**PROYECTO FINAL DATABRICKS**

El objetivo del siguiente proyecto es disponibilidad la información de las atenciones presenciales que se realizan en un Centro de Atención, para la cual se está desarrollando la siguiente arquitectura en Azure Databricks.

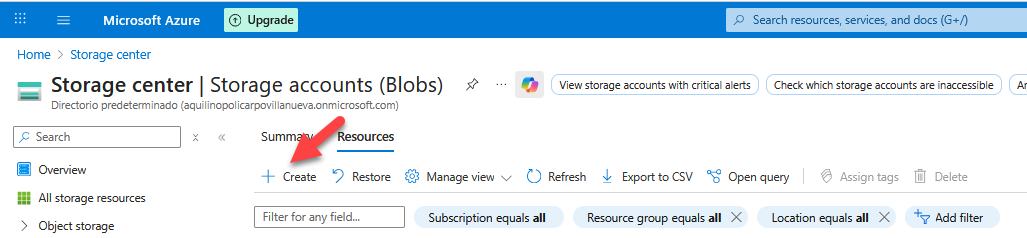
Interfaz de usuario gráfica, Aplicación

El contenido generado por IA puede ser incorrecto.

Los primeros pasos es el aprovisionamiento de los servicios en la Nube y las configuraciones para la conectividad entre ellos.

**SERVICIO: Azure datalake storage Gen 2**

Buscamos el servicio Storage Accounts, y damos click en el botón crear.

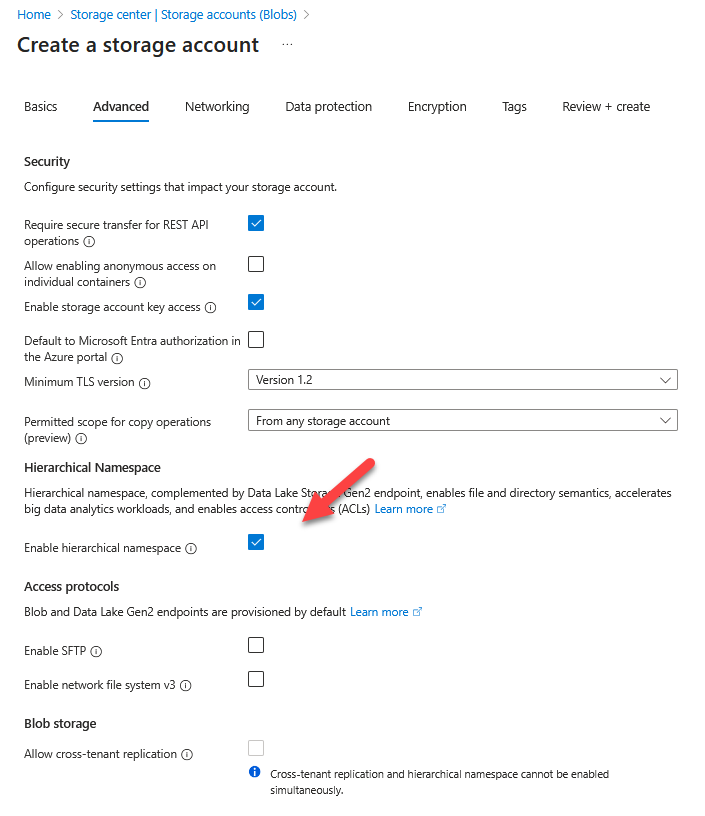


Seleccionamos el grupo de recursos: RG-PROYECTO01.

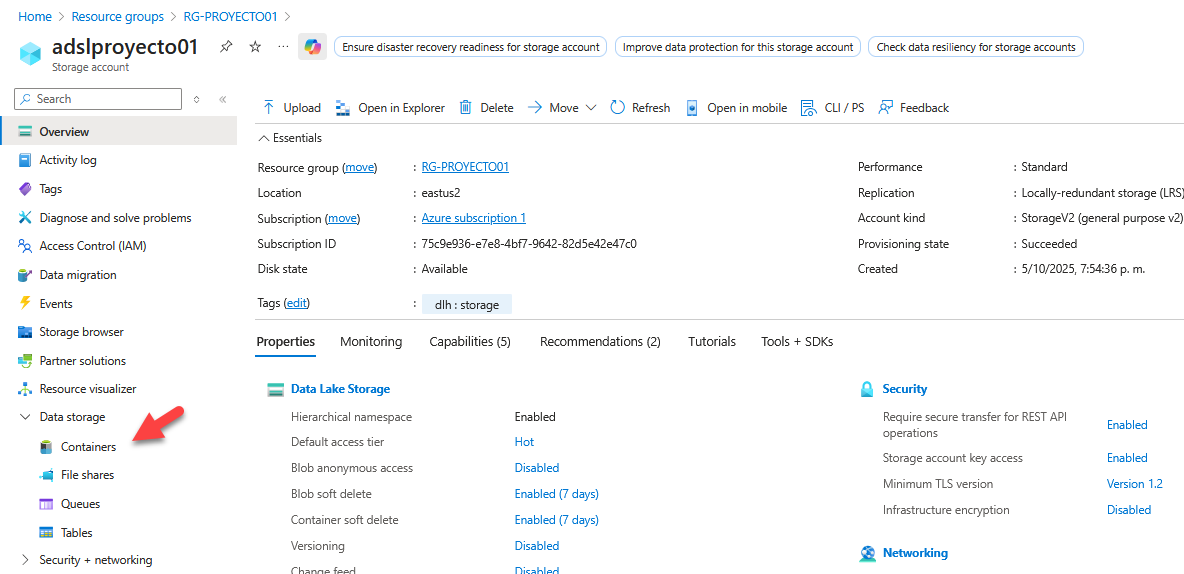
Colocamos el nombre del Storage Account: adlsproyecto01



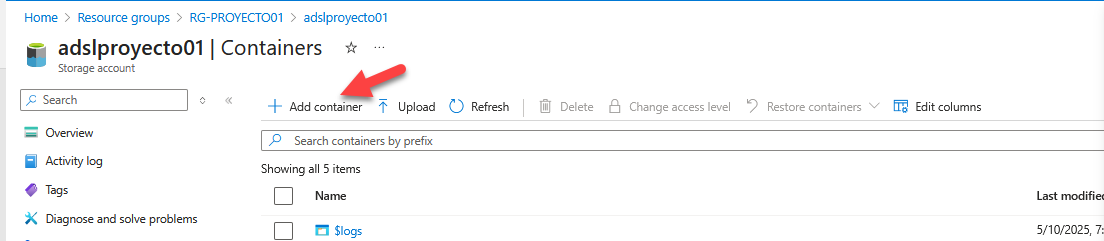
En la pestaña Advanced, seleccionamos el espacio de nombre jerárquico, para que le storage account se comporte como un Data Lake.

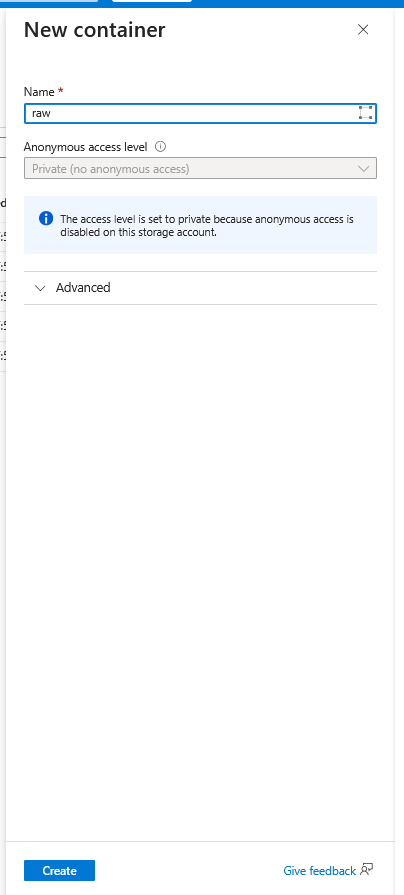


Luego de crear el servicio ADLS, se crea los contenedores, en la opción Data Storage, Containers.

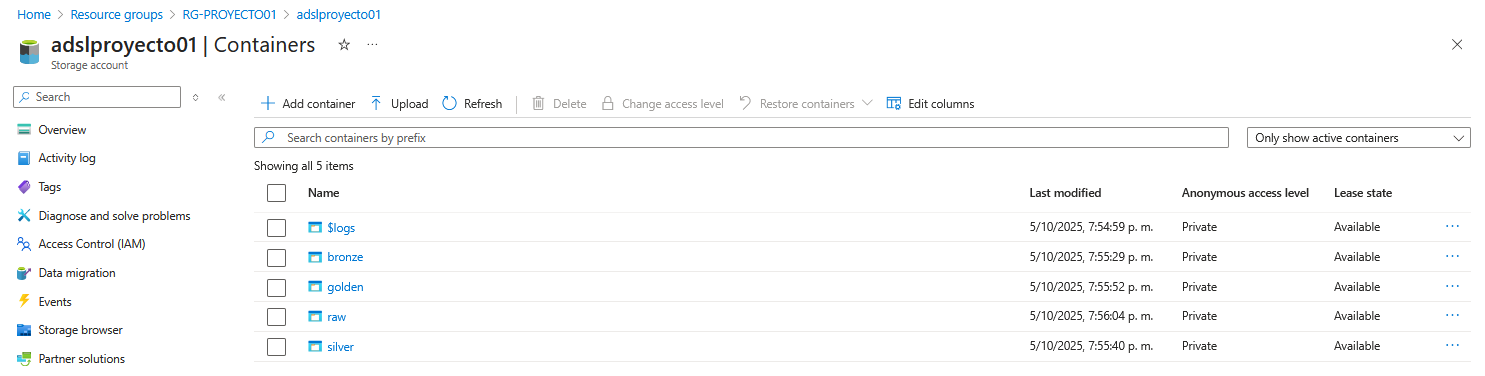


Creamos el contenedor raw:



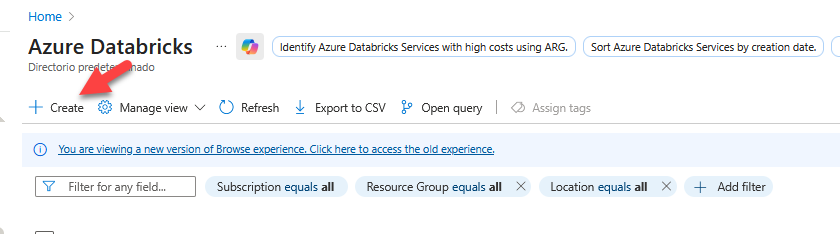


Creamos el resto de contenedores: bronze, silver, Golden



**SERVICIO: Azure Databricks**

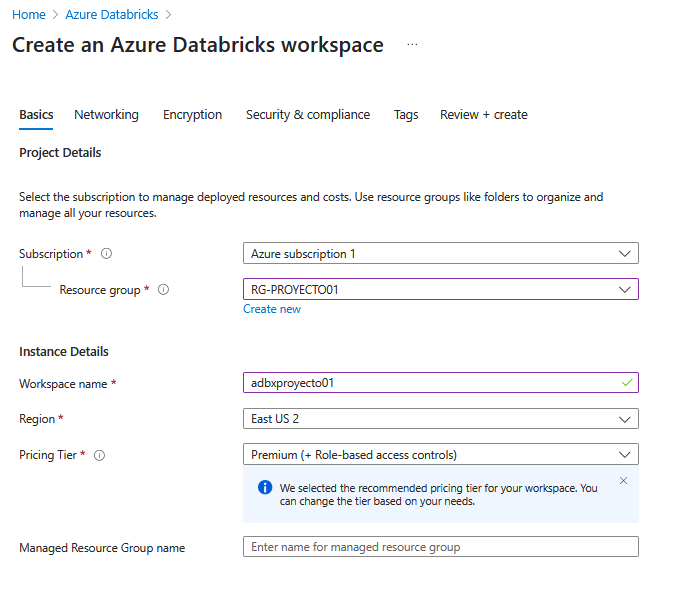
Buscamos el servicio Azure databricks, y damos click en la opción crear.



Colocamos el Grupo de Recursos: GR-PROYECTO01

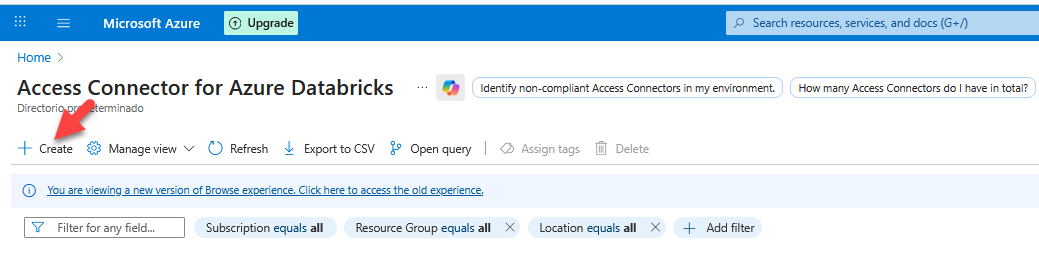
Colocamos el nombre del servicio: adbxproyecto01

Colocamos la región: East US 2



**SERVICIO: Azure Access Connector for Azure Databricks**

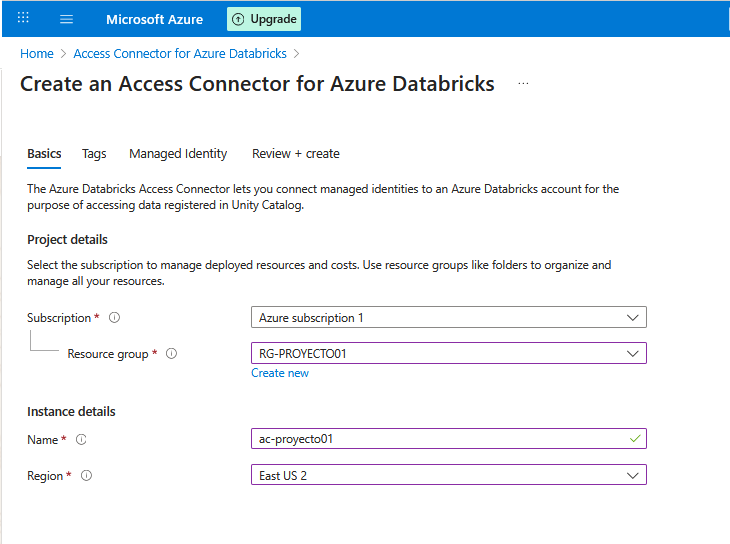
Buscamos el servicio Access Connector. Y damos click en el botón crear.



Seleccionamos el grupo de recursos: GR-PROYECTO01

Colocamos un nombre al conector: ac-proyecto01

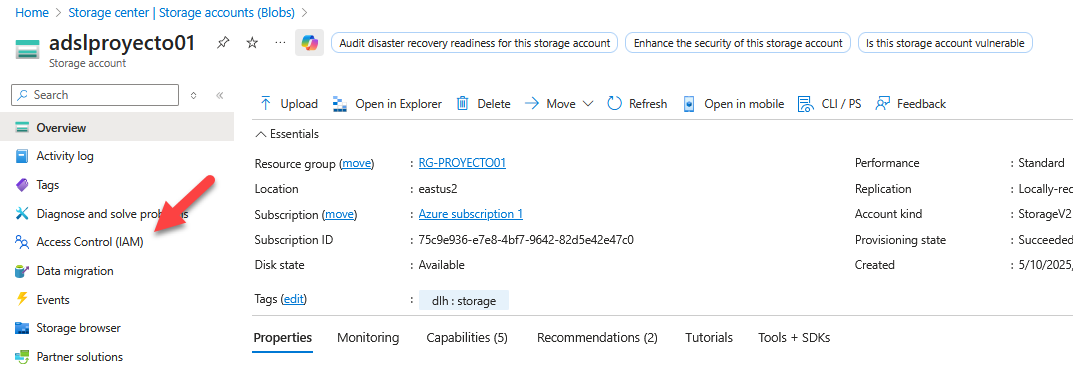
Colocamos la región: East US 2



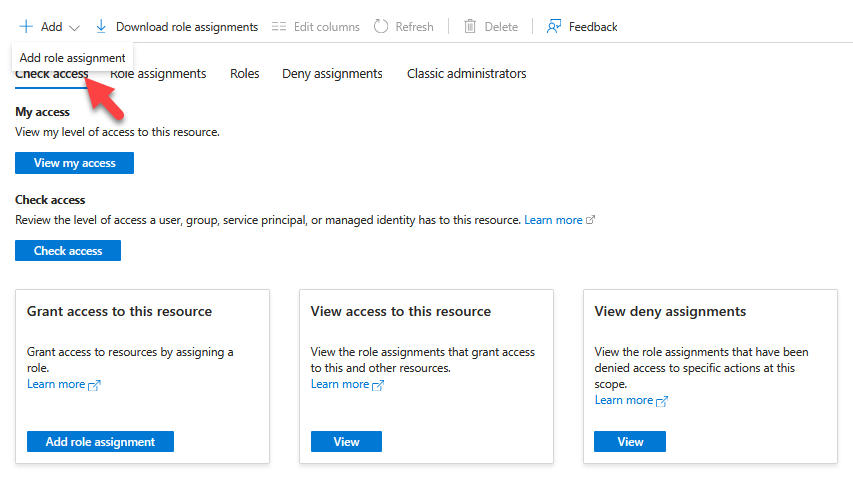
**Permisos de Data contributor al ADSL.**

Procedemos a vincular el Access Connector al ADSL.

