

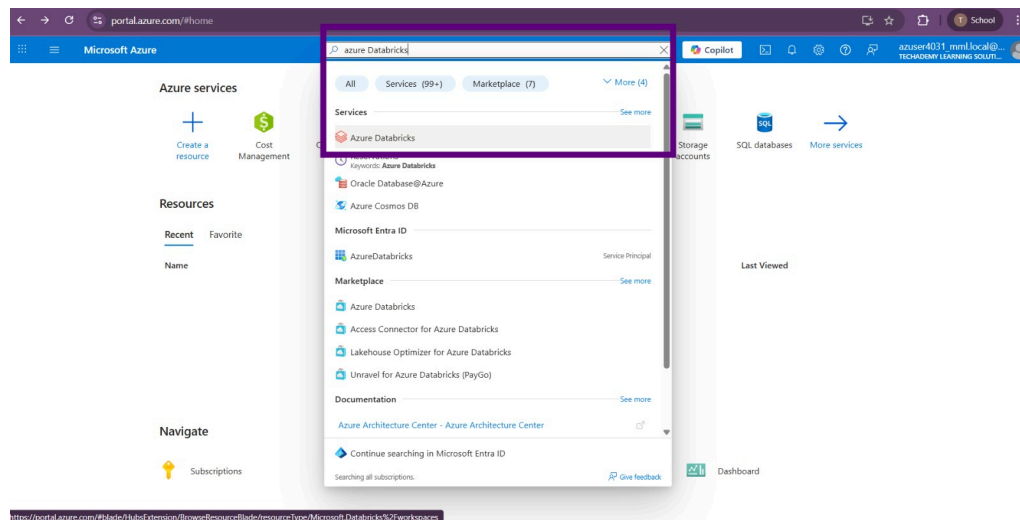
# Azure Databricks Setup Assignment

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

## 1. Create an Azure Databricks Workspace

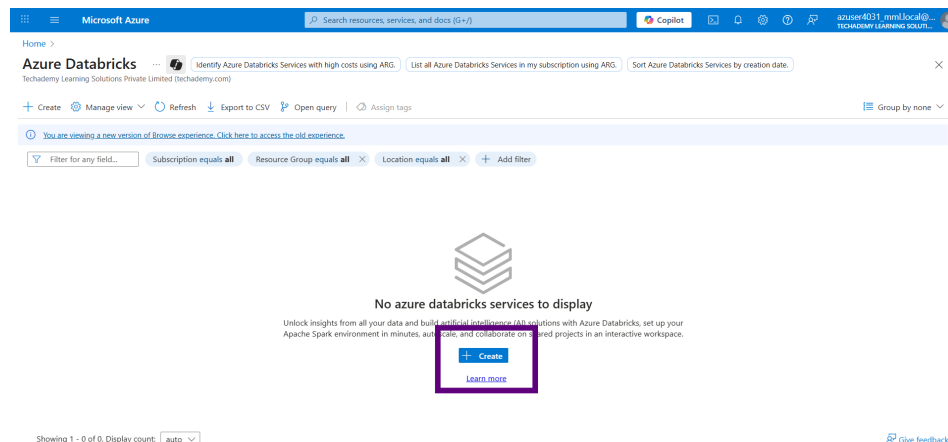
### Step 1: Search for Azure Databricks

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

Microsoft Azure

Search resources, services, and docs (G+)

[Home](#) > [Azure Databricks](#) >

## Create an Azure Databricks workspace ...

Basics

Networking

Encryption

Security & compliance

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 \* ⓘ

MML Learners

Resource group \* ⓘ

rg-azuser4031\_mml.local-yVJeu

[Create new](#)

### Instance Details

Workspace name \*

hexadatabrickwp

Region \*

East US

Pricing Tier \* ⓘ

Standard (Apache Spark, Secure with Microsoft Entra ID)

