

Cloud **Essentials**

# Build Machine Learning Solutions with Oracle's Services and Tools

Remove complexity from machine learning and data science projects with cloud-based technologies

**ORACLE**  
Cloud







# Machine Learning Gains Ground

Machine learning technology is growing in popularity—and it is fueling one of the hottest segments of the software industry. According to a report by Market Research Future, the global machine learning market will see a compound annual growth rate of 42 percent between 2018 and 2024, driven in part by the rise in the ability to ingest and process unstructured data at scale, and the widespread adoption of cloud-based services.<sup>1</sup>

Data scientists use machine learning algorithms to extract knowledge and insights from a given data set. These insights can be used by applications and business analytics systems to drive decision-making.

This brief describes how you can use machine learning technology in conjunction with a comprehensive data platform that includes mature Oracle technologies for data management, data preparation, and analysis—backed by complementary applications, tools, frameworks, and infrastructure.



<sup>1</sup> Machine Learning Market Research Report - Global Forecast to 2024, September, 2019



# Who Uses Machine Learning?

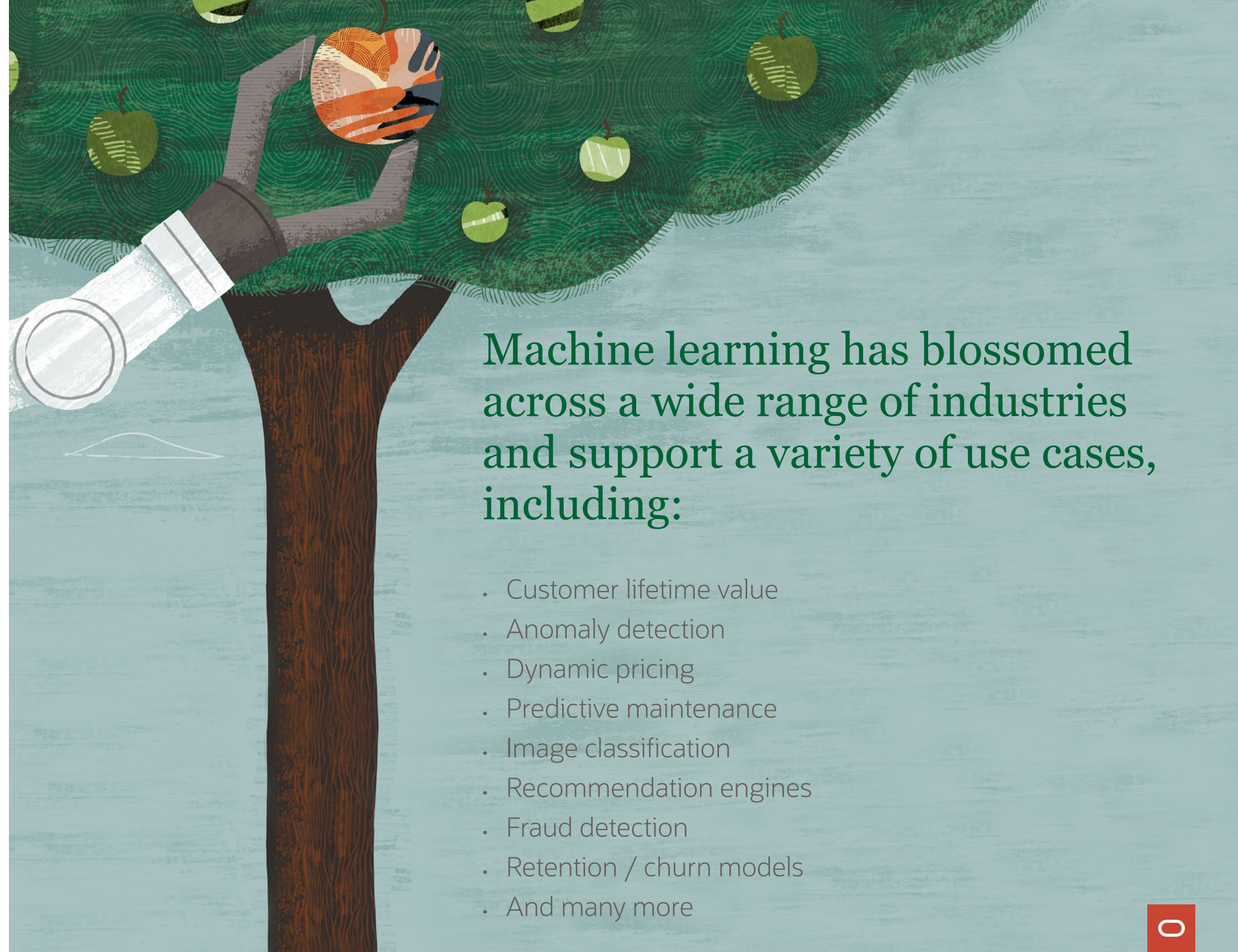
Whenever we interact with banks, shop online, or use social media networks, machine learning algorithms are making our experiences more valuable, efficient, and secure.

