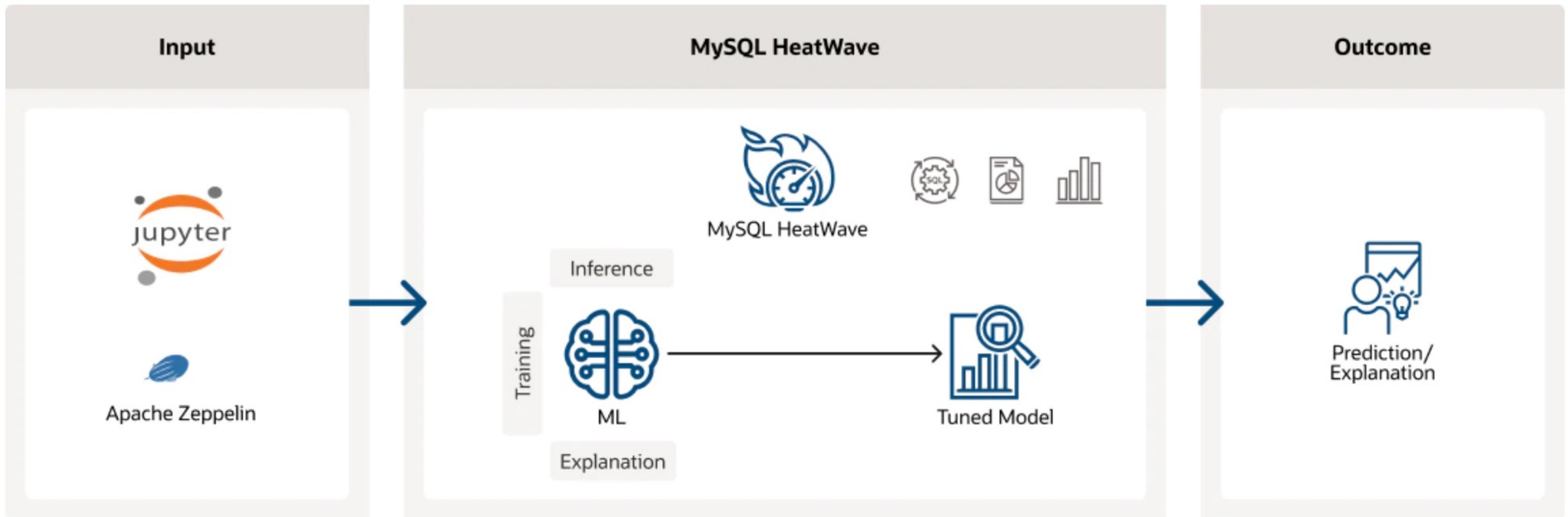
Applications work without changes

Faster than comparable services, at a fraction of the cost

Scales to thousands of cores

MySQL HeatWave ML

In-database machine learning



<https://www.oracle.com/mysql/heatwave/>

OML Services

Supports lightweight model scoring using REST endpoints for application integration

