How to setup Databricks workspace with Unity Catalog

Contents

Databricks Setup

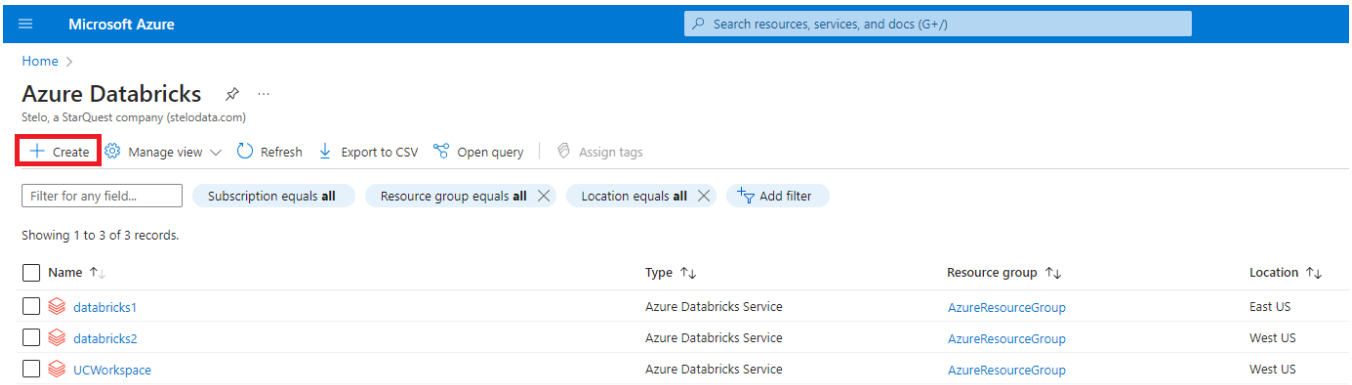
General Notes and Requirements

- (**Required**) All Databricks resources that will be created should all reside in the same region.
- (**Required**) Before setting up Databricks with Unity Catalog support, double check and ensure which regions support Databricks using the link down below:

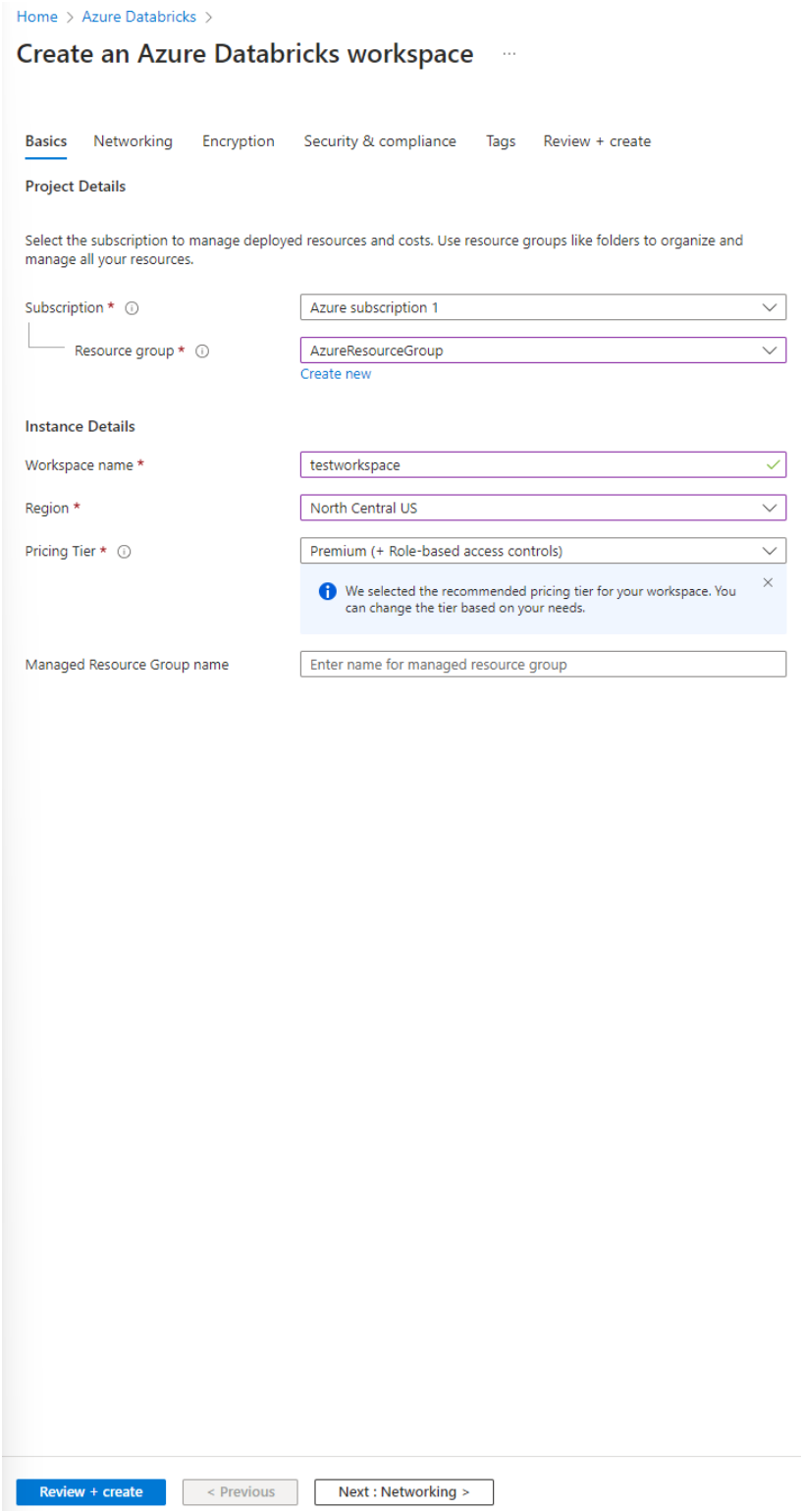
  https://learn.microsoft.com/en-us/azure/databricks/resources/supported-regions

Create a Databricks workspace

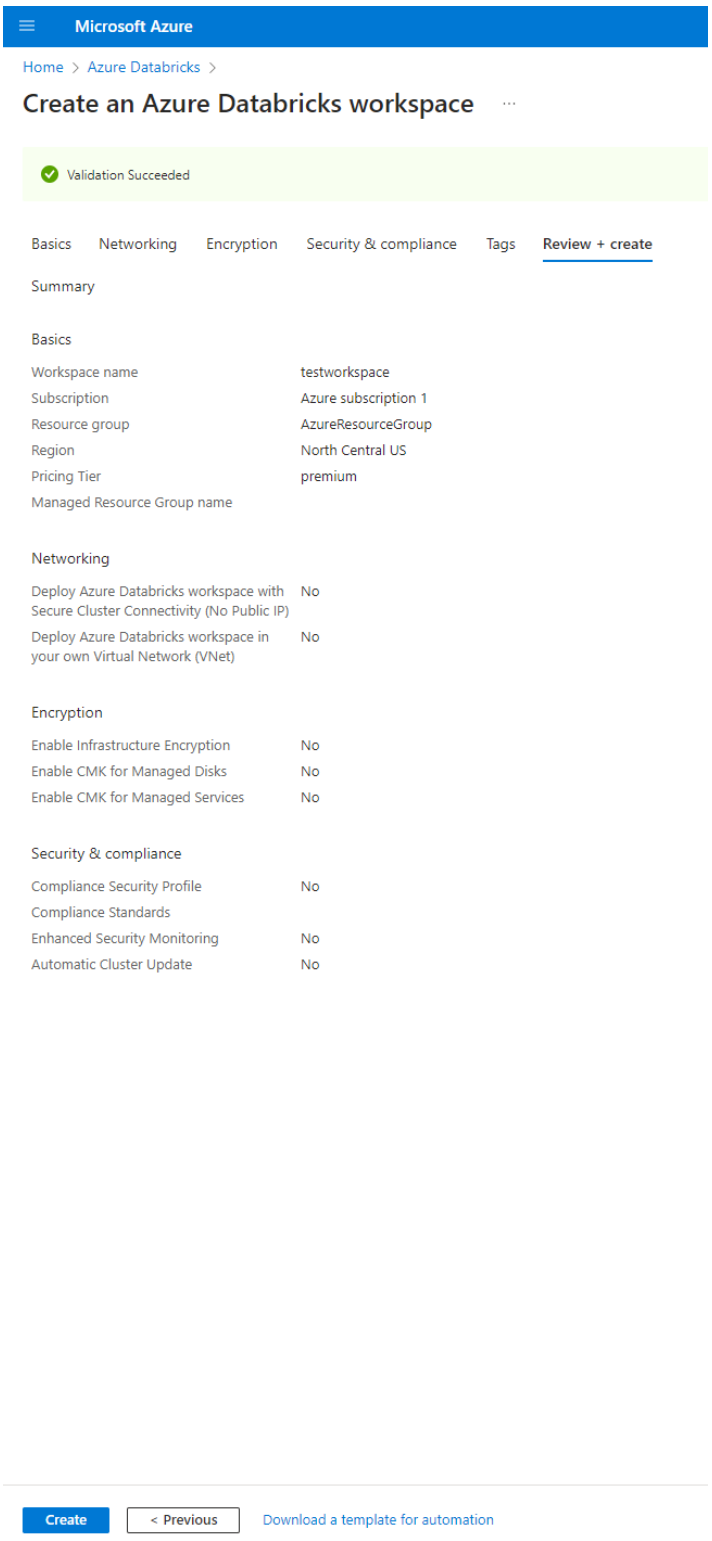
1) In the Azure Portal, select **Create**.



2) In the current page, select and/or fill in the following resources (note that the fields with * are mandatory). After filling in the necessary fields, select **Review + create**.