According to The Data Warehouse Institute (TDWI), which surveyed a wide range of companies about their use of AI and machine learning, 92 percent of today's companies use machine learning technology in some fashion and 85 percent are building predictive models with machine learning tools.<sup>2</sup>

For example, financial institutions use machine learning to determine a person's credit score to aid in loan approval decisions. Manufacturers use machine learning to monitor production equipment and avert potential downtime, or identify root causes of product defects.

From improving crop yields to predicting fraud, forward-looking organizations depend on machine learning technologies to gain better insights, make better decisions, and improve their competitive advantage in rapidly evolving markets.

<sup>2</sup>Halper, Fern, Ph. D., Best Practices Report: "Driving Digital Transformation Using AI and Machine Learning" (tdwi.org/bpreports – September 25, 2019).



## Machine learning has blossomed across a wide range of industries and support a variety of use cases, including:

- Customer lifetime value
- Anomaly detection
- Dynamic pricing
- Predictive maintenance
- Image classification
- Recommendation engines
- Fraud detection
- Retention / churn models
- And many more



# Hurdles to Machine Learning Adoption

## Managing data, managing tools, and deploying models

In the past, businesses have typically made decisions based solely on historical data, gut feelings, or intuition. Today, machine learning enables business leaders to make data-driven decisions that are more insightful, forward-looking, and proactive.

To derive the most value from these machine learning initiatives, you need data management systems to store and secure your data. You also need hardware and software systems to run data science tools. Finally, you need to deliver results in a form that can be consumed by nontechnical users, as well as by other information systems. These challenges can be summed up in four primary ways:

- **Access:** Machine learning algorithms require lots of data, which is often spread across different applications and databases and controlled by different business units.
- **Complexity:** Preparing data and developing machine learning models are often time-consuming tasks and may involve the use of multiple specialized software tools, open-source libraries, and the coordination of multiple contributors and business units.
- **Scalability:** Data scientists need the ability to analyze more data, build more models, and score more data at scale. All of this drives data productivity.
- **Deployment:** To achieve tangible results from machine learning initiatives, models must be put in an operational context, i.e., *deployed*, and then monitored and updated as circumstances change.

## Defining Terms

- **Data Science** uses scientific methods and algorithms to derive knowledge and insights from structured and unstructured data.
- **Artificial Intelligence (AI)** is a branch of computer science dealing with the simulation of intelligent behavior in computers.
- **Machine Learning (ML)** entails the use of algorithms and statistical models to discover patterns and predict outcomes from data.



# Oracle Knows Machine Learning

Oracle has infused machine learning technology throughout nearly every facet of the company. Machine learning helps automate internal operations and has a huge and growing impact on the development of new products and services.

Oracle is accelerating the adoption of machine learning technology as part of a larger focus on data science and AI. Oracle creates machine learning technology that offers:

## Increased Productivity:

Enabling data scientists to create higher-quality models, faster

## Greater Flexibility:

Offering flexibility and the ability to work with data at scale across the enterprise, so data scientists can work with needed data to gain better results

## Easier Deployment:

Putting models into production faster to support data-driven business objectives



This is not a new venture for Oracle or its customers, many of whom have been developing machine learning solutions using Oracle Machine Learning-based technology for decades.

At the core of our data science platform are Oracle Cloud Infrastructure Data Science and Oracle Machine Learning. Cloud Infrastructure Data Science helps enterprises collaboratively build, evaluate, manage, and deploy machine learning. Designed for expert data scientists, Cloud Infrastructure Data Science provides access to the best of open-source algorithms and libraries, best-in-class hardware for CPUs, and fast networking and storage. It is also data-source agnostic and allows users to utilize Python for machine learning.

