

How to setup Databricks workspace with Unity Catalog

Contents

Databricks Setup	2
General Notes and Requirements	2
Create a Databricks workspace	2
Create an Access Connector for the Databricks workspace	5
Creating the Databricks Storage account	5
Create a Storage Container	11
Create a Unity metastore	13
Create a Unity Catalog	15
(Optional) Create a Unity Catalog volume	18
Create an Azure key vault and secrets (Optional)	20
Configure the Azure key vault instance for Azure Databricks	21
Create a Databricks cluster	22

Databricks Setup

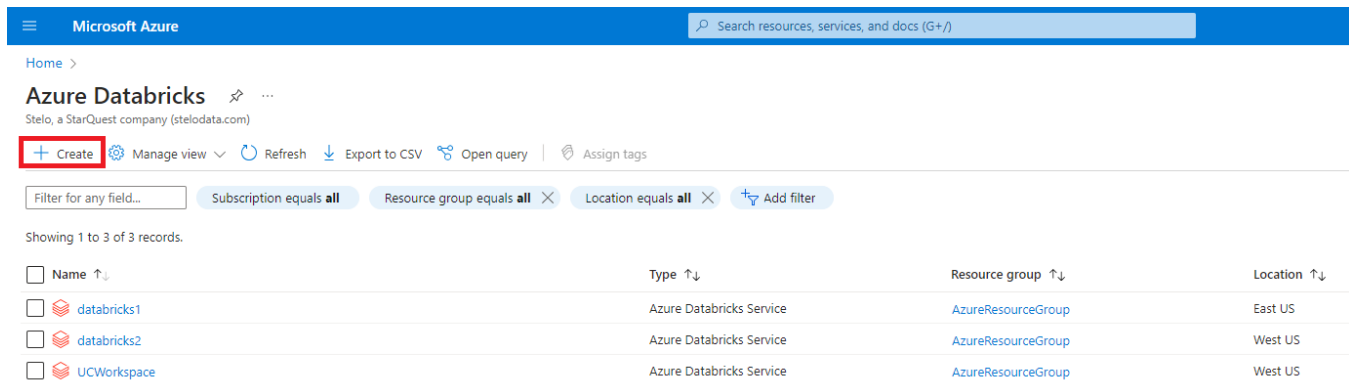
General Notes and Requirements

- **(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**.








Microsoft Azure

Search resources, services, and docs (G+)

Home >




Azure Databricks  ...

Stelo, a StarQuest company (stelodata.com)

+ Create  Manage view  Refresh  Export to CSV  Open query  Assign tags

Filter for any field... Subscription equals all Resource group equals all Location equals all Add filter

Showing 1 to 3 of 3 records.

<input type="checkbox"/>	Name ↑↓	Type ↑↓	Resource group ↑↓	Location ↑↓
<input type="checkbox"/>	 databricks1	Azure Databricks Service	AzureResourceGroup	East US
<input type="checkbox"/>	 databricks2	Azure Databricks Service	AzureResourceGroup	West US
<input type="checkbox"/>	 UCWorkspace	Azure Databricks Service	AzureResourceGroup	West US

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**.

[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="Azure subscription 1"/>
Resource group * ⓘ	<input type="text" value="AzureResourceGroup"/>

[Create new](#)

Instance Details

Workspace name *	<input type="text" value="testworkspace"/>
Region *	<input type="text" value="North Central US"/>
Pricing Tier * ⓘ	<input type="text" value="Premium (+ Role-based access controls)"/>

i We selected the recommended pricing tier for your workspace. You can change the tier based on your needs. ×

Managed Resource Group name	<input type="text" value="Enter name for managed resource group"/>
-----------------------------	--------------------------------------------------------------------

[Review + create](#)

[< Previous](#)

[Next : Networking >](#)

3) Select **Create**.

