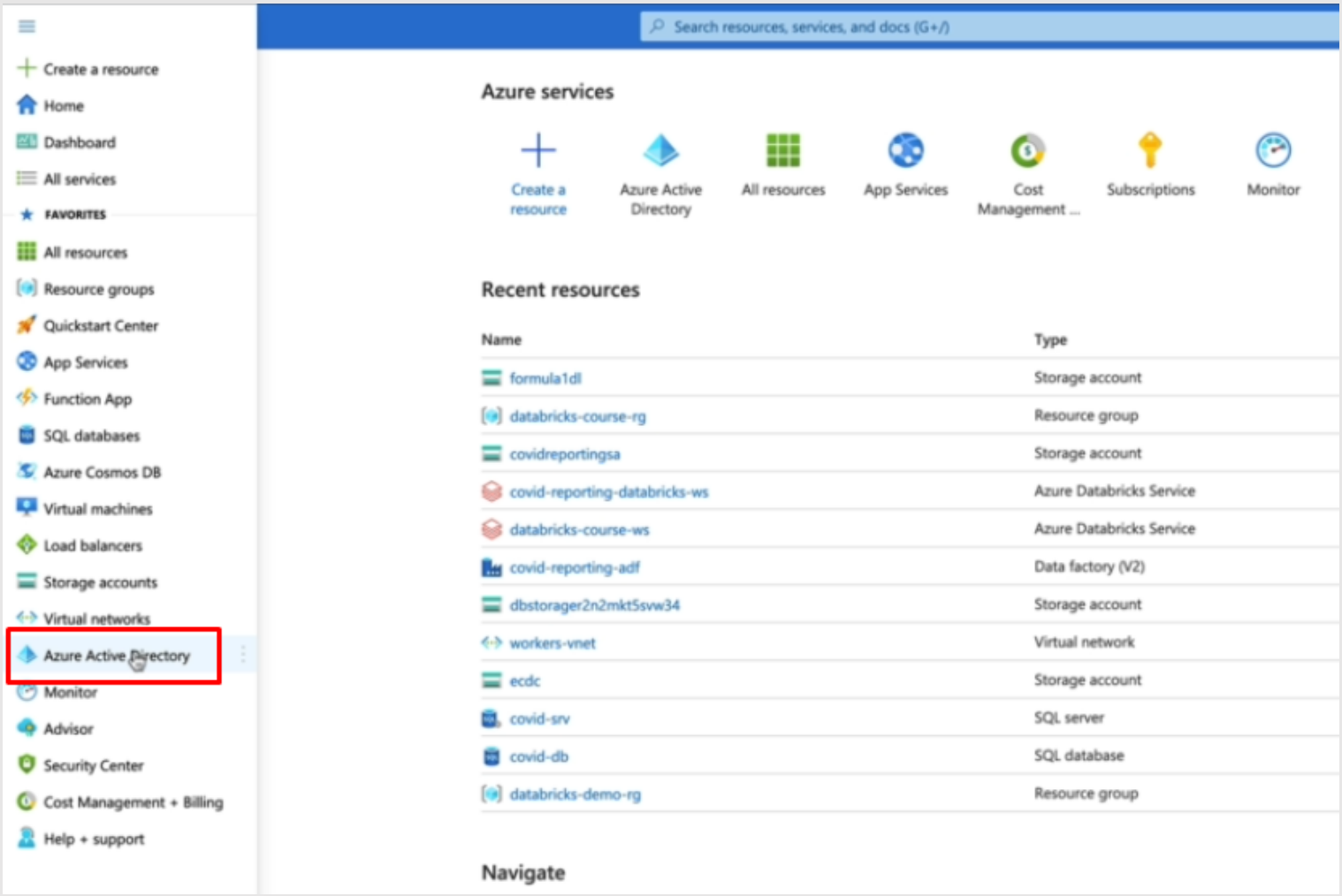
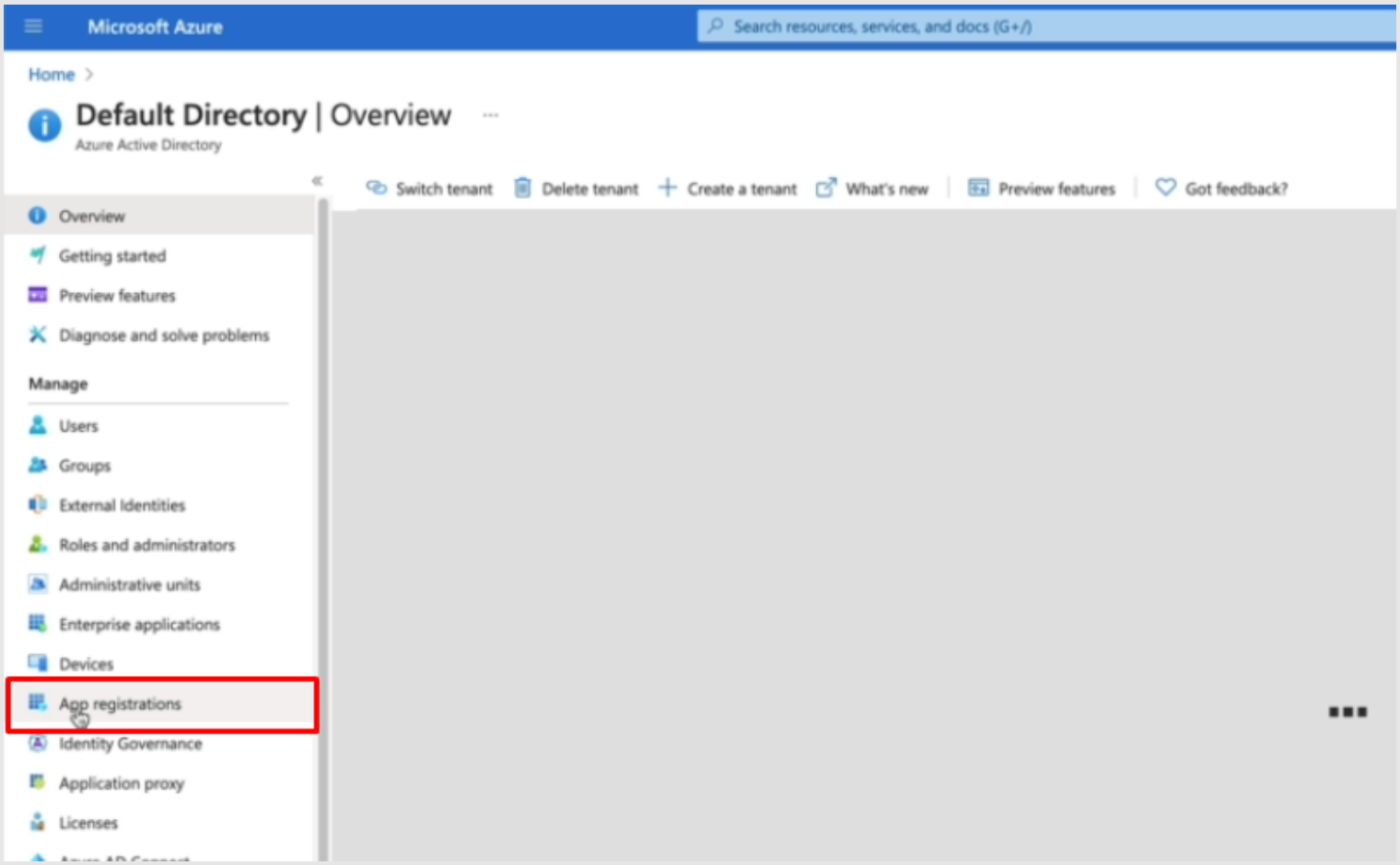
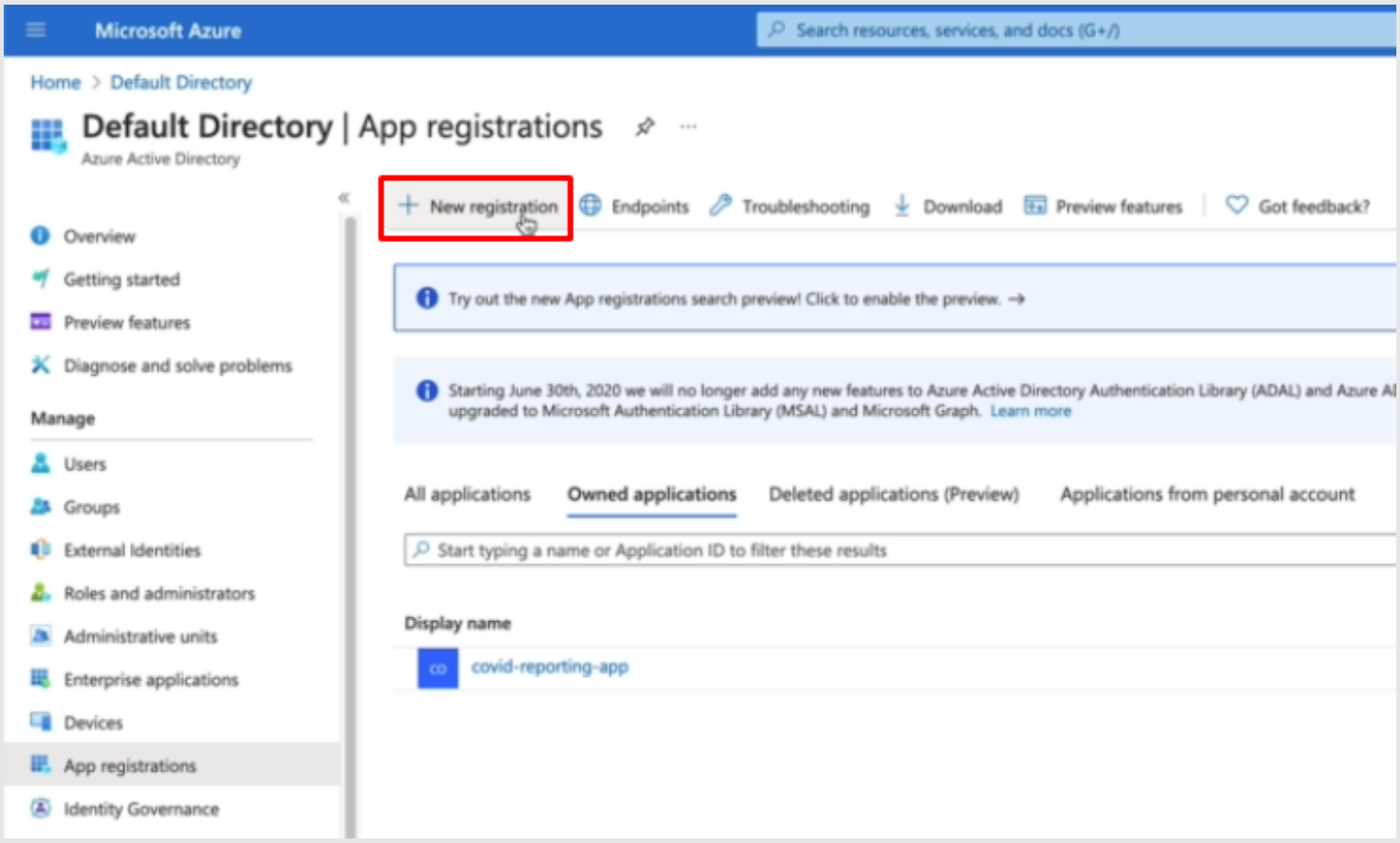
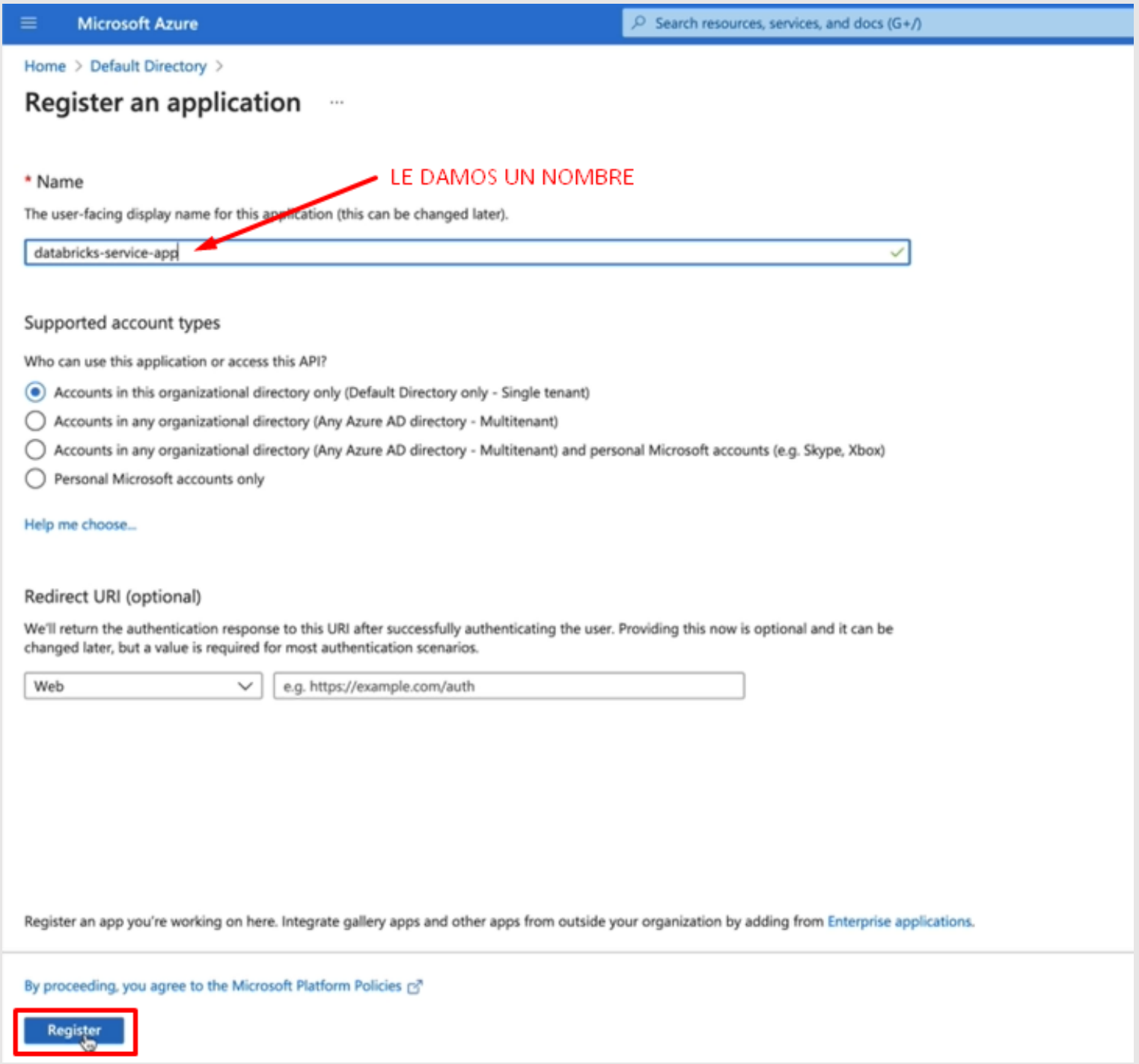
Azure Databricks Integracion con ADLS Usar Active Directory – Utilizar Service Principal / Databricks Scoped Secret