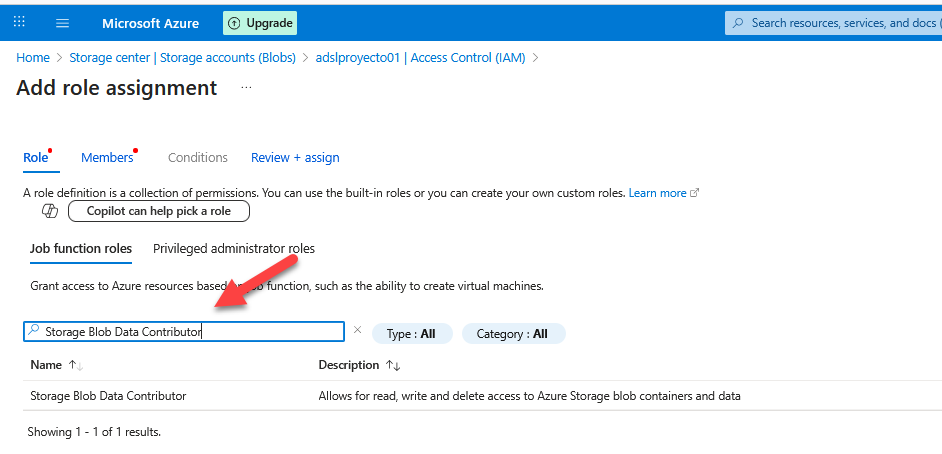
Vamos a la opción de Access Control (IAM)



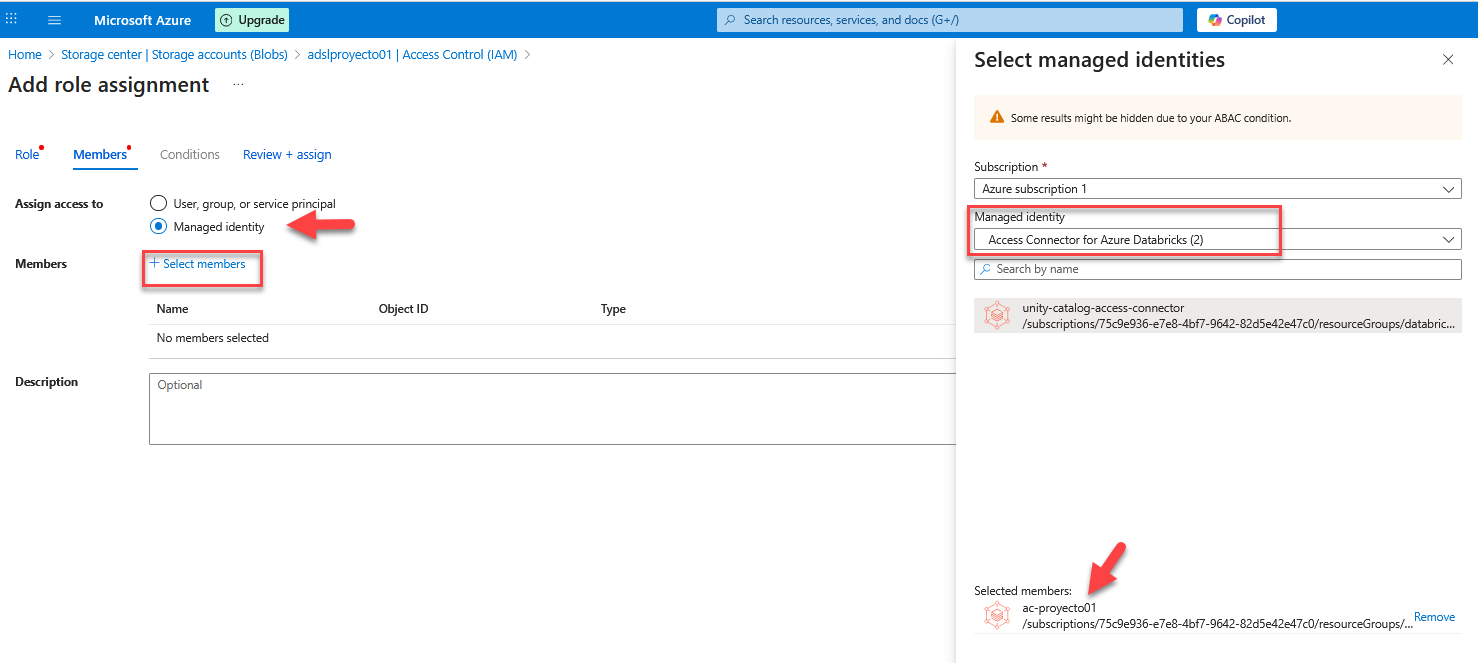
Luego la opción de Añadir Rol.



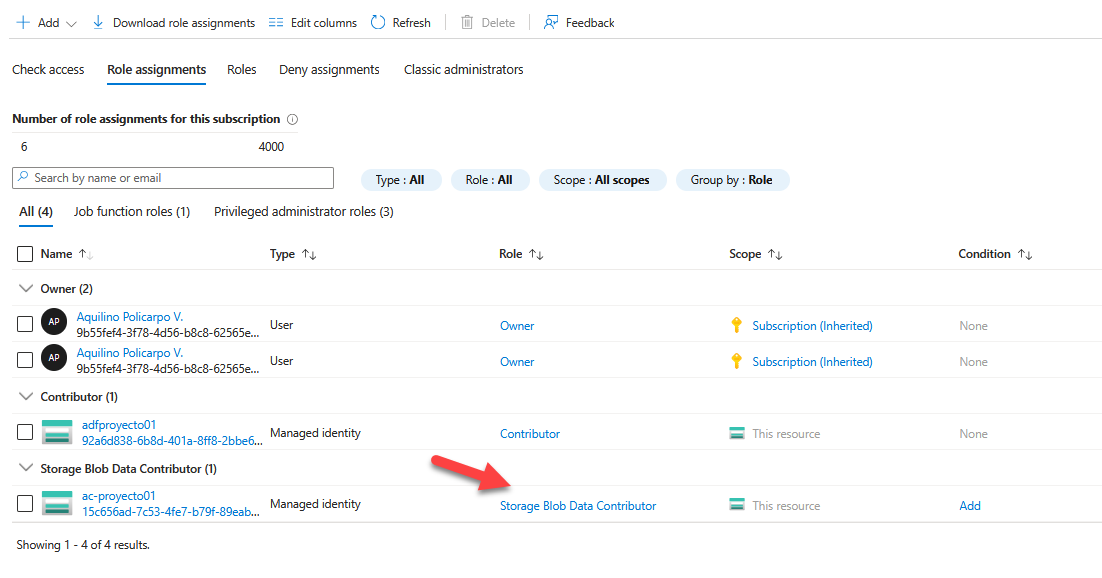
Buscamos el rol “Storage Blob Data Contributor” y luego damos click en Next



Asignamos el permiso de tipo “Managed Identity”, y seleccionamos el Access Control creado.

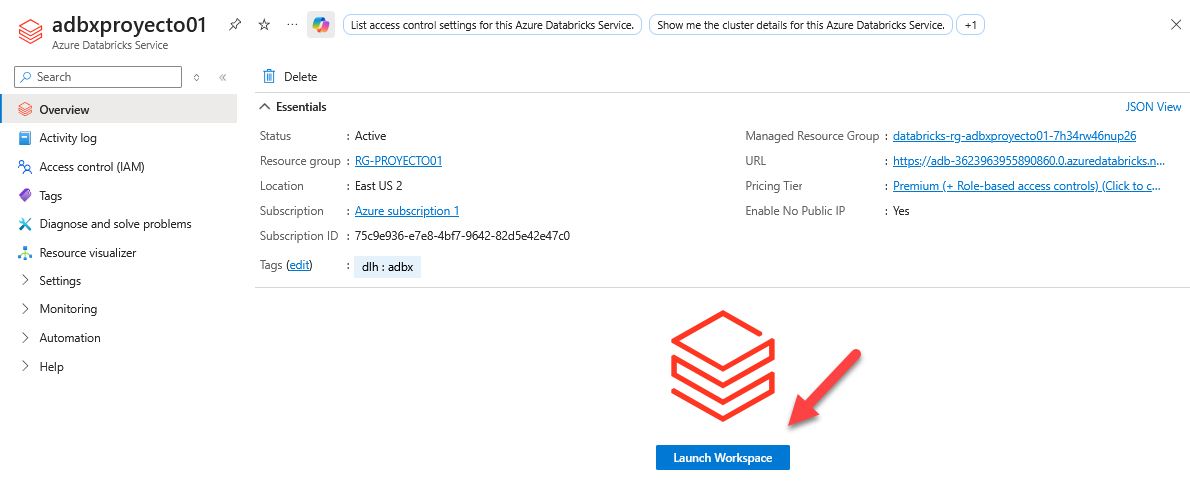


Validamos que el rol se ha asignado al Blob Storage.

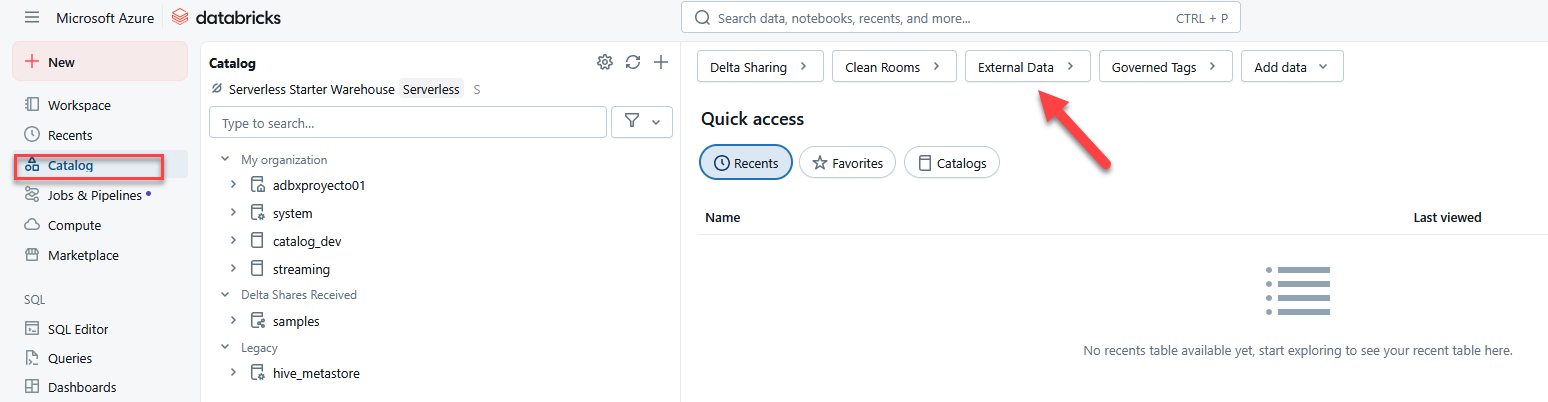


**Creamos los External Credential y External Location en el servicio de Databricks.**

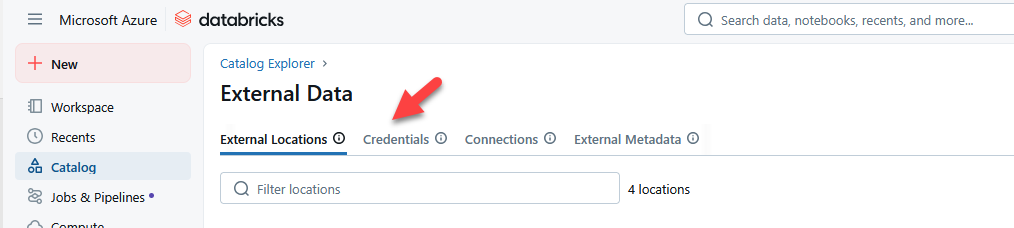
Entramos al servicio de Azure Databricks.



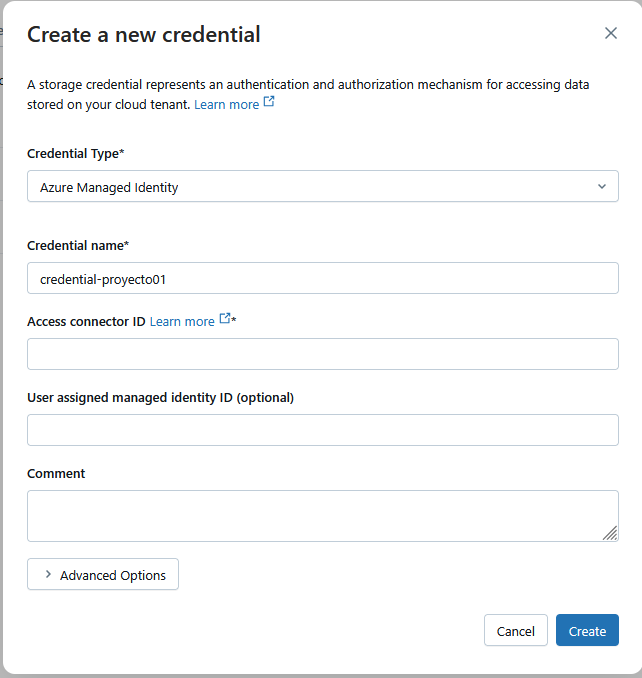
Luego vamos a la opción de Catalog, y el botón External Data.



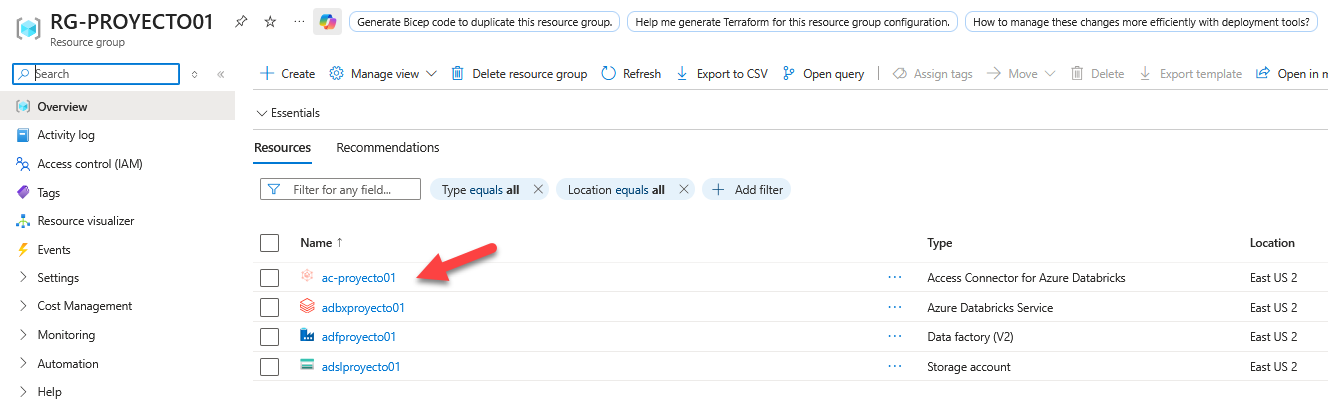
Primero creamos el External Credential.