3) Select **Create**.



4) If Unity Catalog is not required, skip to **Create a Databricks cluster**.

Create an Access Connector for the Databricks workspace

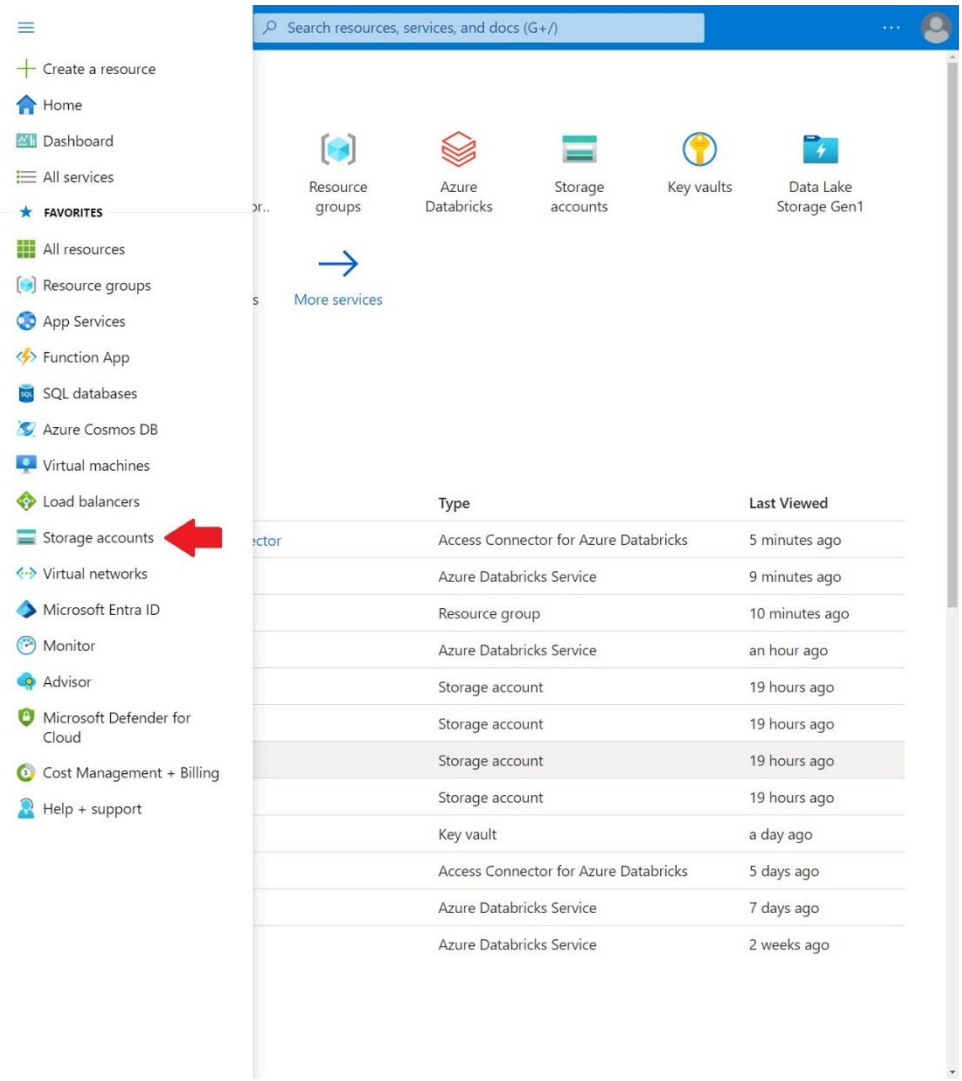
An Access Connector is typically created when creating a Databricks workspace. This connector is typically used for setting up a Unity Metastore. For users that want to create additional Access Connectors, refer to Step 1 of the following document down below for more information.

https://learn.microsoft.com/en-us/azure/databricks/data-governance/unity-catalog/azure-managed-identities#--step-1-create-an-access-connector-for-azure-databricks

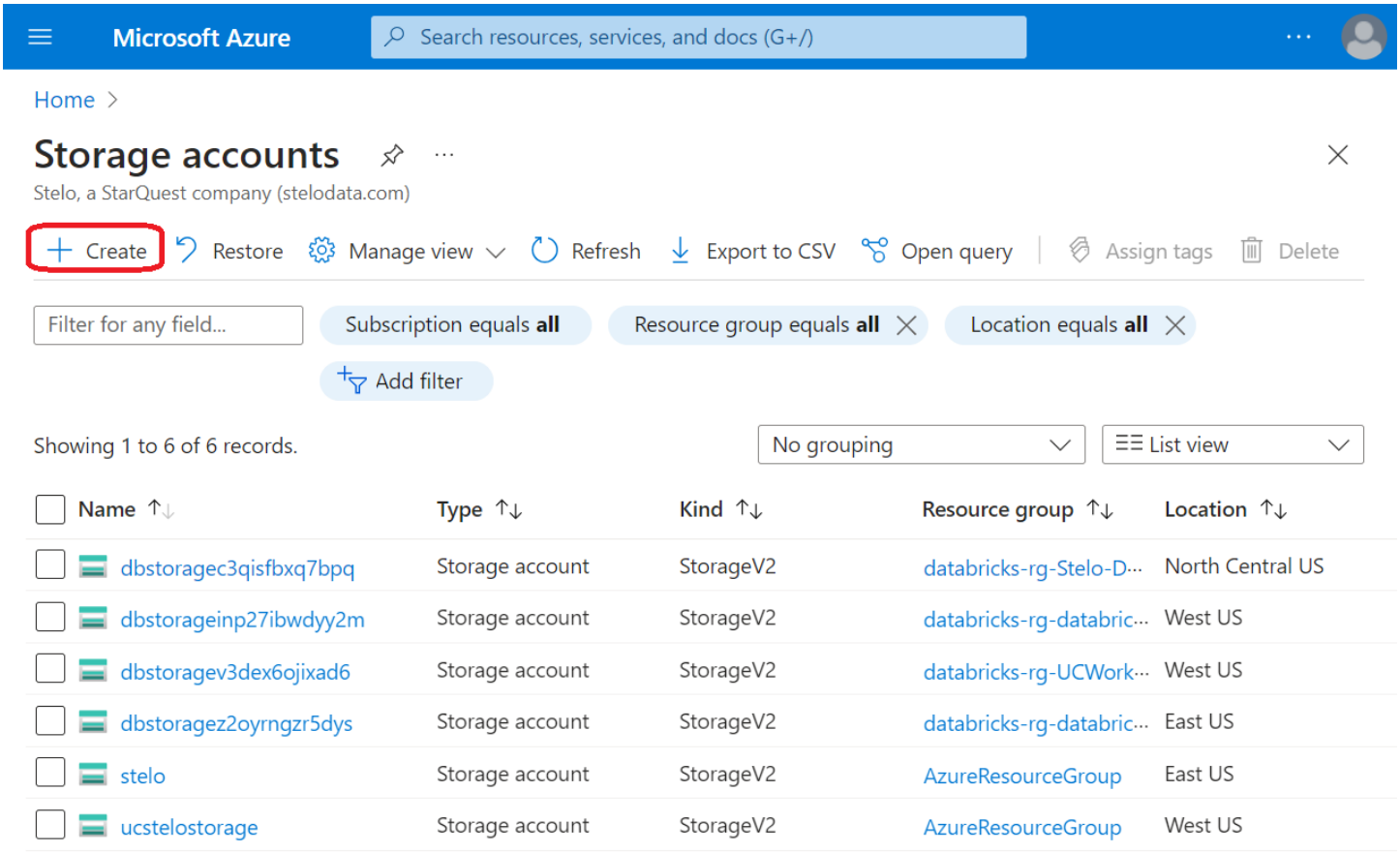
Creating the Databricks Storage account

By default, a storage account is created for the Databricks workspace upon the workspace's creation. However, this storage account will not be eligible for upgrading to a Data Lake Gen 2 storage account. To enable UC support for the Databricks workspace, a Data Lake Gen 2 storage account must be created. To create a Data Lake Gen 2 storage account, do the following:

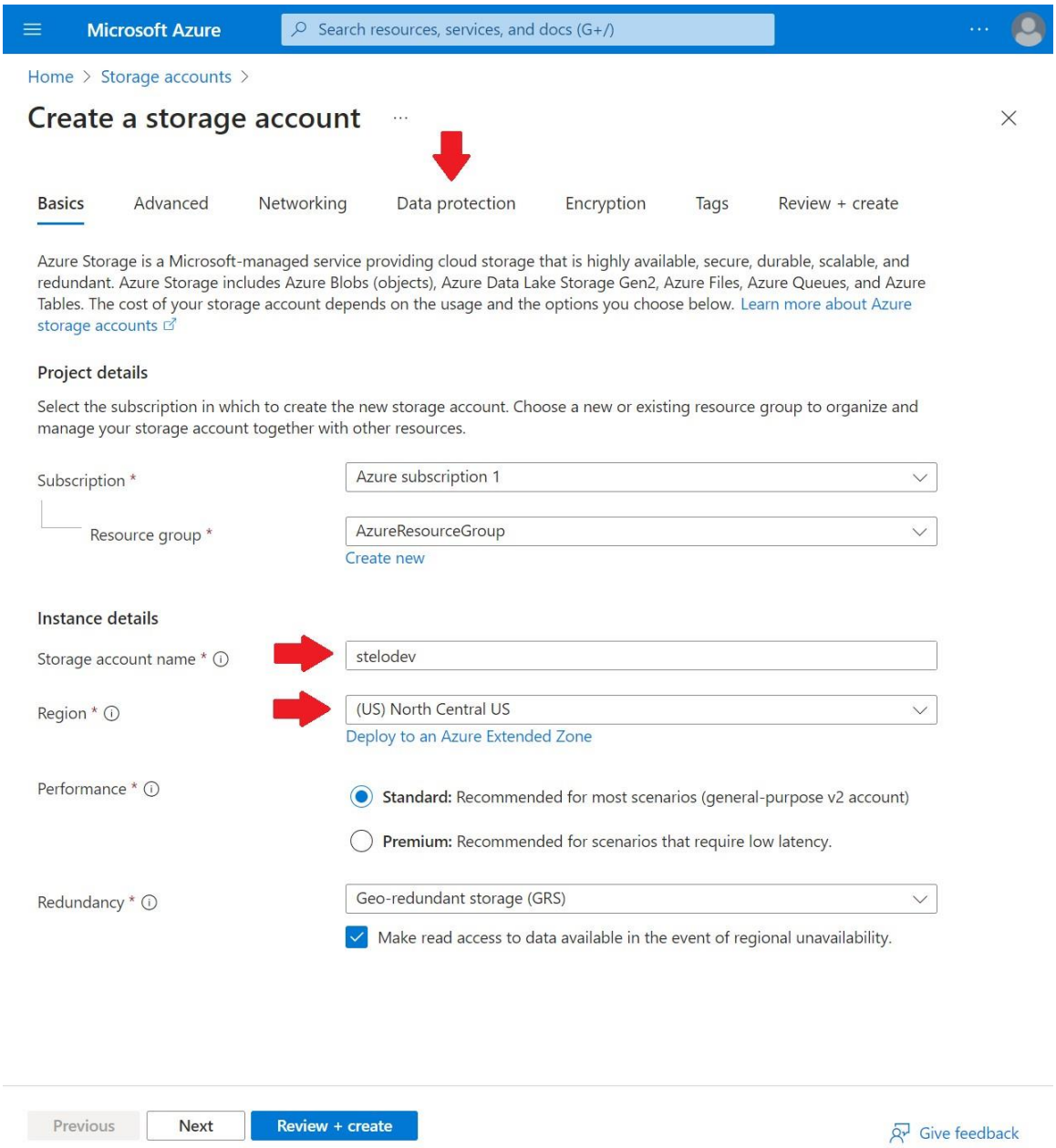
1) Select Storage Accounts



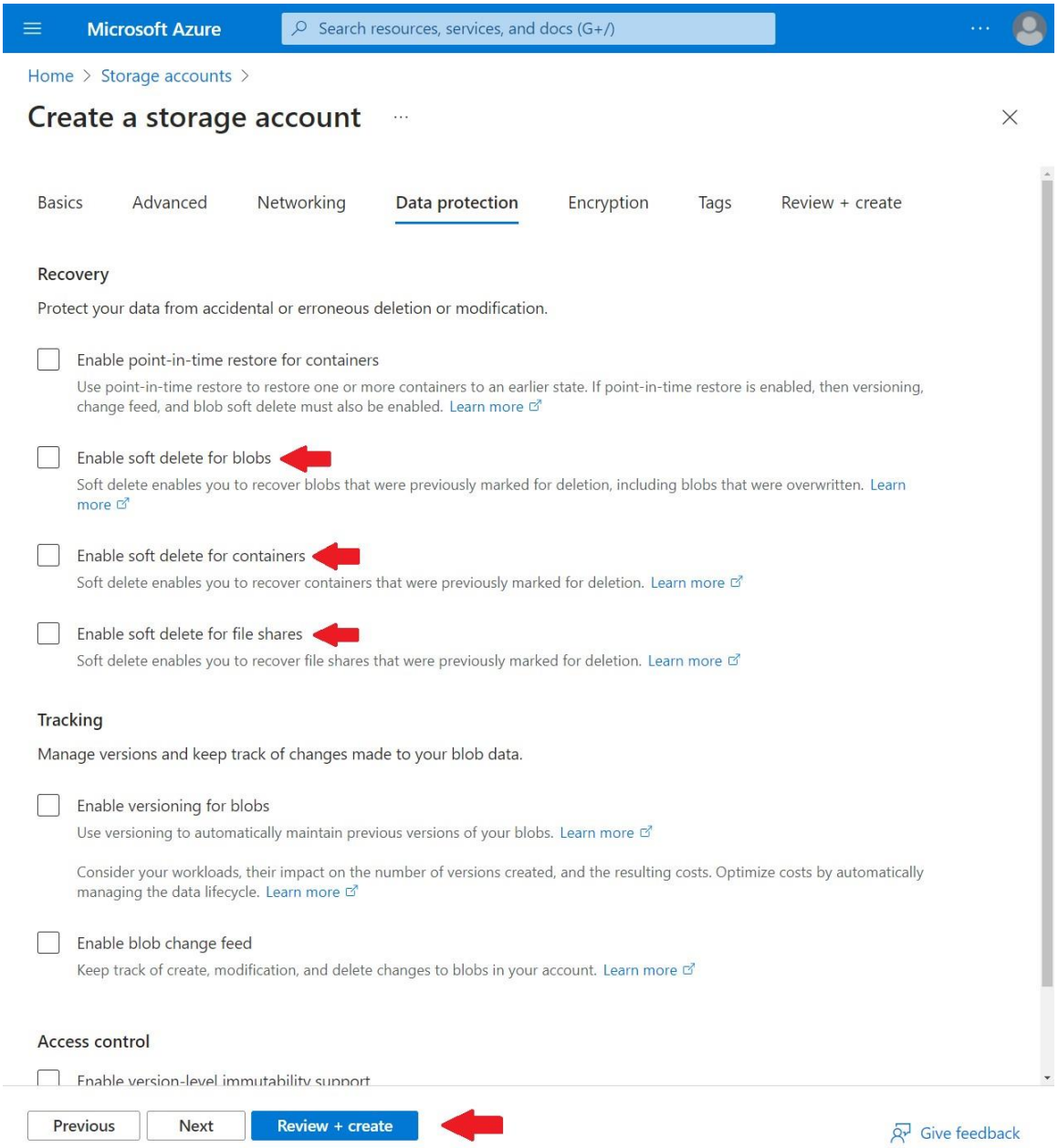
2) Select **Create.**



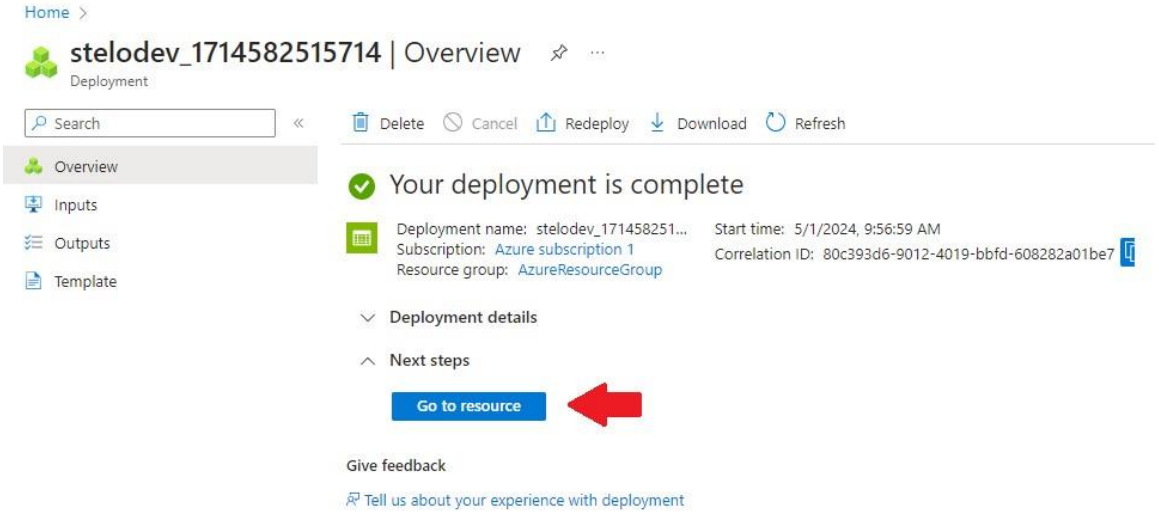
3) Select the Subscription and Resource group of interest. Then enter a name for the Storage account and select the same region in which all the Databricks resources are deployed. As an example, if the Databricks resources are deployed in West US, then select West US for the region. An example of this can be seen down below.



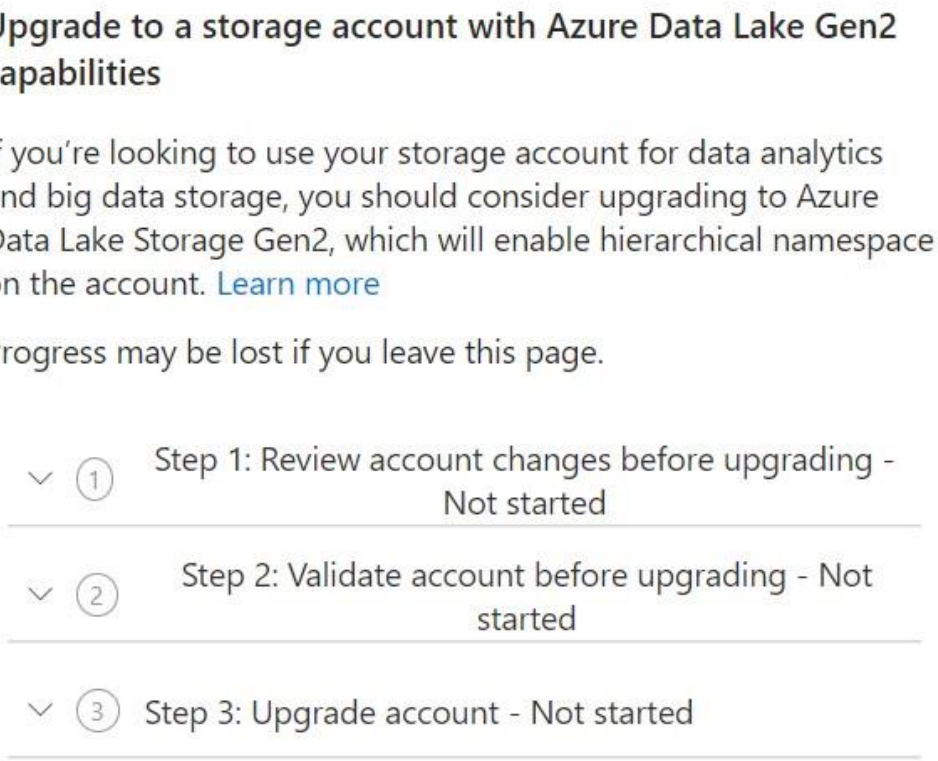
4) Once the above data has been filled, select **Data protection** and uncheck the boxes for **Enable soft delete for containers**, **Enable soft delete for file shares**, and **Enable soft delete for blobs**. Select **Review + create** to review and then create.