Managed Resource Group name

hexacluster

Review + create

< Previous

Next : Networking >

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

Skip this step and select next

Home > Azure Databricks >

### Create an Azure Databricks workspace

Basics **Networking** Encryption Security & compliance Tags Review + create

Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP) ☒ Yes ☐ No

Deploy Azure Databricks workspace in your own Virtual Network (VNet) ☐ Yes ☒ No

Review + create

< Previous

Next : Encryption >

Home > Azure Databricks >

### Create an Azure Databricks workspace

Basics Networking **Encryption** Security & compliance Tags Review + create

#### Data Encryption

For additional control of your data, you can add your own key to protect and control access to some types of data. Enabling customer-managed key encryption for Managed Services or Managed Disks is an irreversible action. The key, key vault, and key version may be updated but the features cannot be disabled after being enabled.

##### Managed Disks

Use your own key

☐

⚠The current pricing tier does not support customer-managed key encryption.

##### Managed Services

Use your own key

☐

⚠The current pricing tier does not support customer-managed key encryption.

#### Double encryption for DBFS root

In addition to your choice of the default encryption or your own managed key encryption, Azure Databricks DBFS root can also be encrypted with a second layer of encryption called infrastructure encryption using platform-managed key to achieve Double Encryption for DBFS root.

Review + create

< Previous

Next : Security & compliance >

Microsoft Azure Search resources, services, and docs (G+)

Home > Azure Databricks >

### Create an Azure Databricks workspace ...

Basics Networking Encryption **Security & compliance** Tags Review + create

#### Enhanced Security & Compliance

Enhanced Security and Compliance Add-On helps simplify the complexity of meeting security and regulatory requirements.

Enable compliance security profile ⓘ ☐  
⚠The current pricing tier does not support the Enhanced Security and Compliance add-on.

Enable enhanced security monitoring ⓘ ☐  
⚠The current pricing tier does not support the Enhanced Security and Compliance add-on.

Enable automatic cluster update ⓘ ☐  
⚠The current pricing tier does not support the Enhanced Security and Compliance add-on.

Review + create < Previous **Next : Tags >**

## 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.

Microsoft Azure Search resources, services, and docs (G+)

Home > Azure Databricks >

### Create an Azure Databricks workspace ...

Basics Networking Encryption Security & compliance **Tags** Review + create

Name ⓘ	Value ⓘ	Resource
cluster	Hexa	Azure Databricks Service 
		Azure Databricks Service

Review + create < Previous **Next : Review + create >**


## 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.

[Home](#) > [Azure Databricks](#) >

## Create an Azure Databricks workspace ...

 Validation Succeeded

Basics
Networking
Encryption
Security & compliance
Tags
Review + create

### Summary

#### Basics

Workspace name	hexadatabrickwp
Subscription	MML Learners
Resource group	rg-azuser4031_mml.local-yVJeu
Region	East US
Pricing Tier	standard
Managed Resource Group name	hexacluster

#### Networking

Deploy Azure Databricks workspace with	Yes
Secure Cluster Connectivity (No Public IP)	
Deploy Azure Databricks workspace in your own Virtual Network (VNet)	No

**Encryption**

Create

[< Previous](#)
[Download a template for automation](#)

## 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.

Microsoft Azure
Search resources, services, and docs (G+)

[Home](#) >

**rg-azuser4031\_mml.local-yVJeu\_hexadatabrickwp** | Overview ...

Deployment

Delete Cancel Redeploy Download Refresh


Overview

Inputs

Outputs

Template

### Deployment is in progress

 Deployment name : rg-azuser4031\_mml.local-yVJeu\_hexadatabrickwp

Subscription : MML Learners

Resource group : rg-azuser4031\_mml.local-yVJeu

Start time : 8/6/2025, 10:47:25 AM


Correlation ID : 43a3c73e-34d2-4f89-b040-9a6750634339

Deployment details

Resource	Type	Status	Operation details
There are no resources to display.			

