

Databricks

-Mohammad R Islam

Steps to Create a Databricks Workspace

1) Login to Azure

- a) Azure portal: [Link](#)
- b) Sign in with your Azure credentials.

2) Create Azure Databricks

a) Configure Basic Settings

- i) **Subscription:** Select the Azure subscription in which you want to create the workspace.
- ii) **Resource Group:** Choose an existing resource group or create a new one.
- iii) **Workspace Name:** Enter a unique name for your Databricks workspace.
- iv) **Region:** Select the Azure region where the workspace will be hosted (choose a region close to your data sources).

[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 *	<input type="text" value="AzureTraining"/>
Resource group *	<input type="text" value="BigData"/> Create new

Instance Details

Workspace name *	<input type="text" value="Sampletest"/>
Region *	<input type="text" value="UK South"/>
Pricing Tier *	<input type="text" value="Premium (+ Role-based access controls)"/> Standard (Apache Spark, Secure with Microsoft Entra ID) Premium (+ Role-based access controls) Trial (Premium - 14-Days Free DBUs)
Managed Resource Group name	

[Review + create](#)

[< Previous](#)

[Next : Networking >](#)

b) Networking

- Deploy Azure Databricks workspace with secure cluster connectivity (No public IP) – **Yes/No**
- Deploy Azure Databricks workspace with you own Virtual network (VNet) – **Yes/No**
- If you choose a **VNet**, you will need to specify the following:
 - Private Subnet** with CIDR range within the chosen VNet.
 - Public Access:** with CIDR range within the chosen VNet.

[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

Virtual Network * ⓘ

Two new subnets will be created in your Virtual Network

Implicit delegation of both subnets will be done to Azure Databricks on your behalf

Public Subnet Name *

Public Subnet CIDR Range * ⓘ

Private Subnet Name *

Private Subnet CIDR Range * ⓘ

[Review + create](#)

[< Previous](#)

[Next : Encryption >](#)

c) Encryption

- i) **Encryption at Rest:** Azure Databricks automatically encrypts data at rest. You can choose to use Azure-managed keys or **your own** customer-managed keys (**CMK**).

(1) If using a **CMK**, specify the Key Vault and the key you want to use for encryption.

- ii) **Configure DBFS Encryption:** Databricks will use the specified CMK for encrypting the data in DBFS. You do not need to make additional configurations to enable this, as Databricks will automatically encrypt the data stored in DBFS using the specified key.

- iii) **Encryption in Transit:** Databricks uses TLS/SSL for encrypting data in transit by default.

[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.

Enable Infrastructure Encryption ⓘ

☐

⚠The current pricing tier does not support infrastructure encryption.

[Review + create](#)

[< Previous](#)

[Next : Security & compliance >](#)

d) Workspace Security and compliance Settings

- i) **Enable Compliance security profile**

Databricks does not have a specific "Compliance Security Provider," you can enhance compliance in your Databricks environment by enabling audit logging, integrating with external compliance and security solutions, configuring access controls, setting up encryption, enabling IP access lists, and monitoring compliance metrics. This multi-layered approach will help you meet various regulatory requirements and maintain a secure environment.

- ii) **Enable Enhanced security monitoring**

Enabling Enhanced Security Monitoring in Databricks involves configuring audit logs, access control, IP access lists, encryption, user activity monitoring, and integration with external security solutions. These steps help ensure that your Databricks environment is secure and compliant with best practices for data protection and security monitoring.

iii) Enable automatic cluster update

Enabling automatic cluster updates in Databricks helps ensure that your clusters are always running the latest version of Databricks Runtime, which includes performance improvements, security updates, and new features.

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 >

e) Review and Create

- i) Once you have filled in all the necessary fields, review your settings on the **Review + create** tab.
- ii) Check the relevant summary for any errors or required fields.
- iii) Click the **Create** button to deploy your Databricks workspace.

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

Create an Azure Databricks workspace ...

✖ Validation failed. Required information is missing or not valid on the Networking tab.

Basics ✖ Networking Encryption Security & compliance Tags Review + create

Summary

Basics

Workspace name	Sampletest
Subscription	AzureTraining
Resource group	BigData
Region	UK South
Pricing Tier	standard
Managed Resource Group name	databricksTest

Networking

Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP)	Yes
Deploy Azure Databricks workspace in your own Virtual Network (VNet)	Yes
Virtual Network	Not Selected Yet
Public Subnet Name	Not Added Yet
Public Subnet CIDR Range	Not Added Yet
Private Subnet Name	Not Added Yet
Private Subnet CIDR Range	Not Added Yet

