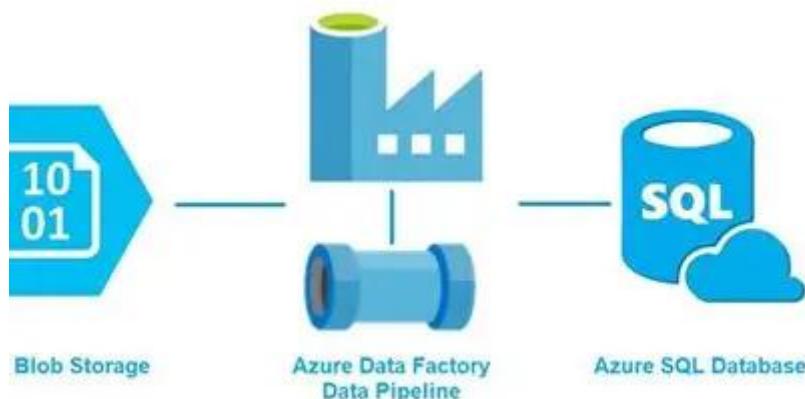


1. Azure Data Factory (ADF)

- A cloud-based ETL (Extract, Transform, Load) and data integration service that orchestrates and automates data movement and transformation.
- **Key Properties:**
 - **Data Integration:** Connects to 90+ data sources (on-prem & cloud).
 - **Pipeline Orchestration:** Build workflows for data movement.
 - **Mapping Data Flows:** Visual interface for transformations.
 - **Triggering & Scheduling:** Time-based or event-based triggers.
- **Use Cases:**
 1. **Data Migration:** Move data from on-prem SQL Server to Azure SQL Database.
Example: Migrating ERP data to cloud.
 2. **ETL for Analytics:** Extract data from multiple sources, transform, and load into Azure Synapse.
Example: Retail company consolidating sales data.
 3. **Data Pipeline Automation:** Automate ingestion from APIs to Data Lake.
Example: IoT sensor data ingestion every hour.



2. Azure Data Lake

- **Definition:** A scalable data storage service optimized for big data analytics; stores structured and unstructured data.
- **Key Properties:**
 - **Hierarchical Namespace:** Organize data like a file system.
 - **Massive Scalability:** Handles petabytes of data.
 - **Integration:** Works with HDInsight, Databricks, Synapse.

- **Use Cases:**

1. **Big Data Storage:** Store raw logs, images, and sensor data. *Example:* Autonomous vehicle company storing telemetry data.
2. **Data Lakehouse Architecture:** Combine raw and curated data for analytics. *Example:* Finance firm storing transaction history for fraud detection.
3. **Machine Learning Data Prep:** Store training datasets for ML models. *Example:* Healthcare company storing patient records for predictive analysis.

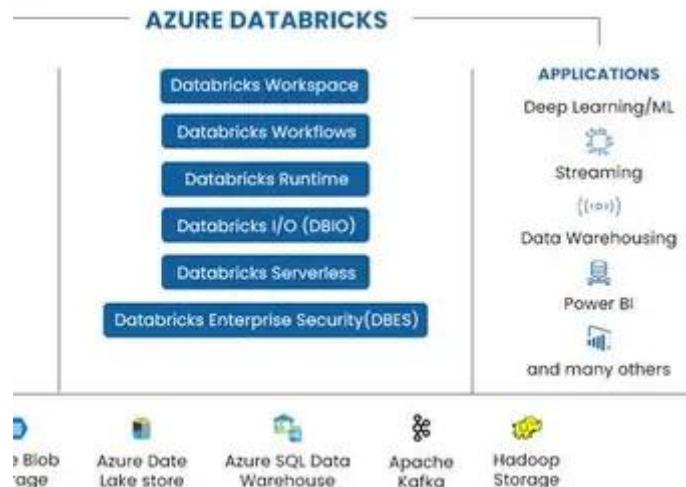


3. Azure Databricks

- **Definition:** An Apache Spark-based analytics platform for big data and AI workloads.
- **Key Properties:**
 - **Unified Analytics:** Combines data engineering, data science, and ML.
 - **Scalable Compute:** Auto-scaling clusters for large workloads.
 - **Integration:** Works with Data Lake, Synapse, ML services.

- **Use Cases:**

1. **Data Engineering:** Clean and transform large datasets. *Example:* ETL pipeline for e-commerce clickstream data.
2. **Machine Learning:** Train predictive models on big data. *Example:* Predicting customer churn using historical data.
3. **Real-Time Analytics:** Stream processing for IoT data. *Example:* Smart factory analyzing sensor data in real time.



4. Azure Synapse Analytics

- A cloud-based analytics service that combines **big data** and **data warehousing** for querying and analyzing large datasets at scale.
- **Purpose:** Enables integration of structured and unstructured data for BI and advanced analytics.

Key Properties

- **Massive Parallel Processing (MPP):** Distributes queries across multiple nodes for speed.
- **Integrated with Azure Ecosystem:** Works with Data Lake, Power BI, ADF.
- **Hybrid Querying:** Supports both on-demand (serverless) and provisioned (dedicated) SQL pools.
- **Security & Governance:** Built-in encryption, role-based access, and compliance.

Use Cases

1. **Enterprise Data Warehouse:** Store and analyze structured business data.
Example: Retail company analyzing sales and inventory trends.
2. **Big Data Analytics:** Query petabytes of data from Data Lake using serverless SQL.
Example: Telecom analyzing call detail records for patterns.
3. **BI & Reporting:** Connect Synapse to Power BI for dashboards.
Example: Finance team visualizing revenue and expense KPIs.

Examples

- **Scenario 1:** A healthcare organization uses Synapse to combine patient records from SQL databases and IoT health devices for predictive analytics.

- **Scenario 2:** A global e-commerce company uses Synapse for real-time sales reporting integrated with Power BI.
- **Scenario 3:** A bank uses Synapse to run fraud detection queries on historical transaction data stored in Data Lake.

Azure Synapse Analytics

