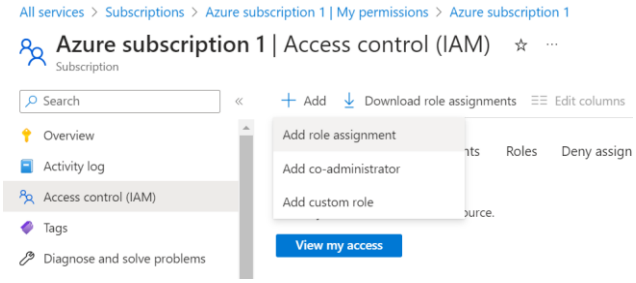
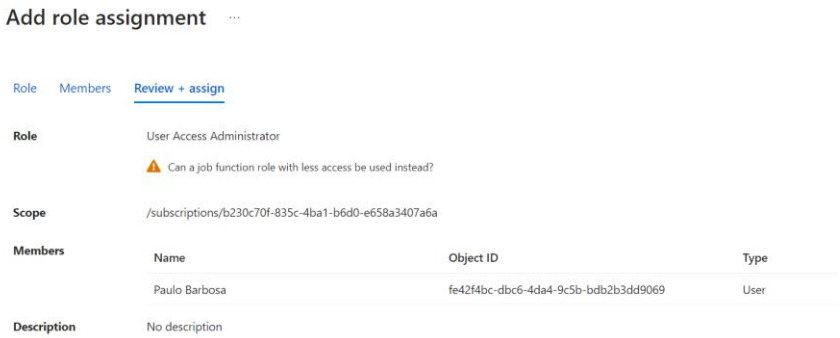
Step by step to run the application

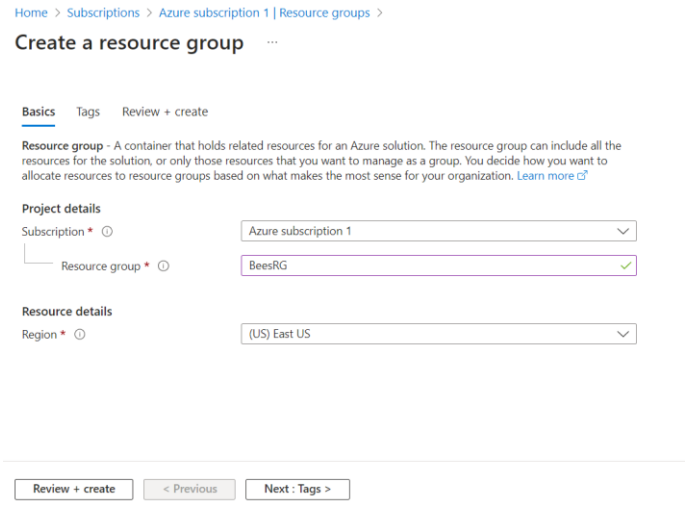
1. Open Azure Portal
2. Go to Access Control (IAM) in your Subscription options
3. Click on Add Role Assignment



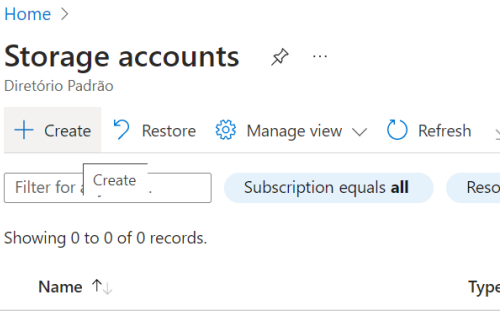
4. Add “User Access Administrator”, “Storage Blob Data Contributor” and “Key Vault Administrator” Roles



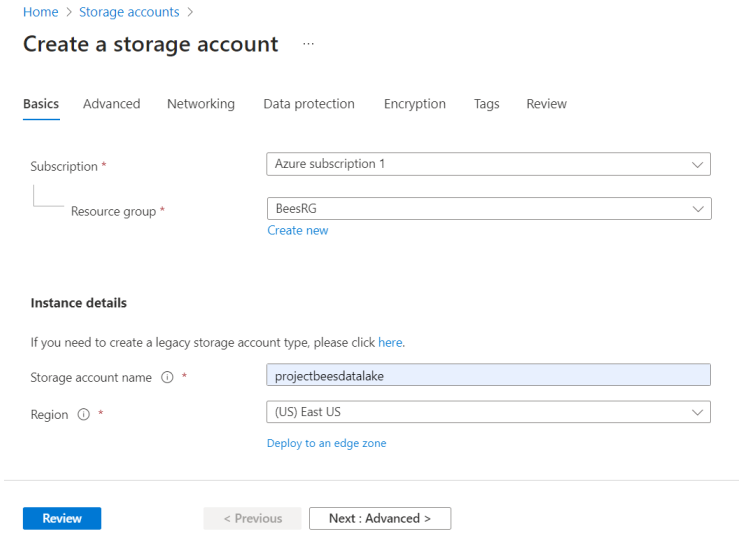
5. Go to “Resource Groups” and create a resource group



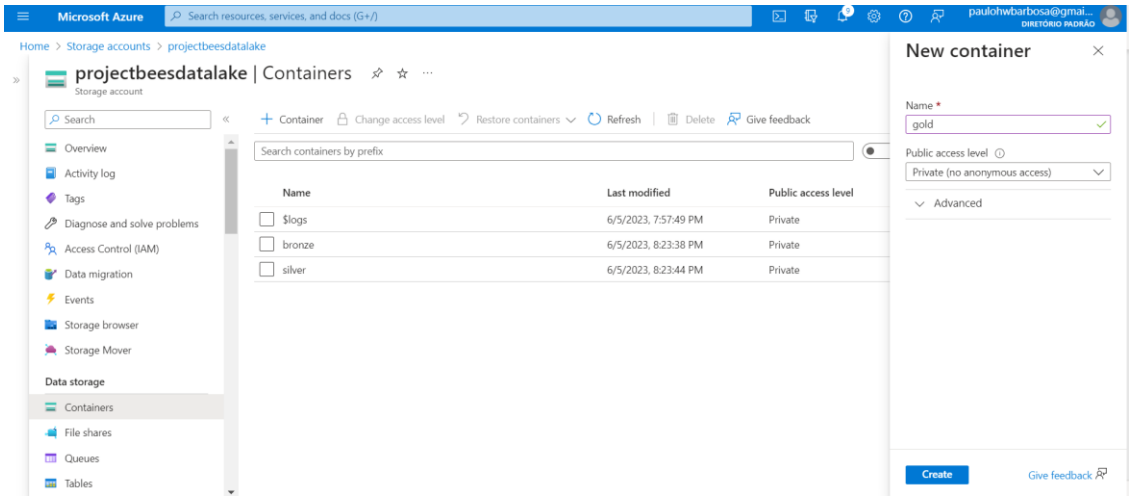
6. Go to Storage accounts and click on “create”



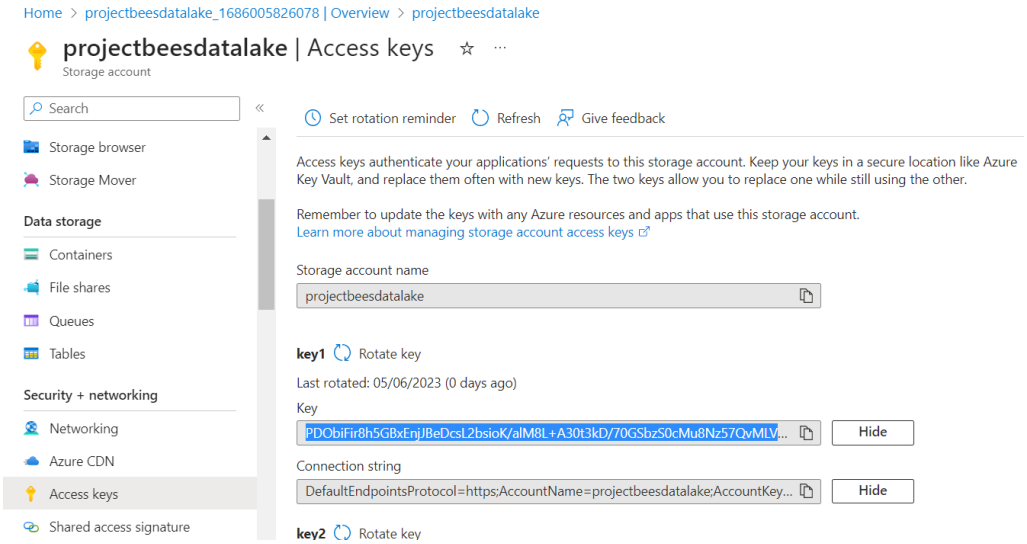
7. Create the Storage Account as the image below, the Storage account name must be unique:



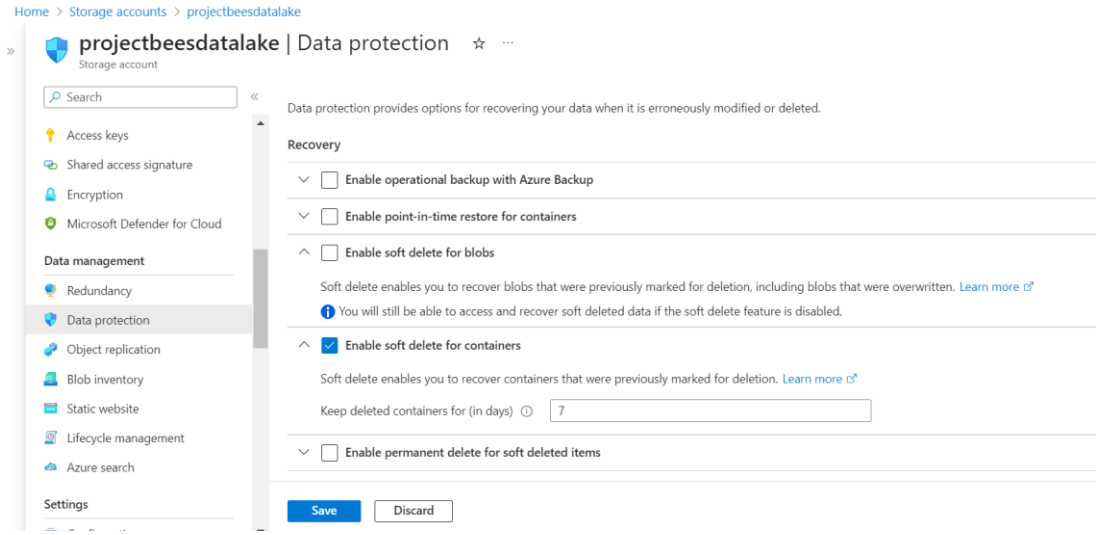
8. Access this Storage Account, go to “Containers” and create a bronze, a silver and a gold container



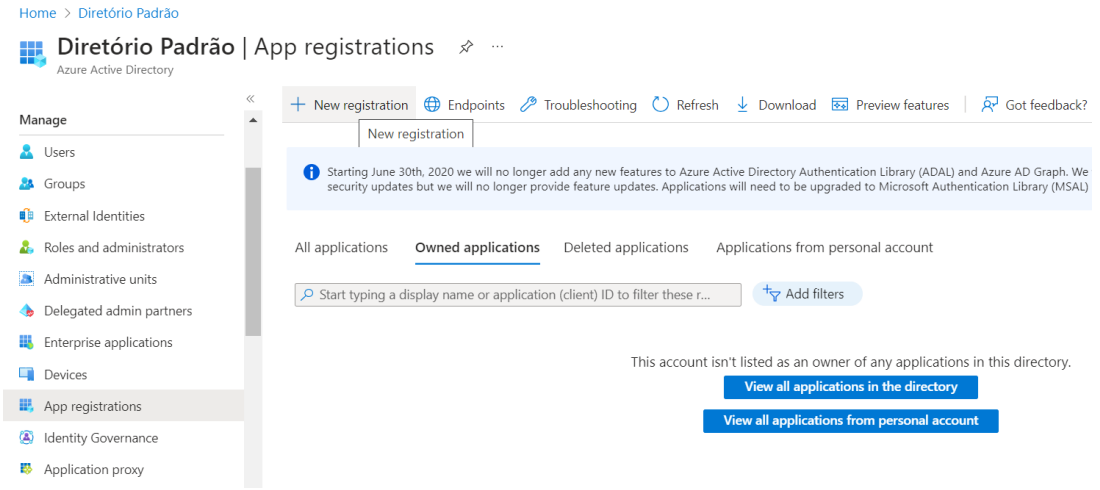
9. Go to “Access Keys” and copy the Key value to a notepad.



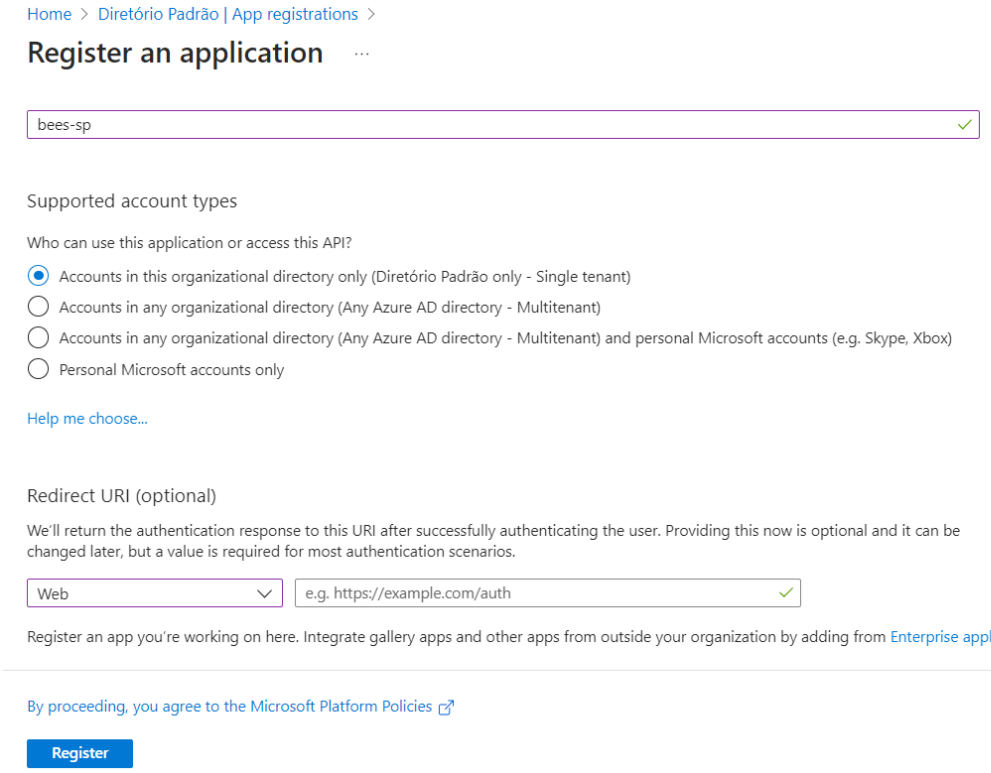
10. Go to Data protection and disable Soft Delete for blobs



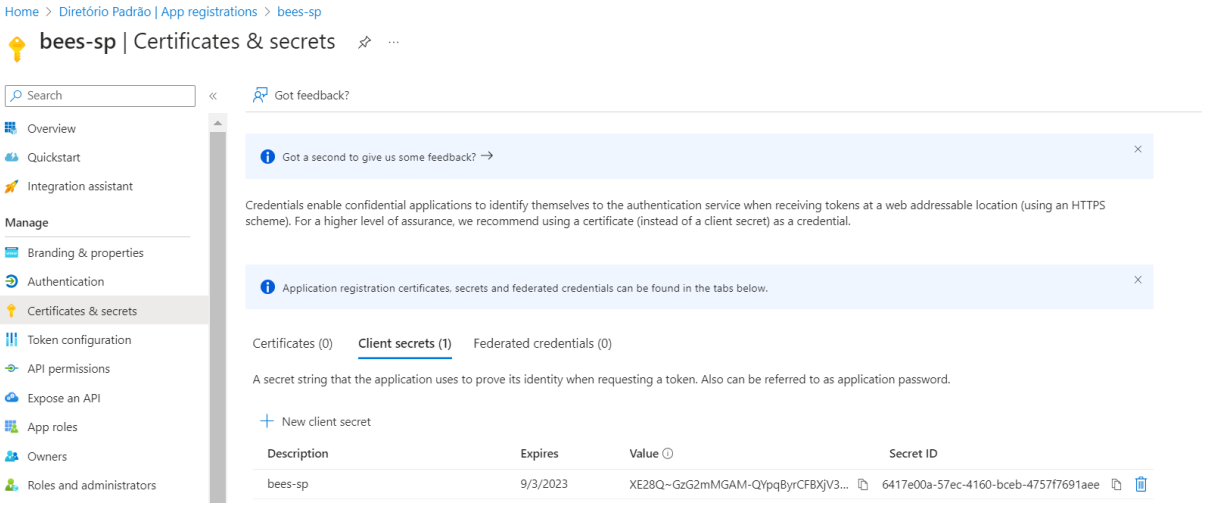
11. Go to Azure Active Directory, App registrations and “New registration” to create a service principal



