

GUÍA: RETOS BÁSICOS AZURE DATABRICKS

Axayacatl Valenzuela Faddul

MICROSOFT

INDICE

Ingresando al portal de administración de Azure	2
Creando el grupo de recursos	Error! Bookmark not defined.
Azure Data Lake Storage	2
Crear contenedor.....	5
Subir archivos	6
Generar SAS	6
Databricks	8
Crear Workspace	8
Crear cluster	10
Importar el código	11

INGRESANDO AL PORTAL DE ADMINISTRACIÓN DE AZURE

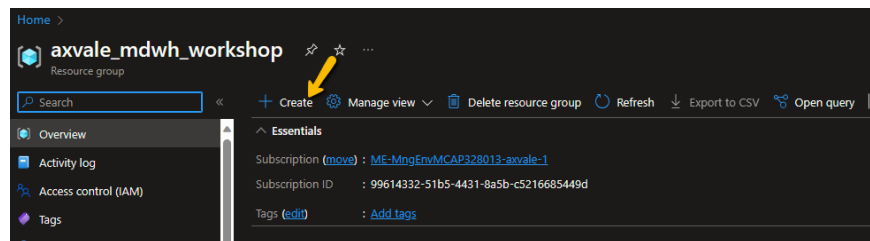
El primer paso para nuestro taller es ingresar al portal de administración de azure, donde crearemos los componentes de nuestra arquitectura.

1. Ingresar a la página <https://portal.azure.com>
2. Ingresar con las credenciales provistas

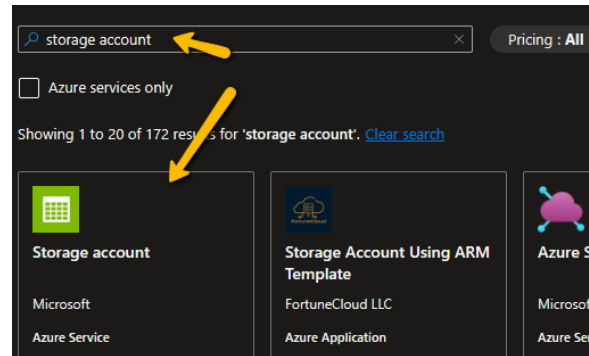
AZURE DATA LAKE STORAGE

Una vez dentro del portal de azure, procederemos a identificar y entrar a nuestro grupo de recursos (pj: dbrikcs_axa.valenzuela)

1. Estando en el grupo de recursos, clic en el botón “Create”



2. En el recuadro de búsqueda, teclear “storage account”. Seleccionar la opción Storage Account, Microsoft, Azure Service y clic en “Create”



3. En la página de creación seleccionar las siguientes opciones
 - a. Resource Group: El grupo de recursos que se creó para el workshop
 - b. Storage Account name: tusinicialesretooals
 - i. Ejemplo: axvaleretooids
 - c. Region: EastUS2
 - d. Performance: Standard
 - e. Redundancy: Locally-redundant storage (LRS)

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription *

Resource group * [Create new](#)

Instance details

Storage account name ⓘ *

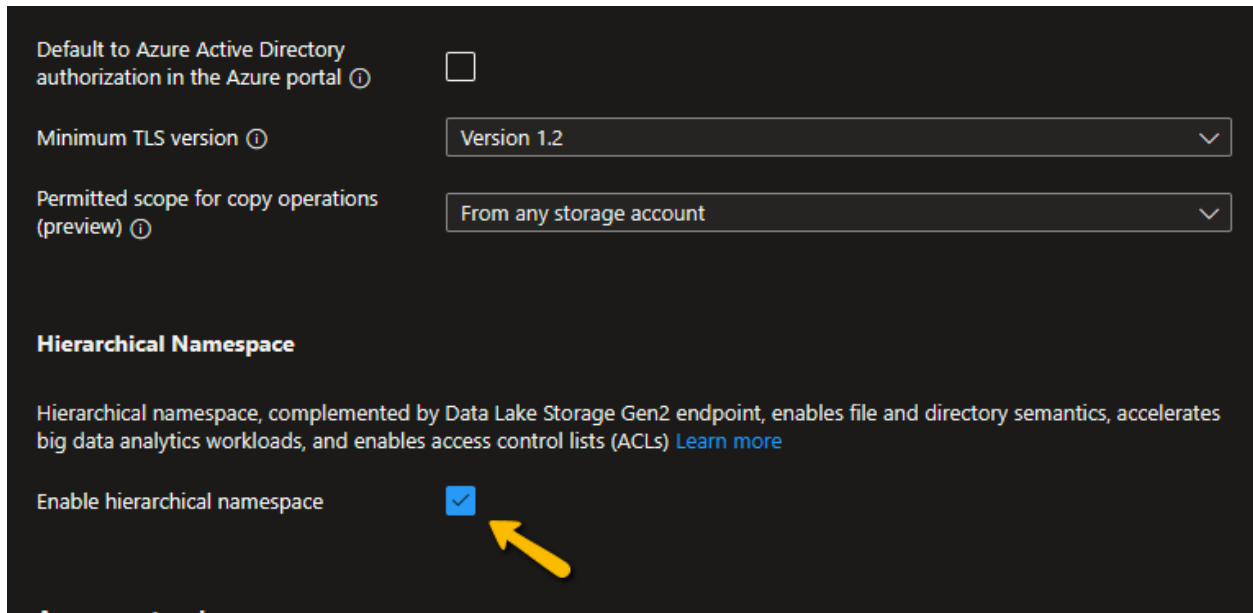
Region ⓘ * [Deploy to an edge zone](#)

