



BIG DATA
ACADEMY

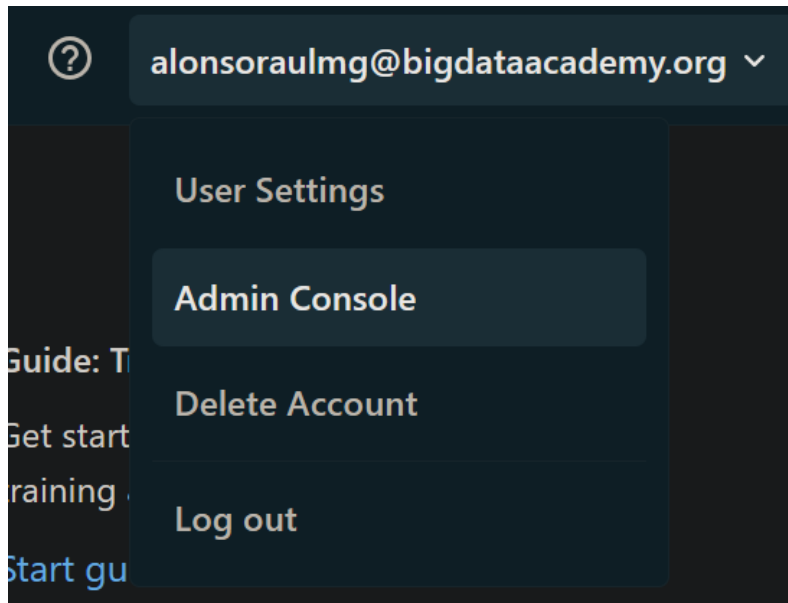
LABORATORIO 03

PREPARACIÓN DE CLÚSTER DE ALMACENAMIENTO

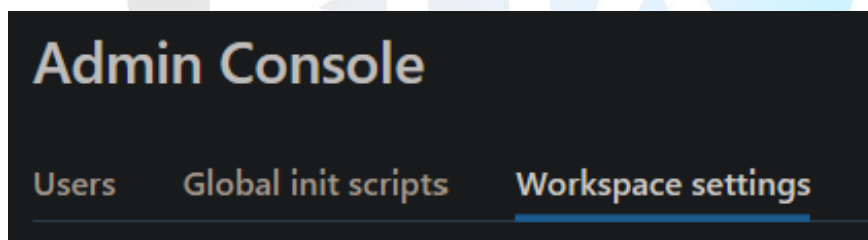
FORMADOR: ALONSO MELGAREJO
alonsoraulmgs@gmail.com

LABORATORIO 03 – PREPARACIÓN DE CLÚSTER DE ALMACENAMIENTO

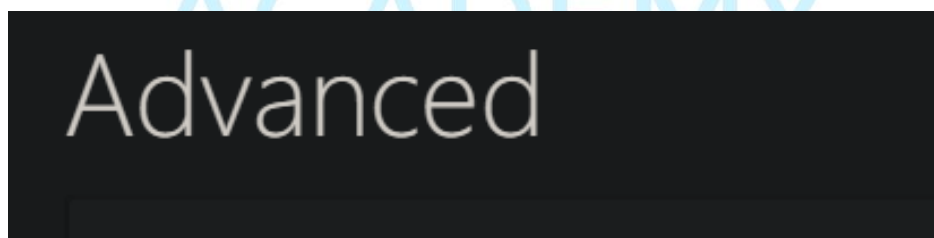
1. En Databricks, en la esquina superior derecha, sobre el nombre de nuestro usuario seleccionamos la opción “Admin Console”



2. Seleccionamos la pestaña “Workspace settings”



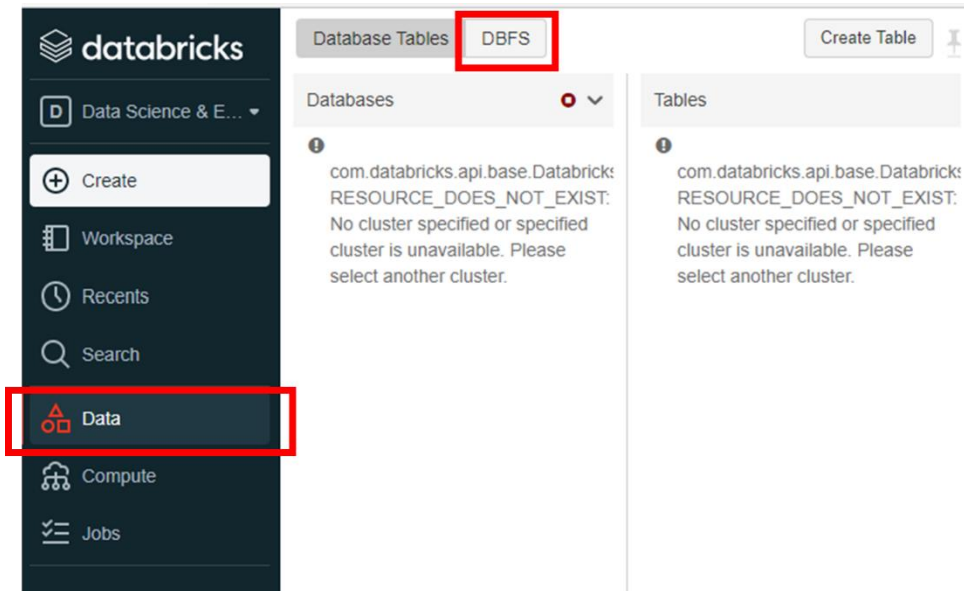
3. Vamos a la sección “Advanced”



