



Azure Data Factory vs Databricks: 4 Critical Key Differences

Amit Kulkarni • Last Modified: October 13th, 2023

In the world of unstructured information, raw data does not have a proper context to provide valuable insights. Unorganized data often stored in **Relational, Non-Relational**, and other storage systems require a service that can orchestrate processes to refine information into actionable business insights. **Azure Data Factory (ADF)** and **Databricks** are two such Cloud services that handle these complex and unorganized data with **Extract-Transform-Load (ETL)** and **Data Integration** processes to facilitate a better foundation for analysis. While **ADF** is used for **Data Integration Services** to **monitor data movements** from various sources at scale, **Databricks** simplifies Data Architecture by **unifying Data, Analytics, and AI workloads in a single platform**.

This article describes the key differences between Azure Data Factory and Databricks. It briefly explains Azure Data Factory and Databricks along with its benefits to gain ideas for the underlying differences relatively.

Read along to find out in-depth information about **Azure Data Factory vs Databricks**.

Table of Contents

- [Prerequisites](#)
- [What is Azure?](#)
- [What is Azure Data Factory?](#)
 - [Key Benefits of Azure Data Factory](#)
- [What is Databricks?](#)

- Key Benefits of Databricks
- Azure Data Factory vs Databricks: Key Differences
 - Azure Data Factory vs Databricks: Purpose
 - Azure Data Factory vs Databricks: Ease of Usage
 - Azure Data Factory vs Databricks: Flexibility in Coding
 - Azure Data Factory vs Databricks: Data Processing
- Conclusion

Prerequisites

- Understanding of Big Data.
- An idea of Big Data Analytics.

What is Azure?

Microsoft Azure is a public **Cloud Computing** Microsoft platform that provides Software as a Service (**SaaS**), Platform as a Service (**PaaS**), and Infrastructure as a Service (**IaaS**) for a wide range of information technology tasks. These services provide **analytics, virtual computing, storage, and networking solutions** using various programming languages, tools, and frameworks that offer language extensibility. **Microsoft Azure** also offers a wide range of intelligent solutions for Data Warehousing, Advanced Analytics on Big Data to turn raw data into actionable insights.

What is Azure Data Factory?

Azure Data Factory (ADF) is a **Cloud-based PaaS** offered by the Azure platform for integrating different data sources. Since it comes with pre-built connectors, it provides a perfect solution for hybrid **Extract-Transform-Load (ETL)**, **Extract-Load-Transform (ELT)**, and other Data Integration pipelines.

Typically, an ETL tool **Extracts** data from various sources, **Transforms** collected data for intended analytical use cases and **Loads** it into a destination that can be a Database or Data Warehouse. ADF provides a **code-free ETL tool** on the

Cloud for users to quickly perform complex ETL processes. It helps users define a dataset, create Data Pipelines to transform data and map them with various destinations. Below are some essential components of ADF:

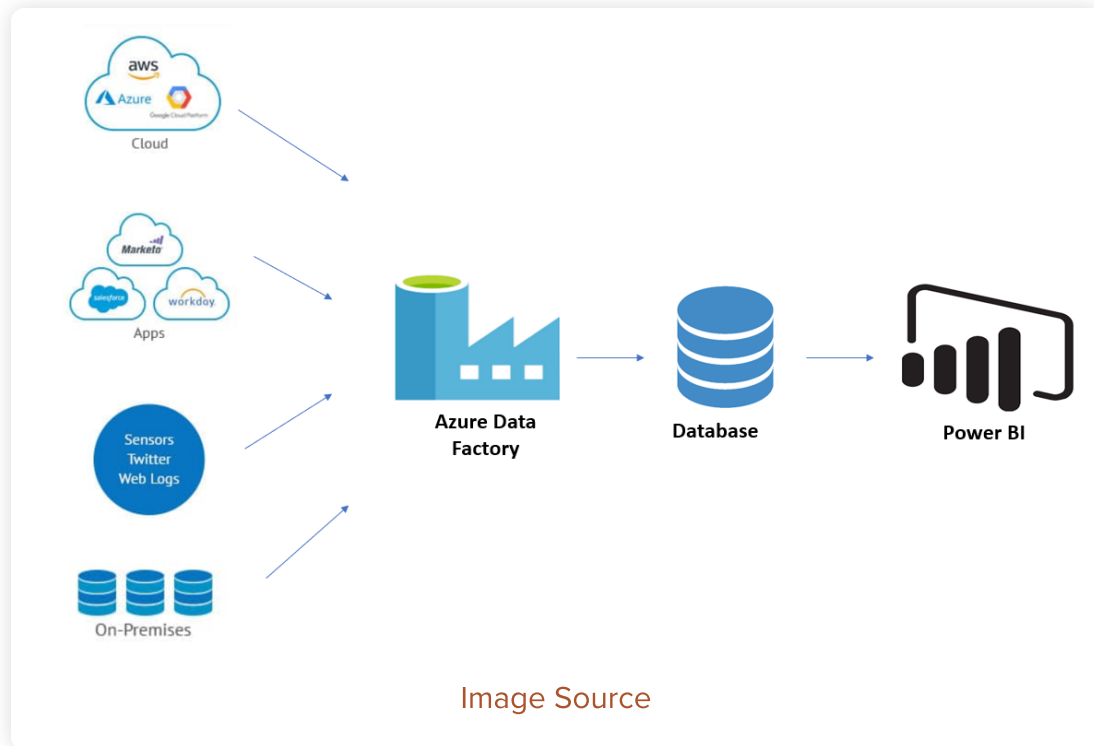
- **Pipeline:** It is a logical group of activities built to perform a unit of work. A single Pipeline performs different actions like ingesting data from either blob storage or querying the SQL database.
- **Activities:** It represents a unit of work in a Pipeline. It includes activities that copy blob data to a storage table or transform JSON data in a storage blob into SQL Table records.
- **Datasets:** It represents Data Structures within the Data Stores. Datasets point to data that 'activities' need to use as inputs or outputs.
- **Triggers:** It is a way to run execution in a pipeline. Triggers determine when a Pipeline execution should begin. Presently, ADF supports three types of triggers:
 - **Schedule Trigger:** It is a trigger that invokes a pipeline at a scheduled time.
 - **Tumbling Window Trigger:** It is a trigger that operates on a periodic interval.
 - **An Event-based Trigger:** It is a trigger that invokes a pipeline during a specific event.
- **Integration Runtime (IR):** It is the computing infrastructure that provides **Data Integration** capabilities like Data Flow, Data Movement, Activity Dispatch, and SSIS (SQL Server Integration Services) package execution. The IR is available on Azure, self-hosted, or Azure SSIS platforms.

ADF offers a **graphical overview** to create or manage activities and pipelines that do not require coding skills. However, a user must possess enough ADF experience while dealing with **complex transformation**. Below are some crucial features offered by ADF:

- **Data ingestion:** ADF provides default connectors with almost all on-premise data sources, including MySQL, SQL Server, or Oracle database.
- **Data Pipeline:** ADF allows running pipelines up to one run per minute. However, it does not allow a real-time run.

- **Data Monitoring:** ADF provides you to monitor pipelines with various alert rules. The execution of various pipelines can be monitored through UI and even set up alerts if anything fails using Azure Monitor.

Key Benefits of Azure Data Factory



ADF is a **highly scalable, cost-effective, and agile ETL service** that provides Data Integration solutions to businesses. Enterprises integrate their systems to harness the power of data generated by every other digital software with Business Intelligence tools to make informed decisions. However, to streamline the process of insight generation with analytics tools, organizations rely on ETL processes for transforming collected data and improving the quality of information for further analysis.

- **Fully managed:** Traditional ETL tools have complex deployment processes. Organizations require experts to install, configure, and maintain Data Integration environments carefully. On the other hand, ADF is fully managed by Microsoft that leverages Azure Integration Runtime to handle data movements, Spark Cluster to map data flows, developer tools, and API to ensure peak performance.
- **Low-code:** The most challenging aspect of the ETL pipeline is the transformation stage. Enterprises develop customized scripts written in

different programming languages like **C#, SQL, and Python** based on the business requirements. Although such practices help build complex Data Pipelines, it is tedious to fix bugs with tens of thousands or more lines of code. However, ADF enables developers to transform data by mapping data flows based on industry-standard on the Apache Spark platform. It helps users to create code-free transformations to reduce the turnaround time for analytics, thereby improving productivity.

- **Graphical User interface:** Traditional ETL platforms are either scripting-based or UI-based. They not only lock users to specific and proprietary tools but also fail to deliver the same performance. However, ADF provides a Graphical User Interface (GUI) that allows drag-and-drop features to create a Data Integration pipeline with ease. Such features are utilized by calling an API at the backend. As a result, these developments avoid configuration issues.

What is Databricks?