Microsoft Azure

Home > Azure Databricks >

Create an Azure Databricks workspace

Validation Succeeded

Basics Networking Encryption Security & compliance Tags Review + create

Summary

Basics

Workspace name	testworkspace
Subscription	Azure subscription 1
Resource group	AzureResourceGroup
Region	North Central US
Pricing Tier	premium
Managed Resource Group name	

Networking

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

Encryption

Enable Infrastructure Encryption	No
Enable CMK for Managed Disks	No
Enable CMK for Managed Services	No

Security & compliance

Compliance Security Profile	No
Compliance Standards	
Enhanced Security Monitoring	No
Automatic Cluster Update	No

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

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

The screenshot shows the Azure portal interface. On the left, the navigation pane includes 'Storage accounts' with a red arrow pointing to it. The main content area features a search bar at the top, followed by a grid of service tiles: Resource groups, Azure Databricks, Storage accounts, Key vaults, and Data Lake Storage Gen1. Below the tiles is a 'More services' button. A table below lists various resources:

Type	Last Viewed
Access Connector for Azure Databricks	5 minutes ago
Azure Databricks Service	9 minutes ago
Resource group	10 minutes ago
Azure Databricks Service	an hour ago
Storage account	19 hours ago
Storage account	19 hours ago
Storage account	19 hours ago
Storage account	19 hours ago
Key vault	a day ago
Access Connector for Azure Databricks	5 days ago
Azure Databricks Service	7 days ago
Azure Databricks Service	2 weeks ago

2) Select **Create**.

Microsoft Azure

Home >

Storage accounts

Stelo, a StarQuest company (stelodata.com)

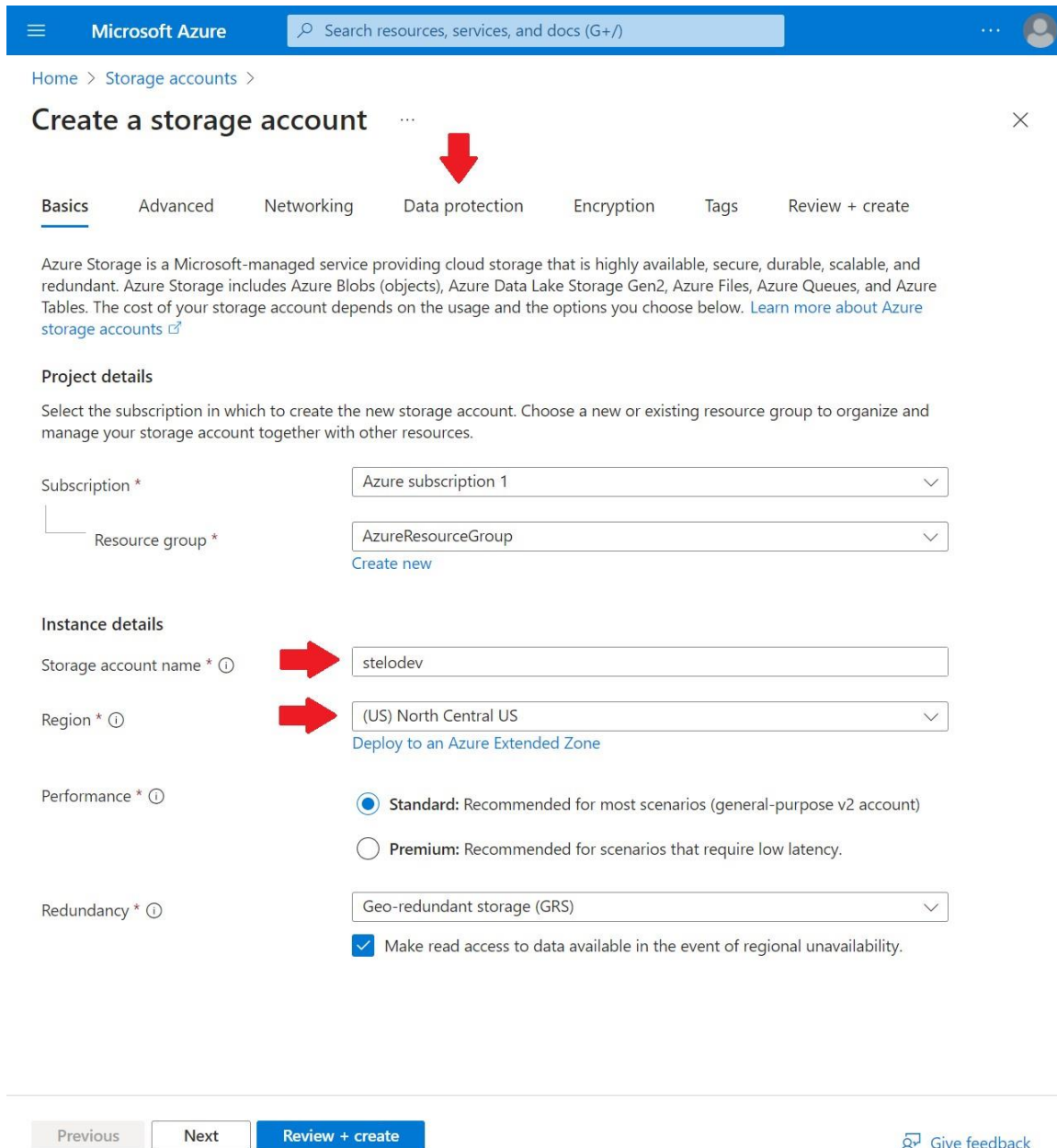
+ Create Restore Manage view Refresh Export to CSV Open query Assign tags Delete