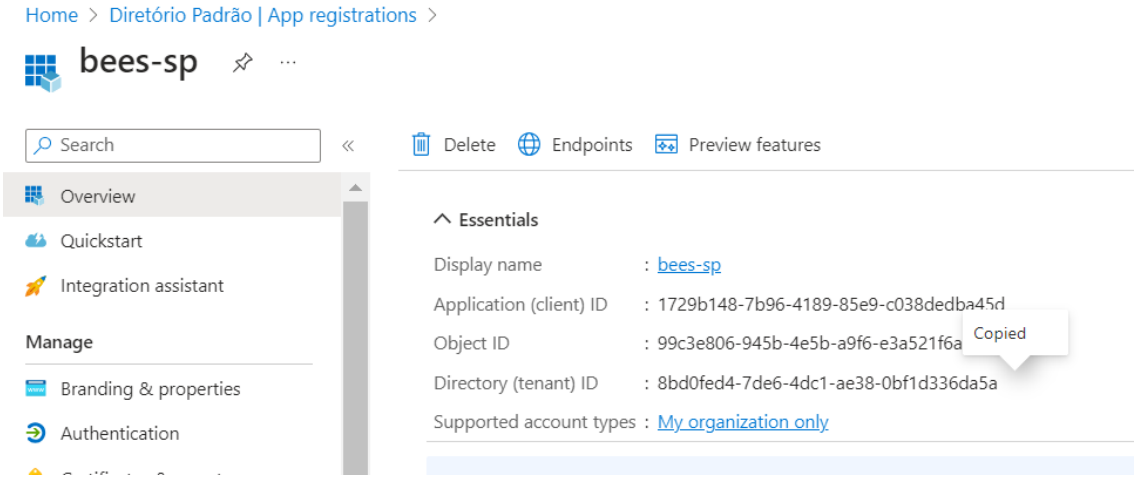
12. Fill the gaps as the image below:



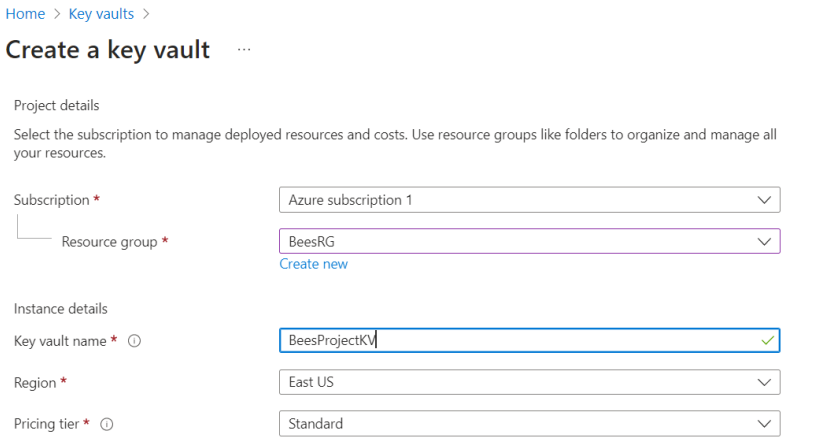
13. Go to “Certificates & secrets” and create a client secret, and copy those values to a notepad



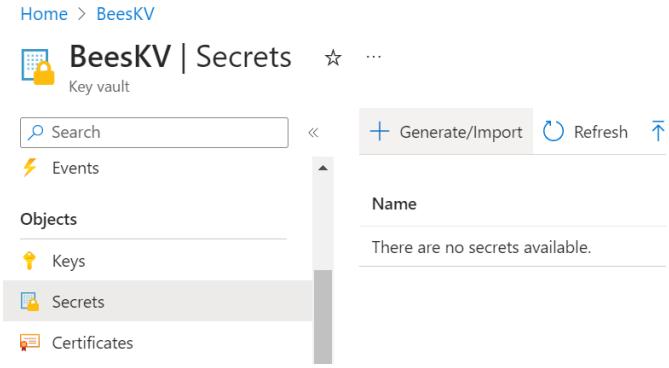
14. Go to Overview and copy the Application (client) ID and the Directory (tenant) ID to a notepad



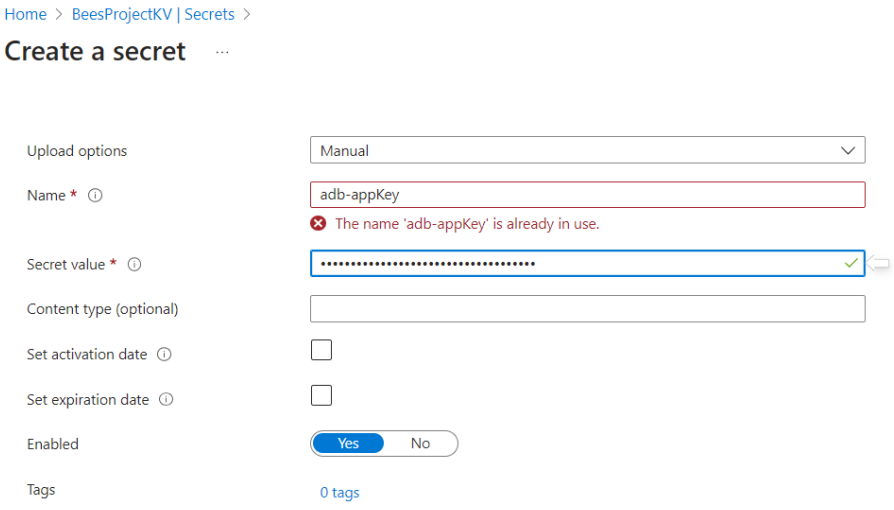
15. Go to Key Vaults and create a key vault



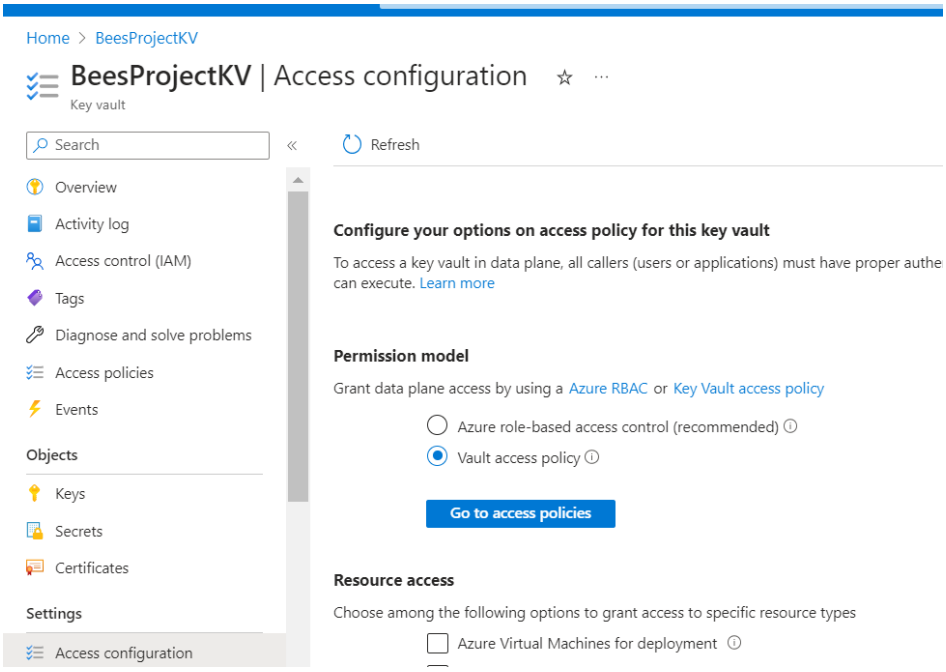
16. Access the created Key Vault and click on “Generate/Import”, in the Secrets section



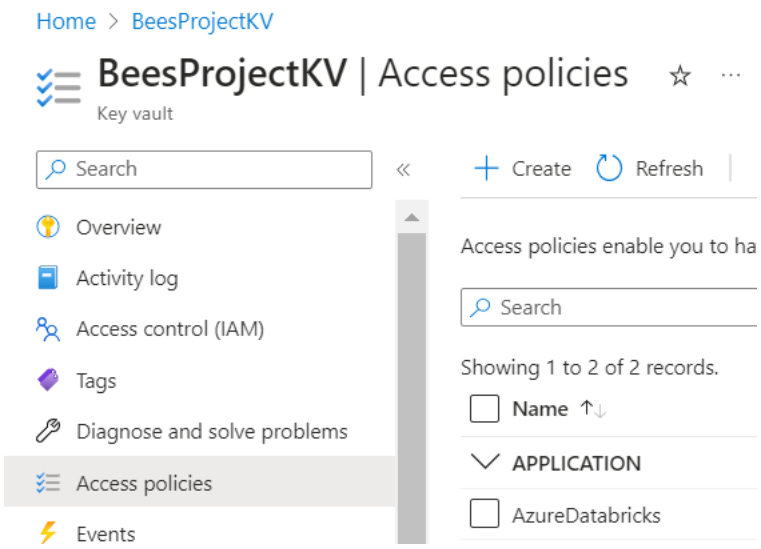
17. We are going to create some secrets:
    Secret adb-tenantId using Directory (tenant) Id value copied at step 14.
    Secret adb-appId using Application (client) Id value copied at step 14.
    Secret adb-appSecret using App’s Secret value copied at step 13.



