

# Criando Unity Catalog na AWS:

Como visto no tutorial de criação de Workspaces, ao visualizar o status **CREATE\_COMPLETE** durante CloudFormation, o UC já vem de forma automatizada:

| Timestamp                    | Logical ID                       | Status             | Status reason               |
|------------------------------|----------------------------------|--------------------|-----------------------------|
| 2023-11-28 10:29:58 UTC-0300 | databricks-workspace-stack-63120 | CREATE_COMPLETE    | -                           |
| 2023-11-28 10:29:55 UTC-0300 | createWorkspace                  | CREATE_COMPLETE    | -                           |
| 2023-11-28 10:29:55 UTC-0300 | createWorkspace                  | CREATE_IN_PROGRESS | Resource creation initiated |
| 2023-11-28 10:26:53 UTC-0300 | createWorkspace                  | CREATE_IN_PROGRESS | -                           |
| 2023-11-28 10:26:53 UTC-0300 | createCredentials                | CREATE_COMPLETE    | -                           |
| 2023-11-28 10:26:52 UTC-0300 | createCredentials                | CREATE_IN_PROGRESS | Resource creation initiated |
| 2023-11-28 10:26:52 UTC-0300 | createStorageConfiguration       | CREATE_COMPLETE    | -                           |
| 2023-11-28 10:26:52 UTC-0300 | createStorageConfiguration       | CREATE_IN_PROGRESS | Resource creation initiated |
| 2023-11-28 10:26:49 UTC-0300 | createCredentials                | CREATE_IN_PROGRESS | -                           |
| 2023-11-28 10:26:49 UTC-0300 | createStorageConfiguration       | CREATE_IN_PROGRESS | -                           |
| 2023-11-28 10:26:48 UTC-0300 | databricksApiFunction            | CREATE_COMPLETE    | -                           |
| 2023-11-28 10:26:42 UTC-0300 | databricksApiFunction            | CREATE_IN_PROGRESS | Resource creation initiated |
| 2023-11-28 10:26:41 UTC-0300 | databricksApiFunction            | CREATE_IN_PROGRESS | -                           |

E então exibirá o workspace criado junto ao Unity Catalog também configurado com link para acesso ao Databricks Workspace. Interessante notar que a automatização do template por default irá nomear o Unity Catalog de acordo com a região escolhida:

| Name                | Status  | Pricing tier | Region    | Bucket name          | Credential name      | Created           | Metastore          |
|---------------------|---------|--------------|-----------|----------------------|----------------------|-------------------|--------------------|
| databricksworksp... | Running | Premium      | us-west-1 | databricks-worksp... | databricks-worksp... | today at 10:43 AM | metastore_aws_u... |
| databricksseq...    | Running | Premium      | us-west-2 | databricks-worksp... | databricks-worksp... | today at 10:26 AM | metastore_aws_u... |

Válido lembrar que um Metastore/UC pode agregar vários Workspace de mesma região, mas não é possível um Workspace ter dois ou mais UC diferentes anexados. Mas a criação de novos UC/Metastore também são possíveis através de Quickstart, que irão apresentar um template com configurações IARN de IAM e bucket criado (será visto na criação de External Location).

# Criando Unity Catalog na GCP:

Para criar seu Unity Catalog na GCP, primeiramente você deve estar criando um bucket que servirá como espaço de armazenamento do metastore:

Google Cloud | My First Project | storage | Search

### Create a bucket

✓ Name your bucket

Name: databricks-gcp-unitycatalog

• Choose where to store your data

This choice defines the geographic placement of your data and affects cost, performance, and availability. Cannot be changed later. [Learn more](#)

Location type

☐ Multi-region

Highest availability across largest area

☐ Dual-region

High availability and low latency across 2 regions

☒ Region

Lowest latency within a single region

us-central1 (Iowa)

CONTINUE

• Choose a storage class for your data

Default storage class: Standard

• Choose how to control access to objects

Public access prevention: On

Access control: Uniform

• Choose how to protect object data

Protection tools: None

Data encryption: Google-managed

CREATE CANCEL

Good to know

Location pricing

Storage rates vary depending on the storage class of your data and location of your bucket. [Pricing details](#)

