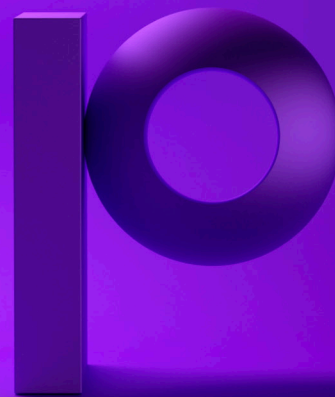




© 2020 Databricks, Inc. All rights reserved.



precisely

# Databricks and Precisely Connect

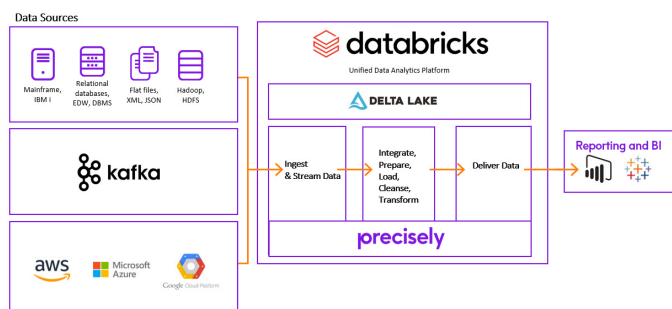
Breaking down data silos by integrating legacy, mainframe and IBM i data into Databricks Unified Data Analytics platform for cloud-based AI and ML projects

The richest data in your organization often lives on siloed mainframe, IBM i systems, Hadoop, Oracle, Db2 and others. Data silos, combined with growing data volumes and incompatible data formats, are all roadblocks to using data for innovative projects. Do not let your analytics, AI, and ML projects stall or fail because of these barriers.

Together, Databricks and Precisely help you to facilitate high-performance analytics, AI, and machine learning regardless of where enterprise data lives. Find success by integrating data sources to deliver innovation to your business without the need for specialized skills.

## Pipelines for Legacy Data to the Cloud

Liberate data from legacy sources for use within Databricks Unified Data Analytics Platform and Delta Lake by building data pipelines with Connect. Connect offers a design once, deploy anywhere approach to ETL workflows, and the interface allows you to define both batch and streaming ETL workflows from the same view. This visual approach to data integration means there is no reformatting or code generation needed to construct high-performance data pipelines. With point-and-click transformations, you can free time and attention to focus on your rapidly expanding and evolving technology stack.



## Build a Data Lakehouse

Connect helps you to build a data lakehouse by efficiently offloading data from legacy data stores to Databricks Unified Analytics Platform. Onboard data from almost any source, including:

- **Mainframe data:** VSAM, COBOL Copybooks, mainframe fixed and sequential files
- **RDBMS:** Oracle, SQL, Db2, MySQL, Sybase, PostgreSQL
- **Semi-structured data:** JSON, XML
- **Enterprise data warehouses:** Teradata, IBM Netezza, Vertica, Greenplum
- **Cloud:** Amazon AWS, Microsoft Azure, Google Cloud Platform
- **Big Data:** Hadoop, Hive
- **Streaming platforms:** Apache Kafka
- **Flat files:** Fixed length, variable length, delimited

Connect takes an end-to-end managed approach to offloading data. Regardless of which source you choose, you can replicate hundreds or thousands of tables – including whole database schemas – into Databricks quickly and easily.

## Scale for Machine Learning Projects

Your machine learning projects require collection and native integration of complex legacy data stores at scale. Connect collects the data you need from all your legacy data stores and sends it to Databricks, which provides a scalable framework for machine learning. Connect has native integration with the Databricks Runtime. Directly in your Databricks cluster, use Connect's intelligent engine to cleanse, transform, and prepare petabyte-scale datasets for analytics within your tight SLAs.

