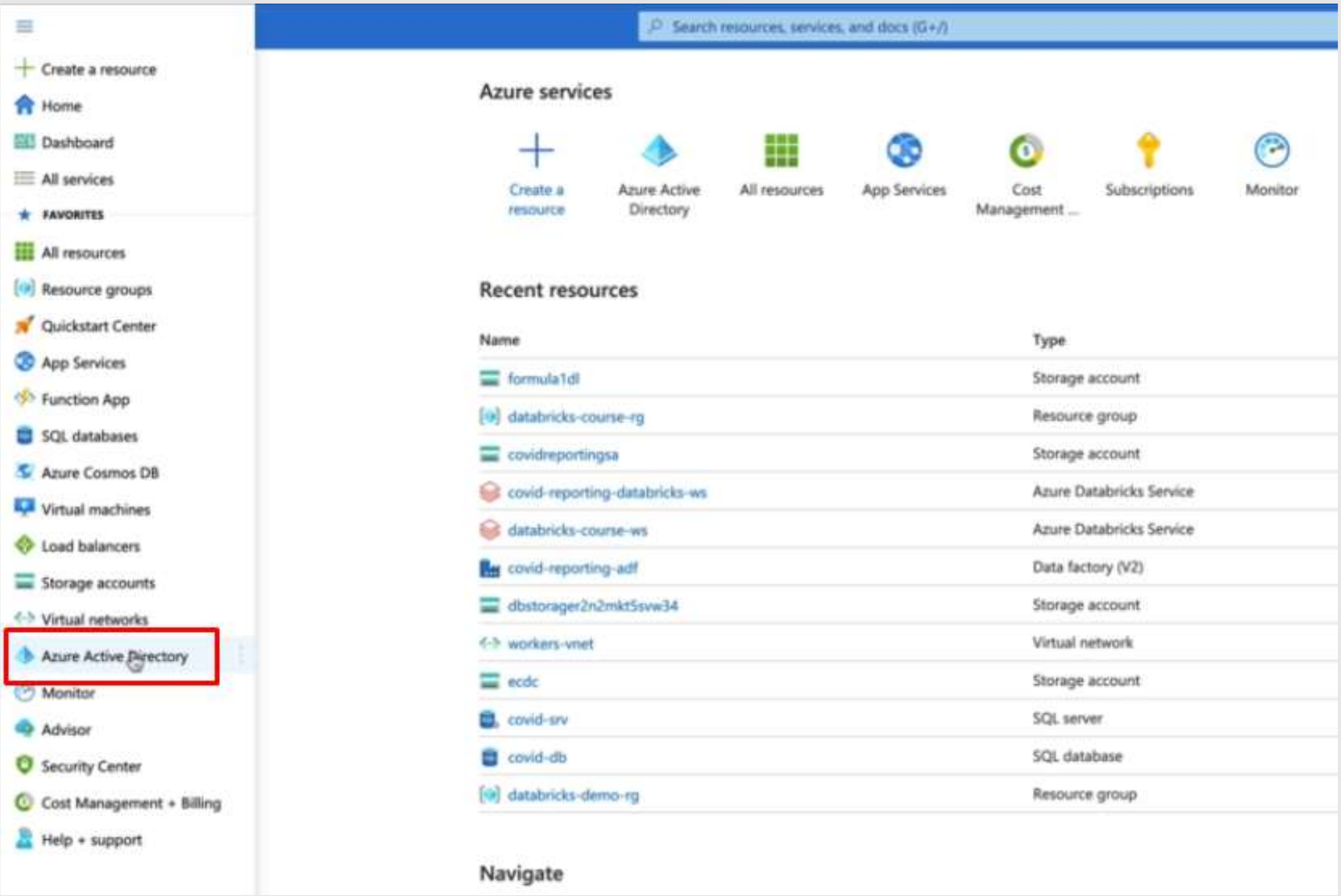
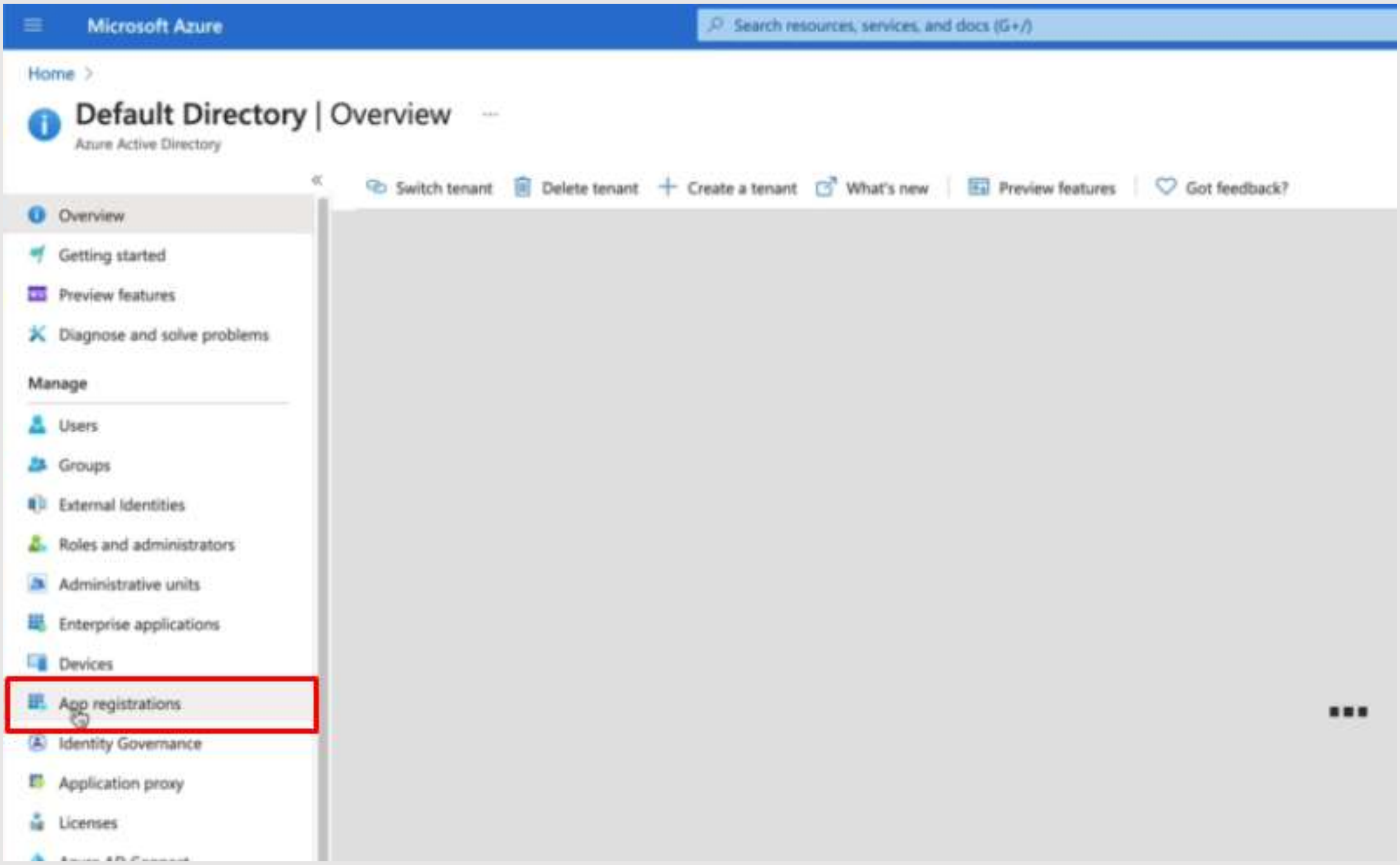
Azure Databricks Integracion con ADLS Usar Active Directory – Utilizar Service Principal / Databricks Scoped Secret

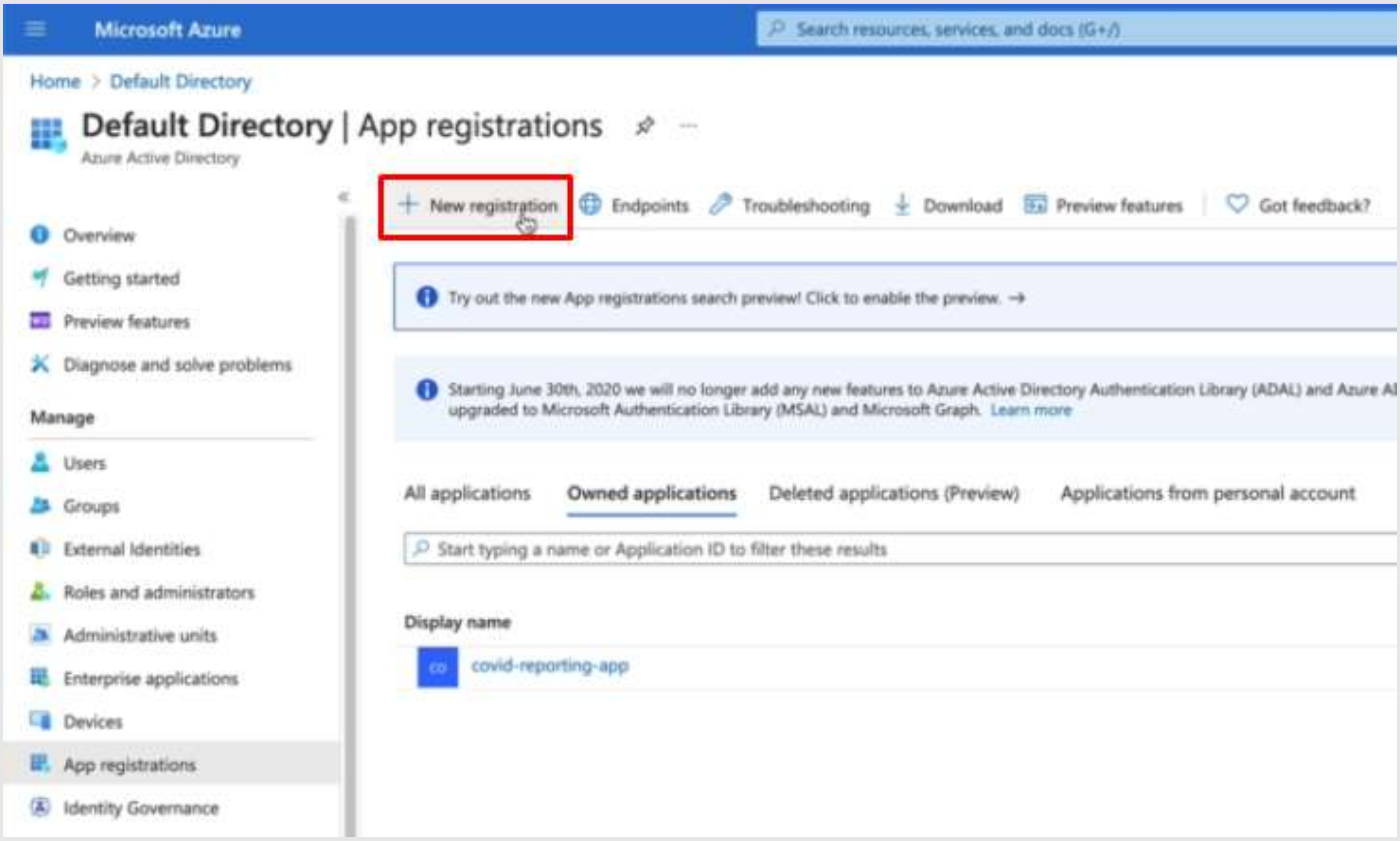
Creando un Azure Service Principal



Recent resources

Name	Type
formula1dl	Storage account
databricks-course-rg	Resource group
covidreportingsa	Storage account
covid-reporting-databricks-ws	Azure Databricks Service
databricks-course-ws	Azure Databricks Service
covid-reporting-adf	Data factory (V2)
dbstorager2n2mkt5svw34	Storage account
workers-vnet	Virtual network
ecdc	Storage account
covid-srv	SQL server
covid-db	SQL database
databricks-demo-rg	Resource group

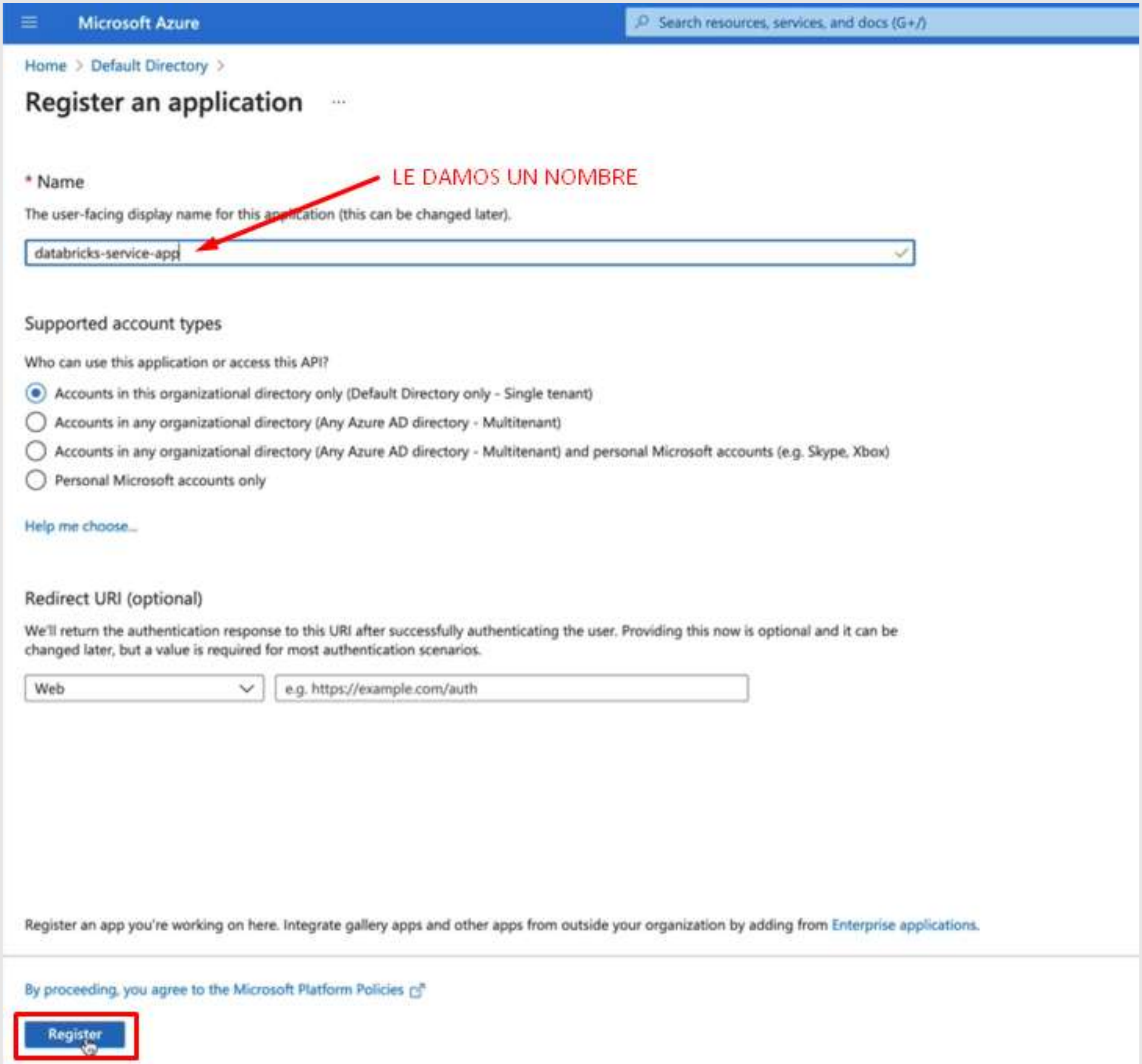




Try out the new App registrations search preview! Click to enable the preview. →

Starting June 30th, 2020 we will no longer add any new features to Azure Active Directory Authentication Library (ADAL) and Azure AI upgraded to Microsoft Authentication Library (MSAL) and Microsoft Graph. Learn more

Start typing a name or Application ID to filter these results



Microsoft Azure
Search resources, services, and docs (G+/)
Home > Default Directory >

Register an application

LE DAMOS UN NOMBRE

* Name

The user-facing display name for this application (this can be changed later).

databricks-service-app

Supported account types

Who can use this application or access this API?