Current configuration: Region / Standard

| Item               | Cost                 |
|--------------------|----------------------|
| us-central1 (Iowa) | \$0.020 per GB-month |

ESTIMATE YOUR MONTHLY COST

**LEMBRE-SE:** o seu bucket deve estar sempre em uma **mesma REGIÃO** que o seu Workspace!

Já na página da sua Account Console, você pode então ir na aba de Data e visualizar a opção de criar um Metastore:

databricks Account | vilton.work@gmail.com

Get started with Unity Catalog

Unity Catalog is a fine-grained governance solution for data and AI on the Lakehouse

As an account admin you can create and administer a metastore. You can then add specific workspaces to this metastore. A metastore and assigned workspaces must all be in the same region.

Follow the steps below to get started

As an account admin

Create a metastore

Add users and groups

Data

Metastores

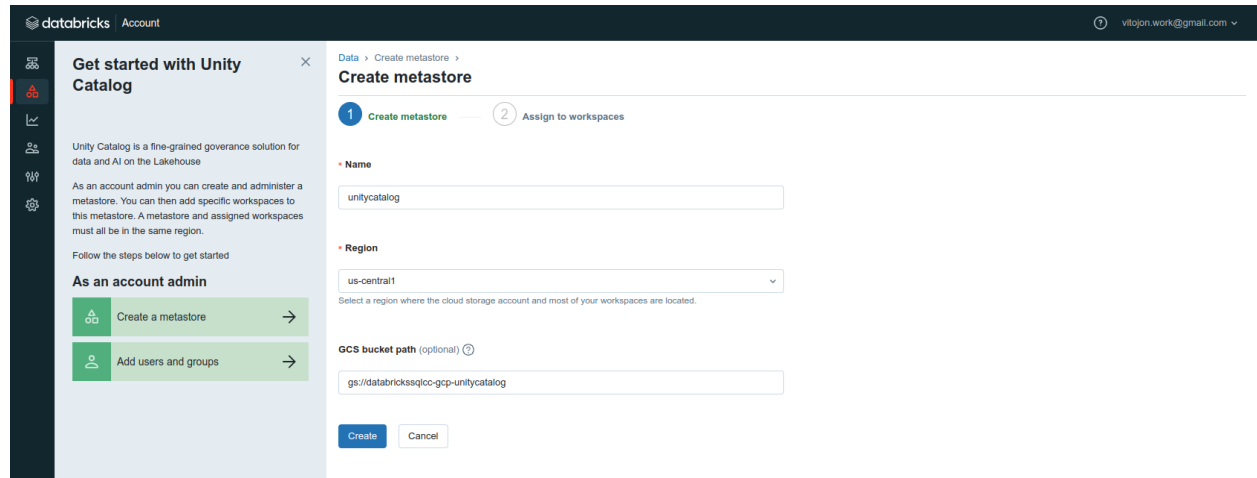
A metastore is the top-level container for data in Unity Catalog. Within a metastore, Unity Catalog provides a 3-level namespace for organizing data: catalogs, schemas (also called databases), and tables / views. [Learn More](#)

Create metastore

Create a Metastore

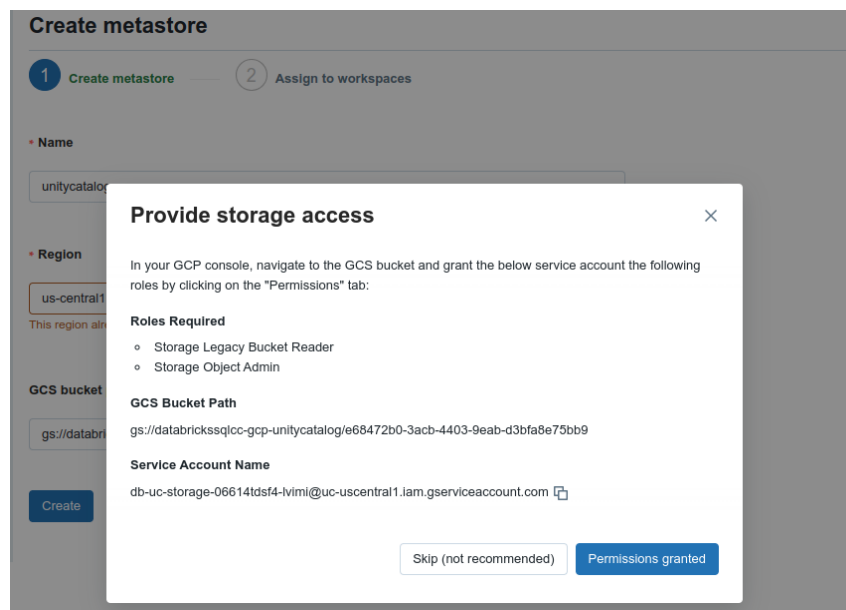
| Name          | Region | Path | Created at |
|---------------|--------|------|------------|
| No metastores |        |      |            |