For those who wish to use the power of databases and data lakes for machine learning, Oracle Machine Learning provides in-database data preparation and exploration as well as model building, evaluation, and deployment. Designed for multiple user roles and collaboration, Oracle Machine Learning has Oracle-optimized algorithms, uses data from Oracle's extended data platform, and enables users to work with SQL and R.

Both Oracle Cloud Infrastructure Data Science and Oracle Machine Learning can be used at the same time and will work together so users can take advantage of either option, depending upon their needs and desires. Both of these services take advantage of several other services that complement them, and extend support for the rest of the data science lifecycle and data science team at large.





# About Oracle Cloud Infrastructure Data Science

Oracle has launched Oracle Cloud Infrastructure Data Science, based on the acquisition of DataScience.com. It was built to accelerate the workload of individual data scientists while also supporting the data science team.

Cloud Infrastructure Data Science is data source agnostic and gives data scientists flexible tools to build, manage, deploy, and monitor machine learning models. These skilled technology professionals can also use their favorite Python machine learning or deep learning frameworks while writing their code interactively within the JupyterLab notebook environment.

Cloud Infrastructure Data Science delivers powerful team capabilities including:

**Shared projects** that help data scientists organize and segregate work in different compartments with specific access control policies. In a project, data scientists can view and share models, access notebook sessions, and write version-controlled code.

Inside notebook sessions, users can also connect to Git providers and pull code, make changes tracked with version control, and sync back their changes with Git. All of this ensures that everyone can work off the same versioned code base.

**Notebook Sessions** that give data scientists the power to specify the compute and storage resources that will be used as they work in their JupyterLab notebooks. A notebook session represents a virtual machine containing the tools and libraries that data scientists use to build, train, and evaluate

models. The virtual machine is preloaded with the JupyterLabs user interface for Jupyter notebooks. It also comes with over 300 open-source machine learning libraries including:

- sci-kit learn
- XGBoost
- LightGMB
- Keras
- TensorFlow

- Dask
- pandas

Samples and tutorial notebooks are also included.

In addition to open source, Oracle provides other capabilities to enhance user productivity, such as:

**Accelerated Data Science (ADS) SDK** to make common data science tasks faster, easier, and less error-prone. This is a Python library unique to

Oracle that offers capabilities for accessing, profiling, and manipulating data. ADS also offers a simple interface for model evaluation and interpretation, and Oracle's AutoML engine for automated model training.

**Model Catalogs** to enable team members to reliably share their models and capture artifacts needed for model deployment. The model catalog also enables model auditability and reproducibility, allowing data scientists to re-run the same code in the same environment with the same dataset that was used to train a given model. The model catalog in Cloud Infrastructure Data Science tracks model metadata (including the creator, created data, name, and provenance), and allows for saving model artifacts in service-managed object storage, loading models into notebook sessions for testing, and deployment of the model to Oracle Functions.

Users can also pull the model from a catalog that a colleague has created, compare the model with one that has already been created, perform all operations on the model, and have all work on models and projects documented in a governed way.





# About Oracle Machine Learning

Oracle Machine Learning refers to the family of products used to analyze and prepare data, and build, evaluate, and deploy machine learning models in Oracle data management systems.

Oracle Machine Learning has had its roots in Oracle Database since the acquisition of Thinking Machines Corporation in 1999. Initially designed to help database users prepare data, and build and deploy machine learning models from data stored in Oracle databases using SQL, Oracle Machine Learning has expanded to include R APIs, as well as leverage new technologies such as big data, Oracle Autonomous Database, and other Oracle Database cloud services.

Oracle provides you with multiple ways to take advantage of the machine learning in Oracle Database. Oracle Machine Learning's in-database algorithms have been designed to take advantage of parallelism (multi-threading) and distributed execution (across multiple nodes). This provides scalability and performance benefits because model building and scoring can take advantage of multiple CPUs and compute nodes on large-scale hardware such as Oracle Exadata and cloud-based solutions such as Oracle Autonomous Database.

