**AZURE DATA FACTORY:**

What is ETL

What is the Data Factory Version: V2

What are the differences between SSIS and Azure Data Factory

What are different components in Azure Data Factory

**Linked Services:** Linked service will stores the connection information.

Establishing the connection to the Data Sources.

Ex-Storage account, Azure SQL Database, SQL Server

**Integration Runtimes:** Providing Compute infrastructure to data factory pipeline activities to execute and connecting to different networks.

There are 3 types of Integration Run Times.

1. **Self-hosted Integration Runtime—**Connecting to On-premises or Private networks.
2. **Auto resolve or Azure Integration Runtime**—Connecting to Cloud services.

**Auto Reslove**: Data factory will be created by default.

**Azure Integration**: Customized integration Runtime to connect.

1. **SSIS integration Runtime**—Lifting and shifting of ssis packages from on-premises to cloud environment.

**Datasets:** Preparation of data by using Linked Services.

**Activities:** Copy of Data

**Pipelines:**Data factory work flow

**Triggers**: 3 types of triggers

1. Normal Scheduling Trigger
2. Event Based Trigger
3. Tumbling window Trigger

**SCENARIO 1: COPY DATA FROM ON-PREMICES SQL SERVER DATABASE TO BLOB STORAGE**

1. Integration Runtimes:

* Self-hosted Integration run times required for --on premise SQL Server Database.
* Auto Resolve integration runtimes required for--Blob Storage.

1. Linked Services:

* One for on premises SQL Server database—Self hosted Integration Runtimes
* One more for Blob Storage – Auto Resolve Integration Runtime.

1. Dataset:

* One Data set for Source SQL table: Linked Services of Onprem SQL.
* One more Dataset for Destination Blob Storage: Linked services of Blob Storage.

1. Pipeline: Contains work flow
2. Activity:

Copy of Data activity

1. Triggers: Scheduling of Data factory pipeline to execute on particular time.

**SCENARIO 2: COPY DATA FROM AZURE BLOB STORAGE TO AZURE SQL DATABASE (CLOUD ENVIRONMENT)**

1. Integration Runtimes:

* Auto Resolve Integration run times required for -- Blob Storage.
* Auto Resolve integration runtimes required for—AZURE SQL Database.

1. Linked Services:

* One for on AZURE SQL database—Auto Resolve Integration Runtime.
* One more for Blob Storage – Auto Resolve Integration Runtime.

1. Dataset:

* One Data set for Source is Blob Storage: Linked Services of Blob Storage.
* One more Dataset for Destination AZURE SQL Database.: Linked services of AZURE SQL Database.

1. Pipeline: Contains work flow
2. Activity:

Copy of Data activity

1. Triggers: Scheduling of Data factory pipeline to execute on particular time.

**Activities:**

**Copy Data Activity:**

Copy data activity is using for copying the data from source to destination.

**Get Metdata Activity:**

Get Metadata activity is used for get met data information about the files

Ex: all the file names in ur folder, whether file existing or not, last modified date and time of file.

**For Each loop Activity:**

if we want to executive any activity for multiple times by looping we can use each loop activity.

**SCENARIO 3: COPYING ALL DATA IN MULTIPLE FILES FROM AZURE BLOB STORAGE TO AZURE SQL TABLE(DATABASE) IN CLOUD ENVIRONMENT**

**Blob Storage:**

**Multiple files-**emphyd.txt, empbng.txt

**AZURE SQL DATABASE**

**SQL Table**- employee

BELOW STEPS NEEDS TO BE CREATED:

1. Create Data factory
2. Check Auto Resolve Integrations
3. Create 2 linked services, one for Blob storage and other is one for Azure SQL Database
4. 2 Datasets based on Activities.