Azure Data Factory

What is Azure Data Factory?

→ It is a cloud based data integration service that allows you to create data driven workflows in the cloud for orchestrating and automating data movement and data transformation.

→ It does not store data itself.
→ It basically helps to transfer data from one place to another.
→ It allows you to monitor and manage workflows using both programmatic and UI mechanisms.
→ Used for data movement and transformation.

ADF can be used for:-

→ Supporting data migration.
→ Creating data from a client's server or azure data to an azure data lake.
→ Carrying out various data integration processes.
→ Integrating data from different ERP systems and loading it into Azure Synapse for reporting.

hosted on Azure

Service / tools for azure

sql python pyspark sparkSql.	⇒	azure databricks azure data lake / adls gen 2 azure data factory azure Synapse

azure DevOps

(2)

classmate
Date
Page

How Azure Data Factory Works?

- Basically it will move, transform or run data in pipeline.
  ~~used to transfer~~ data from source to destination.
- ADF allows you to create data pipelines which move & transform data and then ~~move~~ run the pipelines on a specified schedule (hourly, daily, weekly, etc.)

~~Step1:- Connect and Collect & Transform~~

Step 1:- Connect and Collect.

→ Connect to all the required sources of data and processing such as SaaS services, file share, FTP and web services.

→ Then, move the data as needed to a centralized location for subsequent processing by using the Copy Activity in a data pipeline to move data from both on-premise and cloud source data stores to a centralization data store in the cloud for further analysis.

Step 2:- Transform and Enrich.

→ Once data is present in a centralized data ~~storage~~ store in the cloud, it is transformed using compute services such as HDInsight Hadoop, Spark, Azure Data Lake Analytics, and Machine Learning.

Step 3:- Publish

→ Deliver transformed data from the cloud to on-premise sources like SQL server or keep it in your cloud storage sources for consumption by BI and analytics tools and other applications.

Data Migration Activities with Azure Data.

→ By using Microsoft Azure Data Factory, data migration occurs between two cloud data stores & between an on-premise data store and a cloud data store.

→ Copy activity in ADF copies data from a source data store to a sink data store.

→ Azure supports various data stores such as source or sink data stores like Azure Blob Storage, Azure Cosmos DB (DocumentDB API).

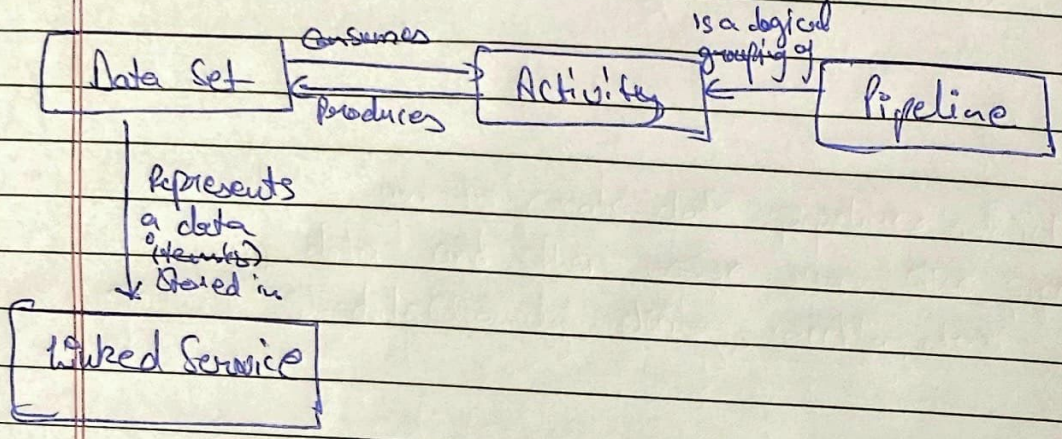
Pipeline is a group of activities:- They are used to group activities into a unit that together performs a task. A Data factory may have one or more pipelines.

- Activities define the actions to perform your data. Currently, Azure data factory supports two types of activities:- data movement and transformation.

- Linked services define the information needed for Azure data factory to connect to external resources. For eg:- an Azure Storage linked service specifies a connection string to connect the Azure Storage Account.

4.

How the Azure Data Factory Components work together.



Following tools or API's can be used to create data pipelines in Azure Data Factory.

- Azure Portal
- Visual Studio
- PowerShell
- .Net API
- REST API
- Azure Resource Manager Template.