Enable key elements of overall enterprise MLOps strategy

Fast data scoring performance for streaming and real-time applications

Pay only for actual scoring compute – no pre-provisioned VM

Facilitate collaboration across data science team

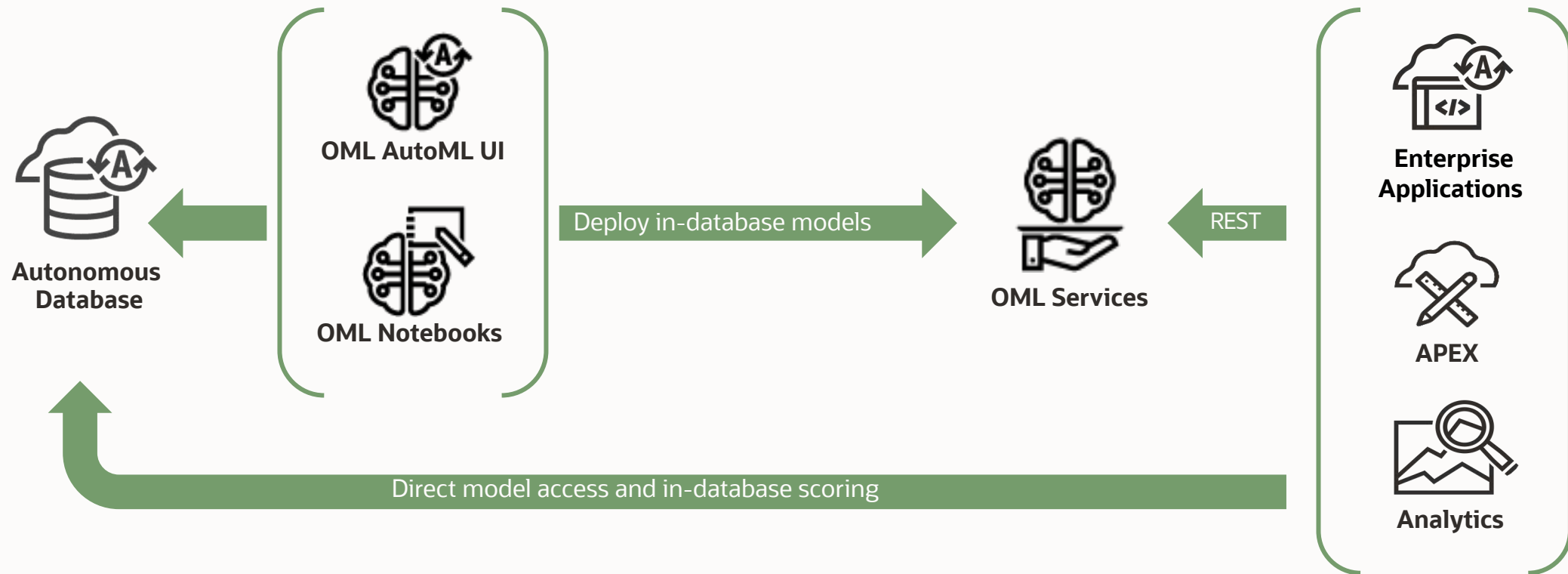
Model Management and Deployment Services

- Deploy in-database (native format) and third-party (ONNX format) models
- Import ONNX for Tensorflow, PyTorch, MXNet, scikitlearn, etc.
- Store, version, compare ML models
- Organize models within namespaces

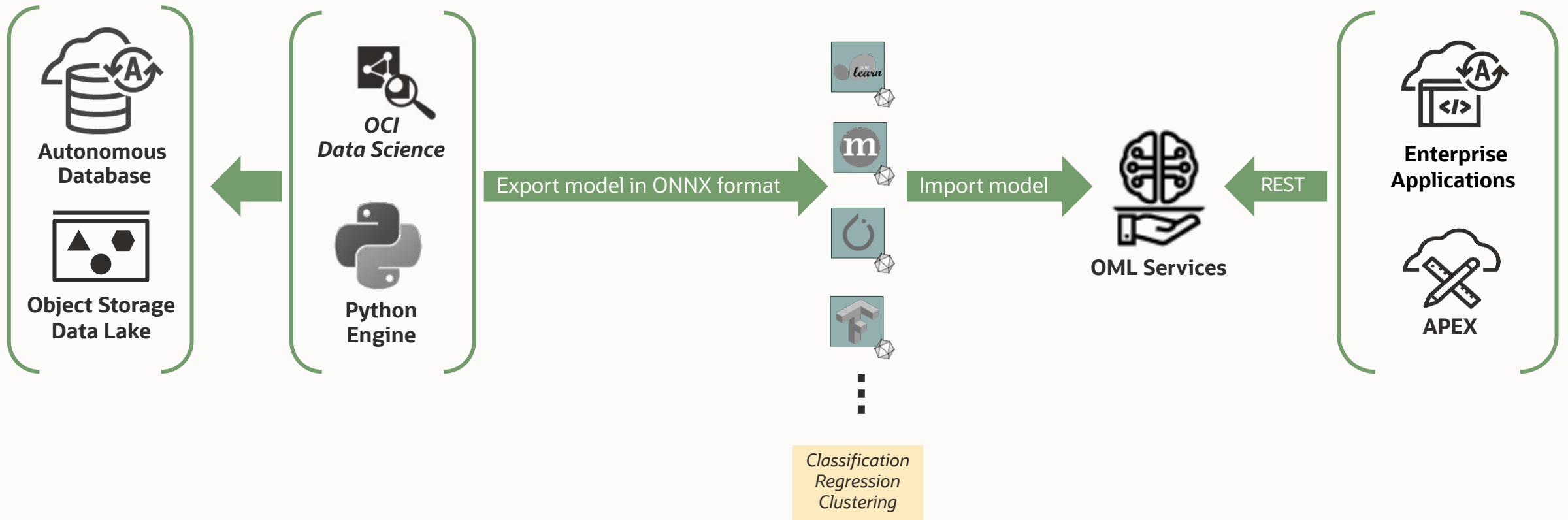
Built-in cognitive text services (English, French, Spanish, Italian)

- Extract topics and keywords
- Sentiment analysis (English only)
- Text summary and similarity

