Azure data factory

**Pipelines in azure**

Pipeline is logical grouping of activities that together perform a task.

It has a set of activities that can take data from ADLS and perform some transformation on data and load into SQL DB

A pipeline could contain a set of activities that ingest and clean log data, and then kick off a mapping data flow to analyse the log data.

**Activities**

Activities defines actions to perform on the data in a pipeline. For example you use copy activity to copy data from azure SQL to azure blob storage, then use a data flow activity or data brick notebook activity to process or transform data from blob storage to data warehouse for analytics purpose

An activity can take zero or more input [datasets](https://learn.microsoft.com/en-us/azure/data-factory/concepts-datasets-linked-services) and produce one or more output [datasets](https://learn.microsoft.com/en-us/azure/data-factory/concepts-datasets-linked-services).

**Data sets**

**dataset** is a named view of data that simply points to or references the data you want to use in your **activities** as inputs and outputs.

 Data set identifies the data within the data stores such as tables, files, folders,

documents

Data sets are the reference points for the data in which you want to use in your

Activities as input or output

**Linked services:**

It act like a bridge between azure data factory and data sources that we want interact with. Linked services stores the necessary connection strings, authentication details and other configuration required to establish connection with data sources

We can parametrize the linked services and add dynamic values during runtime

**Triggers in ADF**

In Azure Data Factory (ADF), triggers are a key feature that enable you to automate and schedule the execution of data pipelines and other activities. Triggers play a crucial role in orchestrating data movement and transformation workflows based on predefined time schedules or external events.

There are three types of triggers in execution of pipelines

**Scheduled trigger**

**Thumbling window trigger**

**Event based trigger**

**Integration runtime in ADF**

The integration runtime is a compute infrastructure used by azure data factory to provide the following data integration capabilities across different network environments.

* **Data flow**: execute a data flow in managed azure compute environment
* **Data movement**: copy data across data stores in public network and private network
* **Activity dispatch**: dispatch and monitor transformation activities running on a variety of computed services.
* **Siss package execution**: execute siss package in a managed compute environment assignment