● Accounts in this organizational directory only (Default Directory only - Single tenant)

○ Accounts in any organizational directory (Any Azure AD directory - Multitenant)

○ Accounts in any organizational directory (Any Azure AD directory - Multitenant) and personal Microsoft accounts (e.g. Skype, Xbox)

○ Personal Microsoft accounts only

Help me choose...

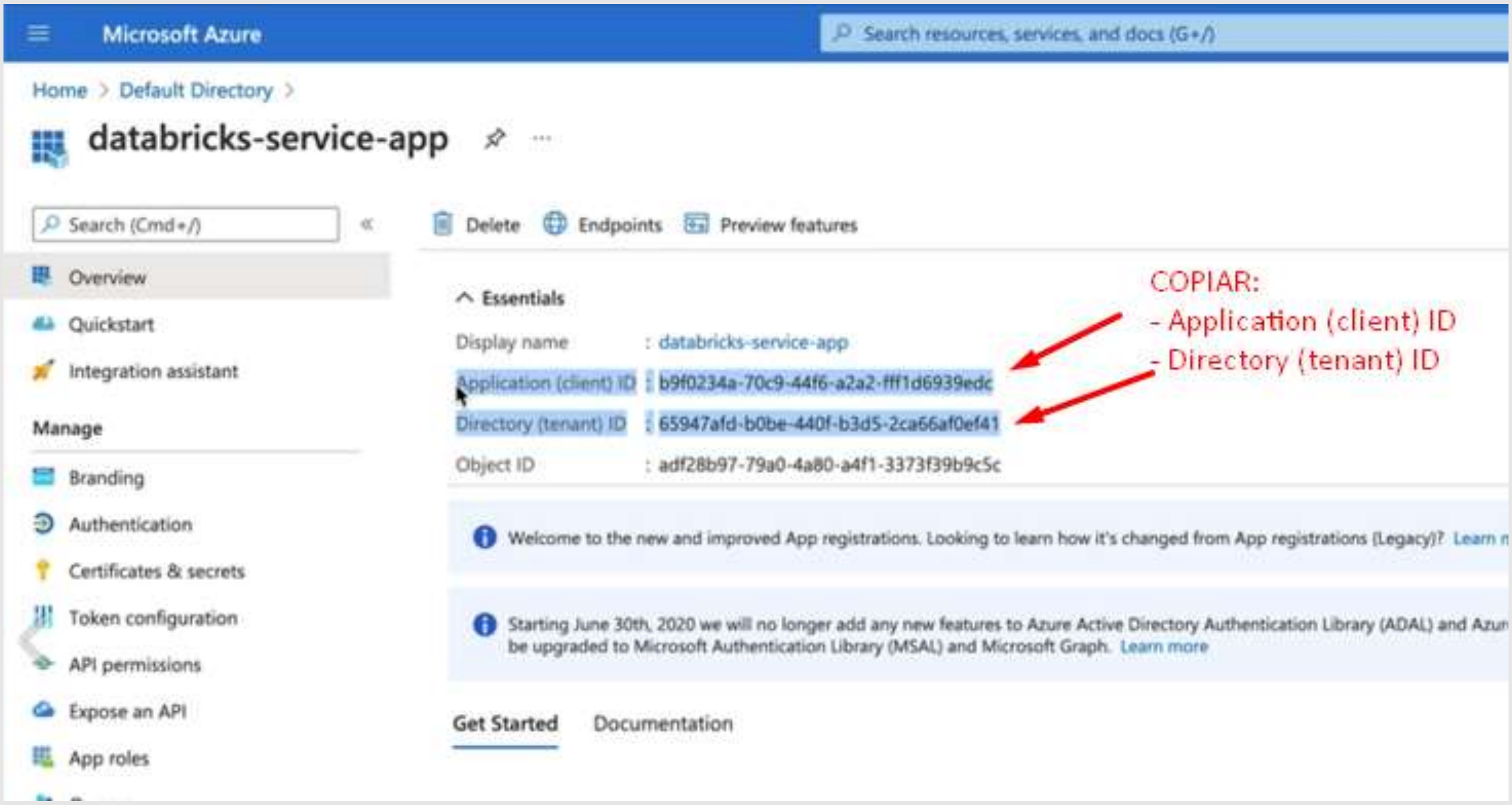
Redirect URI (optional)

We'll return the authentication response to this URI after successfully authenticating the user. Providing this now is optional and it can be changed later, but a value is required for most authentication scenarios.

Web | e.g. https://example.com/auth

Register an app you're working on here. Integrate gallery apps and other apps from outside your organization by adding from Enterprise applications.

By proceeding, you agree to the Microsoft Platform Policies



databricks-service-app

Essentials

Display name : databricks-service-app

Application (client) ID : b9f0234a-70c9-44f6-a2a2-fff1d6939edc

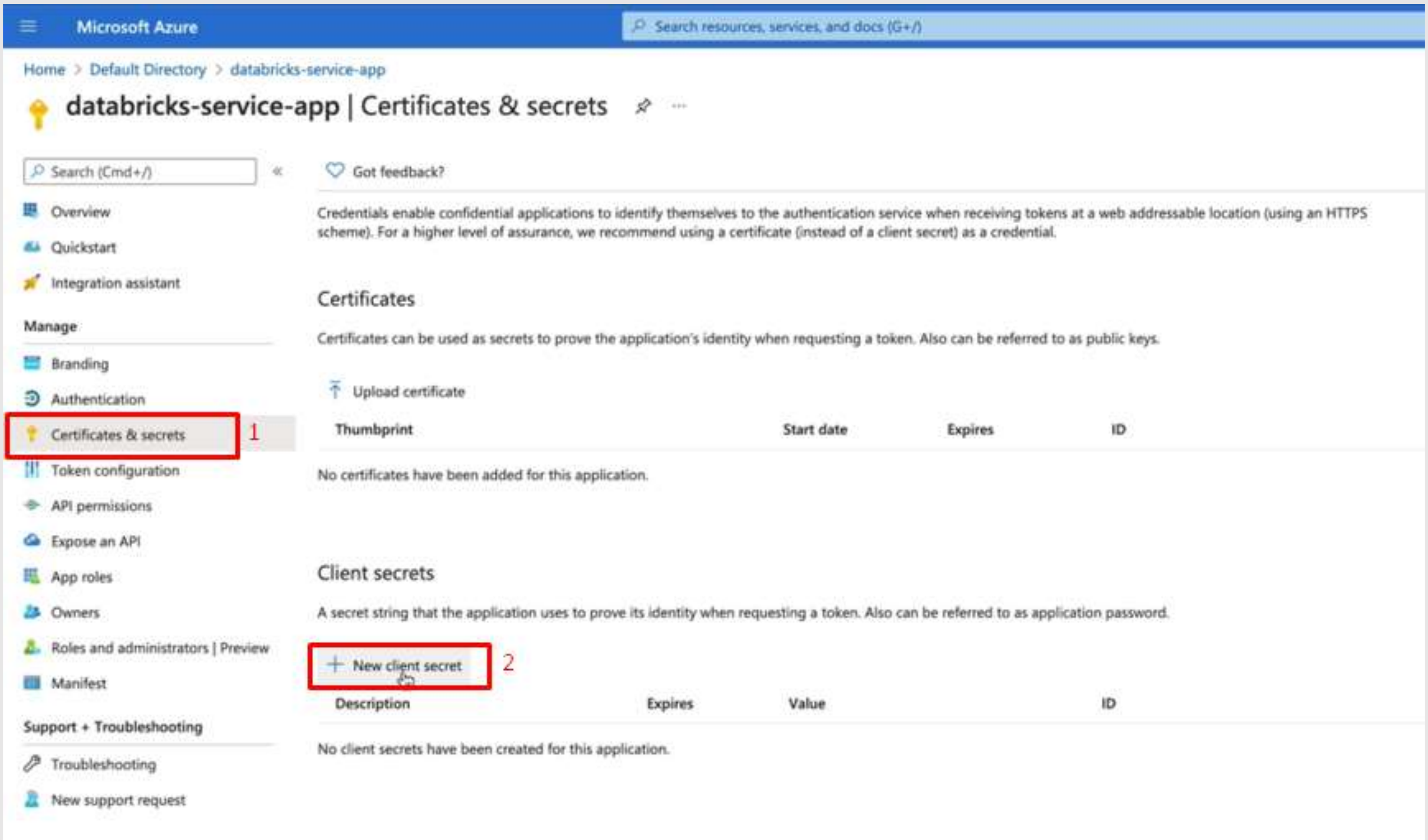
Directory (tenant) ID : 65947afd-b0be-440f-b3d5-2ca66af0ef41

Object ID : adf28b97-79a0-4a80-a4f1-3373f39b9c5c

Welcome to the new and improved App registrations. Looking to learn how it's changed from App registrations (Legacy)? Learn m

Starting June 30th, 2020 we will no longer add any new features to Azure Active Directory Authentication Library (ADAL) and Azur be upgraded to Microsoft Authentication Library (MSAL) and Microsoft Graph. Learn more

COPIAR:
- Application (client) ID
- Directory (tenant) ID



databricks-service-app | Certificates & secrets

Credentials enable confidential applications to identify themselves to the authentication service when receiving tokens at a web addressable location (using an HTTPS scheme). For a higher level of assurance, we recommend using a certificate (instead of a client secret) as a credential.

Certificates

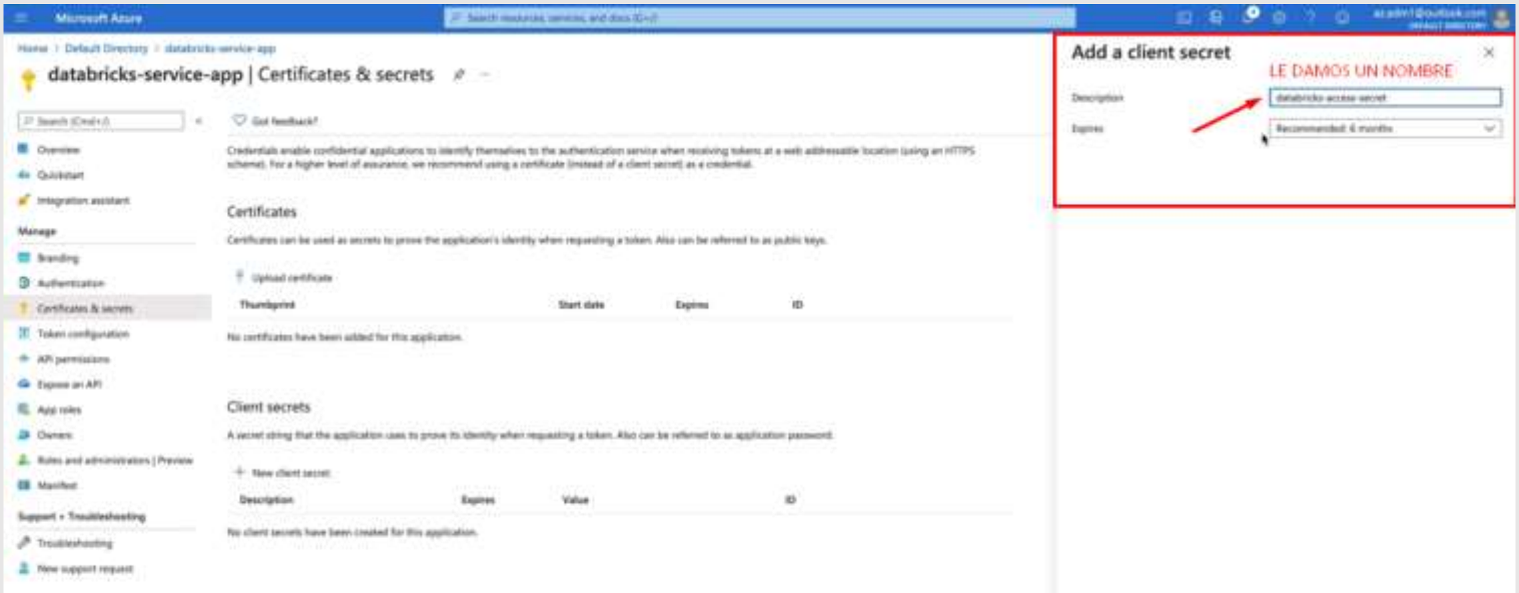
Certificates can be used as secrets to prove the application's identity when requesting a token. Also can be referred to as public keys.

Thumbprint	Start date	Expires	ID
No certificates have been added for this application.			

Client secrets

A secret string that the application uses to prove its identity when requesting a token. Also can be referred to as application password.

Description	Expires	Value	ID
No client secrets have been created for this application.			



databricks-service-app | Certificates & secrets

Credentials enable confidential applications to identify themselves to the authentication service when receiving tokens at a web addressable location (using an HTTPS scheme). For a higher level of assurance, we recommend using a certificate (instead of a client secret) as a credential.

Certificates

Certificates can be used as secrets to prove the application's identity when requesting a token. Also can be referred to as public keys.

Thumbprint	Start date	Expires	ID

No certificates have been added for this application.

Client secrets

A secret string that the application uses to prove its identity when requesting a token. Also can be referred to as application password.

Description	Expires	Value	ID

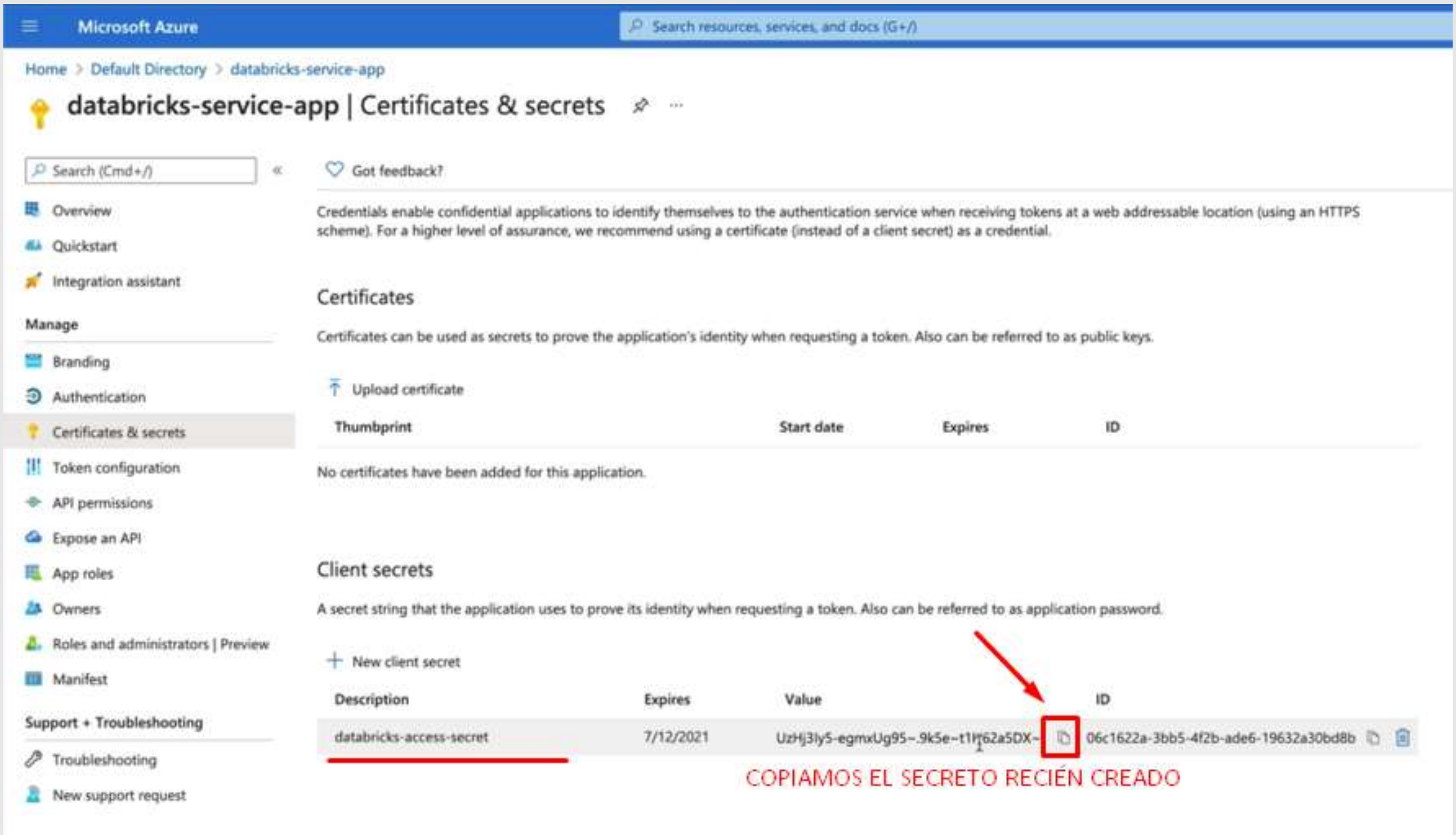
No client secrets have been created for this application.