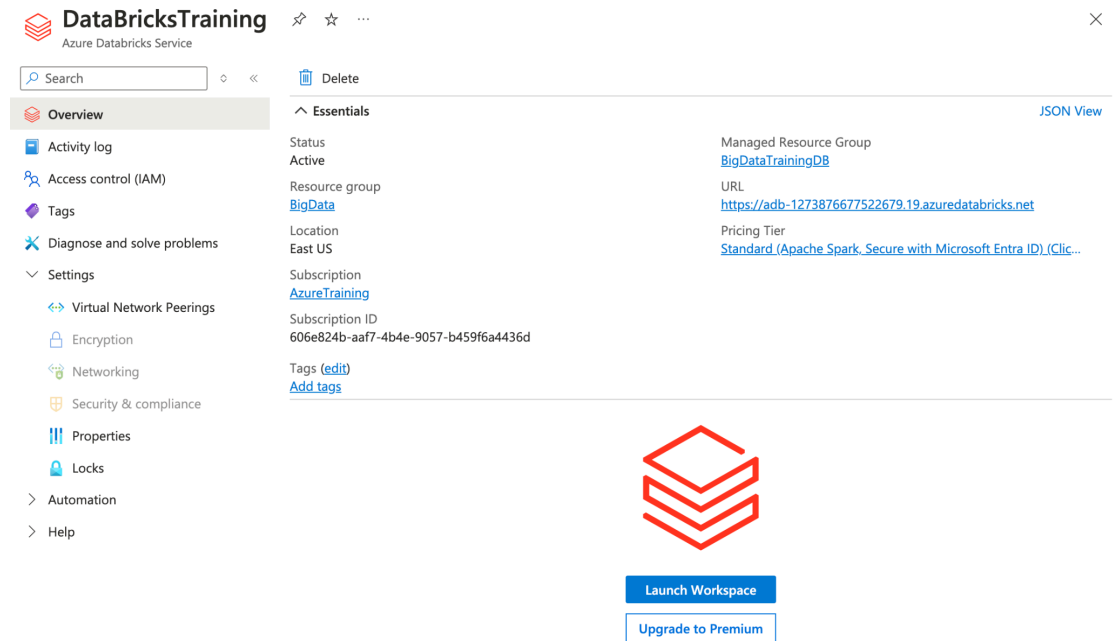
Create

< Previous

f) Accessing the Workspace

- i) After the deployment is complete, go to the resource group where your Databricks workspace is located.

- ii) Click on the Databricks workspace resource.
- iii) You can now launch the workspace by clicking on the **Launch Workspace** button.



Databricks workspace

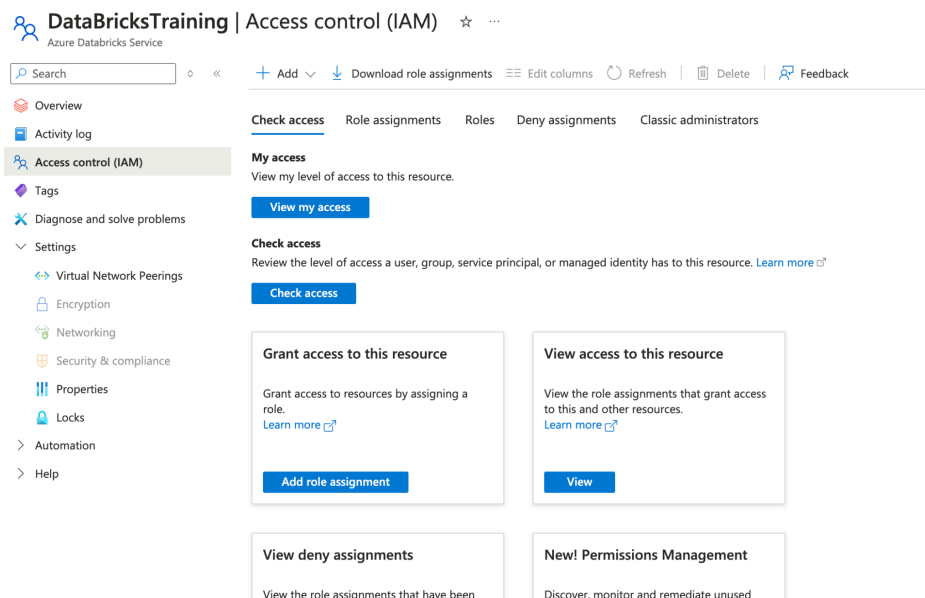
1. Activity logs

Databricks activity logs are crucial for tracking and auditing various events and activities occurring within your Databricks workspace. These logs help with monitoring for security, operational efficiency, and compliance purposes. You can configure activity logging and access the logs via Azure, AWS, or your preferred cloud provider, depending on your Databricks environment.

2. Access Control (IAM)

Access control in Databricks is essential for managing who can access your workspace, clusters, data, and other resources. Databricks provides various levels of access control mechanisms to ensure security and compliance, including

- Workspace access control
- cluster access control
- Job access control
- Notebook permissions, and
- Data access control.



3. Diagnose and solve problems

Diagnosing and solving problems in Databricks involves a combination of monitoring logs, debugging code, optimizing performance, and managing resources effectively.

- Cluster-Related Issues
- Job and Notebook Issues
- Performance Tuning
- Data Access and Storage Issues