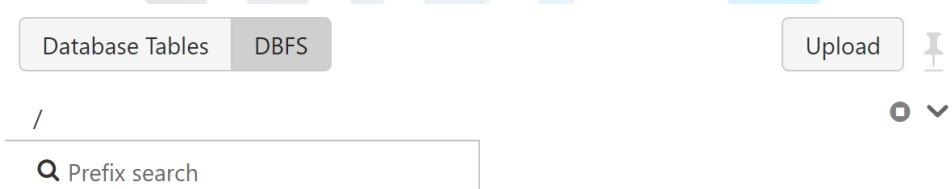
4. Habilitamos la opción “DBFS File Browser”.



5. Para que los cambios tomen efecto, refrescamos la página presionando F5.
6. El clúster de almacenamiento es permanente, nunca se borrarán los archivos que se suban en él. Subiremos los archivos de la carpeta “data” de este laboratorio. Seleccionamos la opción “Data” y hacemos clic sobre “DBFS”



7. Damos clic sobre la opción “Upload”



8. Se nos preguntará en qué ruta queremos subir los archivos:

Upload Data to DBFS

DBFS Target Directory ?

/FileStore

Files uploaded to DBFS are accessible by everyone who has access to this workspace. [Learn more](#)

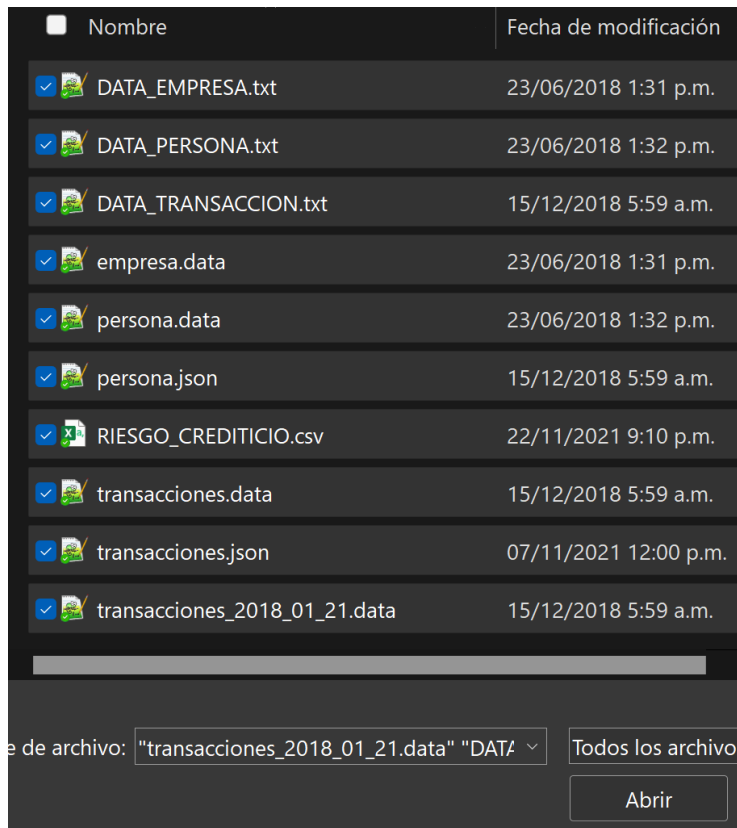
Como ruta (donde dice “XXX”) ingresamos:

IMPORTANTE: LA RUTA DEBE SER ESCRITA EN MINÚSCULA E INCLUIR EL GUIÓN ABAJO (_pyspark) YA QUE LOS SISTEMAS DE ARCHIVOS SON “CASE SENSITIVE”, ES DECIR DIFERENCIAN MINÚSCULAS DE MAYÚSCULAS

9. Para subir los archivos damos clic en “click to browse”

Drop files to upload, or click to browse

10. Seleccionamos todos los archivos del directorio “data” de este laboratorio y damos clic en “Abrir”



11. Los archivos comenzarán a subirse, luego de finalizar la subida damos clic en “Done”