Colocamos el nombre de la Credencial

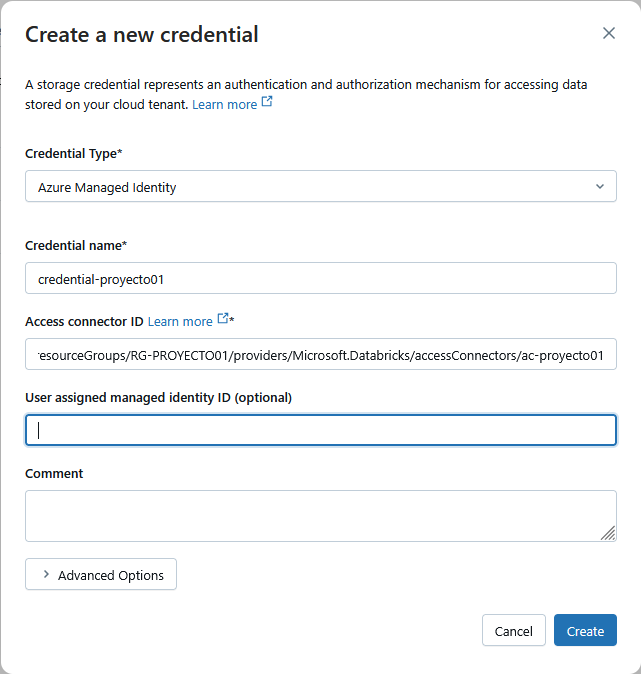


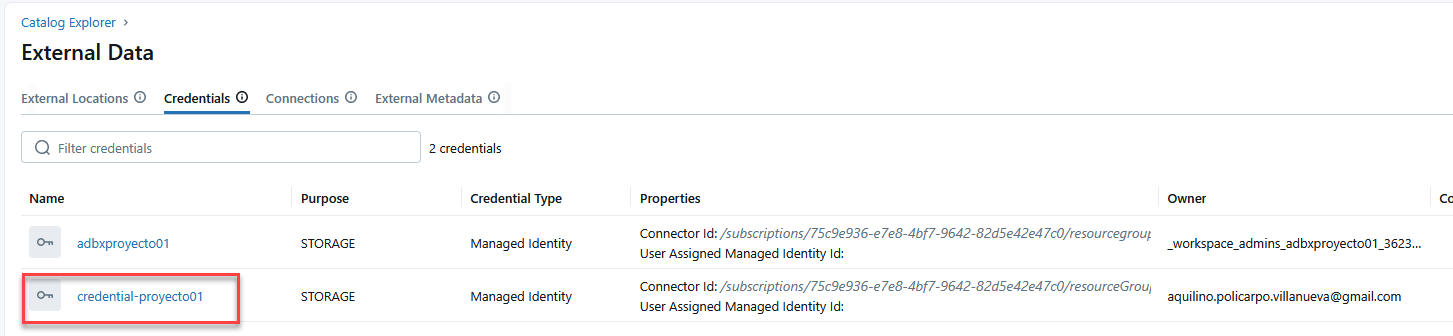
Para colocar el Access connector ID, vamos al Servicio Access Connector.



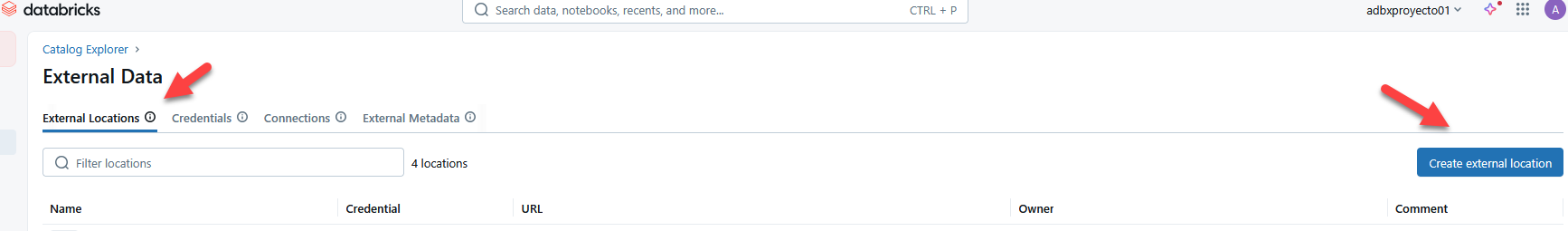


Copiamos el ID y luego damos Crear.





Luego creamos el External Location.

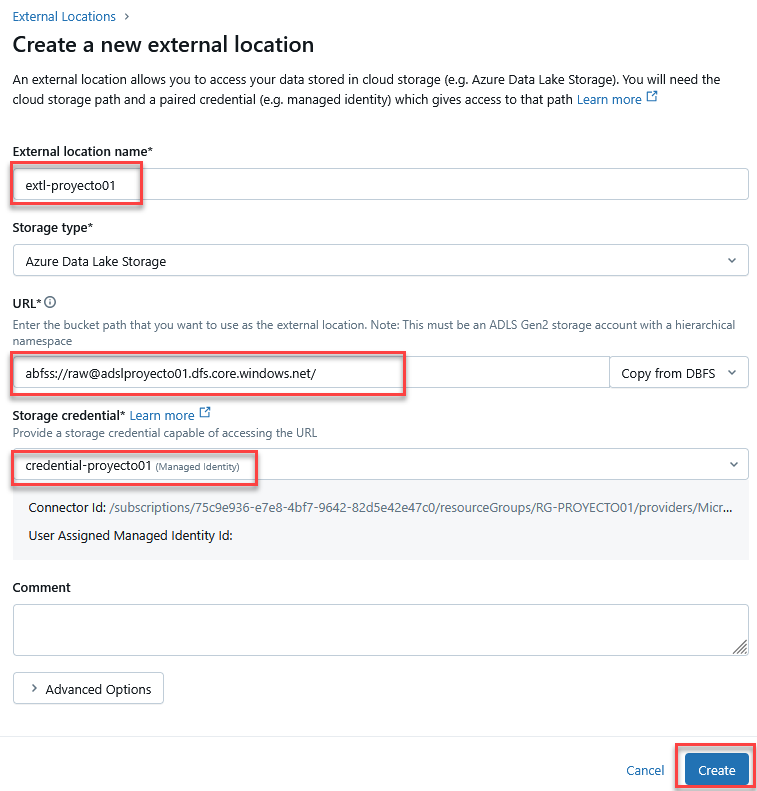


Seleccionamos el nombre del Externla location: extl-proyecto01

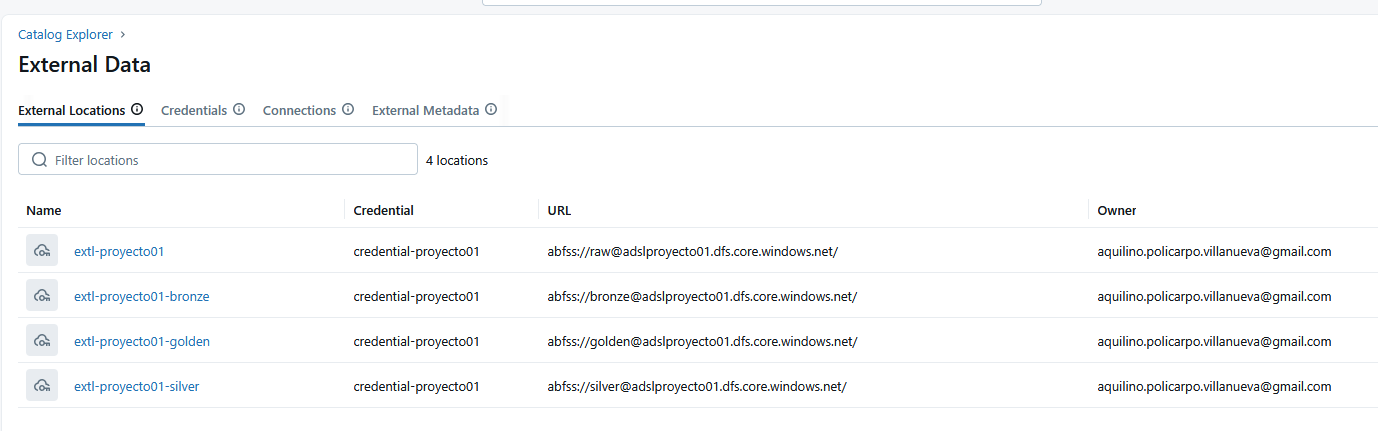
Seleccionamos la ubicación de nuestro contenedor en el ADSL: abfss://bronze@adslproyecto01.dfs.core.windows.net/

Seleccionamos la Credencial configurada: creadential-proyecto01

Luego damos click en crear.

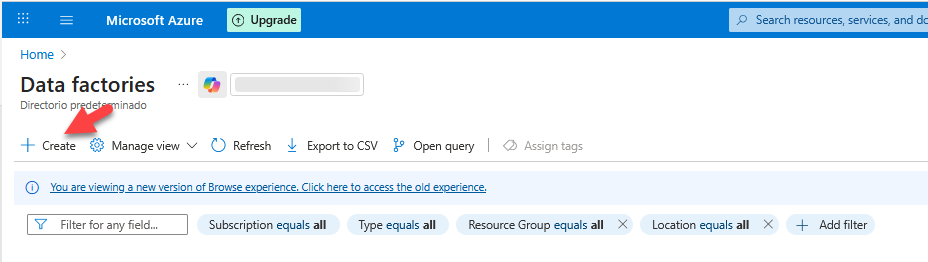


Luego creamos los demás external location para el resto de contenedores.



**SERVICIO: Azure Data Factory**

Buscamos el servicio Azure Data Factory, luego damos click en crear.

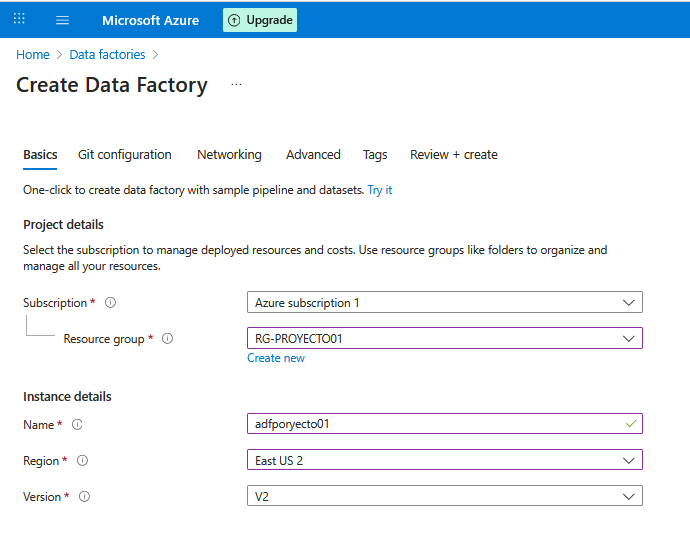


Seleccionamos el grupo de recursos: RG-PROYECTO01

Colocamos el nombre del Servicio data Factory: adfproyecto01

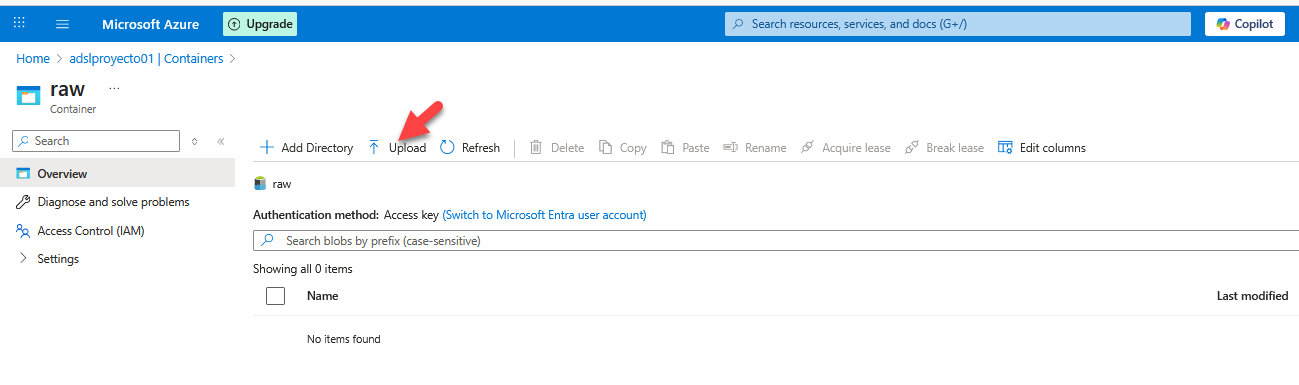
Seleccionamos la región: East US 2

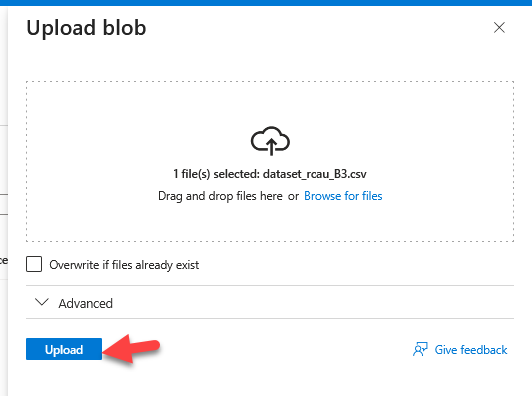
Luego damos click en Crear

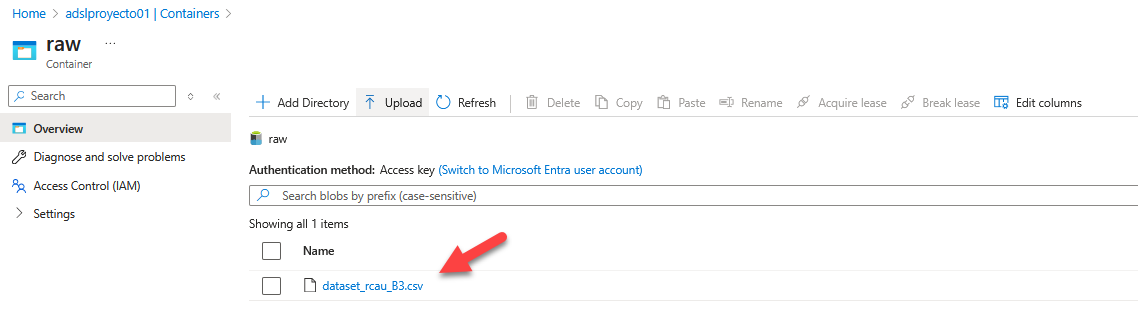


Después de configurar los servicios, vamos a desarrollar los notebooks para el ETL.

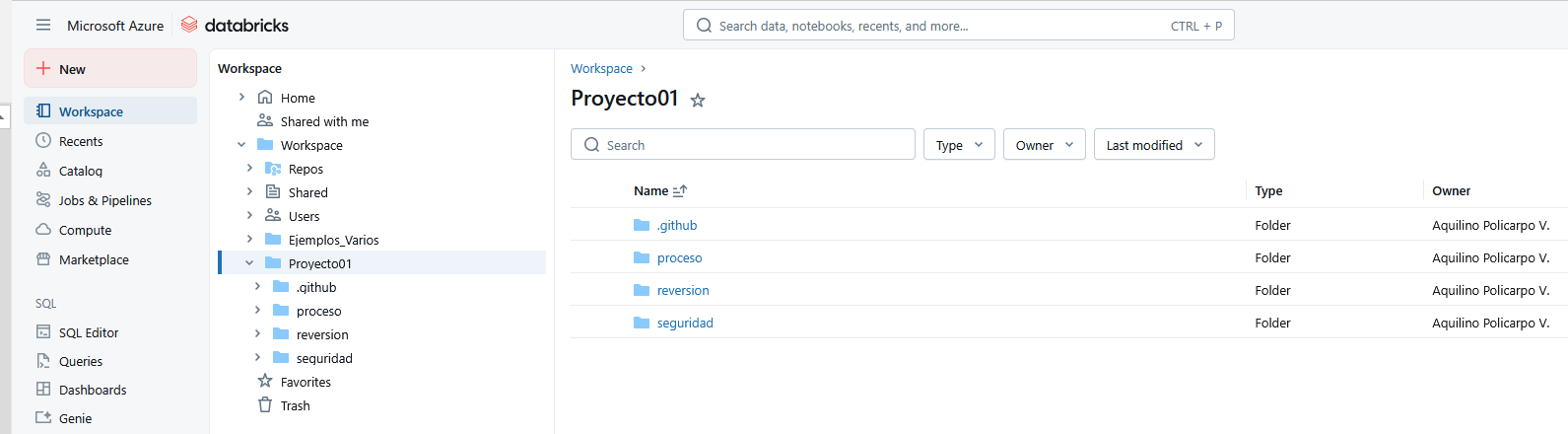
Primero cargamos el Dataset en el contenedor **raw** del ADSL.



