Tejashree G
tejashreeg.ai2021@dce.edu.in

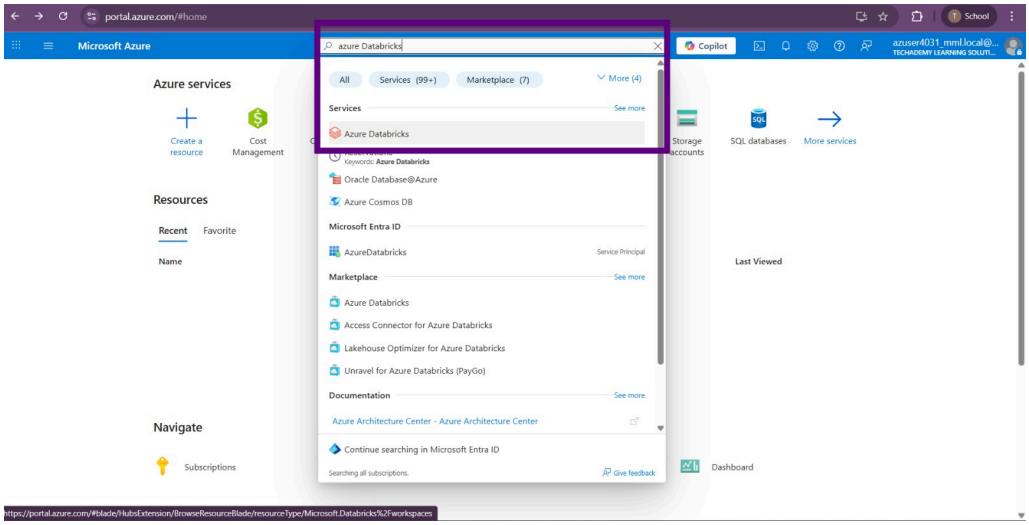
Azure Databricks Setup Assignment

This document outlines the process of setting up an Azure Databricks workspace, including cluster creation and configuration.

1. Create an Azure Databricks Workspace

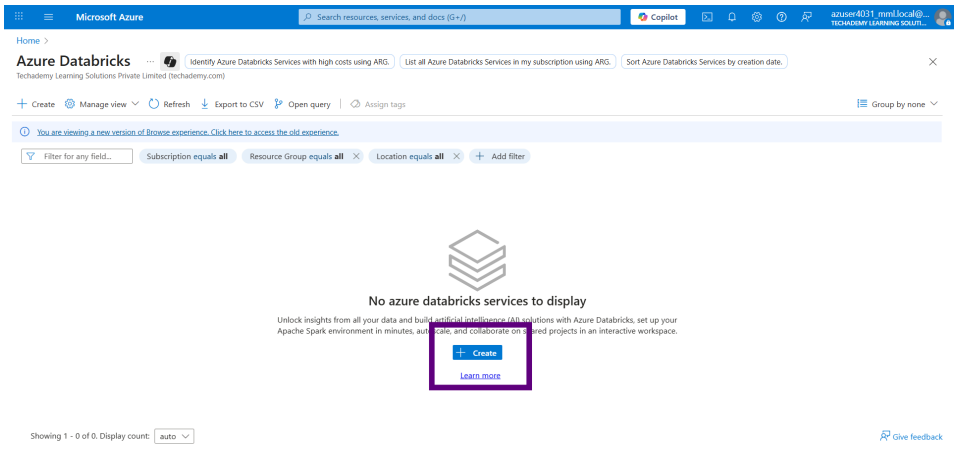
Step 1: Search for Azure Databricks

1. Log in to the Azure Portal.
2. In the search bar at the top, type "Azure Databricks" and select the "Azure Databricks" service from the results.



Step 2: Initiate Workspace Creation

On the Azure Databricks page, click on + Create Azure Databricks workspace.