Databricks is a **SaaS**-based **Data Engineering** tool that processes and transforms massive quantities of data to build Machine Learning models. It supports various Cloud services like **Azure, AWS, and Google Cloud**. For instance, Databricks is optimized for the Microsoft Azure Cloud services platform (**Azure Databricks**) that offers SQL, Data Science, Data Engineering, and Machine Learning environments to develop data-intensive applications. With **Databricks SQL**, analysts can run SQL queries on Data Lakes, create multiple visualizations to explore query results, and build and share dashboards. Databricks also provides an interactive and collaborative workspace for Data Engineers and Machine Learning Engineers to build complex Data Science projects easily.

Key Benefits of Databricks

Databricks is an Apache Spark-based distributed platform that splits workloads among various processors to regulate demands at scale. Below are some benefits of Databricks:

- **Adaptability:** Although Databricks is a Spark-based analytics platform, it still allows multiple programming languages like Python or SQL to interact with Spark. Since it also incorporates Language API at the backend to interact

with Spark, it has higher adaptability in Big Data and Machine Learning domains.

- **Integration:** Databricks integrates with the Azure platform to drive the Azure Big Data solutions with Machine Learning tools in the Cloud. The outcomes of Machine Learning solutions can be visualized in Power BI using Databricks connector to derive valuable insights.
- **Collaboration:** Scripts written in notebooks can be instantly brought into the production phase in Databricks. The collaborative feature provides an environment for multiple members to build Data Modeling and Machine Learning applications effectively.

Simplify Databricks ETL and Analysis with Hevo's No-code Data Pipeline

A fully managed No-code Data Pipeline platform like Hevo Data helps you integrate and load data from 150+ different **data sources (including 40+ free sources)** to a Data Warehouse or Destination of your choice such as **Databricks** in real-time in an effortless manner. Hevo with its minimal learning curve can be set up in just a few minutes allowing the users to load data without having to compromise performance. Its strong integration with umpteenth sources allows users to bring in data of different kinds in a smooth fashion without having to code a single line.

GET STARTED WITH HEVO FOR FREE

Check out some of the cool features of Hevo:

- **Completely Automated:** The Hevo platform can be set up in just a few minutes and requires minimal maintenance.
- **Transformations:** Hevo provides preload transformations through Python code. It also allows you to run transformation code for each event in the Data Pipelines you set up. You need to edit the event object's properties received in the transform method as a parameter to carry out the transformation. Hevo also offers drag and drop transformations like Date and Control Functions, JSON, and Event

Manipulation to name a few. These can be configured and tested before putting them to use.

- **Connectors:** Hevo supports 150+ integrations to SaaS platforms, files, Databases, Analytics, and BI tools. It supports various destinations including Google BigQuery, Amazon Redshift, Snowflake Data Warehouses; Amazon S3, Databricks Data Lakes; and MySQL, SQL Server, TokuDB, DynamoDB, PostgreSQL Databases to name a few.
- **Real-Time Data Transfer:** Hevo provides real-time data migration, so you can have analysis-ready data always.
- **100% Complete & Accurate Data Transfer:** Hevo's robust infrastructure ensures reliable data transfer with zero data loss.
- **Scalable Infrastructure:** Hevo has in-built integrations for 100+ sources **(Including 40+ free sources)** that can help you scale your data infrastructure as required.
- **24/7 Live Support:** The Hevo team is available round the clock to extend exceptional support to you through chat, email, and support calls.
- **Schema Management:** Hevo takes away the tedious task of schema management & automatically detects the schema of incoming data and maps it to the destination schema.
- **Live Monitoring:** Hevo allows you to monitor the data flow so you can check where your data is at a particular point in time.

[SIGN UP HERE FOR A 14-DAY FREE TRIAL!](#)

Azure Data Factory vs Databricks: Key Differences

Interestingly, Azure Data Factory maps dataflows using **Apache Spark Clusters**, and Databricks uses a **similar architecture**. Although both are capable of performing **scalable data transformation, data aggregation, and data movement tasks**, there are some underlying key differences between ADF and Databricks, as mentioned below:

- **Azure Data Factory vs Databricks: Purpose**

- [Azure Data Factory vs Databricks: Ease of Usage](#)
- [Azure Data Factory vs Databricks: Flexibility in Coding](#)
- [Azure Data Factory vs Databricks: Data Processing](#)

Azure Data Factory vs Databricks: Purpose

ADF is primarily used for Data Integration services to perform **ETL processes** and orchestrate data movements at scale. In contrast, Databricks provides a **collaborative platform** for Data Engineers and Data Scientists to perform ETL as well as build Machine Learning models under a single platform.