Na tela de criação do Metastore, certifique-se de colocar um nome para seu Unity Catalog, assim como a região em que está o bucket criado e o seu path/URL:



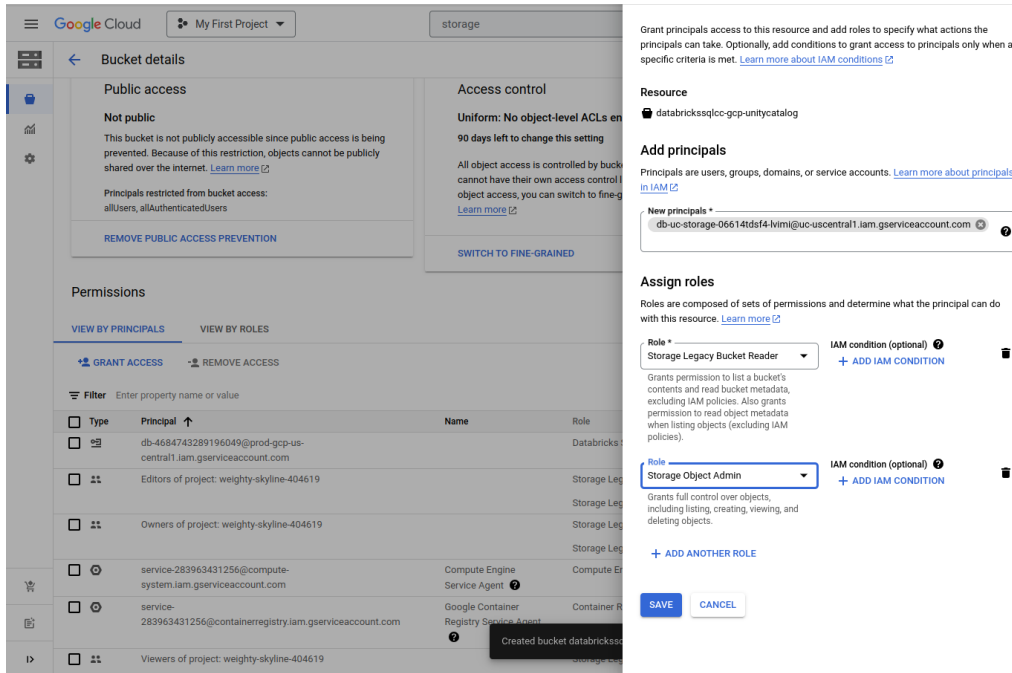
The screenshot shows the Databricks 'Create metastore' interface. On the left, a sidebar titled 'Get started with Unity Catalog' provides instructions and links to 'Create a metastore' and 'Add users and groups'. The main form area is titled 'Create metastore' and includes a progress indicator with two steps: '1 Create metastore' and '2 Assign to workspaces'. The form fields are: 'Name' (text input with 'unitycatalog'), 'Region' (dropdown menu with 'us-central1'), and 'GCS bucket path (optional)' (text input with 'gs://databrickssqlcc-gcp-unitycatalog'). At the bottom of the form are 'Create' and 'Cancel' buttons.