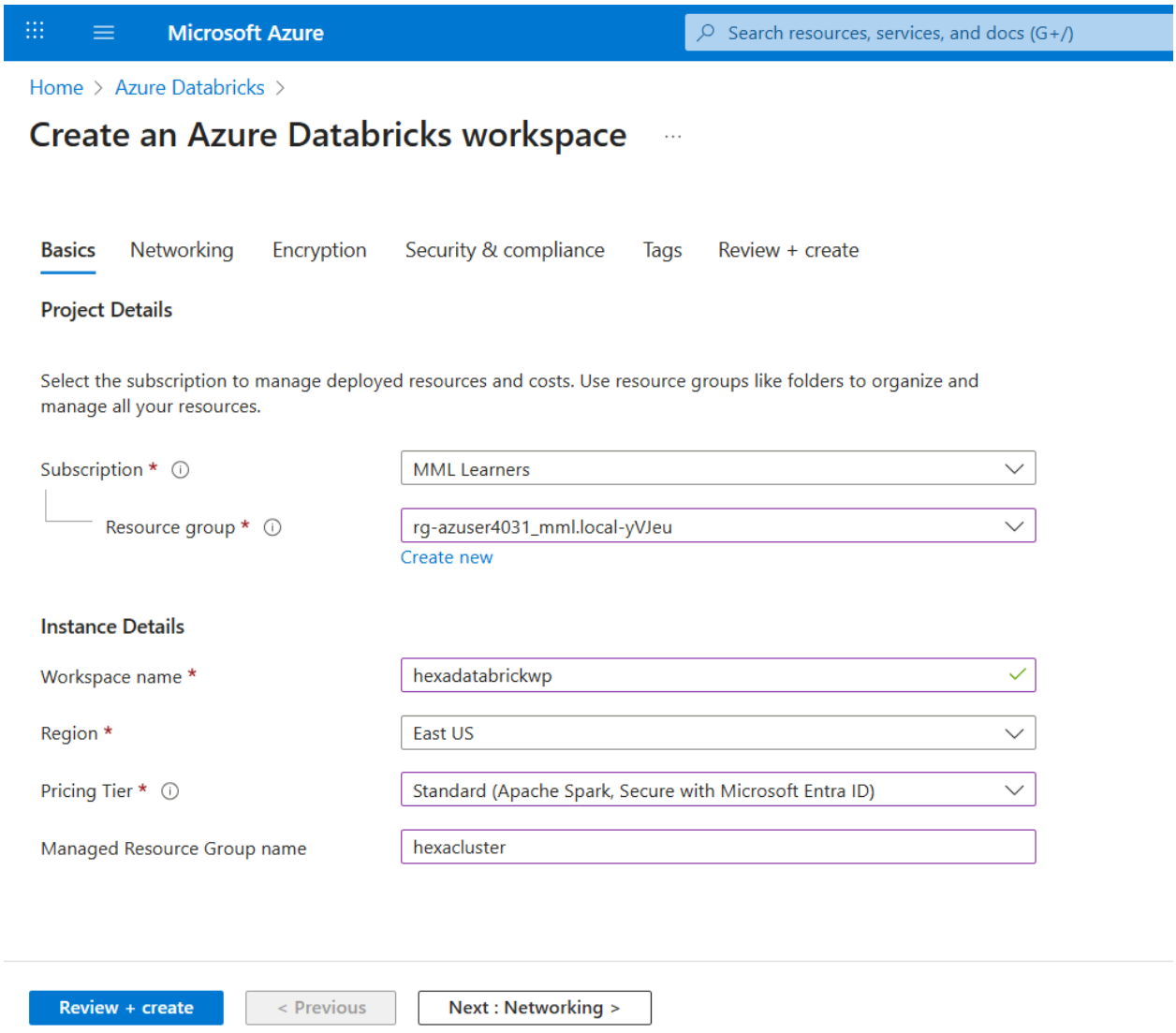
Step 3: Configure the Details

Project Details:

- Subscription: The Azure subscription is selected (e.g., "MML Learners").
- Resource group: An existing resource group is chosen.

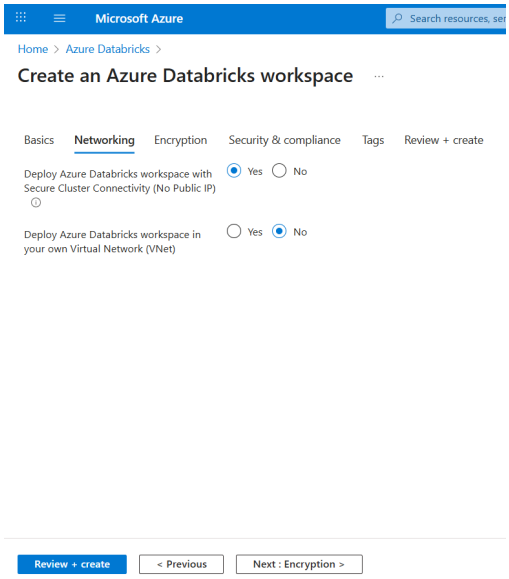
Instance Details:

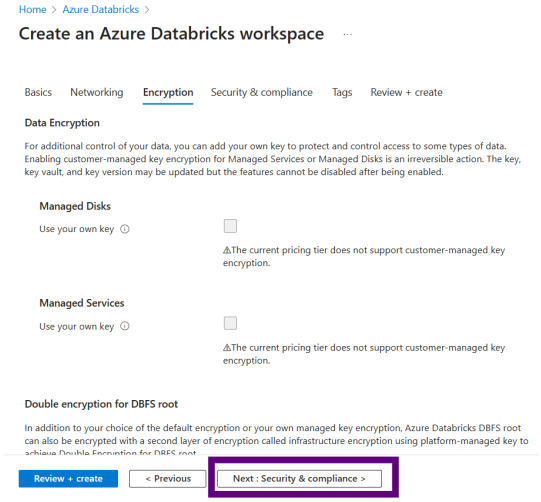
- Workspace name: A unique name for the Databricks workspace is entered (e.g., hexadatabrickwp).
- Region: The Azure region where the workspace is to be deployed is selected (e.g., "East US").
- Pricing Tier: "Standard" (Apache Spark, Secure with Microsoft Entra ID) is selected.
- Managed Resource Group name: This will be automatically populated

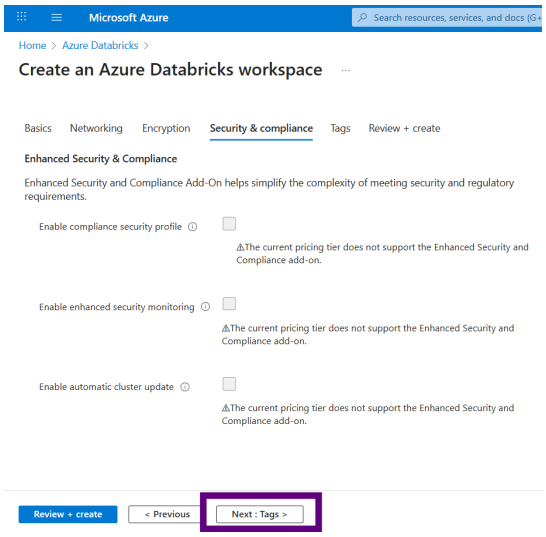


Step 4: Configure Networking, Encryption, Security & Compliance (Optional)