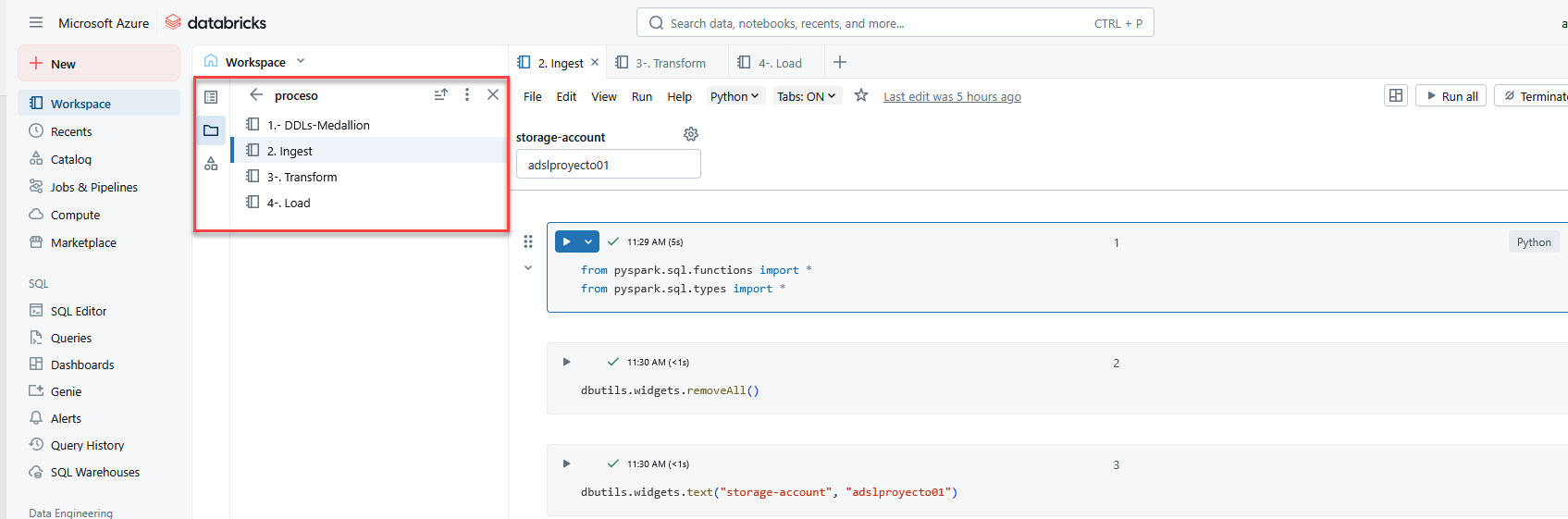




Desarrollamos los Notebooks en Azure Databricks.



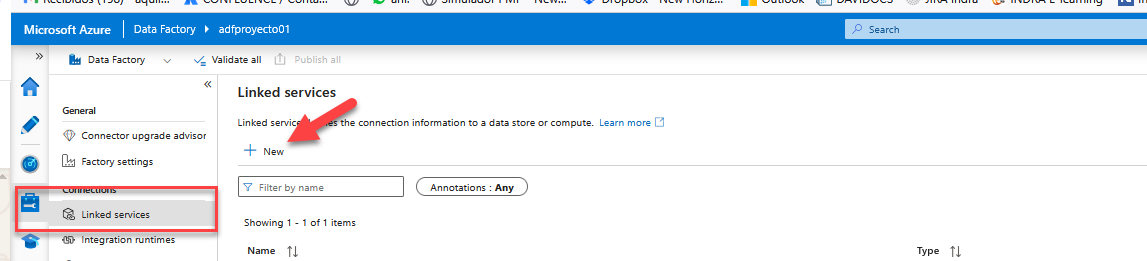
Dentro de la carpeta “**Procesos**” tenemos los Notebooks para el ETL

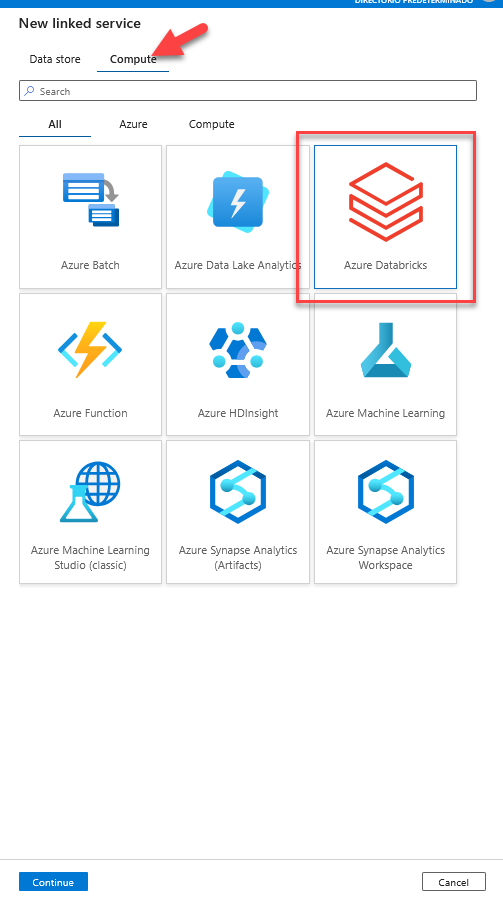


**Implementación del Workflow en Azure Data Factory.**

Como primer paso creamos los Link Services para el Databricks y ADSL.

Creamos el LS para Databricks



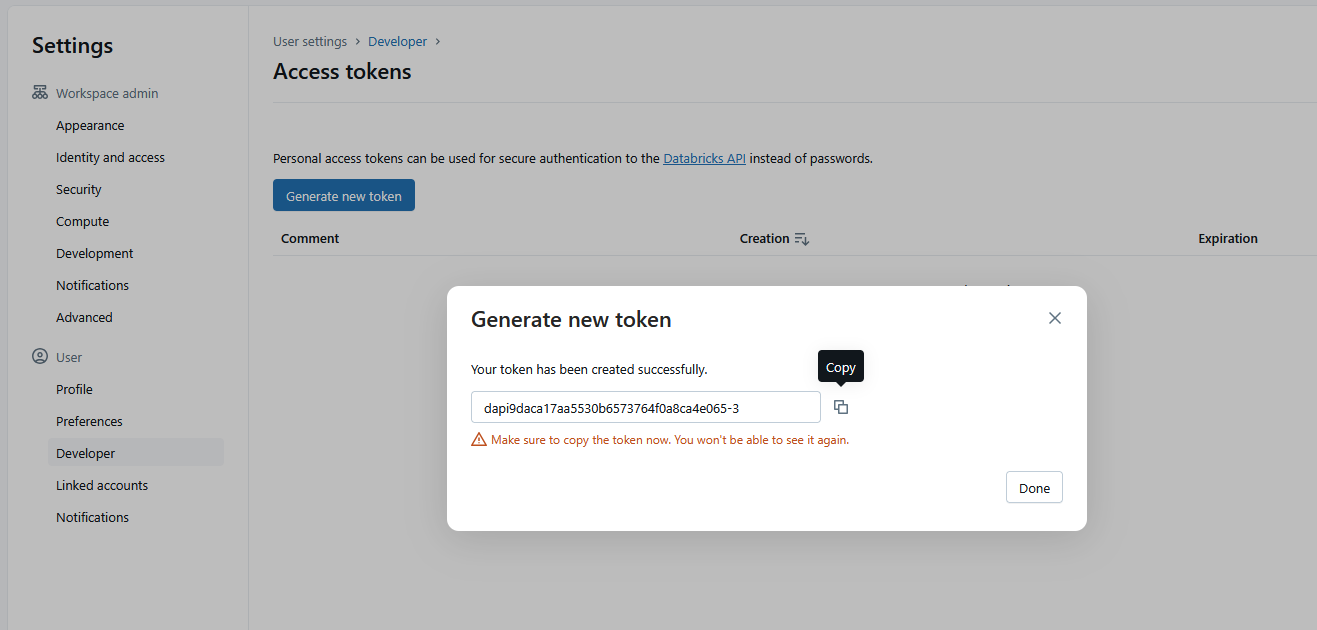


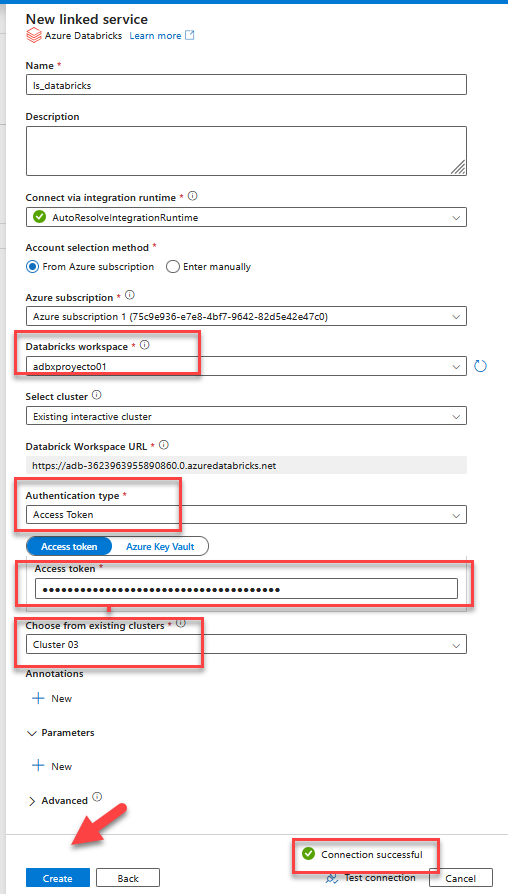
Ingresamos el Workspace.

Seleccionamos “Existing interactive cluster”

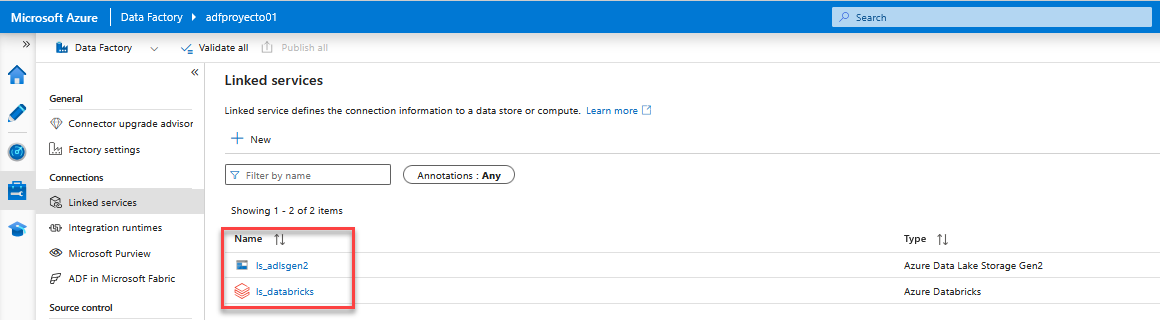
En el tipo de Autenticacion seleccionamos “Access Token” y generamos el Token en Azure Databricks.

Por ultimo probamos la conectividad.

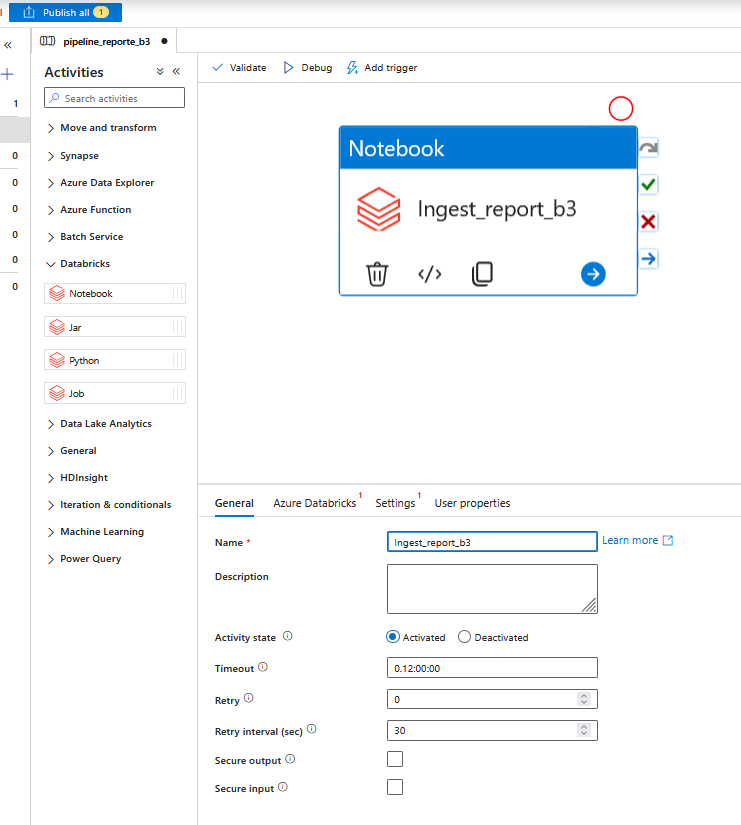




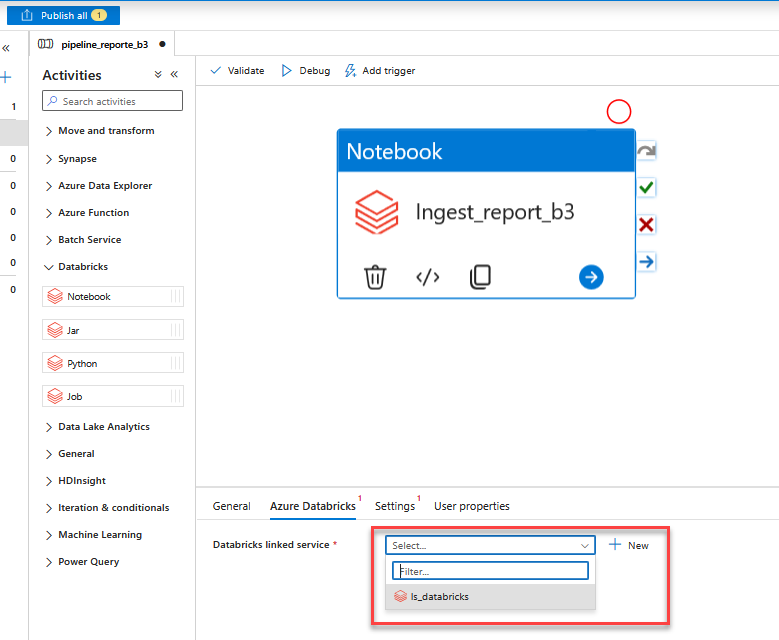
También generamos el Linked Service para el Azure DataFactory.



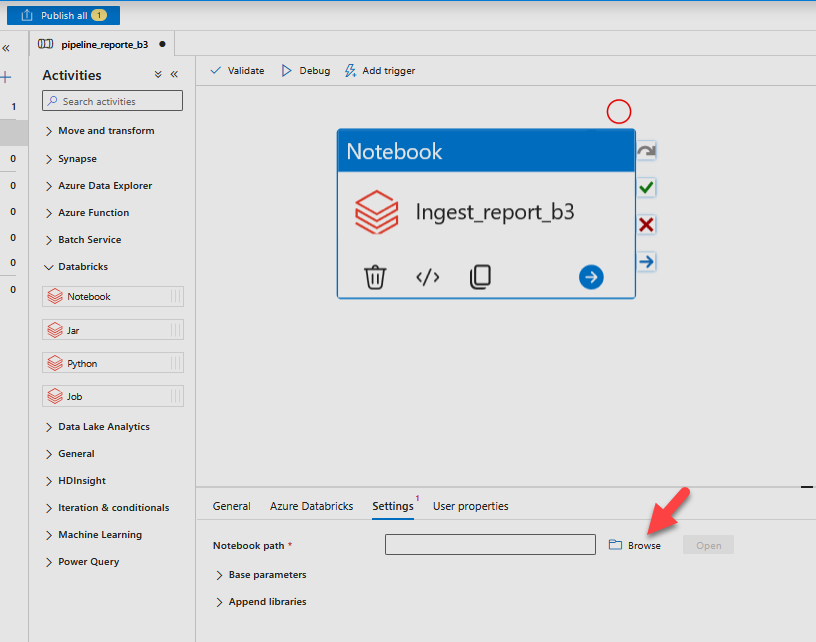
Generamos el Pipeline con los Notebook.

