Azure Synapse Vs Data Factory Vs Data Bricks

Azure Synapse Analytics

Azure Synapse is a cloud-based data warehouse service. It is mainly used for storing and analyzing large volumes of structured data. With Synapse, organizations can run complex SQL queries very quickly, generate business reports, and integrate with tools like Power BI. In short, it is best suited for data analytics and reporting.

Azure Data Factory (ADF)

Azure Data Factory is a data integration and orchestration service. Its purpose is to move data from one place to another, schedule workflows, and perform basic transformations. It can connect to many sources (databases, files, APIs, on-premises systems) and deliver data to services like Synapse or Databricks. In simple terms, ADF acts as the pipeline that brings data together and automates the flow.

Azure Databricks

Azure Databricks is a data processing and machine learning platform built on Apache Spark. It is designed to handle large and complex data sets (both structured and unstructured). Databricks is mainly used by data engineers and data scientists for tasks such as data transformation, big data analytics, and AI/ML model development.

How They Work Together

- **Step 1: Data Factory** collects raw data from multiple sources (databases, files, APIs, on-premises).
- **Step 2: Databricks** processes, cleans, and transforms the raw data; advanced ML and analytics can also be applied here.
- **Step 3: Synapse** stores the final structured data and makes it ready for fast querying, dashboards, and reporting through Power BI.

Detailed Comparison

Feature / Aspect	Azure Synapse Analytics	Azure Data Factory (ADF)	Azure Databricks
What it is	Cloud data warehouse for analytics	Data integration & orchestration service	Data processing & machine learning platform
Primary Purpose	Store and query structured data for reporting	Move, transform, and schedule data pipelines	Process large/complex data and build advanced analytics/ML
Data Type Supported	Mainly structured data (tables, SQL)	Structured & semi-structured (depends on target)	Structured, semi-structured, and unstructured data
Core Functionality	Fast SQL queries, dashboards, reporting	Connects to 90+ sources, ETL/ELT pipelines, automation	Big data processing, data transformation, AI/ML model training
Users	Business analysts, BI teams	Data engineers, ETL developers	Data engineers, data scientists, ML engineers
Integration	Tight integration with Power BI	Integrates with Synapse, Databricks, Blob Storage, SQL DB, etc.	Can push processed data into Synapse, ML models to Azure ML
Ease of Use	SQL-based, user friendly for analysts	Low-code/drag-and-drop interface for pipelines	Requires coding (Python, Scala, R, SQL with Spark)
Best For	Business reporting & analytics	Data movement & orchestration across multiple systems	Heavy data transformation, advanced analytics, AI & ML

Summary

- **Synapse** : For storing and analyzing structured data (reports, dashboards).
- **Data Factory** : For moving, integrating, and scheduling data pipelines.
- **Databricks** : For processing raw/complex data and applying advanced analytics or machine learning.