Ao clicar para criar o Metastore, uma janela é exibida com as instruções necessárias para configuração de autorização no permissionamento do bucket GCS configurado para que a Databricks tenha seus devidos controles (Storage Legacy Bucket Reader e Storage Object Admin) e então uma Service Account da GCP é fornecida:



The screenshot shows a 'Provide storage access' dialog box overlaid on the 'Create metastore' form. The dialog box contains the following information: 'In your GCP console, navigate to the GCS bucket and grant the below service account the following roles by clicking on the "Permissions" tab:', 'Roles Required' (Storage Legacy Bucket Reader, Storage Object Admin), 'GCS Bucket Path' (gs://databrickssqlcc-gcp-unitycatalog/e68472b0-3acb-4403-9eab-d3bfa8e75bb9), and 'Service Account Name' (db-uc-storage-06614tdsf4-lvimi@uc-uscentral1.iam.gserviceaccount.com). At the bottom of the dialog are 'Skip (not recommended)' and 'Permissions granted' buttons.

Vá então até o bucket criado para ser a base do Unity Catalog e, na aba “Permissions”, clique em “Grant Access” fornecendo então as condições de [Storage Legacy Bucket Reader](#) e [Storage Object Admin](#) para a Service Account fornecida e então salve os novos acessos:



Grant principals access to this resource and add roles to specify what actions the principals can take. Optionally, add conditions to grant access to principals only when a specific criteria is met. [Learn more about IAM conditions](#)

**Resource**  
databricksqjcc-gcp-unitycatalog

**Add principals**  
Principals are users, groups, domains, or service accounts. [Learn more about principals in IAM](#)

**New principals \***  
db-uc-storage-06614tdsf4-lvimi@uc-uscentral1.iam.gserviceaccount.com

**Assign roles**  
Roles are composed of sets of permissions and determine what the principal can do with this resource. [Learn more](#)

**Role \***  
Storage Legacy Bucket Reader [+ ADD IAM CONDITION](#)

Grants permission to list a bucket's contents and read bucket metadata, excluding IAM policies. Also grants permission to read object metadata when listing objects (excluding IAM policies).

**Role**  
Storage Object Admin [+ ADD IAM CONDITION](#)

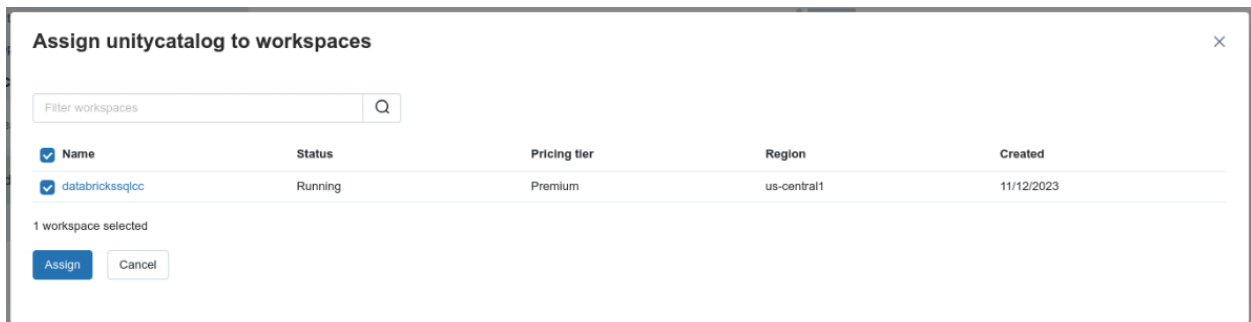
Grants full control over objects, including listing, creating, viewing, and deleting objects.

[+ ADD ANOTHER ROLE](#)

[SAVE](#) [CANCEL](#)

| Type                     | Principal  | Name                                    | Role                             |
|--------------------------|--|---|----------------------------------|
| <input type="checkbox"/> | db-4684743289196049@prod-gcp-us-central1.iam.gserviceaccount.com |   | Databricks                       |
| <input type="checkbox"/> | Editors of project: weighty-skyline-404619                       |   | Storage Legacy Bucket Reader     |
| <input type="checkbox"/> | Owners of project: weighty-skyline-404619                        |   | Storage Legacy Bucket Reader     |
| <input type="checkbox"/> | service-283963431256@compute-system.iam.gserviceaccount.com      | Compute Engine Service Agent            | Compute Engine Service Agent     |
| <input type="checkbox"/> | service-283963431256@containerregistry.iam.gserviceaccount.com   | Google Container Registry Service Agent | Container Registry Service Agent |
| <input type="checkbox"/> | Viewers of project: weighty-skyline-404619                       |   | Storage Legacy Bucket Reader     |

