Azure Databricks: Unity Catalog Walk Through

**Version:** v1, 2023-03-09, [wcarroll@microsoft.com](mailto:wcarroll@microsoft.com)

**Public Documentation:**

<https://learn.microsoft.com/en-us/azure/databricks/data-governance/unity-catalog/>

<https://learn.microsoft.com/en-us/azure/databricks/data-governance/unity-catalog/get-started>

[Create a storage account for Azure Data Lake Storage Gen2 - Azure Storage | Microsoft Learn](https://learn.microsoft.com/en-us/azure/storage/blobs/create-data-lake-storage-account)

[Use Azure managed identities in Unity Catalog to access storage - Azure Databricks | Microsoft Learn](https://learn.microsoft.com/en-us/azure/databricks/data-governance/unity-catalog/azure-managed-identities)

<https://learn.microsoft.com/en-us/azure/active-directory/managed-identities-azure-resources/how-manage-user-assigned-managed-identities>

[Unity Catalog best practices - Azure Databricks | Microsoft Learn](https://learn.microsoft.com/en-us/azure/databricks/data-governance/unity-catalog/best-practices)

[Using Unity Catalog with Structured Streaming - Azure Databricks | Microsoft Learn](https://learn.microsoft.com/en-us/azure/databricks/structured-streaming/unity-catalog)

[Tutorial: Unity Catalog metastore admin tasks for Databricks SQL - Azure Databricks | Microsoft Learn](https://learn.microsoft.com/en-us/azure/databricks/getting-started/uc-metastore-admin-quickstart)

[Unity Catalog privileges and securable objects - Azure Databricks | Microsoft Learn](https://learn.microsoft.com/en-us/azure/databricks/data-governance/unity-catalog/manage-privileges/privileges)

[Working with Unity Catalog in Azure Databricks - Microsoft Community Hub](https://techcommunity.microsoft.com/t5/fasttrack-for-azure/working-with-unity-catalog-in-azure-databricks/ba-p/3693781)

**Wiki Page:**

[Unity Catalog - Overview (azure.com)](https://dev.azure.com/supportability/AzureDataBricks/_wiki/wikis/AzureDataBricks.wiki/597207/Unity-Catalog)

[Key Steps for setting up Unity Catalog Metastore - Overview (azure.com)](https://dev.azure.com/supportability/AzureDataBricks/_wiki/wikis/AzureDataBricks.wiki/645730/Key-Steps-for-setting-up-Unity-Catalog-Metastore)

**Note:** FTE’s can get an MSDN Visual Studio Account with $150 per month charges. This is required because you must be an Azure Active Directory Global Administrator to set up unity catalog.

**Walk Through:**

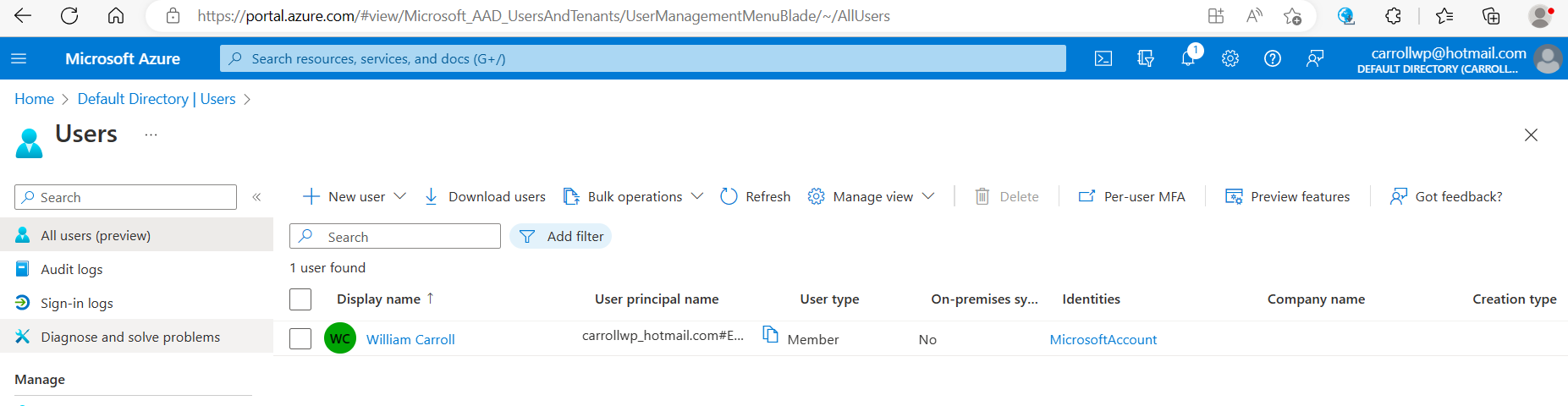
In your Azure Active Directory, your user must be part of the Global Administrators group. I am using [carrollwp@hotmail.com](mailto:carrollwp@hotmail.com) as my user account that is part of the Azure Active Directory Global Administrators group.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated



Graphical user interface, text, application, email

Description automatically generated

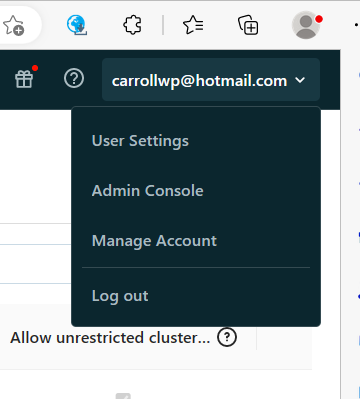
With the user, create a default Databricks workspace. It must be a premium pricing tier to use unity catalog.

Graphical user interface, text, application, email

Description automatically generated

The user must be administrator of the Databricks workspace.

Admin Console:



Graphical user interface, text, application

Description automatically generated

When you click Manage Account it should take you to accounts.auredatabricks.net. If it takes you to the Azure portal your user account is not a Global Administrator in the Azure Active Directory.

Manage Account:

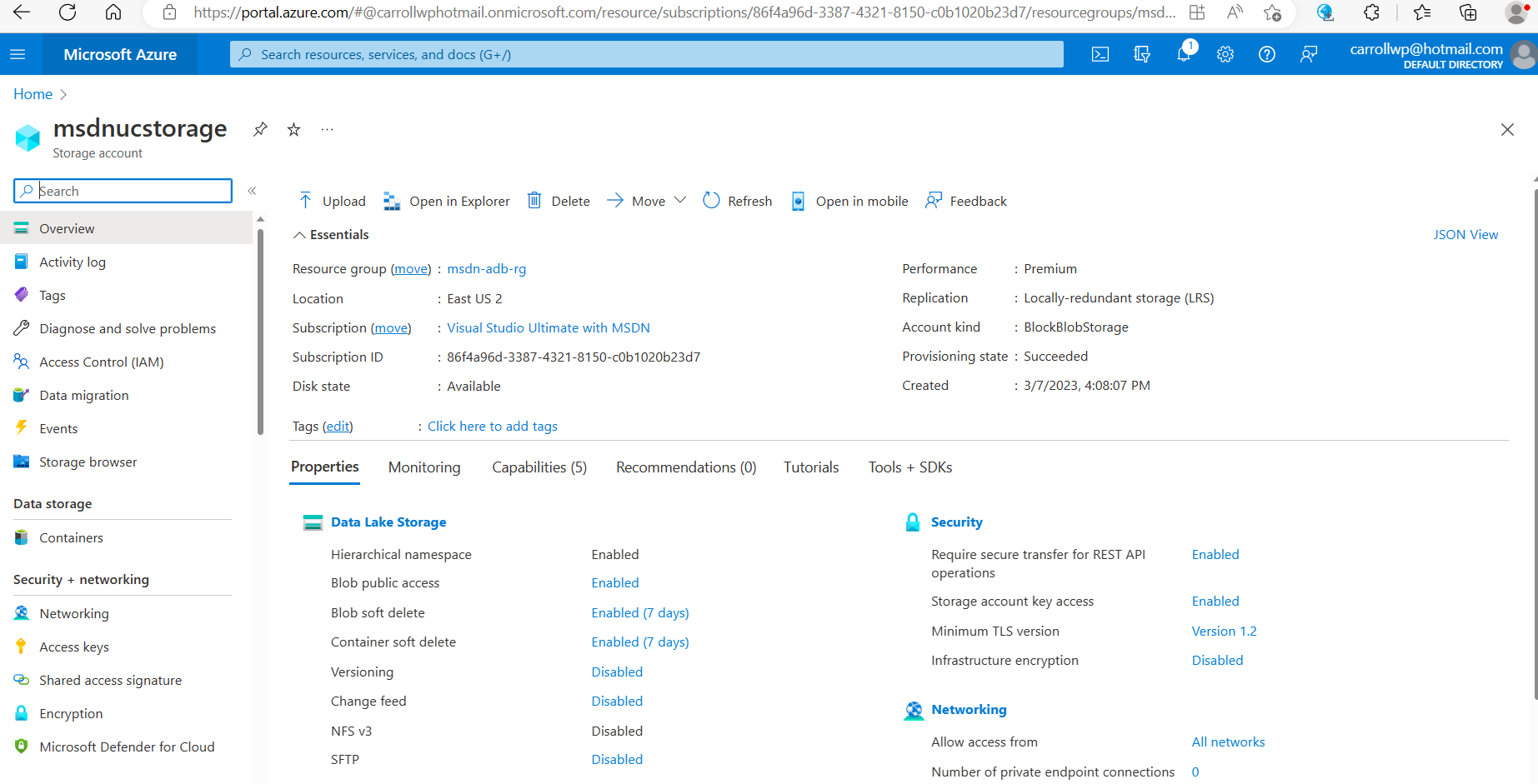
Graphical user interface, application

Description automatically generated

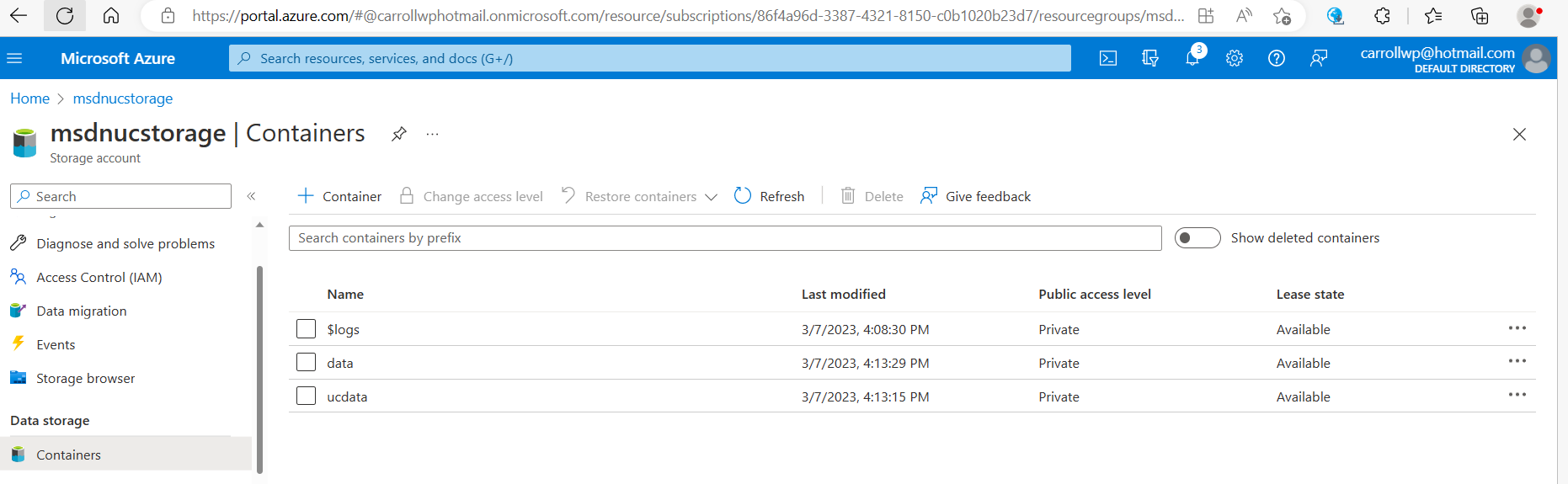
Graphical user interface, text, application, email