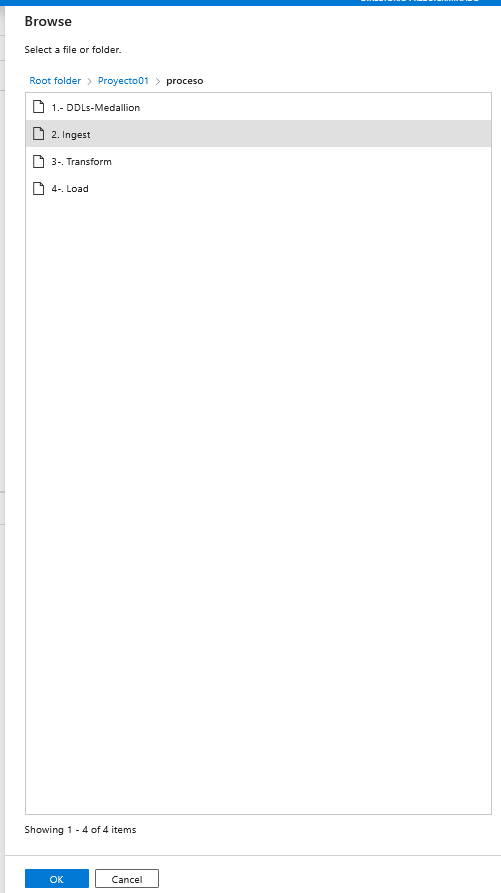


Seleccionamos el Linked Service generado.

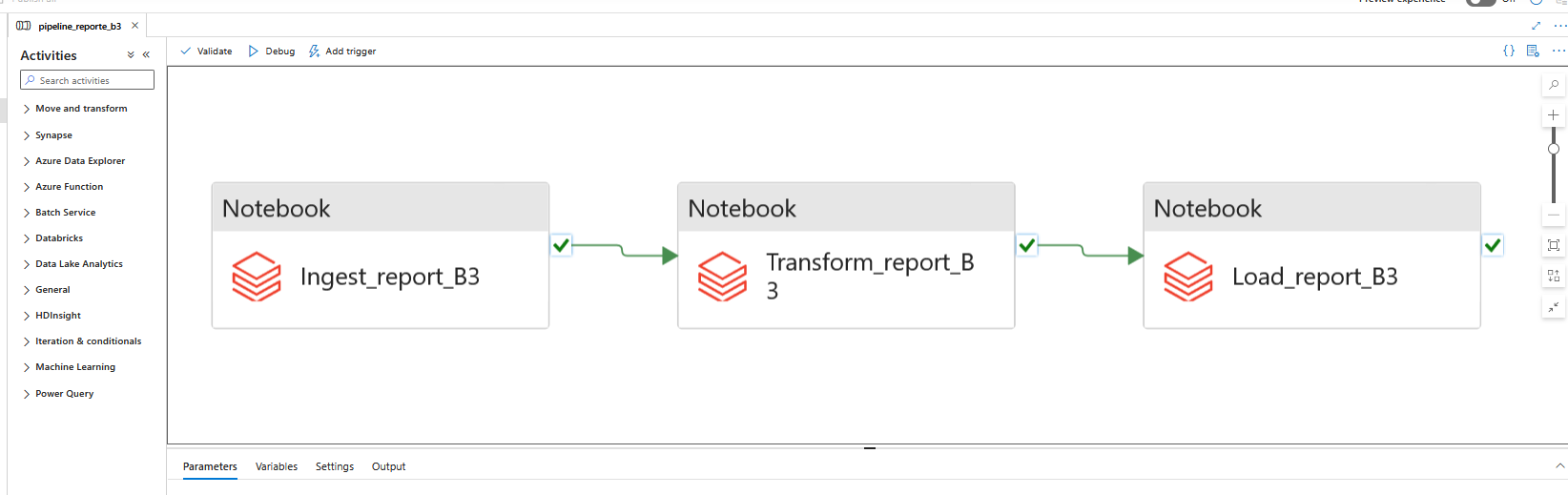


Seleccionamos la ruta del Notebook.



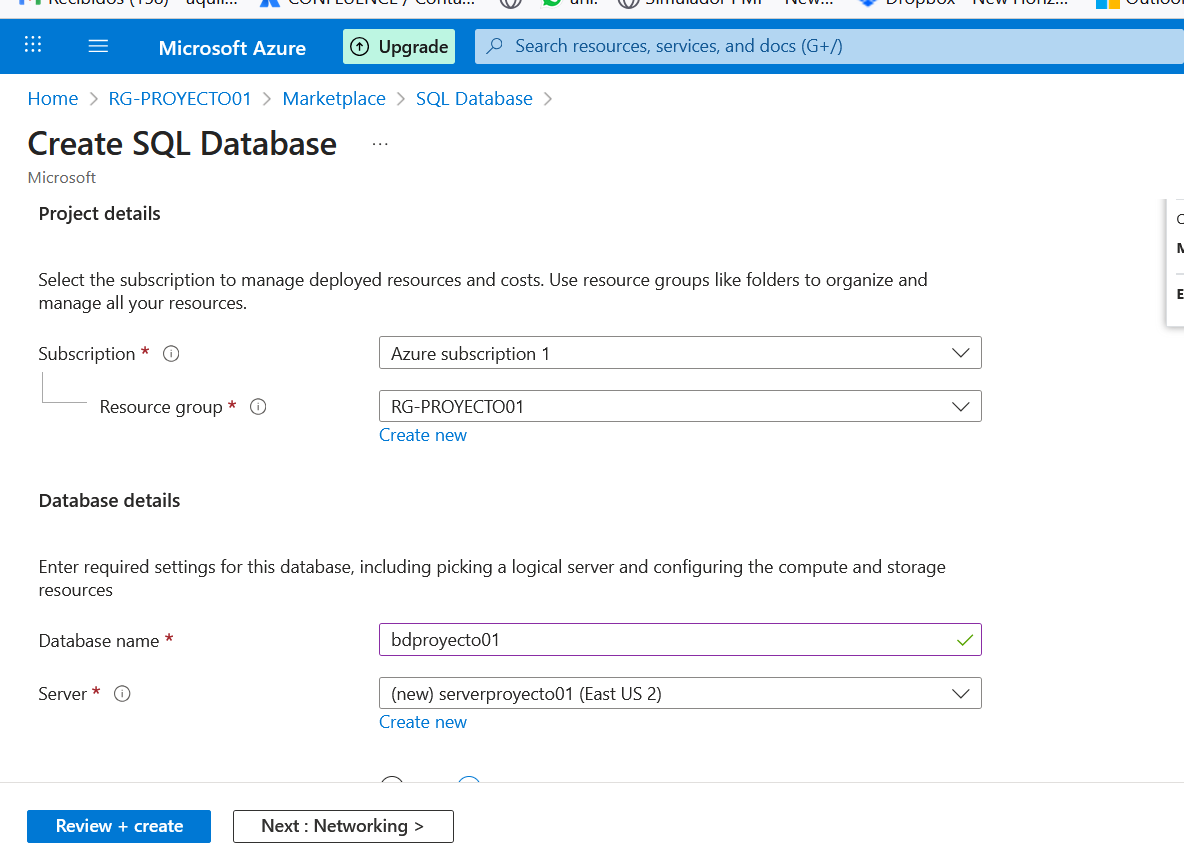


Repetimos los pasos para el resto de los Notebooks.

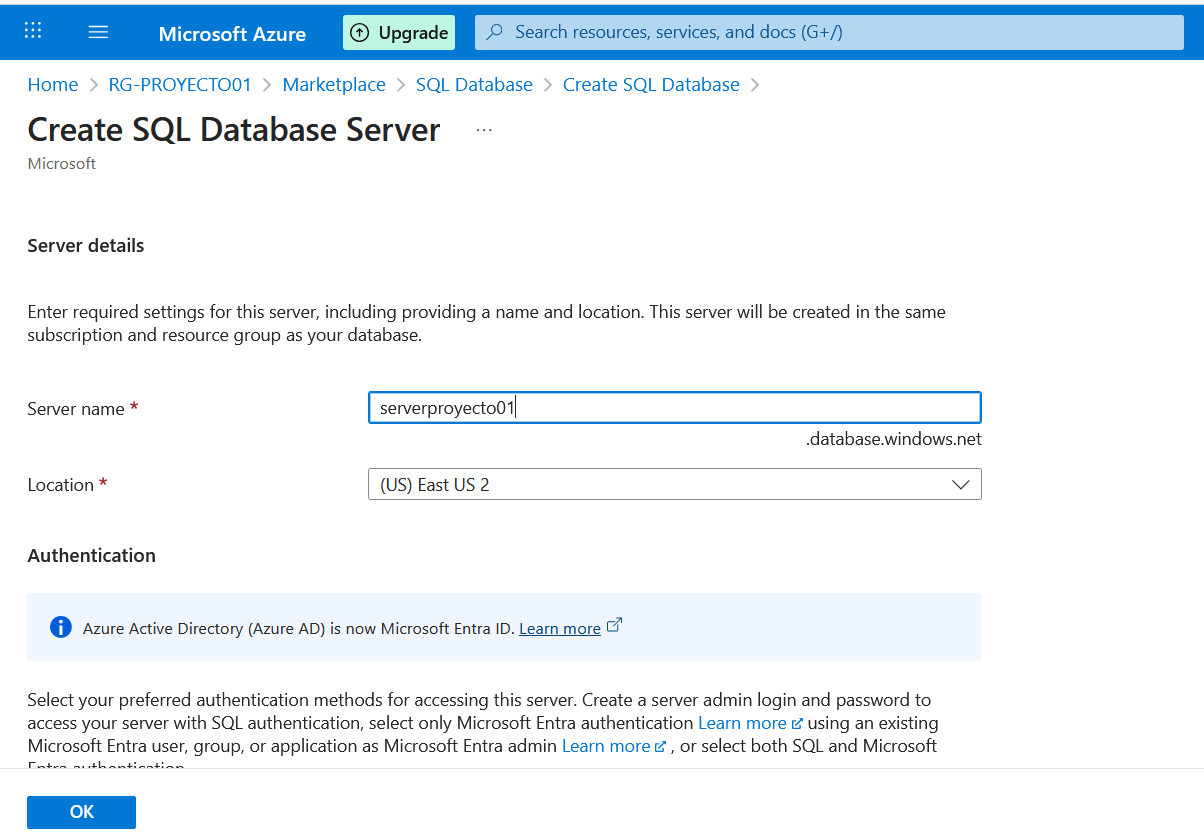


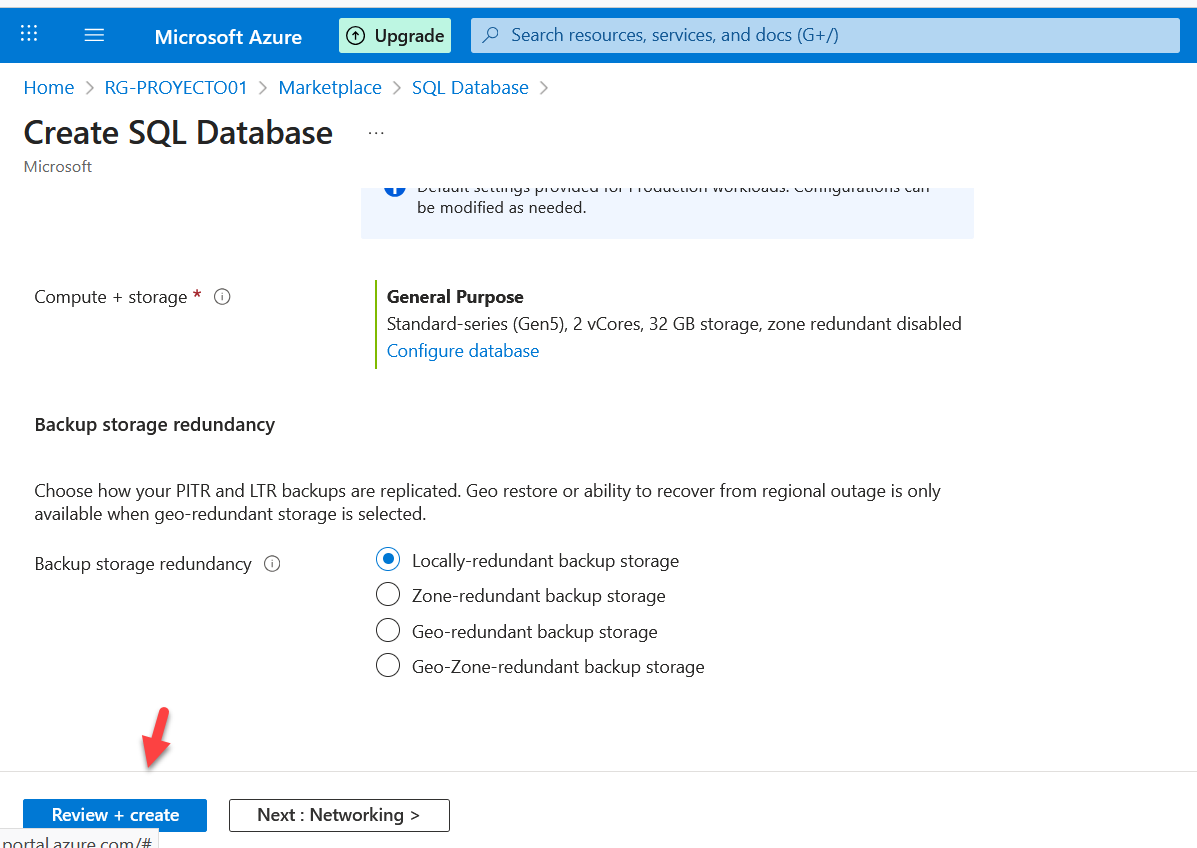
**SERVICIO: Azure SQL database**

Creamos el servicio de Azure SQL database para que consuma la información desde la capa Golden.