Skip this step and select next



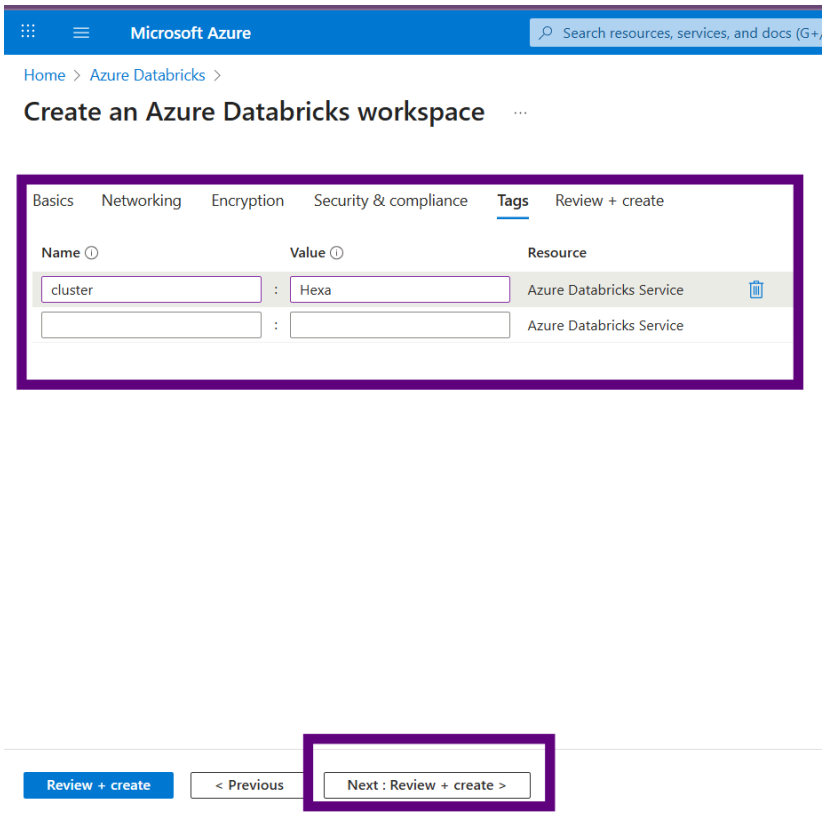




Step 5: Add Tags

1. Name: A tag name is entered (e.g., cluster).
2. Value: A corresponding value is entered (e.g., Hexa).

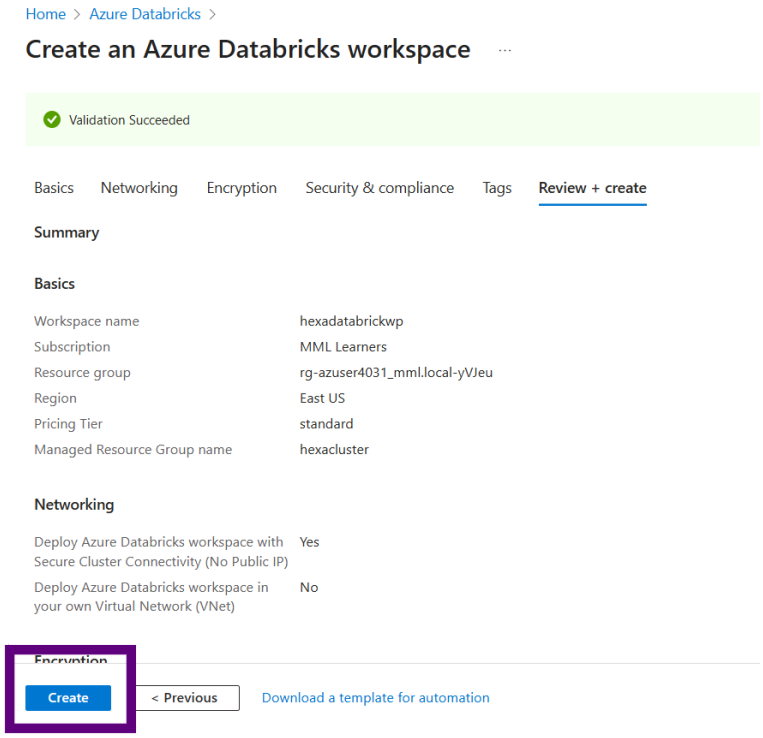
Tags are useful for organizing and managing Azure resources, especially for cost tracking.



Step 8: Review and Create

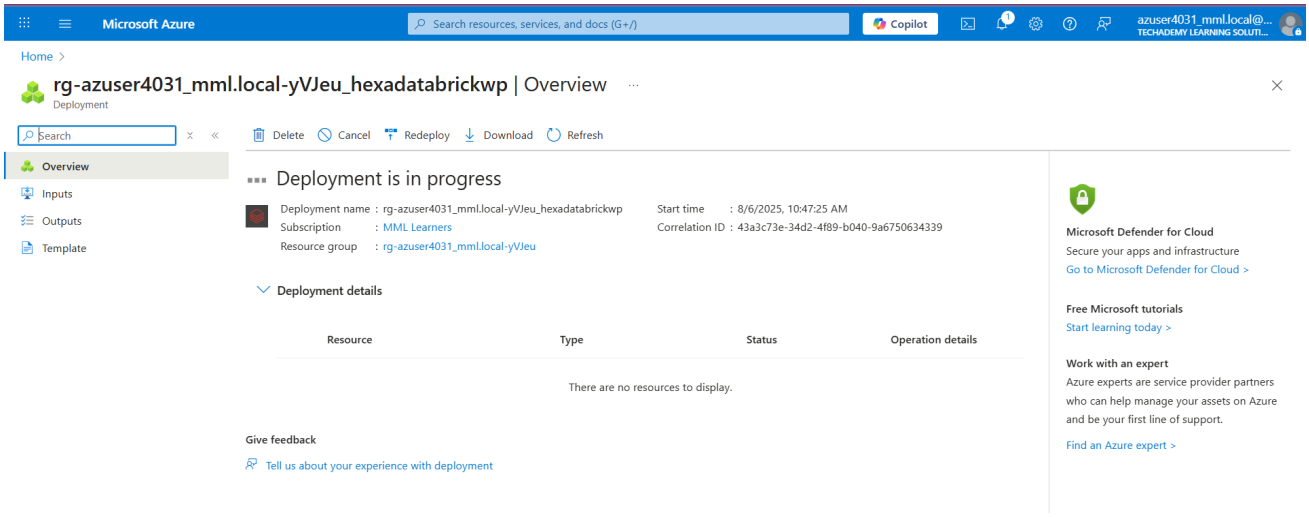
One navigates to the Review + create tab.

