Read and write data by using Azure Databricks

Task 1. Clone the Databricks archive

1. From the Azure portal, go to your Databricks workspace and select Launch workspace.
2. In the left pane, select Workspace, select Users, and then select your username (the entry with the house icon).
3. In the pane that appears, select the downward-pointing chevron next to your name, and then select Import.
4. In the Import Notebooks pane, select URL, and paste in the following URL:

https://github.com/MicrosoftDocs/mslearn-read-and-write-data-using-azure-databricks/blob/master/DBC/04-Reading-Writing-Data.dbc?raw=true

5. Select Import.
6. A folder named after the archive should appear. Select that folder. The folder contains one or more notebooks that you'll use in completing this lab.

Task 2. Complete the following notebooks

To complete the labs:

1. Continue working within your Azure Databricks workspace, and open the new 04-Reading-Writing-Data folder. Within the folder, you'll find Python, Scala, and Spark subfolders
2. Choose the folder for the language you prefer to use, open the corresponding folder, and then open the notebook.
3. Follow the instructions within the notebook until you've completed the entire notebook. Then continue with the remaining notebooks in order.
   a. 01-Getting-Started: This notebook gets you started with your Databricks workspace.
   b. 02-Querying-Files: This notebook contains exercises to help you query large data files and visualize your results.
   c. 03-Joins-Aggregations: You do basic aggregation and joins in this notebook.
   d. 04-Accessing-Data: This notebook lists the steps for accessing data from various sources by using Databricks.
   e. 05-Querying-JSON: In this notebook, you learn how to query JSON and hierarchical data with DataFrames.
   f. 06-Data-Lakes: This notebook lists exercises that show how to create an Azure Data Lake Storage Gen2 instance and use Databricks DataFrames to query and analyze this data.
   g. 07-Azure-Data-Lake-Gen2: In this notebook, you use Databricks to query and analyze data stores in Azure Data Lake Storage Gen2.
   h. 08-Key-Vault-backed-secret-scopes: This notebook lists the steps for configuring a Key Vault-backed secret scope. You'll create a Key Vault-backed secret scope and securely store in it usernames and passwords for a sample SQL database and an Azure Cosmos DB instance to be used in the following notebooks.
   i. 09-SQL-Database-Connect-Using-Key-Vault: In this notebook, you'll connect to a SQL database by using your Azure SQL username and password that you created and securely stored in the Key Vault-backed secret scope in the previous notebook.

j. 10-Cosmos-DB-Connect-Using-Key-Vault: In this notebook, you'll connect to an Azure Cosmos DB instance by using the Azure Cosmos DB username and password that you previously created and securely stored in the Key Vault-backed secret scope.

k. Exploratory-Data-Analysis: This notebook is in the Optional subfolder. It includes a sample project for you to explore later.