AZURE DATA FACTORY

Azure Data Factory (ADF) is a cloud-based data integration service that allows you to create, schedule, and orchestrate ETL (Extract, Transform, Load) and ELT (Extract, Load, Transform) workflows. It is widely used for moving data between various sources and destinations in Azure and on-premises environments.

Key Features of Azure Data Factory

1. **Data Integration** – Connects to multiple data sources, including Azure services, on-prem databases, SaaS applications, and cloud storage.
2. **Data Transformation** – Uses **Mapping Data Flows** or **Azure Databricks** for transforming data before loading it into a destination.
3. **Data Orchestration** – Schedules and automates workflows across different services.
4. **Monitoring & Logging** – Provides real-time monitoring and error-handling capabilities.
5. **Scalability** – Supports big data workloads and high-performance parallel processing.

Core Components of ADF

1. **Pipelines** – A logical grouping of activities that perform a data workflow.
2. **Activities** – Tasks like data movement, transformation, or control flow (e.g., executing stored procedures).
3. **Datasets** – References to data stored in linked services (e.g., Azure Blob Storage, SQL Database).
4. **Linked Services** – Connectors to various data stores (e.g., Azure SQL, Amazon S3, SAP, Oracle).
5. **Integration Runtimes (IR)** – Compute infrastructure that executes data flows. There are three types:
    - **Azure IR** – For cloud-based transformations.
    - **Self-hosted IR** – For on-prem data movement.
    - **SSIS IR** – For running SSIS packages in Azure.

Common Use Cases

- ☑ Data migration from on-prem to cloud
- ☑ ETL & ELT pipeline creation

- ✅ Data warehousing (e.g., loading data into Azure Synapse Analytics)
- ✅ Incremental data load and CDC (Change Data Capture)
- ✅ IoT and streaming data processing

ADF COMPONENTS & TERMINOLOGY

1. activity -
2. dataset -
3. linked service-
4. source & sink
5. trigger
6. integration run time