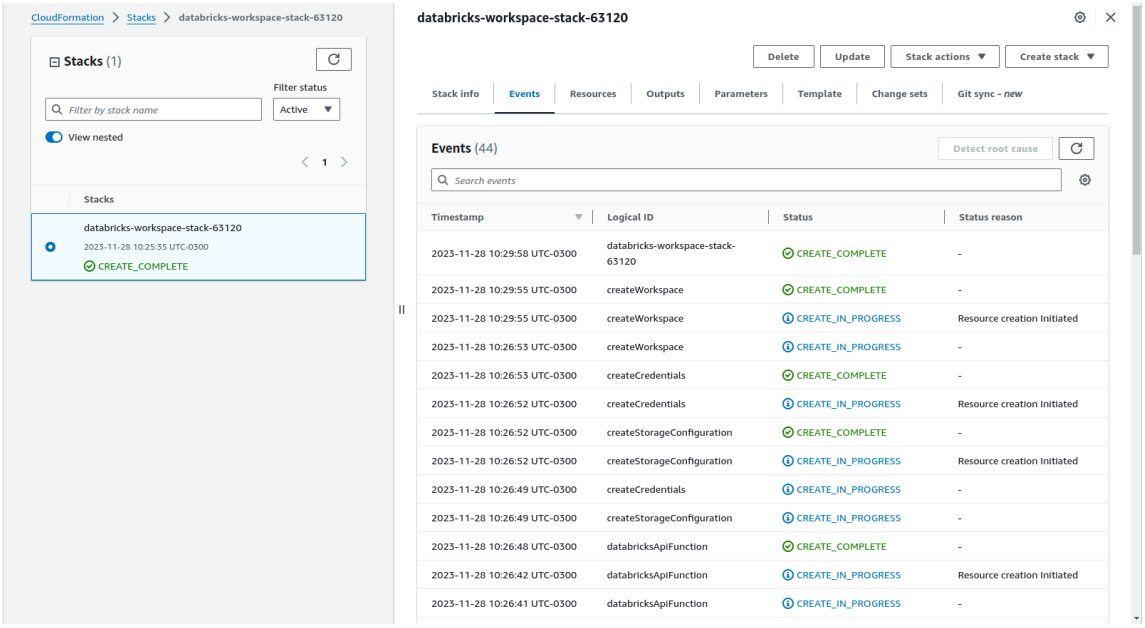
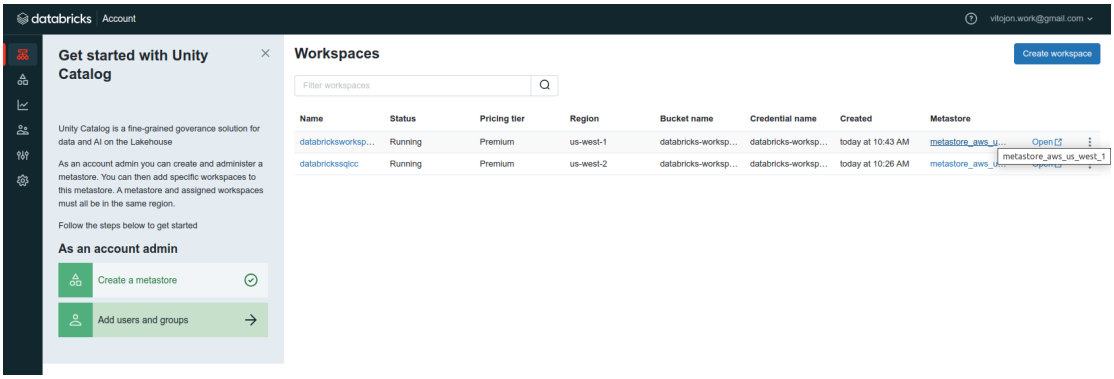
Criando Unity Catalog na AWS:

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



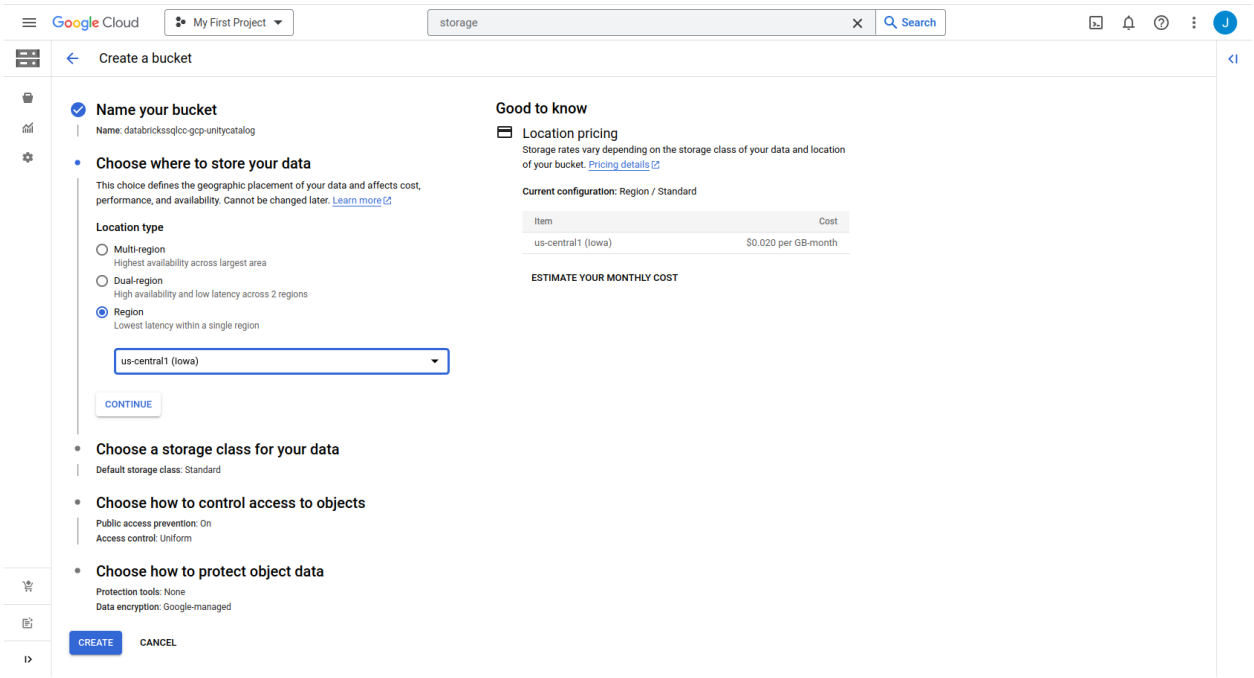
E então exibirá o workspace criado junto ao Unity Catalog também configurado com link para acesso ao Databricks Workspace. Interessante notar que a automatização do template por default irá nomear o Unity Catalog de acordo com a região escolhida:



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).

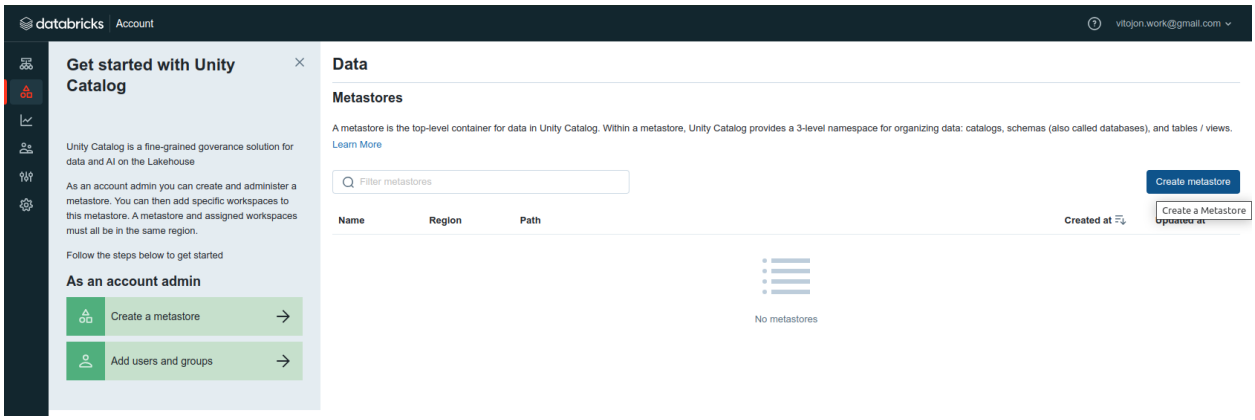
Criando Unity Catalog na GCP:

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

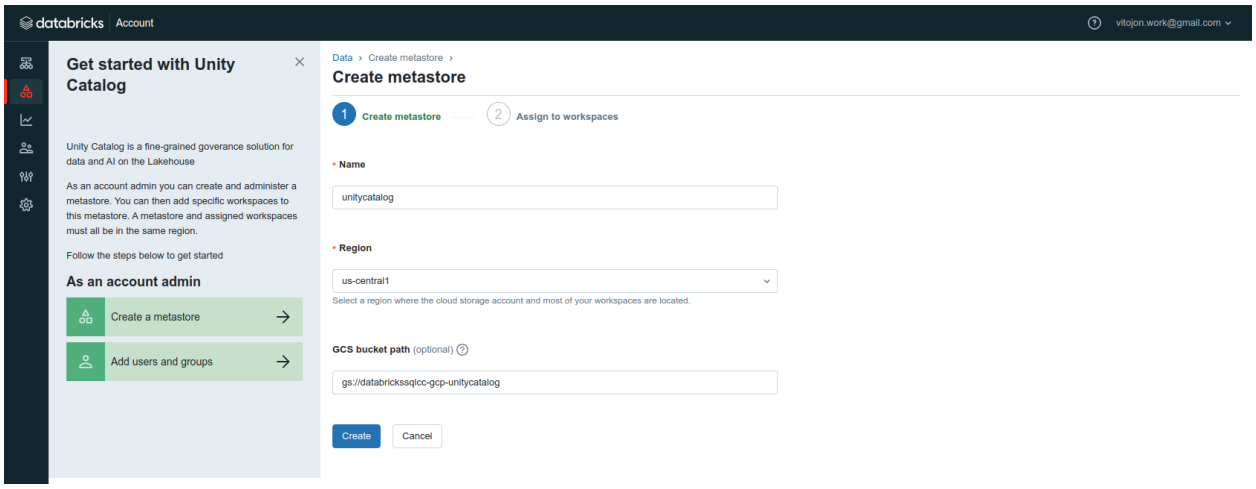


LEMBRE-SE: o seu bucket deve estar sempre em uma mesma REGIÃO que o seu Workspace!