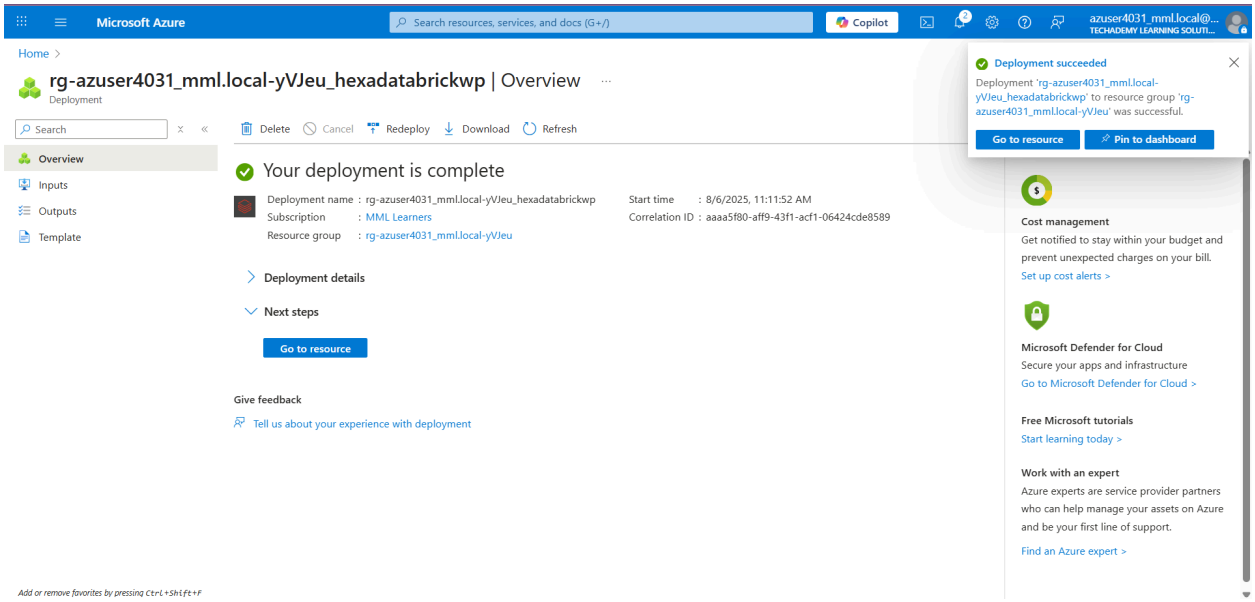
1. All configurations that have been made are reviewed. It is ensured that "Validation Succeeded" is displayed.
2. The Summary section is reviewed to confirm all details are correct.
3. Create is clicked to start the deployment of the Azure Databricks workspace.