Description automatically generated

Create an Azure Gen2 Storage account with BlockBlob, Hierarchal Namespace enabled. This is the storage account that your unity catalog managed and external tables will be stored, instead of in DBFS.



Create a container, ucdata, for unity catalog.



Create a folder in the container to hold managed tables and external tables for the region.

Graphical user interface, text, application, email

Description automatically generated

Create and save off the abfss path to the managed tables location.

abfss://<container-name>@<storage-account-name>.dfs.core.windows.net/<metastore-name>

abfss://ucdata@msdnucstorage.dfs.core.windows.net/eastus2\_uc\_metastore

Click the + resource and search for “Access Connector for Azure Databricks”.

Graphical user interface, application

Description automatically generated

Create the connector in the same region as your Databricks Workspace.

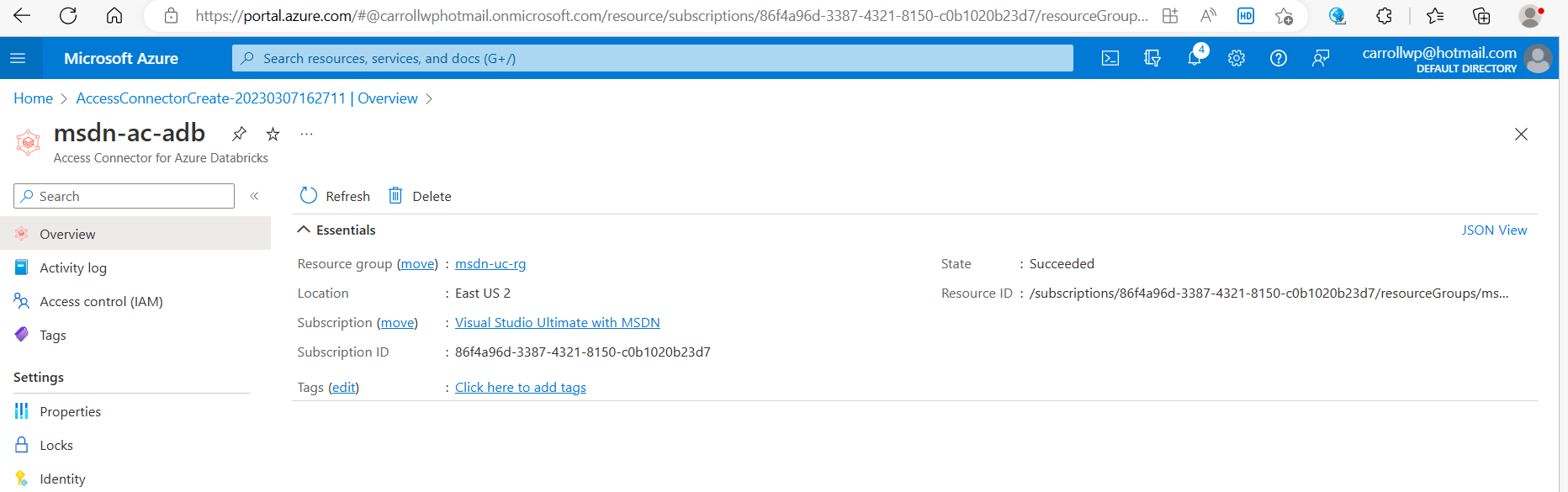
Graphical user interface, text, application, email

Description automatically generated

Can create either a system managed identity or a user defined managed identity. Creates by default a system managed identity. See in documentation, details for a user defined managed identity. In this example I created a new resource group to hold my connector.