Retorne para a Databricks, já confirmando a concessão de acessos e verá então uma janela mostrando os Workspaces disponíveis que podem ser aplicados o Unity Catalog em devida região:



**Assign unitycatalog to workspaces**

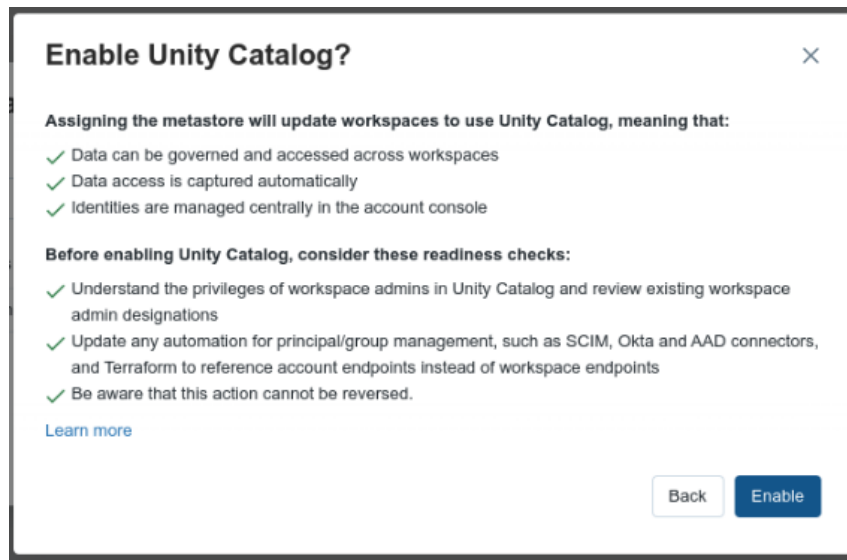
Filter workspaces

| Name           | Status  | Pricing tier | Region      | Created    |
|----------------|---------|--------------|-------------|------------|
| databricksqjcc | Running | Premium      | us-central1 | 11/12/2023 |



1 workspace selected

[Assign](#) [Cancel](#)

Confirme a habilitação do Unity Catalog:



E por fim encontrará o Unity Catalog funcionando em seu Workspace:

**Catalog Explorer** `unitycatalog`   [Send feedback](#)

# Criando Unity Catalog na Azure:

Assim como na GCP, crie um bucket na Azure para criação do Unity Catalog, porém se atentando aos detalhes de Resource Group, a região (mesma que do Workspace, sempre vale relembrar), em performance Premium e de tipo Block Blobs:

Microsoft Azure

Home > Storage accounts >

### Create a storage account

Basics Advanced Networking Data protection Encryption Tags Review

**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 \* databricks\_workshop\_mdw  
[Create new](#)

**Instance details**

Storage account name \* azureucdatabrickssql

Region \* (US) East US 2  
[Deploy to an edge zone](#)

Performance \* ☐ Standard: Recommended for most scenarios (general-purpose v2 account)  
☒ Premium: Recommended for scenarios that require low latency.

Premium account type \* Block blobs

Redundancy \* Locally-redundant storage (LRS)

[Review](#) < Previous Next: Advanced > [Give feedback](#)

Após criação bem sucedida do seu Storage, vá até a configuração de management para Data Protection e desabilite os checks que permitem Soft Delete para Blobs e Containers (Enable ativo irá impedir a criação do Unity Catalog):

azureucdatabrickssql | Data protection

Storage account

Search

Overview  
Activity log  
Tags  
Diagnose and solve problems  
Access Control (IAM)  
Data migration  
Events  
Storage browser  
Storage Mover  
Data storage  
Containers  
Security + networking  
Networking  
Front Door and CDN  
Access keys  
Shared access signature  
Encryption  
Microsoft Defender for Cloud  
Data management  
Redundancy  
Data protection

Data protection provides options for recovering your data when it is erroneously modified or deleted.

