

# 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

Create an Azure Databricks workspace ...

The screenshot shows the 'Create new workspace' form in the Azure portal, specifically the 'Basics' tab. The form is divided into two main sections: 'Project Details' and 'Instance Details'. In the 'Project Details' section, the 'Subscription' is set to 'MSDN Platforms' and the 'Resource group' is 'OReilly'. Below these, there is a 'Create new' link. In the 'Instance Details' section, the 'Workspace name' is 'MyDatabricksWorkspace', the 'Region' is 'East US 2', and the 'Pricing Tier' is 'Standard (Apache Spark, Secure with Azure AD)'. Each field has a dropdown arrow on the right, and the workspace name field has a green checkmark on the right.

Basics Networking Advanced Tags Review + create

Project Details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription \* ⓘ MSDN Platforms

Resource group \* ⓘ OReilly  
[Create new](#)

Instance Details

Workspace name \* MyDatabricksWorkspace ✓

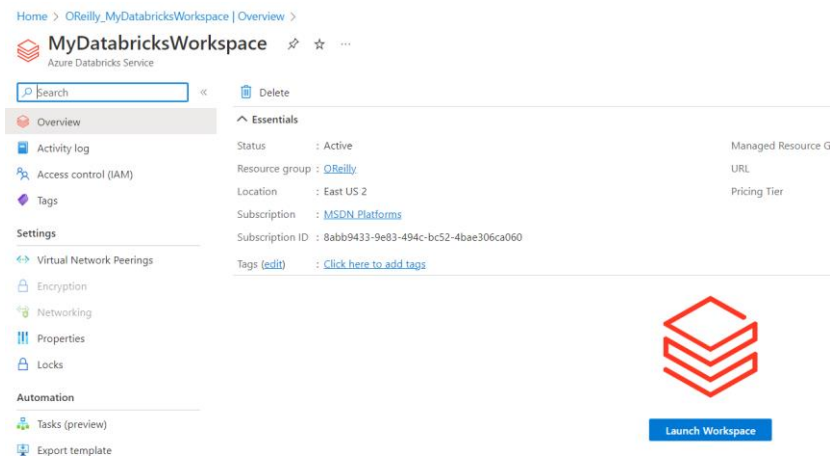
Region \* East US 2

Pricing Tier \* ⓘ Standard (Apache Spark, Secure with Azure AD)

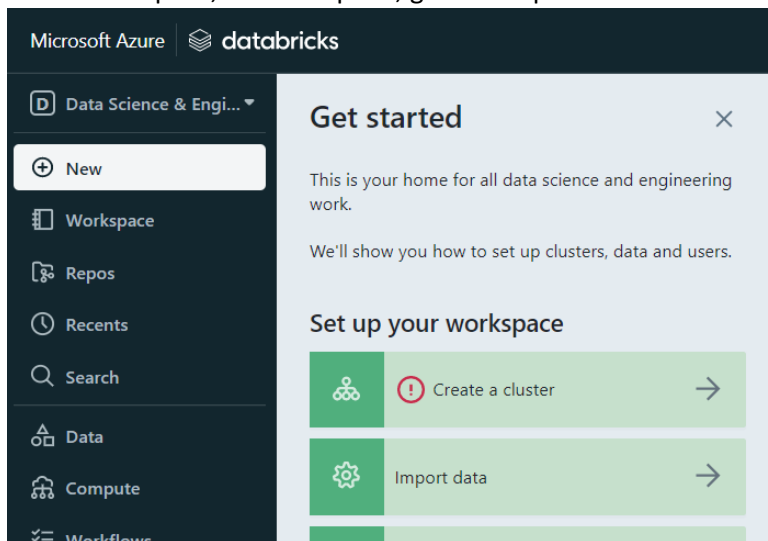
- b. Click Create
- c. This will take few minutes to create

## 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.*

Compute > New compute >

**DemoCluster**

☐ Multi node ☒ Single node

Access mode ⓘ

Single user access ⓘ

Single user

Mohit Batra

### Performance

Databricks runtime version ⓘ

Runtime: 14.3 LTS (Scala 2.12, Spark 3.5.0)

☐ Use Photon Acceleration ⓘ

Node type ⓘ

Standard\_DS3\_v2

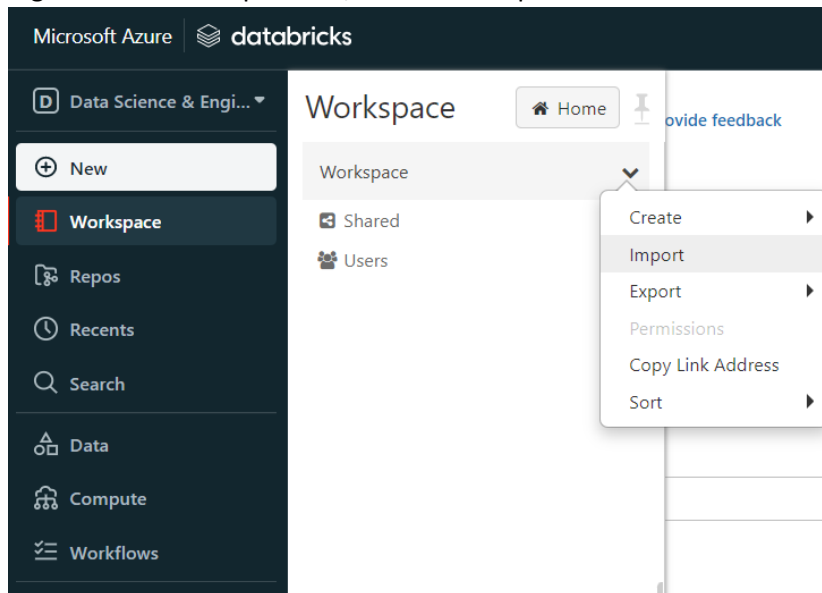
14 GB Memory, 4 Cores

ⓘ

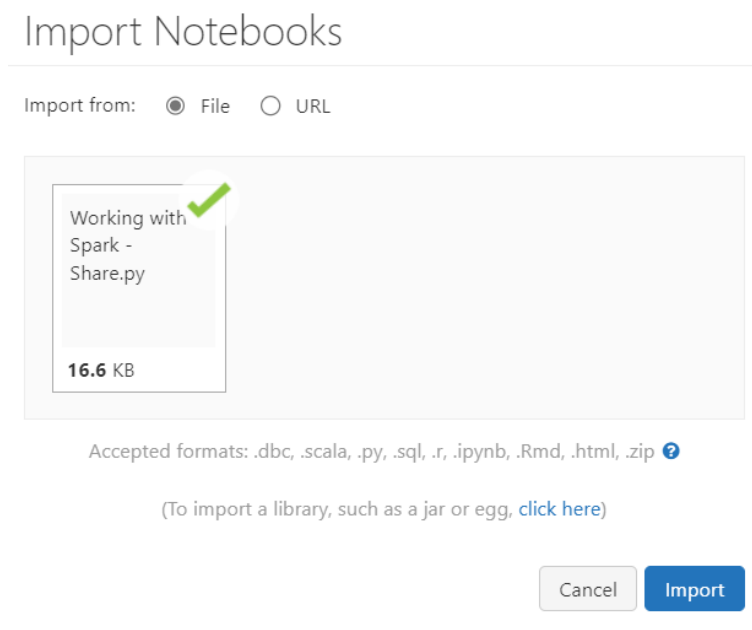
☒ Terminate after  minutes of inactivity ⓘ

## 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.