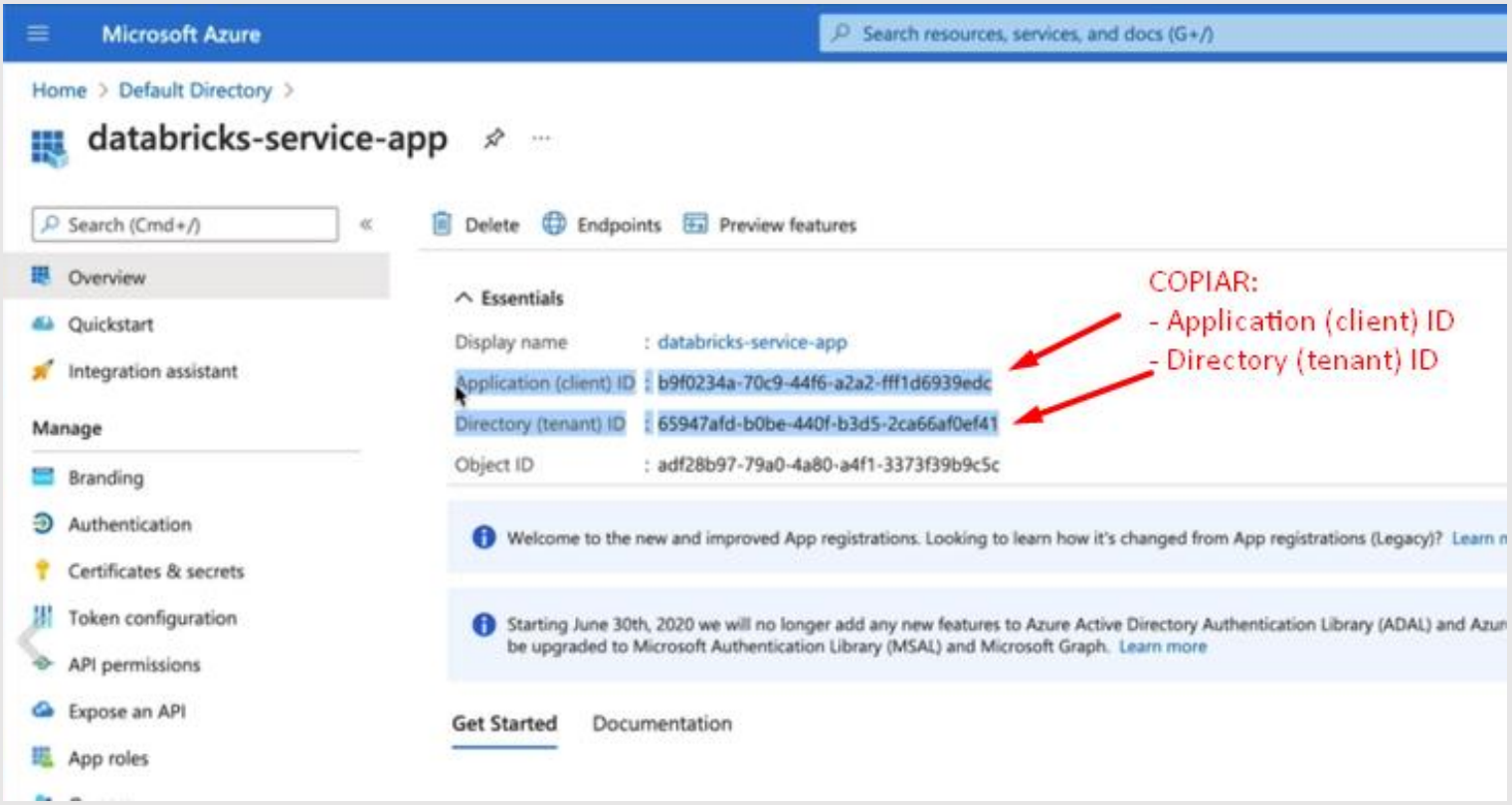
Creando un Azure Service Principal

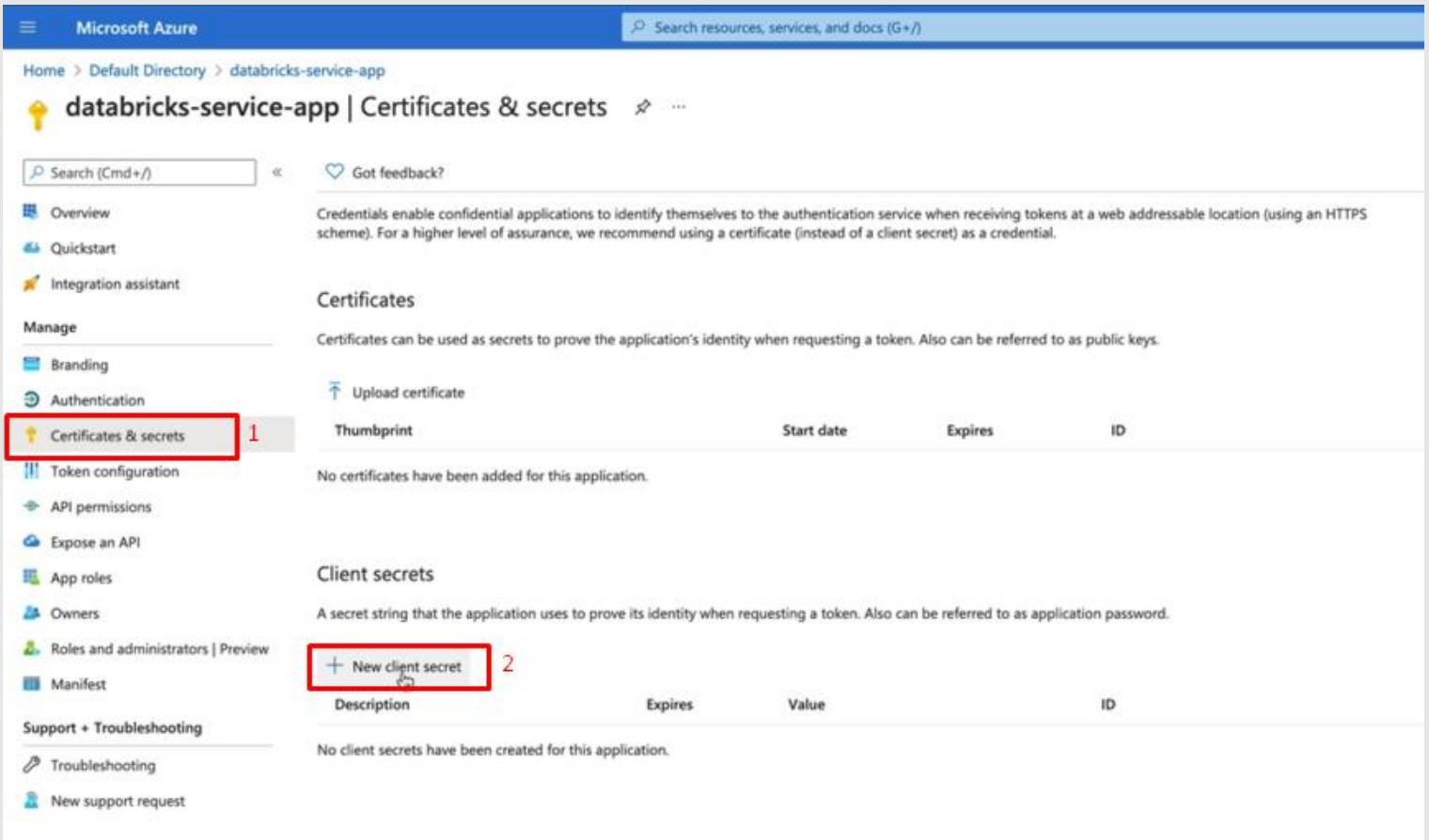


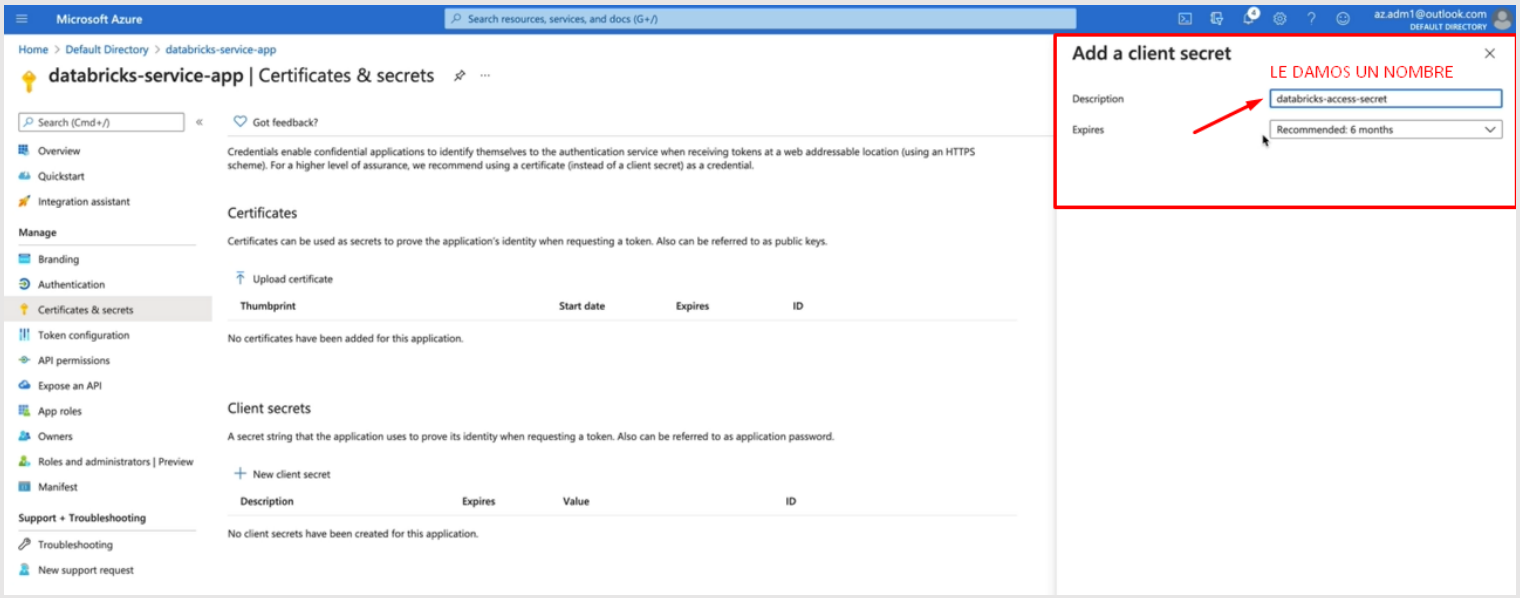

Microsoft Azure
Search resources, services, and docs (G+/)
Home
Default Directory | Overview
Azure Active Directory
Switch tenant
Delete tenant
Create a tenant
What's new
Preview features
Got feedback?
Overview
Getting started
Preview features
Diagnose and solve problems
Manage
Users
Groups
External Identities
Roles and administrators
Administrative units
Enterprise applications
Devices
App registrations
Identity Governance
Application proxy
Licenses

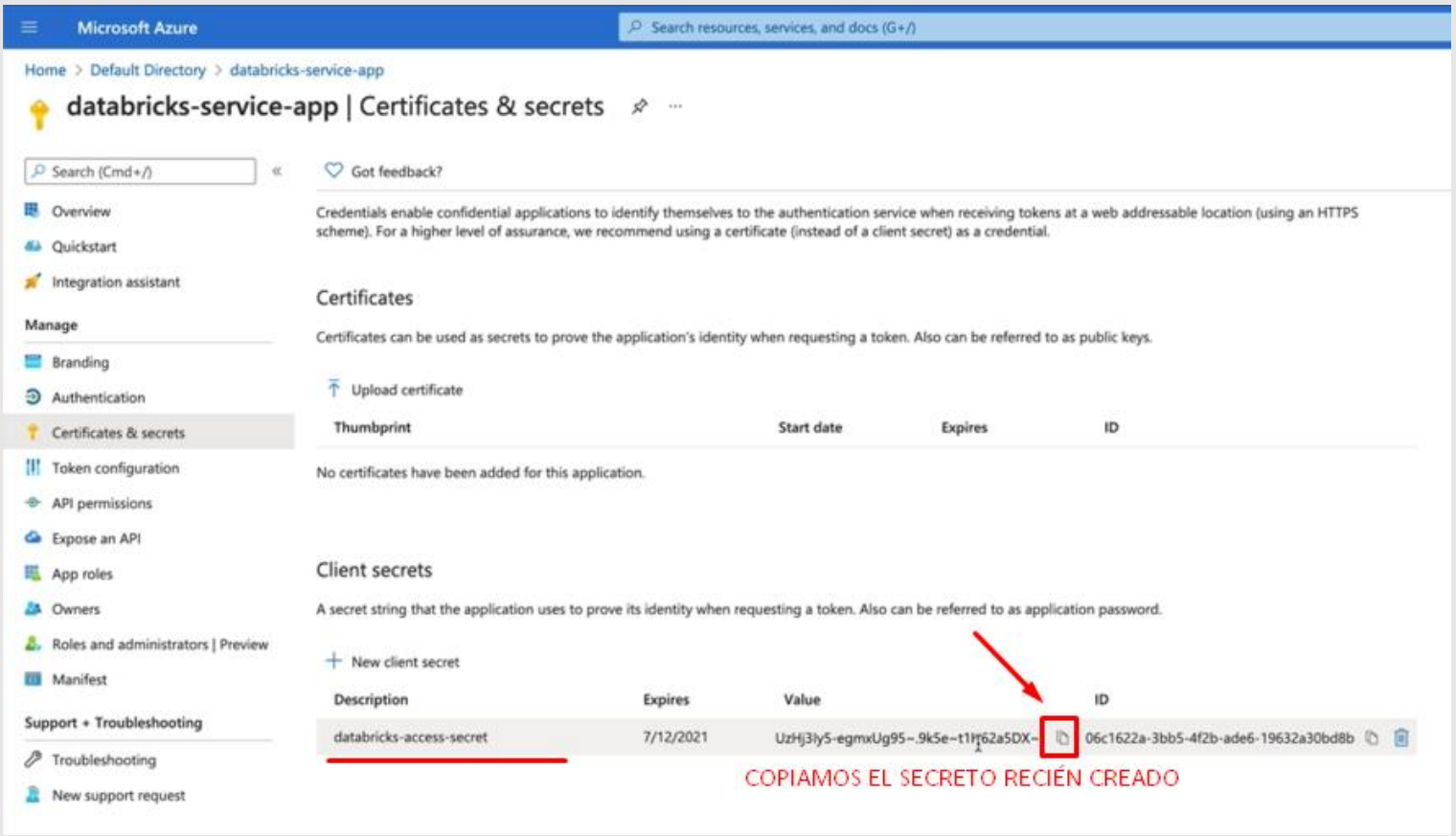

Microsoft Azure
Search resources, services, and docs (G+/)
Home
Default Directory
Default Directory | App registrations
Azure Active Directory
New registration
Endpoints
Troubleshooting
Download
Preview features
Got feedback?
Overview
Getting started
Preview features
Diagnose and solve problems
Manage
Users
Groups
External Identities
Roles and administrators
Administrative units
Enterprise applications
Devices
App registrations
Identity Governance
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
upgraded to Microsoft Authentication Library (MSAL) and Microsoft Graph. Learn more
All applications
Owned applications
Deleted applications (Preview)
Applications from personal account
Start typing a name or Application ID to filter these results
Display name
covid-reporting-app

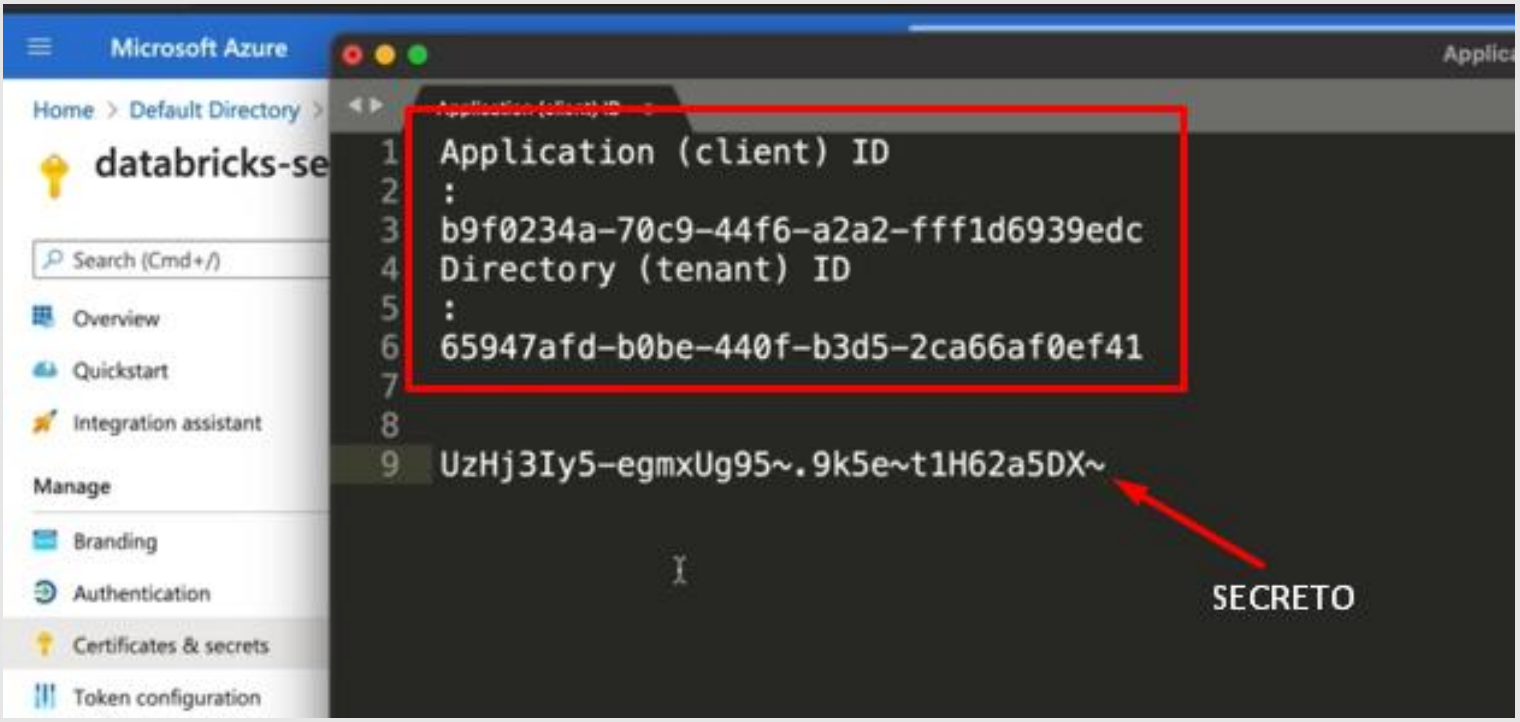

Microsoft Azure
Search resources, services, and docs (G+/)
Home > Default Directory >
Register an application
* Name
LE DAMOS UN NOMBRE
The user-facing display name for this application (this can be changed later).
databricks-service-app
Supported account types
Who can use this application or access 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
Register

