aws marketplace

DATA AND ANALYTICS SOLUTION SPOTLIGHT

Efficiently deploy data and AI workloads while reducing costs

Optimize costs with Databricks on AWS

Enterprise data is often siloed and disconnected, and many data solutions can be technically complex. Together, these factors can hamper productivity and increase operational costs.

The Databricks Data Intelligence Platform allows you to build, deploy, share, and maintain enterprise-grade data, analytics, and AI solutions at scale. The platform integrates with your cloud storage and security to automatically manage and deploy cloud infrastructure.

Using Databricks on Amazon Web Services (AWS), organizations can decrease operational costs and total cost of ownership (TCO), simplify infrastructure management, and reduce data silos.

Exploring core principles behind cost optimization

Databricks and AWS can help you meet the goals outlined below.

Choose the correct resources

Selecting the right compute size for your workload can be a time-consuming process. AWS Compute Optimizer can help you simplify your rightsizing journey. The service applies insights from millions of workloads to make recommendations on optimal resources for your workload. AWS Compute Optimizer allows you to continuously scan your resource usage and match your workload to optimal resources.

Amazon Elastic Compute Cloud (Amazon EC2) also provides different instance types to enable you to choose the CPU, memory, storage, and networking capacity you need for your workload.

Databricks provides guidelines on choosing the correct compute instance based on your workload type. For example, if you have ML workloads, you can use memory optimized VM. These guidelines can help you better manage workload costs.

In collaboration with:

databricks

Reduce workload costs with Databricks

The Databricks platform allows you to better manage the costs of data and AI workloads by focusing on three areas:

- **Track global usage:** Monitor and interpret usage across regions, workspaces, and accounts. Identify anti-patterns and outliers and take corrective action.
- **Measure value:** Track return on investments, ensuring that projects deliver more value than they spend. Justify spending for early-stage developments and deployments.
- **Improve governance:** Establish monitoring and enforce behaviors across diverse requirements. Maximize efforts and articulate the value that can be brought by referencing a Center of Excellence—a delivery framework and methodology that helps solve data and AI problems at scale.

Dynamically allocate and de-allocate resources

Databricks can help you allocate and release resources to match your performance requirements. You can save infrastructure costs while maintaining end-to-end latency SLAs for streaming workloads with Databricks on AWS.

Delta Live Tables is a declarative extract, transform, load (ETL) framework for the Databricks Data Intelligence Platform that helps data teams cost-effectively simplify streaming and batch ETL. Delta Live Tables shuts down nodes when utilization is low while guaranteeing task execution, and only scales up to the needed number of nodes.

Also, you can use serverless compute from Databricks to focus on creating business value instead of managing infrastructure. Databricks can provide idle time removal, performance optimization, and infrastructure management, which can help you save costs.

Monitor and control cost

Databricks can help you improve your cost management practices with system tables, a Databricks-hosted analytical store of your account's operational data. You can use system tables for granular cost reporting, forecasting, and controlling your cost.