Filter for any field... Subscription equals all Resource group equals all Location equals all Add filter

Showing 1 to 6 of 6 records. No grouping List view

<input type="checkbox"/>	Name ↑↓	Type ↑↓	Kind ↑↓	Resource group ↑↓	Location ↑↓
<input type="checkbox"/>	dbstorageec3qisfbxq7bpq	Storage account	StorageV2	databricks-rg-Stelo-D...	North Central US
<input type="checkbox"/>	dbstorageinp27ibwdyy2m	Storage account	StorageV2	databricks-rg-databric...	West US
<input type="checkbox"/>	dbstoragev3dex6ojxad6	Storage account	StorageV2	databricks-rg-UCWork...	West US
<input type="checkbox"/>	dbstoragez2oyrngzr5dys	Storage account	StorageV2	databricks-rg-databric...	East US
<input type="checkbox"/>	stelo	Storage account	StorageV2	AzureResourceGroup	East US
<input type="checkbox"/>	ucstelostorage	Storage account	StorageV2	AzureResourceGroup	West US

- 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.



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

Home > Storage accounts >

Create a storage account

Basics Advanced Networking **Data protection** Encryption Tags Review + create

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. [Learn more about Azure storage accounts](#)

Project details

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription * Azure subscription 1

Resource group * AzureResourceGroup [Create new](#)

Instance details

Storage account name * ① **→** stelodev

Region * ① **→** (US) North Central US [Deploy to an Azure Extended Zone](#)

Performance * ①

Standard: Recommended for most scenarios (general-purpose v2 account)

Premium: Recommended for scenarios that require low latency.

Redundancy * ①

Geo-redundant storage (GRS)

Make read access to data available in the event of regional unavailability.

Previous Next **Review + create** [Give feedback](#)

- 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.

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




Home > Storage accounts >

Create a storage account

Basics Advanced Networking **Data protection** Encryption Tags Review + create

Recovery

Protect your data from accidental or erroneous deletion or modification.

- Enable point-in-time restore for containers
Use point-in-time restore to restore one or more containers to an earlier state. If point-in-time restore is enabled, then versioning, change feed, and blob soft delete must also be enabled. [Learn more](#)
- Enable soft delete for blobs 
Soft delete enables you to recover blobs that were previously marked for deletion, including blobs that were overwritten. [Learn more](#)
- Enable soft delete for containers 
Soft delete enables you to recover containers that were previously marked for deletion. [Learn more](#)
- Enable soft delete for file shares 
Soft delete enables you to recover file shares that were previously marked for deletion. [Learn more](#)


Tracking

Manage versions and keep track of changes made to your blob data.

- Enable versioning for blobs
Use versioning to automatically maintain previous versions of your blobs. [Learn more](#)
Consider your workloads, their impact on the number of versions created, and the resulting costs. Optimize costs by automatically managing the data lifecycle. [Learn more](#)
- Enable blob change feed
Keep track of create, modification, and delete changes to blobs in your account. [Learn more](#)


Access control

- Enable version-level immutability support






 [Give feedback](#)

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


Home >



stelodev_1714582515714 | Overview  ...

Deployment

Search <<  Delete  Cancel  Redeploy  Download  Refresh


- Overview
- Inputs
- Outputs
- Template

 **Your deployment is complete**


 Deployment name: stelodev_171458251... Start time: 5/1/2024, 9:56:59 AM
Subscription: Azure subscription 1 Correlation ID: 80c393d6-9012-4019-bbfd-608282a01be7 
Resource group: AzureResourceGroup

∨ Deployment details

∧ Next steps

[Go to resource](#) 

Give feedback

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

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

Upgrade to a storage account with Azure Data Lake Gen2 capabilities

If you're looking to use your storage account for data analytics and big data storage, you should consider upgrading to Azure Data Lake Storage Gen2, which will enable hierarchical namespace on the account. [Learn more](#)

Progress may be lost if you leave this page.

