Databricks Workspace
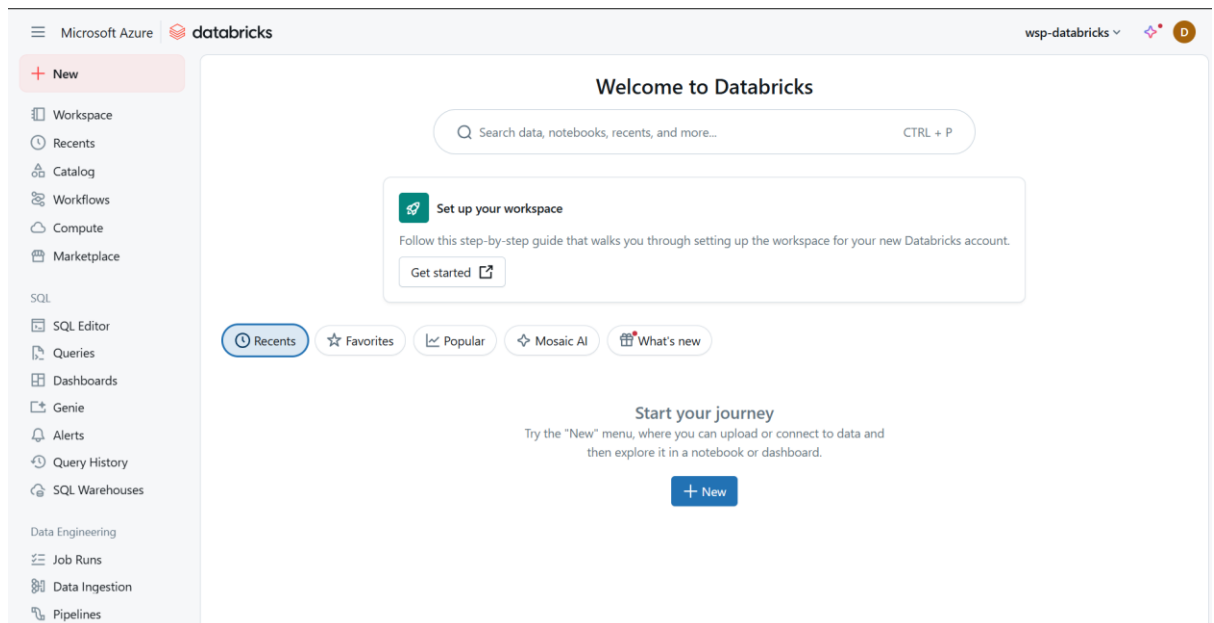
1. Creating Databricks workspace:

Click on resource create and search for databricks, open-> create

Give resource group name, workspace name, select region and for pricing tier select premium

Then review create and create and launch the work space

Along with data bricks there are other resources which will be created

Name ↑↓	Type ↑↓	Resource group ↑↓	Location ↑↓	Subscription ↑↓	
☐ 🔑 dbmanagedidentity	Managed Identity	databricks-rg-wsp-databricks-d…	East US	Azure subscription 1	⋯
☐ 🟩 dbstoragenl2pvwjvxuhpo	Storage account	databricks-rg-wsp-databricks-d…	East US	Azure subscription 1	⋯
☐ 🔴 unity-catalog-access-connector	Access Connector for Azure Dat…	databricks-rg-wsp-databricks-d…	East US	Azure subscription 1	⋯
☐ 🛡 workers-sg	Network security group	databricks-rg-wsp-databricks-d…	East US	Azure subscription 1	⋯
☐ ◈ workers-vnet	Virtual network	databricks-rg-wsp-databricks-d…	East US	Azure subscription 1	⋯

Storage account stores workspace files, logs, and datasets.

Managed Identity: Secure authentication for Databricks without credentials.

Access Connector: Manages secure access to Unity Catalog storage.

Network Security Group (NSG): Controls network traffic for worker nodes.

Virtual Network (VNet) : Enables secure communication between Databricks components.

1. dbmanagedidentity (Managed Identity)

Purpose: A Managed Identity is used for secure authentication without managing secrets or credentials.

Function in Databricks: It allows Databricks to authenticate to Azure services like Storage Accounts, Key Vault, and other resources securely.

How it Works: Instead of using manual credentials, this identity is assigned permissions via Azure Role-Based Access Control (RBAC).

2. dbstoragenl2pvwjvxuhpo (Storage Account)

Purpose: This is the Azure Storage Account automatically provisioned for Databricks File System (DBFS).

Function in Databricks:

Stores logs, libraries, and cluster-related data.

Can be used as a workspace data lake for raw datasets.

Integration: Typically linked with Azure Data Lake Storage (ADLS) Gen2 for structured data storage.

3. unity-catalog-access-connector (Access Connector for Unity Catalog)

Purpose: This is a managed identity-based access connector that allows Databricks Unity Catalog to interact securely with storage.

Function in Databricks:

Manages fine-grained access control for structured and unstructured data.

Allows table-level access control in Unity Catalog.

Key Benefit: Securely connects Databricks to Azure Storage without needing credentials.

4. workers-sg (Network Security Group - NSG)

Purpose: A Network Security Group (NSG) controls inbound and outbound traffic to virtual machines and services.

Function in Databricks:

Secures worker nodes in the Databricks cluster.

Defines rules to allow or deny network traffic between components.

Security Importance: Ensures only authorized communication between Databricks, storage, and other Azure services.

5. workers-vnet (Virtual Network - VNet)

Purpose: A Virtual Network (VNet) is created to enable secure communication between Databricks components.

Function in Databricks:

Provides network isolation for Databricks clusters.

Ensures connectivity between control plane and data plane.

Why It's Needed: Required to enable private link, secure access to Azure Data Lake, and restrict internet exposure.