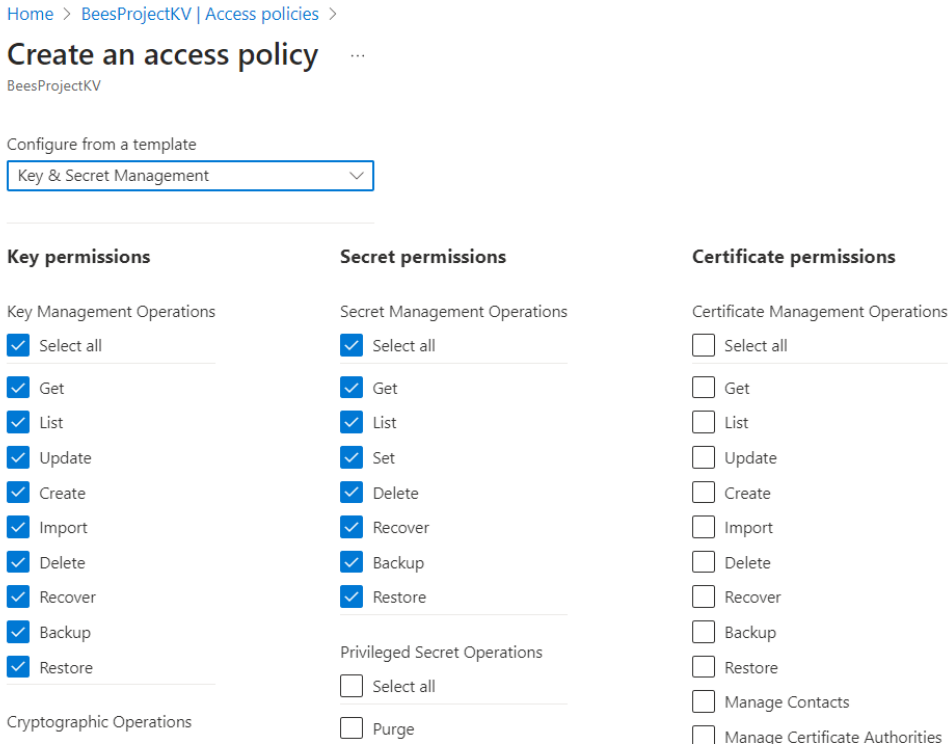
18. Go to Access configuration and select “Vault access policy”

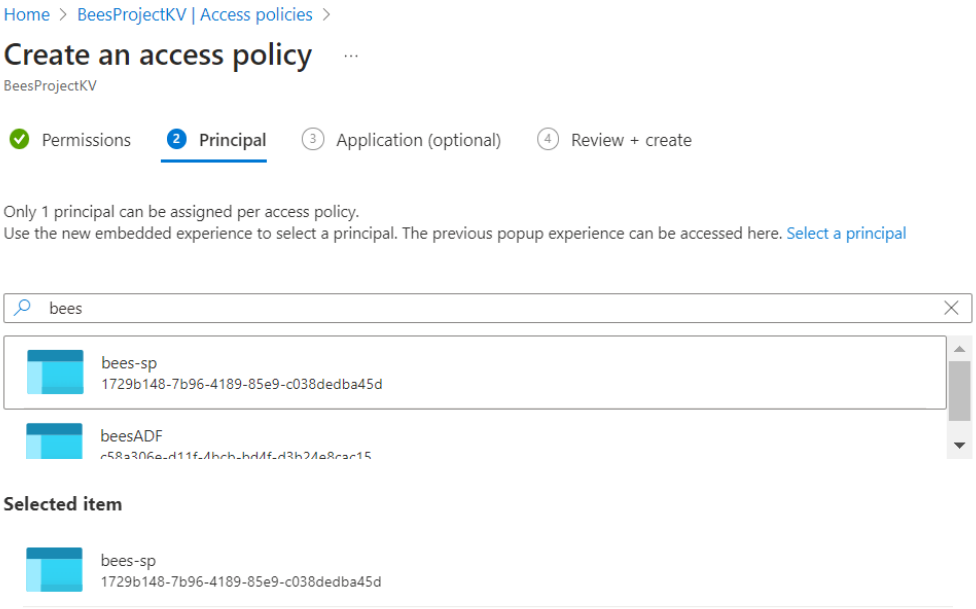


19. Go to Access policies and click on create

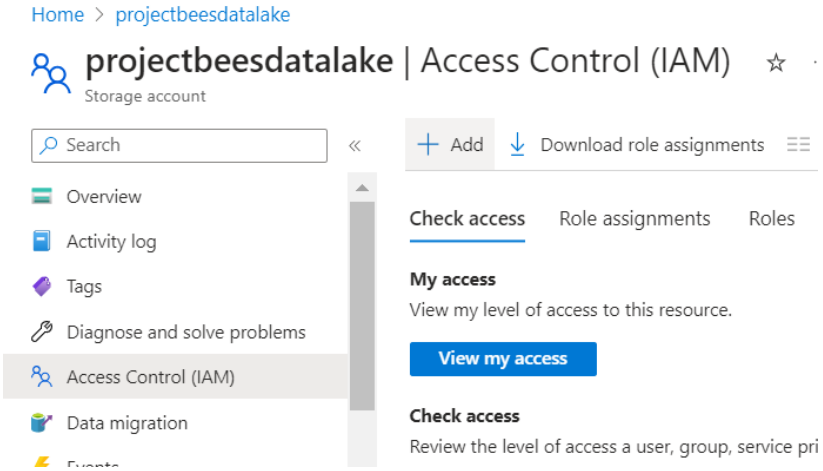


20. Select Key & Secret management to the service principal

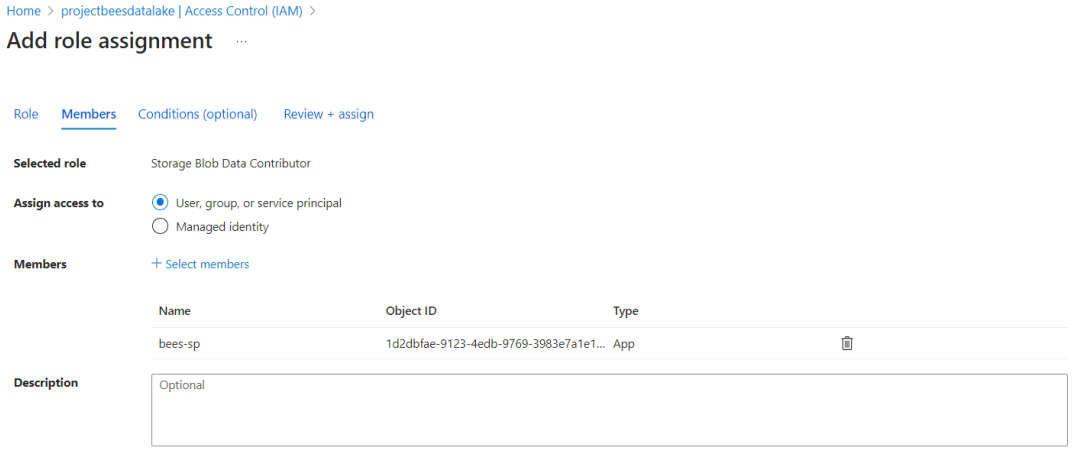




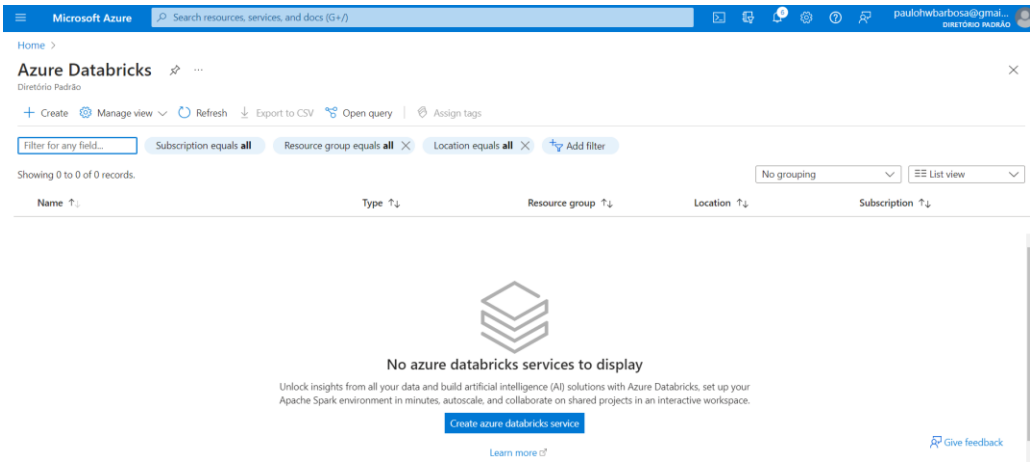
21. Go to the Storage account and go to Access Control (IAM) and click on “Add”



