Setting up Azure Databricks Workspace & Cluster

Exercise 0 – Setup Azure Data Lake Gen2 Account

1. Complete Lab 1 – Working with Azure Data Lake Gen2 account
2. Upload Files if not already uploaded

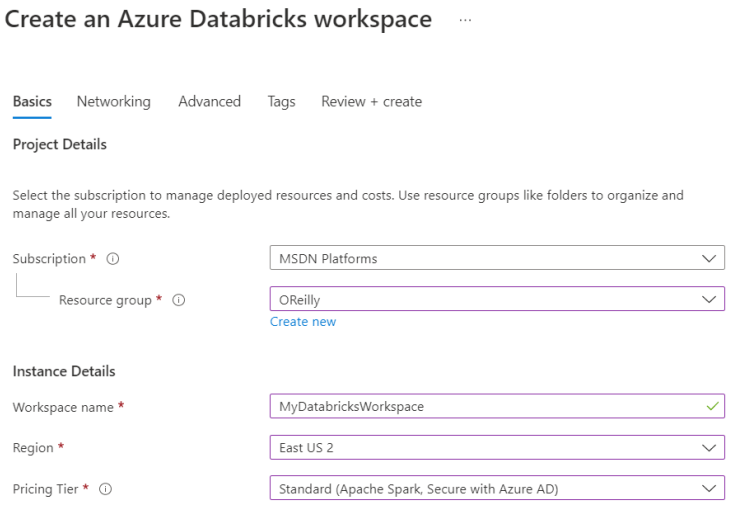
Exercise 1 – Setup Azure Databricks Workspace

1. Go to Azure portal (portal.azure.com)

2. In the search bar, search for Azure Databricks. And select it

3. Click on Create New

4. Fill up the properties to create account
    a. [Basics Tab]
        i. Select subscription
        ii. Select resource group
        iii. Provide a unique name
        iv. Select region of your choice (example – East US 2)
        v. Select pricing tier as Trial or Standard
        vi. Click Review + Create

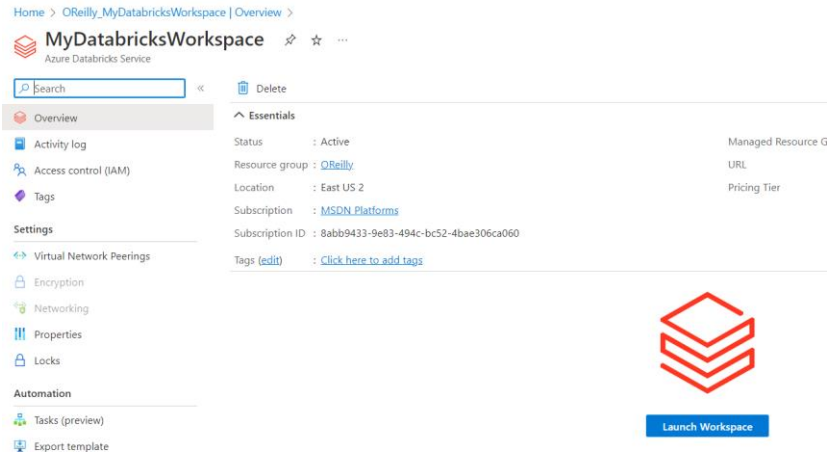


    b. Click Create
    c. This will take few minutes to create

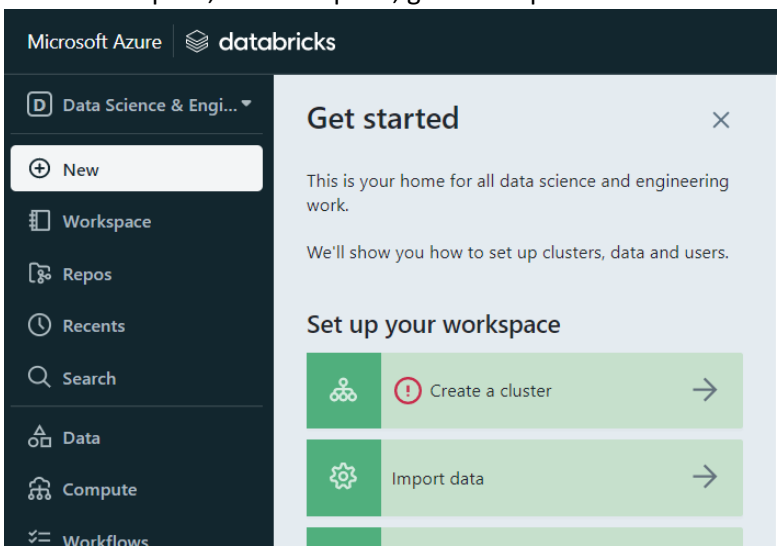
Exercise 2 – Launch Databricks Workspace & Create Cluster