5) Once the Storage account has been created, go into the resource, and select **Data Lake Gen2 upgrade** in the side bar.

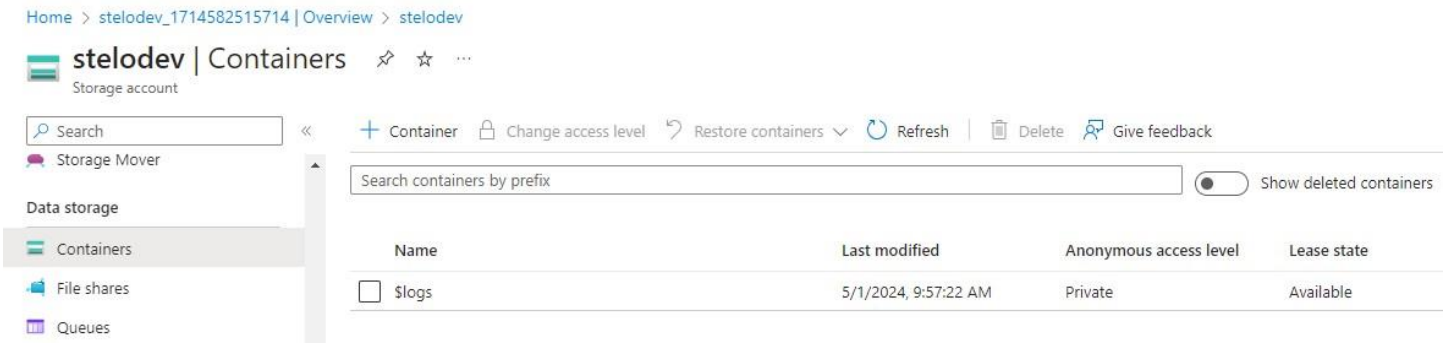


6) Upgrade the storage account to Data Lake Gen2 and follow the prompts carefully.

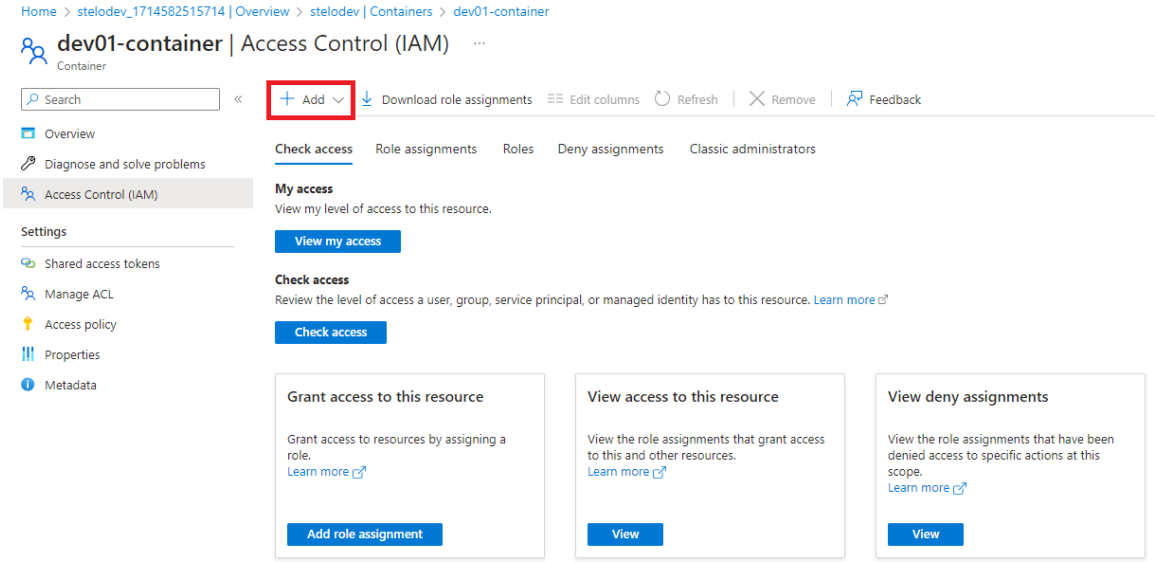


Create a Storage Container

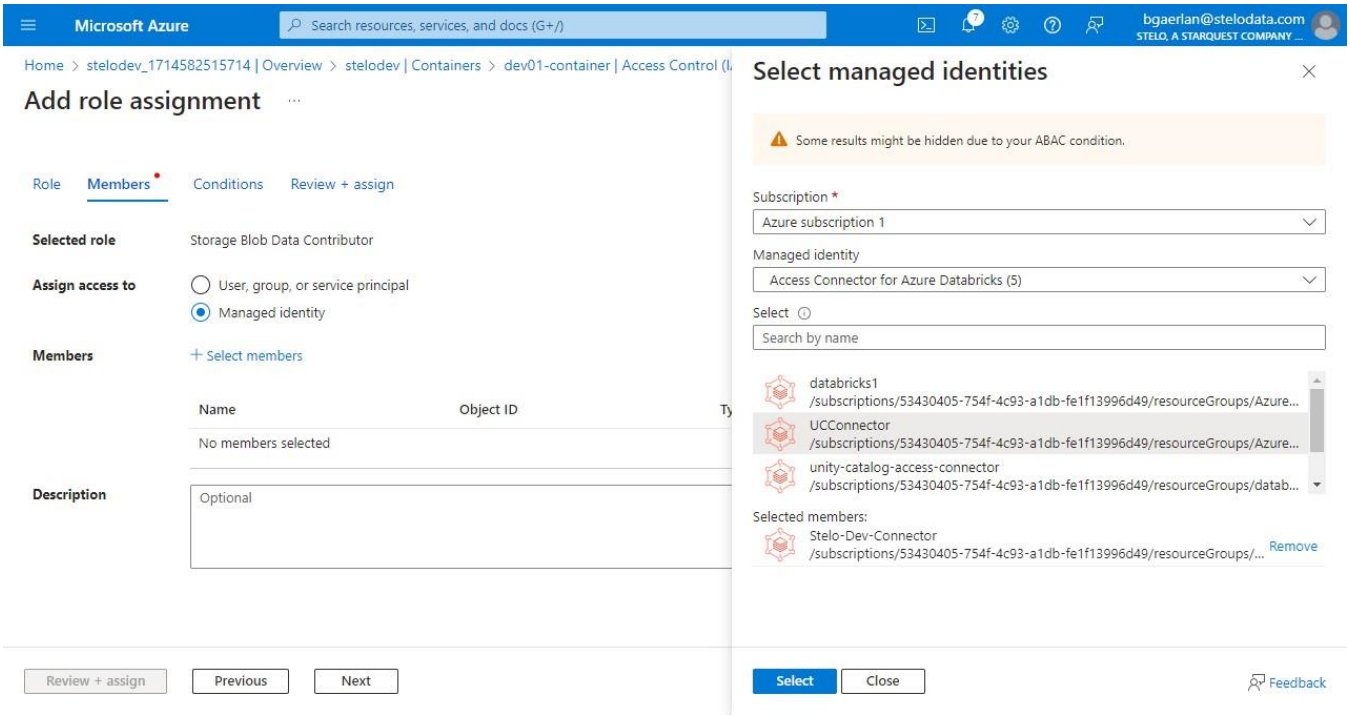
1) Select **+ Container** to add a new container.



2) Add a name for the new container and select **Create**.
3) Once the container has been created, select the newly created container.
4) Select **Access Control (IAM)**.
5) Select **Add**, then select **Add role assignment**.