The Oracle family of products for in-database machine learning include:

**Oracle Machine Learning for SQL (OML4SQL)** supports Oracle Database and Oracle Autonomous Database where data scientists use SQL and PL/SQL code to build machine learning models and score data. You can use Oracle Machine Learning Notebooks on Autonomous Database, or Oracle SQL Developer or another SQL integrated development environment (IDE) to write SQL code. The SQL interface makes machine learning more accessible to users skilled in SQL.

**Oracle Machine Learning Notebooks** are based on Apache Zeppelin notebooks with Autonomous Database they provide a collaborative notebook interface for developing and executing SQL, PL/SQL, and OML4SQL code. Oracle Machine Learning Notebooks support data scientists, data analysts, application developers and DBAs in easily shared notebooks and templates. It delivers a collaborative framework that provides access

permissions, versioning, and execution scheduling. Notebooks are organized at two levels: workspaces that control sharing, and projects for finer grained organization.

**Oracle Machine Learning for R (OML4R)** supports Oracle Database, both in the cloud and on-premise, enabling scalable data exploration, data preparation, model building, and scoring using in-database algorithms—through a natural R interface with the user's R IDE of choice. R users can also deploy user-defined functions that leverage third-party R packages, which can be invoked via SQL for ease of deployment.

**Oracle Data Miner** is a no-code graphical tool for data exploration, preparation, model building and evaluation, data scoring, as well as solution deployment. Users construct analytical workflows in this drag-and-drop extension to Oracle SQL

Developer. This makes machine learning accessible to a broader set of users, including DBAs and other Oracle Database experts.

**Oracle Machine Learning for Spark (OML4Spark)** supports data science projects involving data in the data lake, leveraging the full Spark cluster. It provides an R interface for data exploration and preparation, as well as Spark-based machine learning algorithms—both scalable proprietary implementations and those tightly integrated from third-party libraries.



# How Oracle Simplifies Machine Learning

Machine learning can be complicated, but having the right tools makes the entire process simpler. Here are just a few examples of how Oracle works to continually improve the machine learning process.

## Managing the Entire Model Lifecycle

The process of building a machine learning model is an iterative one. Oracle Cloud Infrastructure Data Science and Oracle Machine Learning make it easier to manage models throughout their entire lifecycle.

## Building Models

Cloud Infrastructure Data Science's JupyterLab environment offers a variety of open-source libraries for building machine learning models. It also includes the Accelerated Data Science (ADS) SDK. Data scientists can automate model training through the ADS AutoML API. ADS also helps data scientists identify the best data transformations to apply to data sets.

Cloud Infrastructure Data Science also has AutoML automated data feature engineering, algorithm selection, and hyperparameter tuning which automates the process of training and selecting model candidates. This saves significant time for data scientists.

Oracle Machine Learning enables building models that reside in Oracle Database using SQL and R—key languages used by data scientists.

In-database algorithms provide automatic data preparation (ADP) to enable producing a quick, yet reasonable first-cut model from data. Automatic

data preparation (ADP) performs normalization, outlier treatment, missing value treatment, and binning as needed by individual algorithms. Any applied transformations are stored in the model itself so scoring is immediate.

Oracle Machine Learning supports partitioned models, which allow users to specify one or more columns of data on which to partition data. Then, it automatically builds one model per partition. This ensemble of models appears as a single model to the user when scoring data—greatly simplifying model use and automating an otherwise repetitive coding exercise.

Data Miner facilitates building multiple models through a convenient drag-and-drop user interface. OML4Spark enables model building directly in big data environments through an R interface using both native Oracle algorithms and those from Apache Spark's MLlib.

## Evaluating Models

Oracle Cloud Infrastructure Data Science's model evaluation generates a comprehensive suite of evaluation metrics and suitable visualizations to measure model performance against new data. It can also rank models over time to enable optimal behavior in production. Model evaluation goes beyond raw performance to take into account expected baseline behavior.

Oracle Machine Learning provides statistics for evaluating model quality, including a cost model to account for different impacts of false positives and

false negatives. Oracle Data Miner provides visual comparison of automatically generated in-database model quality statistics.

## Explaining Models

We often rely on the results of machine learning models. But being able to explain the internal mechanics of them in human terms, or deciphering why the model made the decisions that it did, is another matter entirely. It is not always possible to decipher the model and its decision—but it's becoming increasingly important.