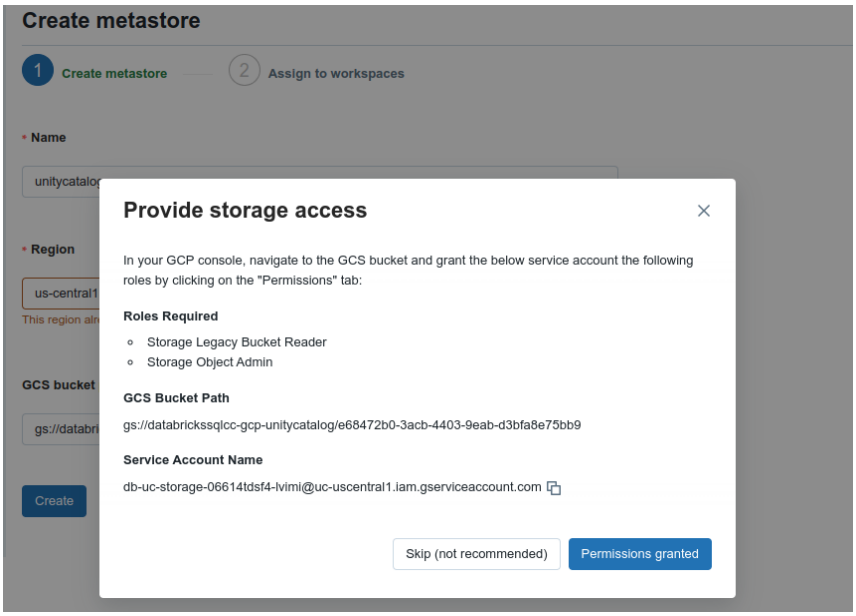
Já na página da sua Account Console, você pode então ir na aba de Data e visualizar a opção de criar um Metastore:



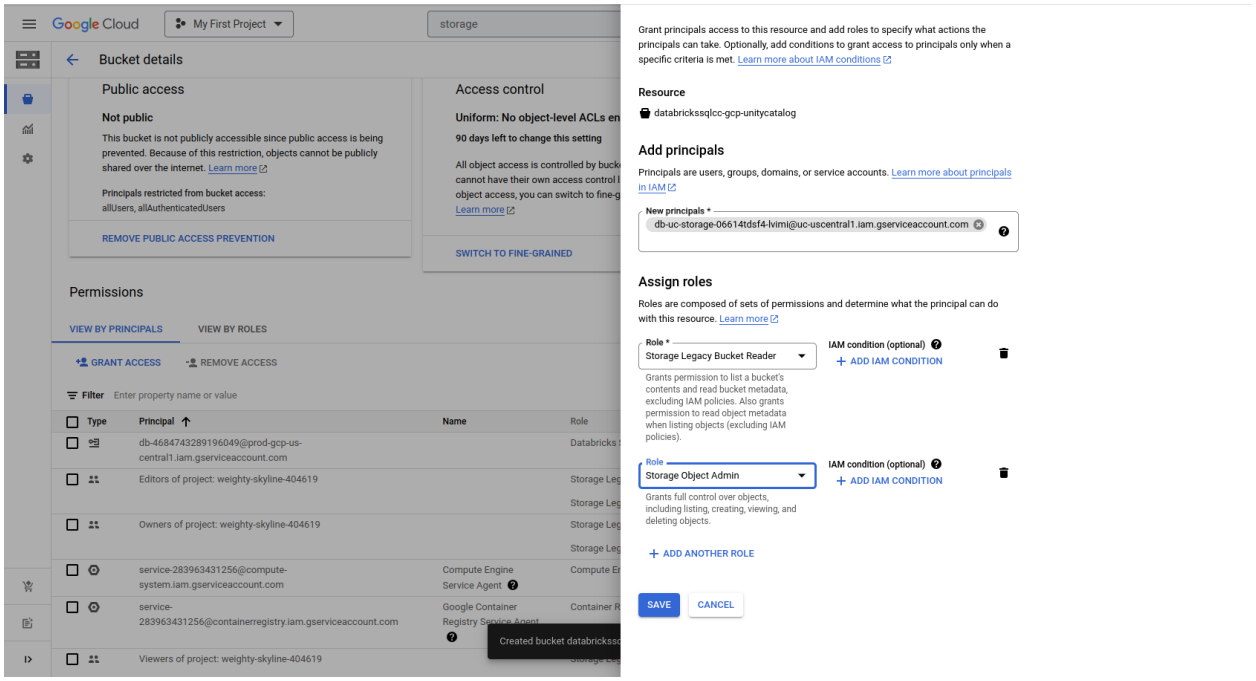
Na tela de criação do Metastore, certifique-se de colocar um nome para seu Unity Catalog, assim como a região em que está o bucket criado e o seu path/URL:



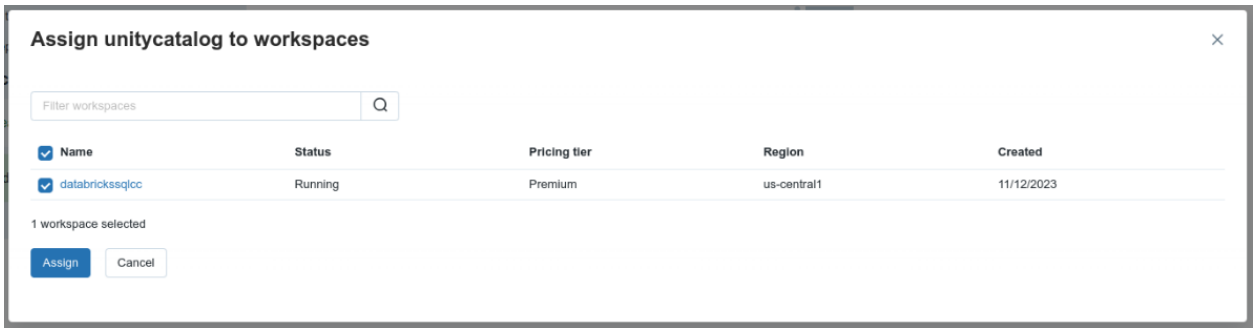
Ao clicar para criar o Metastore, uma janela é exibida com as instruções necessárias para configuração de autorização no permissionamento do bucket GCS configurado para que a Databricks tenha seus devidos controles (Storage Legacy Bucket Reader e Storage Object Admin) e então uma Service Account da GCP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



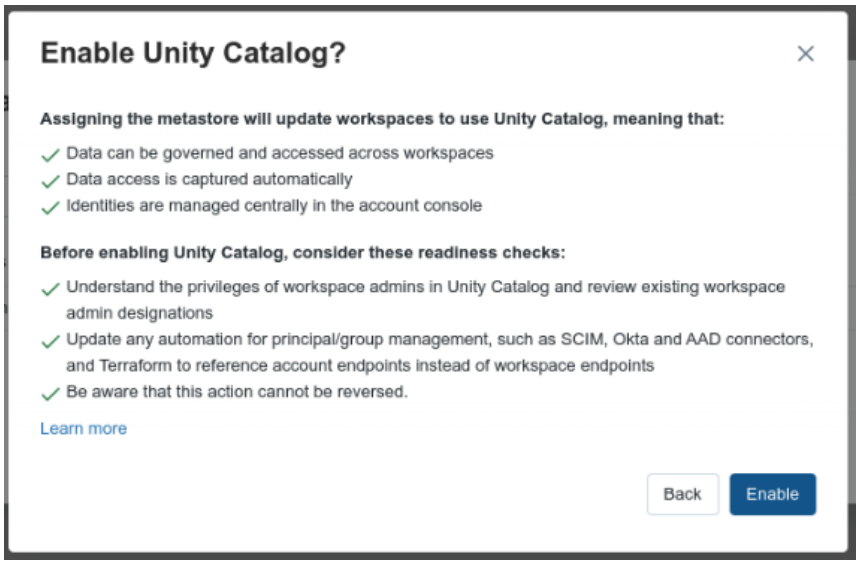
Vá então até o bucket criado para ser a base do Unity Catalog e, na aba "Permissions", clique em "Grant Access" fornecendo então as condições de Storage Legacy Bucket Reader e Storage Object Admin para a Service Account fornecida e então salve os novos acessos:



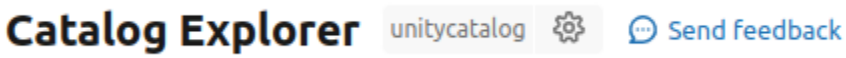
Retorne para a Databricks, já confirmando a concessão de acessos e verá então uma janela mostrando os Workspaces disponíveis que podem ser aplicados o Unity Catalog em devida região:



Confirme a habilitação do Unity Catalog:

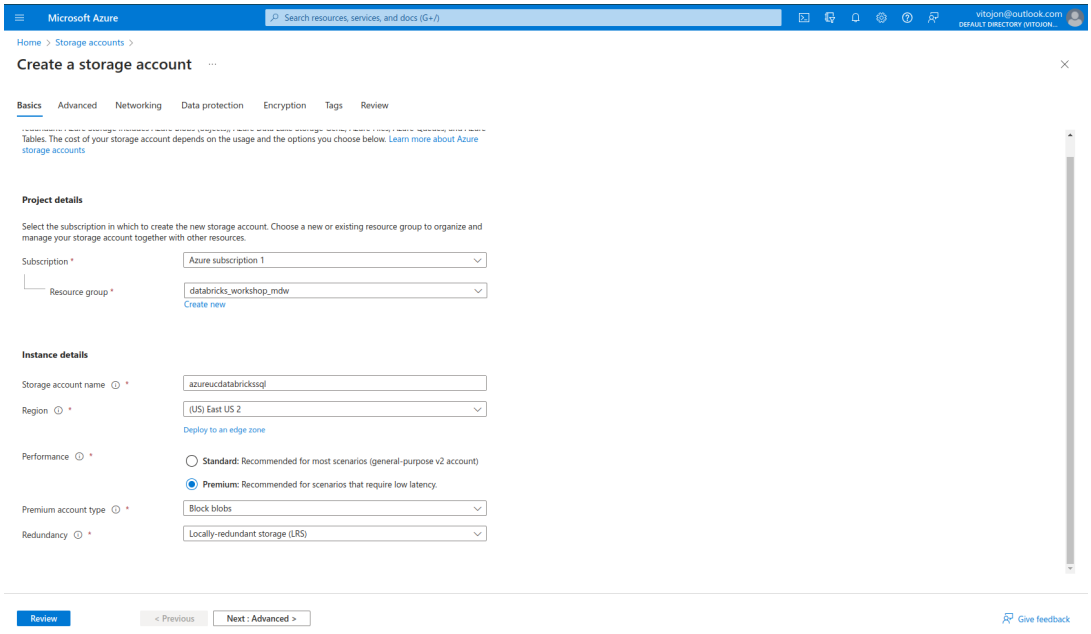


E por fim encontrará o Unity Catalog funcionando em seu Workspace:

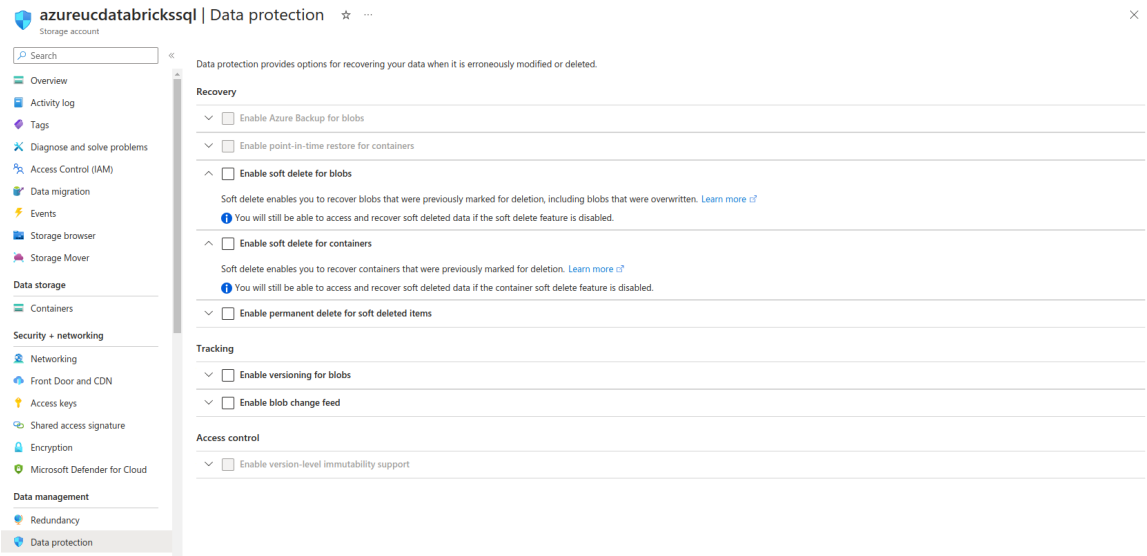


Criando Unity Catalog na Azure:

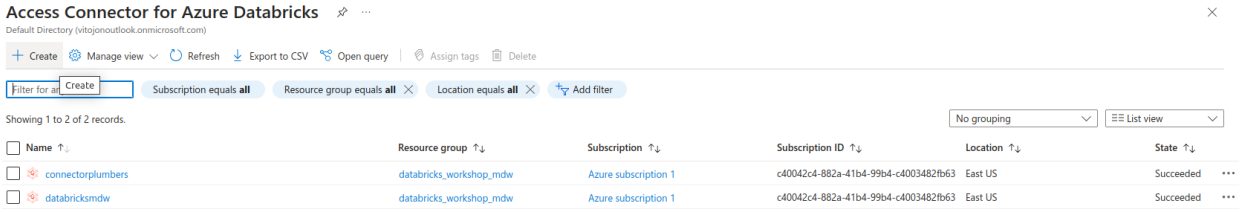
Assim como na GCP, crie um bucket na Azure para criação do Unity Catalog, porém se atentando aos detalhes de Resource Group, a região (mesma que do Workspace, sempre vale relembrar), em performance Premium e de tipo Block Blobs:



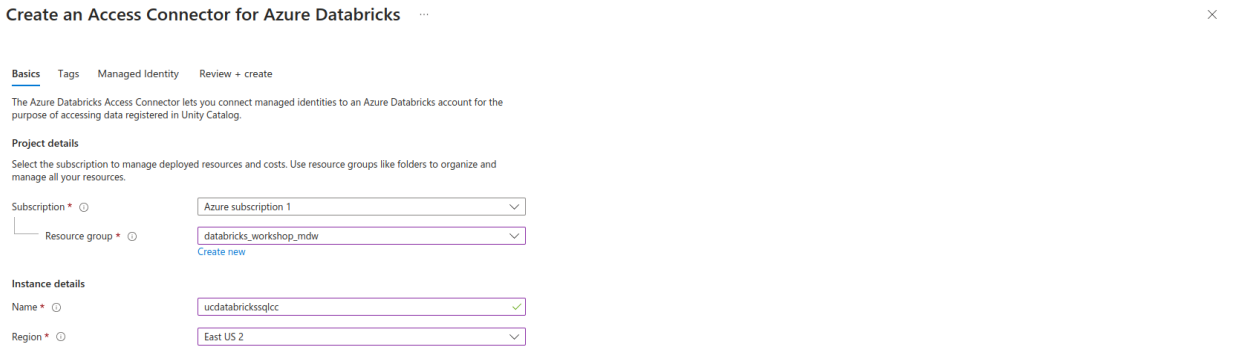
Após criação bem sucedida do seu Storage, vá até a configuração de management para Data Protection e desabilite os checks que permitem Soft Delete para Blobs e Containers (Enable ativo irá impedir a criação do Unity Catalog):