Save off the connectors resource id. This can be found on the overview or properties blade.

/subscriptions/86f4a96d-3387-4321-8150-c0b1020b23d7/resourceGroups/msdn-uc-rg/providers/Microsoft.Databricks/accessConnectors/msdn-ac-adb



Access Connector for Azure Databricks in its resource group.

Graphical user interface, text, application, email

Description automatically generated

In the Azure Active Directory, under Enterprise Applications, you can search for the connectors object\_id or name.

Graphical user interface, text, application, email

Description automatically generated

Add your user account and managed identity to the Azure Storage Account in the Storage Data Blob Contributor role.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated

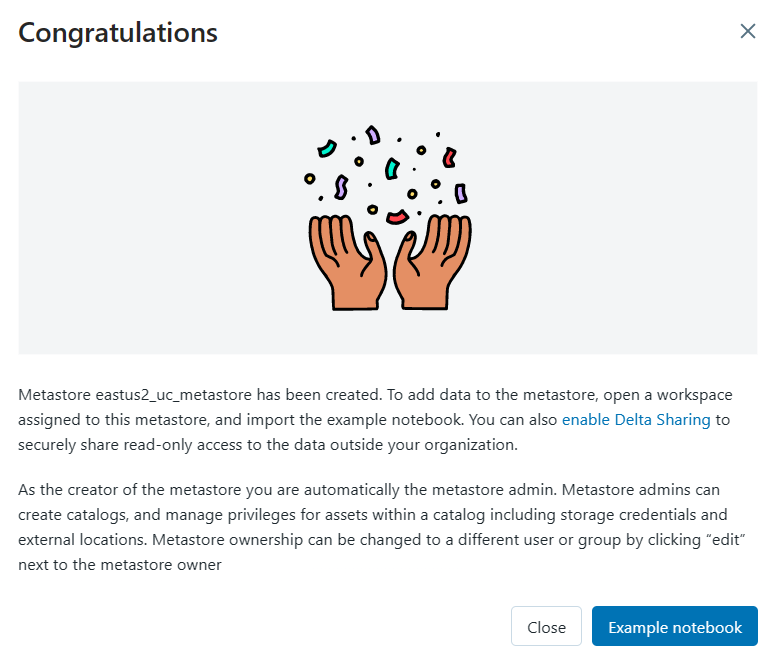
In the Databricks workspace, Managed Accounts UI, assign and enable the workspace to the unity catalog for the eastus2 region.

A screenshot of a computer

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated



Graphical user interface, text, application

Description automatically generated

In the Databricks workspace UI, under the data blade, you will see the storage credential for the system managed identity for the connector to the storage account. Previously we added the system managed identity to the storage accounts Blob Storage Contributor Role in IAM.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

In the Databricks workspace UI, under the data blade, specify the storage location used for external tables.