Cloud Infrastructure Data Science offers a model explanation capability which provides automated explanation of the relative weighting and importance of the factors that go into generating

a prediction. With a fraud detection model, for example, a data scientist can explain which factors are the biggest drivers of fraud so the business can modify processes or implement safeguards.

Oracle Machine Learning's in-database models provide model-specific explanatory details on model predictions to understand why a given prediction was made. In OML4SQL, users can see the predictor variables and specific values for the record being scored. This is given a weight and ranking, with the weight indicating the relative importance of the variable to the specific prediction. Model details are also available for understanding model behavior.

Through Oracle Data Miner, model viewers are available for a variety of in-database models.

## Options for Oracle Data Science and Machine Learning Customers

### Access to Open-Source Technologies

Oracle has partnered with a number of open-source technologies to bring a wide variety of solutions to its data science ecosystem, including big data, analytics, and machine learning tools. These partnerships, based on broadly accepted industry standards, allow you to run your choice of machine learning tools, languages, notebooks, and frameworks as part of Oracle's high-performance, second-generation cloud infrastructure.

### Flexibility in Compute

Oracle offers a variety of compute shapes to suit every need. The compute is elastic—start small and scale out as your model training workload requires more data. Oracle provides compute infrastructure that is suitable for a wide range of applications and use cases.



Through Oracle Data Miner, model viewers are available for a variety of in-database models. Oracle Data Miner offers visual tree viewer for decision tree models, cluster tree and comparison viewer for hierarchical k-means, and coefficient ranking viewers for support vector machine models, among others.

### Deploying Models

Taking a trained machine learning model and getting it into the right systems is often a difficult and laborious process. Many otherwise successful data science projects fail due to an inability to place solutions into production either in a timely manner, or at all.

Cloud Infrastructure Data Science enables teams to operationalize models as scalable and secure APIs. With Cloud Infrastructure Data Science, the model can be deployed to Oracle Functions, which means that the model can be deployed as a Docker container. This container can be converted to a REST API endpoint through the Oracle Cloud Infrastructure API Gateway service. It can be consumed from anywhere—from the database, inside a notebook session, from an application inside or outside Oracle Cloud Infrastructure—even another cloud so organizations can start making use

of machine learning results faster, and start proving value to the enterprise.

With Oracle Machine Learning, in-database models

are immediately deployable through SQL. User-defined R functions can also be deployed through Oracle Database—invoked from R, SQL or REST APIs. This powerful capability places the spawning and management of R and Python engines under

the control of Oracle Database. Since developers often already use SQL for data access, deployment through Oracle Machine Learning provides an easy way to integrate data scientist results into applications and dashboards.





# Unique Capabilities of Oracle’s Data Science and Machine Learning Services

Oracle designed its data science and machine learning services to manage the full model lifecycle. However, each service was also created to serve different needs and each has distinctive features and capabilities to best suit different use cases.

## Oracle Cloud Infrastructure Data Science: Unique Capabilities

### Ensure a Collaborative Workflow

It’s difficult for data science to be truly successful unless there’s an emphasis on making those data science teams efficient and successful. Cloud Infrastructure Data Science has done everything to ensure data scientists can work in teams and to centralize all work so there’s no loss of knowledge.

Data scientists can work in “projects” where it’s easy to see what’s happening with a high-level view. Data scientists can share and reuse data science assets and test their colleagues’ models. Because they can also check their code in and out from a Git repository, save their models in the model catalog, and easily deploy them as Oracle Functions, this makes it easier than ever for them to work collaboratively with each other and with other external teams to deploy successful machine learning models.

### Use Your Favorite Open Source Tools

With Oracle Cloud Infrastructure Data Science, you can use the best of open source, including:

- Tools and languages like Python and JupyterLab
- Visualization libraries like Plotly, Matplotlib,

and seaborn

- Data manipulation tools like Dask, Pandas, and Numpy.
- Machine-learning libraries like TensorFlow, Keras, SciKit-Learn, and XGBoost
- Version control with Git

Oracle Cloud Infrastructure Data science was designed to take advantage of the benefits of open source technology. For example, users can utilize open source tools and utilities to:

- Train their models by adding a new library that includes the latest and greatest from the AI / machine learning research community
- Get started easily by integrating pre-built use cases created for time-series modeling, Bayesian modeling, deep learning, anomaly detection, supervised modeling, and unsupervised modeling.
- Use client libraries that connect to a variety of data sources and ingest data in a variety of formats
- Use Oracle Cloud Infrastructure’s tutorial content and example notebooks that show users how to connect to these different curated libraries for modeling in a variety of use cases.