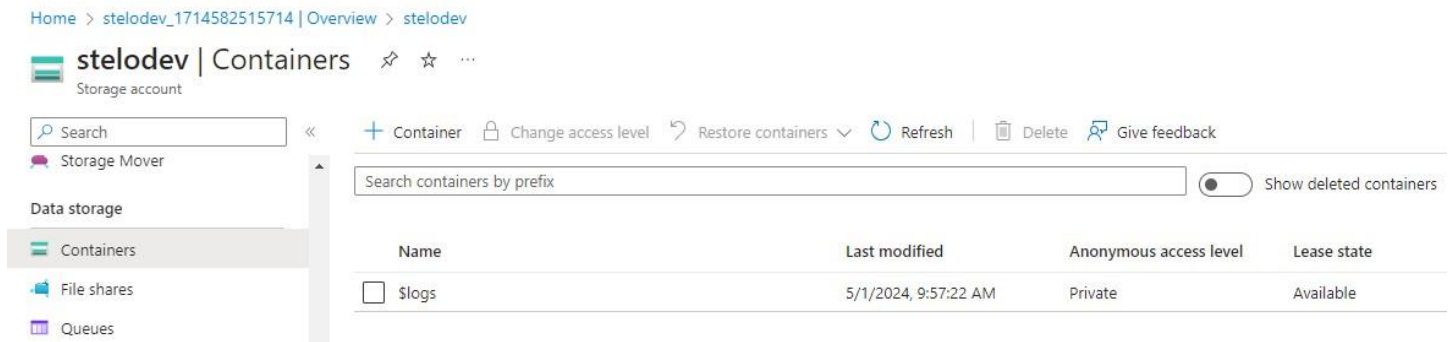
- ∨ ① Step 1: Review account changes before upgrading - Not started

- ∨ ② Step 2: Validate account before upgrading - Not started

- ∨ ③ Step 3: Upgrade account - Not started

Create a Storage Container

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

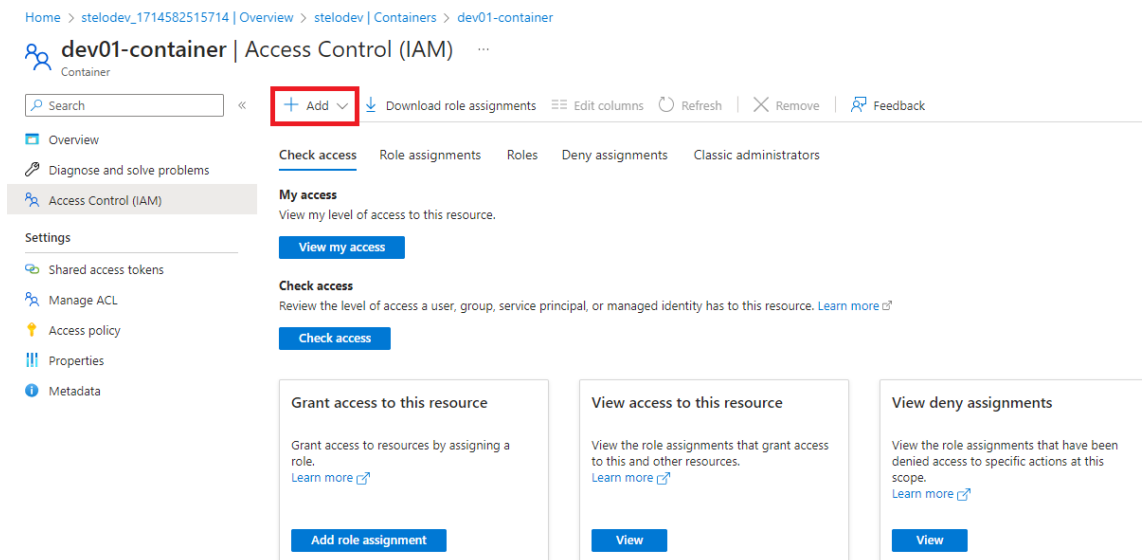


2) Add a name for the new container and select **Create**.

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 Contributor**.