**Recovery**

☐ Enable Azure Backup for blobs

☐ Enable point-in-time restore for containers

☐ 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 >](#)

**i** You will still be able to access and recover soft deleted data if the soft delete feature is disabled.

☐ Enable soft delete for containers

Soft delete enables you to recover containers that were previously marked for deletion. [Learn more >](#)

**i** You will still be able to access and recover soft deleted data if the container soft delete feature is disabled.

☐ Enable permanent delete for soft deleted items

**Tracking**

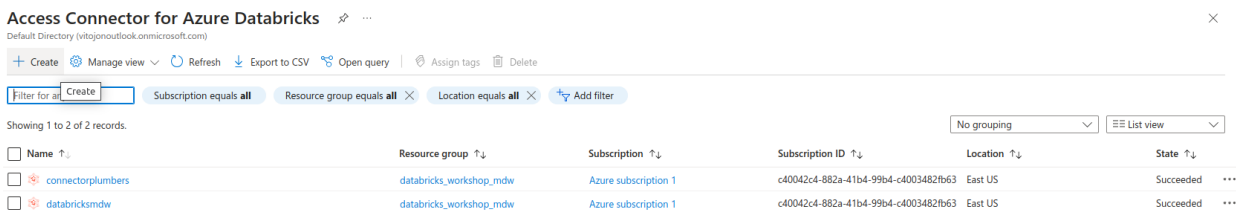
☐ Enable versioning for blobs

☐ Enable blob change feed

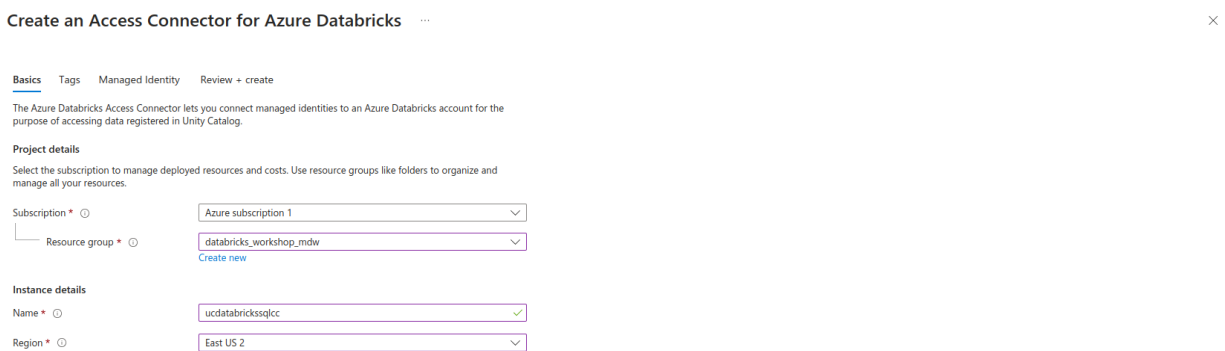
**Access control**

☐ Enable version-level immutability support

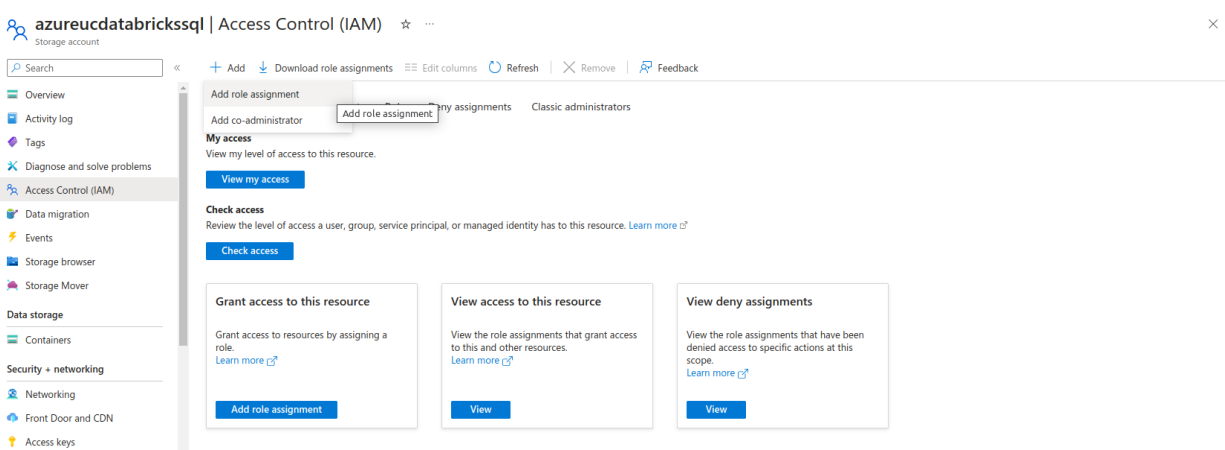
Procure na barra de pesquisas da Azure pelo “Access Connector for Azure Databricks” e clique na opção de criar para obter o seu conector, necessário para integrar permissionamento entre Storage e Workspace:



Aponte seu Resource Group e região indicada para criação desse conector, assim como seu nome para identificação:



Uma vez com o conector criado, compareça até o IAM do seu Storage e adicione uma role, lá você irá incluir a opção de **Storage Blob Data Contributor** para o conector criado na aba de Roles:



## Add role assignment

Role Members Conditions (optional) Review + assign

A role definition is a collection of permissions. You can use the built-in roles or you can create your own custom roles. [Learn more](#) <sup>d</sup>

Assignment type

**Job function roles** Privileged administrator roles

Grant access to Azure resources based on job function, such as the ability to create virtual machines.

blob Type: All Category: All

| Name                              | Description   | Type        | Category | Details              |
|-----------------------------------|---|-------------|----------|----------------------|
| Defender for Storage Data Scanner | Grants access to read blobs and update index tags. This role is used by the data scanner of Defender for Storage. | BuiltInRole | None     | <a href="#">View</a> |
