Context:

Azure Databricks facilitates the usage of SQL queries on unstructured data stored in a data lake, transform the data as required and store it in a structured manner using Azure Synapse Analytics

In this exercise, you will load an unstructured database into a Spark cluster created in Databricks, and store it in an SQL data warehouse after transformation.

Learning Outcomes:

1) Deploy a Spark cluster in Databricks
2) Upload sample data into a file system created in Azure Data Lake Storage Gen2
3) Extract data from the storage account into Databricks and transform it using Spark queries
4) Load the transformed data into Azure Synapse for storage

Steps to be performed

1) Create the following pre-requisite resources
   a) An Azure Synapse SQL Pool
   b) An Azure Blob storage account
   c) An Azure DataLake storage Gen2 account
2) Deploy a Databricks service in a region of your choice
3) Authenticate the databricks service to access the created storage account
4) Load the following dataset into the storage account
   https://raw.githubusercontent.com/Azure/usql/master/Examples/Samples/Data/json/radiowebsite/small_radio_json.json
5) Extract the above dataset into Databricks
6) Perform the following transformations in Databricks on the dataset
   a) Retrieve the following columns: firstname, lastname, location
   b) Rename the column "location" to address
7) Load the transformed data into the created Synapse service.

Note:

● Other required values can be set at your discretion.
● Submission of this assessment shall be done in the form of a pdf document containing the labeled screenshots as outlined in the marks distribution section.

Marks Distribution:

1) Deployment of Spark Databricks cluster   10 marks
2) Loading of the dataset into the storage account   10 marks
3) Extraction of data into Databricks   10 marks
4) Output of transformation queries   10 marks
5) Loading of trqansformed data into Synapse   10 marks

Reference links:

1) https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/create-data-warehouse-portal
2) https://docs.microsoft.com/en-us/azure/storage/blobs/storage-quickstart-blobs-portal
3) https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-quickstart-create-account
4) https://docs.microsoft.com/en-us/azure/databricks/scenarios/quickstart-create-databricks-workspace-portal?tabs=azure-portal