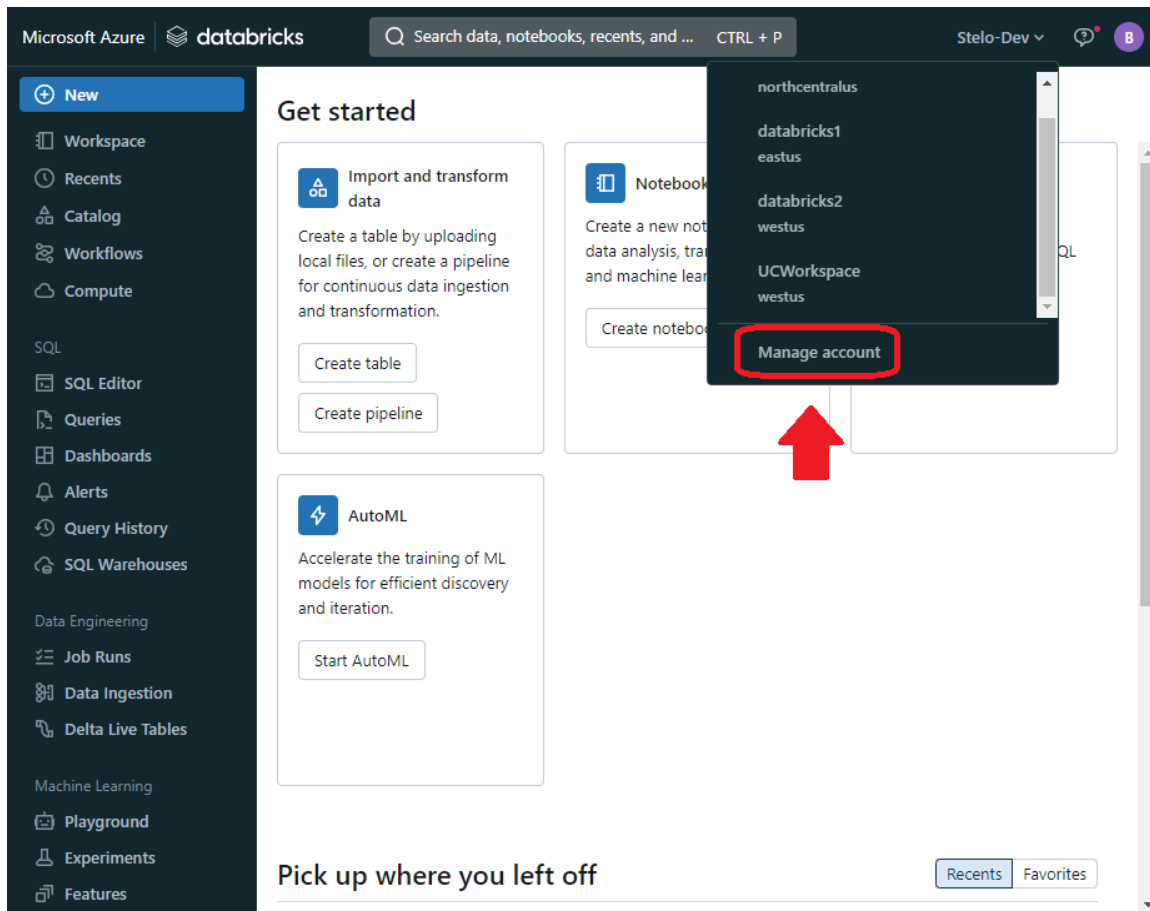
The screenshot shows the Microsoft Azure portal interface. The main page is titled "Add role assignment" and is in the "Members" tab. The "Selected role" is "Storage Blob Data Contributor". Under "Assign access to", the "Managed identity" option is selected. The "Members" section shows "No members selected". The "Description" field contains the text "Optional".

On the right side, a modal window titled "Select managed identities" is open. It displays a warning: "Some results might be hidden due to your ABAC condition." Below this, there are dropdown menus for "Subscription" (set to "Azure subscription 1") and "Managed identity" (set to "Access Connector for Azure Databricks (5)"). A search box is present with the text "Search by name". A list of managed identities is shown, with "UCConnector" selected. Below the list, the "Selected members" section shows "Stelo-Dev-Connector" with a "Remove" link.

At the bottom of the modal, there are "Select" and "Close" buttons. A "Feedback" link is also visible in the bottom right corner of the modal.

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.

Microsoft Azure | databricks | Account

Catalog > Create metastore >

Create metastore

1 Create metastore — 2 Assign to workspaces

* Name

* Region

Select the region for your metastore. You will only be able to assign workspaces in this region to this metastore.

ADLS Gen 2 path (optional) ?

Required if Access Connector Id provided

Access Connector Id ?

