Databricks Hadoop Migration

databricks

Shift from Hadoop to the Databricks Unified Data Analytics Platform

Why Databricks Hadoop Migration?

- Focus on business not infrastructure.
- Save cost from elastic scaling.
- Improve productivity.
- Separate storage and compute.
- Faster time to market.
- Accelerate your innovation.
- No lock-in to aging hardware.
- No more upgrade hassles.

Additional Resources

- Hadoop migration whitepaper
- Schedule a demo

Product overview

The Databricks Hadoop Migration Package gives you a concrete migration outcome through a prescriptive approach that maximizes the value of your existing data and pipeline investments. The Databricks Unified Data Analytics Platform enables your data and artificial intelligence (AI) team to holistically analyze data from data warehouses, data lakes and other data stores. It does this by utilizing a single engine for batch, ML, streaming and real-time queries. By migrating to the Unified Data Analytics Platform, your teams get access to data in one place and the ability to automate complex data pipelines. This enables you to run and maintain machine learning at scale — so you can innovate faster.

Product features

Discovery, planning, and roadmap

- Gather all necessary background on the current Hadoop environment.
- Build an inventory.
- Develop a phased migration plan based on data and pipeline priority, dependency and project criticality.

Target delta architecture

Develop a phased migration plan based on data and pipeline priority, dependency, and project criticality.

Migration

- Data migration and cluster migration planning.
- Set up Databricks environment and metastore migration.
- Migrate code/pipeline and data validation.
- Continuous integration and continuous development (CI/CD) and automation readiness.

How it works

Over the past year, we have seen an acceleration in customers migrating from a Hadoop architecture to a modern cloud architecture. Many organizations have made the move to reduce the operational costs of licenses and maintenance, but they've also discovered that the power of a modern cloud-based analytics platform quickly outweighs the cost of migration.
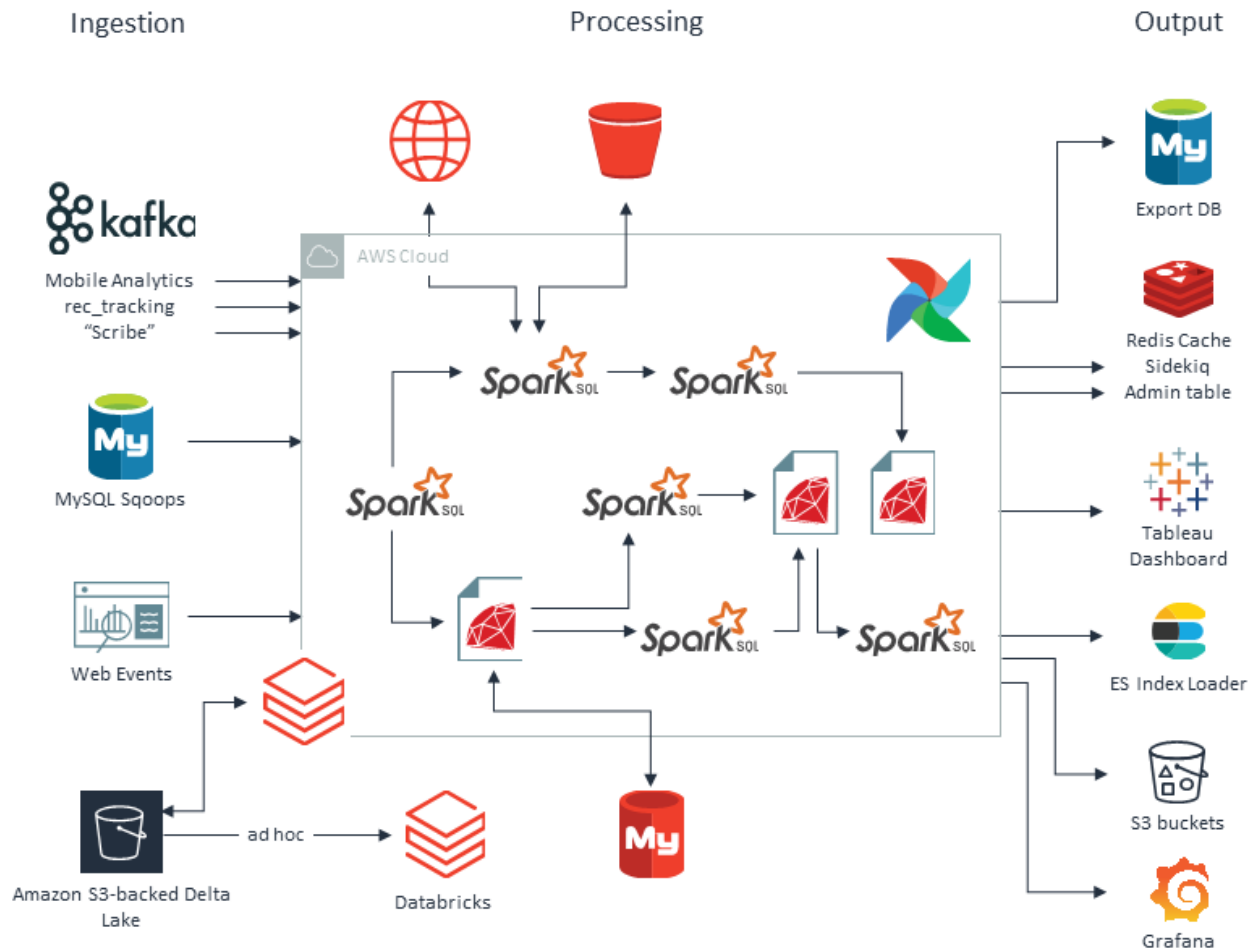
Differentiators

- Data migration; code, extract, transform, load (ETL) pipeline; and job orchestration.
- Pilot implementation for ingestion and integration.
- Productionization hardening best practice.
- Code and performance optimization.

Data Points

95%
Databricks customers who meet objectives and timelines

417%
ROI for companies switching to Databricks

47%
Cost-savings from retiring legacy infrastructure

Solution available in AWS Marketplace