CONFLUENT + DATABRICKS

Modernize Your Data Platform with Databricks and Confluent Cloud

Confluent and Databricks form a powerful and complete data platform focused on helping companies scale real-time analytics and applications. Connecting and migrating real time data from multiple on prem and cloud data sources to Databricks using Confluent Cloud's fully managed Databricks Delta Lake connectors helps speed up moving data to the lakehouse and reduce the TCO associated with migrating to Databricks. That will accelerate development for real-time analytics and predictive applications like fraud detection, customer churn, or preventative maintenance.

Scalability and performance
Confluent Cloud scales and reliably delivers trillions of events per day that can be analyzed at petabyte scale by Databricks for data teams across ML/AI and BI use cases

Fast, efficient, reliable data
App developers, data analysts, and data scientists enjoy increased efficiency, productivity and data availability while lowering costs because everyone can work from the same data store without copying data

Business transformation
Digital transformation with Confluent and Databricks breaks down silos, enabling access to data across the organization, team collaboration and enabling real time application development

Why Databricks and Confluent Cloud?

Cloud has drastically changed the data analytics space as new technology has decoupled storage from compute to power new analytics ranging from traditional BI to machine learning aligning infrastructure costs to actual usage.

As organizations migrate their analytics data from existing on-prem data storage and analytics platforms (Teradata, Cloudera, Oracle, etc.) they are discovering the benefits of landing analytics data in cloud-based data lakehouses like Delta Lake on Databricks.

As these organizations rethink how to move their data platforms to the cloud, they have the opportunity to set data in motion to power real-time event streaming and ETL pipelines that connect data from any environment (cloud or on-prem).

Features

Connect to any app or data and analyze data that was once siloed

Access a library of 120+ connectors like Oracle, Teradata, SAP, and more to put your data in motion to insert fresh, real time data into your AI workloads in Databricks Lakehouse Platform.

Process data in milliseconds, not minutes

Build materialized views, aggregations (e.g. windowing), and large scale table joins with ksqlDB or Databricks to reduce processing times from minutes to milliseconds.

Unite hybrid and multi-cloud environments

Link multiple Confluent clusters together to accelerate migrations to Databricks using cluster linking. This allows you to access real-time, distributed data in Databricks no matter where the data resides.

Solution

Reduce the TCO of hybrid and multi-cloud data pipelines to Databricks and provide you organization the agility to innovate faster on a lakehouse

Confluent helps reduce your TCO associated with managing hybrid and mult-icloud data pipelines with its fully managed cloud service for Apache Kafka, Confluent Cloud. Confluent, with its cluster linking and data replication features, allows you to move resource intensive ETL jobs out of your on-prem data platform directly into Databricks. Organizations can become more productive with fewer resources and focus on innovation and developing analytics that provide new business value.

Confluent allows you to connect your cloud with unified network and security infrastructure and quickly move more data to Databricks without having to worry about hybrid or multicloud networking challenges.

