Azure Synapse Vs Data Factory Vs Data Bricks

Azure Data Factory (ADF) :

Azure Data Factory is the data integration and orchestration service that acts as a pipeline manager in the Azure ecosystem. Its role is to collect data from various sources, apply initial transformations, and load it into destinations like data lakes, warehouses, or analytics platforms.

Key Features:

- Serverless & Dedicated SQL Pools (pay-per-query or reserved).
- Integrates with Power BI for reporting.
- Handles petabytes of data with optimized query performance.

For example, in a retail company, ADF can pull data from point-of-sale systems, supplier databases, and e-commerce platforms, then schedule pipelines to load this data into Azure Data Lake every night. By automating the movement and transformation of data, ADF ensures that all other systems always have the latest and consistent information to work with.

Azure Databricks:

Azure Databricks is designed for big data processing, advanced analytics, and machine learning. It runs on Apache Spark and allows developers and data scientists to process large volumes of structured and unstructured data efficiently. Unlike ADF, which just moves and lightly transforms data, Databricks provides the environment to clean, enrich, and analyze it deeply.

Key Features:

- 90+ connectors (SQL, Blob, Salesforce, SAP, etc.).
- No-code/low-code pipeline builder.
- Triggers, scheduling, and monitoring of data pipelines.

For example, a bank might use Databricks to process millions of daily transactions, detect unusual spending behaviour, and build fraud detection models using machine learning. Databricks can handle the scale and complexity of such operations and produce valuable insights that help the bank protect its customers.

Azure Synapse Analytics :

Azure Synapse Analytics is the data warehousing and analytics platform that allows businesses to query and analyse large datasets using SQL. Its strength lies in making processed and structured data accessible for reporting and business intelligence.

Key Features:

- Apache Spark–based distributed computing.
- Supports Python, R, Scala, Java, SQL.
- Integrates with MLflow, TensorFlow, PyTorch, etc.
- Great for advanced analytics & predictive modeling.

For example, in a healthcare company, after patient and hospital records are collected through ADF and processed in Databricks to remove inconsistencies, the refined data is stored in Synapse. Analysts and doctors can then run queries in Synapse to understand treatment outcomes, patient trends, and hospital performance. The results can be visualized directly in Power BI dashboards, giving decision-makers clear, data-driven insights

COMPARISION TABLE:

Feature	**Synapse**	**ADF**	**Databricks**
Primary Use	Data Warehousing & Analytics	Data Movement & Orchestration	Big Data Processing & ML
Data Handling	Structured, Semi-structured	All types	Structured, Semi-structured, Unstructured
Processing Model	SQL queries	ETL/ELT workflows	Spark-based Distributed Computing
Best For	BI & Reporting	Data Integration	AI/ML, Advanced Analytics
Skillset Needed	SQL	Low-code / Data Eng	Python, R, Scala, ML/AI