Graphical user interface, text, application

Description automatically generated

Fill in the information to create the external location for unity catalog,

Graphical user interface, text, application, email

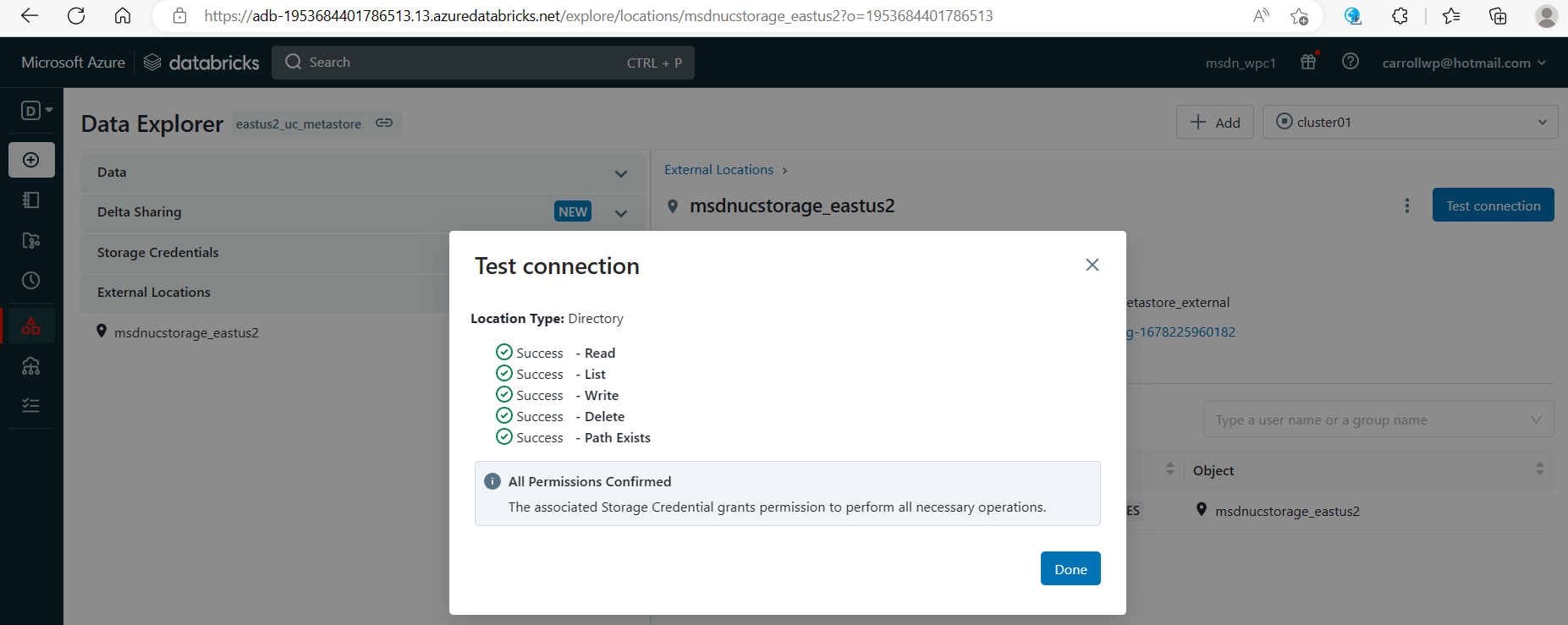
Description automatically generated

Once the external location is done, grant privileges to your user.

A screenshot of a computer

Description automatically generated

You can test the connection and should see success.



You can now create and start a cluster.

If you create a single user cluster, you will get the following error. Create a shared cluster.

Graphical user interface, text, application

Description automatically generated

You can create a python notebook and view the unity catalog objects and create managed and external tables. Some sample commands to try are below.

spark.sql("create catalog quick\_start")

spark.sql("show catalogs").show()

spark.sql("use catalog quickstart\_catalog")

spark.sql("show databases").show()

spark.sql("create table t1 (id int, name string)")

display(spark.sql("show tables"))

display(spark.sql("describe table extended t1"))

spark.sql("insert into t1 (id, name) values (1, 'Delaware')")

spark.sql("select \* from t1").show()

display(spark.sql("select \* from information\_schema.tables where table\_name = 't1'"))

spark.sql("create external table t2 (id int, name string) using delta location 'abfss://ucdata@msdnucstorage.dfs.core.windows.net/eastus2\_uc\_metastore\_external/t2'")

spark.sql("insert into t2 (id, name) values (2, 'Virginia')")

spark.sql("select \* from t2").show()

In the Databricks workspace UI under the data blade, you can see the catalogs, databases, schemas, tables and views. You can see the details and permissions.