Shell

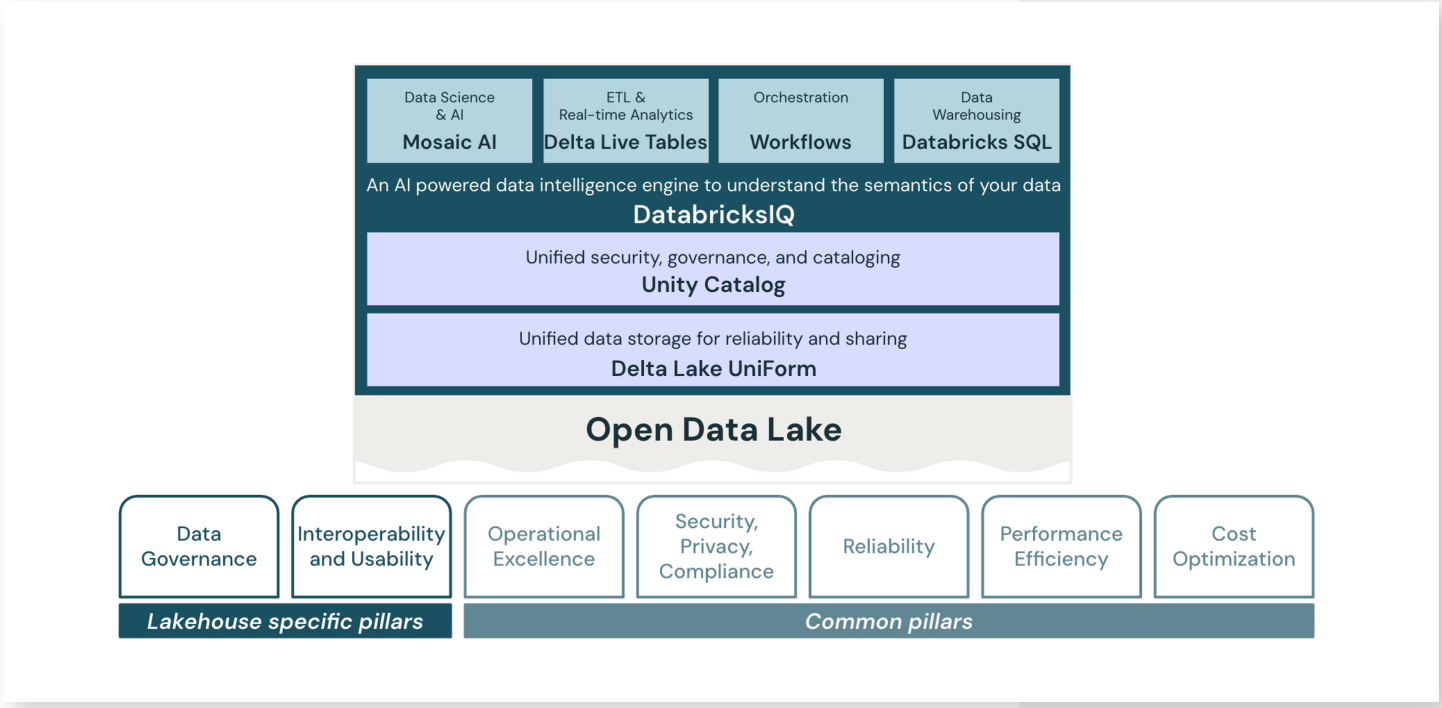
"Delta Lake Liquid Clustering improved our time series queries up to 10x and was remarkably simple to implement on our Lakehouse."

Bryce Bartmann, Chief Digital Technology Advisor, Shell

YipitData

"Using Databricks innovative Liquid Clustering, we have observed remarkable improvements in query performance compared to the traditional z-order methods."

Edward Goo, Director of ETL Engineering, YipitData



AWS featured products

Amazon Redshift is a fully managed, petabyte-scale data warehouse service in the cloud.

Amazon Bedrock is a fully managed service that allows you to build and scale generative AI applications with foundation models.

AWS Glue is a serverless data integration service that makes it seamless for analytics users to discover, prepare, move, and integrate data from multiple sources.

Amazon QuickSight is a business intelligence (BI) service that you can use to deliver easy-to-understand insights to the people who you work with, wherever they are.

Amazon Kinesis is a fully managed service that ingests, buffers, and processes streaming data in real time.

Amazon Athena is a serverless, interactive analytics service that makes it seamless to analyze data directly in Amazon Simple Storage Service (Amazon S3) using standard SQL.

Learn more ›

See how you can better monitor usage to control workload costs.
Purchase or start a free trial of Databricks on AWS today.

About Databricks

Databricks is a data and AI company. More than 10,000 organizations worldwide — including Comcast, Condé Nast, Grammarly, and over 60 percent of the Fortune 500 — rely on the Databricks Data Intelligence Platform to unify and democratize data, analytics, and AI. Databricks is headquartered in San Francisco, with offices around the globe, and was founded by the original creators of Lakehouse, Apache Spark™, Delta Lake, and MLflow.