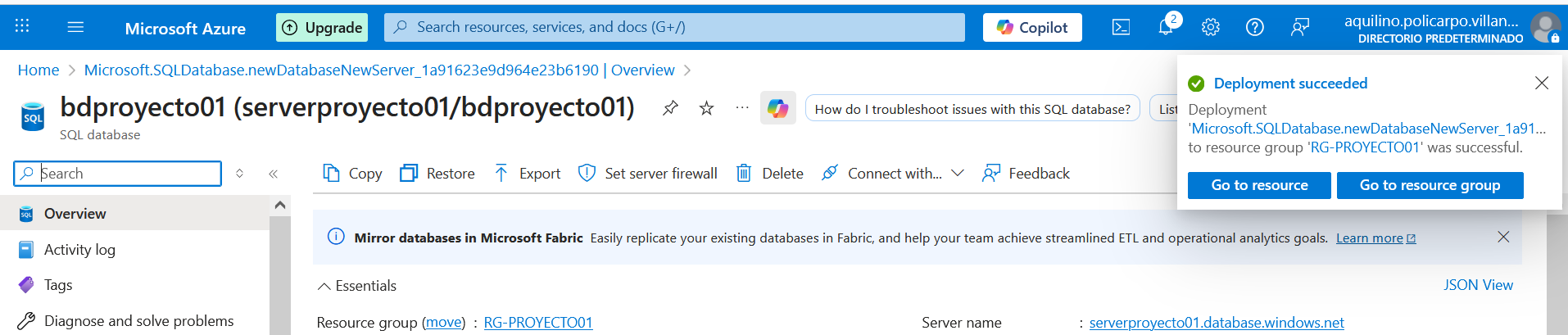


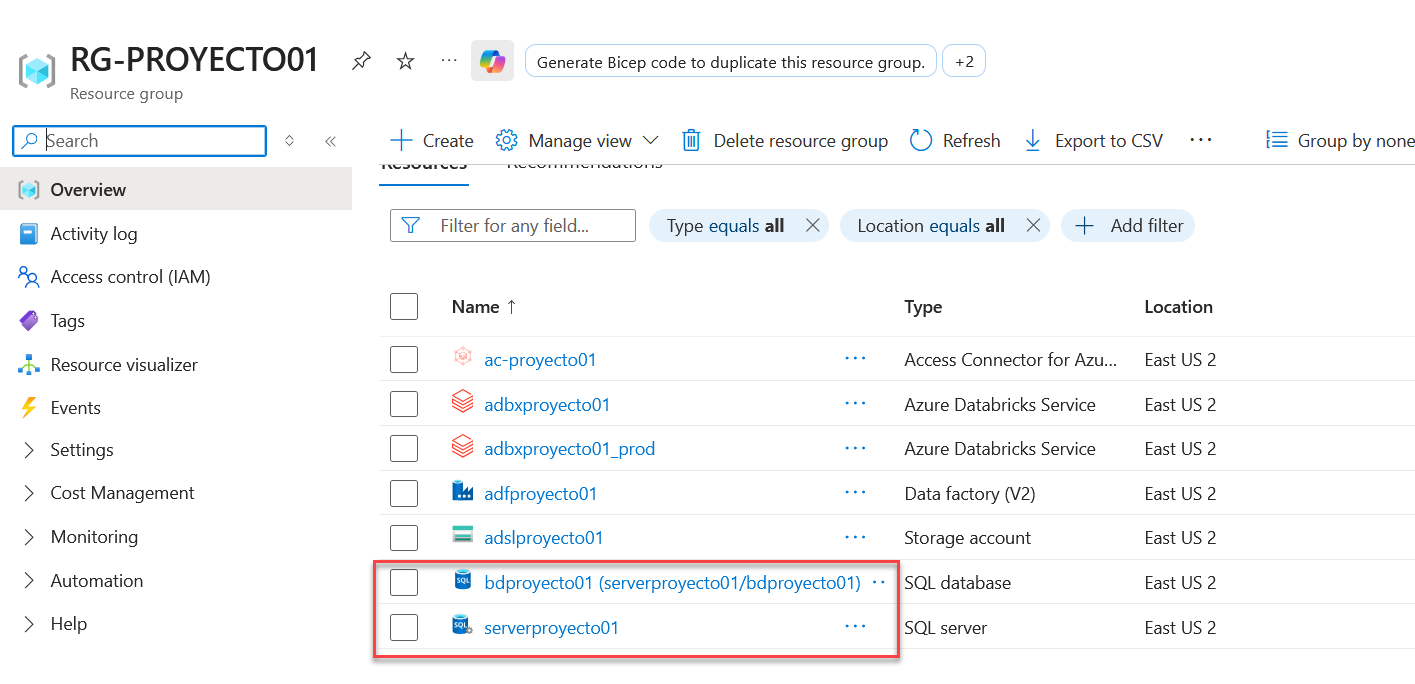
Creamos el Servicio



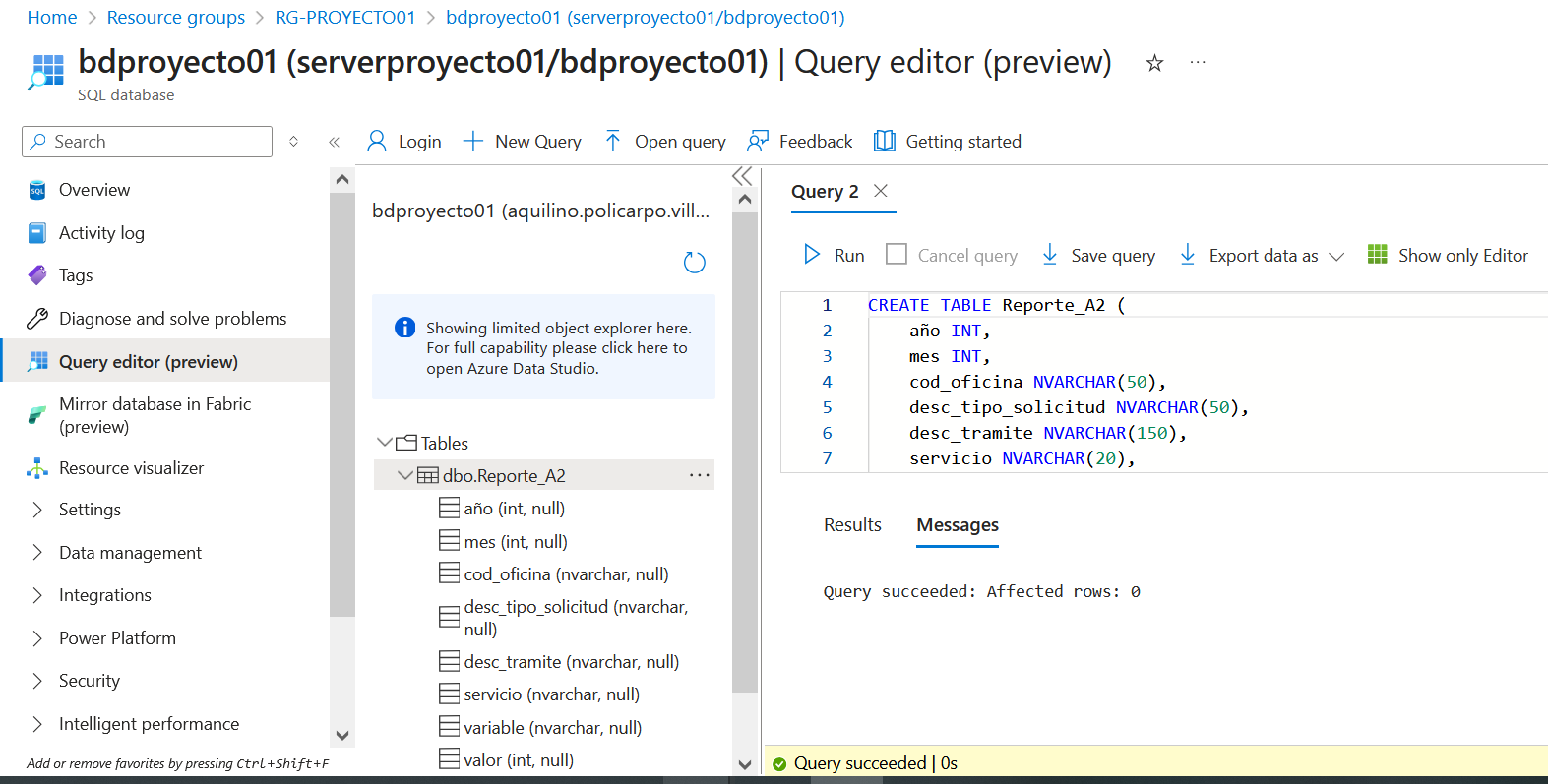
Luego en la Redundancia colocamos como local y damos click en revisar y crear

Creamos el servicio

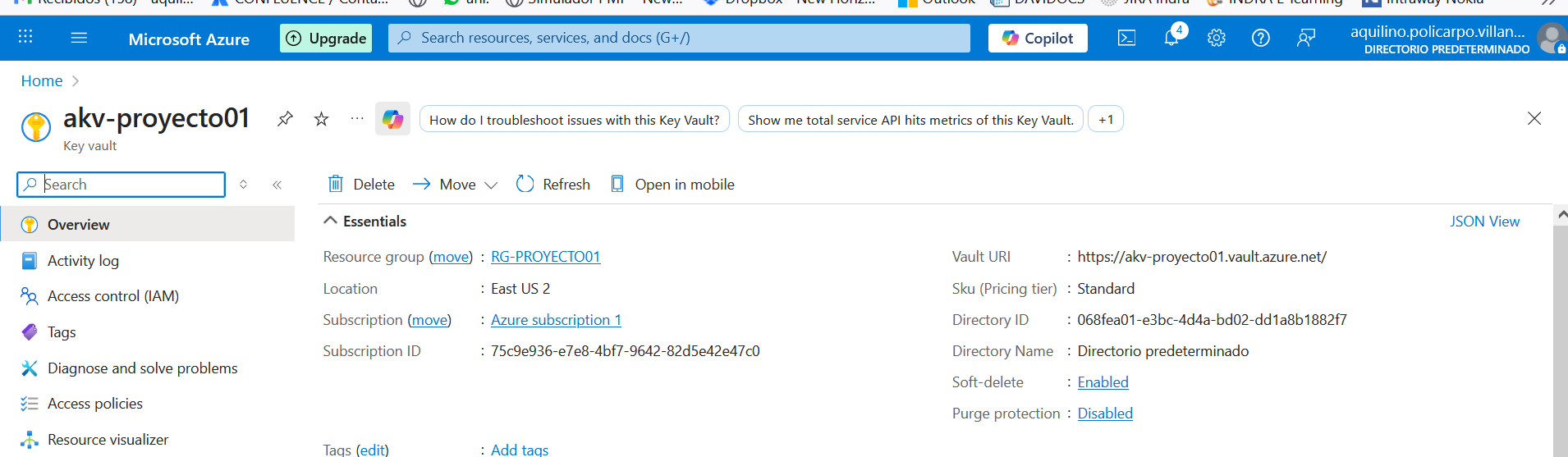




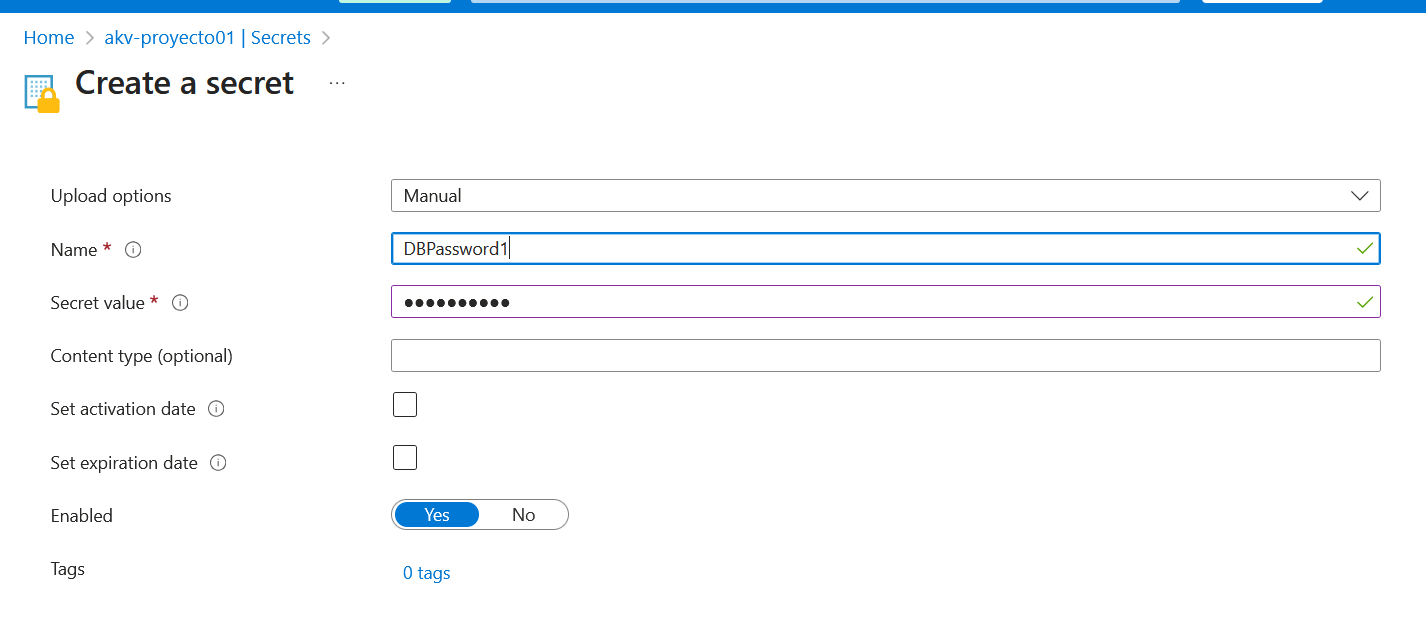
Luego creamos la tabla en la BD creada.



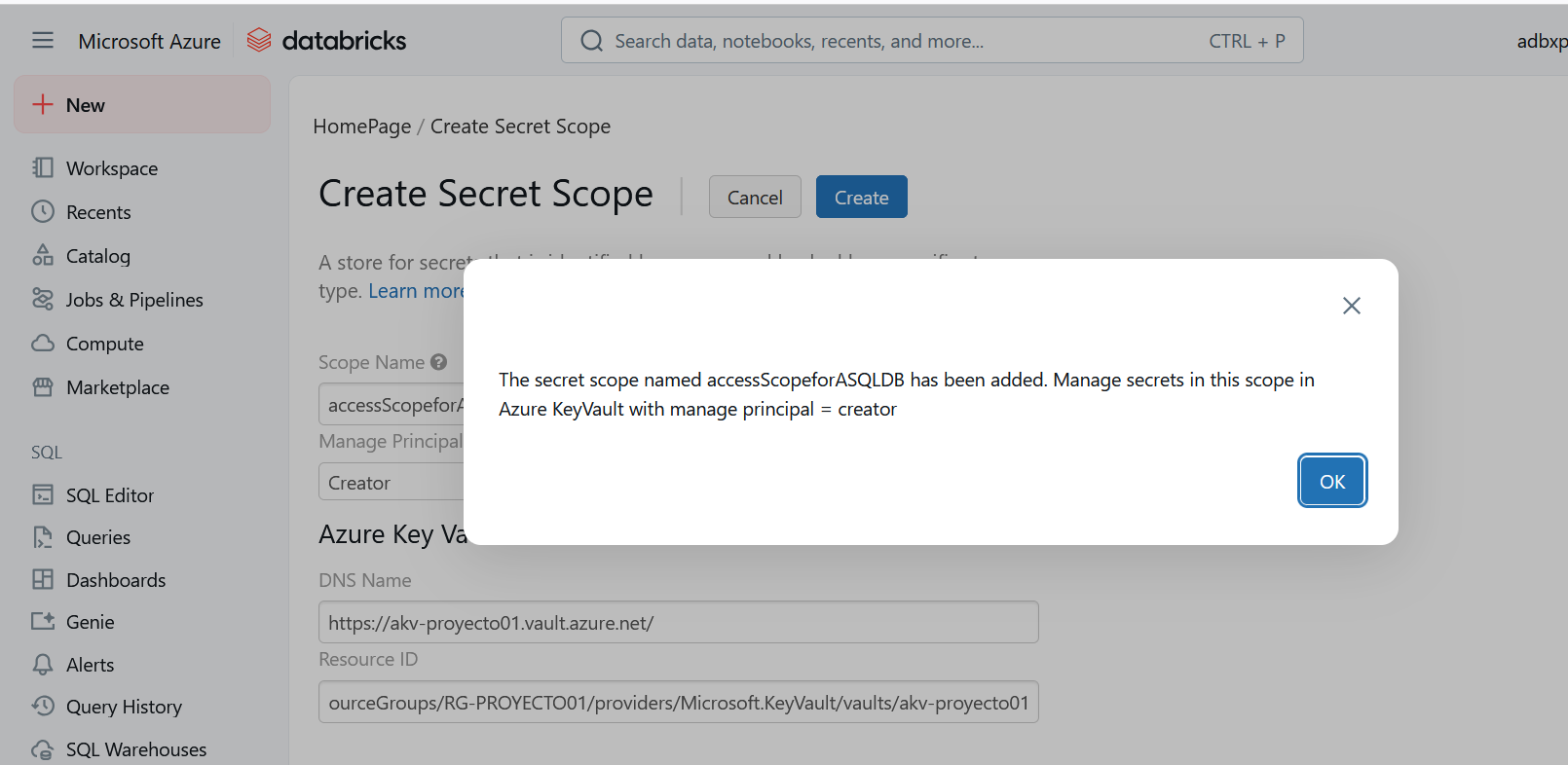
Para realizar la conectividad entre el Databricks y SQL Database, creamos un Key Vault, para ocultar la contraseña de la BD.



Creamos un secreto con la contraseña.