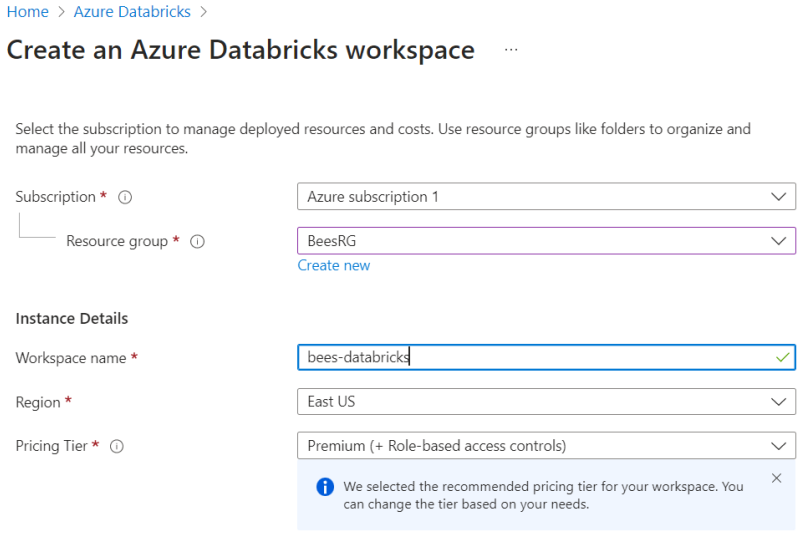
22. Add Storage Blob Data Contributor and Key Vault Administrator roles to the service principal



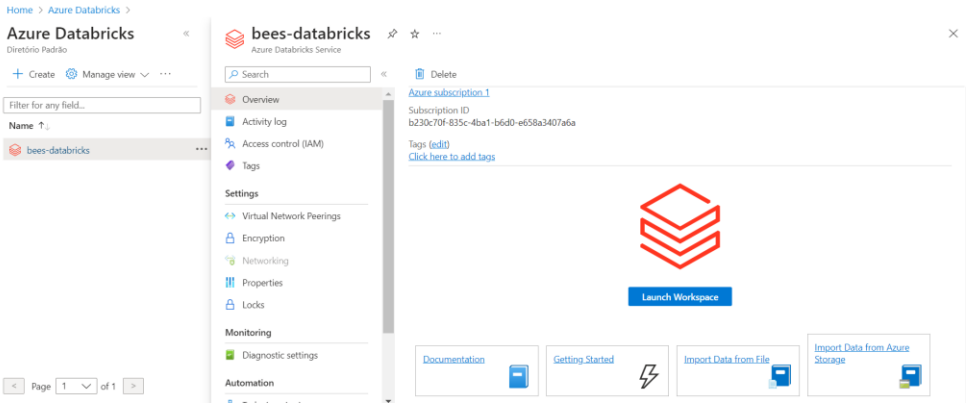
23. Go to Azure Databricks and click on create Azure Databricks service



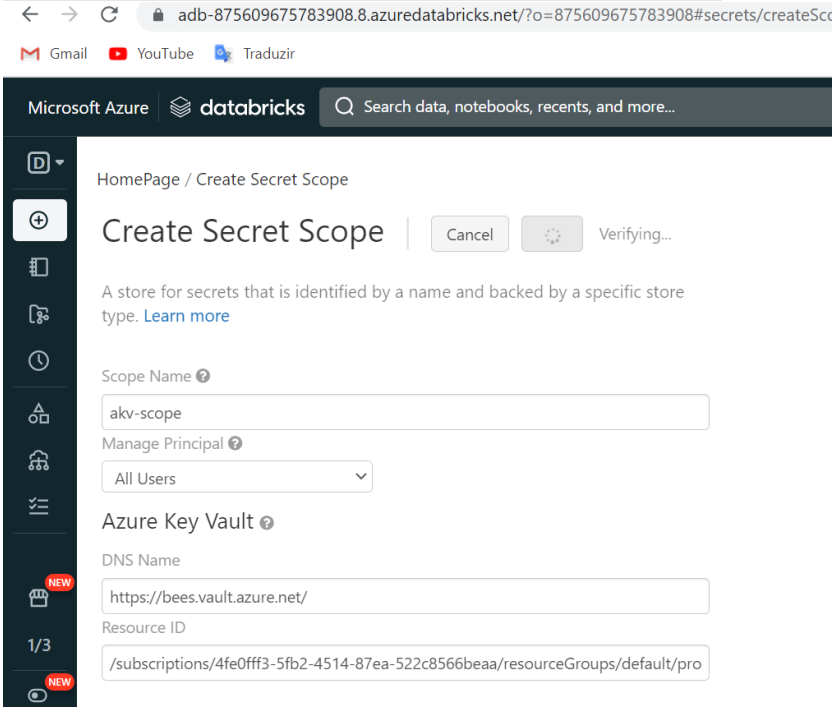
24. Create the Databricks Workspace as the image below:



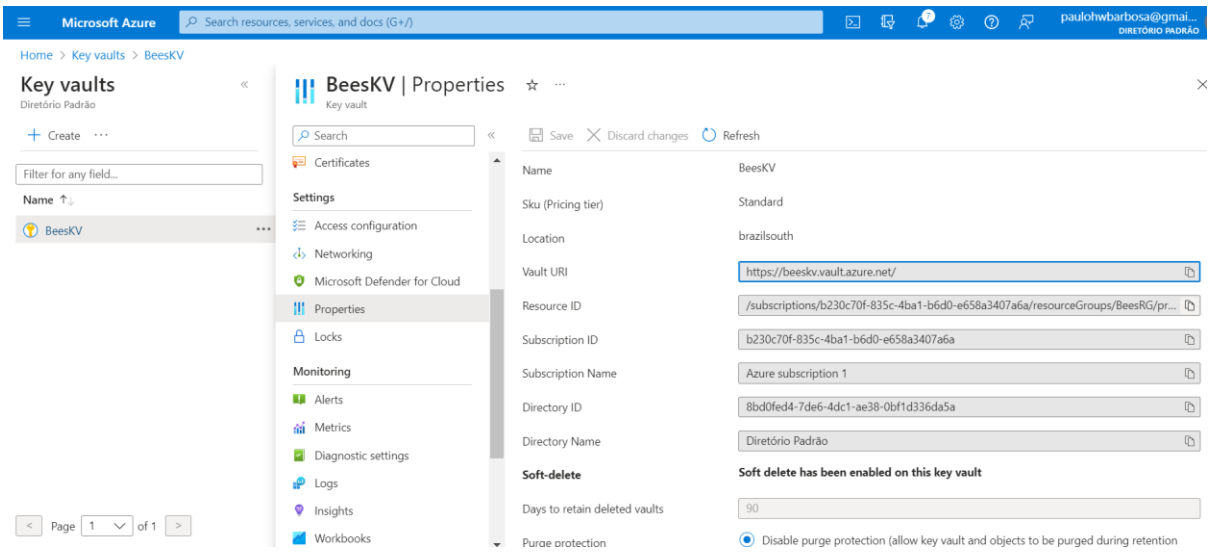
25. Go to Azure Databricks and click on “Launch Workspace”



