



Azure Databricks

A fast, unified, and collaborative Apache Spark-based data and AI platform optimized for the Microsoft Azure cloud.

<https://azure.microsoft.com/en-us/products/databricks>

Overview

Azure Databricks is a unified, cloud-based data analytics platform jointly developed by Microsoft Azure and Databricks, built on the open-source Apache Spark engine and the Lakehouse architecture. It provides a secure, scalable, and collaborative environment for data engineering, data science, machine learning, and business intelligence (BI) workloads. The platform simplifies large-scale data processing and workflow integration through managed cloud infrastructure and seamless integration with Azure services.

Key capabilities include high-performance ETL/ELT pipelines, real-time streaming analytics, and the development/deployment of production-ready AI and Generative AI models. It supports multiple languages (SQL, Python, R, Scala) within its collaborative notebooks and offers managed services like Delta Lake for data reliability and Unity Catalog for centralized governance and security. It is an excellent tool for data processing and analysis, particularly for organizations moving large-scale data and legacy Hadoop workloads to the cloud.

Target users are data engineers, data scientists, and business intelligence analysts in mid-market and enterprise companies. While users praise its scalability, advanced features, and seamless Azure integration, they often note a steep learning curve and the necessity for careful cost monitoring due to its usage-based (DBU and VM) pricing model.

Key Features

- Lakehouse Architecture (Unified Data, Analytics, AI)
- High-Performance ETL/ELT Pipelines (Delta Live Tables)
- Collaborative Notebooks (SQL, Python, R, Scala)

- Managed Apache Spark Compute (Autoscaling, Serverless)
- Unity Catalog (Unified Governance and Security)
- MLflow for Machine Learning Lifecycle Management
- Generative AI Model Building and Deployment
- Integrated SQL Analytics and BI

Pricing

Model: usage_based

Pay-As-You-Go based on Databricks Units (DBUs) consumed per second and Azure Virtual Machine (VM) costs. Tiers (Standard, Premium) and Committed Use Contracts (DBCUs) are available for discounts.

Starting at: USD \$0.07

Target Company Size: medium, enterprise

Integrations

Microsoft Azure Data Lake Storage (ADLS), Azure Data Factory (ADF), Microsoft Power BI, Microsoft Fabric, Microsoft Entra ID (Identity), Microsoft Purview (Governance), GitHub/GitLab (Version Control), RDBMS (General)

Compliance & Certifications

HIPAA, HITRUST CSF, FedRAMP High, PCI-DSS, GDPR, CCPA

This document was generated by IntuitionLabs.ai with the assistance of AI. While we strive for accuracy, please verify critical information independently.