

b'

## Microsoft Azure Introduction to Azure Data Factory

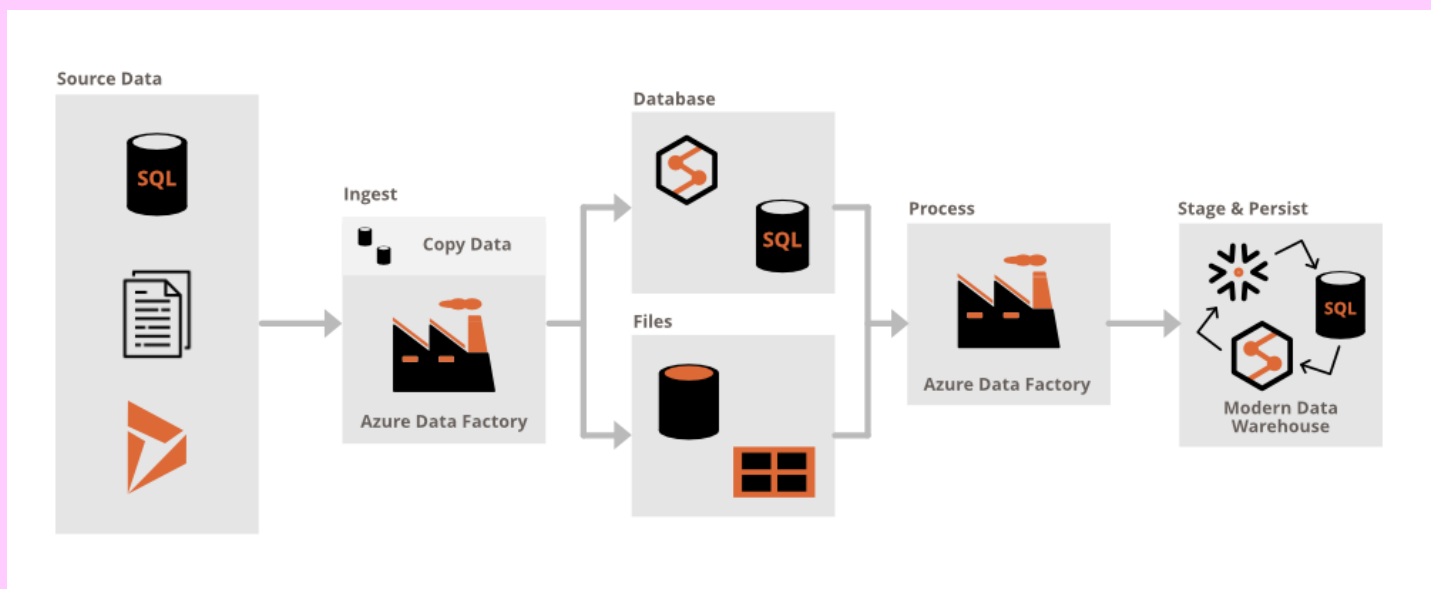
- Last Updated : 09 Jul, 2021

Azure data factory as commonly known as ADF is a ETL(Extract-Transform- load ) Tool to integrate data from various sources of various formats and sizes together, in other words, It is a fully managed, server less data integration solution for ingesting, preparing, and transforming all your data at scale.

As the data is increasing day by day around the world many enterprises and businesses are shifting towards the usage of cloud-based technology to make their business scalable. Because of the increase in cloud adaption, there is a need for reliable ETL tools in the cloud to make the integration. The Azure data factory stands out when compared to other ETL tools because of **features** such as **Easy to Use**, **Cost-Effective solution** , **Powerful and intelligent code free service**.

### The architecture of Azure data Factory:

The figure below describes the Architecture of the data engineering flow using the Azure data factory



### Azure data Factory

The various components of the Azure data factory are as follows:

1. Activities
2. Pipelines
3. Linked Services
4. Datasets
5. Data Flows
6. Integration Runtimes

All these components work together in runtime to help extract and transform the source data.

Before understanding what a pipeline is it is necessary to understand what an activity is.

- **Activity** : Activities in a pipeline define actions to perform on data. For example, copy data activity can read from one location of Blob storage and loads it to another location on Blob storage
- **Pipeline** : Pipeline is a logical grouping of activities that together perform a task. For example, Pipeline can have a set of activities that take data from ADLS and perform some transformation of data using U-SQL and load data in SQL DB
- **Linked Services**: Linked services are used to connect to other sources with the Azure data factory. Linked services act as connection strings for resources to connect. For example, Connecting an AWS S3 to the Azure Data Factory

- **Datasets:** Datasets are simply points or reference the data, which we want to use in our activities as input or output
- **Dataflows:** Data flows feature in the Azure data factory will allow users to develop graphical data transformation logic that can be executed as activities in ADF pipelines
- **Integration Runtimes:** The Integration Runtime(IR) is to compute infrastructure used by ADF to provide capabilities such as Data Flow, Data Movement, Activity Dispatch, and SSIS Package Execution across different network environments.

## Pricing Of Azure Data Factory:

- No upfront cost
  - No termination fees
  - Pay only for what you use
1. **Data Pipelines:** Helps to Integrate data from cloud and hybrid data sources, at scale. Pricing starts from \$972.046 / 1,000 activity runs per month
  2. **SQL Server Integration Services:** Helps to easily move your existing on-premises SQL Server Integration Services projects to a fully-managed environment in the cloud. -Pricing for SQL Server Integration Services integration runtime nodes start from \$960.498 /hour

My Personal Notes

Add your personal notes here

Save