### Be Data Source Agnostic

There are many advantages to doing machine learning within a database. But often, data scientists will want the ability to connect to data from many sources, and the majority of data doesn’t reside in databases. Instead it’s found in file systems, Hadoop, object storage, and more. Data scientists generally want access to more data, and more kinds of data.

Oracle Cloud Infrastructure Data Science is data-source agnostic. The service contains a variety of client libraries to connect to multiple data sources on Oracle Cloud Infrastructure but also on other clouds, including:

- Oracle Cloud Infrastructure Object Storage
- Oracle Autonomous Database
- Oracle Database
- AWS S3
- AWS Redshift
- Elasticsearch
- MongoDB
- And many more

This means that data scientists can draw from a range of data sources to build a rich and comprehensive data set to use for training machine learning models. Because Cloud Infrastructure Data

Science is data source agnostic, users can pull data not just from Oracle Cloud, but other clouds as well..

And Cloud Infrastructure Data Science isn’t just data source agnostic—it’s also data format agnostic. Data scientists can use all of the unstructured data that doesn’t fit nicely into rows and tables, and that includes images, PDFs, text, videos, sensor data, and more.

### Use the Power of Oracle Cloud Infrastructure

Cloud Infrastructure Data Science is Oracle Cloud Infrastructure OCI native—which means it can use the best of the newest, best cloud on the market.

Due to the inherent scalability of Oracle Cloud Infrastructure, data scientists can specify the precise amount of CPU and storage resources they need for each machine learning workload. Data scientists can pick ten CPUs or GPUs one day and twenty the next — and they can even change the number of CPUs or GPUs mid-project, thus allowing them to start small and grow as they need to for more managed costs.

Oracle’s broad range of compute options, fast networking speeds, scalable independent storage, modern architecture, and competitive pricing make it an ideal infrastructure platform for machine learning and AI.





**Provide Security Through Oracle Cloud Infrastructure**

Any enterprise using the Cloud Infrastructure Data Science service will benefit from the OCI security-first approach. IT admins can define the boundaries (via compartments, user groups and policies) inside which data scientists can operate. This gives administrators flexibility and allows, for example, to let multiple users to open the same notebook session or limit access to only the user who created the notebook session. It allows administrators to segregate groups of data scientists (e.g. consultants vs. data scientists vs. interns) and limit their respective access to compute,

data sources, networking, projects, notebook sessions and models. This is all part of a truly enterprise-class offering.

**Economics**

With a simplified pricing model, low pricing on compute and storage, and straight percentage discounts across services, Cloud Infrastructure is a powerful but affordable option. Oracle Cloud Infrastructure Data Science in particular is a zero-premium service, meaning users only pay for the compute and storage they use.

“Employing data scientists or using augmented intelligence isn’t enough to create a successful AI deployment. AI requires a modern data infrastructure to support new data types and often massive amounts of data. Many organizations are moving to the cloud for data management.”

**Fern Halper**  
Ph. D., The Data Warehouse Institute



**Oracle Machine Learning: Unique Capabilities**

**Enable Data Science Contributors Across Roles**

Data science is a team sport, requiring a group of individuals with different skills to collaborate to solve complex business problems. Oracle Machine Learning supports a range of user roles, including: data scientists, data and business analysts, DBAs and IT professionals, and application and dashboard developers.

The Oracle Data Miner user interface empowers both experts and non-experts to use in-database machine learning through a drag-and-drop interface.

**Designed for Performance and Scaling**

With Autonomous Database, users can take advantage of auto scaling with Oracle Machine Learning. With auto scaling enabled, the database can use up to three times more CPU and IO resources than specified by the number of OCPUs explicitly allocated. When auto scaling is enabled, if your workload requires additional CPU and IO resources the database automatically uses needed resources without any manual intervention required.