Give feedback

[Tell us about your experience with deployment](#)

 **Microsoft Defender for Cloud**

Secure your apps and infrastructure

[Go to Microsoft Defender for Cloud >](#)

**Free Microsoft tutorials**

[Start learning today >](#)

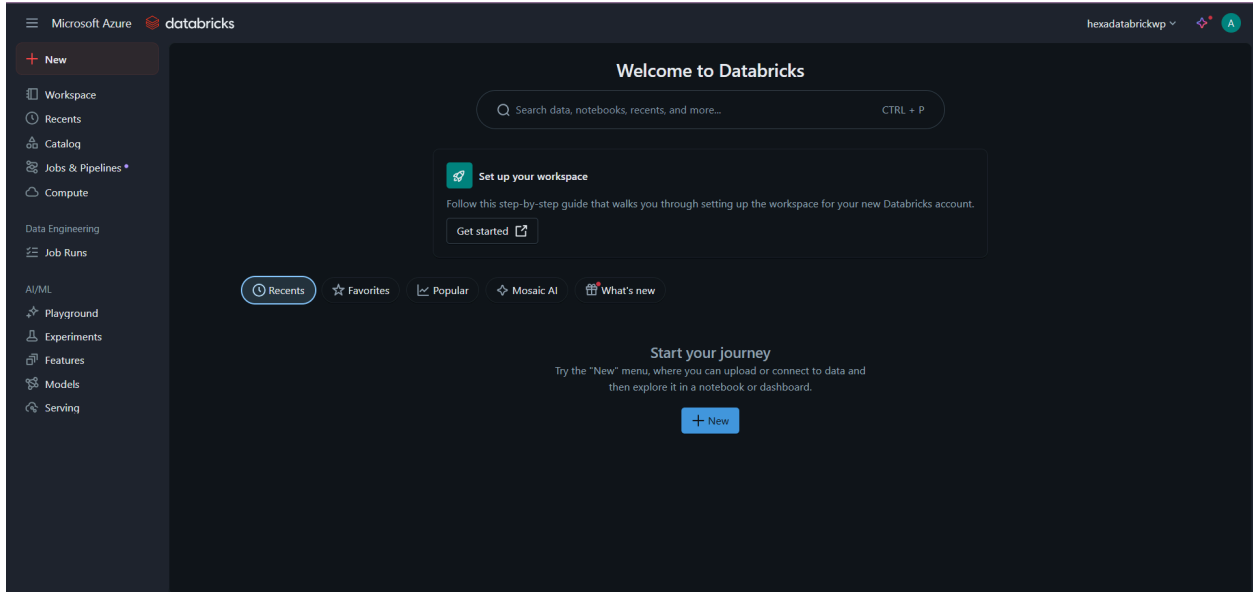
**Work with an expert**

Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.

[Find an Azure expert >](#)