Step 9: Monitor Deployment

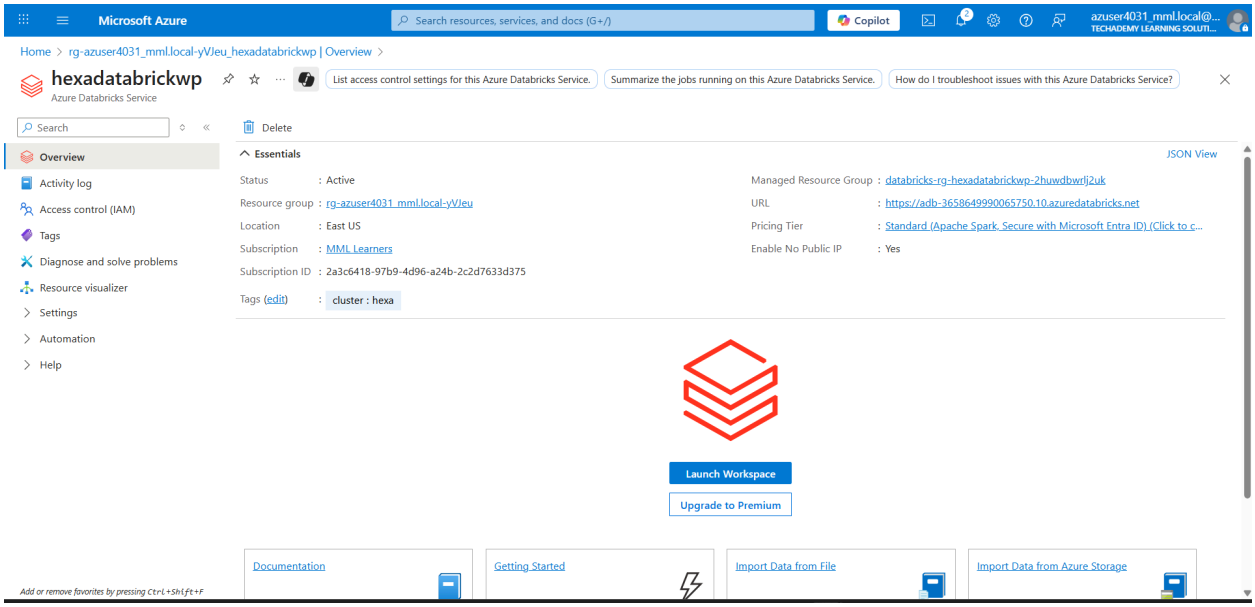
1. The Azure portal will display the deployment progress. "Deployment is in progress" will be seen.
2. Once the deployment is complete, the status will change to "Deployment succeeded."
3. Go to resource is clicked to navigate to the newly created Databricks workspace.

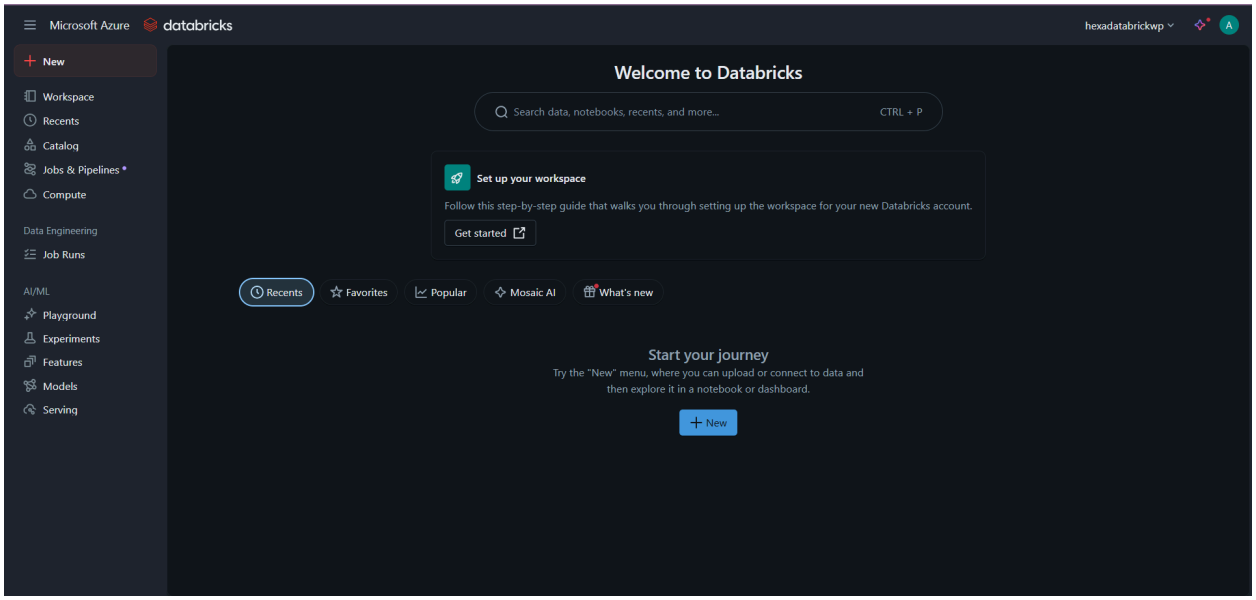




2. Launching the Databricks Workspace

1. From the Azure Databricks workspace overview page, the Launch Workspace button is clicked.
2. This will open the Databricks workspace in a new browser tab, where the "Welcome to Databricks" page will be seen.