| Storage Blob Data Contributor     | Allows for read, write and delete access to Azure Storage blob containers and data                                | BuiltInRole | Storage  | <a href="#">View</a> |
| Storage Blob Data Owner           | Allows for full access to Azure Storage blob containers and data, including assigning POSIX access control.       | BuiltInRole | Storage  | <a href="#">View</a> |
| Storage Blob Data Reader          | Allows for read access to Azure Storage blob containers and data  | BuiltInRole | Storage  | <a href="#">View</a> |
| Storage Blob Delegator            | Allows for generation of a user delegation key which can be used to sign SAS tokens                               | BuiltInRole | Storage  | <a href="#">View</a> |

Showing 1 - 5 of 5 results.

Já na aba de Members, adicione uma Managed Identity, onde poderá ser encontrado o seu conector para Databricks:

## Add role assignment

Role Members Conditions (optional) Review + assign

**Selected role** Storage Blob Data Contributor

**Assign access to** ☐ User, group, or service principal ☒ Managed identity

**Members** + Select members

| Name                | Object ID | Type |
|---------------------|-----------|------|
| No members selected |           |      |

**Description** Optional

Some results might be hidden due to your ABAC condition.

Subscription \*  
Azure subscription 1

Managed identity  
Access Connector for Azure Databricks (3)

Select

Search by name

connectorplumbers  
/subscriptions/c40042c4-882a-41b4-99b4-c4003482fb63/resourceGroups/databri...

databricksmdw  
/subscriptions/c40042c4-882a-41b4-99b4-c4003482fb63/resourceGroups/databri...

Selected members:  
ucdatabrickssqlcc  
/subscriptions/c40042c4-882a-41b4-99b4-c4003482fb63/resourceGroups/... [Remove](#)

Após configuração de IAM, crie um container dando-lhe o nome que for conveniente em seu Storage:

Microsoft Azure Search resources, services, and docs (G+/I) vitojon@outlook.com DEFAULT DIRECTORY (VITOJON...)