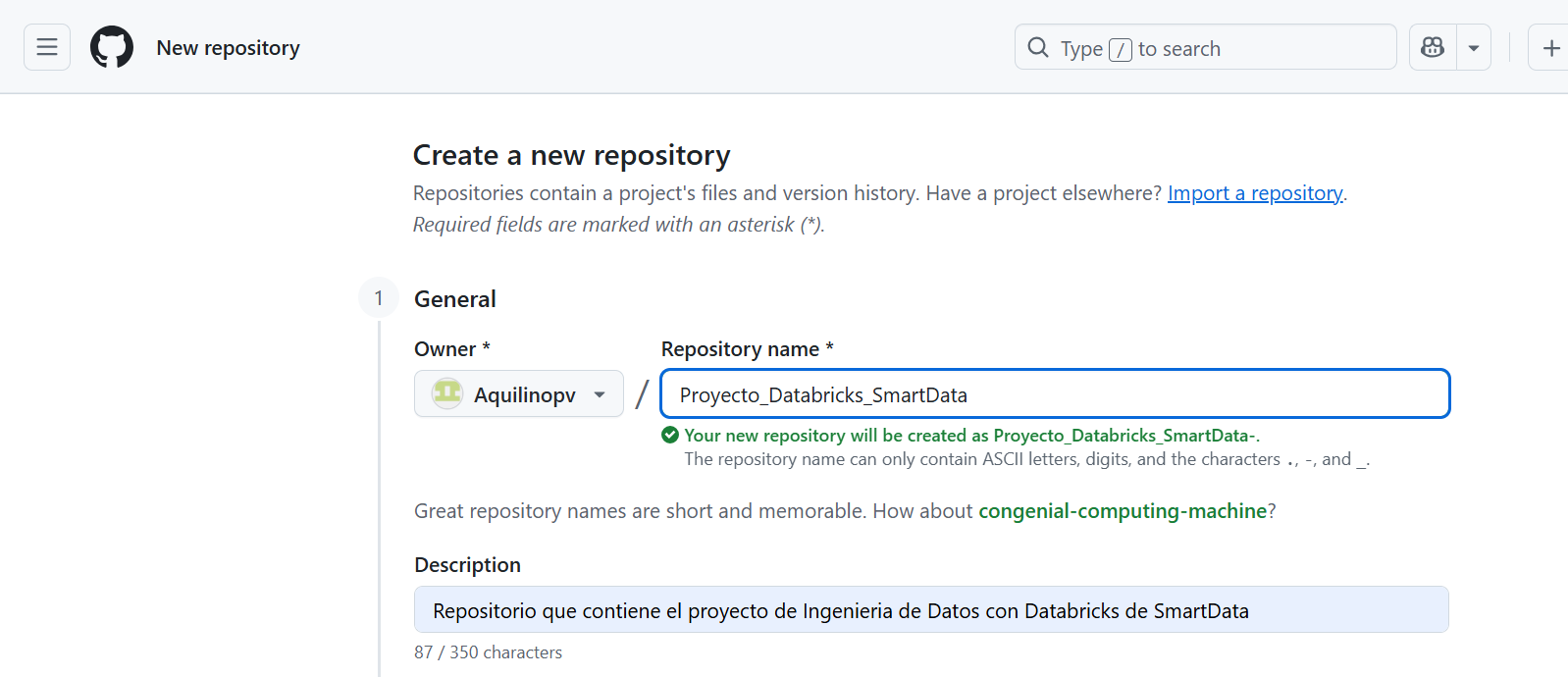
En Databricks, creamos el Secret Scope para poder acceder a los screts creados.

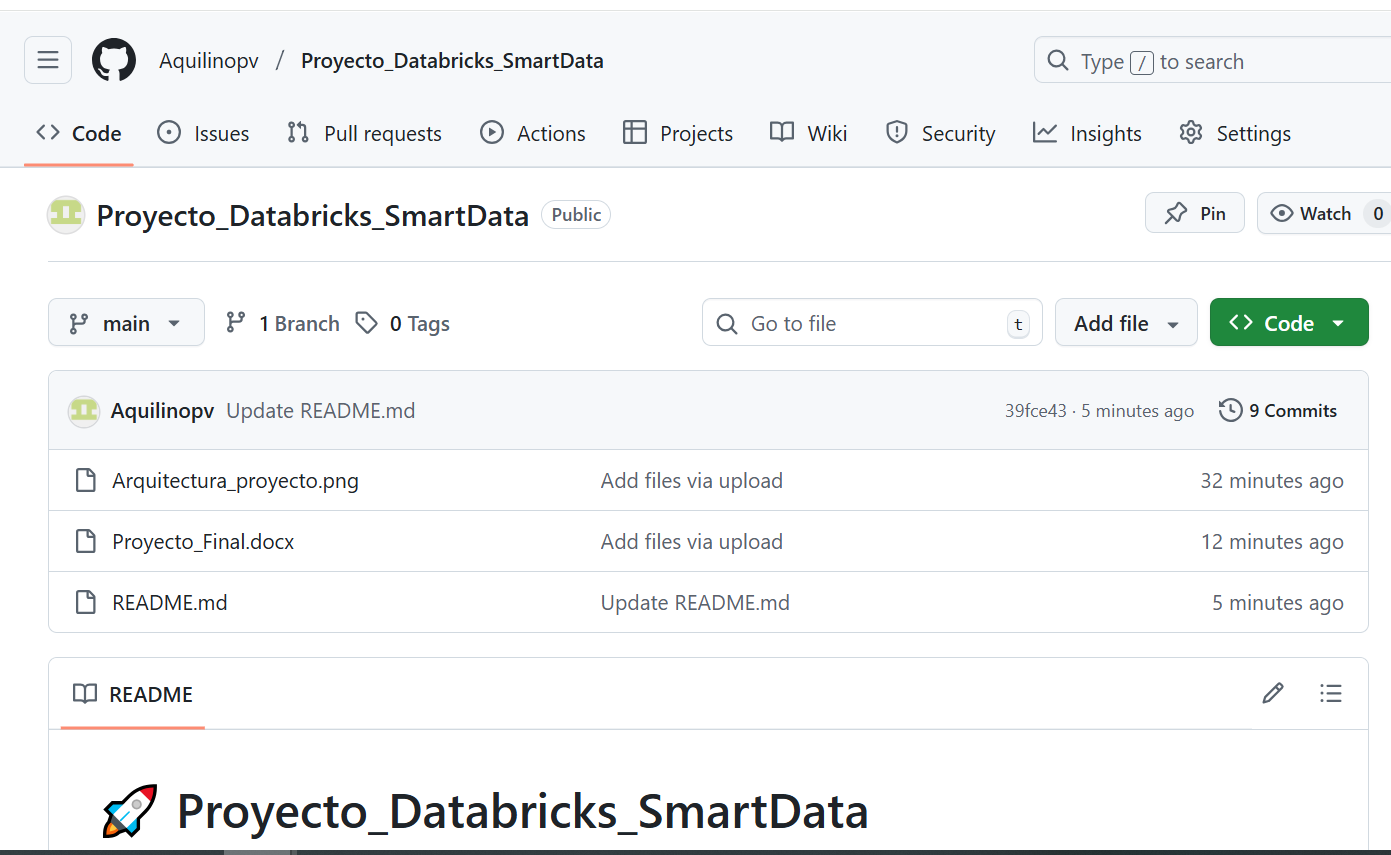


Cuando se ejecute el Proceso, se guarda la información en la tabla **Reporte\_A2** del Azure SQL Database.

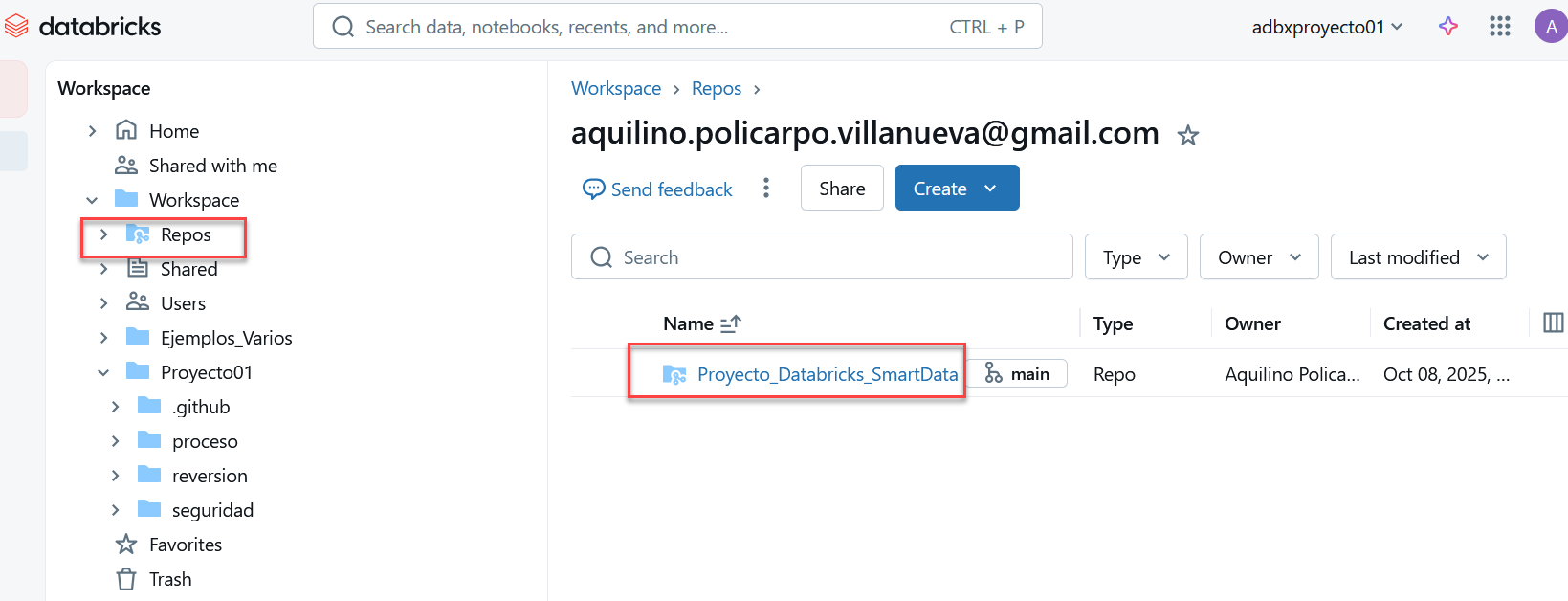
**Configuramos el CI/CD con GitHub.**

Como primer paso Creamos el Repositorio en GitHub.

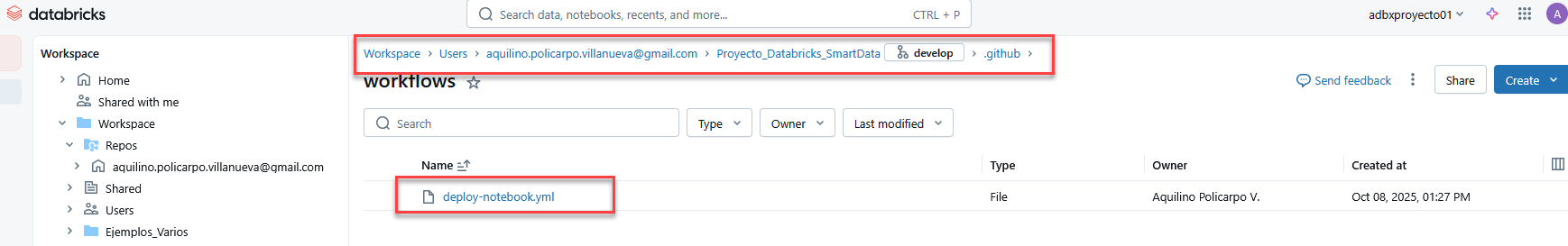




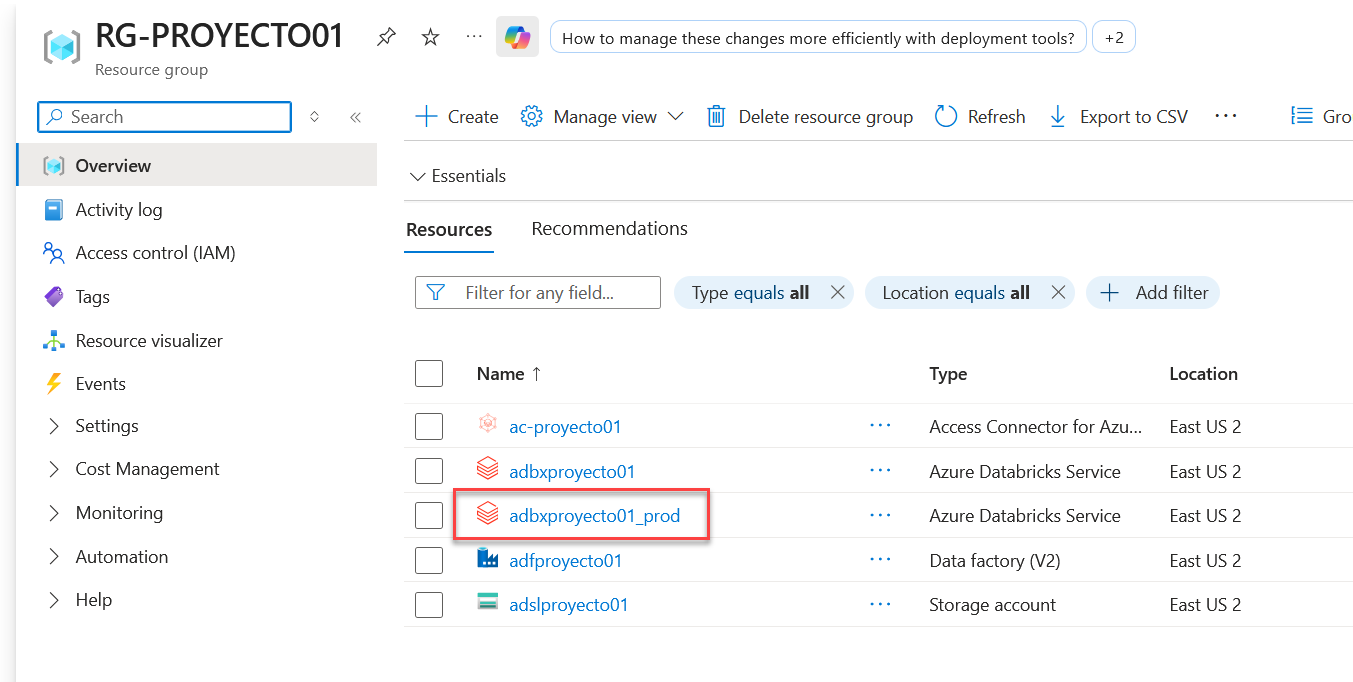
Enlazamos el Repositorio en Databricks.



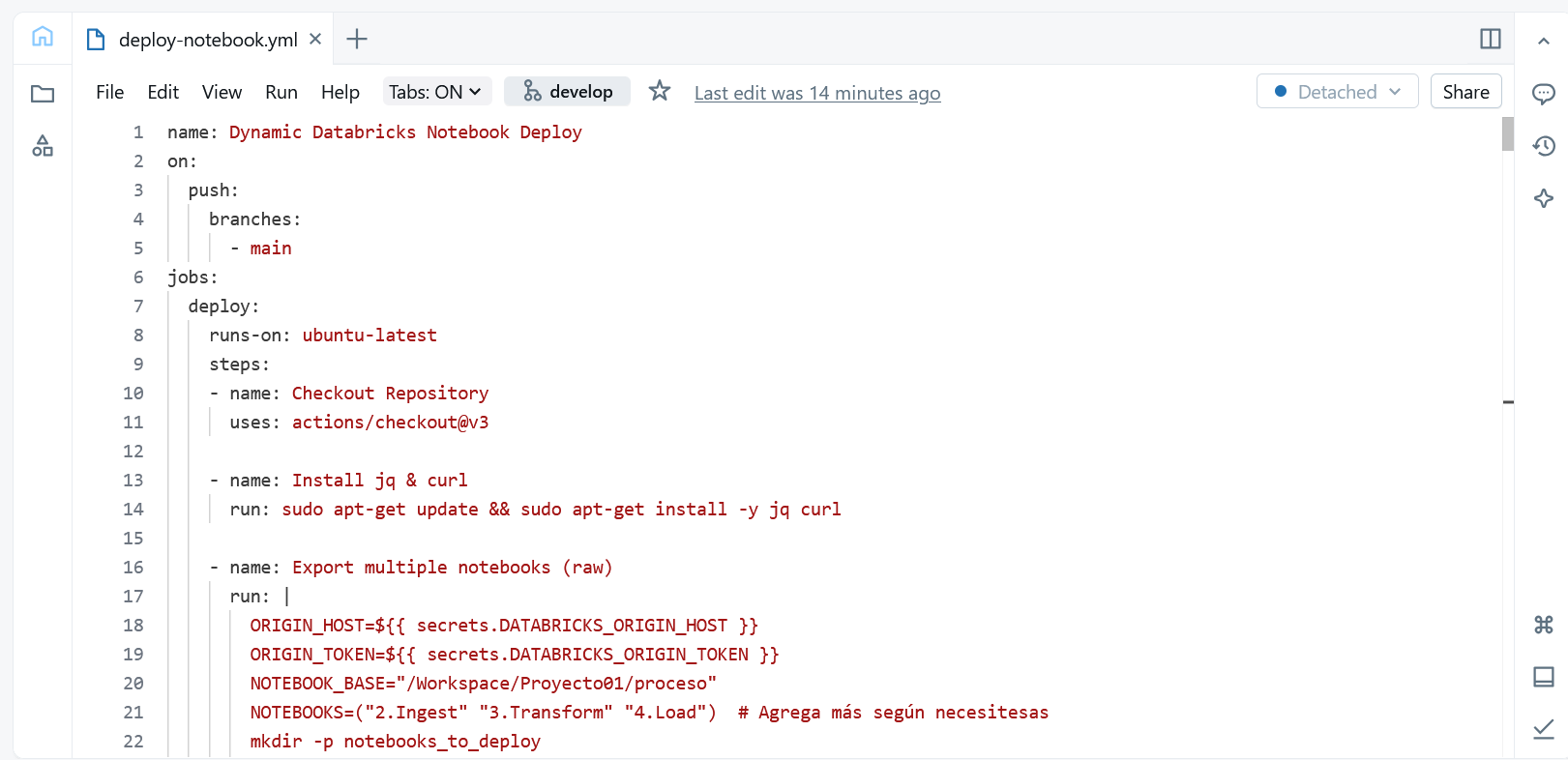
Creamos la carpeta “.gitgub/workflows” y copiamos el archivo deploy-notebook.yml