Advanced options

Create Cancel

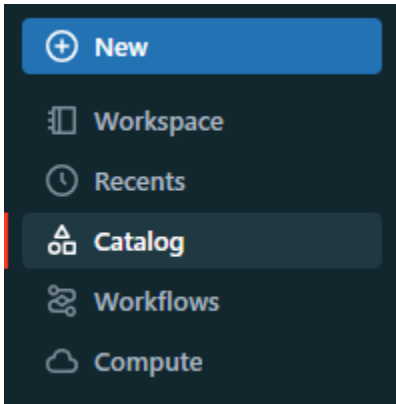
4) Select **Create**.

5) Once the metastore has been created, assign the metastore to the Databricks workspace that was created by selecting the checkbox of the workspace.

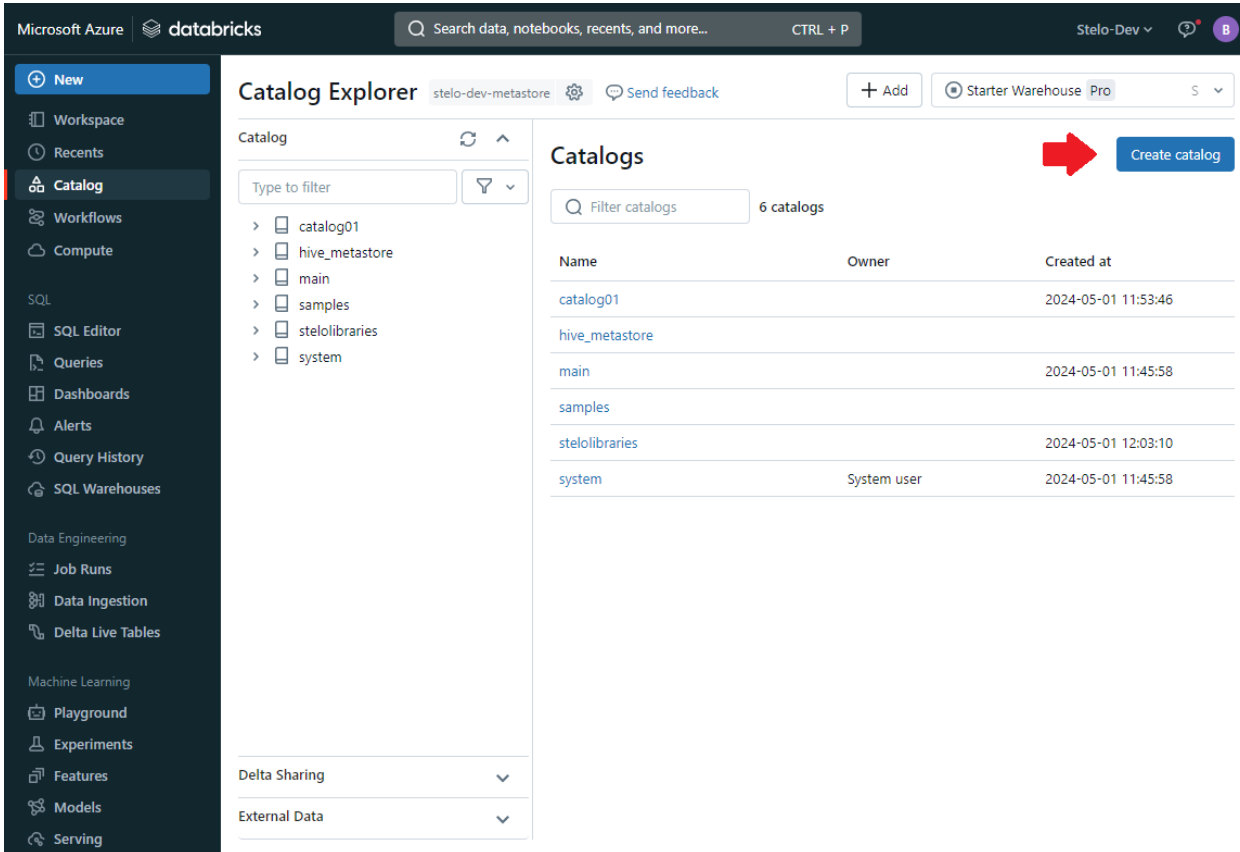
6) Enable Unity Catalog by selecting **Enable**.

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**.

Create a new catalog



A catalog is the first layer of Unity Catalog's three-level namespace and is used to organize your data assets. [Learn more](#)

* Catalog name

* Type

Storage location

Location in cloud storage where data for managed tables will be stored. If not specified, the location will default to the metastore root location.