Home > azureucdatabrickssql | Containers Storage account

Search

+ Container Change access level Restore containers Refresh Delete Give feedback

Search containers by prefix

| Name                                  | Last modified          | Anonymous access level |
|---------------------------------------|------------------------|------------------------|
| <input type="checkbox"/> \$logs       | 11/27/2023, 8:11:32 PM | Private                |
| <input type="checkbox"/> unitycatalog | 11/27/2023, 8:16:24 PM | Private                |

New container

Name \*

Anonymous access level ☐ Private (no anonymous access)

The access level is set to private because anonymous access is disabled on this storage account.

Advanced



Por fim, se preferir, pode ir até o campo de Endpoints no Storage para copiar seu endereço de Storage (será pedido na Databricks):

The screenshot shows the Azure portal interface for the storage account 'azureudatabrickssql'. The left sidebar contains navigation options like 'Front Door and CDN', 'Access keys', 'Shared access signature', 'Encryption', 'Microsoft Defender for Cloud', 'Data management', 'Redundancy', 'Data protection', 'Object replication', 'Blob inventory', 'Static website', 'Lifecycle management', 'Azure search', 'Settings', 'Configuration', 'Data Lake Gen2 upgrade', 'Resource sharing (CORS)', 'Advisor recommendations', and 'Endpoints'. The main content area displays the 'Endpoints' page with a table of endpoints. The table has columns for 'Provisioning state' (Succeeded), 'Created' (11/27/2023, 8:11:07 PM), 'Storage account resource ID', 'Blob service', 'Data Lake Storage', and 'Static website'. The 'Data Lake Storage' endpoint is highlighted with a blue selection bar.

| Provisioning state | Created                | Storage account resource ID  | Blob service  | Data Lake Storage                                 | Static website  |
|--------------------|------------------------|--|---|---|---|
| Succeeded          | 11/27/2023, 8:11:07 PM | /subscriptions/c40042c4-882a-41b4-99b4-c4003482fb63/resourceGroups/databricks_workshop_mdw/providers/Microsoft.Storage/storageAccounts/azureudatabrickssql | /subscriptions/c40042c4-882a-41b4-99b4-c4003482fb63/resourceGroups/databricks_workshop_mdw/providers/Microsoft.Storage/storageAccounts/azureudatabrickssql/blobServices/default | https://azureudatabrickssql.dfs.core.windows.net/ | https://azureudatabrickssql.z20.web.core.windows.net/ |

Assim como também deve copiar o Resource ID do seu conector (também será pedido na criação do Unity Catalog):

The screenshot shows the Azure portal interface for the Access Connector 'ucdatabrickssqlcc'. The left sidebar contains navigation options like 'Overview', 'Activity log', 'Access control (IAM)', 'Tags', 'Settings', 'Properties', 'Locks', and 'Identity'. The main content area displays the 'Overview' page with a table of essential information. The table has columns for 'Resource group', 'Location', 'Subscription', 'Subscription ID', 'Tags', 'State', and 'Resource ID'. The 'Resource ID' is highlighted with a blue selection bar.

| Resource group          | Location  | Subscription         | Subscription ID                      | Tags     | State     | Resource ID  |
|-------------------------|-----------|----------------------|--------------------------------------|----------|-----------|--|
| databricks_workshop_mdw | East US 2 | Azure subscription 1 | c40042c4-882a-41b4-99b4-c4003482fb63 | Add tags | Succeeded | /subscriptions/c40042c4-882a-41b4-99b4-c4003482fb63/resourceGroups/databricks_workshop_mdw/providers/Microsoft.Storage/storageAccounts/azureudatabrickssql |

Compareça na página de Metastore da Account Console em Data:

Data > Create metastore >

## Create metastore

1 Create metastore

2 Assign to workspaces

▪ Name

▪ Region

Select a region where the storage account and most of your workspaces are located.

**ADLS Gen 2 path** (optional) (?)

Do not grant users direct access to this path.

**Access Connector Id** (?)

▼ Advanced options

Create

Cancel

Region: mesma que do Workspace

ADLS Gen 2 Path formato: **<container\_name>@<storage\_account\_name>.dfs.core.windows.net**  
(possível ser copiado o trecho em Endpoints precisando adicionar apenas o container@)

Access Connector Id: copiar resource ID da página do seu conector.