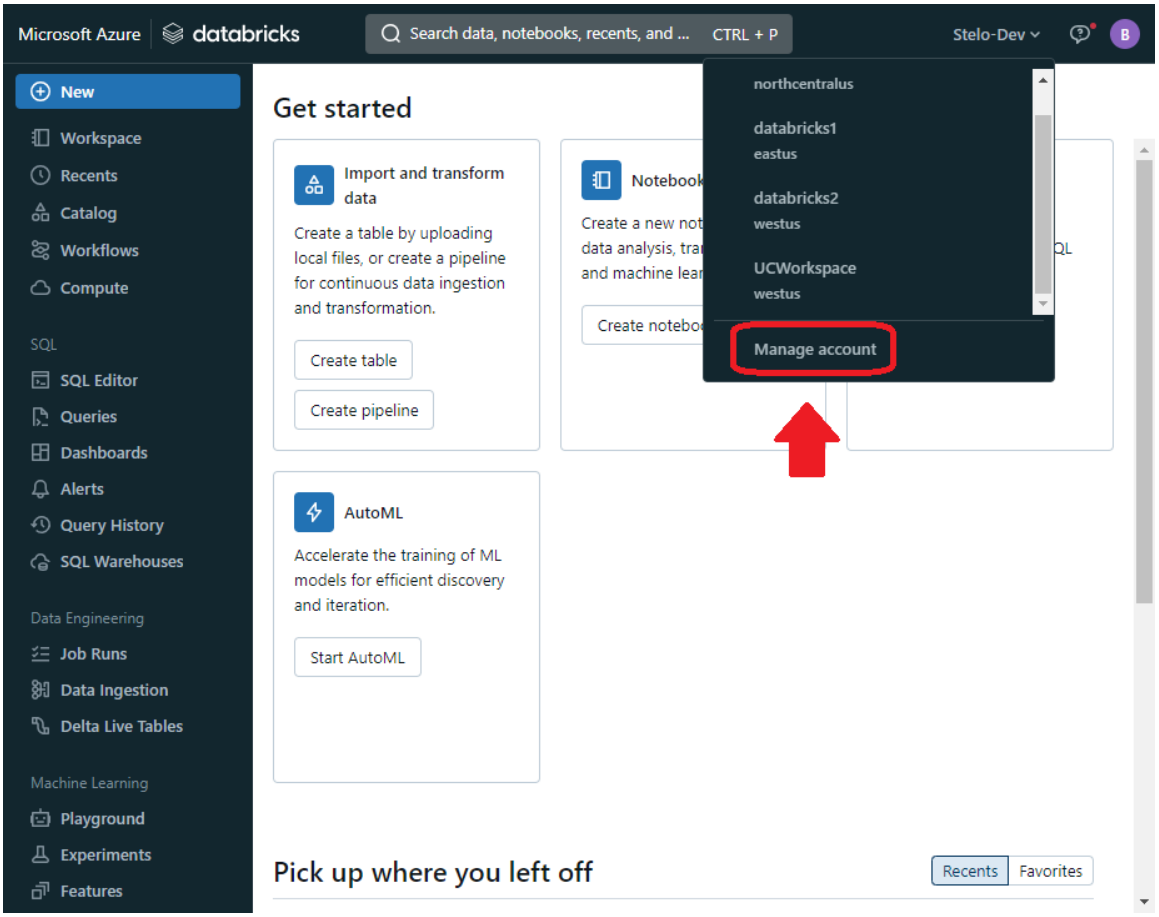


6) Assign the Access Connector that was created in "**Create an Access Connector for the new workspace**" a role of **Storage Blob Data Contributer**.



Create a Unity metastore

1) Select the workspace dropdown, then select Manage account. This will automatically open a new tab.

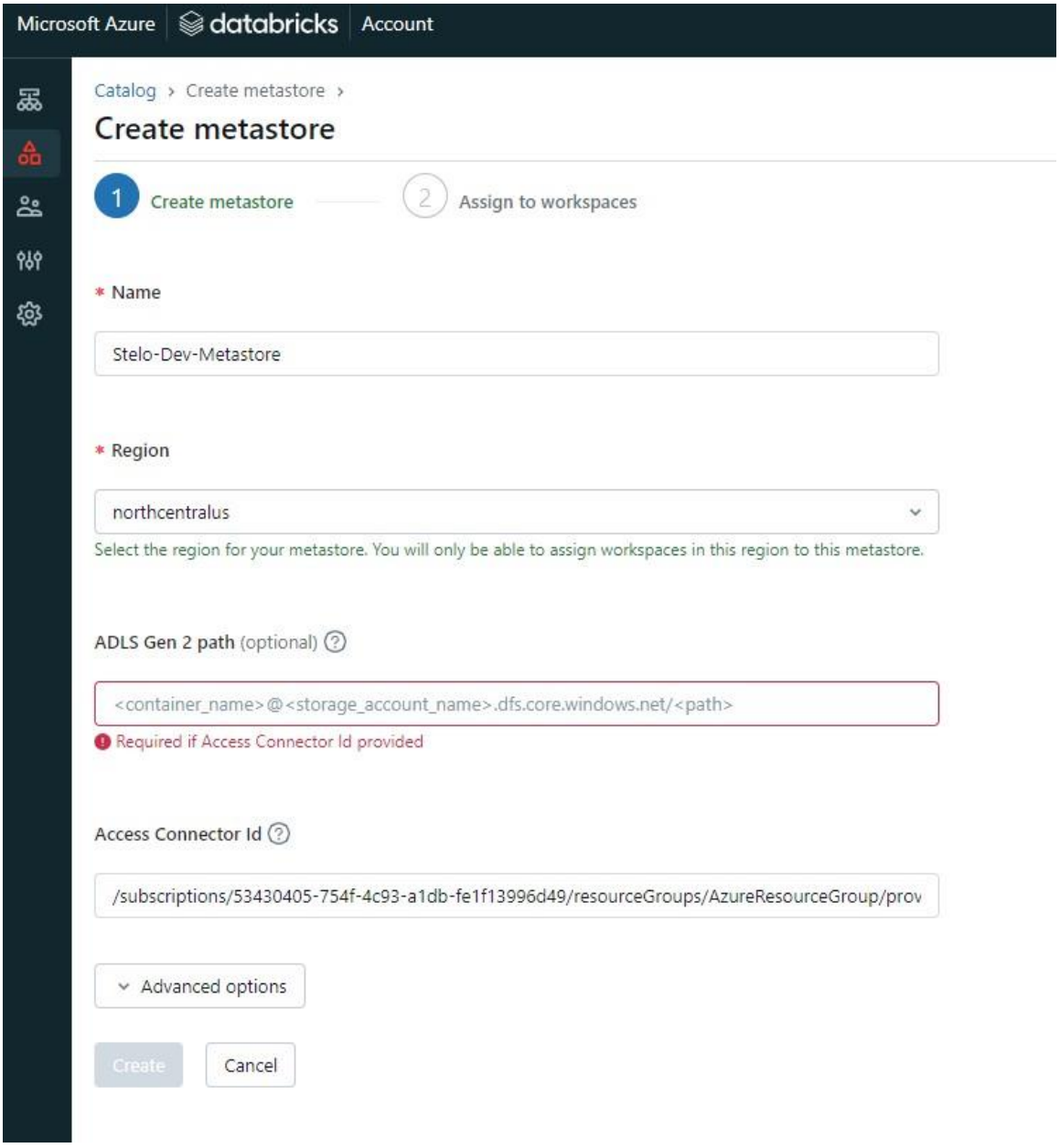


2) Select **Catalog**, then select **Create metastore**.

3) Enter the following information
- A name for the metastore
- A region in which all the Databricks resources are located
- Enter the **Access Connector Id**

Note: If the Access Connector is provided, then the ADLS Gen 2 path **must** be provided in the next step.
- Enter the **ADLS Gen 2 path**.

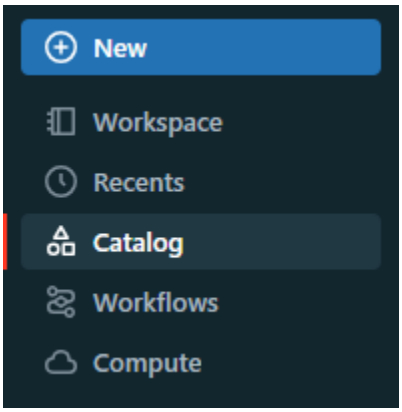
Note: Must be provided if the Access Connector Id is specified.



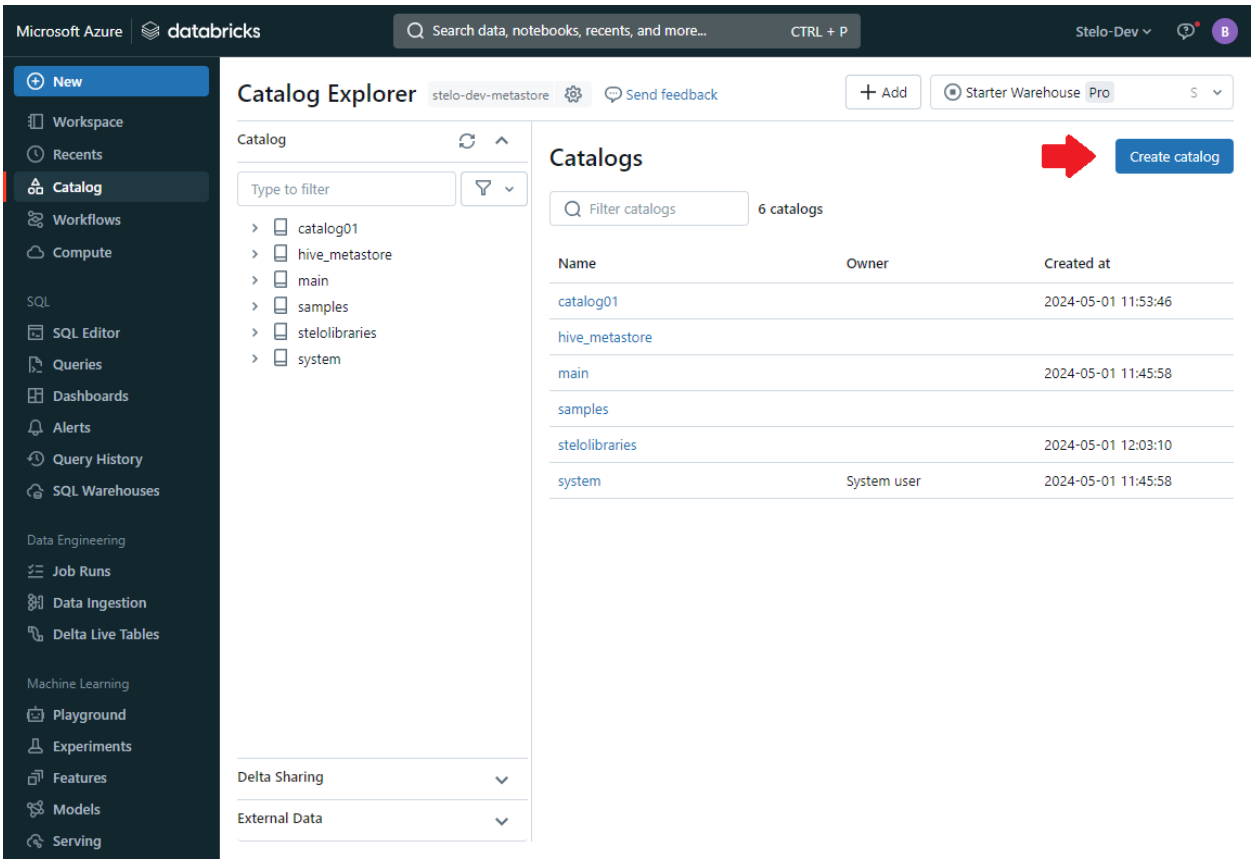
4) Select **Create**.
5) Once the metastore has been created, assign the metastore to the Databricks workspace that was created by selecting the checkbox of the workspace.
6) Enable Unity Catalog by selecting **Enable**.