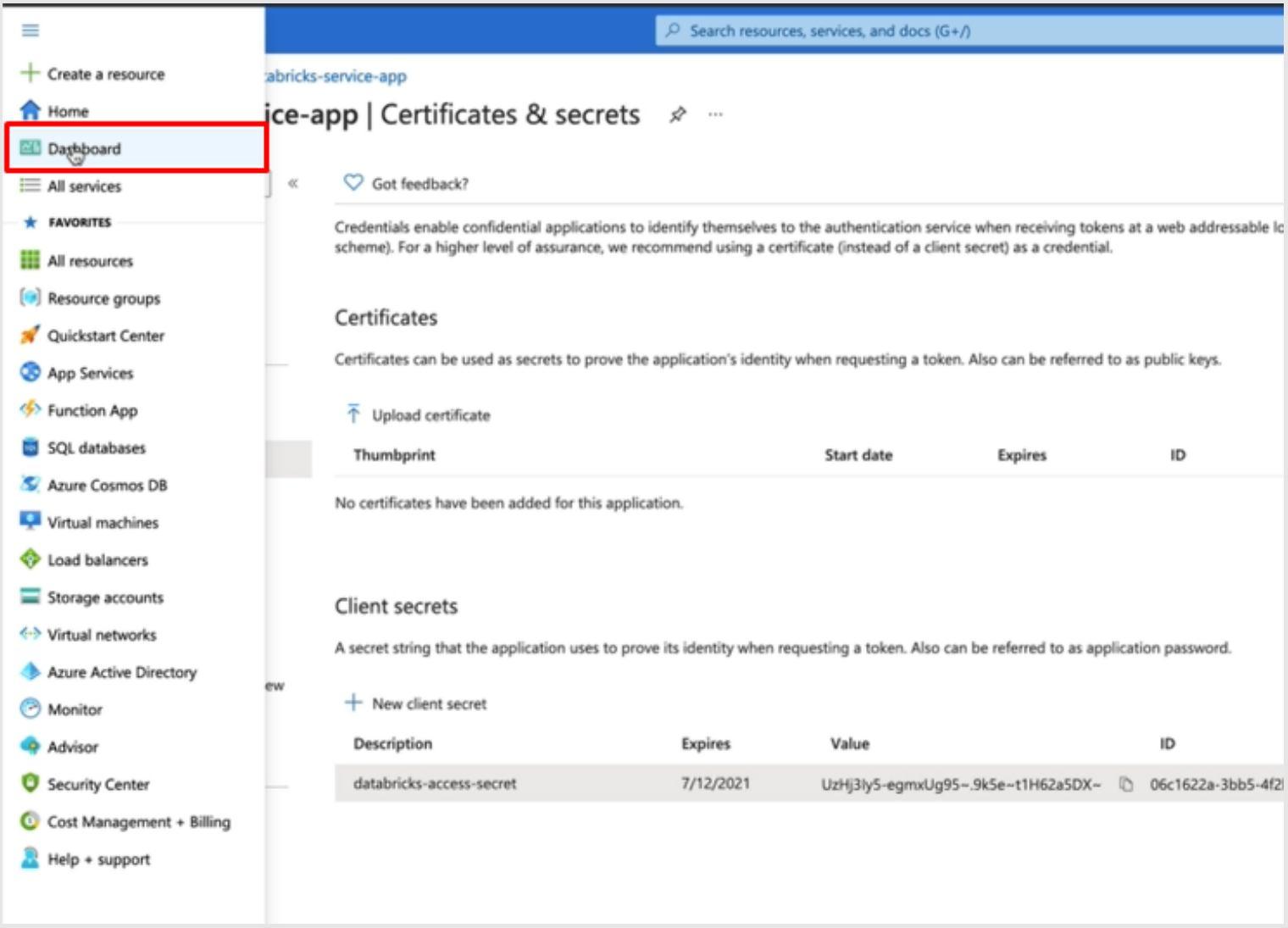

Microsoft Azure
Search resources, services, and docs (G+/)
Home > Default Directory >
databricks-service-app
Search (Cmd+/)
Delete
Endpoints
Preview features
Overview
Quickstart
Integration assistant
Manage
Branding
Authentication
Certificates & secrets
Token configuration
API permissions
Expose an API
App roles
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
COPIAR:
- Application (client) ID
- Directory (tenant) ID
Welcome to the new and improved App registrations. Looking to learn how it's changed from App registrations (Legacy)?
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
Get Started
Documentation

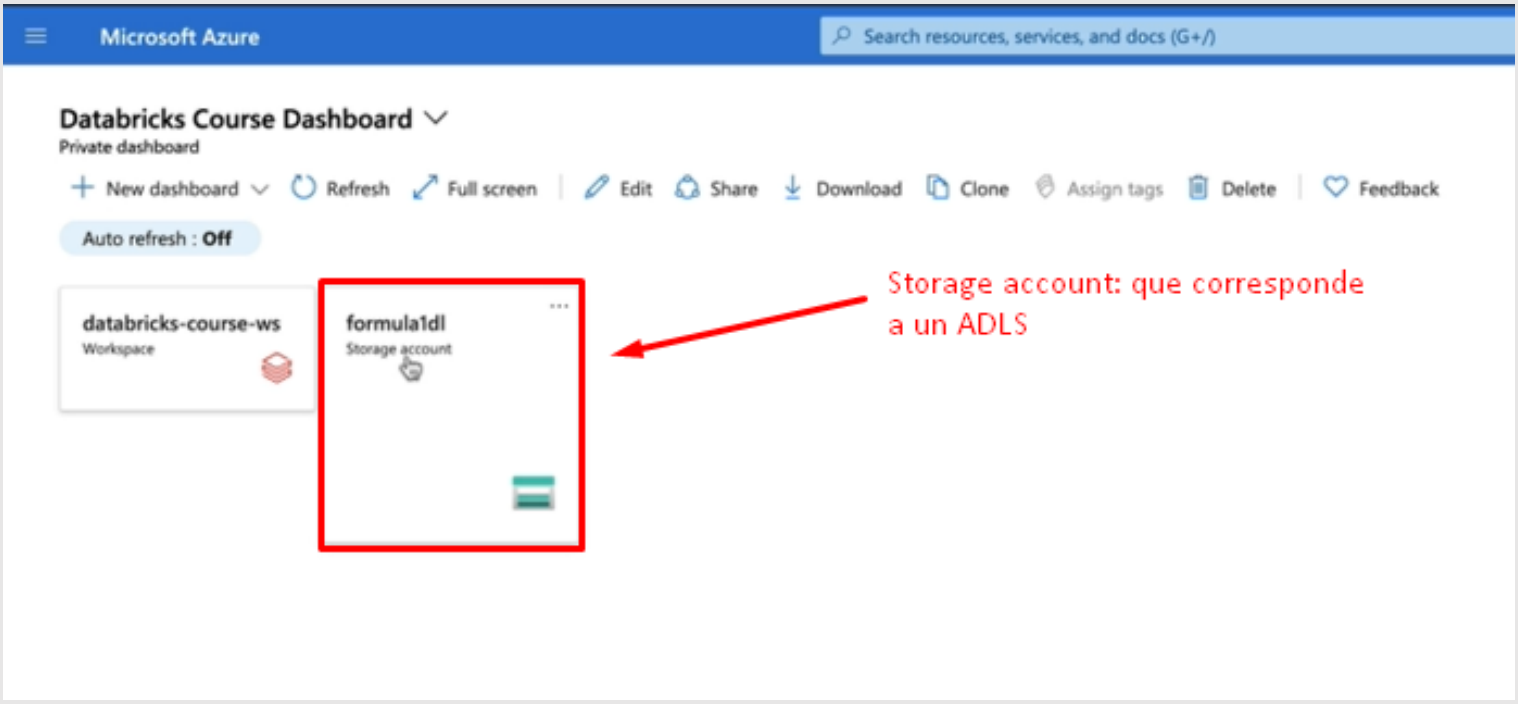

Microsoft Azure
Search resources, services, and docs (G+/)
Home > Default Directory > databricks-service-app
databricks-service-app | Certificates & secrets
Search (Cmd+/)
Got feedback?
Overview
Quickstart
Integration assistant
Manage
Branding
Authentication
Certificates & secrets
1
Token configuration
API permissions
Expose an API
App roles
Owners
Roles and administrators | Preview
Manifest
Support + Troubleshooting
Troubleshooting
New support request
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.
Upload certificate
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.
New client secret
2
Description
Expires
Value
ID
No client secrets have been created for this application.

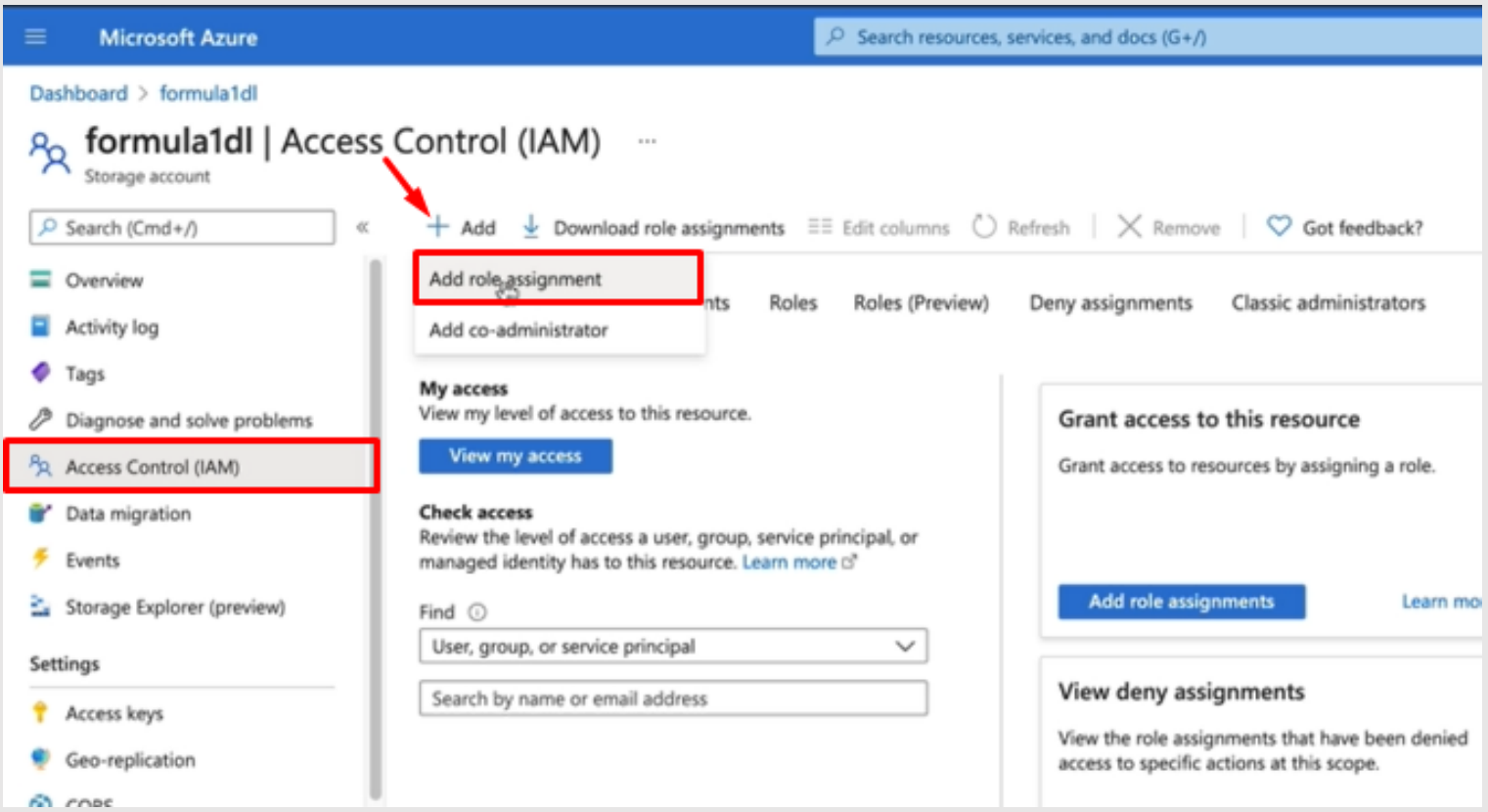

Microsoft Azure
Search resources, services, and docs (G+/)
az.adm1@outlook.com
DEFAULT DIRECTORY
Home > Default Directory > databricks-service-app
databricks-service-app | Certificates & secrets
Search (Cmd+/)
Got feedback?
Overview
Quickstart
Integration assistant
Manage
Branding
Authentication
Certificates & secrets
Token configuration
API permissions
Expose an API
App roles
Owners
Roles and administrators | Preview
Manifest
Support + Troubleshooting
Troubleshooting
New support request
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.
Upload certificate
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.
New client secret
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

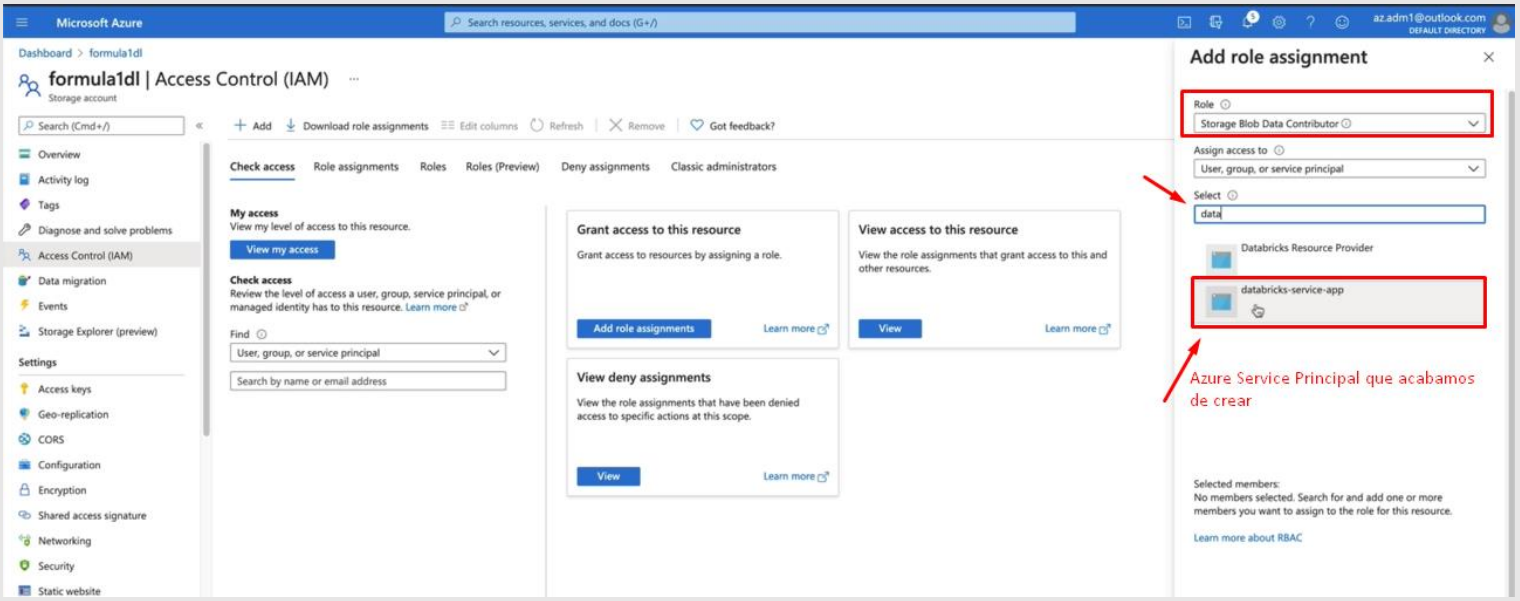

Microsoft Azure
Search resources, services, and docs (G+/)
Home > Default Directory > databricks-service-app
databricks-service-app | Certificates & secrets
Search (Cmd+/)
Got feedback?
Overview
Quickstart
Integration assistant
Manage
Branding
Authentication
Certificates & secrets
Token configuration
API permissions
Expose an API
App roles
Owners
Roles and administrators | Preview
Manifest
Support + Troubleshooting
Troubleshooting
New support request
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.
Upload certificate
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.
New client secret
Description
Expires
Value
ID
databricks-access-secret
7/12/2021
UzHj3Iy5-egmxUg95~.9k5e~t1Hr62a5DX~
06c1622a-3bb5-4f2b-ade6-19632a30bd8b
COPIAMOS EL SECRETO RECIÉN CREADO