Add a client secret

LE DAMOS UN NOMBRE

Description: databricks-access-secret

Expires: Recommended: 6 months



databricks-service-app | Certificates & secrets

Credentials enable confidential applications to identify themselves to the authentication service when receiving tokens at a web addressable location (using an HTTPS scheme). For a higher level of assurance, we recommend using a certificate (instead of a client secret) as a credential.

Certificates

Certificates can be used as secrets to prove the application's identity when requesting a token. Also can be referred to as public keys.

Thumbprint	Start date	Expires	ID

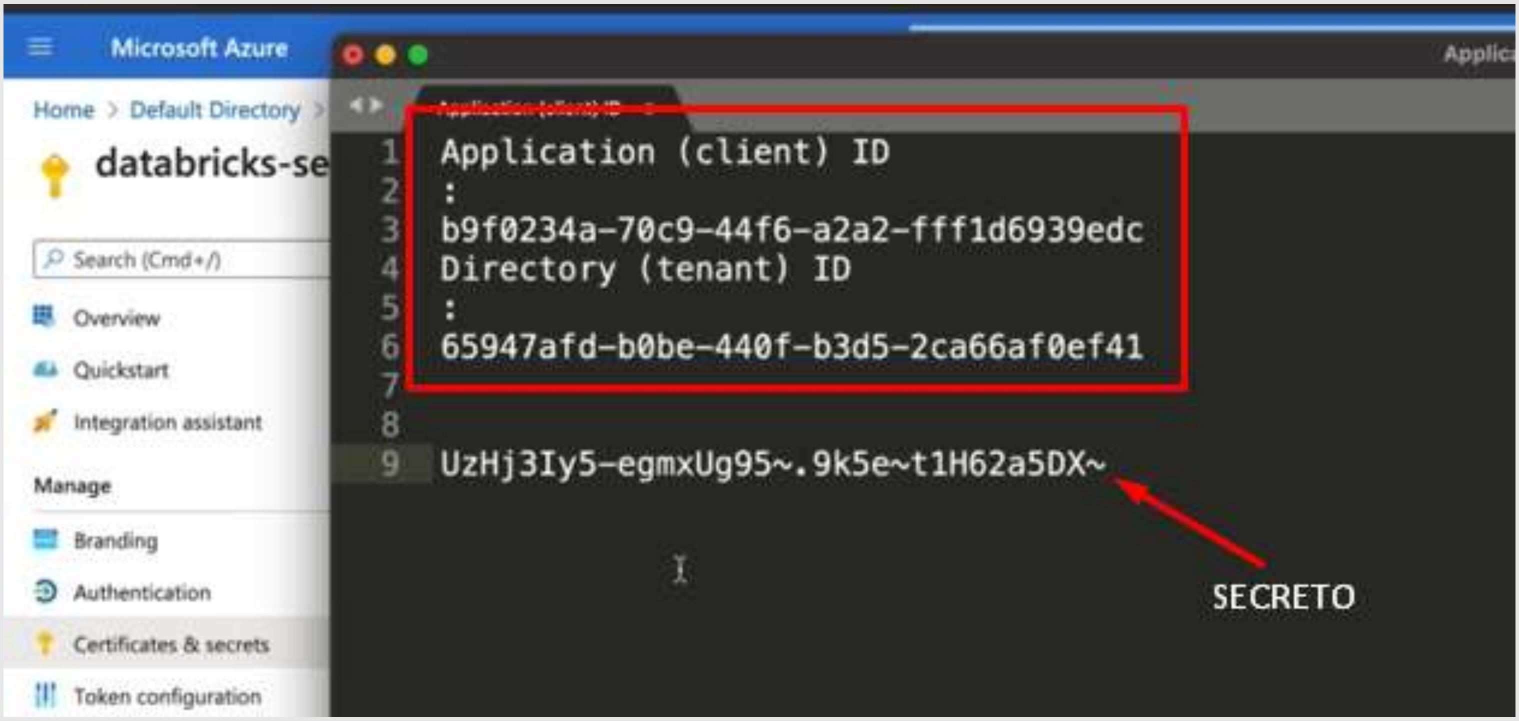
No certificates have been added for this application.

Client secrets

A secret string that the application uses to prove its identity when requesting a token. Also can be referred to as application password.

Description	Expires	Value	ID
databricks-access-secret	7/12/2021	UzHj3ly5-egmxUg95~.9k5e~t1H62a5DX~	06c1622a-3bb5-4f2b-ade6-19632a30bd8b

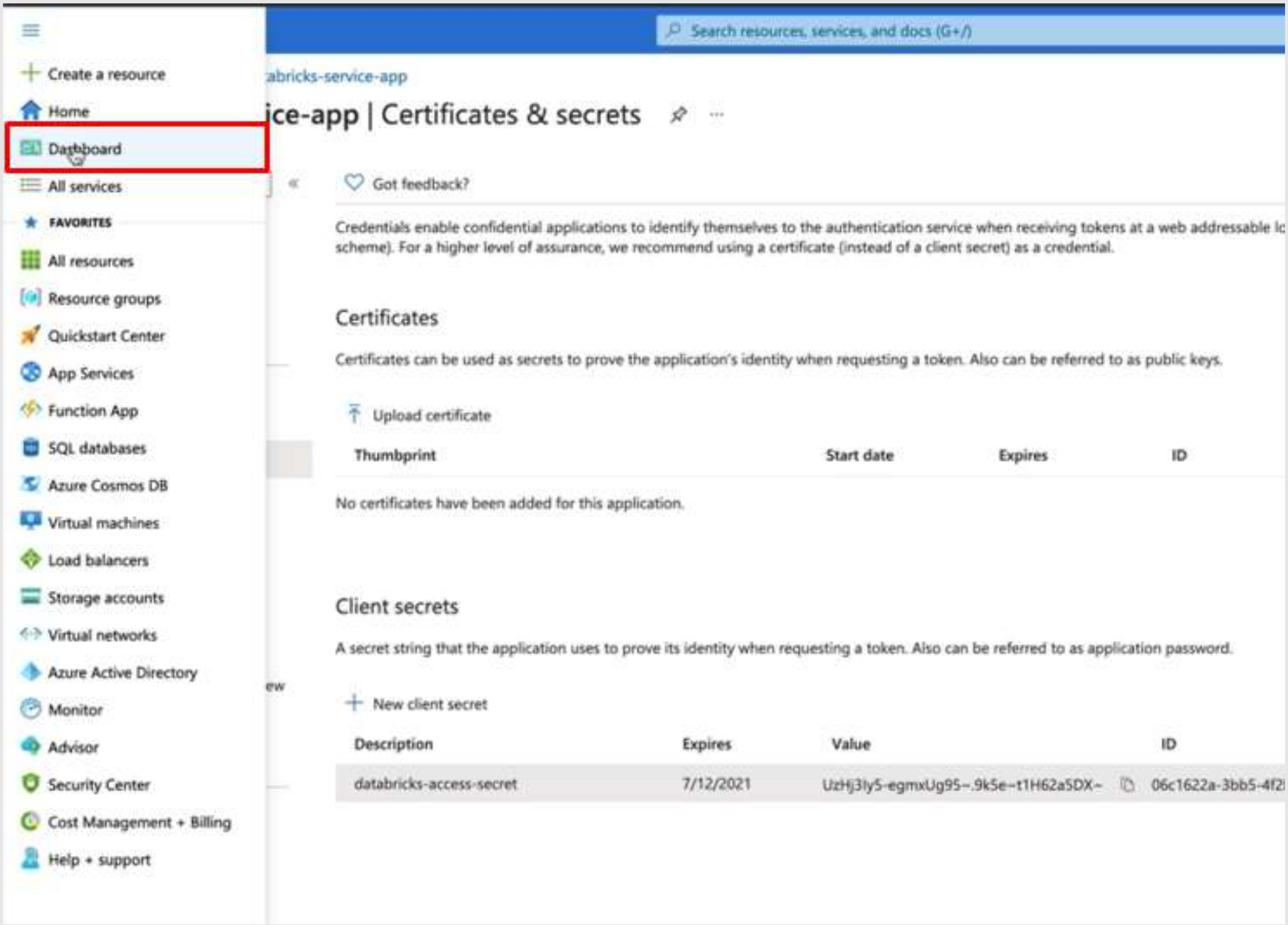
COPIAMOS EL SECRETO RECIÉN CREADO



```
Application (client) ID
:
b9f0234a-70c9-44f6-a2a2-fff1d6939edc
Directory (tenant) ID
:
65947afd-b0be-440f-b3d5-2ca66af0ef41


UzHj3Iy5-egmxUg95~.9k5e~t1H62a5DX~
```

SECRETO



Credentials enable confidential applications to identify themselves to the authentication service when receiving tokens at a web addressable lo scheme). For a higher level of assurance, we recommend using a certificate (instead of a client secret) as a credential.

Certificates

Certificates can be used as secrets to prove the application's identity when requesting a token. Also can be referred to as public keys.

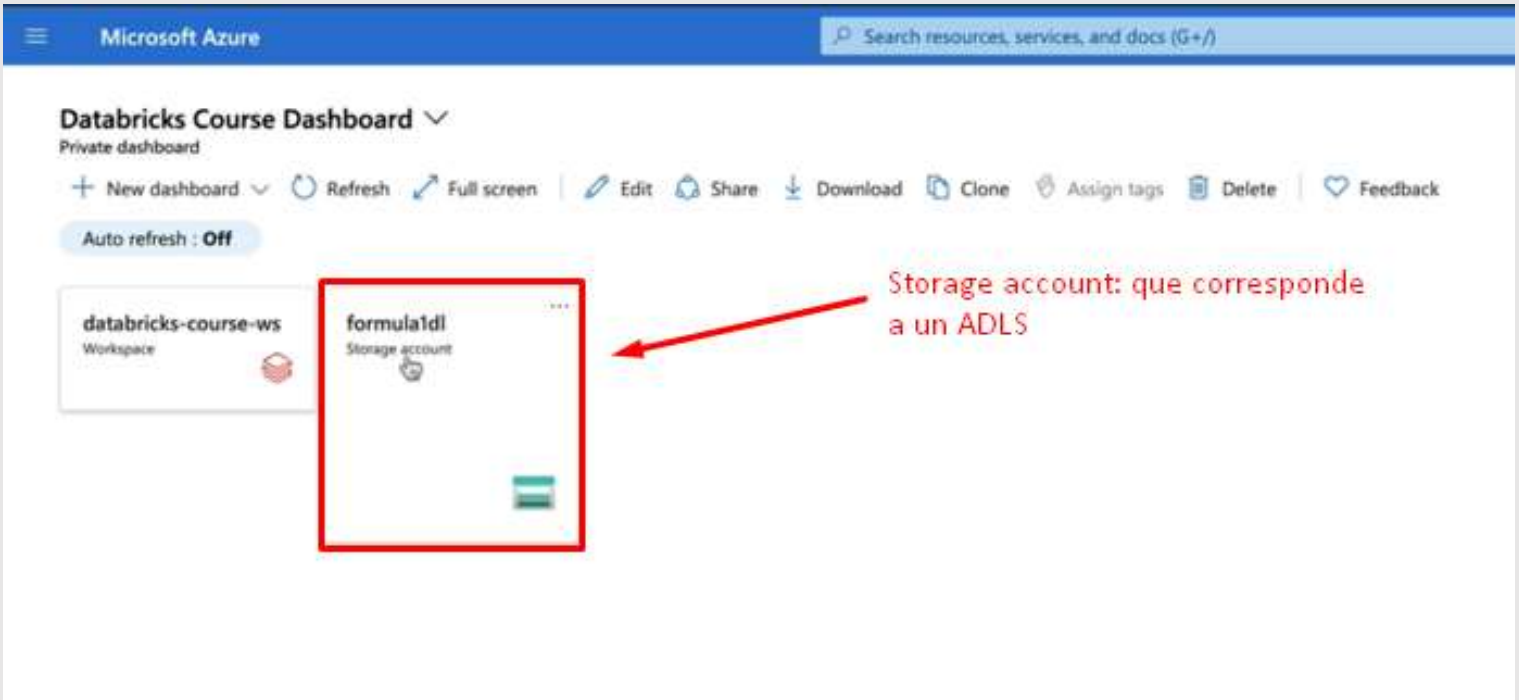
Thumbprint	Start date	Expires	ID

No certificates have been added for this application.