## 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.



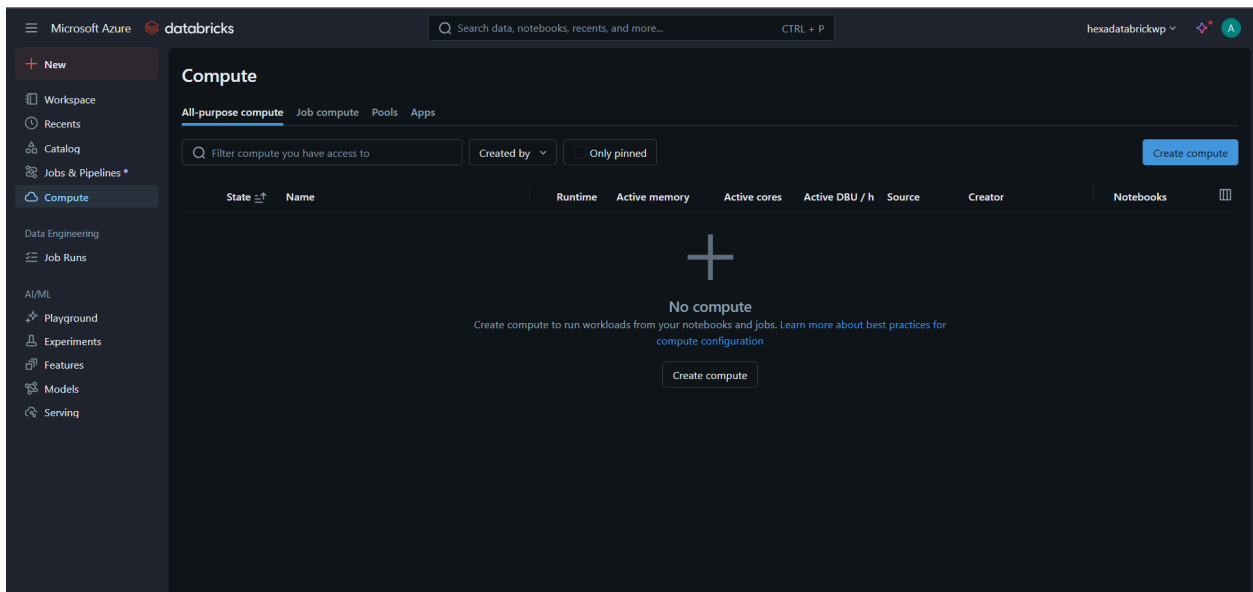
### 3. Create a New Compute Cluster

#### 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

This screenshot shows the 'Create new compute' form in the Databricks interface, specifically the 'General' tab. The form is titled 'Create new compute' and has a breadcrumb trail: 'Compute > New compute > Simple form: ON'. The 'Compute name' field is populated with 'azuser4031\_mmllocal's Cluster 2025-08-06 11:30:41'. The 'Policy' dropdown is set to 'Unrestricted'. In the 'Performance' section, the 'Machine learning' checkbox is unchecked. The 'Databricks runtime' dropdown is set to '16.4 LTS (Scala 2.12)', and the 'Node type' dropdown is set to 'Standard\_DS3\_v2'. The 'Photon acceleration' checkbox is checked. The 'Single node' checkbox is also checked. The 'Terminate after' field is set to '30 minutes of inactivity'. The 'Advanced performance' dropdown is set to 'Standard'. The 'Tags' section is empty. The 'Create' button is highlighted in blue.

This screenshot shows the 'Create new compute' form in the Databricks interface, specifically the 'Advanced' tab. The form is titled 'Create new compute' and has a breadcrumb trail: 'Compute > New compute > Simple form: ON'. The 'Tags' section shows a table with two columns: 'Key' and 'Value'. The first row has 'hexa' as the key and 'cluster' as the value. The 'Advanced' section is expanded, showing the 'Access mode' dropdown set to 'Auto', the 'Single user or group' dropdown set to 'azuser4031\_mmllocal', and the 'Logging' checkbox checked. The 'Create' button is highlighted in blue.



## 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.

The screenshot displays the 'Create Cluster' configuration interface in the Azure Databricks workspace. The interface is divided into several sections:

- General:** Contains the 'Compute name' field with the value 'azuser4031\_mmml.local's Cluster 2025-08-06 11:30:41' and a 'Policy' dropdown set to 'Unrestricted'.
- Performance:** Includes a 'Machine learning' checkbox (unchecked), a 'Databricks runtime' dropdown set to '16.4 LTS (includes Apache Spark 3.5.2, Scala 2.12)', and a 'Node type' dropdown set to 'Standard\_DS3\_v2' (14 GB Memory, 4 Cores). It also features a 'Photon acceleration' checkbox (checked) and a 'Single node' checkbox (checked). A 'Terminate after' field is set to '30 minutes of inactivity'.
- Summary:** A sidebar on the right showing '14 GB Memory, 4 Cores' and 'Price 1.5 DBU/h'.
- Tags:** A section at the bottom for adding tags, with columns for 'Key' and 'Value'.

The top navigation bar includes the Microsoft Azure logo, the Databricks logo, a search bar, and a user profile icon. The left sidebar contains navigation links for 'New', 'Workspace', 'Catalog', 'Jobs & Pipelines', 'Compute', 'Data Engineering', 'Job Runs', 'AI/ML', 'Playground', 'Experiments', 'Features', 'Models', and 'Serving'.