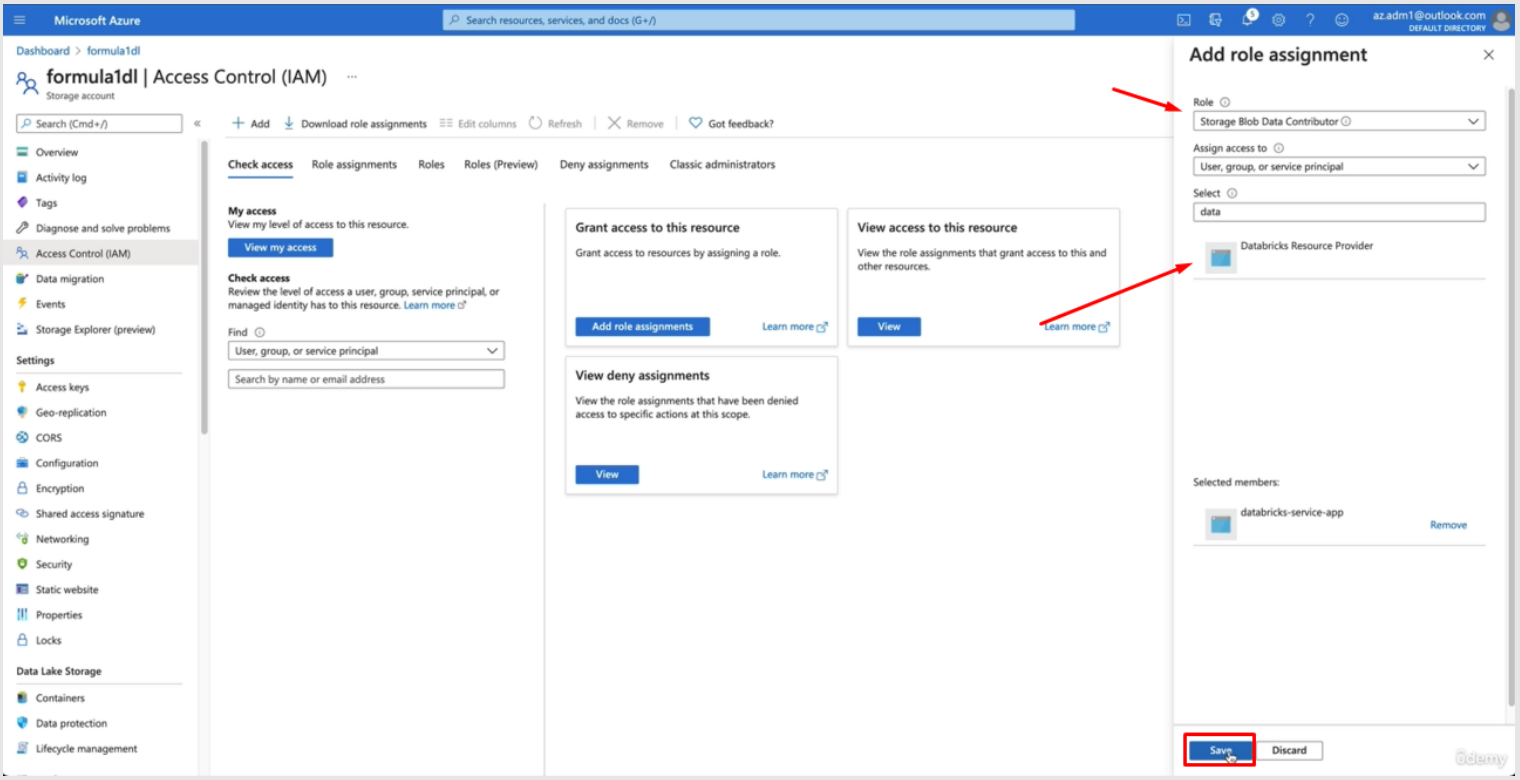

Microsoft Azure
Home > Default Directory >
databricks-se
Search (Cmd+/)
Overview
Quickstart
Integration assistant
Manage
Branding
Authentication
Certificates & secrets
Token configuration
Application (client) ID
:
b9f0234a-70c9-44f6-a2a2-fff1d6939edc
Directory (tenant) ID
:
65947afd-b0be-440f-b3d5-2ca66af0ef41
UzHj3Iy5-egmxUg95~.9k5e~t1H62a5DX~
SECRETO


Search resources, services, and docs (G+/)
Create a resource
Home
Dashboard
All services
FAVORITES
All resources
Resource groups
Quickstart Center
App Services
Function App
SQL databases
Azure Cosmos DB
Virtual machines
Load balancers
Storage accounts
Virtual networks
Azure Active Directory
Monitor
Advisor
Security Center
Cost Management + Billing
Help + support
abricks-service-app
ice-app | Certificates & secrets
Got feedback?
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.
Upload certificate
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.
New client secret
Description
Expires
Value
ID
databricks-access-secret
7/12/2021
UzHj3Iy5-egmxUg95~.9k5e~t1H62a5DX~
06c1622a-3bb5-4f2


Microsoft Azure
Search resources, services, and docs (G+/)
Databricks Course Dashboard
Private dashboard
New dashboard
Refresh
Full screen
Edit
Share
Download
Clone
Assign tags
Delete
Feedback
Auto refresh : Off
databricks-course-ws
Workspace
formula1dl
Storage account
Storage account: que corresponde a un ADLS

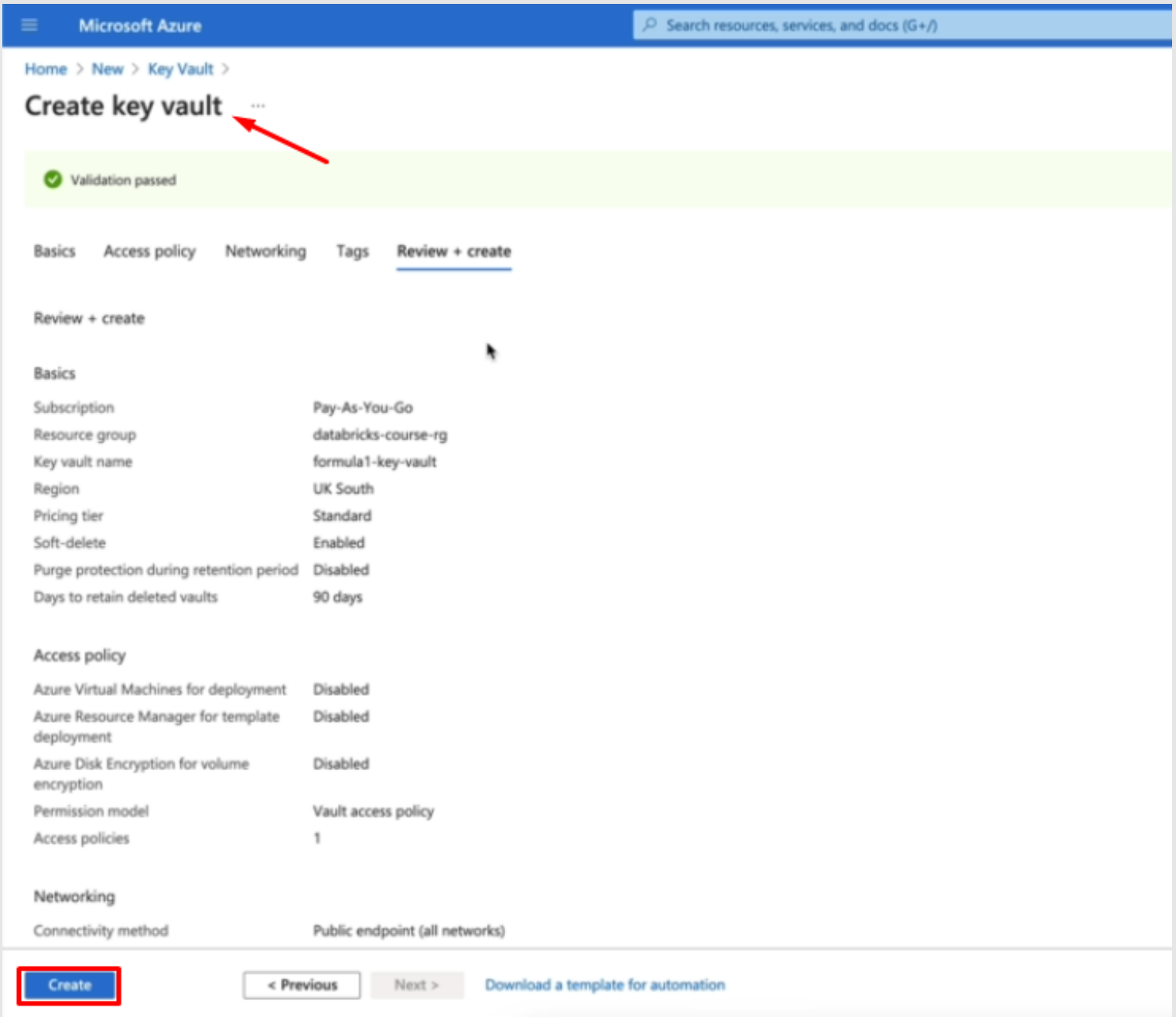

Microsoft Azure
Search resources, services, and docs (G+/)
Dashboard > formula1dl
formula1dl | Access Control (IAM)
Storage account
Search (Cmd+/)
Add
Download role assignments
Edit columns
Refresh
Remove
Got feedback?
Overview
Activity log
Tags
Diagnose and solve problems
Access Control (IAM)
Data migration
Events
Storage Explorer (preview)
Settings
Access keys
Geo-replication
Add role assignment
Add co-administrator
Roles
Roles (Preview)
Deny assignments
Classic administrators
My access
View my level of access to this resource.
View my access
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.
Add role assignments
View deny assignments
View the role assignments that have been denied access to specific actions at this scope.


Microsoft Azure
Search resources, services, and docs (G+/)
az.adm1@outlook.com
DEFAULT DIRECTORY
Dashboard > formula1dl
formula1dl | Access Control (IAM)
Storage account
Search (Cmd+/)
Add
Download role assignments
Edit columns
Refresh
Remove
Got feedback?
Overview
Activity log
Tags
Diagnose and solve problems
Access Control (IAM)
Data migration
Events
Storage Explorer (preview)
Settings
Access keys
Geo-replication
CORS
Configuration
Encryption
Shared access signature
Networking
Security
Static website
Check access
Role assignments
Roles
Roles (Preview)
Deny assignments
Classic administrators
My access
View my level of access to this resource.
View my access
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.
Add role assignments
Learn more
View access to this resource
View the role assignments that grant access to this and other resources.
View
Learn more
View deny assignments
View the role assignments that have been denied access to specific actions at this scope.
View
Learn more
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

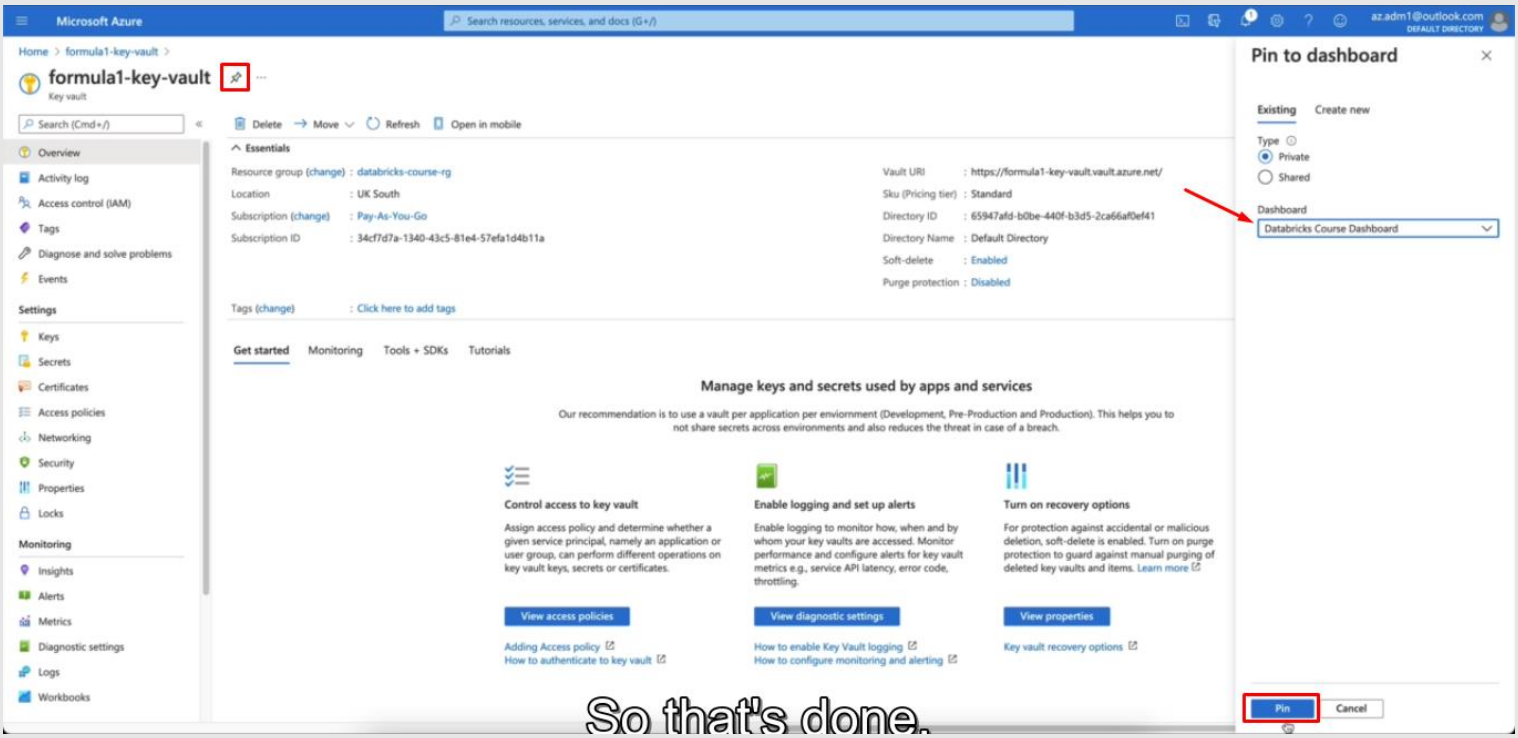

Microsoft Azure
Search resources, services, and docs (G+/)
az.adm1@outlook.com
DEFAULT DIRECTORY
Dashboard > formula1dl
formula1dl | Access Control (IAM)
Storage account
Search (Cmd+/)
Add
Download role assignments
Edit columns
Refresh
Remove
Got feedback?
Overview
Activity log
Tags
Diagnose and solve problems
Access Control (IAM)
Data migration
Events
Storage Explorer (preview)
Settings
Access keys
Geo-replication
CORS
Configuration
Encryption
Shared access signature
Networking
Security
Static website
Properties
Locks
Data Lake Storage
Containers
Data protection
Lifecycle management
Check access
Role assignments
Roles
Roles (Preview)
Deny assignments
Classic administrators
My access
View my level of access to this resource.
View my access
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.
Add role assignments
Learn more
View access to this resource
View the role assignments that grant access to this and other resources.
View
Learn more
View deny assignments
View the role assignments that have been denied access to specific actions at this scope.
View
Learn more
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
Save
Discard