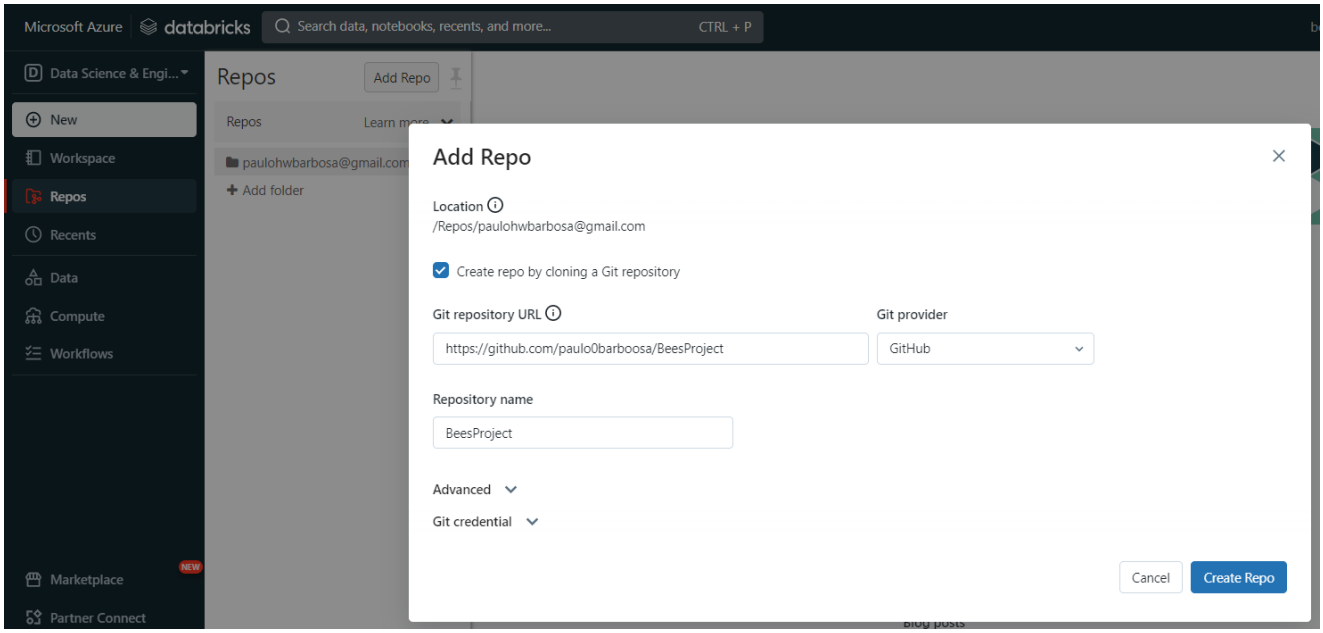
26. Go to the following URL and fill de gaps as shown on the image bellow:
`https://<databricks-instance>#secrets/createScope`



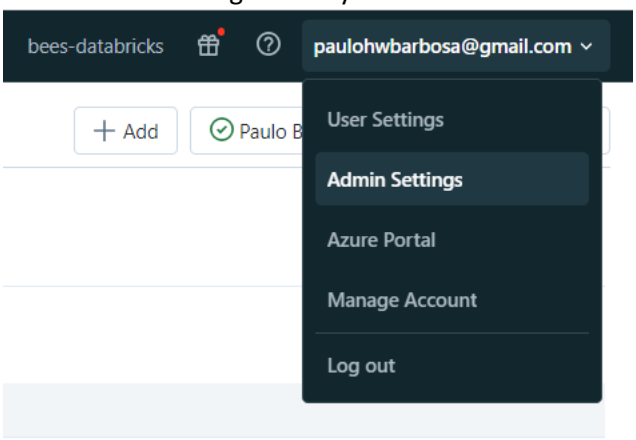
DNS Name field fill with Vault URI value and Resource ID with Resource ID Value, found in the Properties section of the created Key Vault



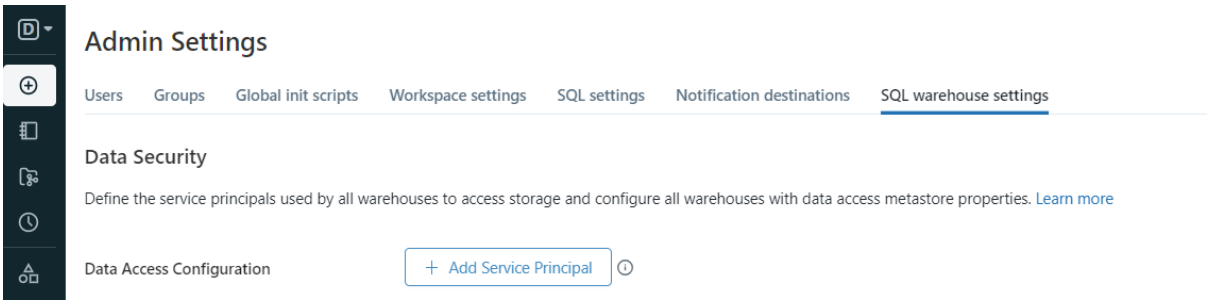
27. Go to “Repos” Section of Databricks, select “Add Repo” and fill the fields as the image below:



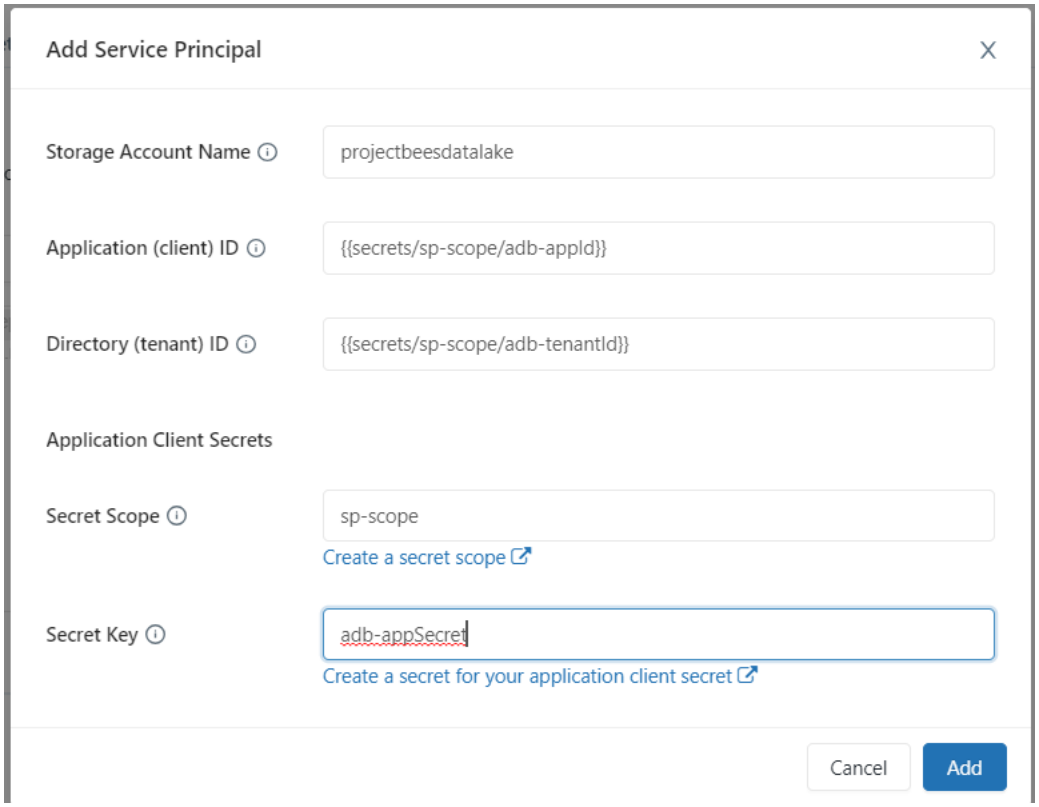
28. Go to Admin Settings under your name account



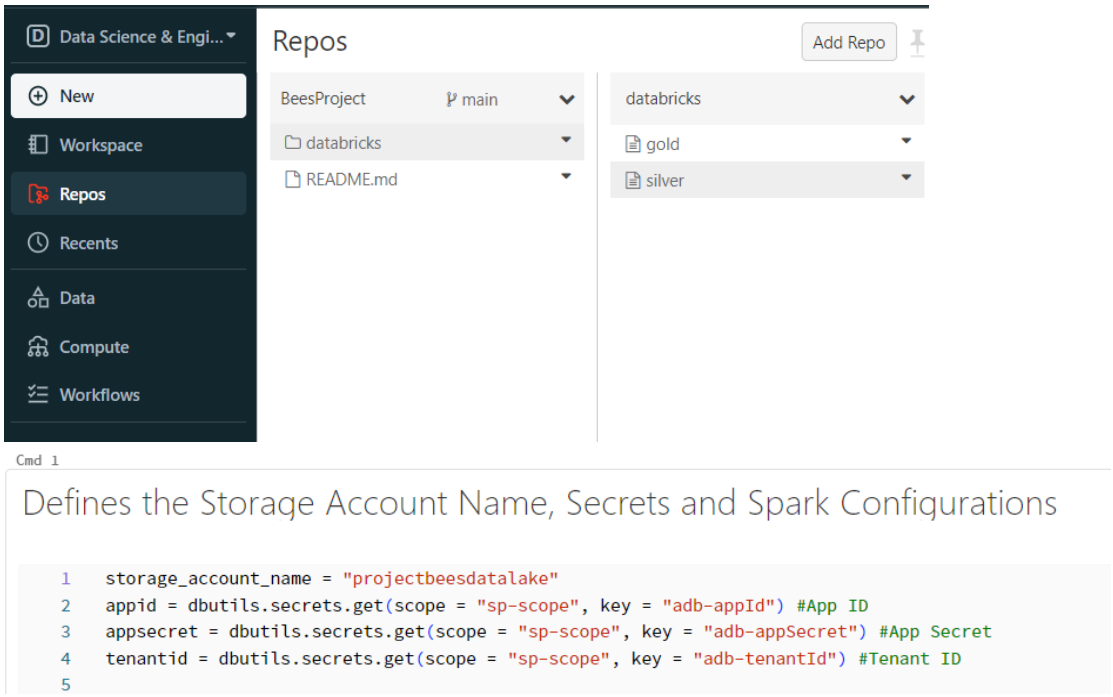
29. Go to SQL warehouse settings and add a Service Principal