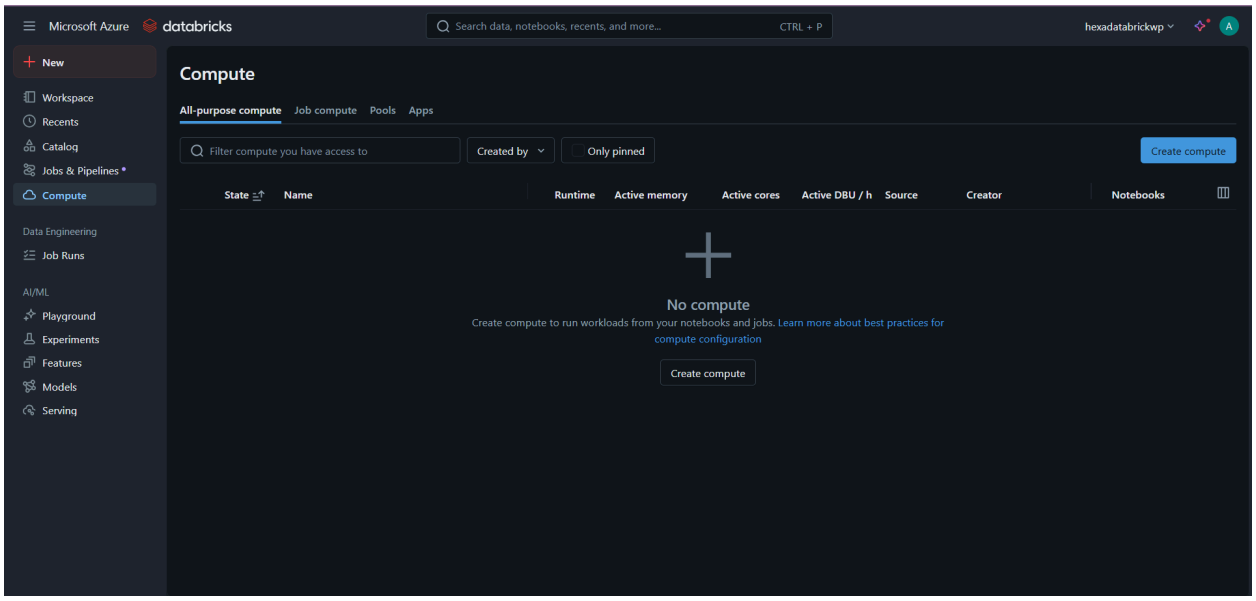
3. Create a New Compute Cluster

Step 1: Navigate to Compute

In the Databricks workspace, Compute is clicked in the left-hand navigation bar.

Step 2: Create New Compute

On the Compute page, the Create Compute button is clicked.