Comment

Cancel

Create

- 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.

Create a new catalog ×

A catalog is the first layer of Unity Catalog's three-level namespace and is used to organize your data assets. [Learn more](#)

* Catalog name

* Type

Storage location

abfss://dev01-container@stelodev.dfs.core.windows.net/directory01

Location in cloud storage where data for managed tables will be stored. If not specified, the location will default to the metastore root location.

Comment

Cancel

Create

- 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.

Catalog Explorer stelo-dev-metastore Send feedback Add data Starter Warehouse Pro S

Catalog catalog01

Type to filter Filter

- catalog01
 - default
 - information_schema
 - hive_metastore
 - main
 - samples
 - stelolibraries
 - system

Catalogs > **catalog01** ☆ ⋮ Create schema

Overview Details Permissions Workspaces

About this catalog


Owner: ✎

Tags: Add tags

Comment: Add comment

Filter schemas 2 schemas

Name	Owner	Created at
default		2024-05-01 11:53:46
information_schema	System user	2024-05-01 11:53:47



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

Create a new schema ✕

A schema is the second layer of Unity Catalog's three-level namespace and organizes tables and views. [Learn more](#)

* Schema name

stelo

Storage location

Select external location ▼

sub/path

Location in cloud storage where data for managed tables will be stored. If the storage location is not specified, it will default to catalog location first and then metastore root location if catalog location is not specified as well.

Comment

Cancel

Create

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

The screenshot shows the Databricks Catalog Explorer interface. On the left, a sidebar lists the catalog structure under 'catalog01', including 'default', 'information_schema', 'stelo', 'hive_metastore', 'main', 'samples', 'stelolibraries', and 'system'. The main area displays the 'catalog01.stelo' schema page. At the top right, there is a 'Create' button with a dropdown menu. A red arrow points to the 'Create volume' option in this menu. Below the 'Create' button, there are tabs for 'Overview', 'Details', and 'Permissions'. The 'About this schema' section shows the owner, tags, and comment fields. At the bottom, there is a table with columns 'Name', 'Owner', 'Created at', and 'Popularity'. The table is currently empty, and a message below it states: 'You either don't have access to any tables or there are no tables in this schema. Either contact your administrator or [create a new sample table.](#)'

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 a new volume ✕

Volume name

Volume type

Managed volume

External volume

Comment (optional)

Cancel Create

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

Microsoft Azure

Home > Key vaults > SteloUCKeyVault

Key vaults
Stelo, a StarQuest company (stelodata.com)

+ Create ...

Filter for any field...

Name ↑

SteloUCKeyVault

SteloUCKeyVault | Access configuration

Search

Refresh

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Access policies

Events

Objects

Keys

Secrets

Certificates

Settings

Access configuration

Networking

Microsoft Defender for Cloud

Properties

Locks

Monitoring

Alerts

Metrics

Diagnostic settings

Logs

Insights

Workbooks

Automation

CLI / PS

Tasks (preview)

Export template

Help

Resource health

Support + Troubleshooting

Configure your options on access policy for this key vault

To access a key vault in data plane, all callers (users or applications) must have proper authentication and authorization. Authentication establishes the identity of the caller. Authorization determines which operations the caller can execute. [Learn more](#)

Permission model

Grant data plane access by using a Azure RBAC or Key Vault access policy

Azure role-based access control (recommended) [ⓘ](#)

Vault access policy [ⓘ](#)

[Go to access control\(IAM\)](#)

Resource access

Choose among the following options to grant access to specific resource types

Azure Virtual Machines for deployment [ⓘ](#)

Azure Resource Manager for template deployment [ⓘ](#)

Azure Disk Encryption for volume encryption [ⓘ](#)