Performance ⓘ * ☒ **Standard:** Recommended for most scenarios (general-purpose v2 account)
☐ **Premium:** Recommended for scenarios that require low latency.

Redundancy ⓘ *

4. Clic en el boton “Next: Advanced>”

5. Estando en la pestaña “Advanced” habilitar la casilla “Hierarchical NameSpace”



Default to Azure Active Directory authorization in the Azure portal ⓘ ☐

Minimum TLS version ⓘ Version 1.2 ▼

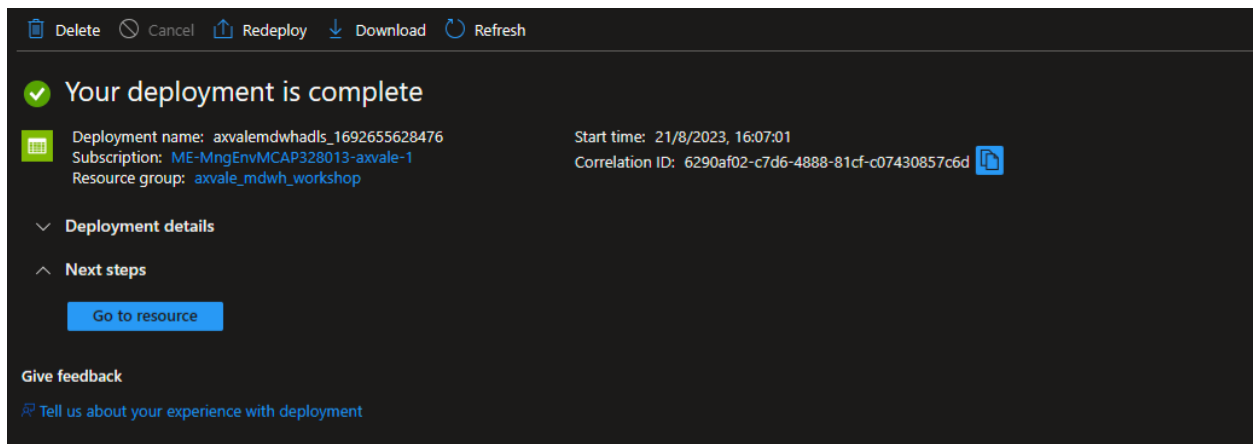
Permitted scope for copy operations (preview) ⓘ From any storage account ▼

Hierarchical Namespace

Hierarchical namespace, complemented by Data Lake Storage Gen2 endpoint, enables file and directory semantics, accelerates big data analytics workloads, and enables access control lists (ACLs) [Learn more](#)

Enable hierarchical namespace ☒

6. Clic en el botón “Review + Create”
7. Clic en el botón “Create”
8. Esperar a que el proceso de creación termine y dar clic en el botón “Go to Resource”



Delete Cancel Redeploy Download Refresh

✓ Your deployment is complete

Deployment name: axvalemmdwhadls_1692655628476
Subscription: ME-MngEnvMCAP328013-axvale-1
Resource group: axvale_mdwh_workshop

Start time: 21/8/2023, 16:07:01
Correlation ID: 6290af02-c7d6-4888-81cf-c07430857c6d

