AZURE DATA BRICKS (DB)

BLOB STORAGE

Binary large object

ETL thorough linked services

AZURE SQL

**Steps to create Pipeline:**

1. Create blob storage, container and upload all documents.
2. Create azure SQL DB
3. Create DB under analytics. Pls give only the resource group name and the Databricks name.
4. Click Azure Data bricks launch workspace
5. Create new notebook, create new resource based on the minimal configuration required.
6. Once the cluster is created, do the following:

* Import pyspark libraries.
* Try accessing blob and list all the files using the access key, by selecting access keys from the side menu in the blob storage. Also ensure while creating Blob storage soft delete option should not be enabled.
* Then do the data processing of the required files.
* Once done, create connection of ASQL server by getting all the details from the Azure SQL already created.
* Then write the dataframe data to the table in ASQL.

**Points to remember:**

1. **Access Key methos is also know an “FULL ACCESS” method because it lists all the files in the container and user can access any documents.**
2. **Also, the access key is used in Databricks with any out any encryption while we access the container from databricks.**