1. Open Azure Databricks instance created in the previous step

2. Click on Launch workspace, to open Databricks UI



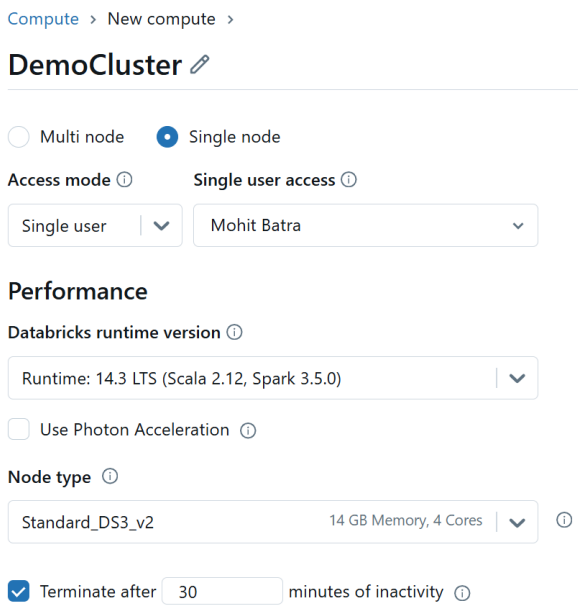
3. In the workspace, from left pane, go to Compute tab.



4. Click on Create Compute to create a cluster.

5. Fill up cluster properties as shown below, and click on Create Cluster. This will take few minutes to setup a single node cluster.

   [Note]: If you want to setup multi-node cluster, select multi node option from UI.

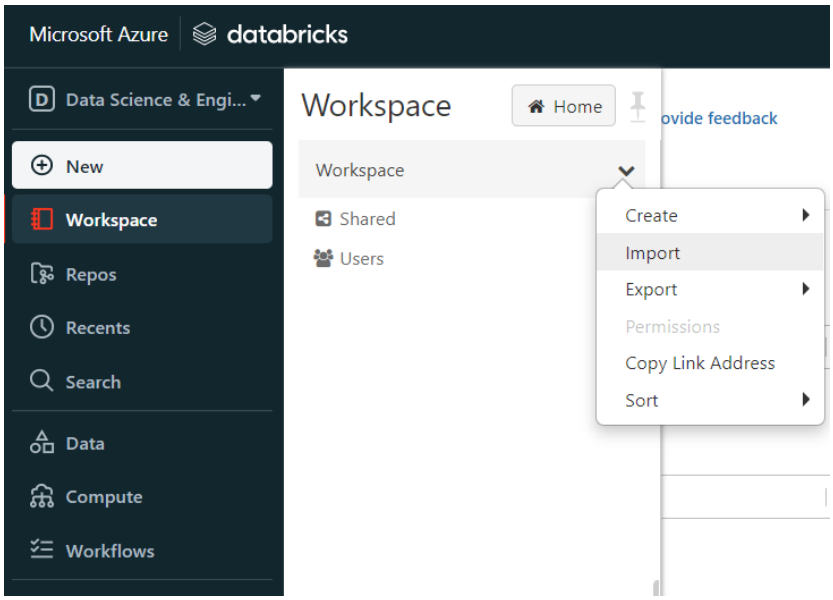


Exercise 3 – Import Notebook & Run Commands

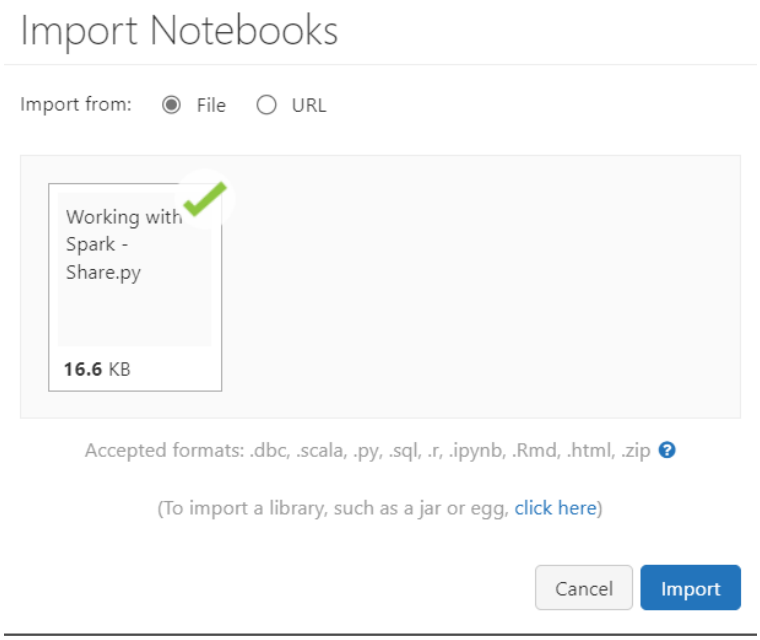
1. Download notebook – “Working with Spark.py” from Github repository.

2. Once cluster is ready, from left pane, go to Workspace tab.

3. Right-click in Workspace tab, and select Import



4. Upload the notebook - “Working with Spark.py” and click Import



5. Open the notebook and run commands 1 to 12.