Create a Unity Catalog

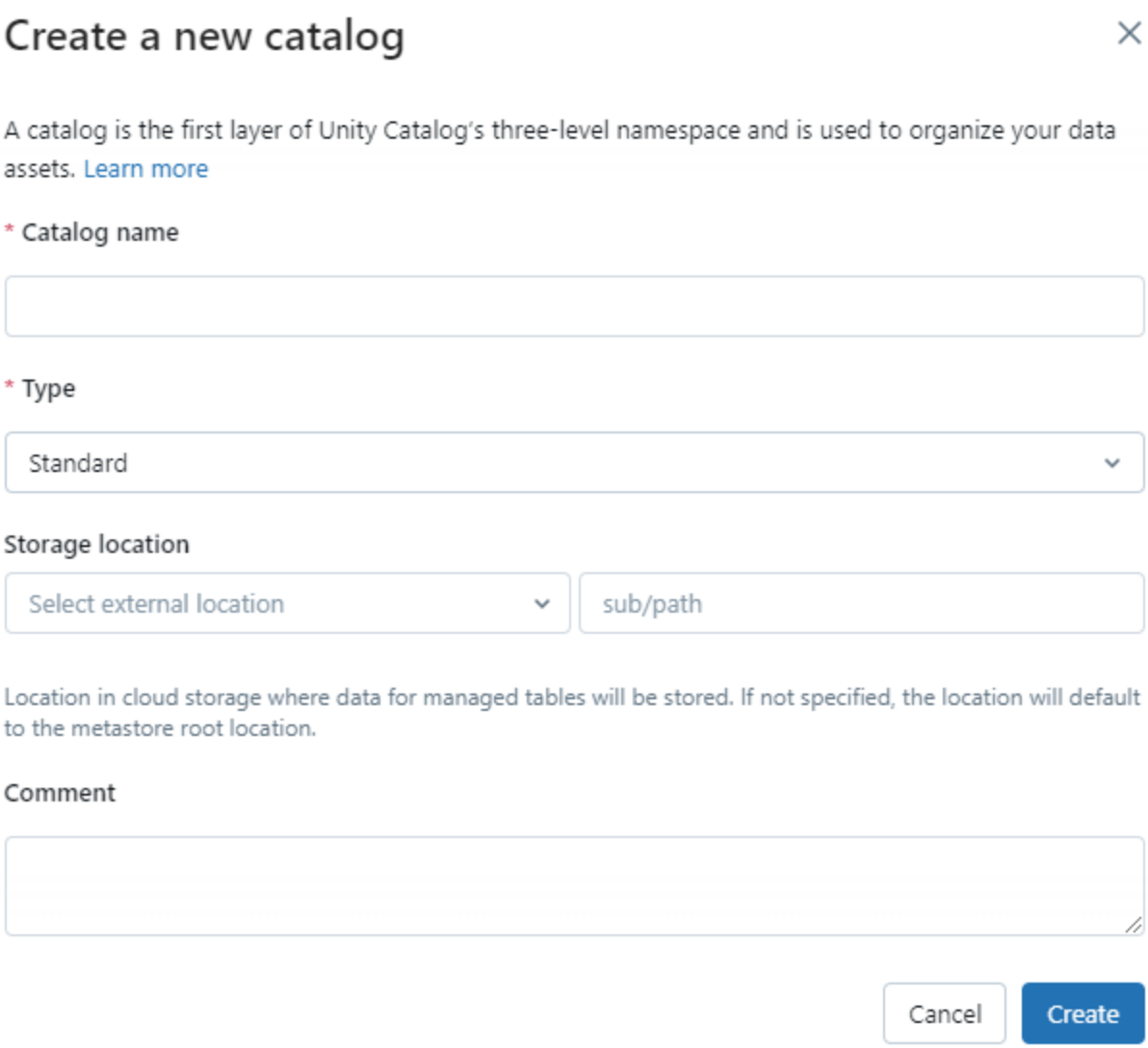
1) In the Databricks workspace, select **Catalog**.



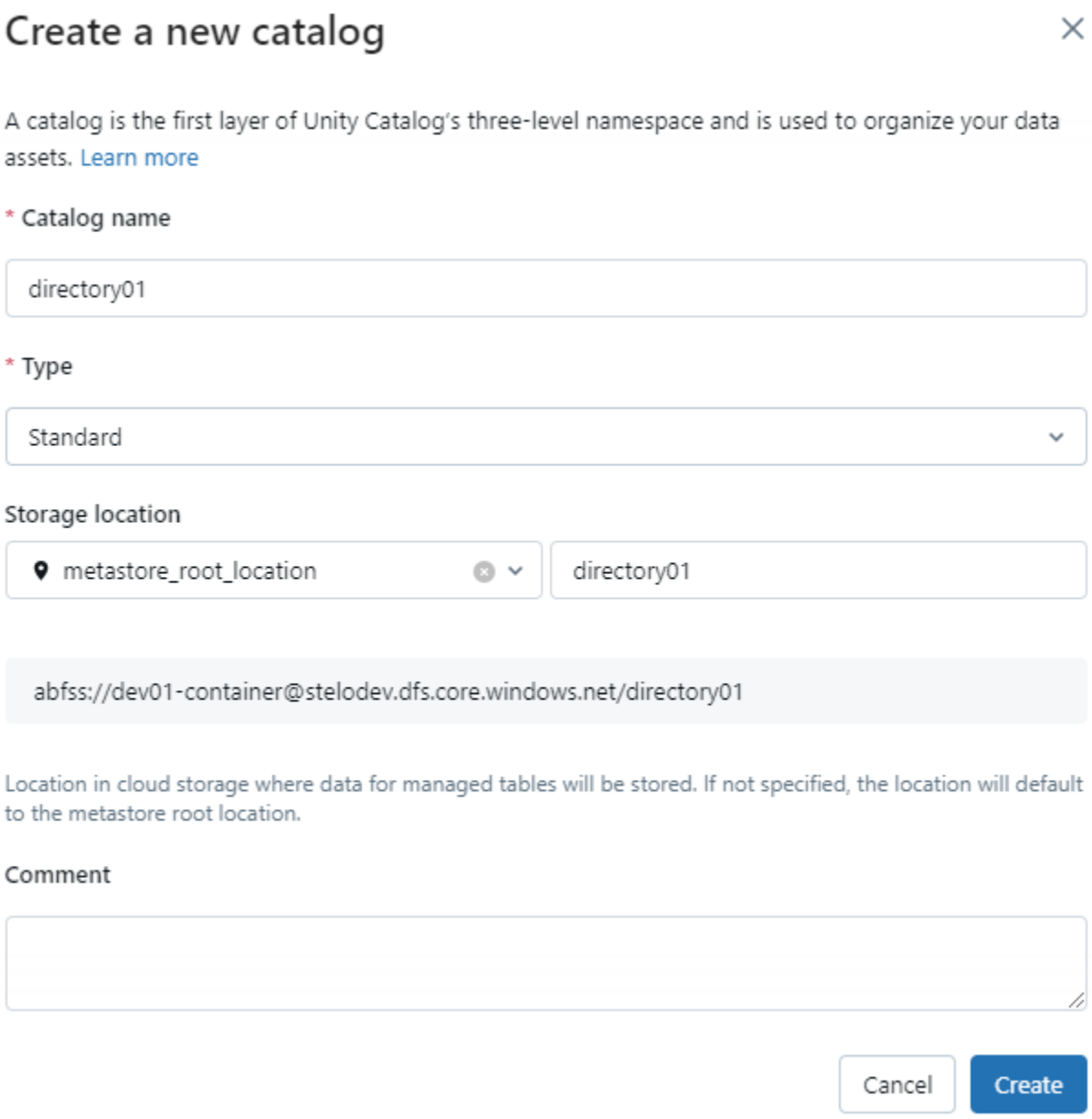
2) Select **Create catalog**.



3) Enter the **catalog name**. The catalog type can be left as **Standard**.