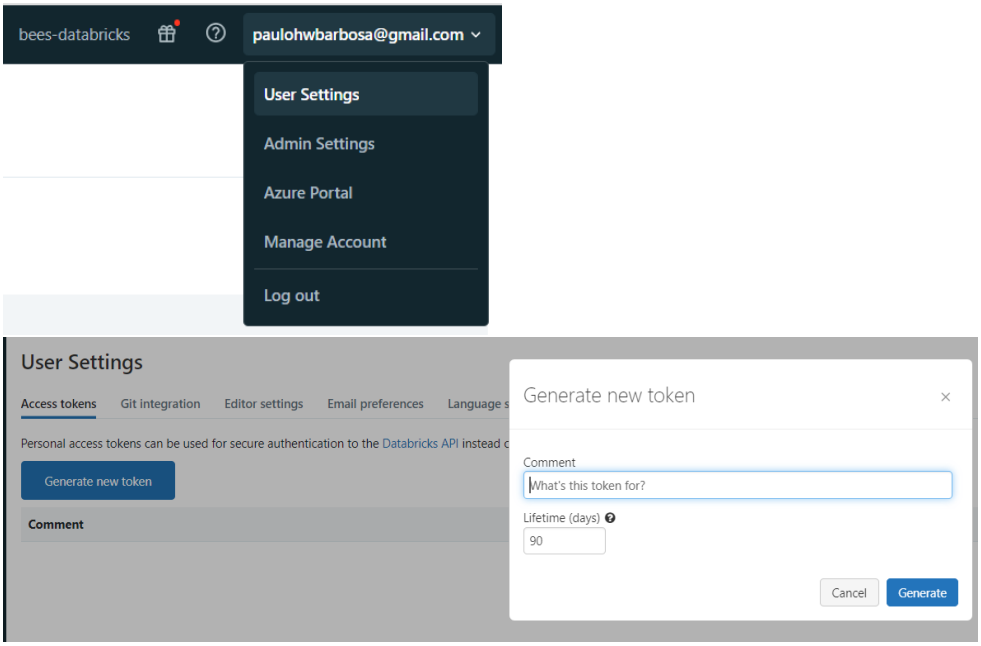
30. Fill the gaps as the image below (change Storage Account Name to your Storage Account):



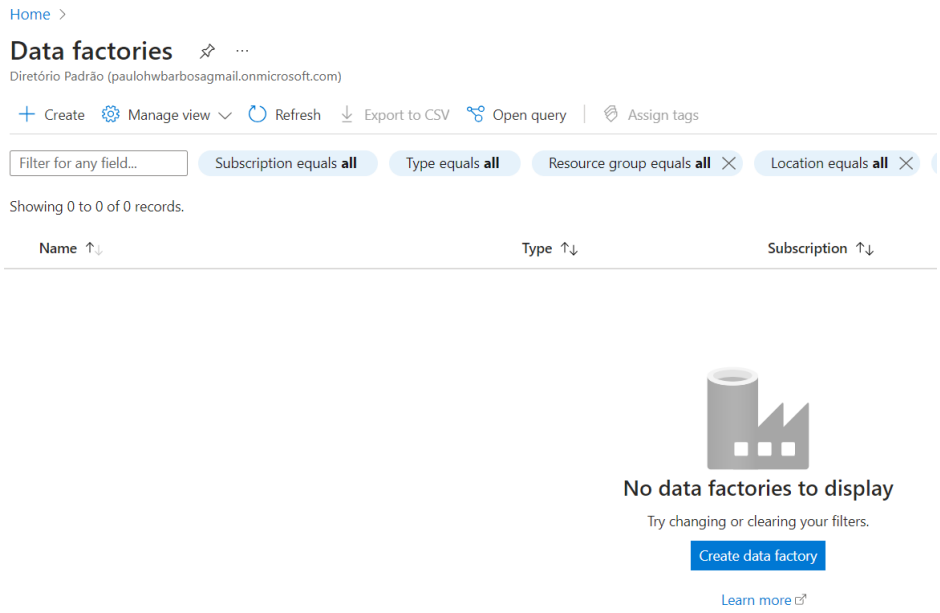
31. Inside both silver and gold notebooks, change the first variable to your Storage Account Name.



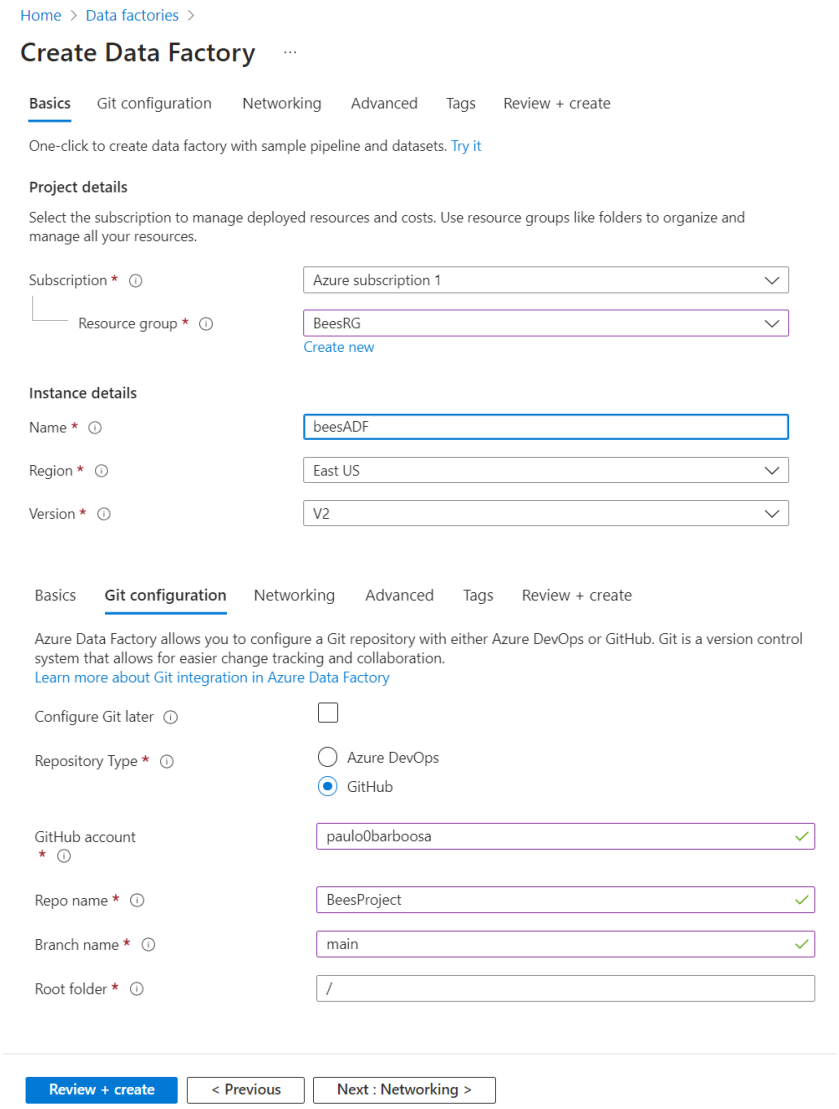
32. Go to “User Settings” and generate an access token, copy it and save it on a notepad



