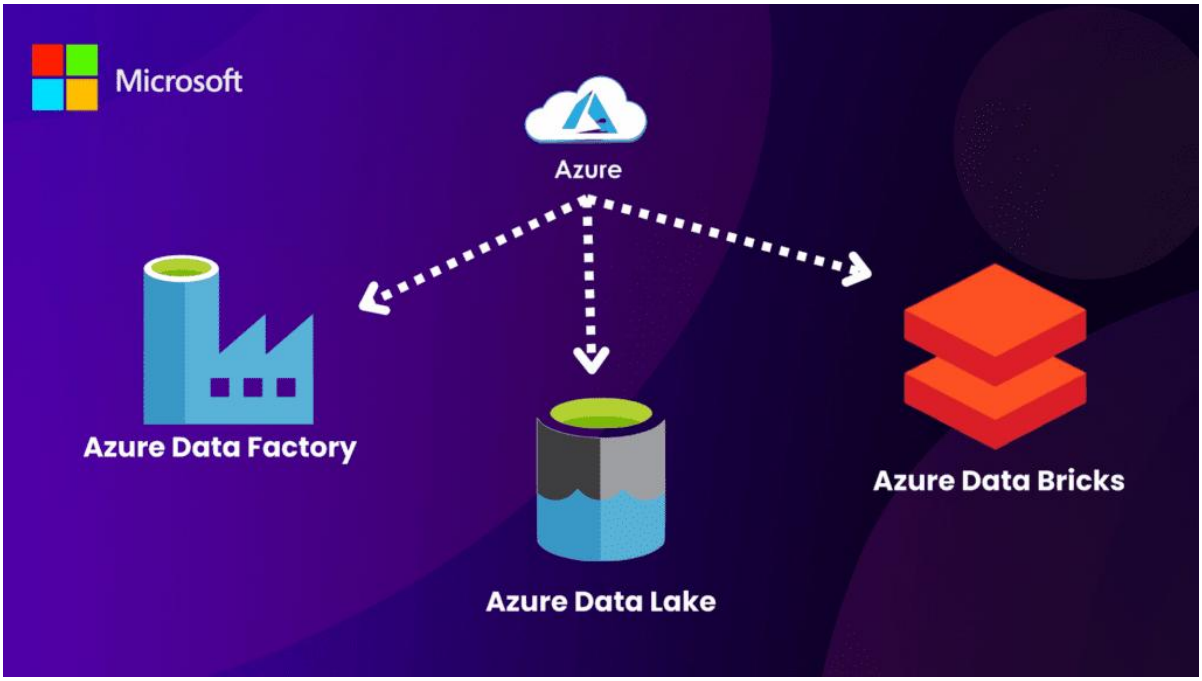
Azure Synapse Vs Data Factory Vs Data Bricks

- Azure Synapse Analytics, Azure Data Factory, and Azure Databricks are three powerful but distinct services in the Microsoft Azure ecosystem, each serving unique purposes within the data lifecycle.
- **Azure Synapse** Analytics is primarily a data warehouse and analytics platform designed for large-scale data storage and querying. It integrates structured and semi-structured data, supports SQL-based querying, and is optimized for business intelligence (BI) and reporting scenarios where fast insights are crucial.
- **Azure Data Factory (ADF)**, on the other hand, is a cloud-based ETL (Extract, Transform, Load) and data integration service used to orchestrate, move, and transform data across different sources. It acts as the pipeline builder, enabling scheduled workflows and connections between on-premises and cloud systems, making it ideal for data ingestion and preparation.
- **Azure Databricks** complements these services as a collaborative big data and machine learning platform built on Apache Spark. It provides a scalable environment for advanced analytics, AI model training, and handling unstructured or streaming data.



- Data Factory is primarily for data integration and ETL pipelines, Synapse is a comprehensive analytics service that combines data warehousing and big data analytics, and Databricks is a collaborative platform built on Apache Spark for data science and machine learning.
- Synapse focuses on querying and reporting, ADF handles data movement and workflow automation, and Databricks enables deep data engineering, data science, and AI. Together, they form a comprehensive ecosystem.

Feature / Service	Azure Synapse Analytics	Azure Data Factory (ADF)	Azure Databricks
Primary Purpose	Data warehousing & analytics	Data integration & orchestration	Big data processing & machine learning
Best For	Business Intelligence (BI), dashboards, reporting	Data ingestion, ETL/ELT pipelines, workflow automation	Data engineering, AI, ML, advanced analytics
Data Type Support	Structured & semi-structured	Structured, semi-structured, unstructured	Structured, semi-structured, unstructured, streaming
Processing Style	SQL-based queries (batch analytics)	Batch data movement & transformation	Distributed computing with Spark (batch + streaming)
Integration	Power BI, ADF, Databricks	Synapse, Databricks, multiple data sources	Synapse, ADF, ML frameworks (TensorFlow, PyTorch, etc.)
Strengths	Fast SQL queries, tight Power BI integration	Easy to build pipelines, serverless data orchestration	Handles massive datasets, advanced ML/AI support
Limitations	Not ideal for unstructured data	Limited for heavy transformations/AI	Requires technical expertise, higher costs for large workloads