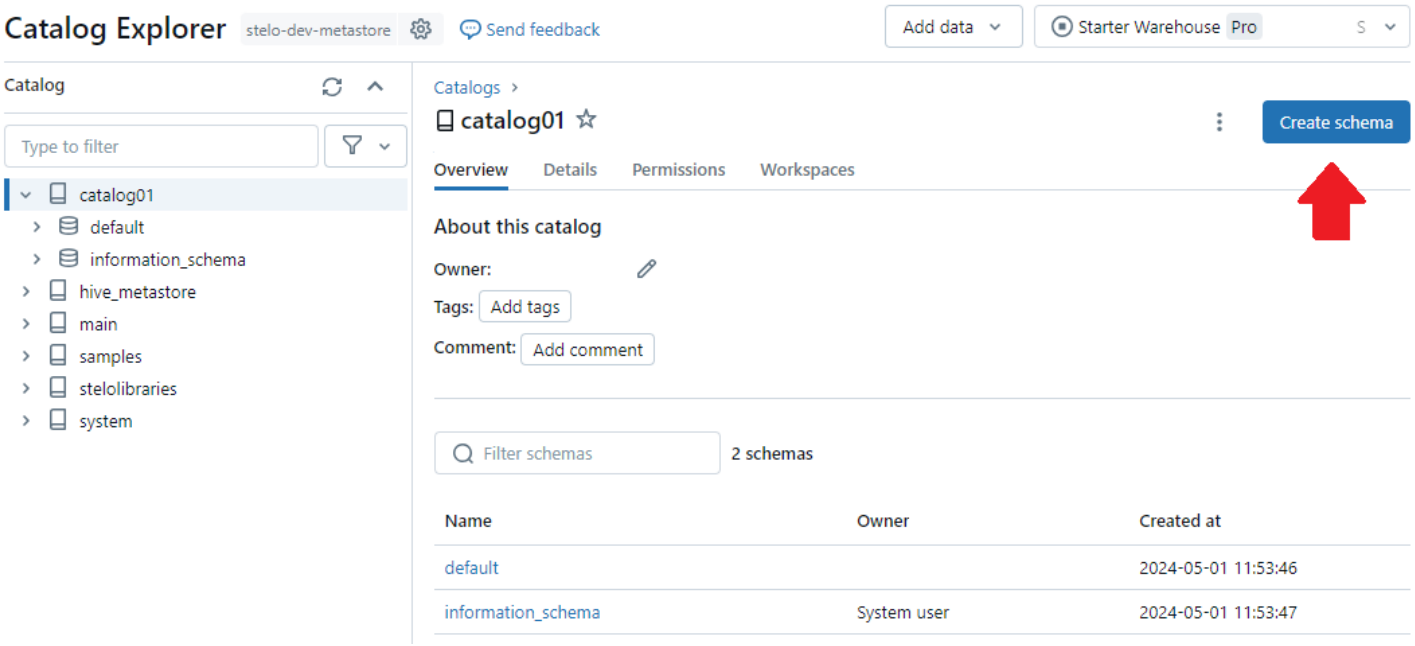
4) Optionally, a storage location can be set for the new catalog using the newly created metastore. However, a path must be supplied if the newly created metastore will be used. An example down below shows the path where all the contents of the catalog will be stored.



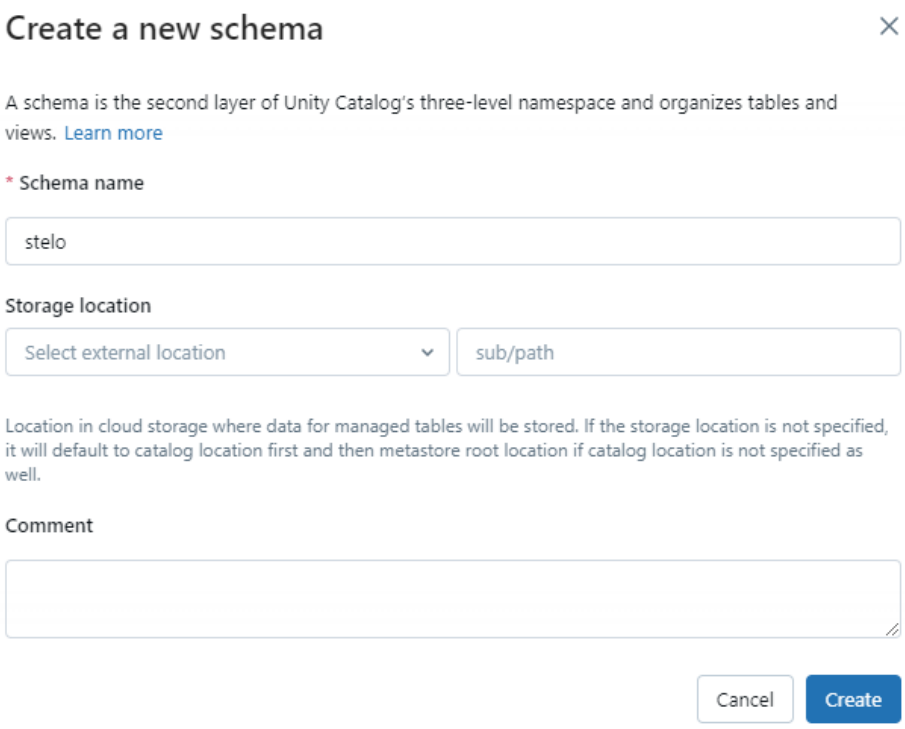
5) Select **Create** to create the Unity Catalog.

(Optional) Create a Unity Catalog volume

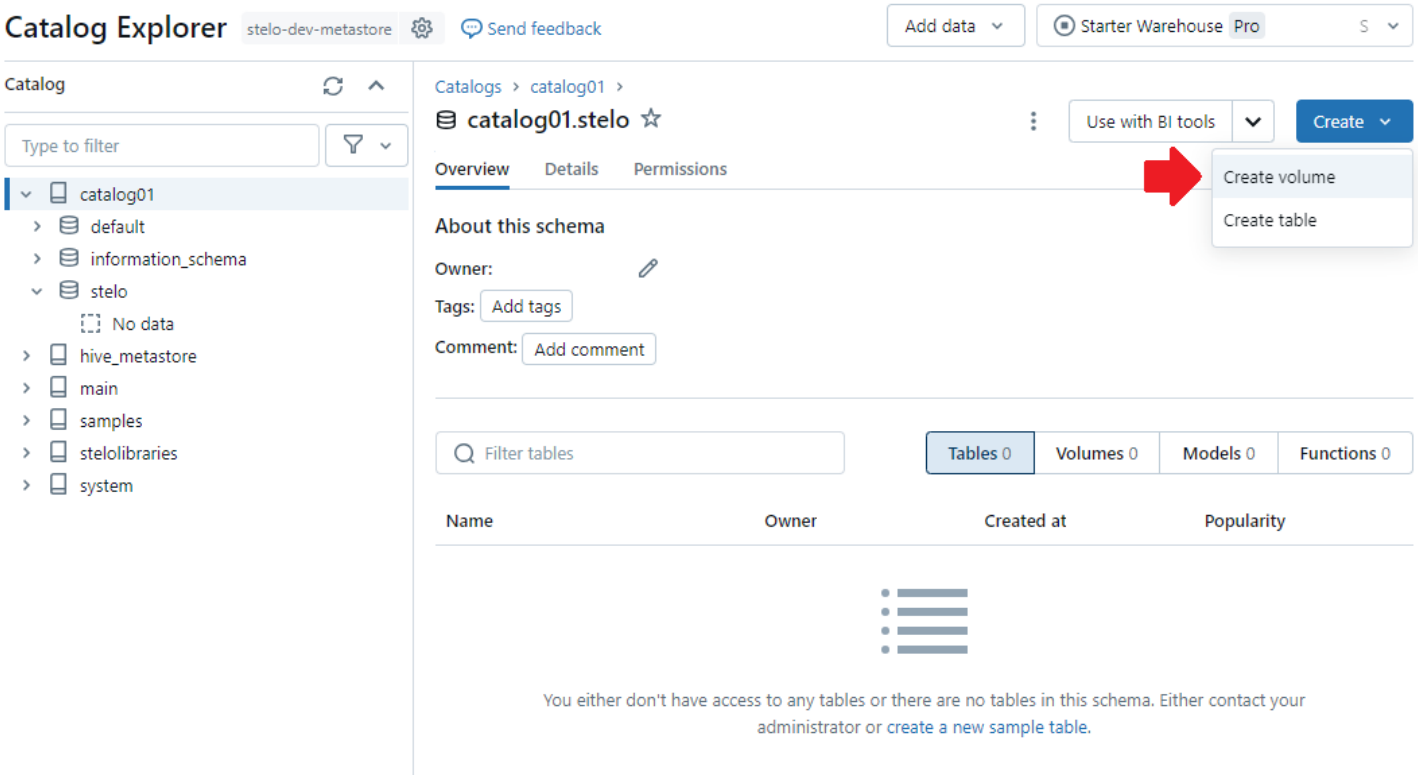
1) Select **Create schema** to create a schema under the newly created Catalog.



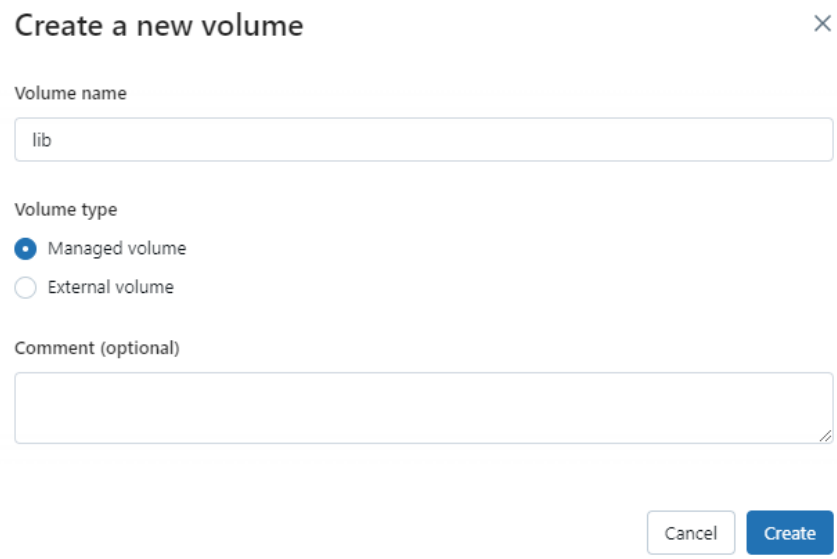
2) Enter a name for the new schema and select **Create**.