Whether on-premises, Oracle Database Cloud Service, or Autonomous Database, users benefit from the underlying Exadata architecture. Oracle Machine Learning’s algorithms are parallel and distributed to leverage multi-node and high CPU/RAM environments. Even scoring data with in-database models takes advantage of storage-level “smart scan” technology for even faster scoring times.

**Build and Execute Machine Learning Within Oracle Database**

Is your organization making full use of the data stored in its Oracle databases and data warehouses? Oracle Machine Learning was designed to satisfy a fundamental need: to maximize the value of your data by making it easy for data science teams to take advantage of machine learning technology, scale to handle large volume data, and deploy data science solutions quickly and easily.

The Oracle Machine Learning platform includes mature technologies to profile, explore, and prepare data, and build, evaluate, and deploy machine learning models within Oracle Database. This opens up machine learning capabilities to application developers, database administrators, DevOps professionals, analysts, and many other people who already work with Oracle Database—enabling them to build on their existing skills and extract value from these corporate assets.

**Minimize Data Movement with In-Database Processing**

According to Forrester, big data is critical to many aspects of AI and machine learning. However, getting the right data in place is a complex process, and some business leaders complain that the amount of time needed to prepare data for AI and machine learning is so great that it’s not worth the effort.<sup>3</sup>

<sup>3</sup> Forrester, PREDICTIONS 2020: On the Precipice of Far (<https://go.forrester.com/predictions/>)





You don't have to move data out of your secure Oracle environment to create machine learning models. Instead, Oracle has brought the algorithms to where the data reside. Oracle allows you to store many types of data in their native form and to apply machine learning algorithms directly to the data, without having to transform it into another format or move it to another analytical engine. These in-database machine learning capabilities include more than 30 algorithms optimized for supporting machine learning projects on data stored right in Oracle Autonomous Data Warehouse and Oracle Database.

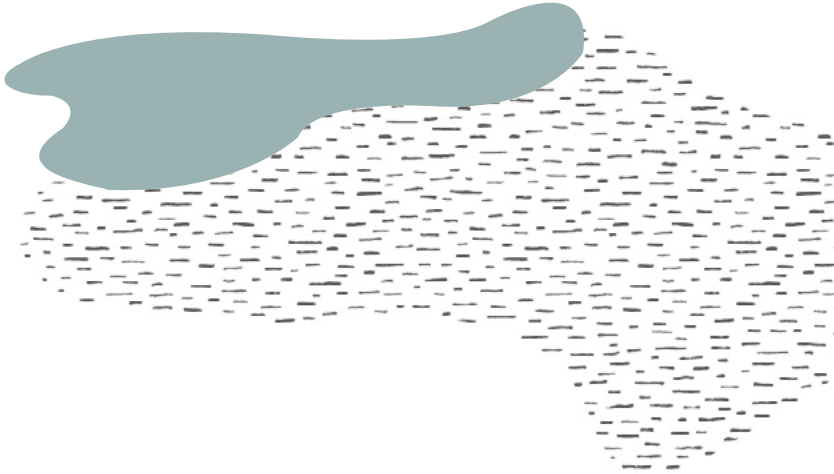
According to TWDI, in-database processing simplifies data integration and speeds up the process of creating, scoring, and updating machine learning models.<sup>4</sup>

**Provide Security through Oracle Database**

Oracle Database enables enterprises to safeguard their most valuable data with Oracle's powerful prevention and detection security controls, including transparent data encryption, encryption key management, privileged user and multifactor access control, data classification and discovery, database activity monitoring and blocking, consolidated auditing and reporting, and data masking. The use of these capabilities is readily available in combination with Oracle Machine Learning. In addition, in-database models such as first-class objects support role-based access to define, manipulate, or apply machine learning to a model.

**Economics**

Oracle Machine Learning in-database features are included in Oracle Autonomous Database—at no extra cost. This includes OML4SQL and Oracle Machine Learning Notebooks.



<sup>4</sup> Russom, Philip, CHECKLIST REPORT: “Data Management Requirements for Data Science, ML, and AI” (January 2020).



# More Machine Learning Tools in Oracle's Data Science Platform

But of course, machine learning with Oracle extends beyond individual services. Oracle offers a wide range of options to suit organizations and their machine learning needs.