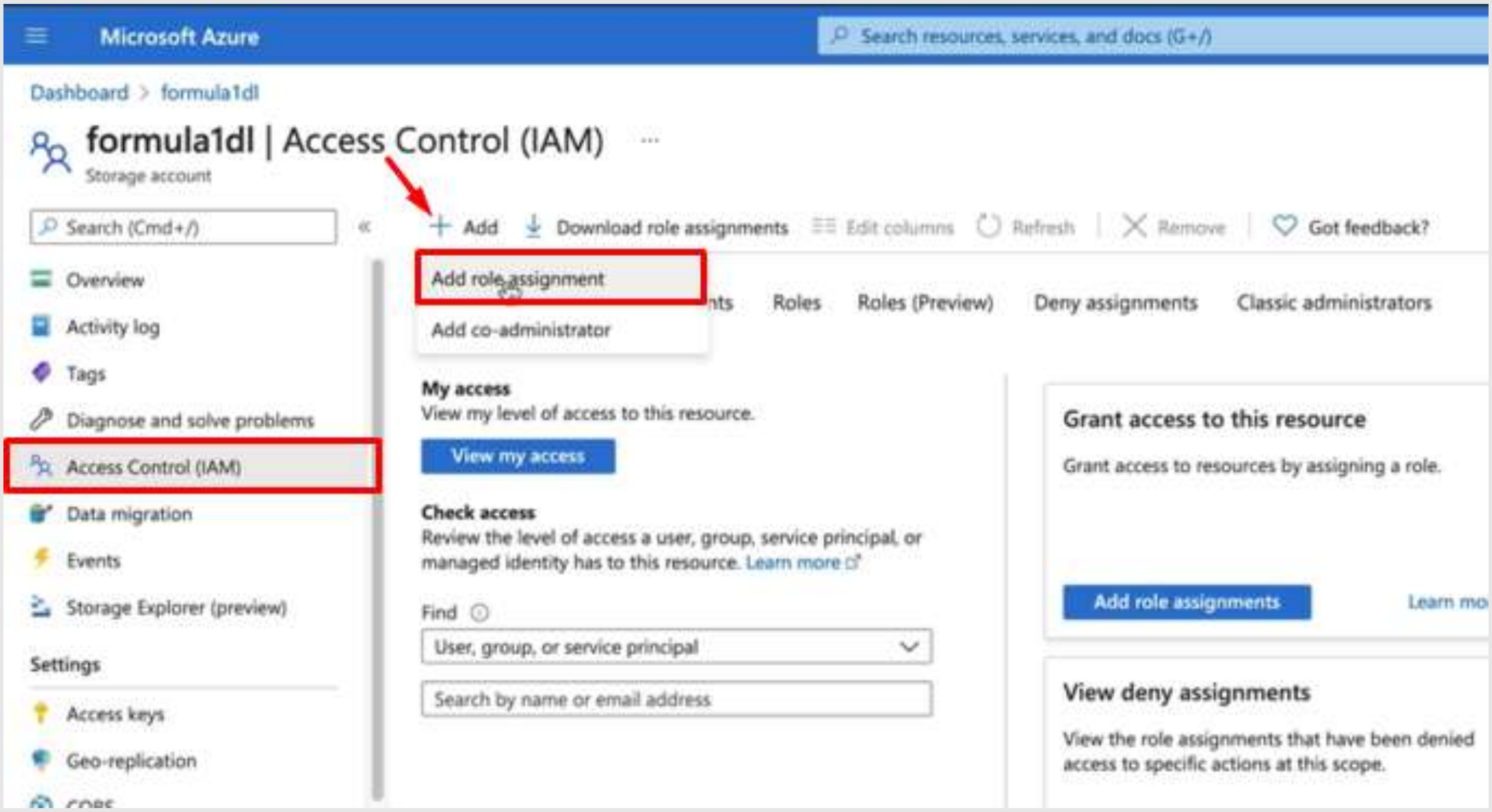
Client secrets

A secret string that the application uses to prove its identity when requesting a token. Also can be referred to as application password.

Description	Expires	Value	ID
databricks-access-secret	7/12/2021	UzHj3Iy5-egmxUg95~.9k5e~t1H62a5DX~	06c1622a-3bb5-4f2



Storage account: que corresponde a un ADLS



My access

View my level of access to this resource.

Check access

Review the level of access a user, group, service principal, or managed identity has to this resource. Learn more

Find: User, group, or service principal

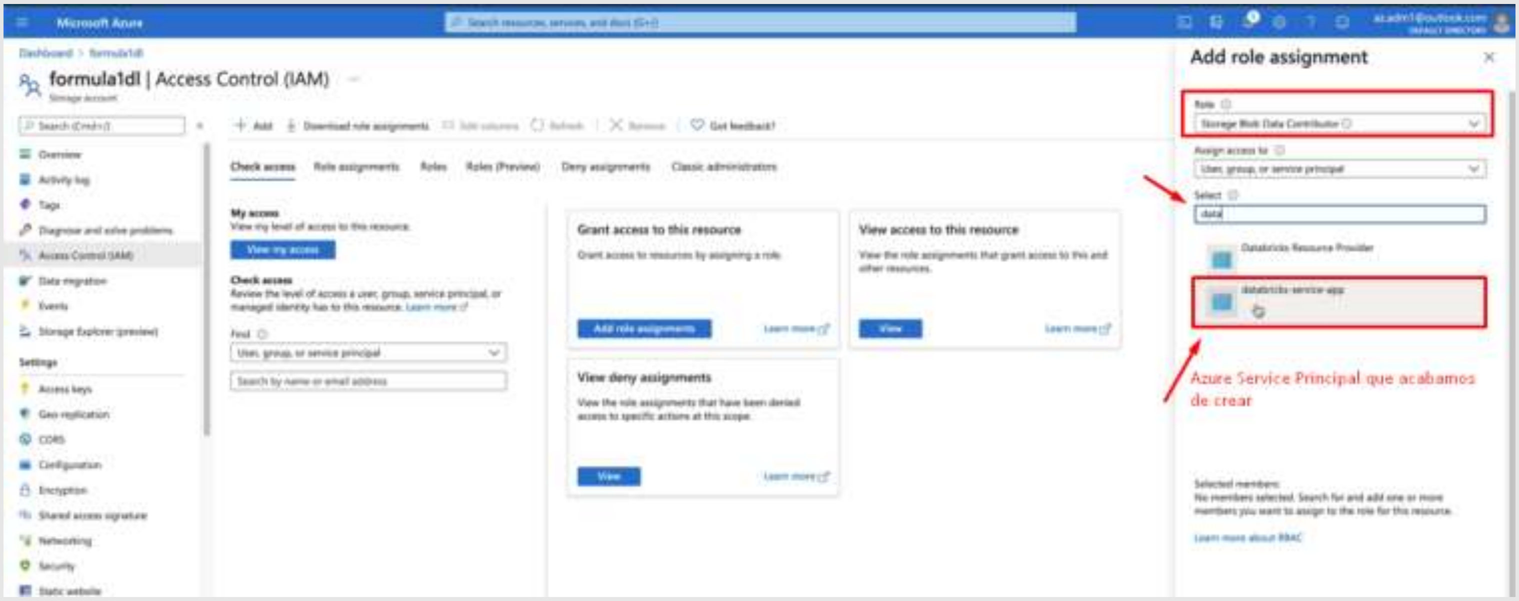
Search by name or email address

Grant access to this resource

Grant access to resources by assigning a role.

View deny assignments

View the role assignments that have been denied access to specific actions at this scope.



formula1dl | Access Control (IAM)

Storage account

My access

View my level of access to this resource.

Check access

Review the level of access a user, group, service principal, or managed identity has to this resource. Learn more

Find: User, group, or service principal

Search by name or email address

Grant access to this resource

Grant access to resources by assigning a role.

View access to this resource

View the role assignments that grant access to this and other resources.

View deny assignments

View the role assignments that have been denied access to specific actions at this scope.

Add role assignment

Role: Storage Blob Data Contributor

Assign access to: User, group, or service principal

Select: data

Databricks Resource Provider

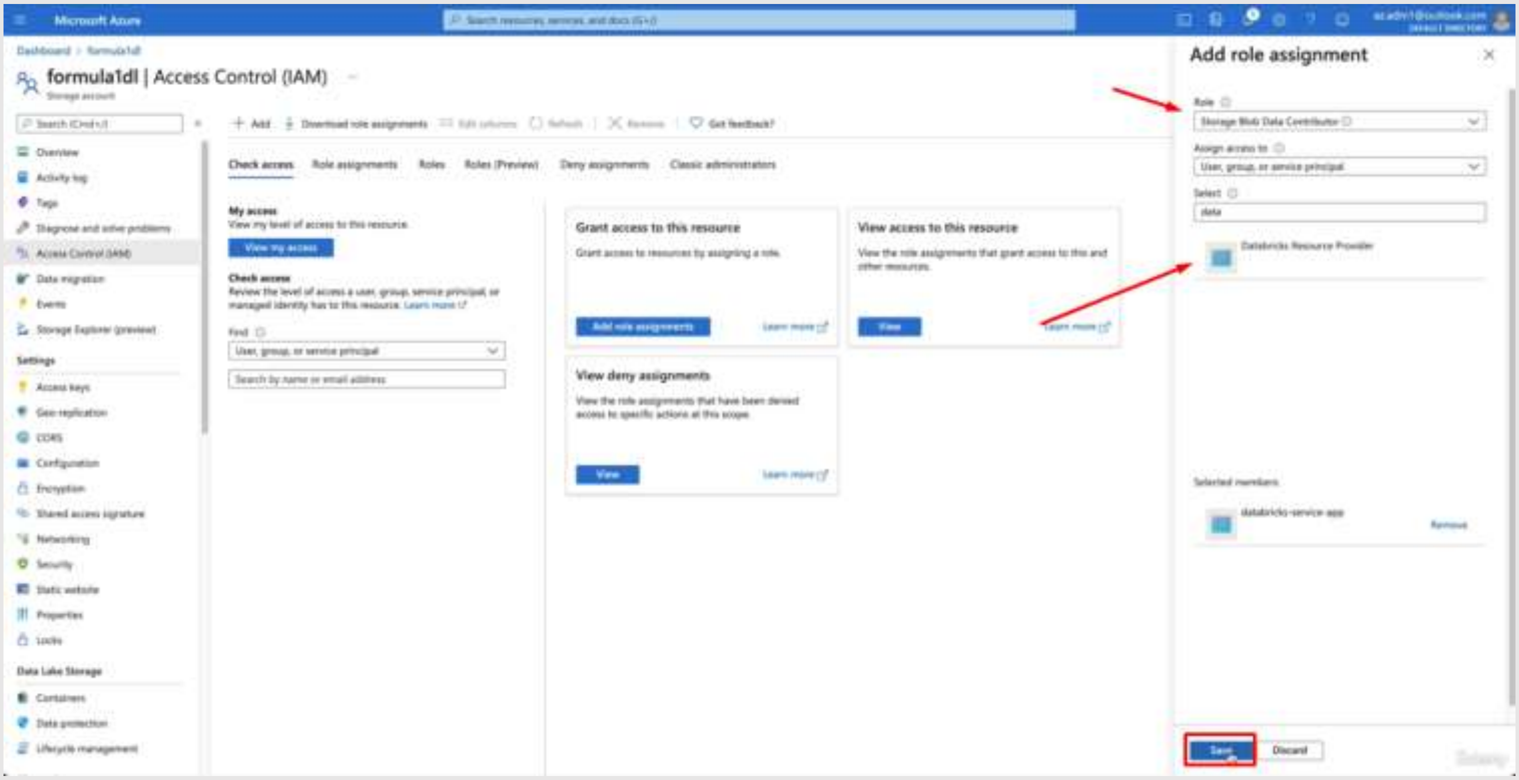
databricks-service-app

Azure Service Principal que acabamos de crear

Selected members:

No members selected. Search for and add one or more members you want to assign to the role for this resource.

Learn more about RBAC



formula1dl | Access Control (IAM)

Storage account

Add role assignment

Role: Storage Blob Data Contributor

Assign access to: User, group, or service principal

Select: data

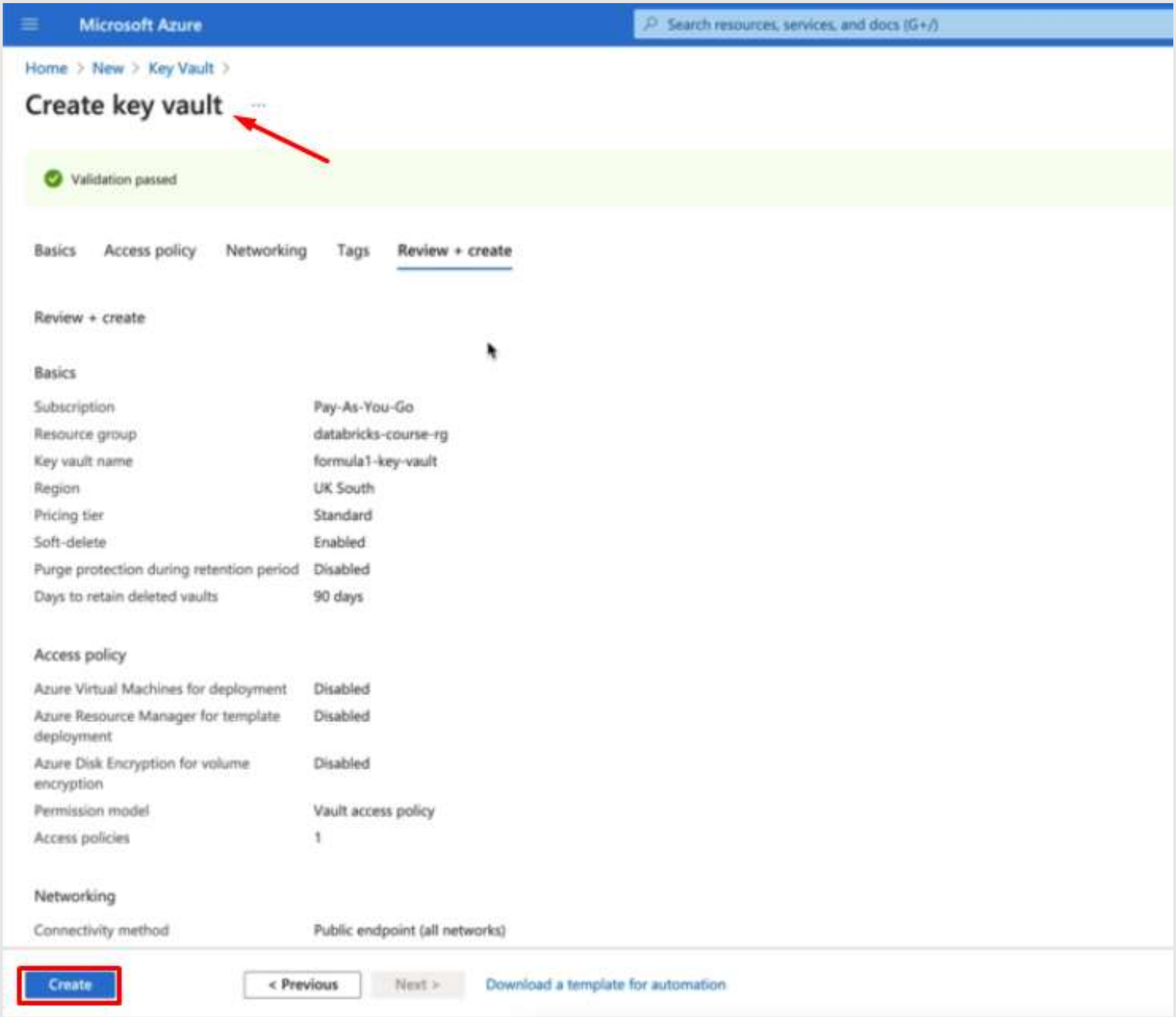
Databricks Resource Provider

Selected members

databricks-service-app Remove

Mounting Azure Data Lake Storage Gen2

Antes que todo, vamos a utilizar Azure Key Vault para almacenar el Client ID, Tenant ID y el Client Secret.