Azure Data Factory vs Databricks: Ease of Usage

Databricks uses Python, Spark, R, Java, or SQL for performing Data Engineering and Data Science activities using notebooks. However, ADF provides a drag-and-drop feature to create and maintain Data Pipelines visually. It consists of Graphical User Interface (GUI) tools that allow delivering applications at a higher rate.

Azure Data Factory vs Databricks: Flexibility in Coding

Although ADF facilitates the ETL pipeline process using GUI tools, developers have **less flexibility** as they cannot modify backend code. Conversely, Databricks implements a **programmatic approach** that provides the flexibility of fine-tuning codes to optimize performance.

Azure Data Factory vs Databricks: Data Processing

Businesses often do Batch or Stream processing when working with a large volume of data. While batch deals with bulk data, streaming deals with either live (**real-time**) or archive data (**less than twelve hours**) based on the applications. ADF and Databricks support both batch and streaming options, but ADF does not support live streaming. On the other hand, Databricks supports both live and archive streaming options through Spark API.

Conclusion

Businesses continuously anticipate the growing demands of **Big Data Analytics** to harness new opportunities. With rising Cloud applications, organizations are often in a dilemma while choosing **Azure Data Factory** and **Databricks**. If an enterprise wants to experience a **no-code ETL** Pipeline for Data Integration, **ADF is better**. On the other hand, **Databricks** provides a **Unified Analytics platform** to **integrate** various ecosystems for **BI reporting, Data Science, and Machine Learning**.

In this article, you have learned about the **comparative understanding of Azure Data Factory vs Databricks**. This article also provided information on Azure Data Factory, Databricks, and their benefits.

Hevo Data, a No-code Data Pipeline provides you with a consistent and reliable solution to manage data transfer between a variety of sources and a wide variety of Desired Destinations such as **Databricks** with a few clicks.

[VISIT OUR WEBSITE TO EXPLORE HEVO](#)

Hevo Data with its strong integration with 150+ **Data Sources (including 40+ Free Sources)** allows you to not only export data from your desired data sources & load it to the destination of your choice but also transform & enrich your data to make it analysis-ready. Hevo also allows integrating data from non-native sources using Hevo's in-built Webhooks Connector. You can then focus on your key business needs and perform insightful analysis using BI tools.

Want to give Hevo a try? [SIGN UP](#) for a **14-day free trial** and experience the feature-rich Hevo suite first hand. You may also have a look at the amazing [price](#), which will assist you in selecting the best plan for your requirements.

Share with us your experience of learning about **Azure Data Factory vs Databricks**. Let us know in the comments section below!

No-code Data Pipeline for Databricks

[TRY FOR FREE](#)

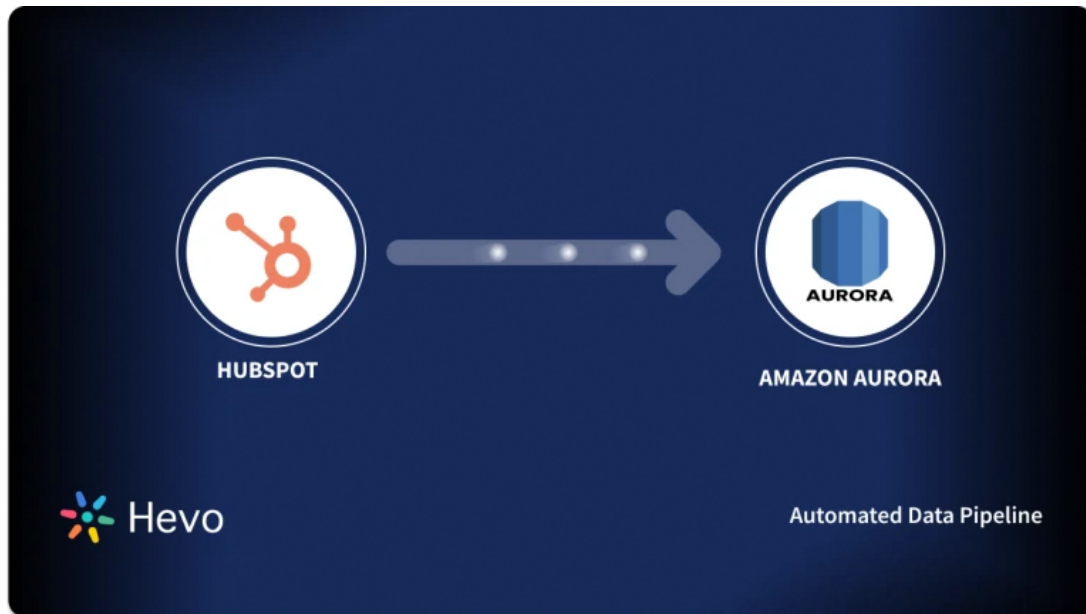
[Azure Data Factory](#)