Deployment details

Next steps

Go to resource

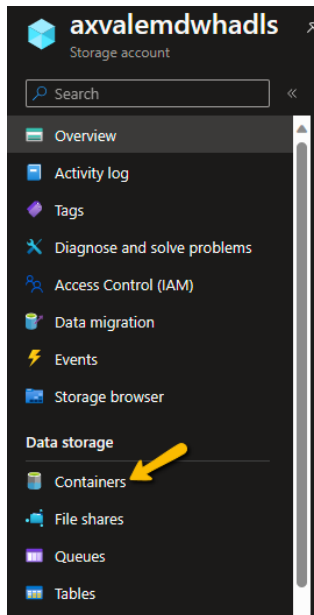
Give feedback

Tell us about your experience with deployment

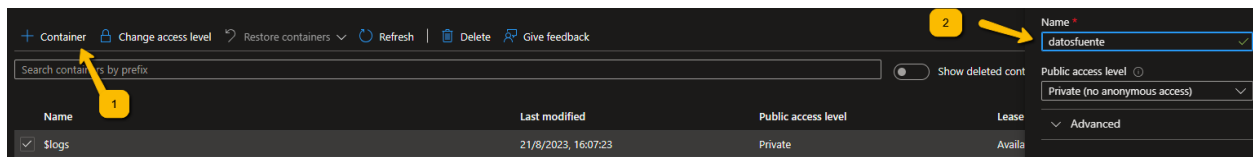
CREAR CONTENEDOR

Estando en la página de administración de la cuenta de almacenamiento

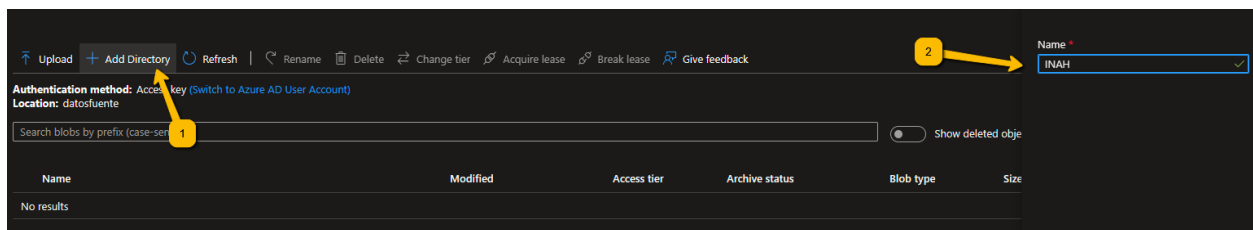
1. Clic en el botón “Containers”, ubicado en el menú de extrema izquierda



2. En el menú del centro
 - a. Clic en el botón “+ Container”
 - b. Nombre: datosfuente
 - c. Public Access level: Private
 - d. Clic en el botón “Create”



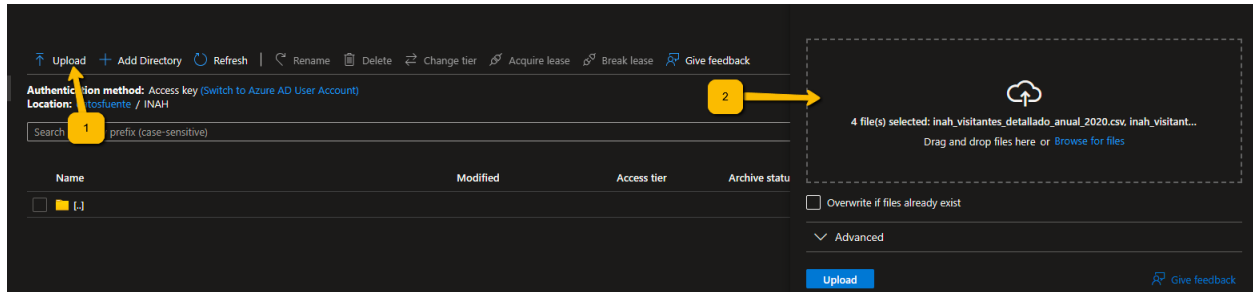
3. Una vez creado el contenedor
 - a. Seleccionar el contenedor
 - b. Clic en el botón “+Add Directory”
 - c. Crear el directorio “INAH”



SUBIR ARCHIVOS

En este paso vamos a subir los archivos de datos al storage Account. Lo primero que tienes que hacer es ubicar la carpeta donde clonaste el repositorio de github

1. Regresa al portal de Azure y en la carpeta que acabamos de crear, da clic en el botón Upload.
2. Da clic en “Browse files” y selecciona los cuatro archivos que están en la carpeta Data/INAH/inah_visitantes en el repositorio clonado
3. Da clic en el botón Upload