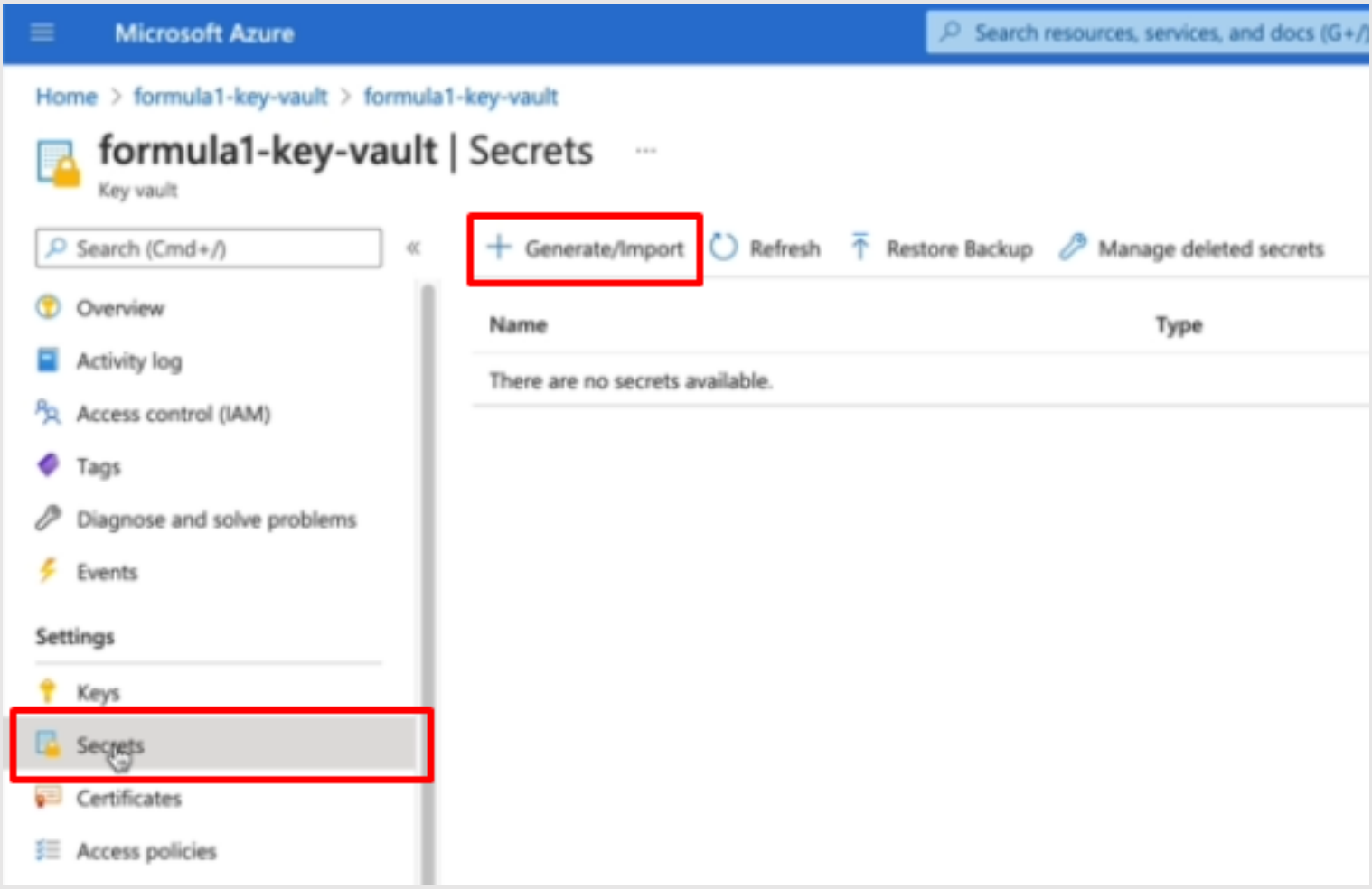
Mounting Azure Data Lake Storage Gen2

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

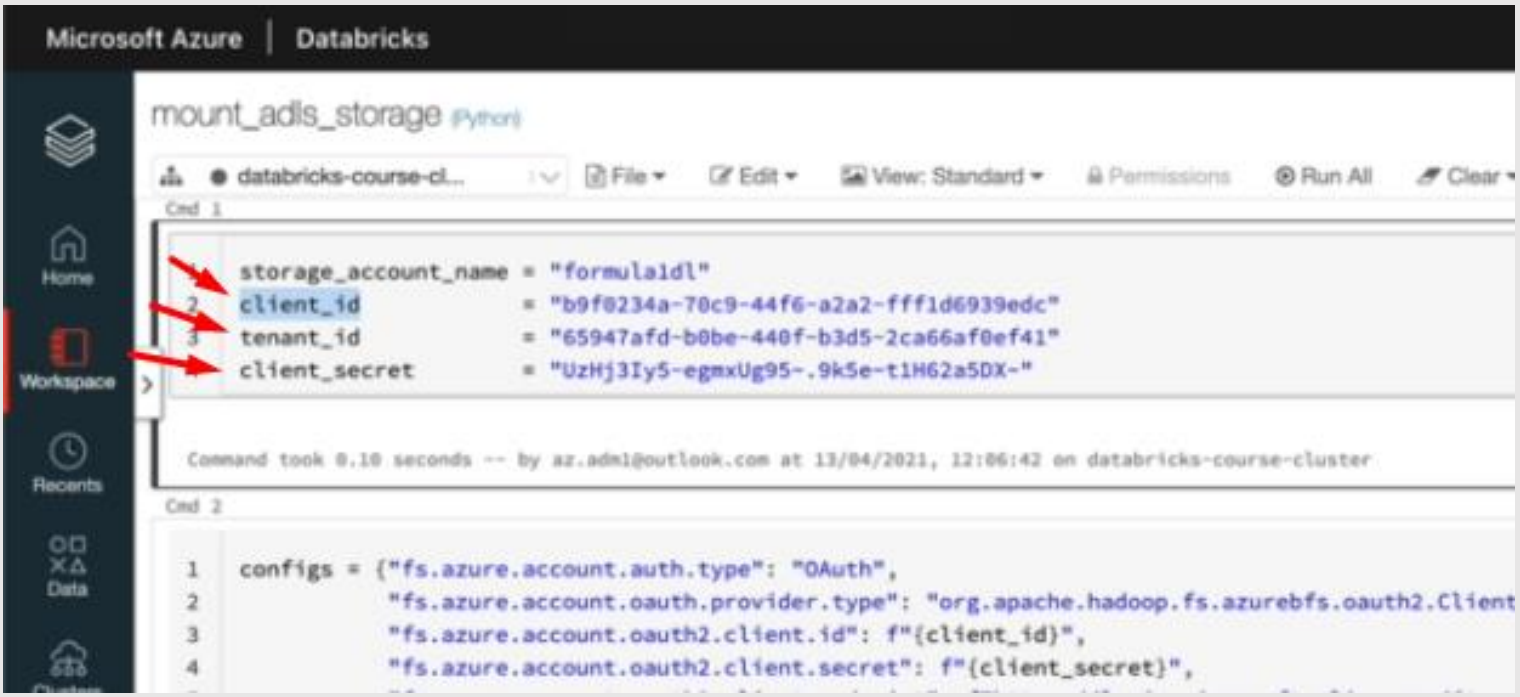


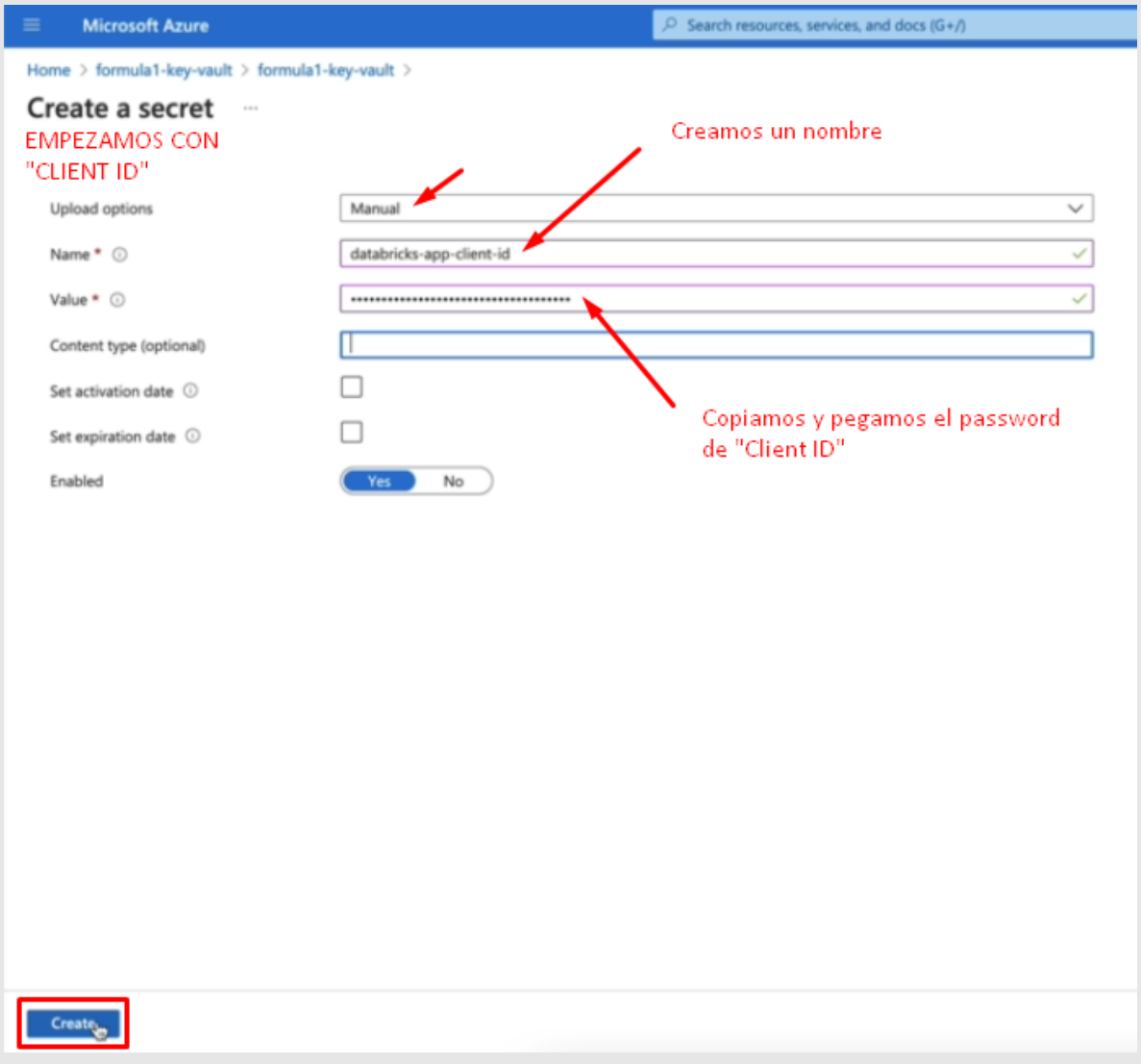
Así agregamos el servicio al dashboard

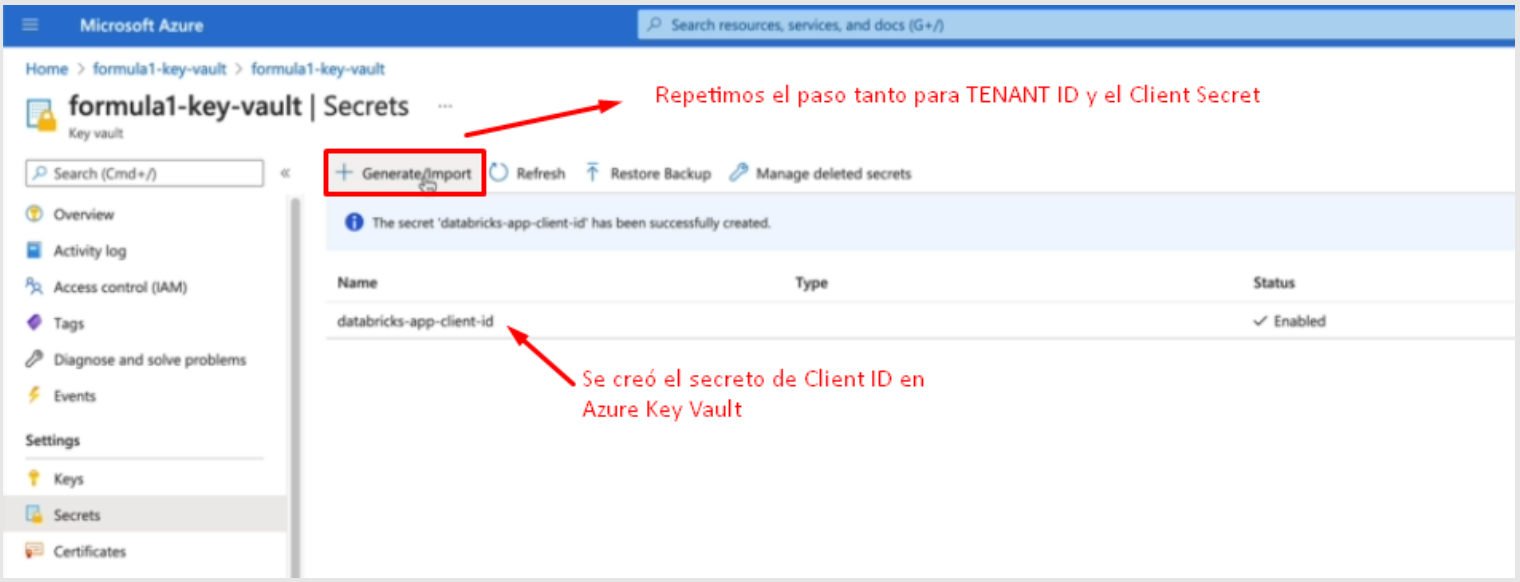




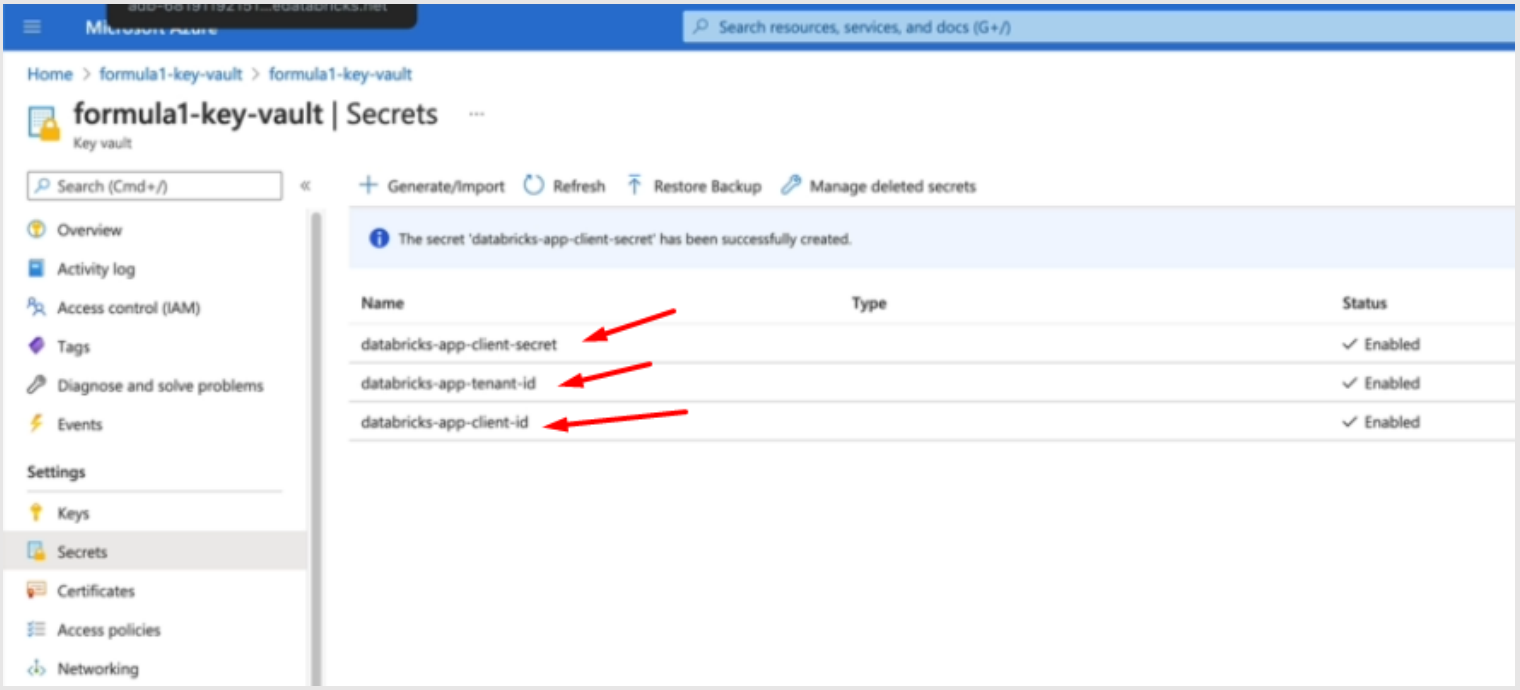
Tenemos que generar 3 secretos en Azure Key Vault