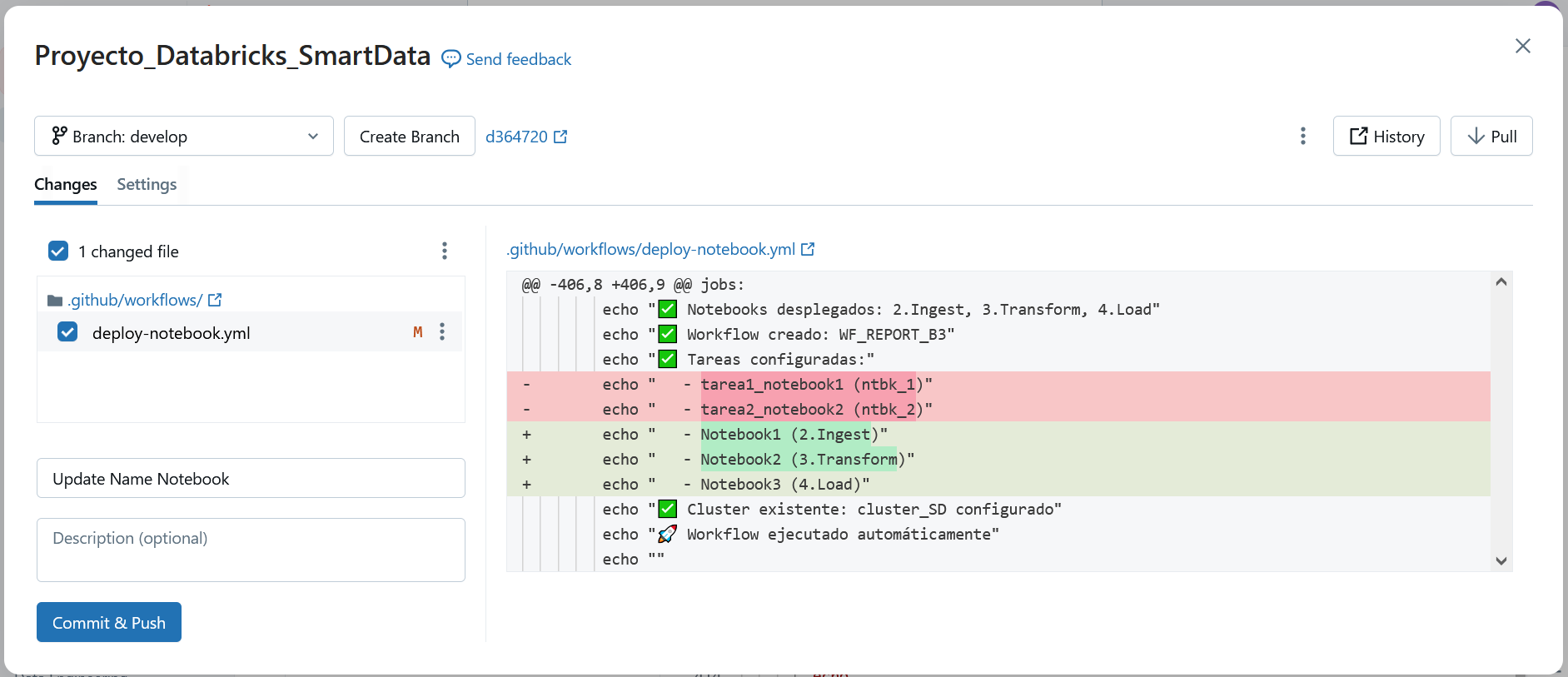
Para probar el despliegue, creamos una nueva instancia en Databricks (adbxproyecto01\_prod).



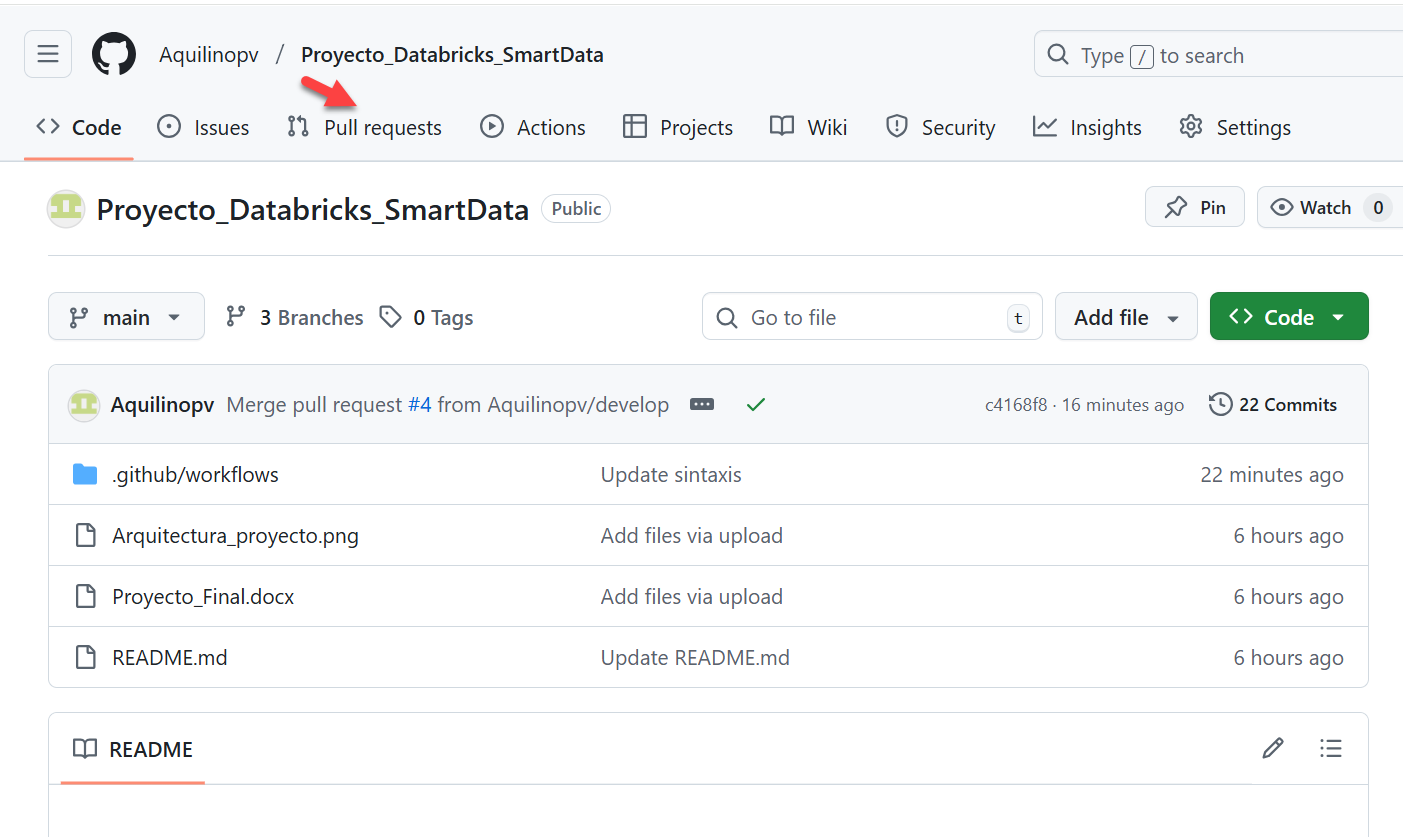
Configuramos el Archivo Yaml para el CD-CI.

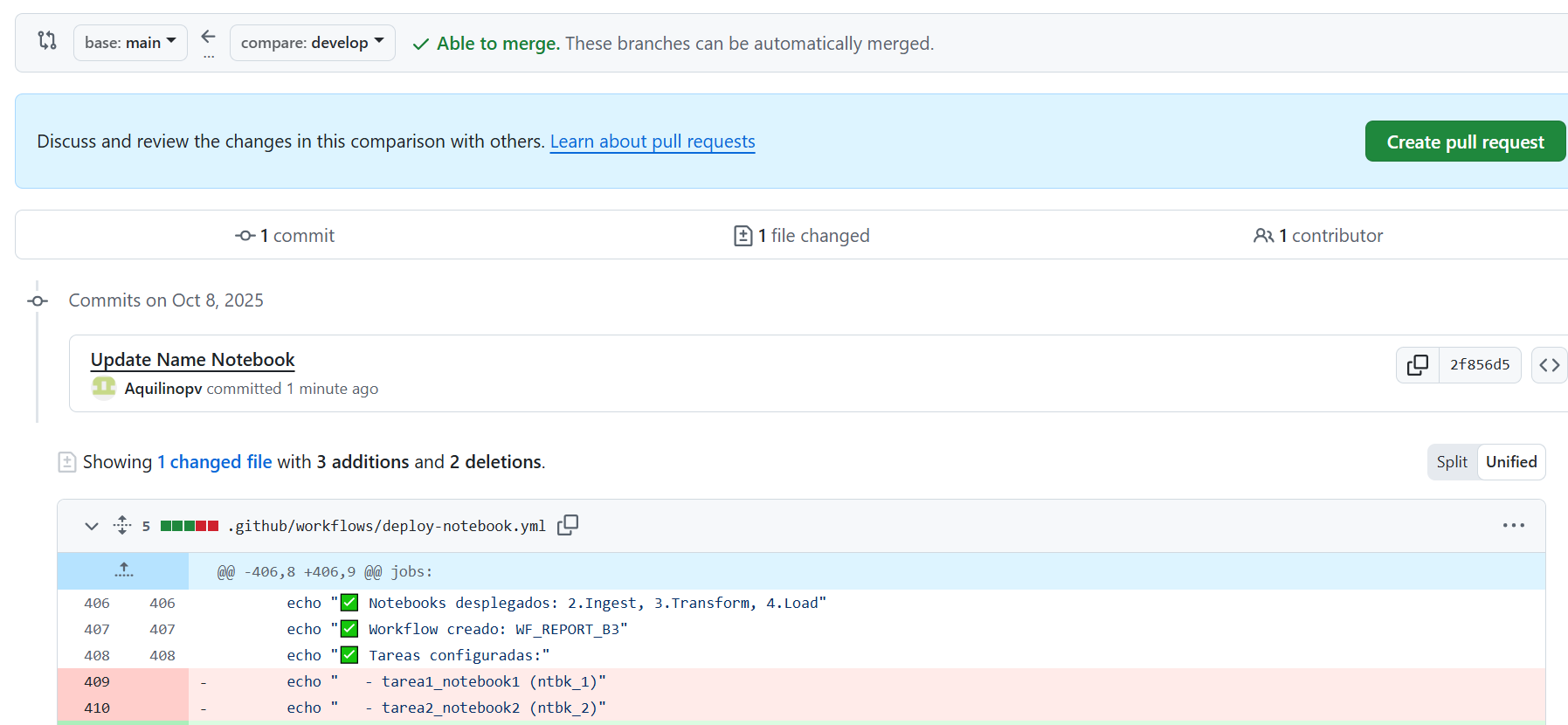


Probamos el “Commit & Push” para probar el despliegue.

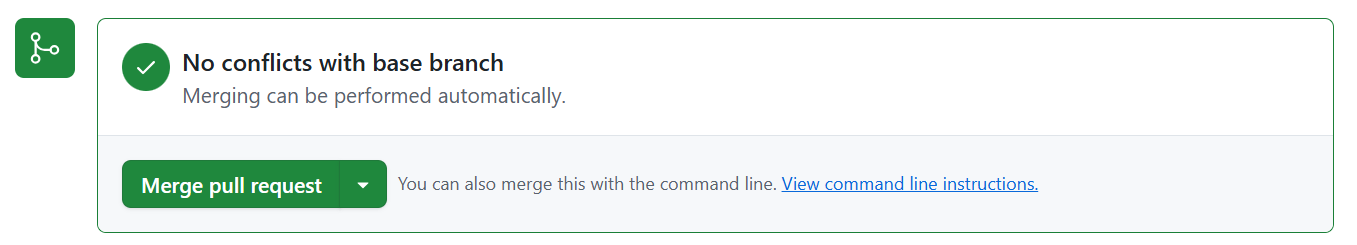


En el repositorio de GitHub hacemos el “pull request”

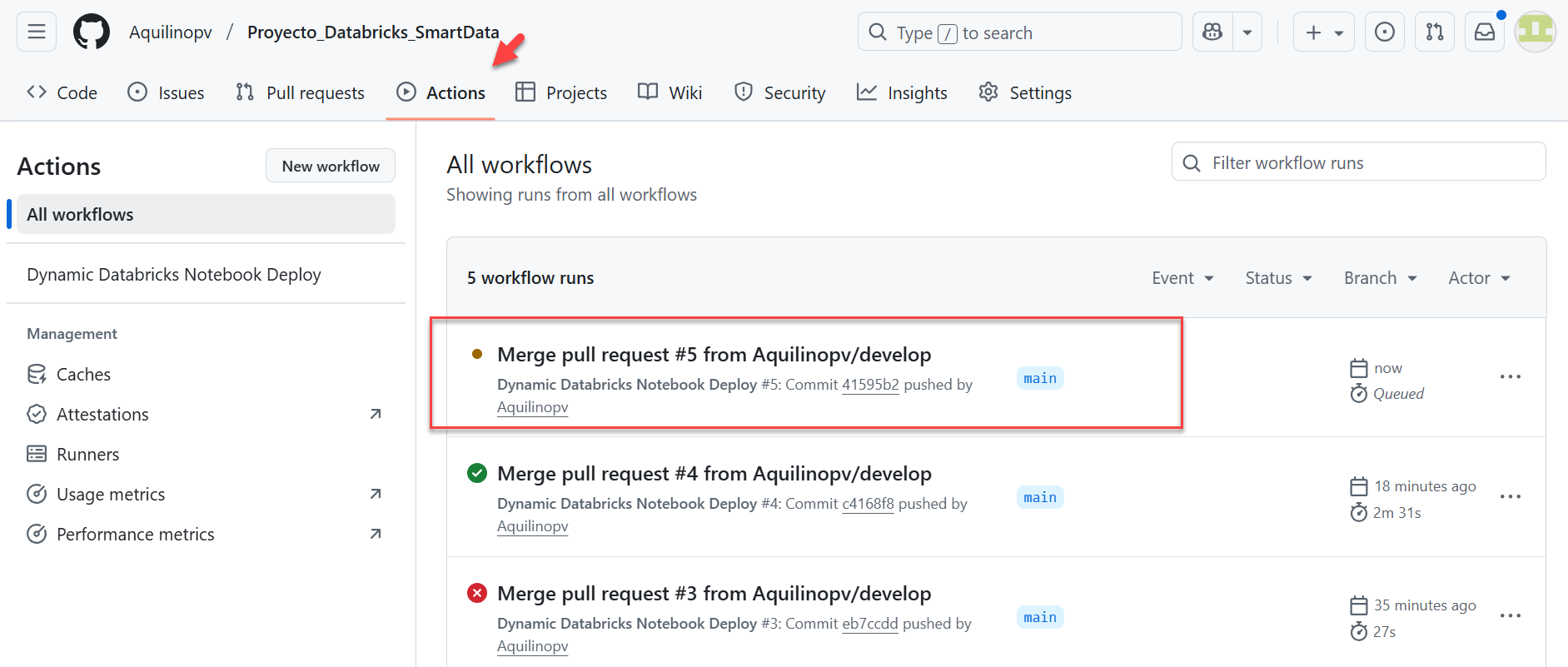




Si no existen conflictos hacemos el “Merge”.



Validamos la ejecución en “Accions”



Validamos si el despliegue se ejecutó correctamente.



Por último, validamos que la tabla en la BD del Servicio SQL Database se ha poblado.