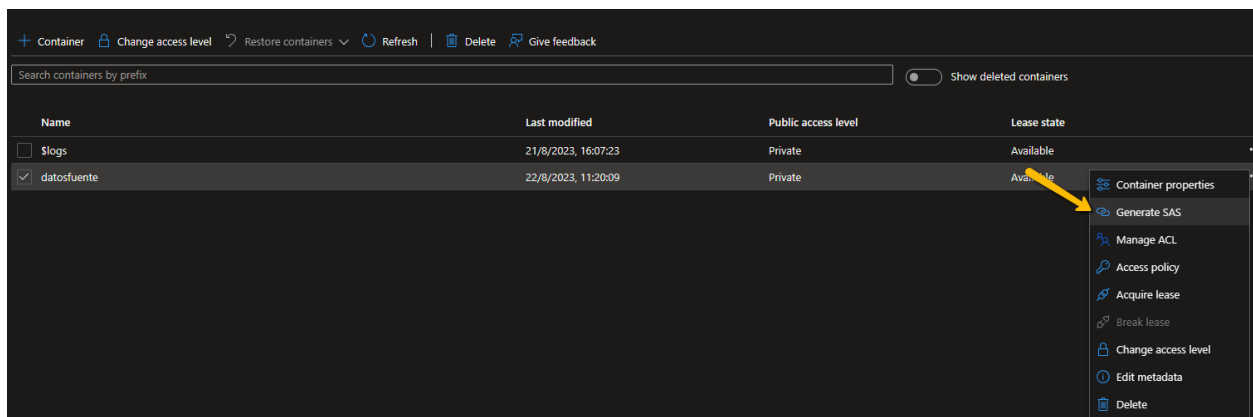


GENERAR SAS

Ahora que has subido los archivos, es necesario otorgar acceso a la carpeta. Para eso utilizaremos una forma de autenticación llamada SAS (Shared Access Sign).

SAS es una forma rápida de otorgar acceso a un usuario, pero no es la más segura. La utilizaremos únicamente para este taller.

1. Ubica el contenedor datos fuente y da clic en el botón con los tres puntos (...) en el extremo derecho
2. Ahora da clic en el botón Generate SAS



3. En la pantalla de generación de SAS, seleccionas las siguientes opciones:
 - a. Signing method: Account key
 - b. Permissions:
 - i. Read
 - ii. Create
 - iii. Write
 - iv. List
 - v. Execute
 - c. Clic en “Generate SAS”. Guarda la llave generada, la ocuparemos después

Generate SAS

A shared access signature (SAS) is a URI that grants restricted access to an Azure Storage container. Use it when you want to grant access to storage account resources for a specific time range without sharing your storage account key. [Learn more about creating an account SAS](#)

Signing method

☒ Account key ☐ User delegation key

Signing key ⓘ

Key 1

Stored access policy

None

Permissions * ⓘ

2 selected

- ☒ Read
- ☐ Add
- ☐ Create
- ☐ Write
- ☐ Delete
- ☒ List
- ☐ Move
- ☐ Execute
- ☐ Ownership
- ☐ Permissions

11:24:47

Mexico City, Monterrey

19:24:47

Mexico City, Monterrey

or 168.1.5.65-168.1....

HTTPS and HTTP

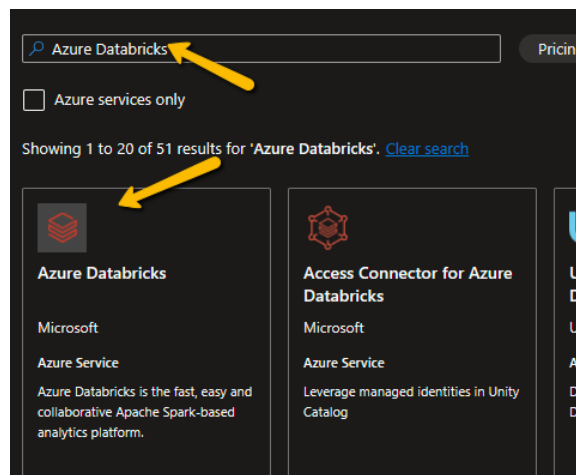
Generate SAS token and URL

DATABRICKS

Si ya tienes un workspace y un cluster, puedes ir directamente a la sección de importación de código

CREAR WORKSPACE

1. Regresa al grupo de recursos y da clic en el botón “+ Create”
2. En recuadro de búsqueda escribe: Databricks
3. De las opciones disponibles selecciona la primera



4. En la pantalla de creación selecciona las siguientes opciones:
 - a. Workspace name: <tusiniciales>_databricks
 - b. Region: East US
 - c. Pricing Tier: Trial
 - d. Clic en “Review + Create”
 - e. Después de la validación, clic en “Create”



Home > axvale_mdwh_workshop > Marketplace > Azure Databricks >



Create an Azure Databricks workspace

Basics Networking Encryption Tags Review + create


Project Details


Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.