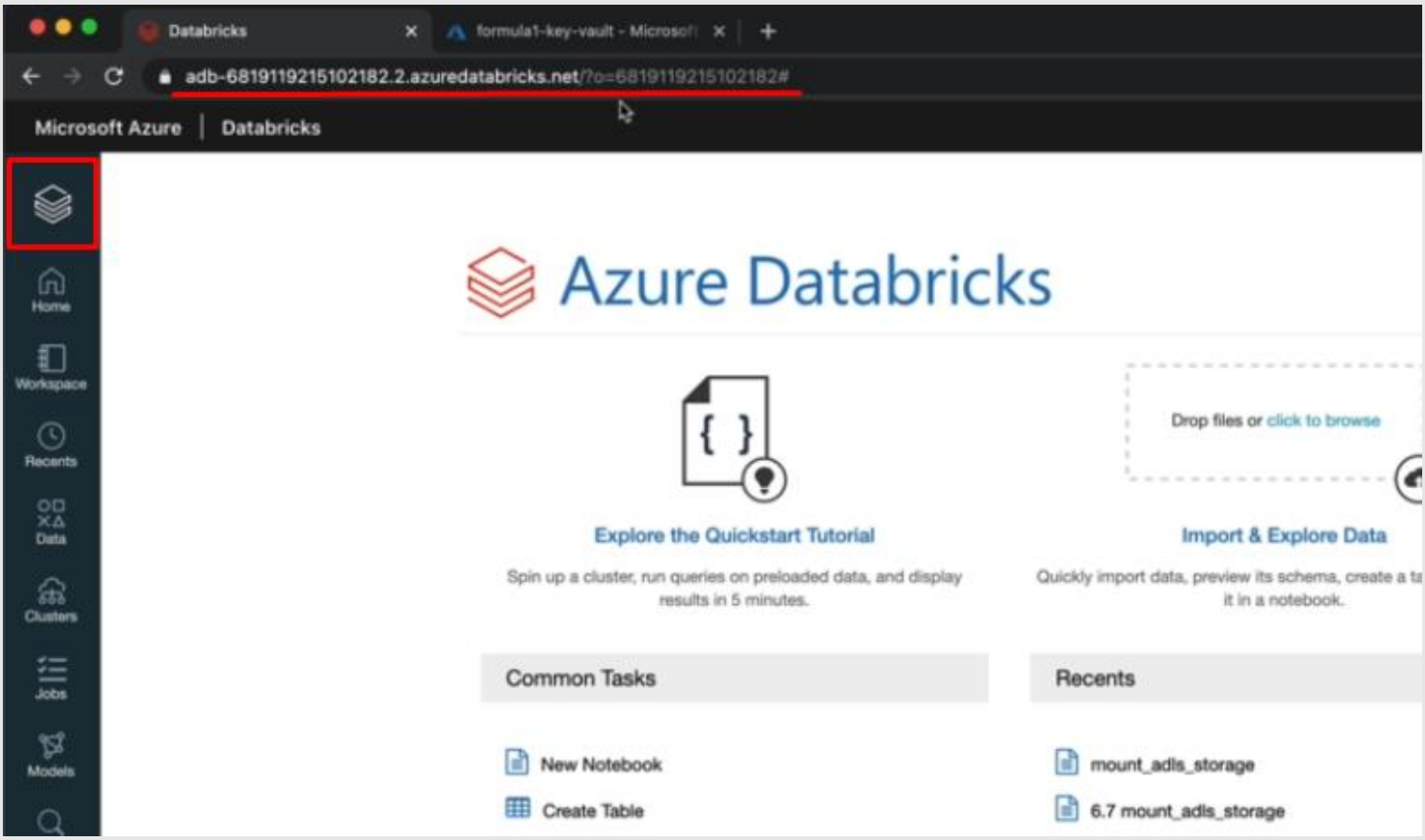


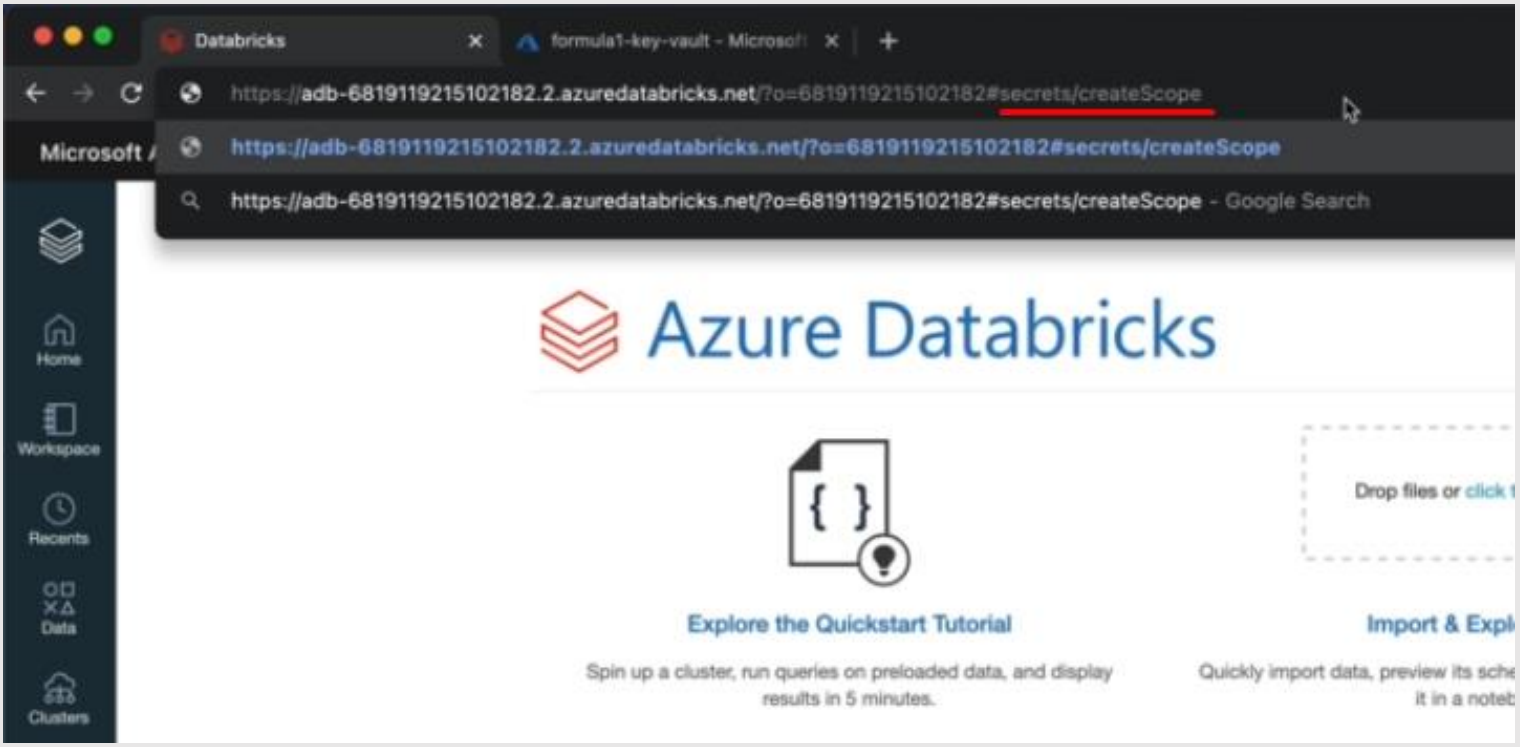


Se han creado los 3 secretos



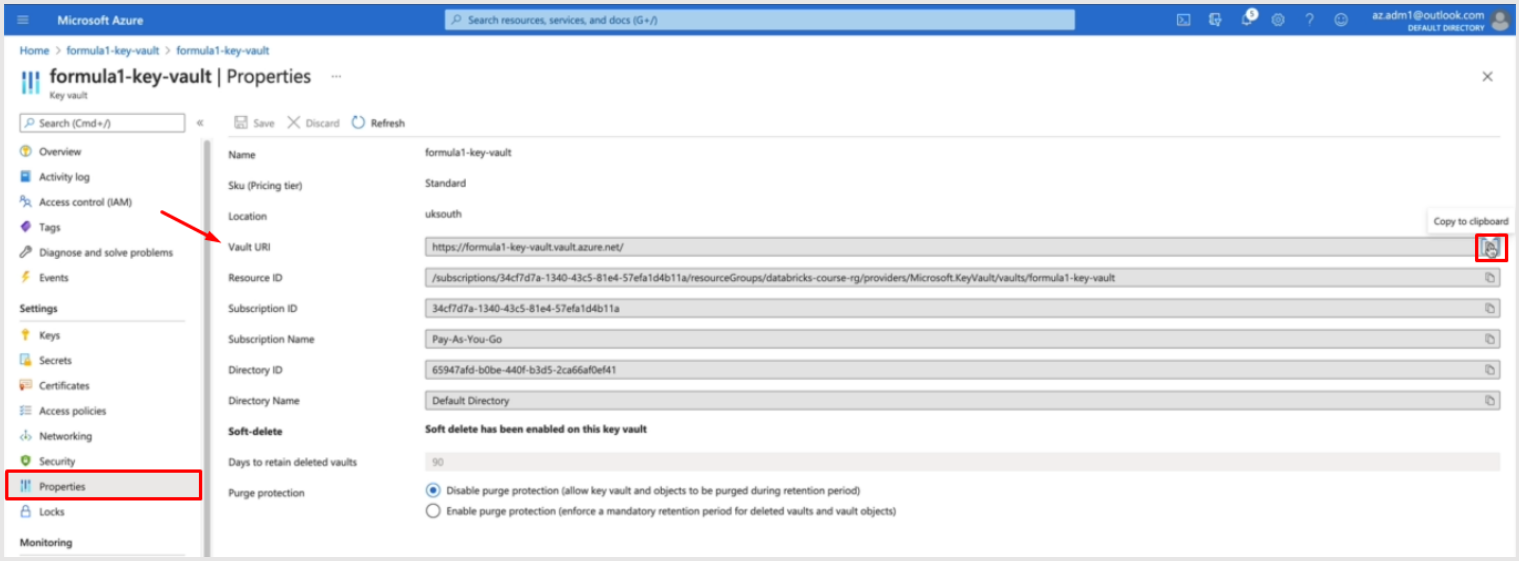
Luego, damos clic en el símbolo de databricks y en la ruta web en el navegador agregaremos **#secrets/createScope**. Vamos a utilizar **Databricks Scoped Secret**