Create key vault

Validation passed

Review + create

Basics

Subscription	Pay-As-You-Go
Resource group	databricks-course-rg
Key vault name	formula1-key-vault
Region	UK South
Pricing tier	Standard
Soft-delete	Enabled
Purge protection during retention period	Disabled
Days to retain deleted vaults	90 days

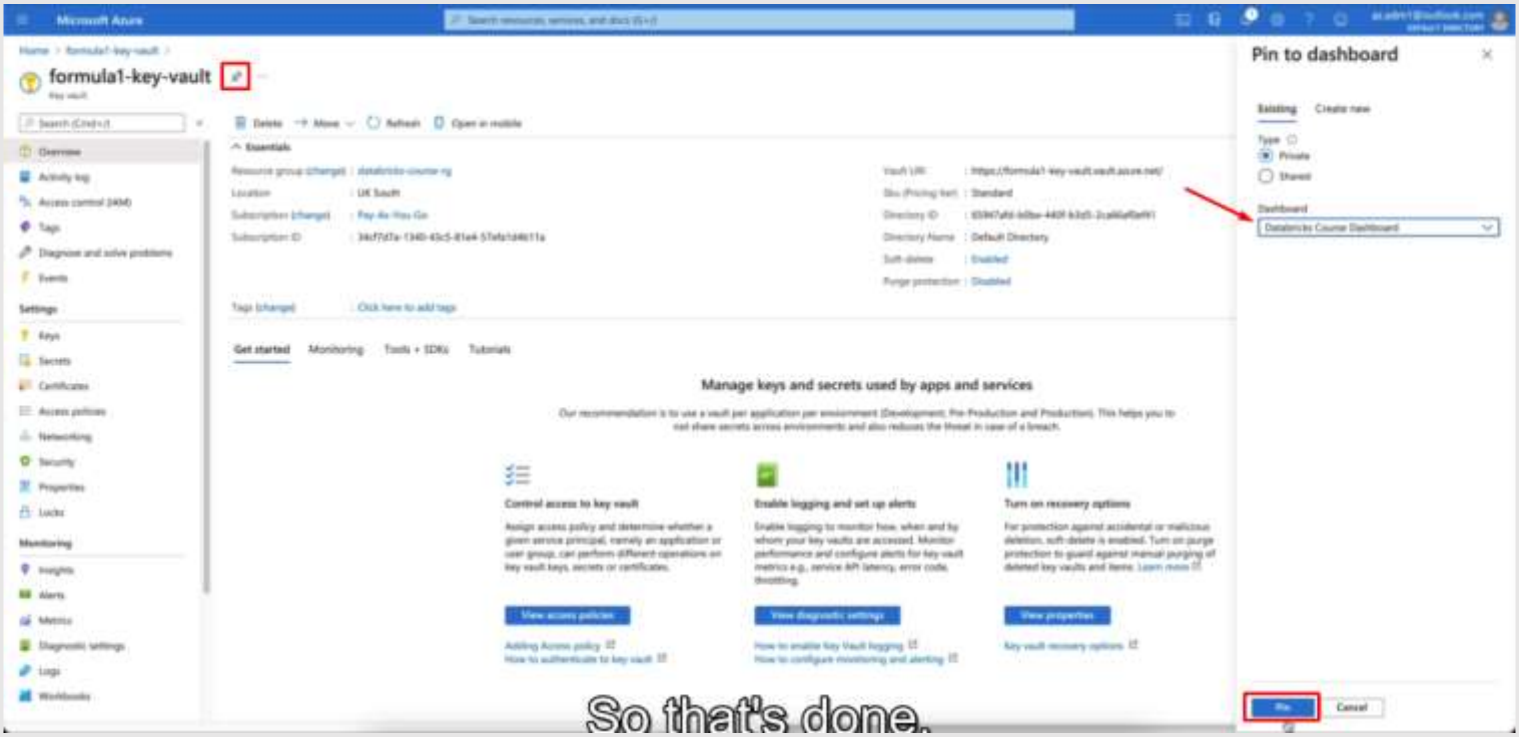
Access policy

Azure Virtual Machines for deployment	Disabled
Azure Resource Manager for template deployment	Disabled
Azure Disk Encryption for volume encryption	Disabled
Permission model	Vault access policy
Access policies	1

Networking

Connectivity method	Public endpoint (all networks)

Así agregamos el servicio al dashboard



formula1-key-vault
Key vault

Resource group (change) : databricks-course-rg
Location : UK South
Subscription (change) : Pay-As-You-Go
Subscription ID : 34cf7d7a-1340-43c5-81e4-57efa1d4611a
Tags (change) : Click here to add tags

Vault URI : https://formula1-key-vault.vault.azure.net/
Sku (Pricing tier) : Standard
Directory ID : 65947af6-b0be-440f-b3d5-2ca66af0ef41
Directory Name : Default Directory
Soft-delete : Enabled
Purge protection : Disabled

Manage keys and secrets used by apps and services

Our recommendation is to use a vault per application per environment (Development, Pre-Production and Production). This helps you to not share secrets across environments and also reduces the threat in case of a breach.

Control access to key vault

Assign access policy and determine whether a given service principal, namely an application or user group, can perform different operations on key vault keys, secrets or certificates.

Enable logging and set up alerts

Enable logging to monitor how, when and by whom your key vaults are accessed. Monitor performance and configure alerts for key vault metrics e.g., service API latency, error code, throttling.

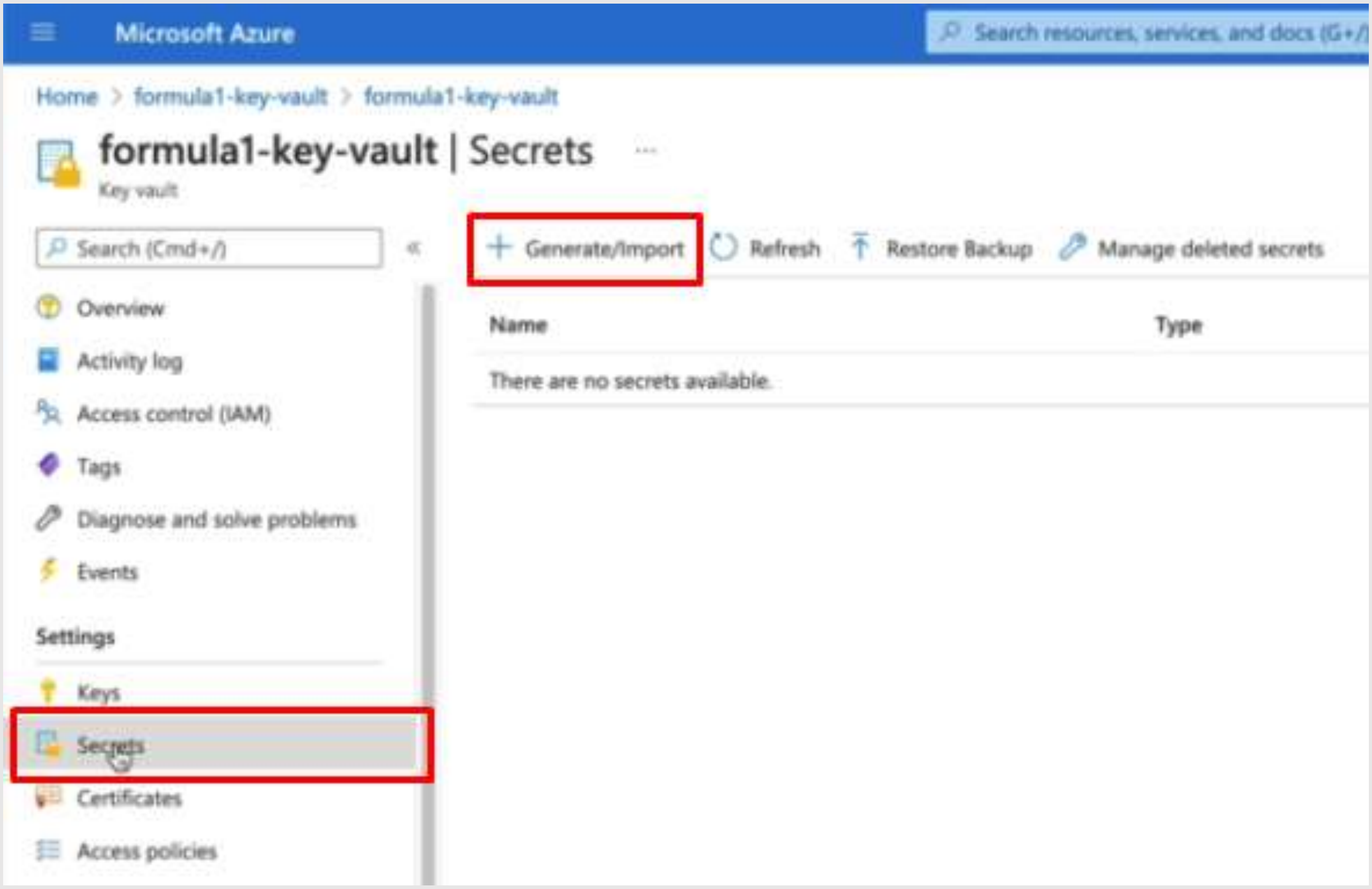
Turn on recovery options

For protection against accidental or malicious deletion, soft-delete is enabled. Turn on purge protection to guard against manual purging of deleted key vaults and items. Learn more

Pin to dashboard

Type
● Private
○ Shared

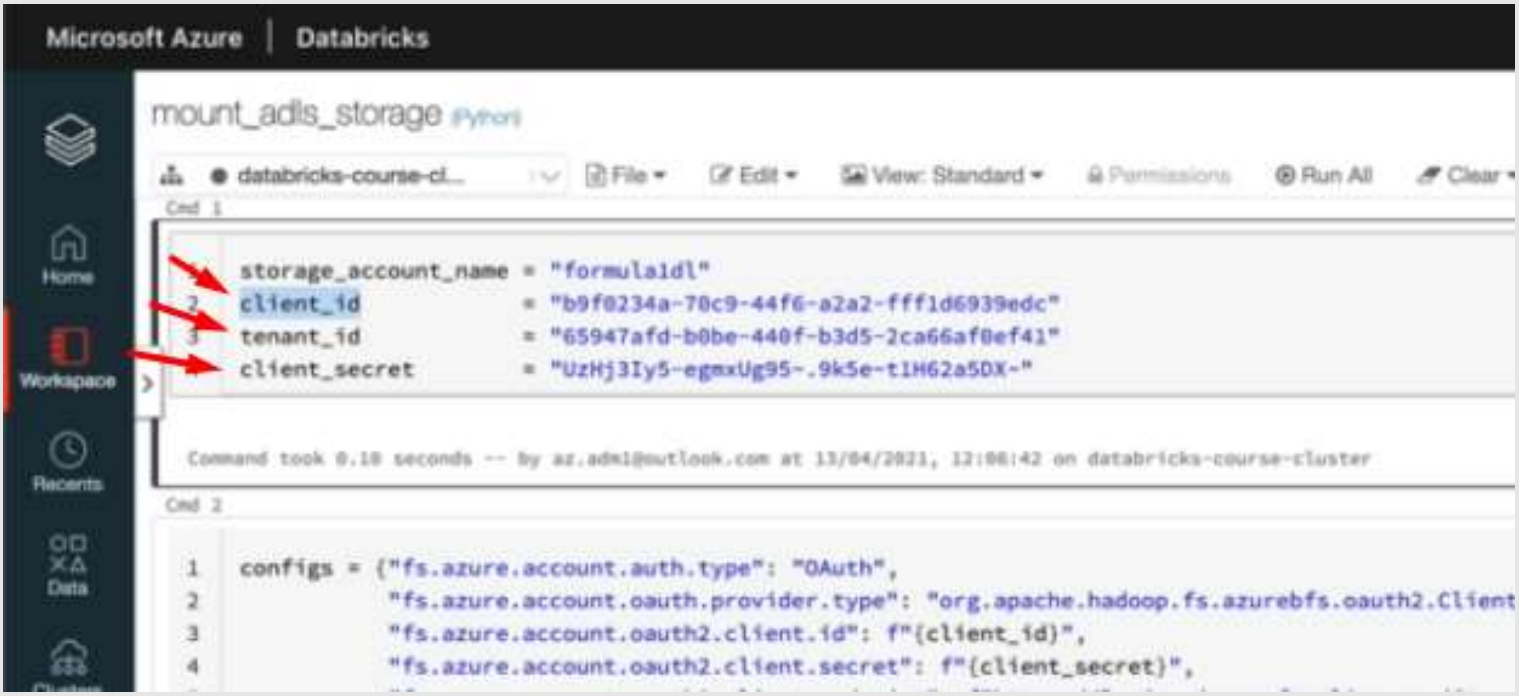
Dashboard: Databricks Course Dashboard



formula1-key-vault | Secrets
Key vault

Name	Type
There are no secrets available.	

Tenemos que generar 3 secretos en Azure Key Vault



mount_adls_storage (Python)

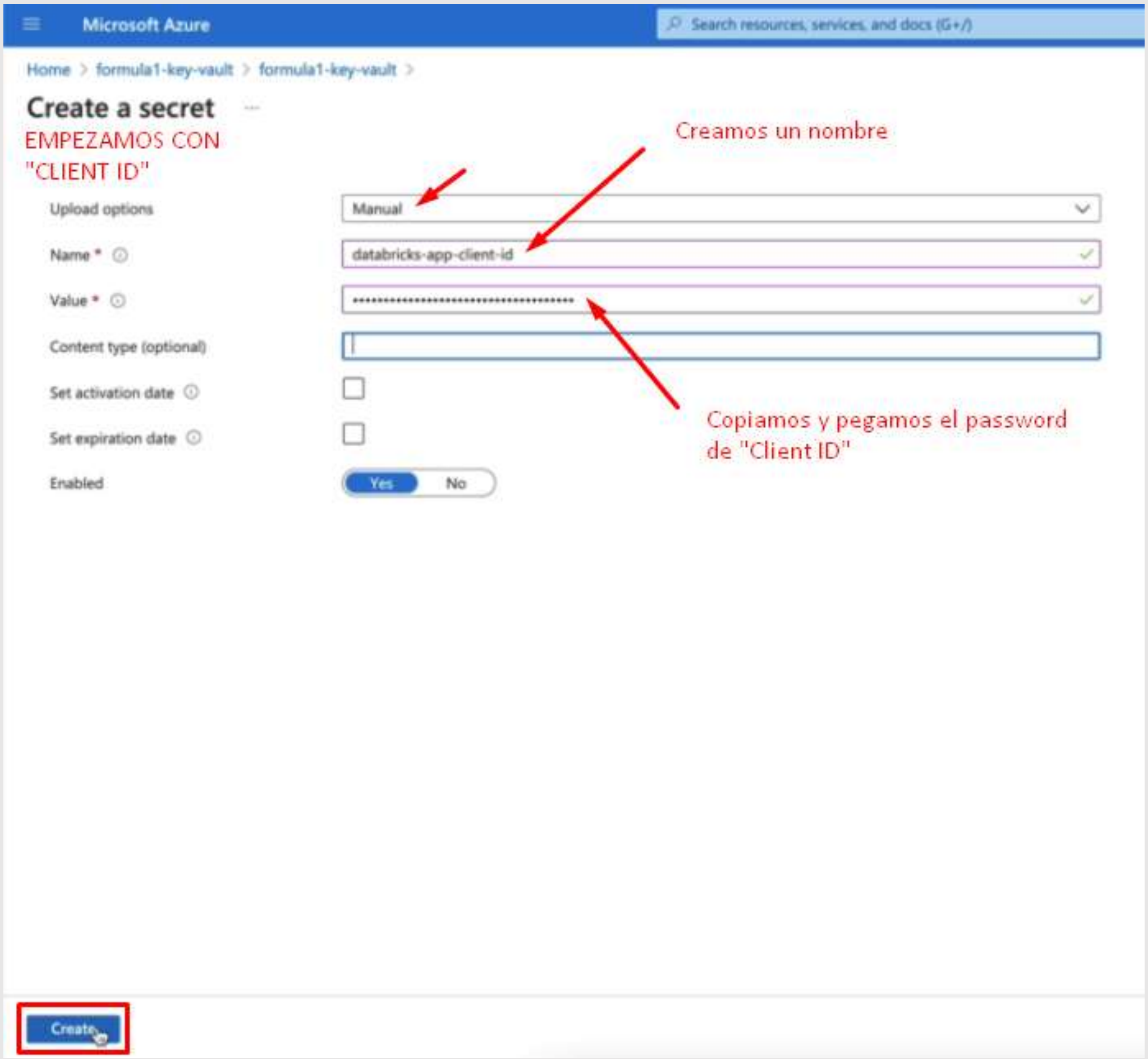
Cmd 1

```
storage_account_name = "formula1dl"
client_id            = "b9f0234a-70c9-44f6-a2a2-fff1d6939edc"
tenant_id            = "65947afd-b0be-440f-b3d5-2ca66af0ef41"
client_secret        = "UzHj3Iy5-egmxUg95-.9k5e-t1H62a5DX-"
```

```
Command took 0.10 seconds -- by az.admi@outlook.com at 13/04/2021, 12:06:42 on databricks-course-cluster
```