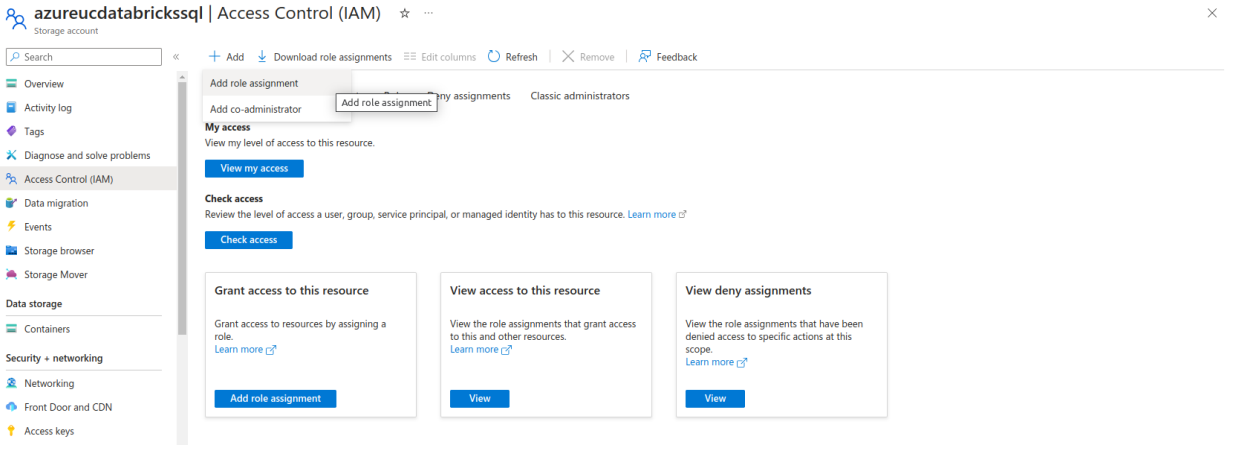
Procure na barra de pesquisas da Azure pelo “Access Connector for Azure Databricks” e clique na opção de criar para obter o seu conector, necessário para integrar permissionamento entre Storage e Workspace:

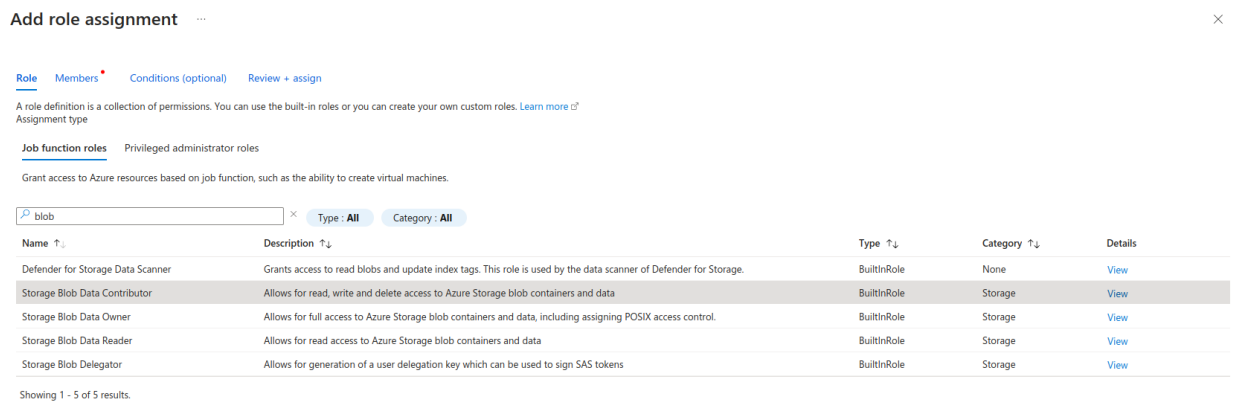


Aponte seu Resource Group e região indicada para criação desse conector, assim como seu nome para identificação:

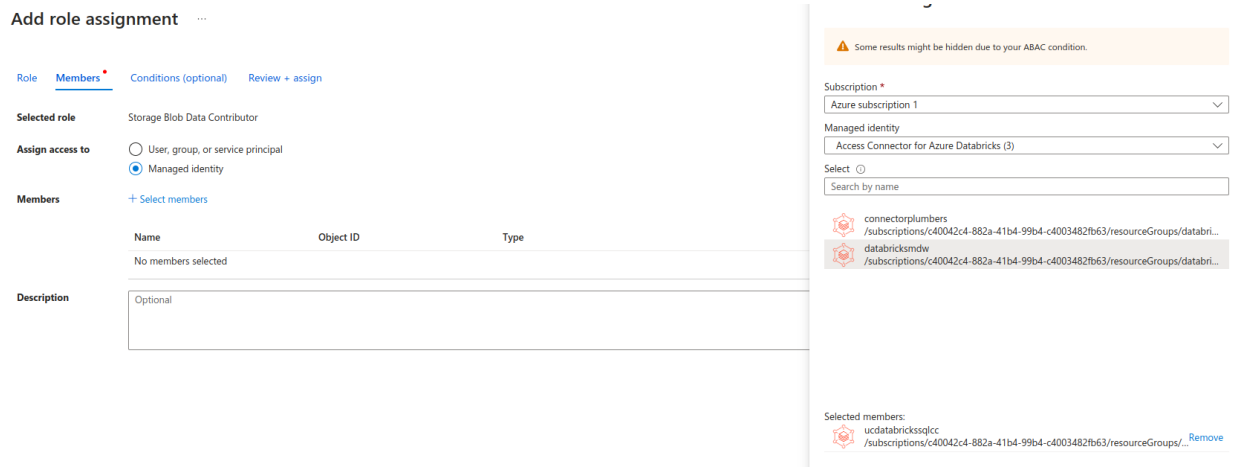


Uma vez com o conector criado, compareça até o IAM do seu Storage e adicione uma role, lá você irá incluir a opção de Storage Blob Data Contributor para o conector criado na aba de Roles:

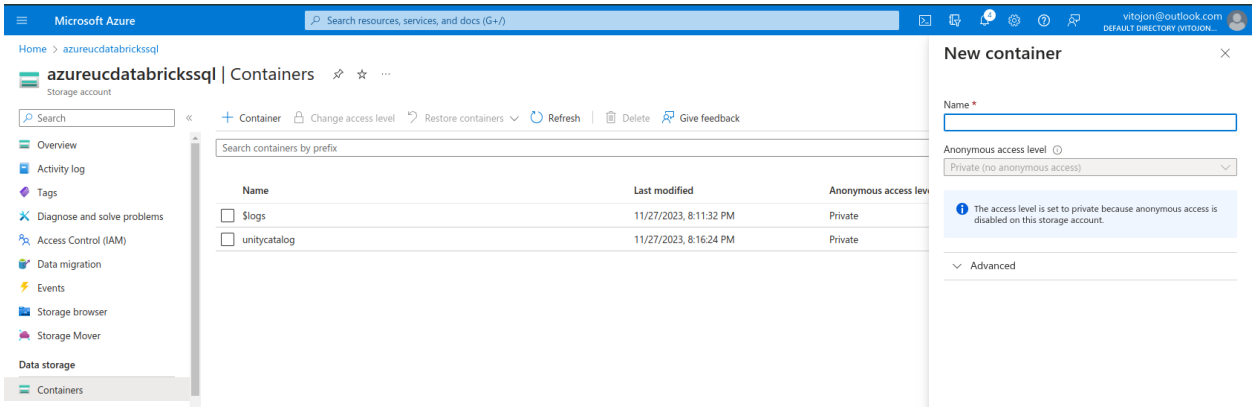




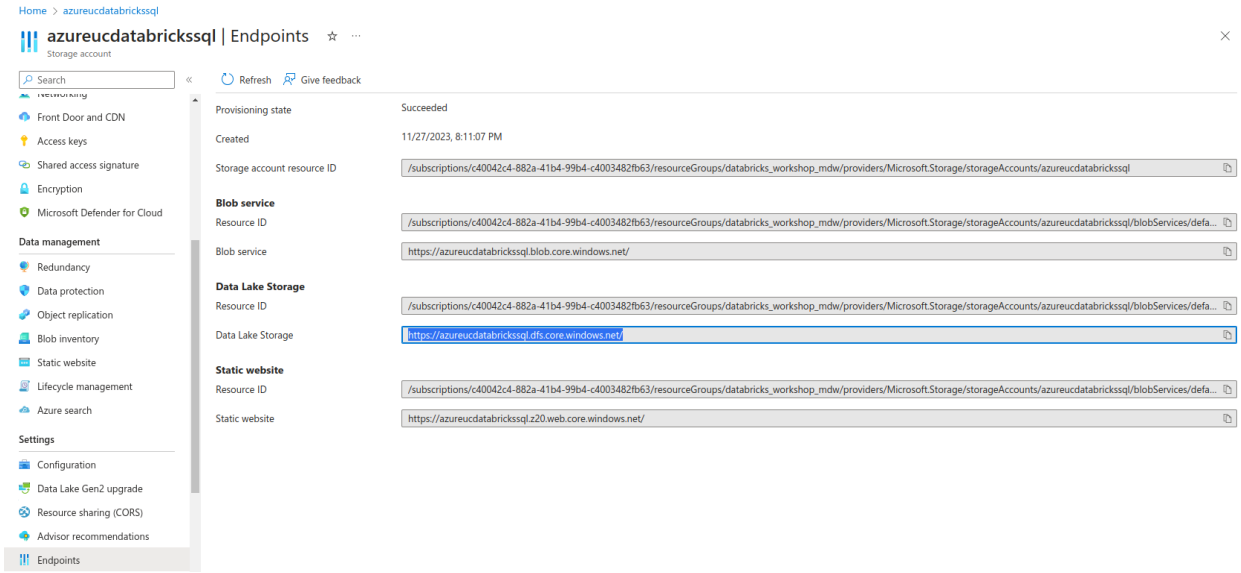
Já na aba de Members, adicione uma Managed Identity, onde poderá ser encontrado o seu conector para Databricks:



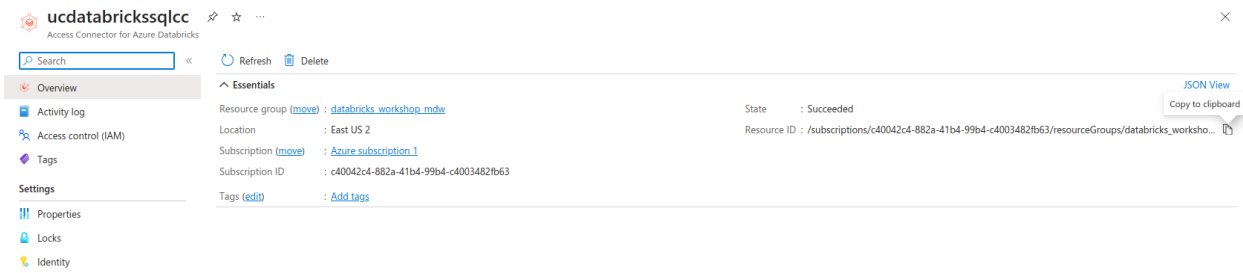
Após configuração de IAM, crie um container dando-lhe o nome que for conveniente em seu Storage:



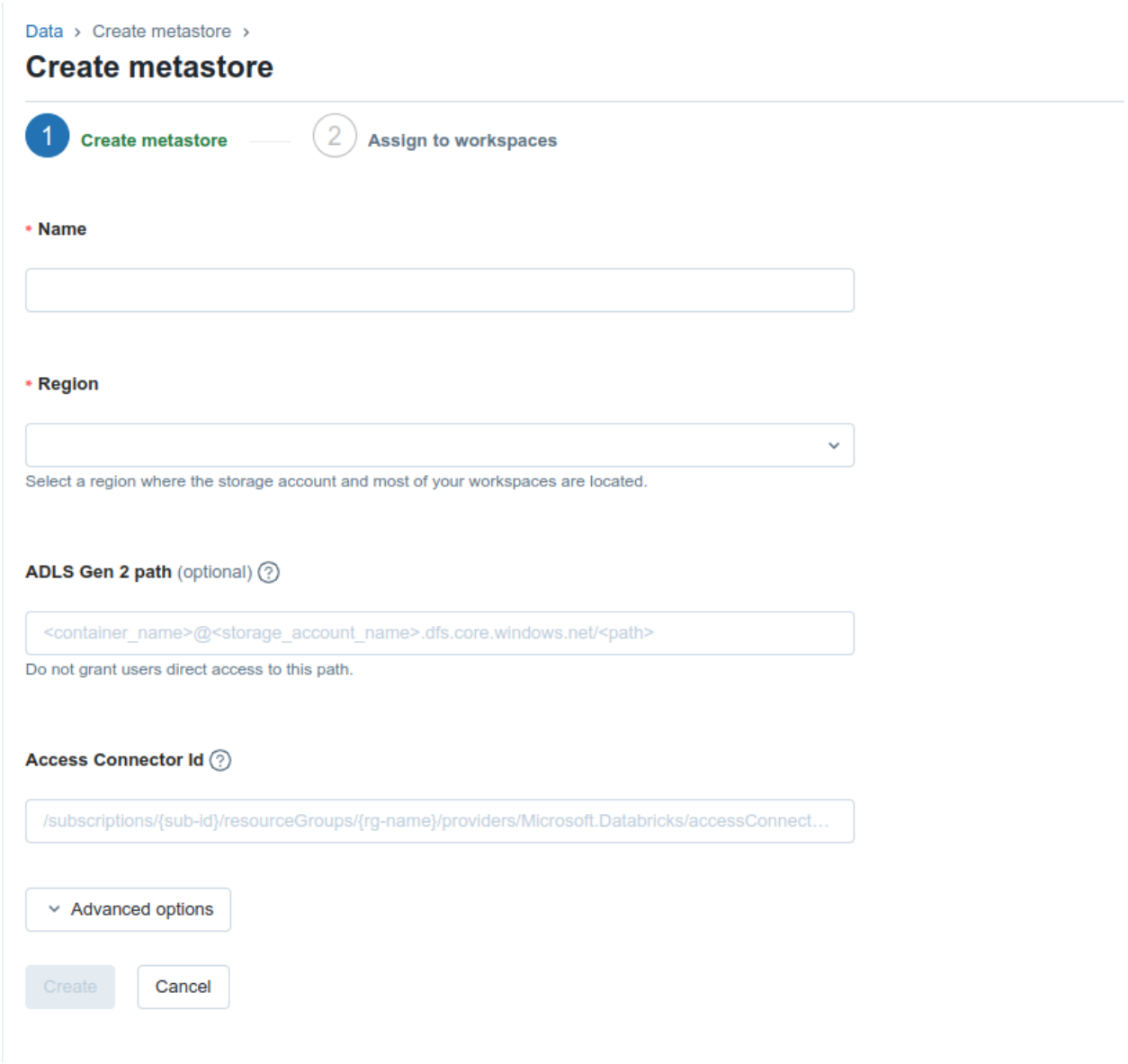
Por fim, se preferir, pode ir até o campo de Endpoints no Storage para copiar seu endereço de Storage (será pedido na Databricks):



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



Compareça na página de Metastore da Account Console em Data:



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.