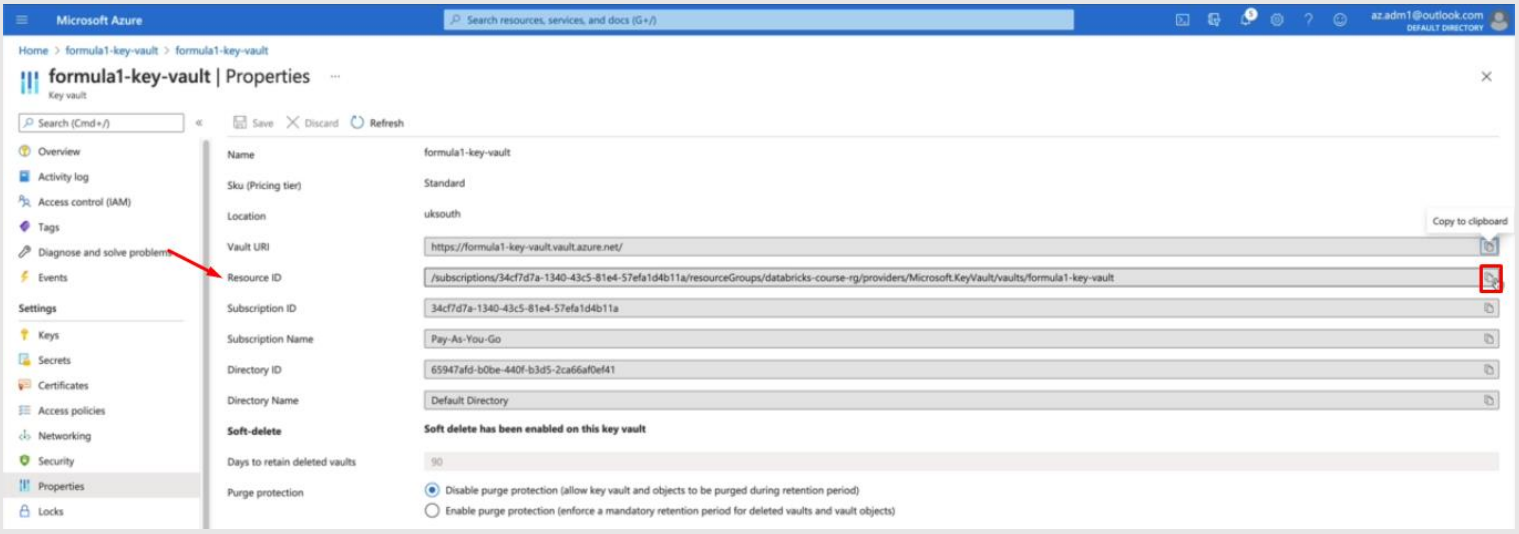


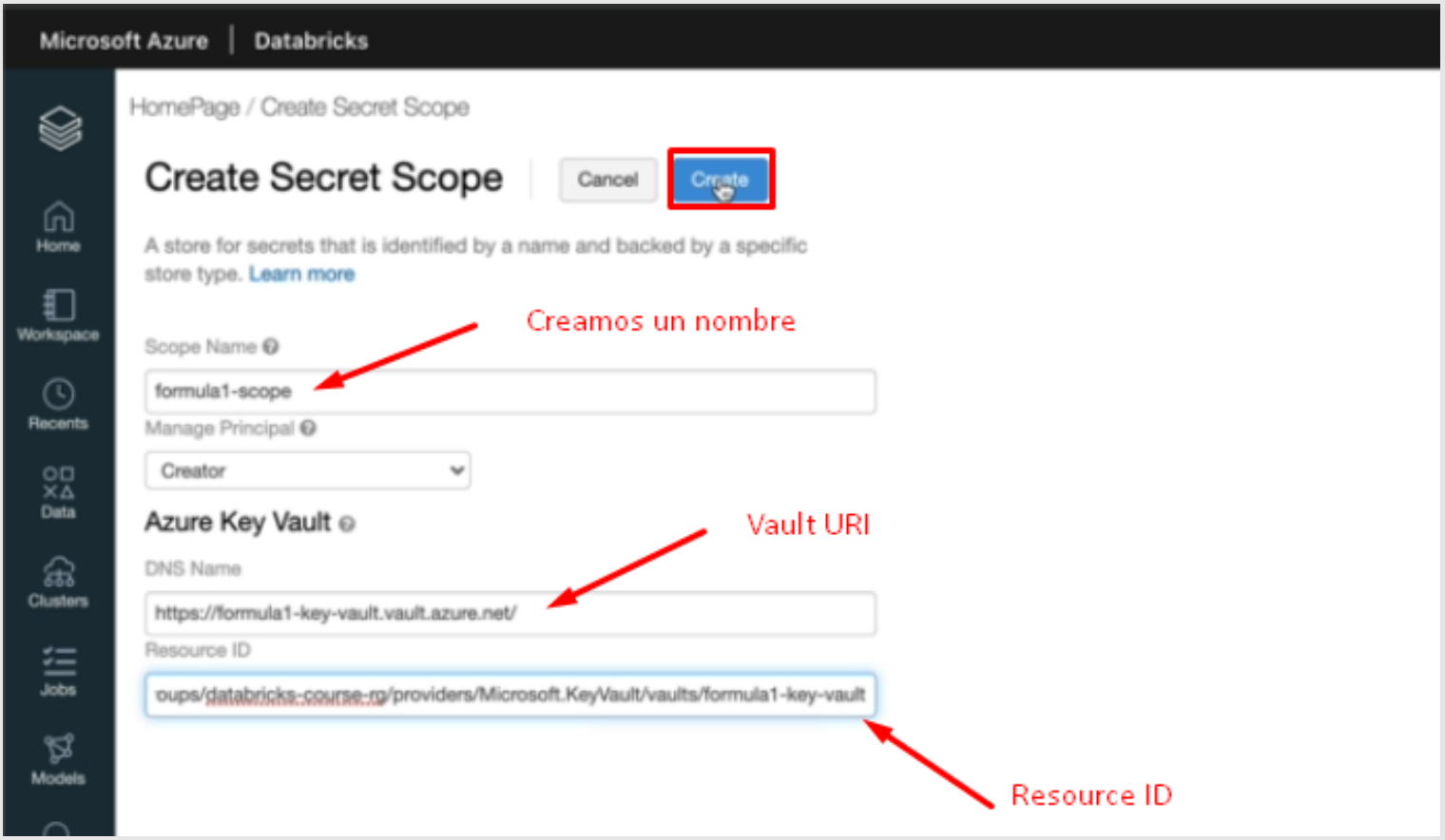
Regresamos al Azure Key Vault

Vault URI

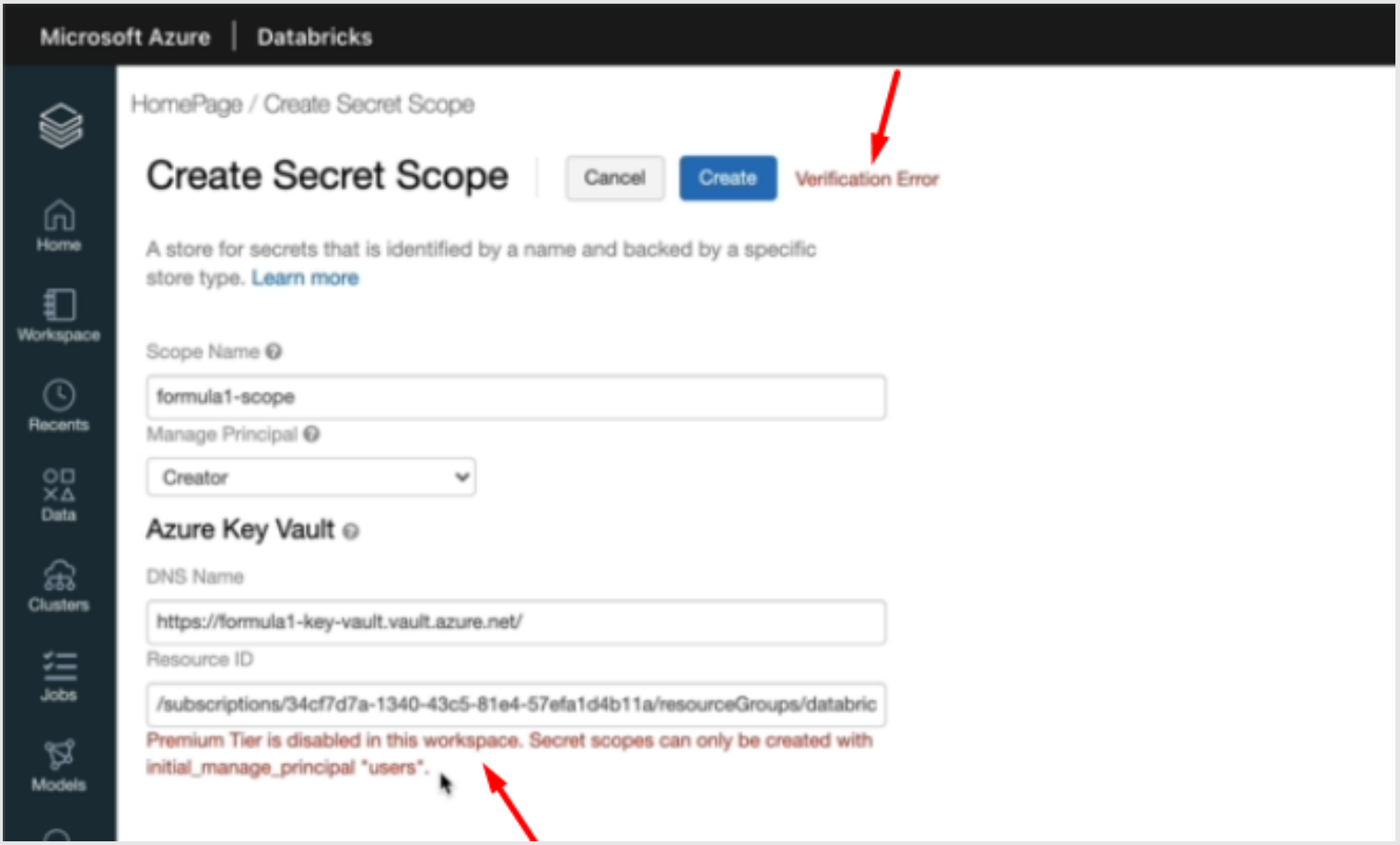


Resource ID

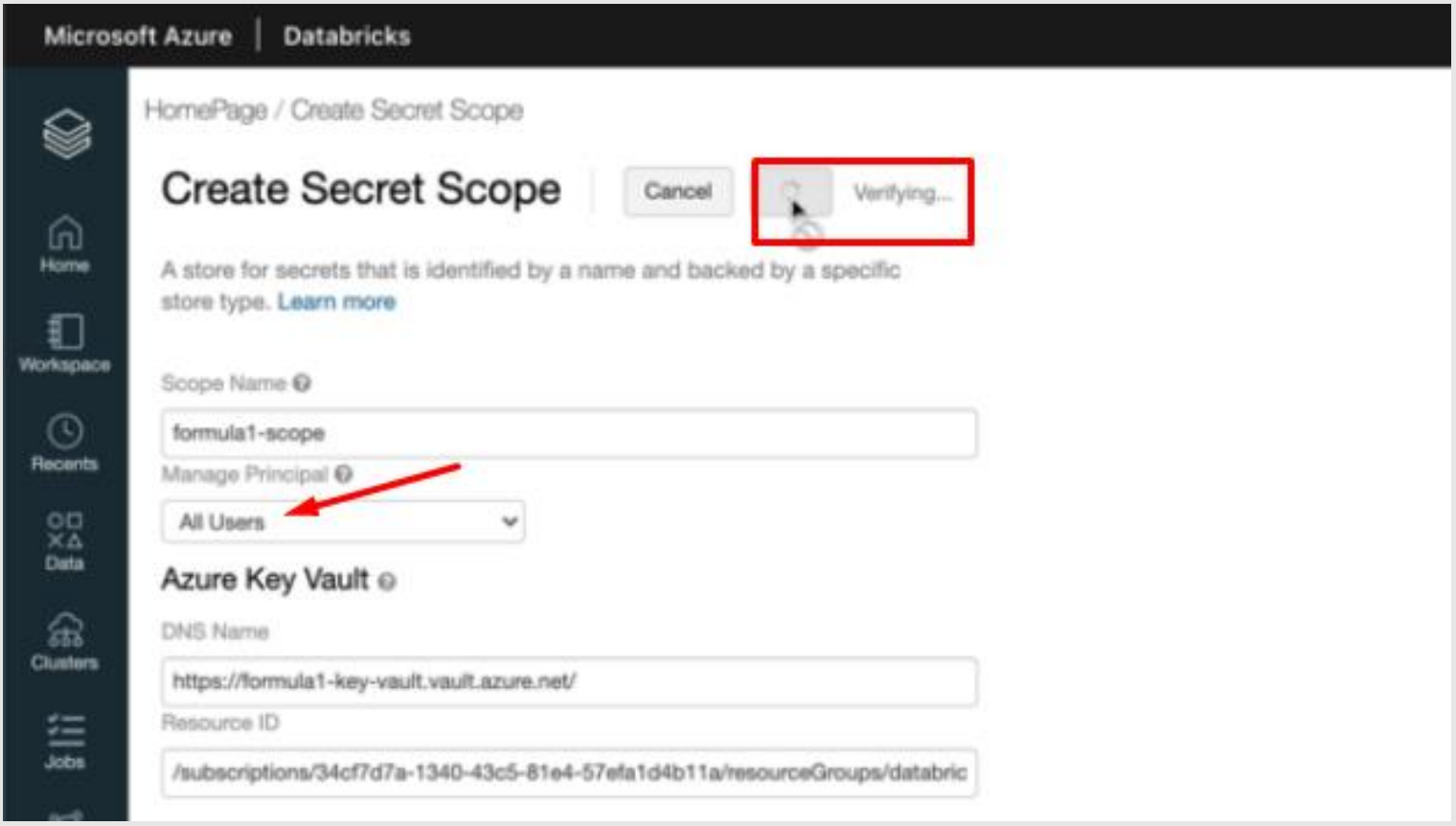


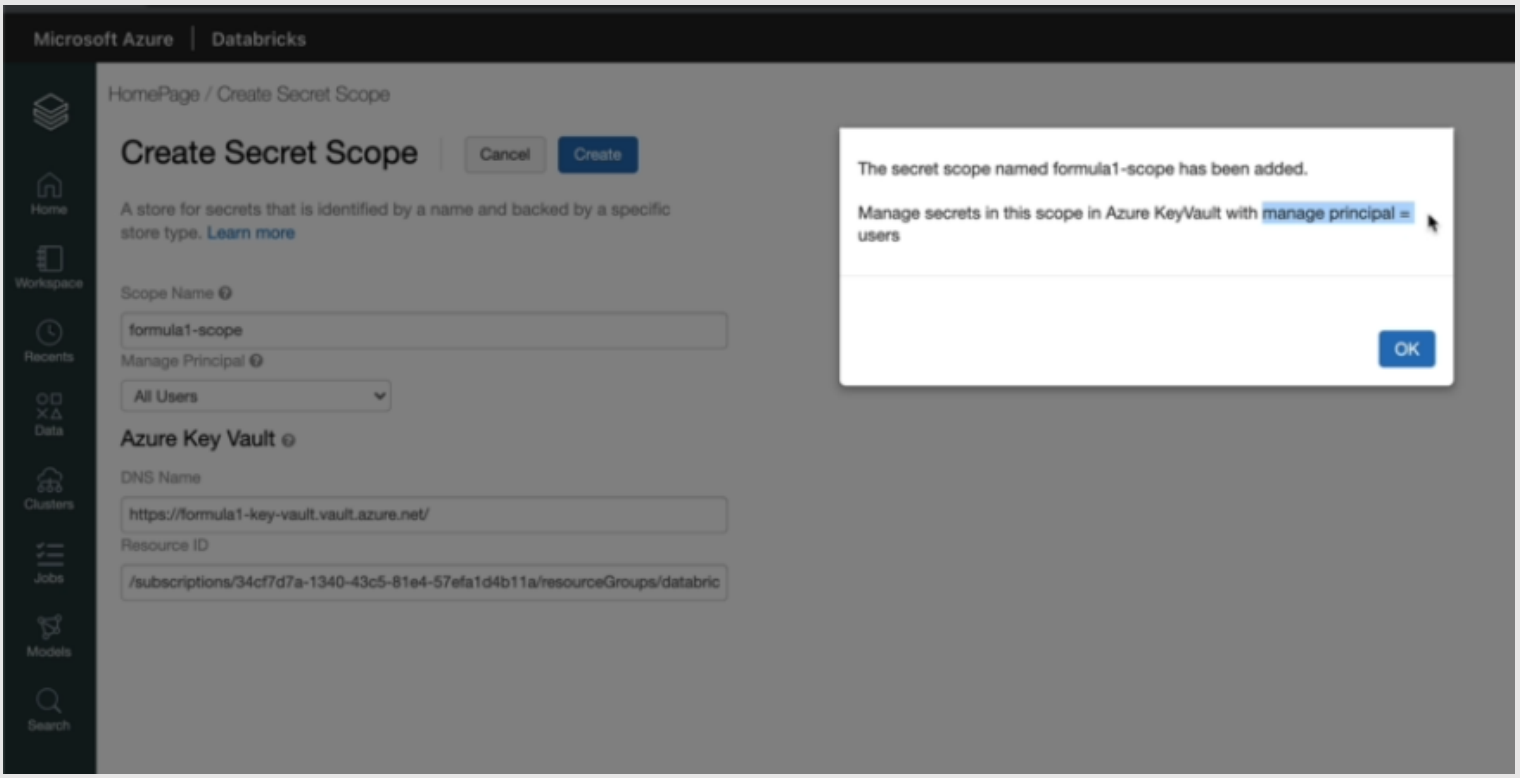


Nos generará un error



Modificamos el “Manage Principal” a “All 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")
```

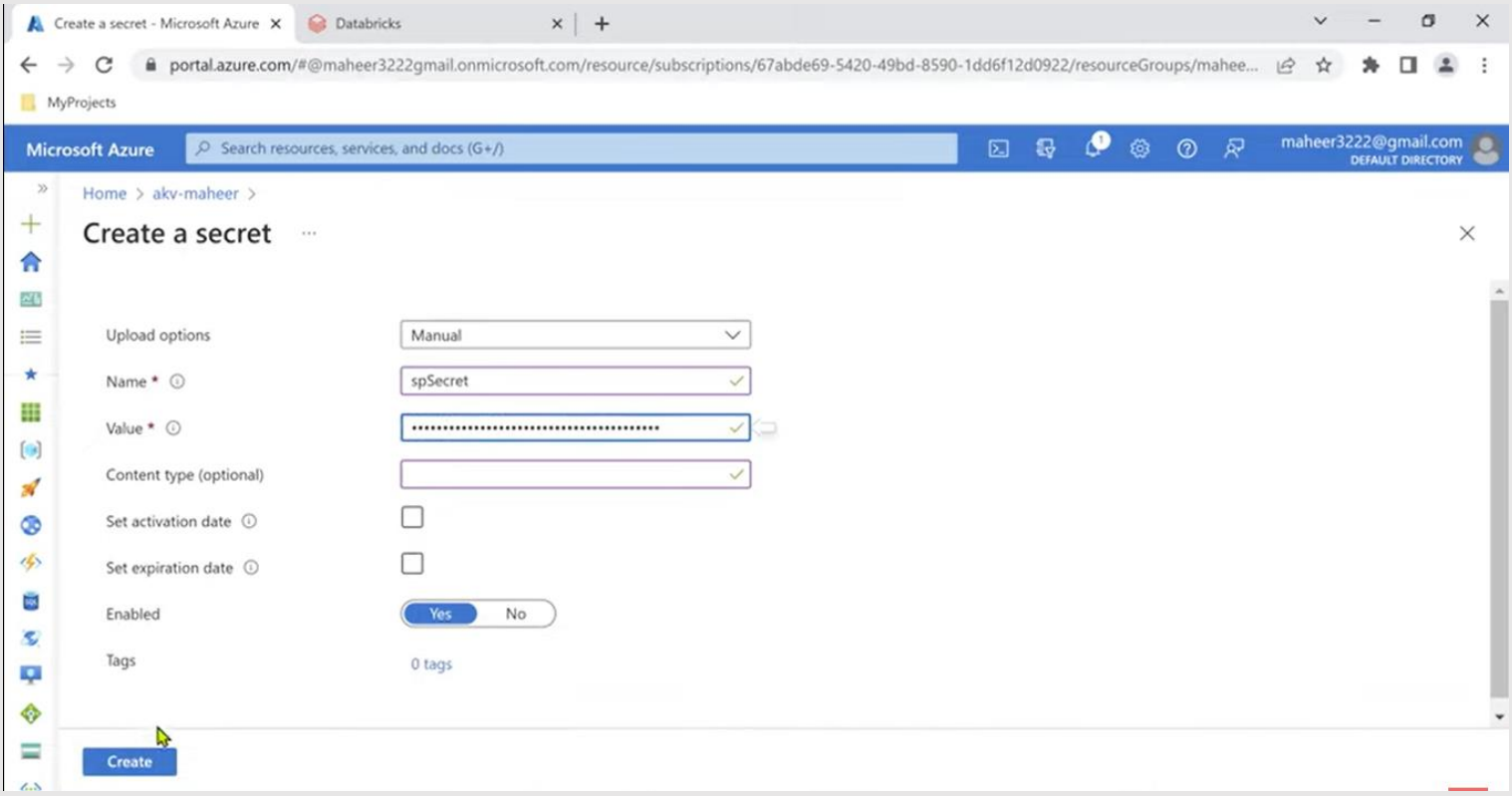
Cmd 6

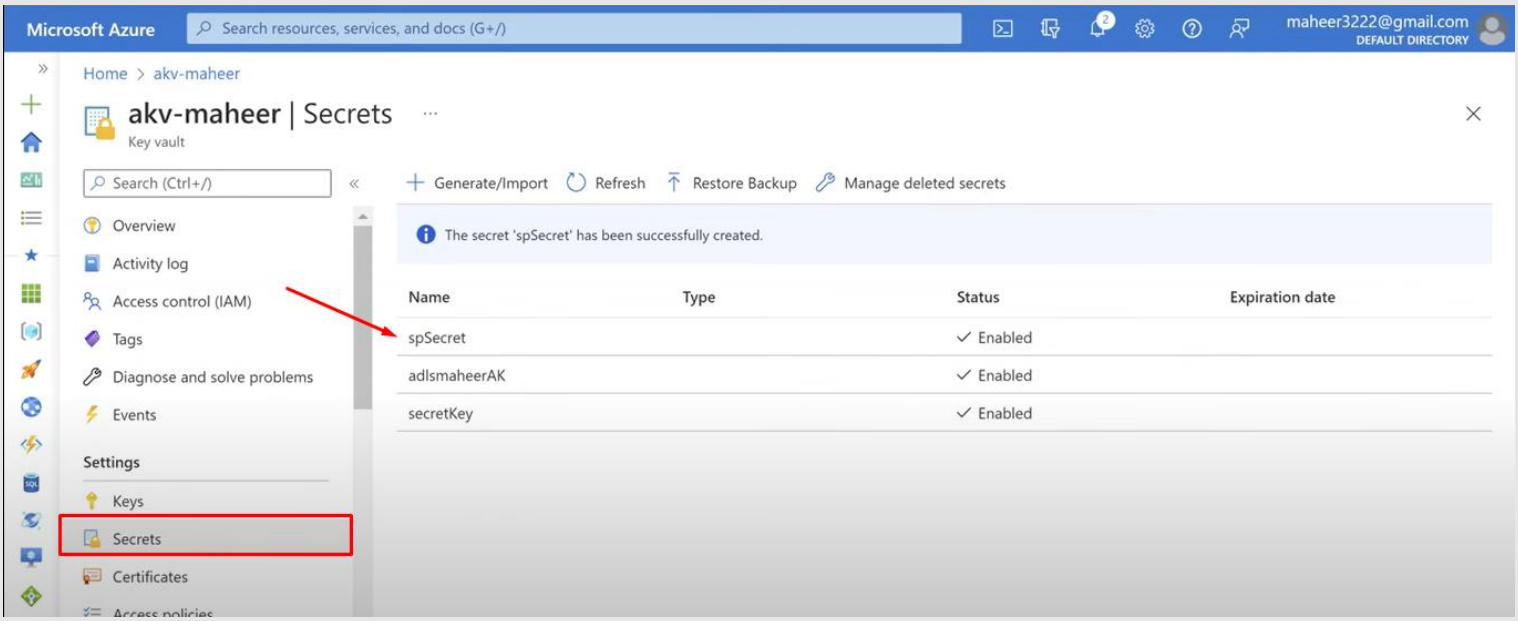
```
dbutils.fs.ls("/mnt/formula1dl/raw")
```

Cmd 7

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

Continuando con imágenes desde otro video, ahora debemos crear un secreto en Azure Key Vault con la clave que obtuvimos al crear nuestro Service Principal. En este ejemplo, no se utilizó **Databricks Scoped Secret**



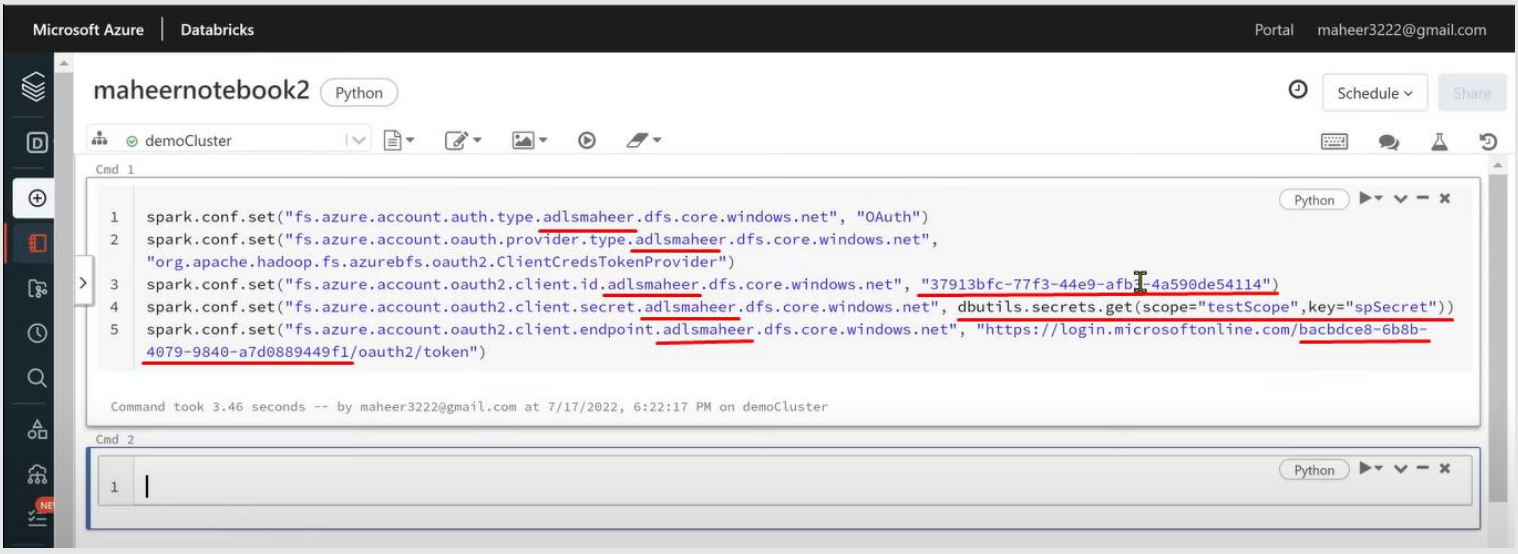


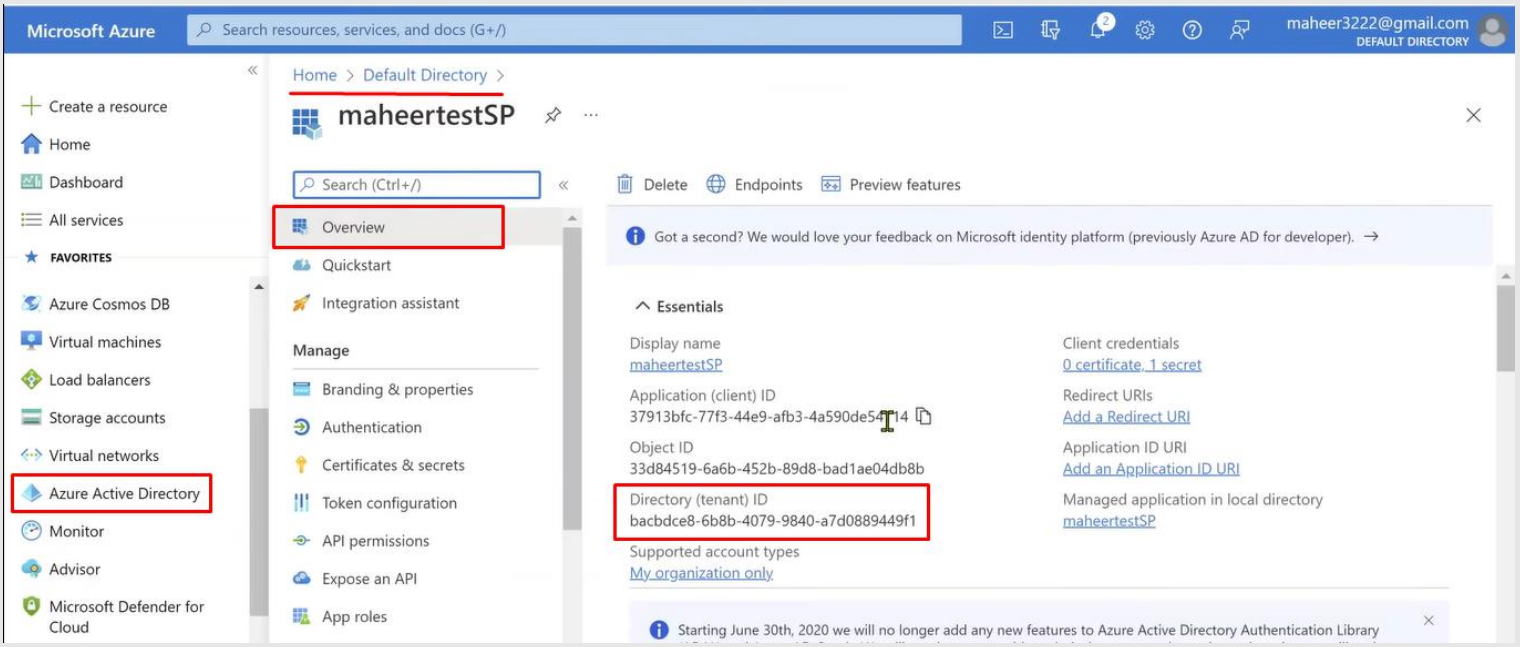
Access ADLS Gen2 with Service Principle

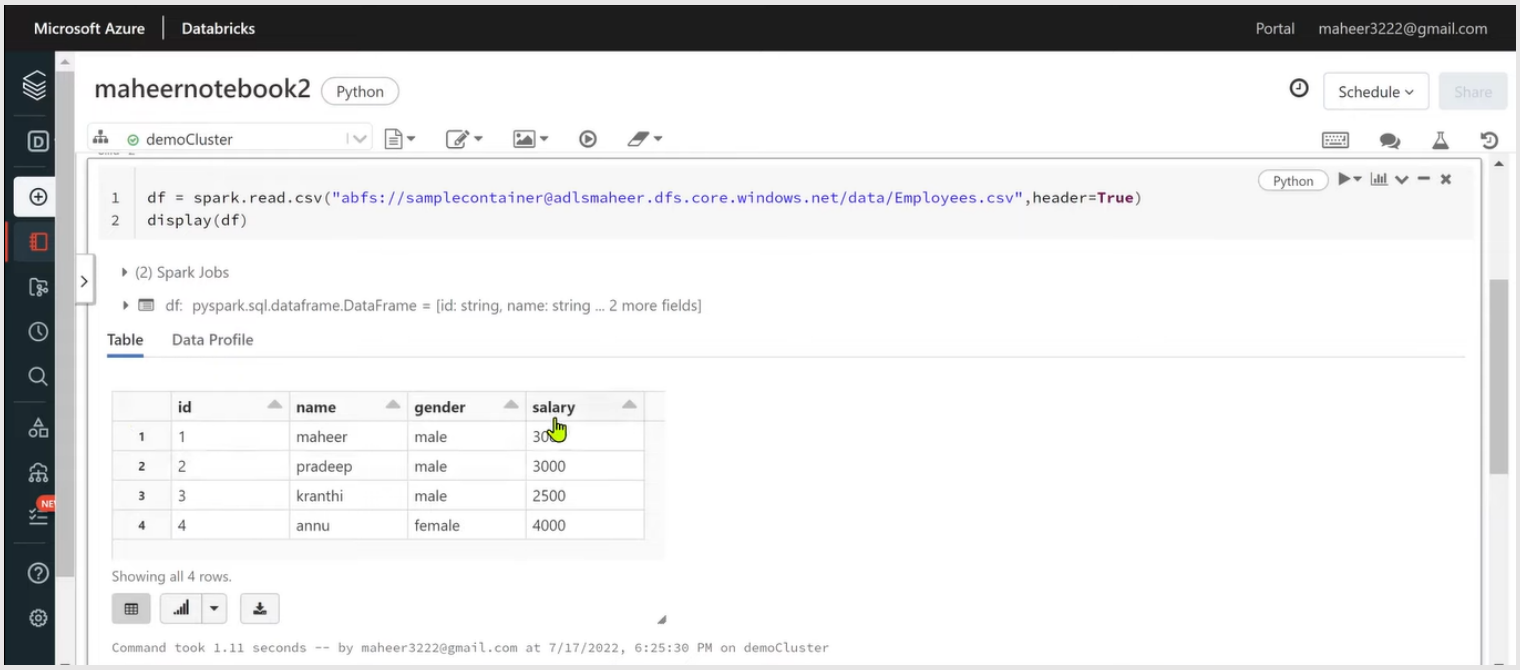
```
service_credential = dbutils.secrets.get(scope="<scope>",key="<service-credential-key>")

spark.conf.set("fs.azure.account.auth.type.<storage-account>.dfs.core.windows.net", "OAuth")
spark.conf.set("fs.azure.account.oauth.provider.type.<storage-account>.dfs.core.windows.net",
"org.apache.hadoop.fs.azurebfs.oauth2.ClientCredsTokenProvider")
spark.conf.set("fs.azure.account.oauth2.client.id.<storage-account>.dfs.core.windows.net",
"<application-id>")
spark.conf.set("fs.azure.account.oauth2.client.secret.<storage-account>.dfs.core.windows.net",
service_credential)
spark.conf.set("fs.azure.account.oauth2.client.endpoint.<storage-account>.dfs.core.windows.net",
"https://login.microsoftonline.com/<directory-id>/oauth2/token")
```

Cuenta de almacenamiento: adlsmaheer
Application id: Corresponde el ID del Service Principal
Service credential: **scope** --> testScope <-- Nombre de nuestro Service Principal