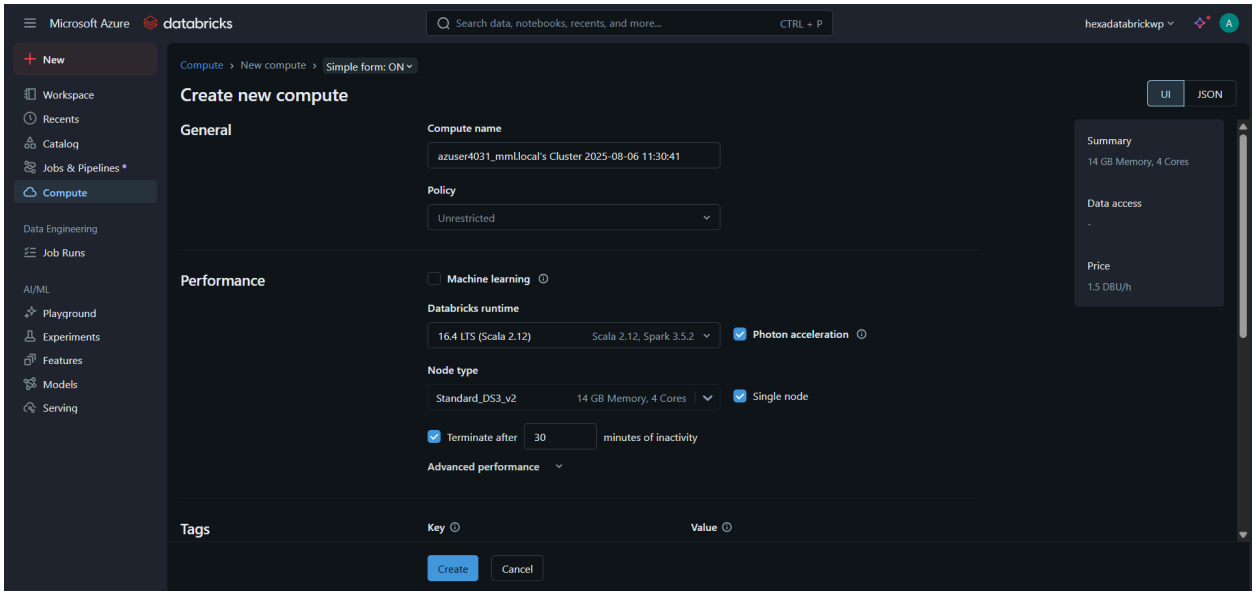
Step 3: Configure Cluster Detail

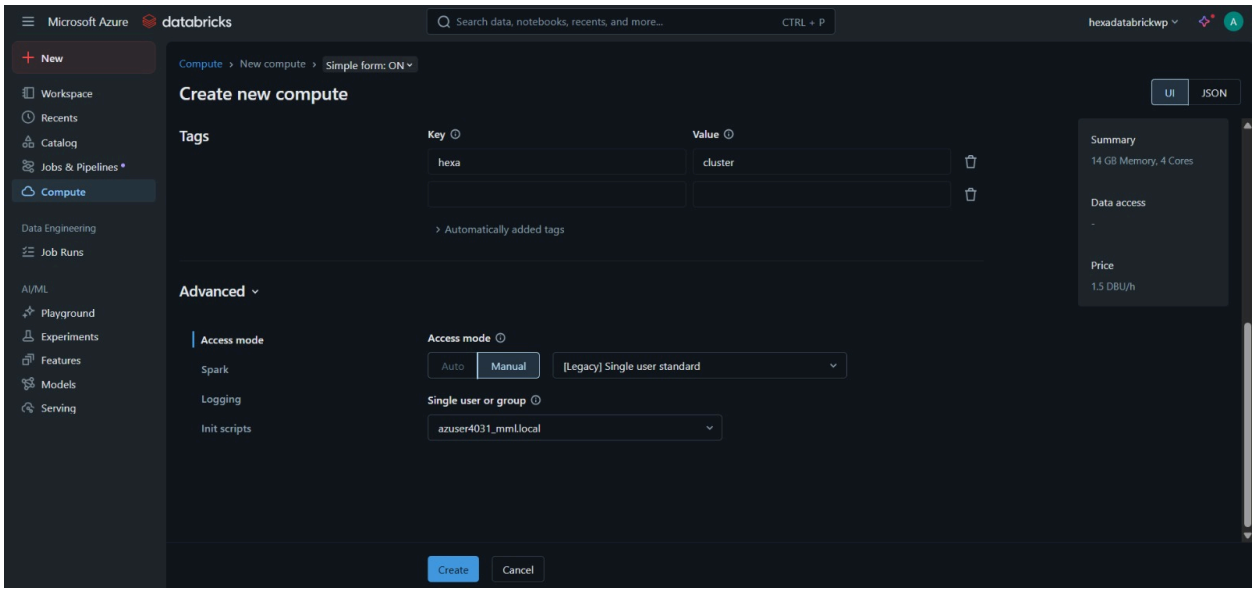
The user will be presented with the "Create new compute" form.

General:

- Compute name: A name for the cluster is provided
- Policy: (Optional) If policies are configured, one is selected.

- Databricks runtime version: The desired Databricks Runtime version is selected (e.g., 14.4 LTS (Scala 2.12, Spark 3.5.0)).
- Node type: The VM size for the worker nodes is chosen (e.g., Standard D532). Cost considerations are important here.
- Terminate after X minutes of inactivity: A reasonable inactivity period after which the cluster will automatically terminate is set (e.g., 30 minutes). This is crucial for cost management.
- Photon acceleration: (Optional) This box is checked if Photon acceleration is desired for faster query performance.
- Single node: (Optional) This box is checked if a single-node cluster is desired





Step 5: Create Cluster

Create compute is clicked to provision the Databricks cluster.

The Azure Databricks workspace and a compute cluster are now set up.