Cmd 2

```
configs = {"fs.azure.account.auth.type": "OAuth",
           "fs.azure.account.oauth.provider.type": "org.apache.hadoop.fs.azurebfs.oauth2.Client
           "fs.azure.account.oauth2.client.id": f"{client_id}",
           "fs.azure.account.oauth2.client.secret": f"{client_secret}",
```


Microsoft Azure
Search resources, services, and docs (G+/)
Home > formula1-key-vault > formula1-key-vault >
Create

Create a secret

EMPEZAMOS CON "CLIENT ID"

Creamos un nombre

Upload options: Manual

Name *: databricks-app-client-id

Value *: ••••••••••••••••••••••••••••••••••••

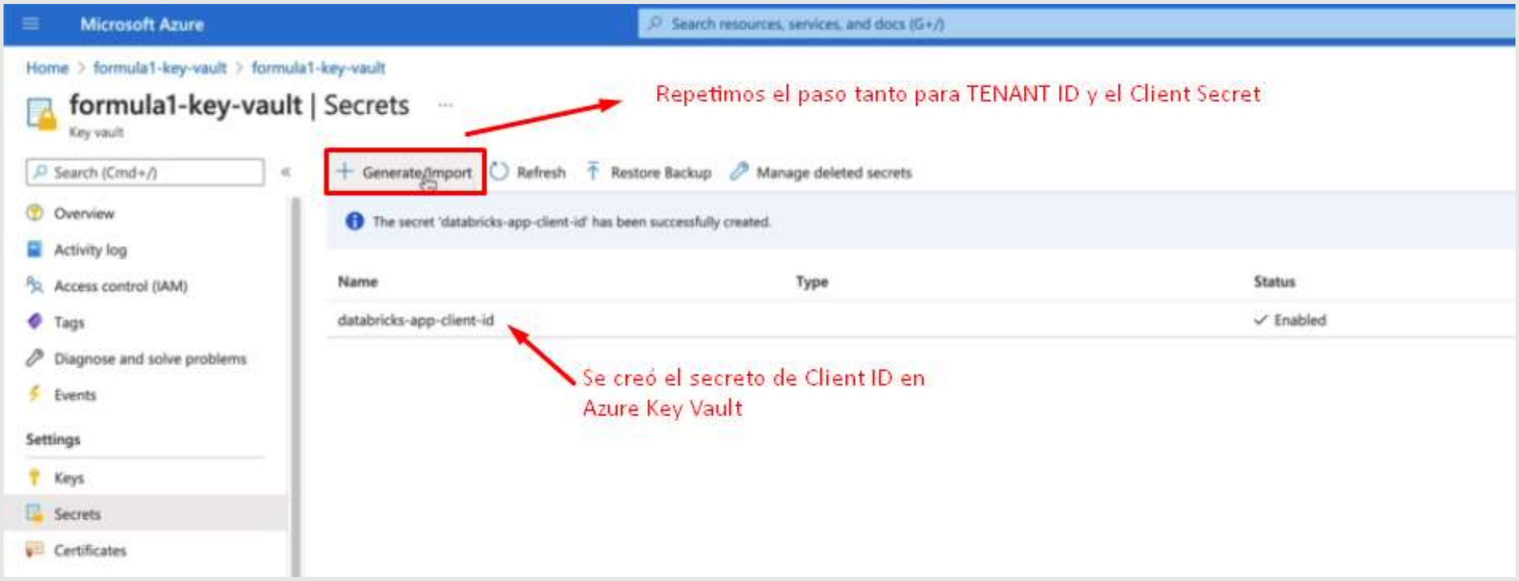
Content type (optional):

Set activation date □

Set expiration date □

Enabled: Yes No

Copiamos y pegamos el password de "Client ID"



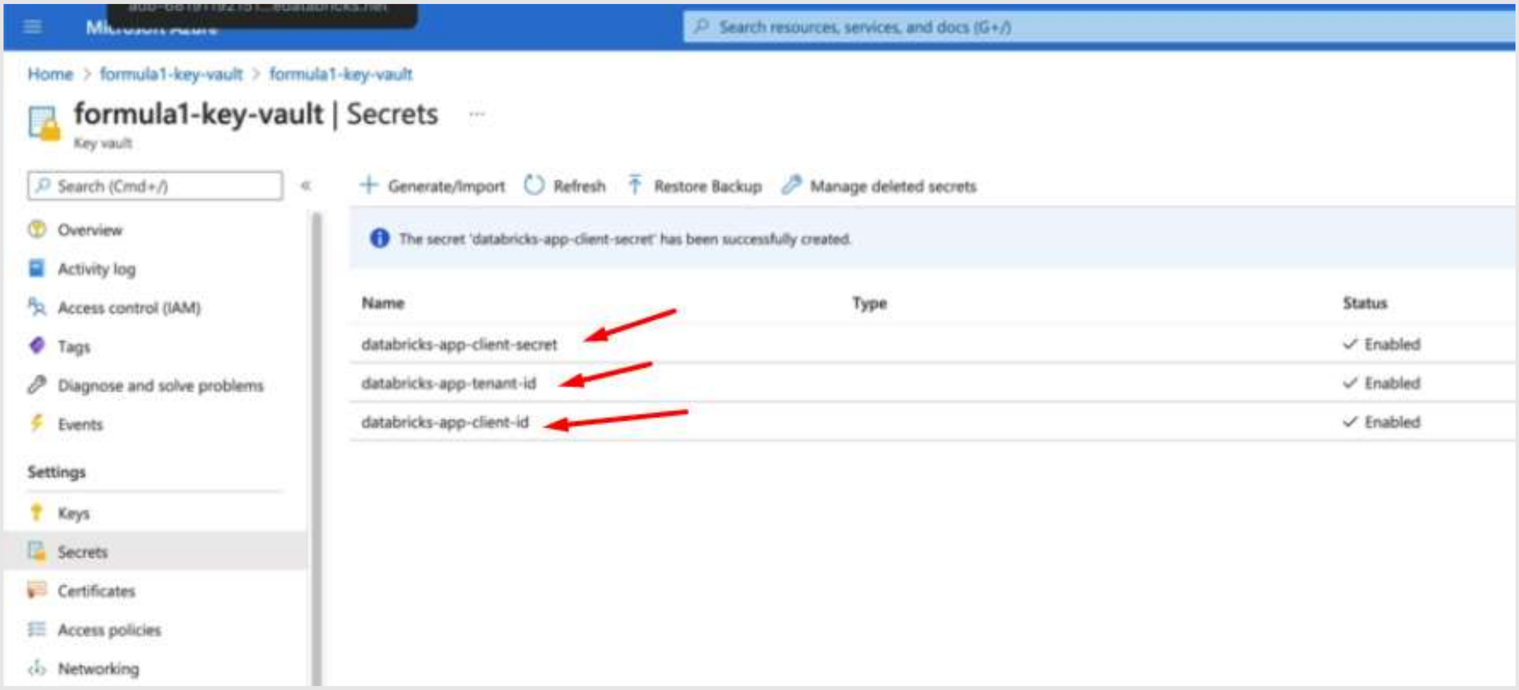
Repetimos el paso tanto para TENANT ID y el Client Secret

The secret 'databricks-app-client-id' has been successfully created.

Name	Type	Status
databricks-app-client-id		✓ Enabled

Se creó el secreto de Client ID en Azure Key Vault

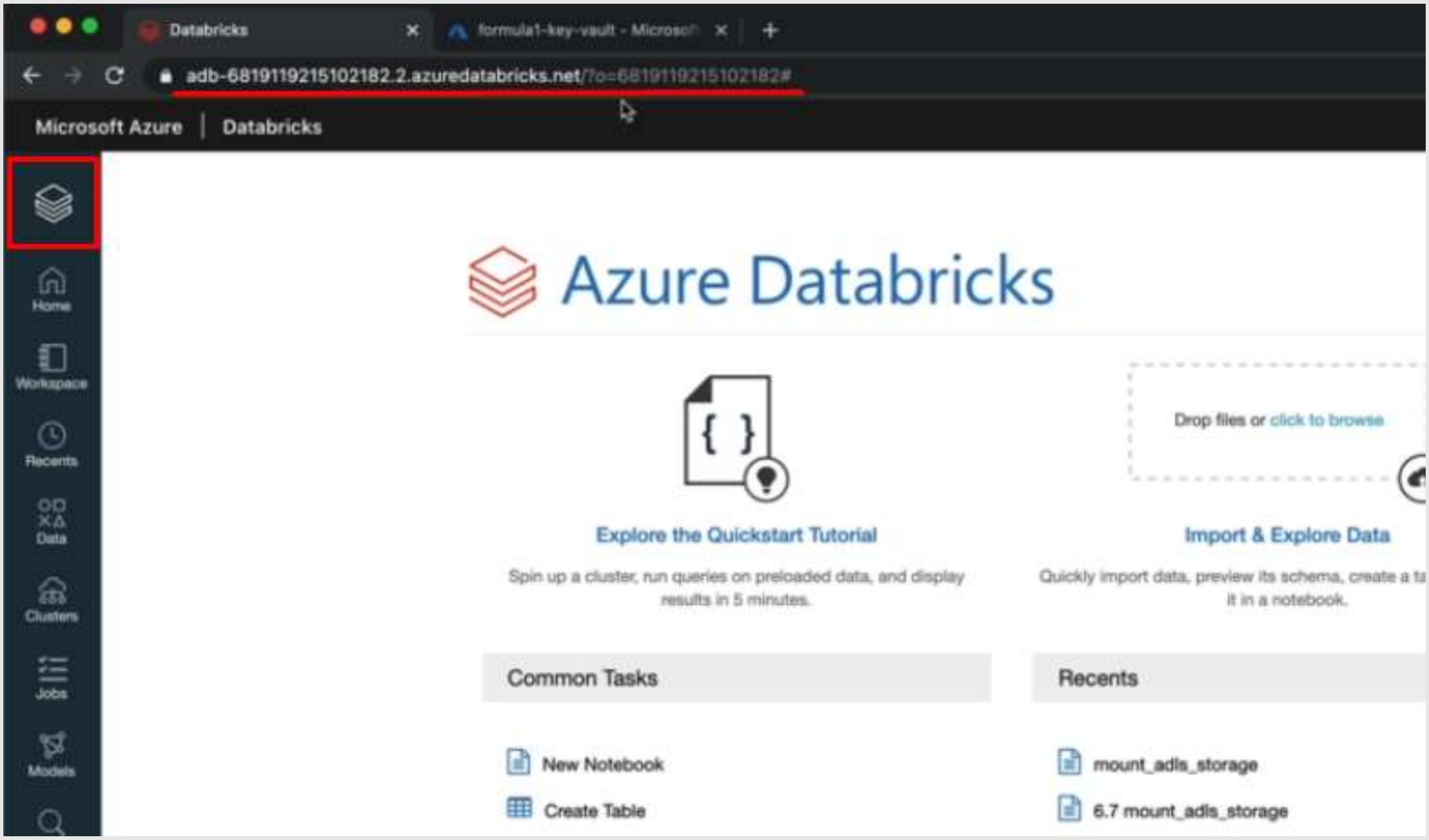
Se han creado los 3 secretos



The secret 'databricks-app-client-secret' has been successfully created.

Name	Type	Status
databricks-app-client-secret		✓ Enabled
databricks-app-tenant-id		✓ Enabled
databricks-app-client-id		✓ Enabled

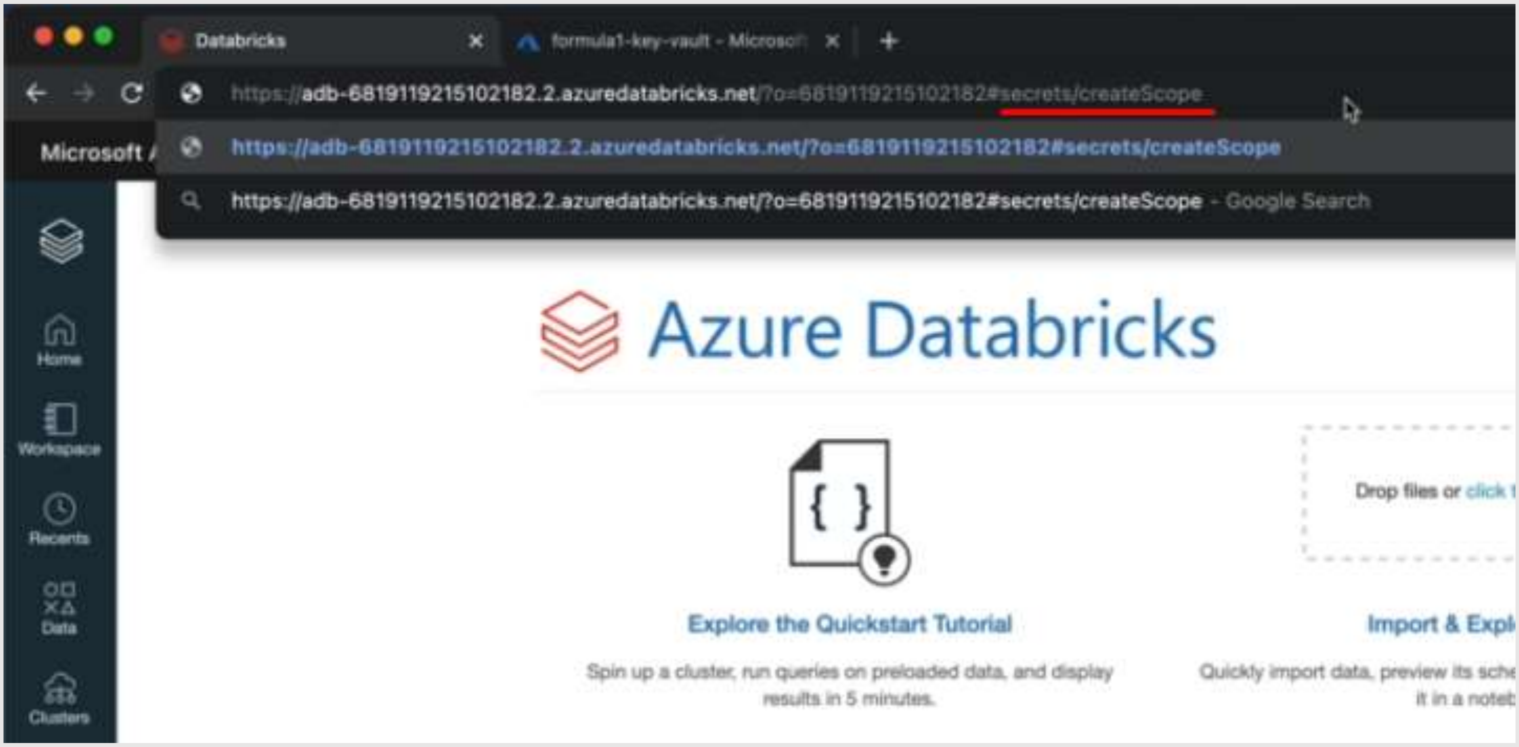
Luego, damos clic en el símbolo de databricks y en la ruta web en el navegador agregaremos **#secrets/createScope**. Vamos a utilizar **Databricks Scoped Secret**



adb-6819119215102182.2.azuredatabricks.net/?o=6819119215102182#

Spin up a cluster, run queries on preloaded data, and display results in 5 minutes.