[Azure Data Factory vs Databricks](#)

[Data Automation](#)

[Databricks](#)

Continue Reading



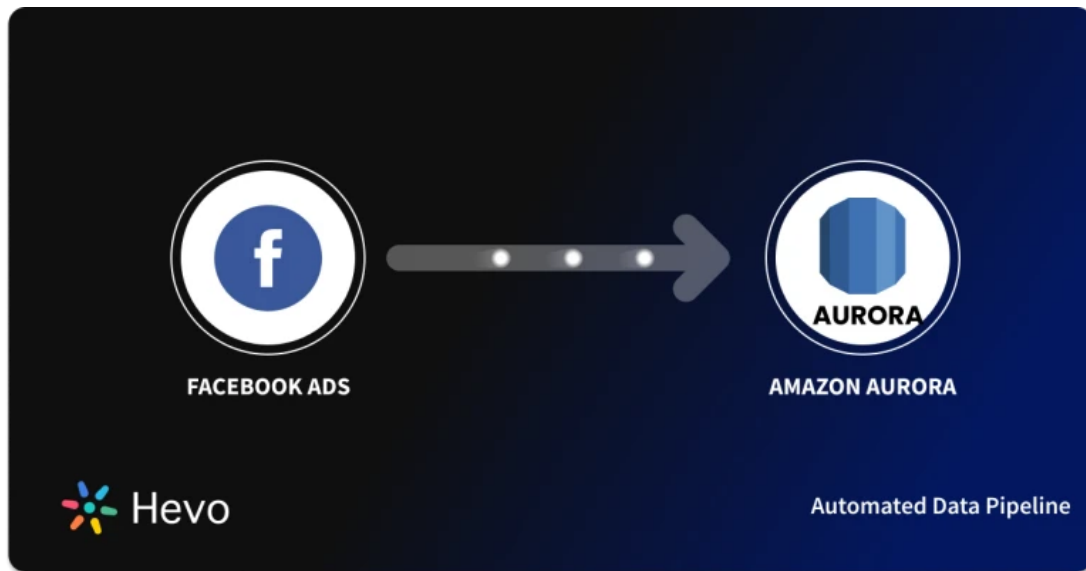
Bhushan Mamtani

Hubspot to Amazon Aurora: 2 Ways to Integrate Data



Suchitra Shenoy

Connect PostgreSQL on Amazon RDS to MySQL: 2 Easy Ways to Integrate Data



Suchitra Shenoy

Facebook Ads to Amazon Aurora: 2 Easy Methods to Load Data

Bring Real-Time Data from Any
Source into your Warehouse

Your Work Email

GET STARTED FOR FREE

Talk to a Product Expert →



Platform

- Hevo Pipeline
- Integrations
- Hevo Pricing
- Security
- Request Demo
- Free Trial
- Upcoming Features
- Changelog
- Hevo Service Status

Resources

- Videos
- Resources Guide
- Success Stories
- Events
- Blog
- Learn
- Documentation
- API Docs
- Engineering Blog
- Sitemap

Industry

- E-commerce
- Finance
- Marketing
- Healthcare
- Sales
- Retail

Company

- Contact Us
- Careers
- Partners
- Hevo Design
- Team

Top Connectors

- MongoDB
- Google Analytics
- Facebook Ads
- PostgreSQL

Top Destinations

- BigQuery
- Snowflake
- Redshift
- PostgreSQL

MySQL

SQL

Popular Integrations

Postgres Export to CSV

Java Connect to Microsoft SQL

Oracle SQL Developer to Excel & CSV

Connect Excel to PostgreSQL

SQLite to MySQL

Our Investors



SEQUOIA

Hevo in News



THE ECONOMIC TIMES



Try Hevo today

BOOK A DEMO



© Hevo Data Inc. 2023. All Rights Reserved. [Privacy policy](#) [Terms of Service](#)