ASSESSMENT

Project Description

Azure Resources Used for the Project:

* Azure Data Lake Storage
* Azure Data Factory
* Azure Databricks

The requirements for this project are

Data Ingestion

Data Transformation

Data Reporting

Data Analysis

Scheduling Requirements

Technologies Used:

* Pyspark
* Spark SQL
* Delta Lake
* Azure Databricks
* Azure Data Factory
* Azure Date Lake Storage Gen2
* Azure Key vault

Creating Delta Lake Tables in Azure Databricks

1. Using a sample CSV file as the source to create a delta format table in Azure Databricks
2. Once the Azure Databricks instance is created, launch the workspace
3. Upload File option and upload the sample file here.
4. To create the tables we need to have a cluster
5. Once the cluster is created, create a new table using the notebook
6. we need to provide an appropriate temporary table name that will create a view in the data frame.
7. we can execute the sample SQL query to ensure that the table can be queried,
8. Once this delta table is created, we need to ensure that it’s appearing in the list of tables.