# **Databricks Unified Data Analytics** significant descriptions described by the description of the description



# All your data, analytics and AI on one platform

# Why Databricks?

- Single, top ranked tool for dev to production collaboration between data engineering and data science at scale.
- Automated performance management eliminates the need for customizing and aggregating from data lake.
- Interactive notebooks support multiple languages, visualizations, model building, training, and libraries.
- Schema enforcement ensures consistent data and only completed writes are committed. Time travel maintains versions of data.

#### **Additional Resources**

- Databricks free trial
- eBooks and webinars
- **Databricks Academy**

### **Product overview**

Databricks Unified Data Analytics Platform is a cloud-based service for running your analytics in one place—from performant data pipelines to state-of-the-art machine learning (ML). From the creators of Apache Spark and MLflow, it provides data science and engineering teams ready-to-use clusters optimized by Apache Spark and ML. Databricks helps companies use data to make better products and services, easily, and more cost efficiently.

# **Product features**

# **Automated infrastructure management**

A fully managed cloud platform on Amazon Web Services (AWS) simplifies operations and delivers superior performance of data pipelines at scale. Simplified cluster management with auto-scaling significantly reduces time spent on data engineering and development.

## **Easier data engineering**

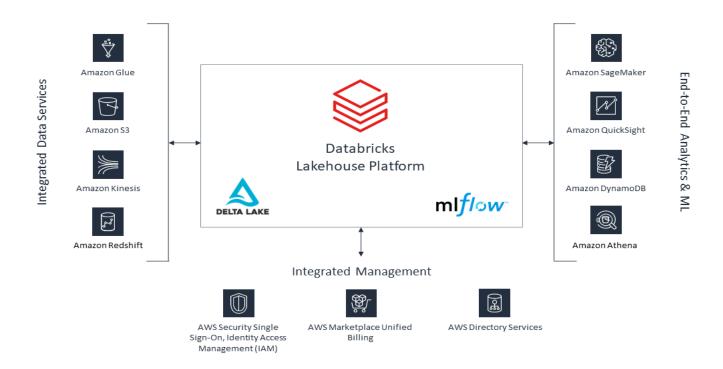
Comprehensive audit logs and GitHub integration make implementation, debugging, testing, and code reviewing process easier, faster, and more interactive.

# **Robust machine learning**

MLflow greatly streamlines the machine learning lifecycle, simplifies model versioning, and enables reproducibility and process repeatability. These models can then be served through Amazon SageMaker providing a fully integrated ML platform solution with the security and integrated management services of AWS.

#### How it works

The Lakehouse Platform combines the data management and performance typically found in data warehouses with the low-cost, flexible object stores offered by data lakes. This unified platform simplifies your data architecture by eliminating the data silos that traditionally separate analytics, data science, and machine learning. It's built on open source and open standards to maximize flexibility. Native collaborative capabilities also accelerate your ability to work across teams and innovate faster.



#### **Differentiators**

- Reliable and performant infrastructure with ready-to-use cluster and enterprise-grade security that simplifies DevOps through autoconfig and auto-scaling capabilities.
- Reliable data pipelines can be built at scale and scheduled interactively or with the Databricks production job scheduler.
- Outstanding collaboration across the data and ML lifecycle powered by notebooks (supporting SQL, R, Python, Scala, and Java). Extensive ecosystem integrations including a plethora of data connectors and ML frameworks (TensorFlow, Keras, XGBoost, and scikit-learn).

### **Data Points**