Subscription *  ME-MngEnvMCAP328013-axvale-1 

Resource group *  axvale_mdwh_workshop 
[Create new](#)

Instance Details

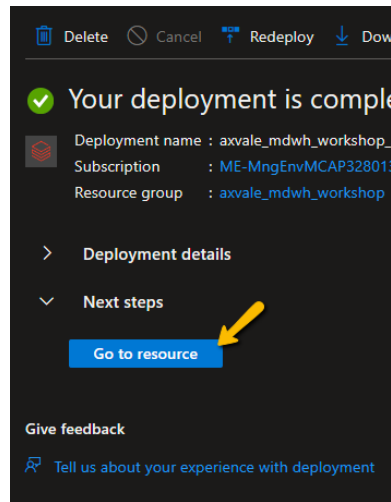
Workspace name * axvale_mdwh_dbricks 

Region * East US 

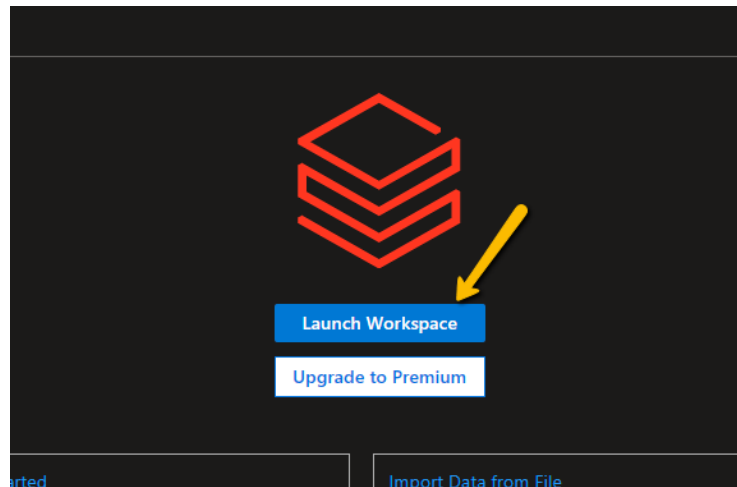
Pricing Tier *  Trial (Premium - 14-Days Free DBUs) 

Managed Resource Group name Enter name for managed resource group

- f. Cuando el proceso de creación termine, da clic en “Go to Resource”

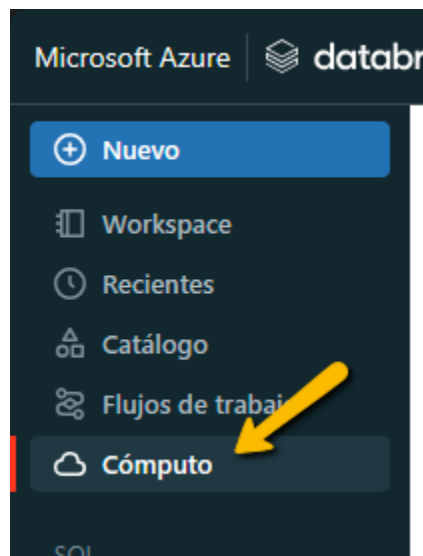


- g. En la pantalla del workspace da clic en “Launch Workspace”



CREAR CLUSTER

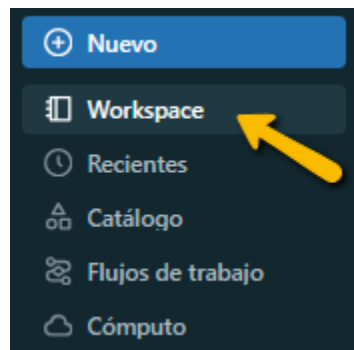
1. Estando en workspace de Databricks, clic en el botón “Cómputo” ubicado en el mené de extrema derecha