Quickly import data, preview its schema, create a ta it in a notebook.



https://adb-6819119215102182.2.azuredatabricks.net/?o=6819119215102182#secrets/createScope

https://adb-6819119215102182.2.azuredatabricks.net/?o=6819119215102182#secrets/createScope

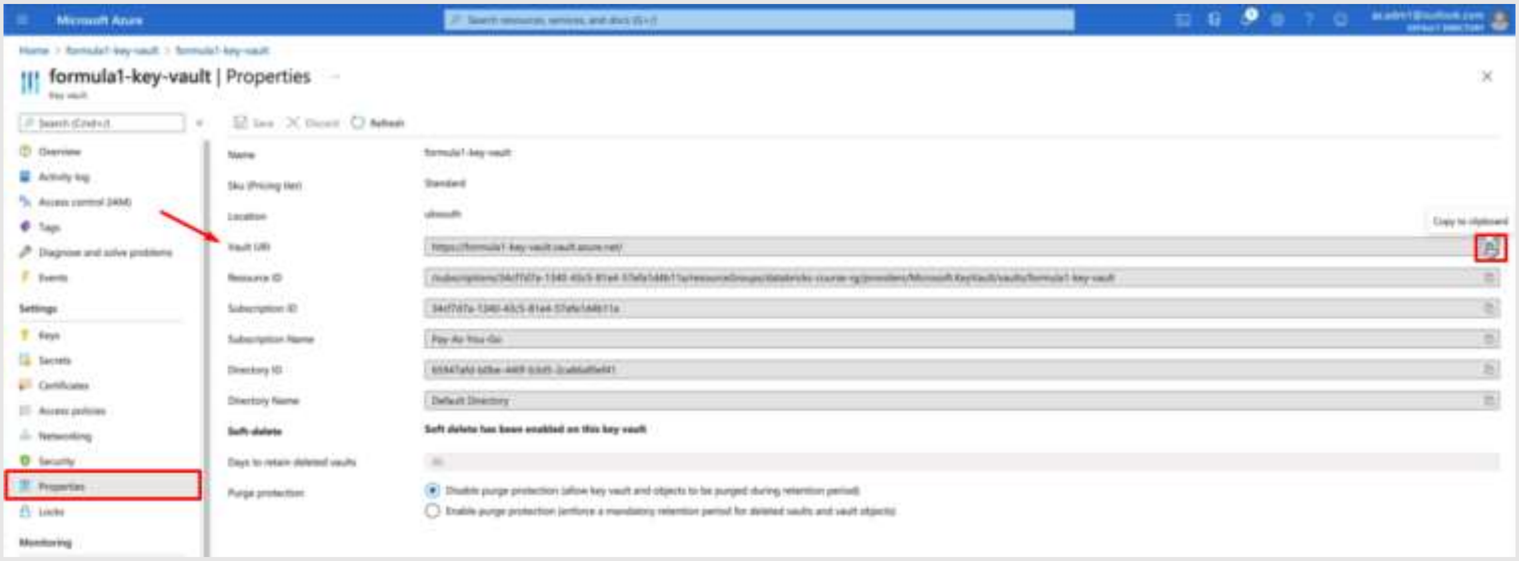
https://adb-6819119215102182.2.azuredatabricks.net/?o=6819119215102182#secrets/createScope - Google Search

Spin up a cluster, run queries on preloaded data, and display results in 5 minutes.

Quickly import data, preview its sche it in a noteb

Regresamos al Azure Key Vault

Vault URI



Name formula1-key-vault

Sku (Pricing tier) Standard

Location uksouth

Vault URI https://formula1-key-vault.vault.azure.net/

Resource ID /subscriptions/34cf7d7a-1340-45c5-81e4-57efa1d4b11a/resourceGroups/databricks-course-rg/providers/Microsoft.KeyVault/vaults/formula1-key-vault

Subscription ID 34cf7d7a-1340-45c5-81e4-57efa1d4b11a

Subscription Name Pay-As-You-Go

Directory ID 65947afd-b0be-4469-b3d5-2ca66af0e041

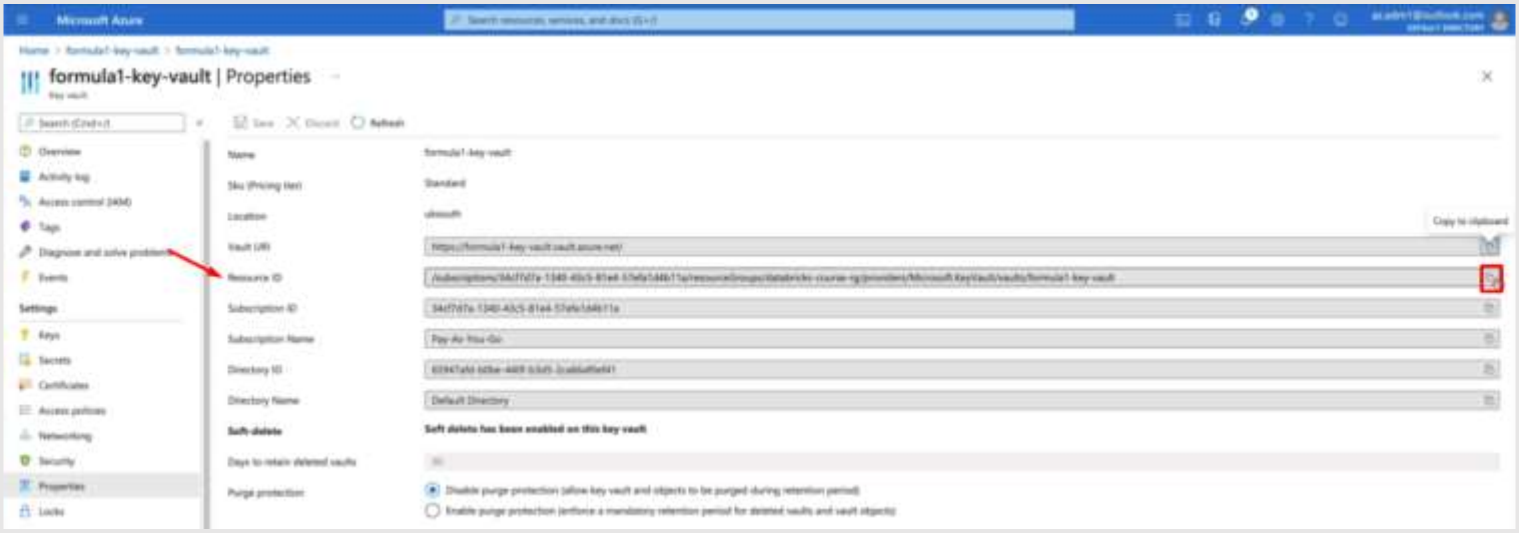
Directory Name Default Directory

Soft-delete **Soft delete has been enabled on this key vault**

Days to retain deleted vaults 90

Purge protection
● Disable purge protection (allow key vault and objects to be purged during retention period)
○ Enable purge protection (enforce a mandatory retention period for deleted vaults and vault objects)

Resource ID



Name formula1-key-vault

Sku (Pricing tier) Standard

Location uksouth

Vault URI https://formula1-key-vault.vault.azure.net/

Resource ID /subscriptions/34cf7d7a-1340-45c5-81e4-57efa1d4b11a/resourceGroups/databricks-course-rg/providers/Microsoft.KeyVault/vaults/formula1-key-vault

Subscription ID 34cf7d7a-1340-45c5-81e4-57efa1d4b11a

Subscription Name Pay-As-You-Go

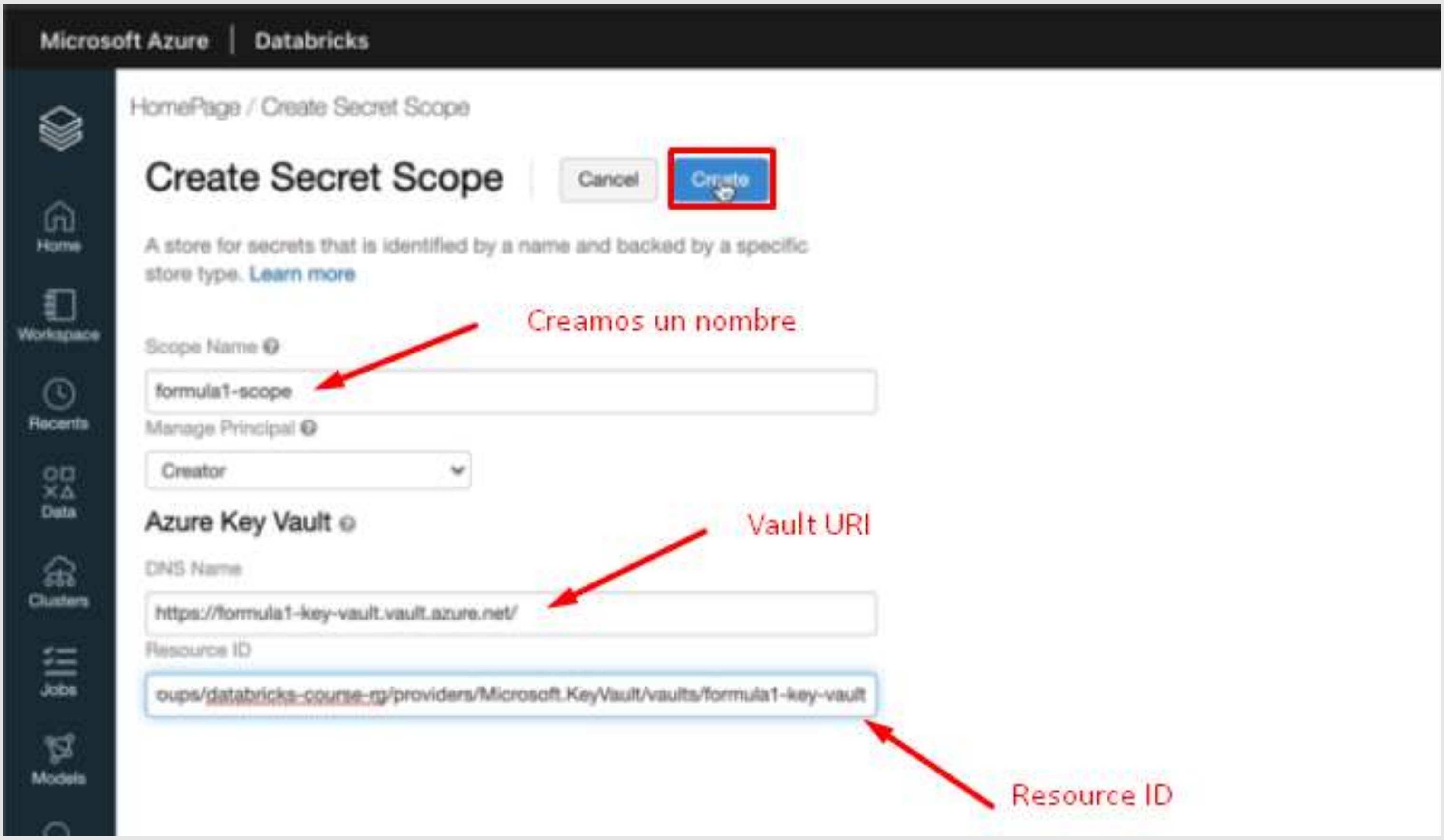
Directory ID 65947afd-b0be-4469-b3d5-2ca66af0e041

Directory Name Default Directory

Soft-delete **Soft delete has been enabled on this key vault**

Days to retain deleted vaults 90

Purge protection
● Disable purge protection (allow key vault and objects to be purged during retention period)
○ Enable purge protection (enforce a mandatory retention period for deleted vaults and vault objects)



HomePage / Create Secret Scope

Create Secret Scope

A store for secrets that is identified by a name and backed by a specific store type. Learn more

Scope Name: formula1-scope

Manage Principal: Creator

Azure Key Vault

DNS Name: https://formula1-key-vault.vault.azure.net/

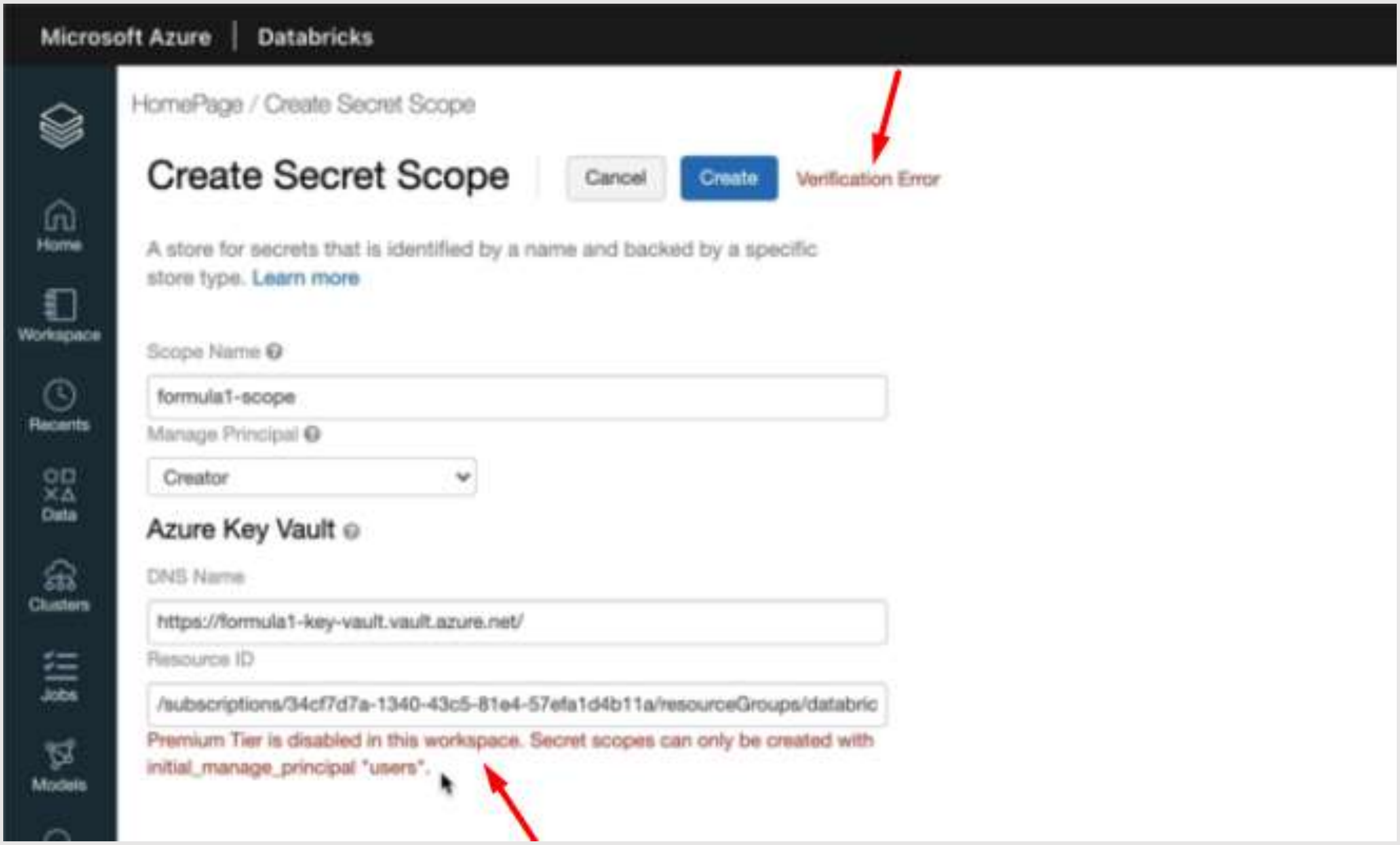
Resource ID: oups/databricks-course-rg/providers/Microsoft.KeyVault/vaults/formula1-key-vault

Creamos un nombre

Vault URI

Resource ID

Nos generará un error



HomePage / Create Secret Scope

Create Secret Scope

Verification Error

A store for secrets that is identified by a name and backed by a specific store type. Learn more