Page 1 of 1

Apply Discard changes

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>

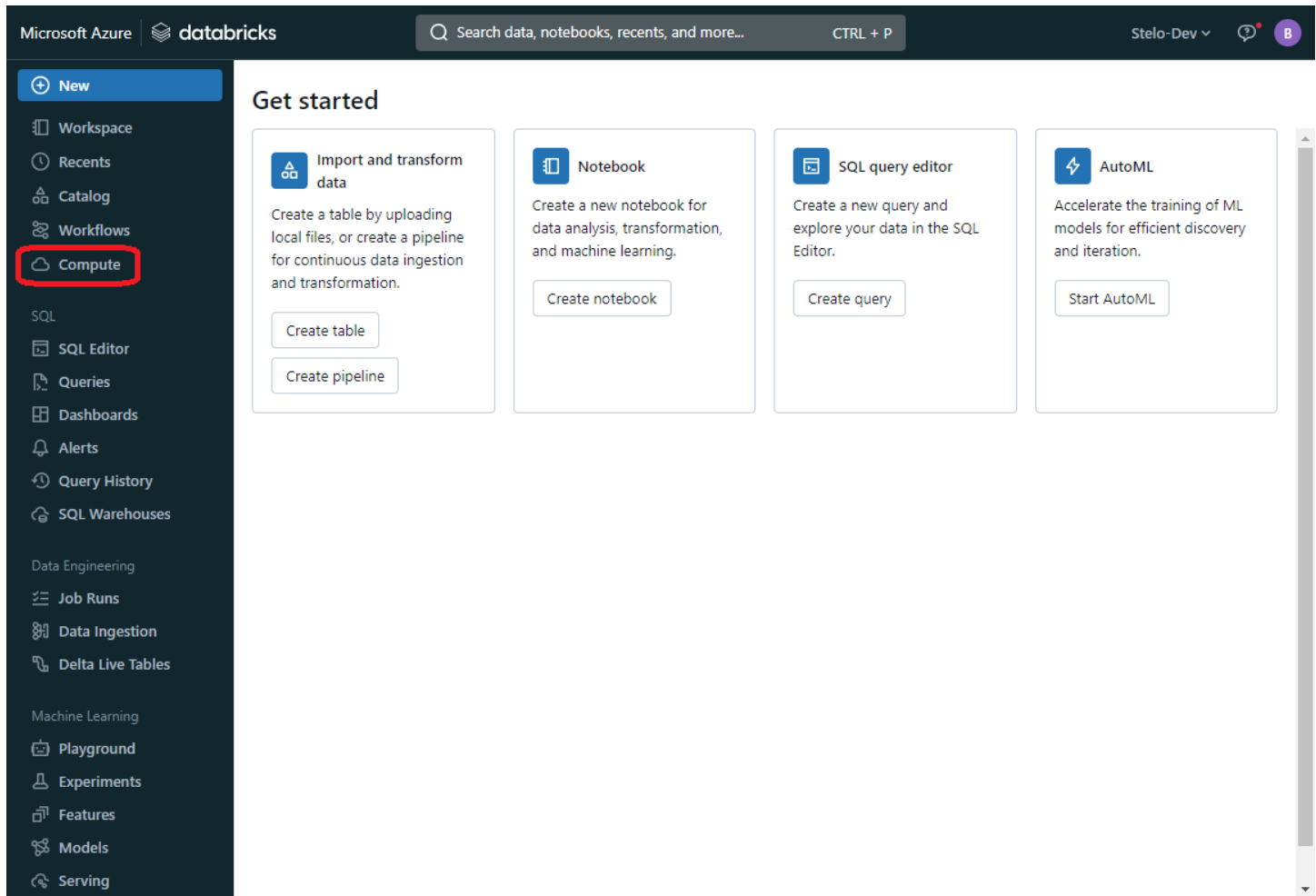
- 3) Enter the Scope Name, DNS Name (Vault URI), and the Resource ID.

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**.

Microsoft Azure | databricks | Search data, notebooks, recents, and more... | CTRL + P | Stelo-Dev | B

Compute > New compute > **DEV Cluster**

Policy [ⓘ]
Unrestricted

Multi node Single node

Access mode [ⓘ]
Shared

Performance

Databricks runtime version [ⓘ]
Runtime: 13.3 LTS (Scala 2.12, Spark 3.4.1)

Use Photon Acceleration [ⓘ]

Worker type [ⓘ]
Standard_DS3_v2 14 GB Memory, 4 Cores

Min workers: 2 | Max workers: 8 | Spot instances [ⓘ]

Driver type
Same as worker 14 GB Memory, 4 Cores

Enable autoscaling [ⓘ]
 Terminate after 120 minutes of inactivity [ⓘ]

Tags [ⓘ]
Add tags
Key Value Add

> Automatically added tags

▶ **Advanced options**

Create compute Cancel

Summary
2-8 Workers 28-112 GB Memory 8-32 Cores
1 Driver 14 GB Memory, 4 Cores
Runtime 13.3.x-scala2.12
Unity Catalog Photon Standard_DS3_v2 4-14 DBU/h