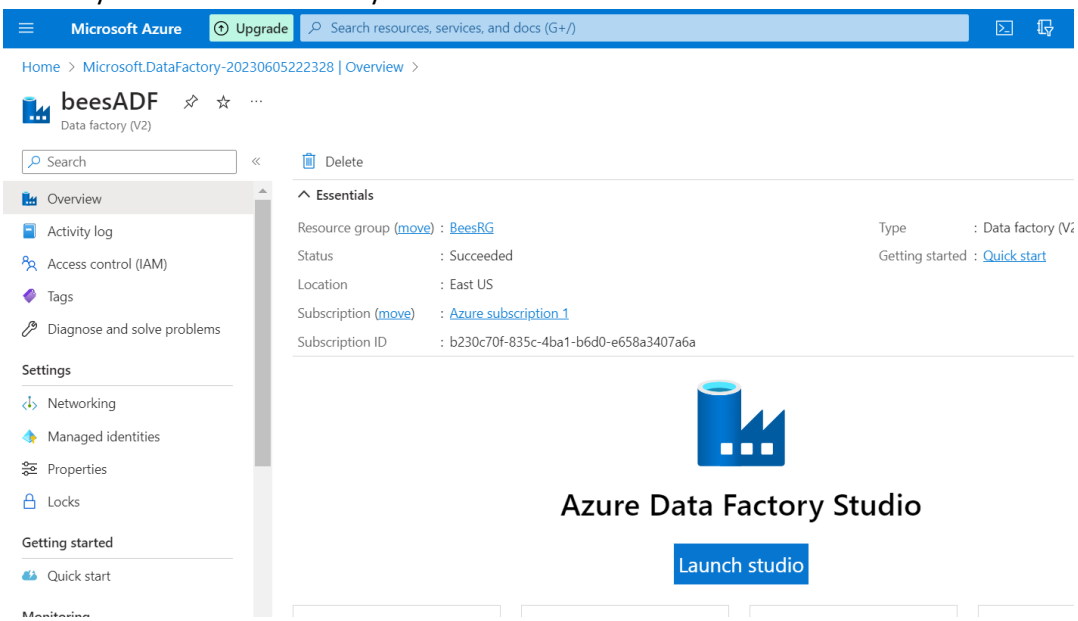
33. Go to “Data Factories” in the Azure Portal and “Create data factory”



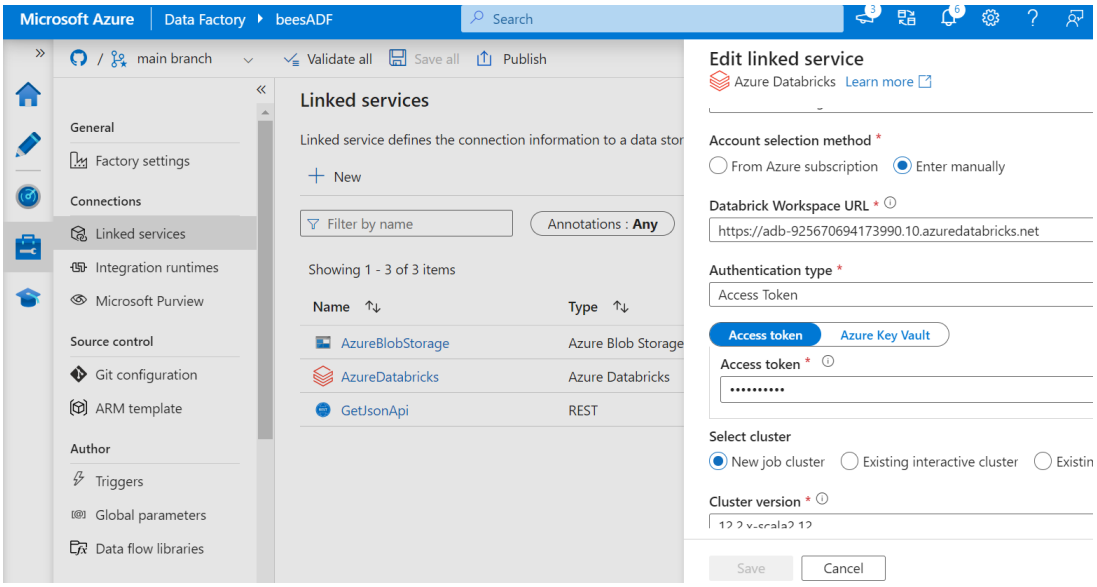
34. Create as the images below:



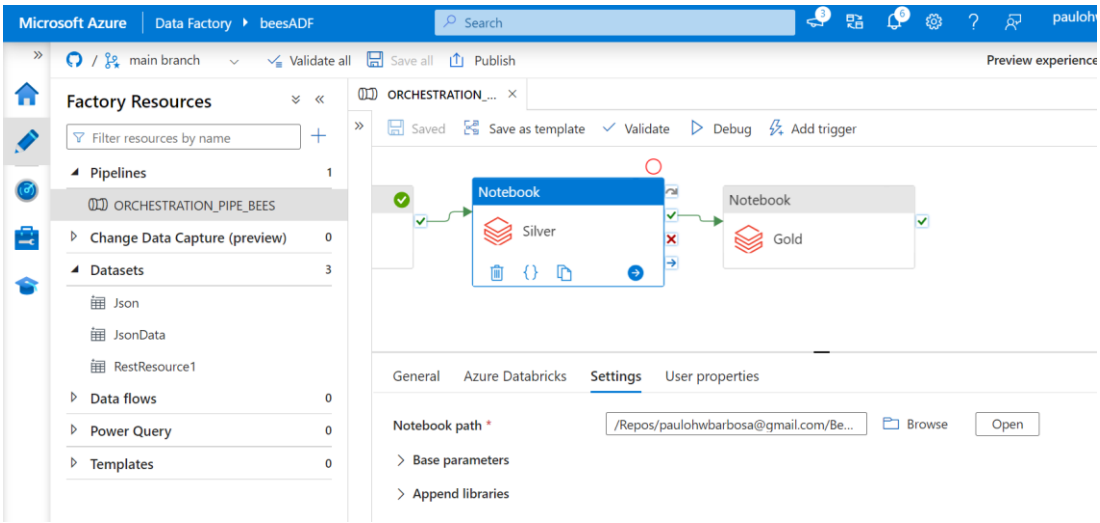
35. Launch your Azure Data Factory



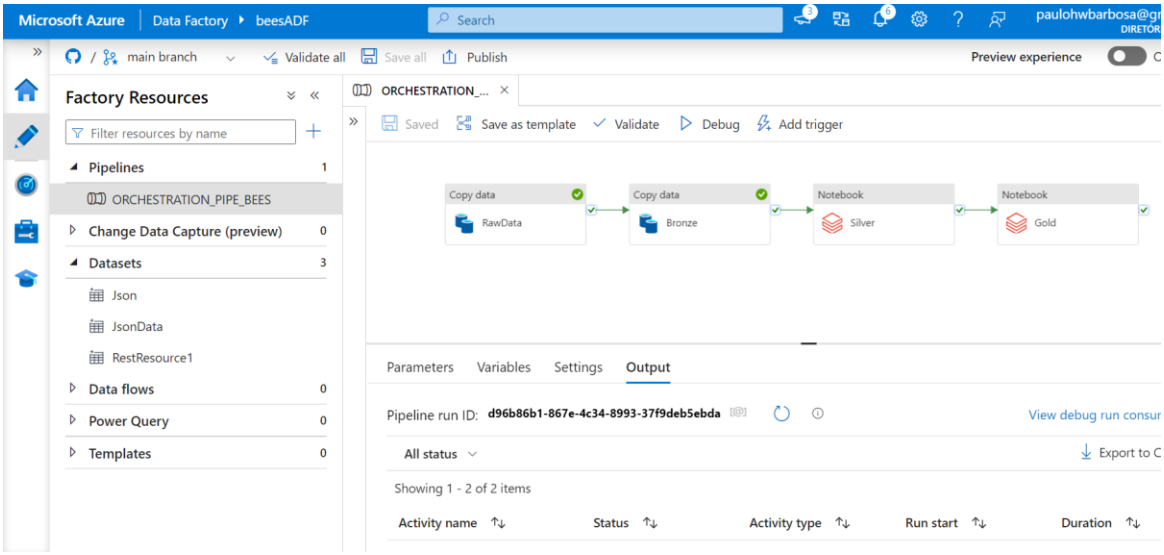
36. In the Manage options, select Linked Services, AzureDatabricks and change the Access Token value to the value copied at the step 32



37. In the Author options, select Pipelines, ORCHESTRATION_PIPE_BEES and change the notebook path at the settings of Silver and Gold Databricks Jobs to your email account
Silver Notebook: /Repos/<account>/BeesProject/databricks/silver
Gold Notebook: /Repos/<account>/BeesProject/databricks/gold



38. Now you can click on “Debug” to run the pipeline


Microsoft Azure
Data Factory
beesADF
Search
paulohwbarbosa@gr
DIRETOR
main branch
Validate all
Save all
Publish
Preview experience
Factory Resources
Filter resources by name
Pipelines 1
ORCHESTRATION_PIPE_BEES
Change Data Capture (preview) 0
Datasets 3
Json
JsonData
RestResource1
Data flows 0
Power Query 0
Templates 0
ORCHESTRATION_...
Saved
Save as template
Validate
Debug
Add trigger
Copy data
RawData
Copy data
Bronze
Notebook
Silver
Notebook
Gold
Parameters
Variables
Settings
Output
Pipeline run ID: d96b86b1-867e-4c34-8993-37f9deb5ebda
View debug run consur
All status
Export to C
Showing 1 - 2 of 2 items
Activity name
Status
Activity type
Run start
Duration