## Create Your Own Machine Learning Setup Through Oracle Cloud Infrastructure

Oracle provides additional solutions through Oracle Cloud Infrastructure, which provides industry-leading compute, network, storage, and database services that have been optimized to run customized computing environments for AI and Machine Learning.

From Oracle Autonomous Database and Oracle Cloud Infrastructure Virtual Machines for Data Science to Oracle Cloud Infrastructure Data Catalog and Oracle Cloud Infrastructure Data Flow, Oracle's services help users enhance and expand upon what they can accomplish with data science. In addition, many of the world's most popular data science solutions are available through the Oracle Cloud Marketplace. The Oracle Quick Start program simplifies deployment and accelerates time-to-productivity on Oracle Cloud Infrastructure, in conjunction with Oracle Kubernetes Engine. Available on GitHub, this collection of Infrastructure-as-Code (IaC) solutions allows customers to easily use partner products.

## Other Important Data Science Platform Services

Cloud Infrastructure Data Science and Oracle Machine Learning are the primary components of our data science platform. But it doesn't just stop there. Oracle has built a comprehensive set of products and capabilities to help organizations achieve their data science goals:

- **Oracle Cloud Infrastructure Data Catalog:** Allows users to discover, find, organize, enrich and trace data assets on Oracle Cloud. Cloud Infrastructure Data Catalog has a built-in business glossary making it easy to curate and discover the right, trusted data.
- **Oracle Big Data Service:** Offers a full Cloudera Hadoop implementation, with dramatically simpler management than other Hadoop offerings, including just one click to make a cluster highly available and to implement security. Big Data Service also includes Oracle Machine Learning for Spark, which allows organizations machine learning in memory with one product and with minimal data movement.
- **Oracle Cloud SQL:** Enables SQL queries on

data in HDFS, Hive, Kafka, NoSQL and Object Storage. Only Oracle CloudSQL enables any user, application, or analytics tool that can talk to Oracle databases to transparently work with data in other data stores, with the benefit of push-down, scale-out processing to minimize data movement.

- **Oracle Cloud Infrastructure Data Flow:** A fully-managed big data service that allows users to run Apache Spark applications with no infrastructure to deploy or manage. It enables enterprises to deliver big data and AI applications faster. Unlike competing Hadoop and Spark services, Cloud Infrastructure Data Flow includes a single window to track all Spark jobs, making it simple to identify expensive tasks and troubleshoot problems.
- **Oracle Cloud Infrastructure Virtual Machines for Data Science:** Preconfigured GPU-based environments with common IDEs, notebooks, and frameworks that can be up and running in under 15 minutes, for \$30 a day.

## Leverage Graph Analytics

Graph analytics, included in Oracle Database and Oracle Big Data Service, augment machine learning systems in the context of social networks, transportation networks, supplier dependencies, financial relationships, and other forms of connected data. The graphs represent an alternative type of model for identifying hidden features or detecting patterns.

Metrics computed from a graph can be added to a predictive model to increase accuracy. Furthermore, the results of a model can be added to a graph for further exploration or additional analysis. Oracle's powerful in-memory analytics engine includes almost 60 graph algorithms to perform ranking, centrality, recommendation, community detection, path finding, and more. It also includes a pattern matching query language.

Running graph analytics against your data can bring a powerful new perspective to machine learning. That's because graphs are all about relationships, and if those relationships are already defined in the data, it can enable machine learning algorithms to discover more connections and better insights.



# A Complete Data Platform for Machine Learning

To make full use of today's machine learning technologies, enterprises need a centralized platform that includes the best data science tools, frameworks, and infrastructure. The Oracle data science platform will transform the way your organization approaches data science projects by helping you extract more value from your data.

**To learn more about Oracle's comprehensive platform for machine learning, visit [oracle.com/data-science](https://oracle.com/data-science).**





## Cloud Essentials

Learn more about **Oracle Gen 2 Cloud**.  
Build, test, and deploy applications on Oracle  
Cloud—for free. Sign up once, get access to two  
free offers. Go to  
**[oracle.com/cloud/free](https://oracle.com/cloud/free)**

---

Copyright © 2020, Oracle and/or its affiliates. All rights reserved. Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.  
VDL51082 191204



**ORACLE**  
Cloud