Scope Name: formula1-scope

Manage Principal: Creator

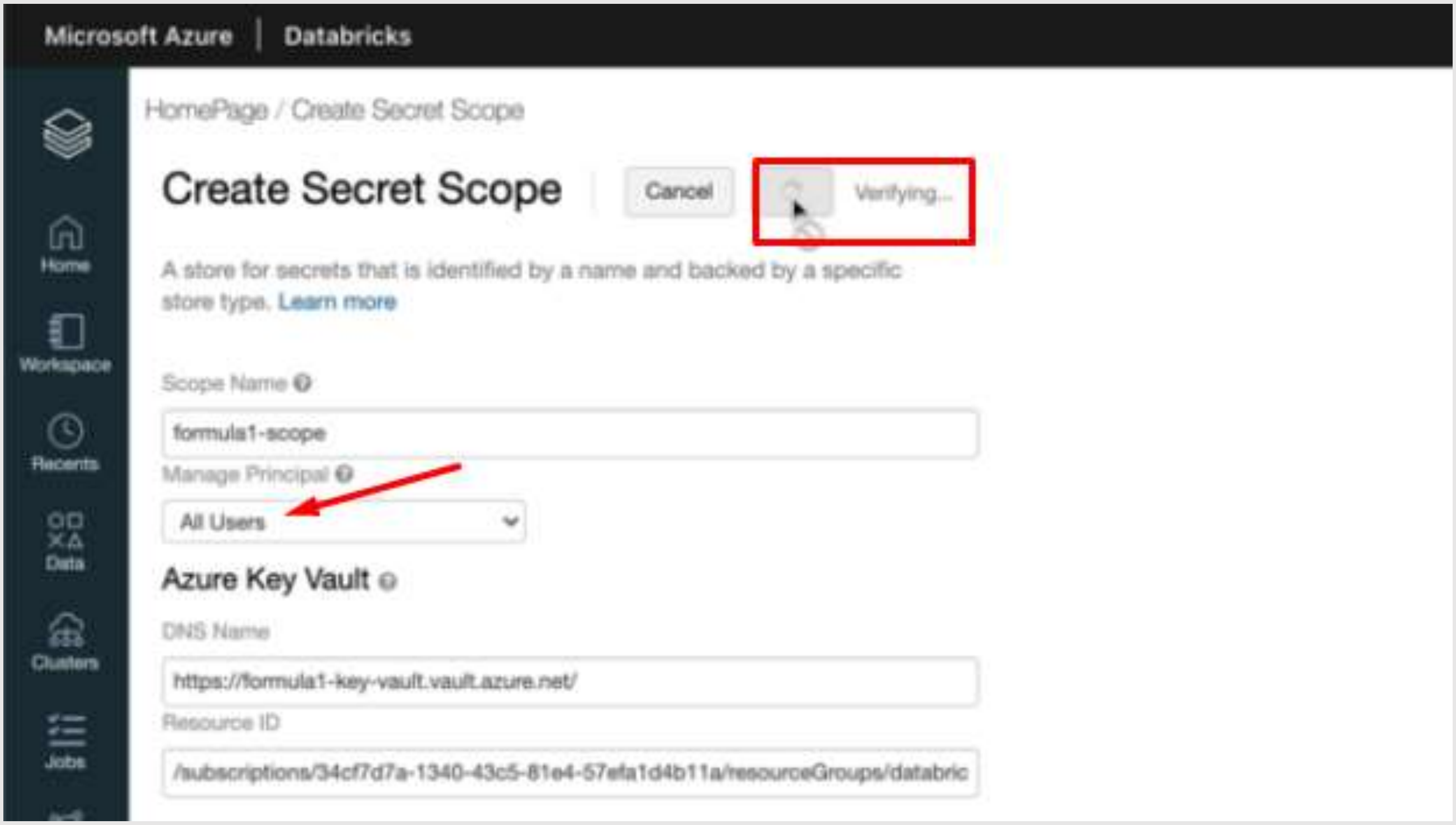
Azure Key Vault

DNS Name: https://formula1-key-vault.vault.azure.net/

Resource ID: /subscriptions/34cf7d7a-1340-43c5-81e4-57efa1d4b11a/resourceGroups/databric

Premium Tier is disabled in this workspace. Secret scopes can only be created with initial_manage_principal "users".

Modificamos el “Manage Principal” a “All Users”



Create Secret Scope

A store for secrets that is identified by a name and backed by a specific store type. Learn more

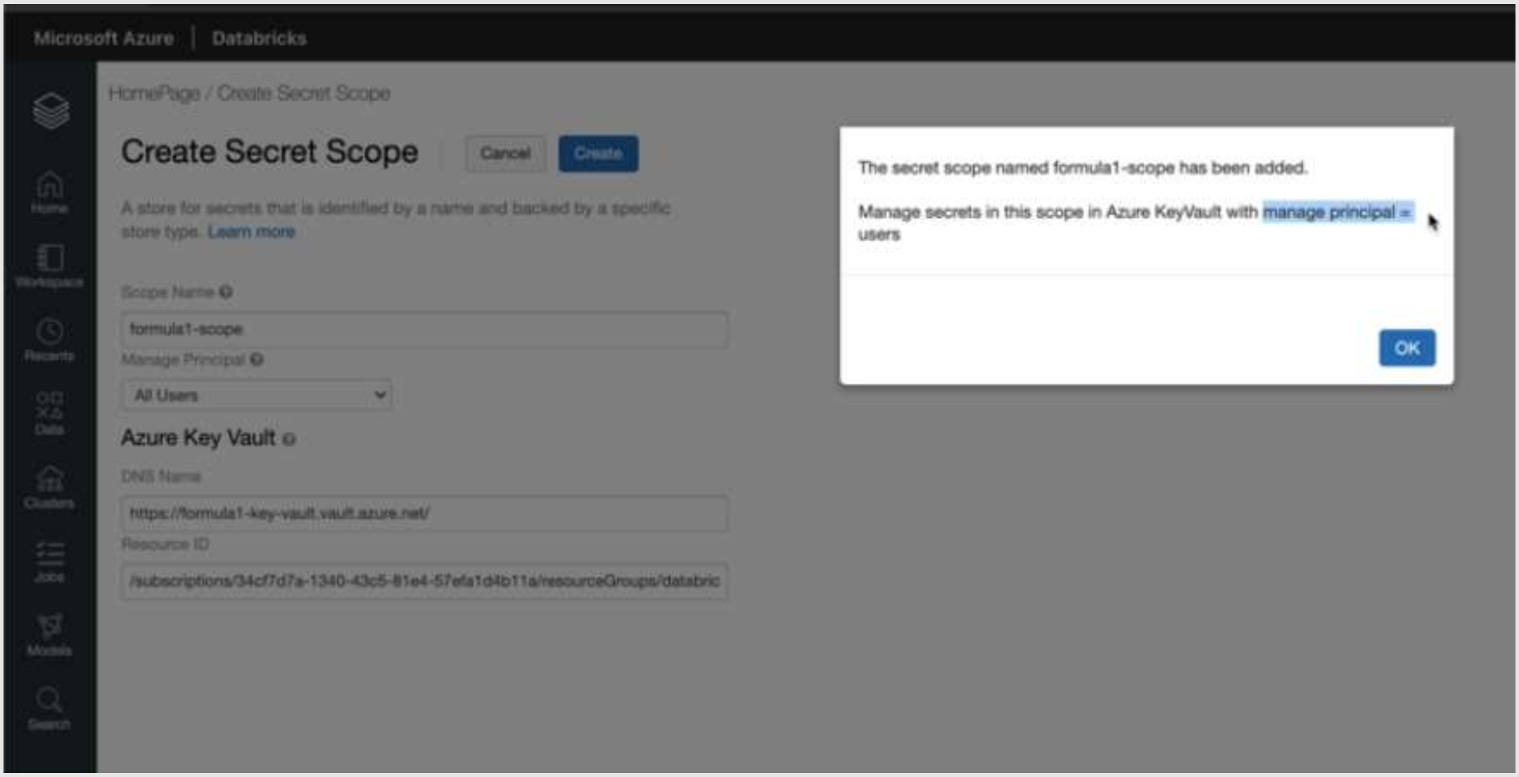
Scope Name: formula1-scope

Manage Principal: All Users

Azure Key Vault

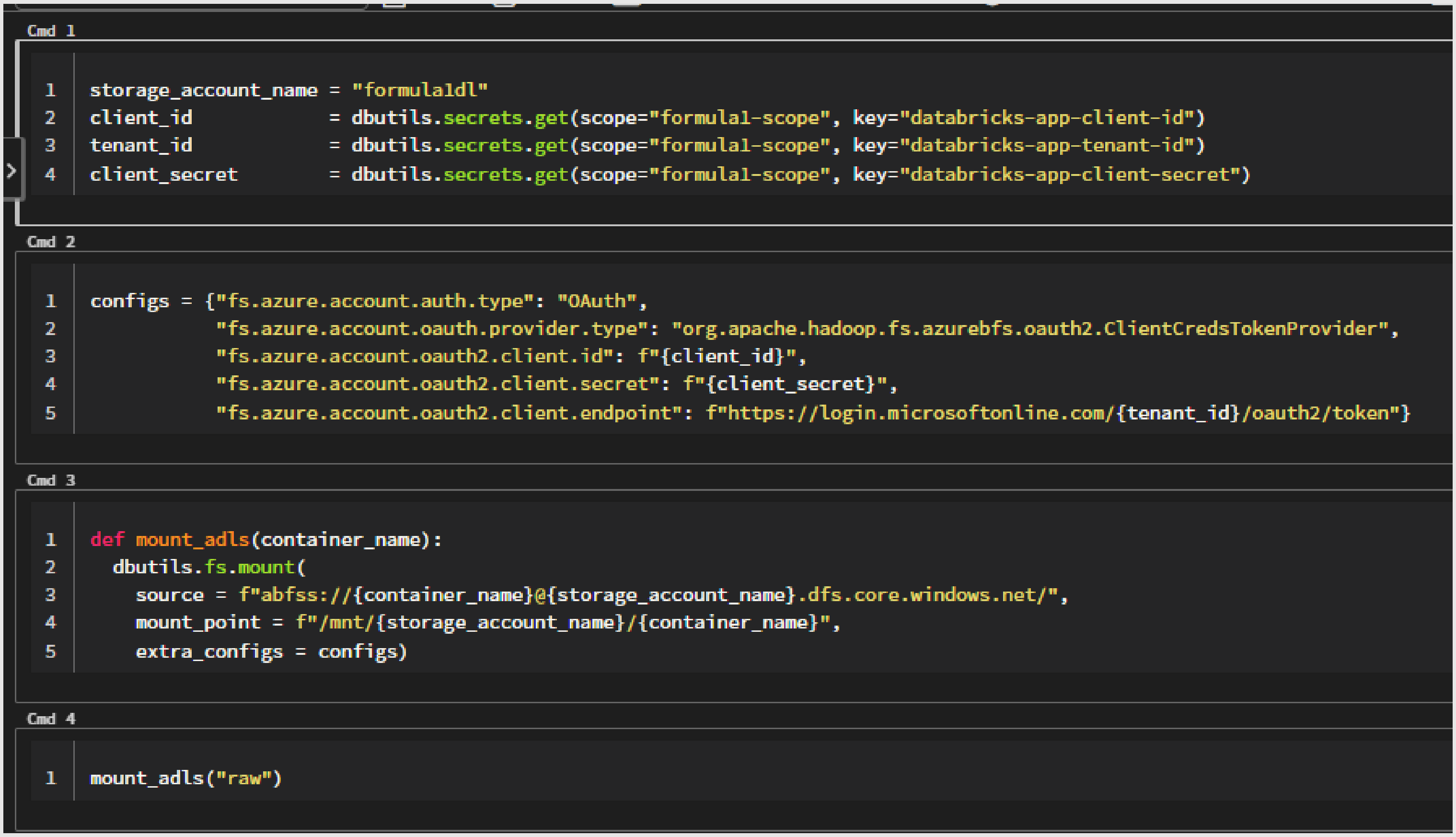
DNS Name: https://formula1-key-vault.vault.azure.net/

Resource ID: /subscriptions/34cf7d7a-1340-43c5-81e4-57efa1d4b11a/resourceGroups/databric



The secret scope named formula1-scope has been added.

Manage secrets in this scope in Azure KeyVault with manage principal = users



Cmd 1

```
storage_account_name = "formula1dl"
client_id            = dbutils.secrets.get(scope="formula1-scope", key="databricks-app-client-id")
tenant_id            = dbutils.secrets.get(scope="formula1-scope", key="databricks-app-tenant-id")
client_secret        = dbutils.secrets.get(scope="formula1-scope", key="databricks-app-client-secret")
```

Cmd 2

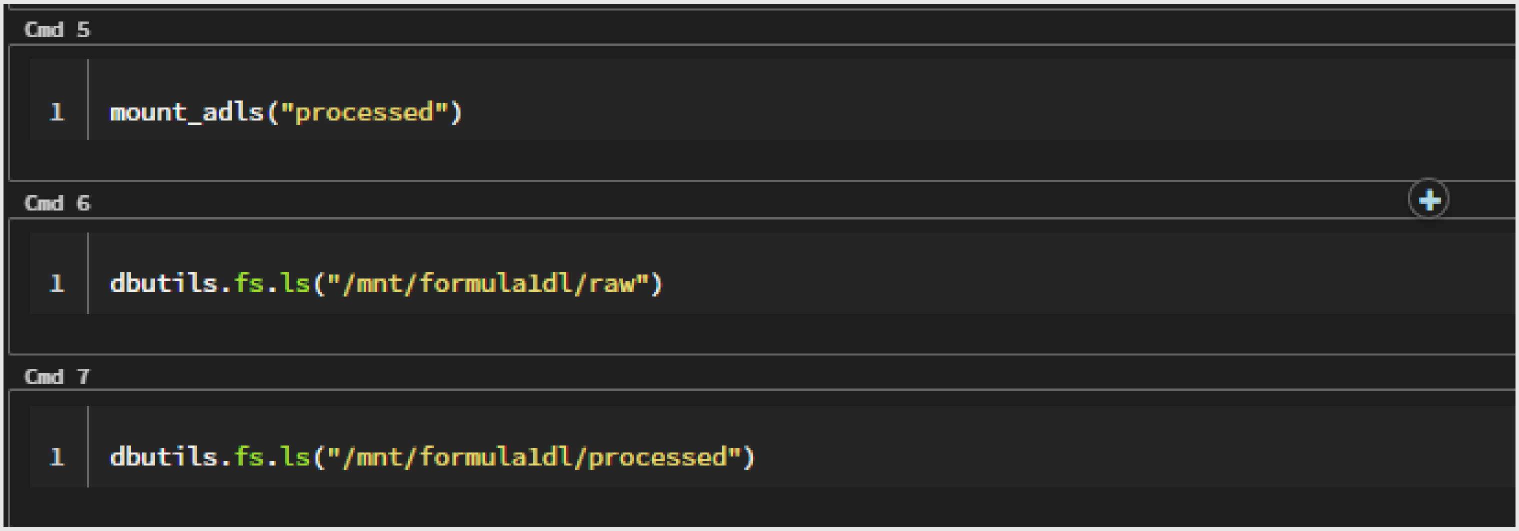
```
configs = {"fs.azure.account.auth.type": "OAuth",
           "fs.azure.account.oauth.provider.type": "org.apache.hadoop.fs.azurebfs.oauth2.ClientCredsTokenProvider",
           "fs.azure.account.oauth2.client.id": f"{client_id}",
           "fs.azure.account.oauth2.client.secret": f"{client_secret}",
           "fs.azure.account.oauth2.client.endpoint": f"https://login.microsoftonline.com/{tenant_id}/oauth2/token"}
```

Cmd 3

```
def mount_adls(container_name):
  dbutils.fs.mount(
    source = f"abfss://{container_name}@{storage_account_name}.dfs.core.windows.net/",
    mount_point = f"/mnt/{storage_account_name}/{container_name}",
    extra_configs = configs)
```

Cmd 4

```
mount_adls("raw")
```



Cmd 5

```
mount_adls("processed")
```

Cmd 6

```
dbutils.fs.ls("/mnt/formula1dl/raw")
```

Cmd 7

```
dbutils.fs.ls("/mnt/formula1dl/processed")
```