In-database model deployment scenarios with OML Services



Third-party model deployment scenarios



Oracle Machine Learning

Benefits summary



- **Minimize or eliminate data movement** for database data
- **Multi-persona, collaborative, democratized** machine learning for data scientists, citizen data scientists, developers
- **Multi-language API** (SQL, Python, R) and **no-code user interfaces**
- **Access from broader data lake data** through external tables and Big Data SQL
- Data and model **governance** via Oracle Database and Autonomous Database security models in development and production
- **Scalable and high-performance** modeling and scoring
- **Elastic scaling** for machine learning as part of OML on Autonomous Database and OML Services
- **Automated Machine Learning** (AutoML) through Python and no-code UI
- **Model explainability** and prediction details to understand models and their predictions better
- **Bridges gap** between development and production with model deployment options
- **MLOps** capabilities include immediate model production deployment from SQL and REST, user collaboration, queryable model repositories, and support for streamlined creation of reproducible ML pipelines
- **Oracle stack, SaaS, PaaS, IaaS** provides a strong environment in which data engineers, ML engineers and architects, corporate developers and others can contribute to the DS and ML workflow
- OCI with Autonomous Database supports **predictable performance at an affordable and predicible cost**
- **On-premises** and **Cloud** availability for ML capabilities
- **Oracle tools** and enterprise **applications integration**, including Oracle Analytics Cloud and Oracle APEX
- **Simple pricing structure** - ML and AutoML capabilities included in core product at **no additional cost**

Resources

OML Webpage

<https://oracle.com/machine-learning>

OML Blog

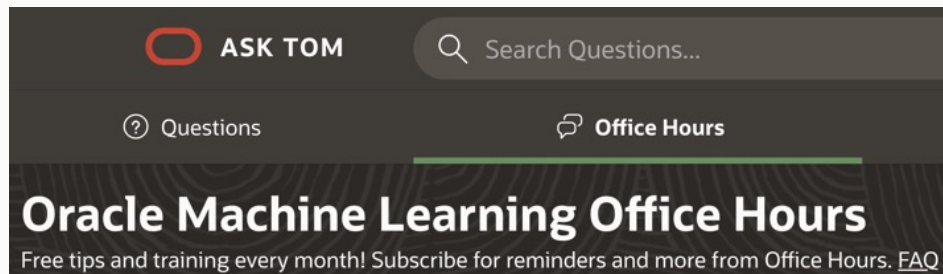
<https://bit.ly/omlblogs>

OML GitHub Repository

<https://bit.ly/omlgithub>

OML Office Hours

<https://bit.ly/omlofficehours>

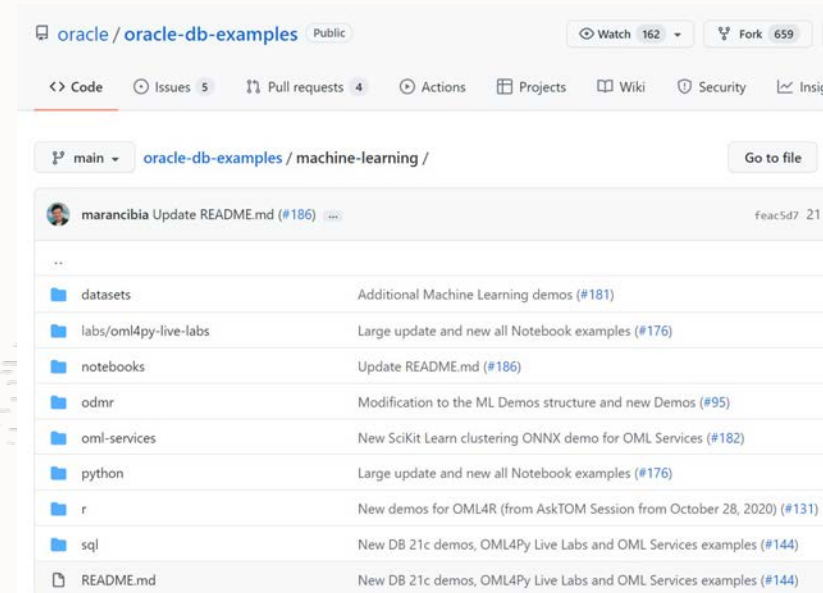


ASK TOM Search Questions...
Questions Office Hours
Oracle Machine Learning Office Hours
Free tips and training every month! Subscribe for reminders and more from Office Hours. [FAQ](#)