key: spSecret <-- Nombre de nuestro secreto en Azure Key Vault
Directory id: Corresponde al Directorio (tenant) ID --> Abajo deje una imagen de donde se obtiene




Microsoft Azure
Search resources, services, and docs (G+/)
maheer3222@gmail.com
DEFAULT DIRECTORY
Create a resource
Home
Dashboard
All services
FAVORITES
Azure Cosmos DB
Virtual machines
Load balancers
Storage accounts
Virtual networks
Azure Active Directory
Monitor
Advisor
Microsoft Defender for Cloud
Home > Default Directory >
maheertestSP
Search (Ctrl+/)
Overview
Quickstart
Integration assistant
Manage
Branding & properties
Authentication
Certificates & secrets
Token configuration
API permissions
Expose an API
App roles
Delete
Endpoints
Preview features
Got a second? We would love your feedback on Microsoft identity platform (previously Azure AD for developer).
Essentials
Display name
maheertestSP
Application (client) ID
37913bfc-77f3-44e9-afb3-4a590de54114
Object ID
33d84519-6a6b-452b-89d8-bad1ae04db8b
Directory (tenant) ID
bacbdce8-6b8b-4079-9840-a7d0889449f1
Supported account types
My organization only
Client credentials
0 certificate, 1 secret
Redirect URIs
Add a Redirect URI
Application ID URI
Add an Application ID URI
Managed application in local directory
maheertestSP
Starting June 30th, 2020 we will no longer add any new features to Azure Active Directory Authentication Library


Microsoft Azure
Databricks
Portal
maheer3222@gmail.com
maheernotebook2
Python
Schedule
Share
demoCluster
Python
1 df = spark.read.csv("abfs://samplecontainer@adlsmaheer.dfs.core.windows.net/data/Employees.csv",header=True)
2 display(df)
(2) Spark Jobs
df: pyspark.sql.dataframe.DataFrame = [id: string, name: string ... 2 more fields]
Table
Data Profile
id
name
gender
salary
1 1 maheer male 30
2 2 pradeep male 3000
3 3 kranthi male 2500
4 4 annu female 4000
Showing all 4 rows.
Command took 1.11 seconds -- by maheer3222@gmail.com at 7/17/2022, 6:25:30 PM on demoCluster