3) Create the Unity Catalog Volume under the newly created schema.

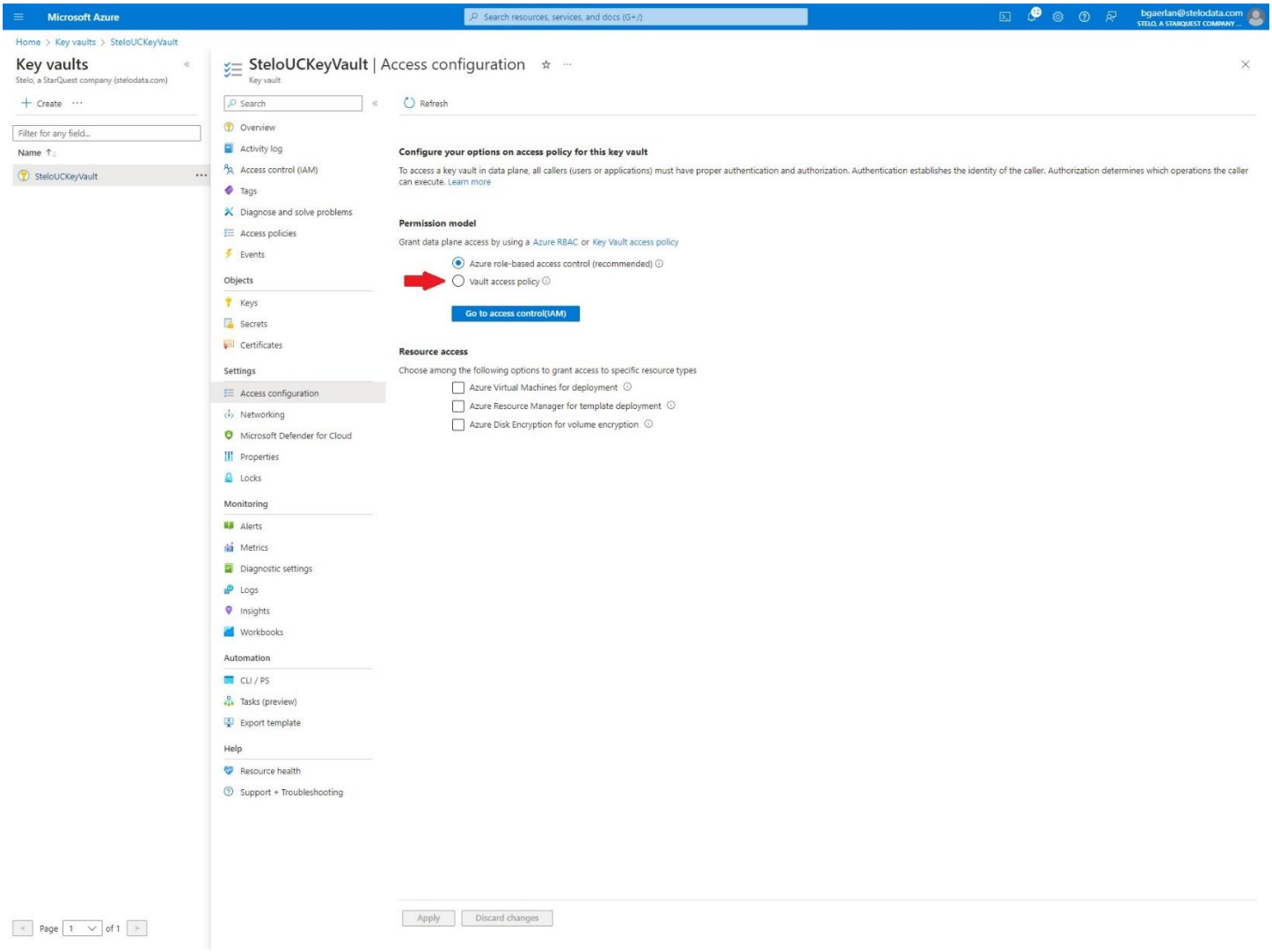


4) Enter a name for the volume and then select **Create** to create the UC volume.
Note: If the user wants to manage the location of the volume, select External volume and enter the path of the new volume. Otherwise, select Managed volume to create the volume under the catalog directory.



Create an Azure key vault and secrets (Optional)

1. Create Key Vault and assign to same region
2. Select Access configuration, then select the **Vault access policy** as shown



3. Select **Secrets** on the side panel.
4. Enter the name of the Secret and the Value.
5. Select **Create**.
6. Repeat steps 3 to 5 for additional secrets.
7. Once all of the secrets have been created, select **Properties** on the side panel.
8. Note down the Vault URI and the Resource ID
   Vault URI:
   Resource ID:

Configure the Azure key vault instance for Azure Databricks

1) Login to the Databricks workspace and copy the URL of the workspace. An example of a workspace URL can be seen below.

   https://adb-2619841996514801.1.azuredatabricks.net/

2) Append #secrets/createScope to the URL as shown in the example below

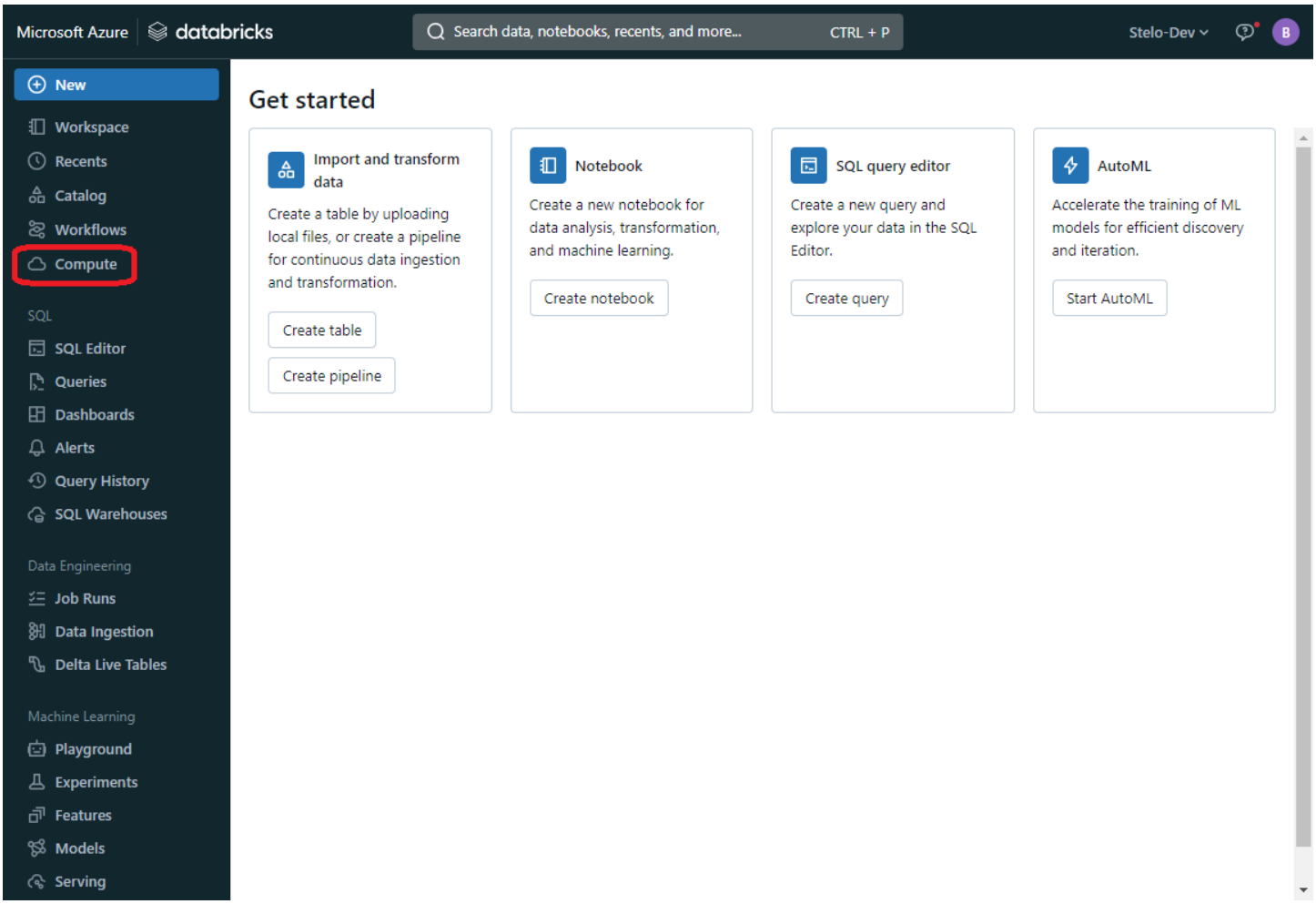
   https://adb-2619841996514801.1.azuredatabricks.net/#secrets/createScope

3) Enter the Scope Name, DNS Name (Vault URI), and the Resource ID.
   Note: The Vault URI and Resource ID taken from step 8 of **Create an Azure Key Vault** should be entered in this step

4) Select **Create**.

Create a Databricks cluster

1) In the Databricks workspace portal, select **Compute**.



2) Set the name of the cluster, then choose whether the cluster will be Single node or Multi node.

Note:

- If Single node is chosen, then the Access mode cannot be **Shared**.

3) Once satisfied, select **Create compute**.