Oracle Machine Learning

- Import Wide Datasets into Nested Columns Using OML4Py**
Jie Liu | 12 minute read
- Oracle Data Miner now Available for Autonomous Database**
Sherry LaMonica | 4 minute read
- Explore Oracle Machine Learning for your NYR**
Mark Horrick | 6 minute read
- Top 10 Reasons to use Machine Learning in Oracle Database**
Mark Horrick | 8 minute read



oracle / oracle-db-examples Public | Watch 162 | Fork 659

Code Issues 5 Pull requests 4 Actions Projects Wiki Security Insights

main oracle-db-examples / machine-learning / Go to file

marancibia Update README.md (#186) feac5d7 21 day

...	
datasets	Additional Machine Learning demos (#181)
labs/oml4py-live-labs	Large update and new all Notebook examples (#176)
notebooks	Update README.md (#186)
odmr	Modification to the ML Demos structure and new Demos (#95)
oml-services	New SciKit Learn clustering ONNX demo for OML Services (#182)
python	Large update and new all Notebook examples (#176)
r	New demos for OML4R (from AskTOM Session from October 28, 2020) (#131)
sql	New DB 21c demos, OML4Py Live Labs and OML Services examples (#144)
README.md	New DB 21c demos, OML4Py Live Labs and OML Services examples (#144)



Resources

OML on ADB Specialist Certification

<https://bit.ly/omlcertification>

Oracle Learning Subscription for OML

<https://bit.ly/omllearning>

OML Live Labs

OML Fundamentals on ADB <https://bit.ly/omlfundamentalshol>

OML4Py <https://bit.ly/oml4pyhol>

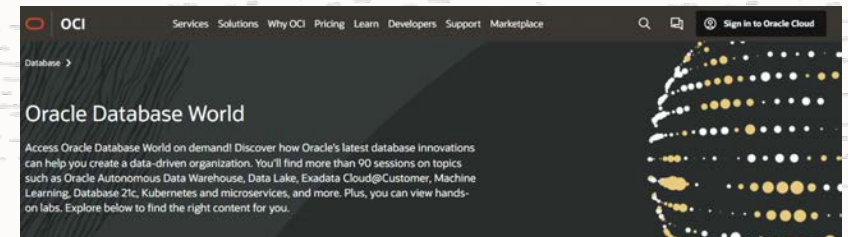
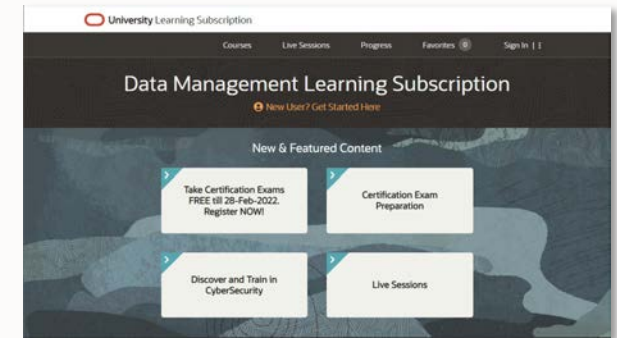
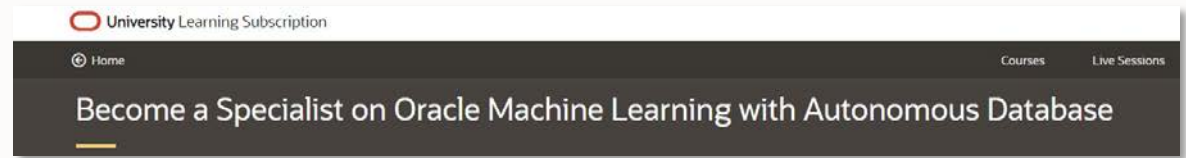
All OML <https://bit.ly/omllivelabs>

OML Sessions from Oracle Database World

<https://blogs.oracle.com/machinelearning/post/join-us-for-machine-learning-sessions-at-oracle-database-world-2021>

LinkedIn Group

Join: [Oracle Machine Learning](#)



Thank you

Umut Nazlıca

umut.nazlica@oracle.com



<https://www.linkedin.com/in/umutnazlica/>



Oracle Data Miner User Interface

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

SQL Developer Extension for Oracle Database, DBCS, and Autonomous Database

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

The screenshot displays the Oracle SQL Developer interface with the Data Miner extension. The main workspace shows a workflow diagram starting with a data source 'CUST_INSUR_LTV1'. This source is processed through several steps: 'Clustering Segmentation 1', 'Filter Columns', 'Multiple Classification Models', and 'Most Likely Customers'. The final output is 'LIKELY_BUY_INSURANCE_CUSTMRS'. The 'Filter Columns' step is expanded to show 'Filter Columns Details 1'. The 'Multiple Classification Models' step is expanded to show a tree of models, including 'Model 10' through 'Model 27'. The 'Most Likely Customers' step is expanded to show 'Model 1' through 'Model 6'. The 'Query Builder' shows a SQL query:

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

 Below the query, the 'Query Result' table shows the top 5 most suspicious fraud policy holder claims:

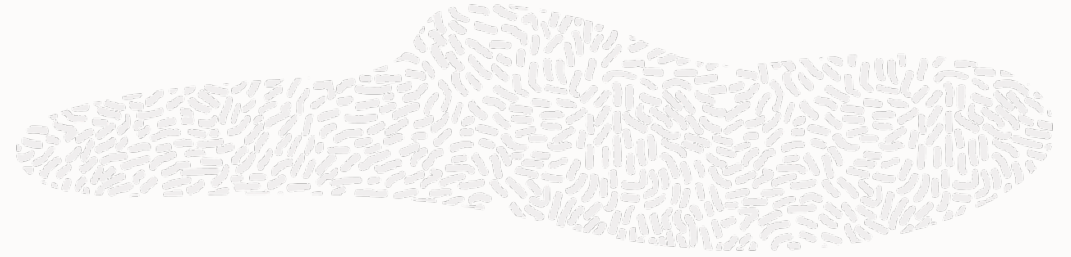
POLICYNUMBER	PERCENT_FRAUD	RNK
654	61.87	1
11068	57.37	2
7435	55.47	3

 The 'Script Output' shows the execution of the SQL code. The 'Components' panel on the right lists various data science tasks like 'Aggregate', 'Filter Columns', 'Join', 'JSON Query', 'Anomaly Detection', 'Association', 'Classification', 'Clustering', 'Feature Extraction', 'Model', 'Model Details', 'Regression', 'Predictive Queries', 'Anomaly Detection Query', 'Clustering Query', 'Feature Extraction Query', and 'Prediction Query'. The 'Evaluate and Apply' and 'Linking Nodes' panels are also visible at the bottom right.



Oracle Machine Learning Notebooks

Autonomous Database as a Data Science Platform



Collaborative UI

- Based on Apache Zeppelin
- Supports data scientists, data analysts, application developers, and DBAs with SQL and Python
- Easy notebook sharing
- Scheduling, versioning, access control

Included with Autonomous Database

- Automatically provisioned and managed
- In-database algorithms and analytics functions
- Explore and prepare, build and evaluate models, score data, deploy solutions

The screenshot displays the Oracle Machine Learning for Python interface. The top navigation bar includes the Oracle logo, 'Machine Learning', and user information 'OMLUSER'. The main content area is divided into three notebook panels:

- Overloaded data visualization functions:** This notebook demonstrates the use of overloaded `boxplot` and `hist` functions. It includes a text description of the `boxplot` function, which performs statistical computations in-database. The code cell shows the following Python code:

```
%python
import matplotlib.pyplot as plt
plt.style.use('seaborn')
plt.figure(figsize=[10,5])

oml.graphics.boxplot(IRIS[:, :4], notch=True, showmeans = True,
                    labels=IRIS.columns[:4])
plt.title('Distribution of IRIS Attributes')
plt.xlabel('cm')
```

The output is a boxplot titled 'Distribution of IRIS Attributes' showing the distribution of 'SEPAL_LENGTH', 'SEPAL_WIDTH', 'PETAL_LENGTH', and 'PETAL_WIDTH'.
- Sepal Length variation in IRIS data set:** This notebook shows a histogram of the 'SEPAL_LENGTH' attribute. The code cell contains:

```
%python
oml.graphics.hist(IRIS['SEPAL_LENGTH'], bins=10, color='red',
                linestyle='solid', edgecolor='black')

plt.title('Sepal Length variation in IRIS data set')
plt.xlabel('Sepal Length')
plt.ylabel('# of iris instances')
plt.show()
```

The output is a red histogram titled 'Sepal Length variation in IRIS data set' with the x-axis labeled 'Sepal Length' and the y-axis labeled '# of iris instances'.
- Create derived variables:** This notebook shows the creation of a derived variable. The code cell contains:

```
%python
is_large_petal = (IRIS['PETAL_LENGTH'] > 5.0) & (IRIS['PETAL_WIDTH'] > 2.0)
```



OML AutoML UI

Enhance data scientist productivity and enable non-expert data professionals

Accelerate new ML projects

Automate repetitive and time-consuming tasks

Generate editable notebooks for selected models

Deploy models as REST endpoints

Featuring

- Monitor experiment progress
- Customize selection quality metric and metrics display
- Even faster data scoring performance for streaming and real-time applications