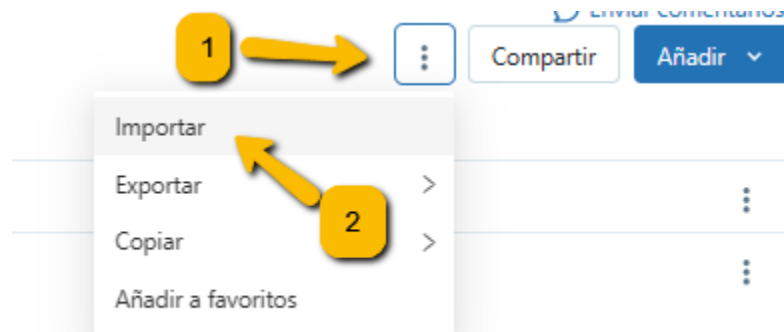
2. En la esquina derecha dar clic en “Crear Computo”
3. En la pantalla de creación de cluster seleccionar las siguientes opciones
 - a. Directriz: Sin restricciones
 - b. Multi-nodo
 - c. Runtime: 13.3 LTS
 - d. Deshabilitar aceleración Photon
 - e. Tipo de worker: Standard_D4a_V4
 - f. Workers:2
 - g. Tipo de driver: El mismo que el worker
 - h. Deshabilitar auto expansión
 - i. Habilitar terminar después de 120 minutos de actividad
4. Clic en crearImportar el código

IMPORTAR EL CÓDIGO

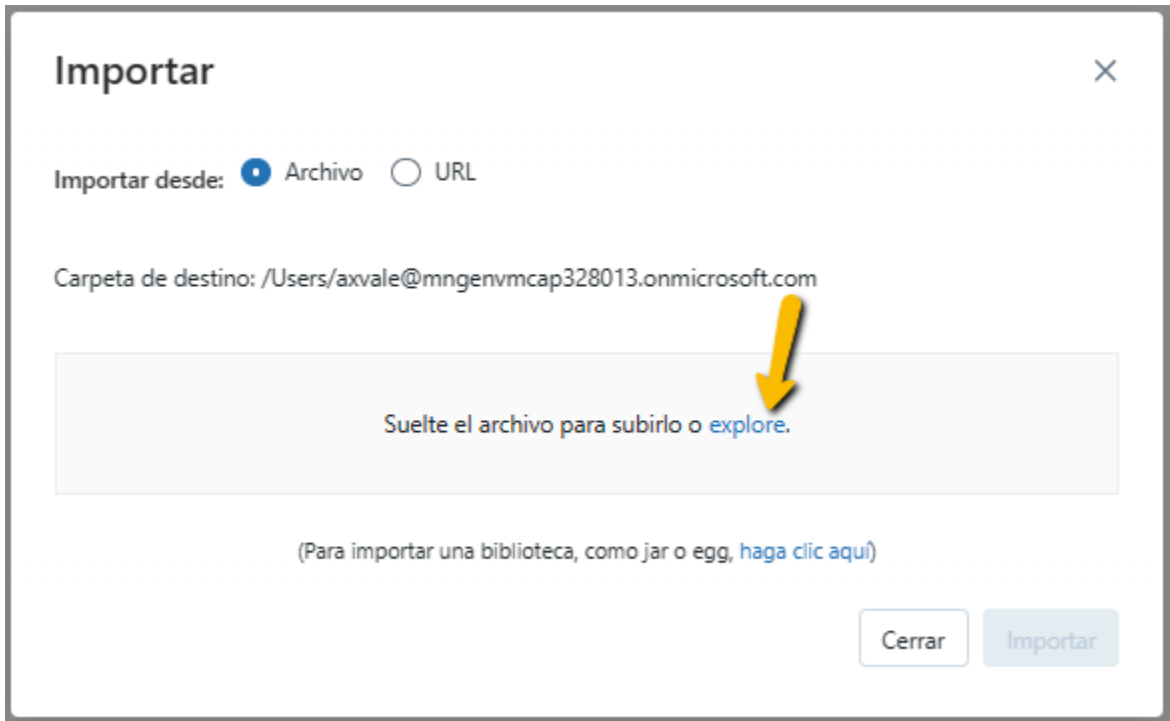
1. En el menú de extrema izquierda clic en “Workspace”



2. En el explorador del workspace, expande las carpetas Workspace>Users hasta encontrar la carpeta de tu usuario
3. Selecciona la carpeta de tu usuario y da clic en el botón de los tres puntos y después selecciona importar



4. En la ventana emergente da clic en explorar y selecciona el archivo en la carpeta Retos Databricks/RetosMedallon.dbc



5. Clic en importar

Al terminar de importar sigue las instrucciones del notebook para completar los retos