

2. Azure Data Factory

Tuesday, February 10, 2026 1:30 PM

What is Azure Data Factory and Its Key Components?

Azure Data Factory (ADF) is a cloud-based data integration service that allows you to create data-driven workflows for orchestrating and automating data movement and data transformation. It is designed to handle complex data pipelines and ETL (Extract, Transform, Load) processes, making it easier to integrate data from various sources.

Key Features:

- Data Movement:
 - Move data between various data stores, both on-premises and in the cloud, with high reliability and performance.
- Data Transformation:
 - Perform data transformation at scale using data flows, mapping data flows, and external processing frameworks like HDInsight, Databricks, and Azure Machine Learning
- Orchestration and Scheduling:
 - Create complex data pipelines to orchestrate data movement and transformation, schedule jobs to run at specific times, and manage dependencies.
- Integration:
 - Integrates seamlessly with a wide range of data sources, including Azure services (Azure Blob Storage, Azure SQL Database, Azure Data Lake), on-premises databases, and third-party services.
- Monitoring and Management:
 - Provides comprehensive monitoring and management capabilities to track pipeline execution, diagnose issues, and ensure data processes run smoothly.
- Code-Free and Code-Based Options:
 - Offers a visual interface for building data pipelines without writing code and supports code-based development using SDKs and APIs.

Key Components of Azure Data Factory:

1. **Pipelines:** Logical grouping of activities that perform a unit of work. ○ Activities: Define the actions to be performed on data (e.g., data movement, data transformation, control flow).
2. **Linked Services:**
 - Linked Services define the connection information needed for Azure Data Factory to connect to external resources. They specify how Data Factory should authenticate and access these resources, such as by using a username and password, or key-based authentication.
3. **Datasets:**
 - Datasets represent the data structures within the data stores. An input dataset represents the data used as input for an activity in the pipeline, while an output dataset represents the data produced by an activity. For example, an Azure Blob dataset might specify the blob container and folder in Azure Blob Storage from which the pipeline should read data, while an Azure SQL Table dataset specifies the table to which the activity writes its output.
4. **Activities:**
 - Activities define the actions that are performed on your data within a pipeline. In Azure Data Factory, activities can be used for data movement, data transformation, or control operations. Examples include copying data from one source to another, transforming data using Azure Data Lake Analytics, or executing stored procedures in a SQL database.
5. **Triggers:**
 - Run pipelines on specified schedules, in response to events, or manually.
6. **Integration Runtimes:**