Power new analytics and apps with Confluent and Databricks at cloud scale

Confluent allows you to operate at enterprise scale with our distributed real-time event streaming and processing capabilities, built on Apache Kafka. Depended upon by over 70% of the Fortune 500 today, Apache Kafka® has become the industry standard for real-time event streaming. It serves as an open foundation, and isn't locked to any one cloud vendor.
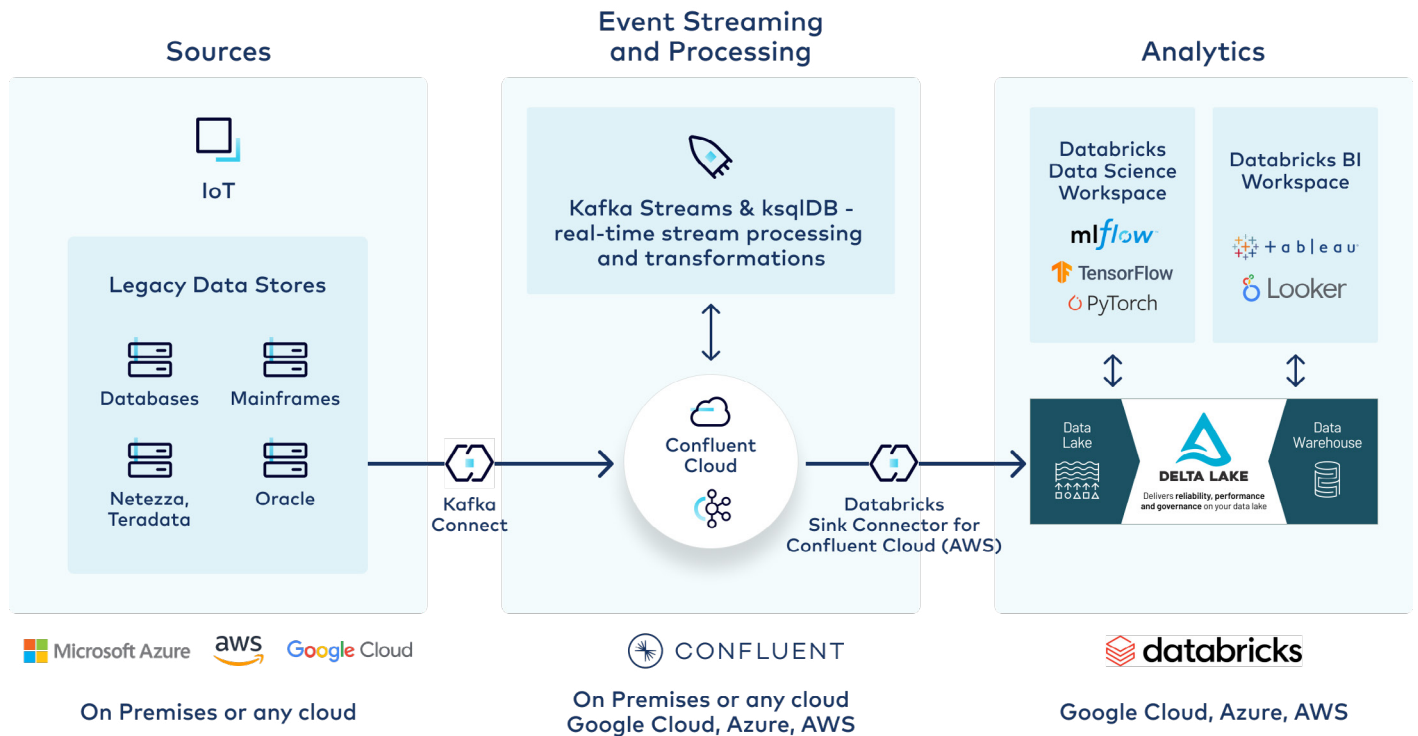
Pairing Confluent with Databricks allows enterprises to use Databricks machine learning and SQL capabilities to build real-time analytics. Migrating and connecting more data to Databricks with Confluent's fully managed service makes it easier to bring new data and analytics online faster with fewer resources.

Get more data to Databricks by connecting multicloud and hybrid

Confluent allows you to quickly connect new data sources from your on-prem data platform to Databricks with our rich ecosystem of 120+ connectors. These sources can reside on-prem or within multiple clouds.

Migrating your data platform or data warehouse to Databricks to unlock machine learning and advanced analytics doesn't need to be a multiyear lift and shift. With Confluent, enterprises can use our fully managed or verified connectors to quickly migrate data off of on-prem systems and into Databricks in real-time. With data delivered in real time, enterprises can use Databricks to deliver up to date analytics and machine learning models for better decision making and action them in apps for superior customer experiences.

Reference Architecture

Sources

IoT

Legacy Data Stores

Databases   Mainframes

Netezza,   Oracle
Teradata

Kafka
Connect

Microsoft Azure   aws   Google Cloud

On Premises or any cloud

**Event Streaming
and Processing**

Kafka Streams & ksqlDB -
real-time stream processing
and transformations

Confluent
Cloud

Databricks
Sink Connector for
Confluent Cloud (AWS)

CONFLUENT

**On Premises or any cloud
Google Cloud, Azure, AWS**

Analytics

Databricks
Data Science
Workspace

ml*flow*

TensorFlow

PyTorch

Databricks BI
Workspace

+ableau

Looker

Data
Lake

DELTA LAKE
Delivers **reliability, performance
and governance** on your data lake

Data
Warehouse

databricks

Google Cloud, Azure, AWS

Additional Information

Easily Migrate Data to the Cloud with Confluent and Databricks

Configure and Launch the Databricks Delta Lake Sink Connector

**How to Consume Data from Apache Kafka Topics and Schema Registry
with Confluent and Azure Databrick**