



BIG DATA
ACADEMY

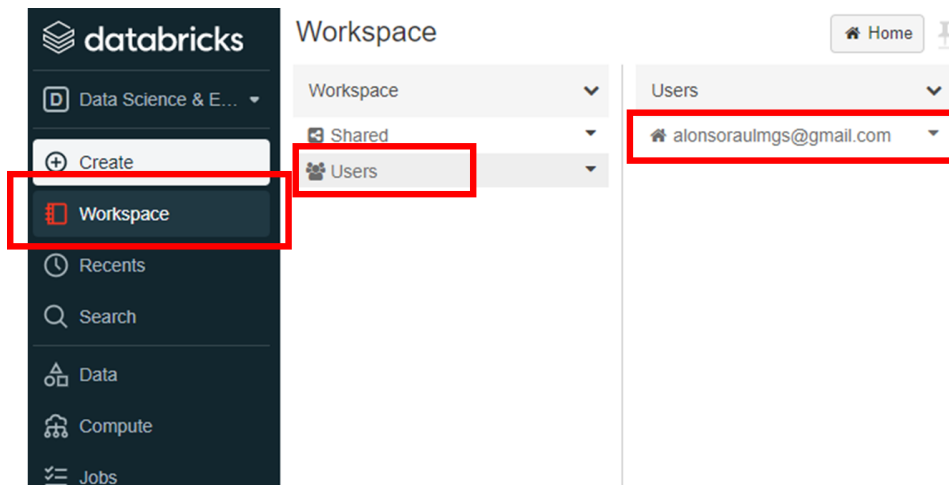
GESTIÓN DE NOTEBOOKS

FORMADOR: ALONSO MELGAREJO
alonsoraulmgs@gmail.com

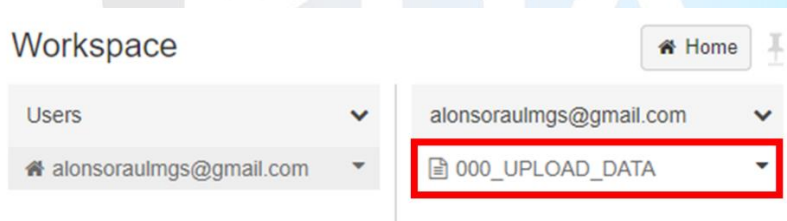
LABORATORIO SUBIDA DE DATOS

1. LISTAR NOTEBOOKS CREADOS

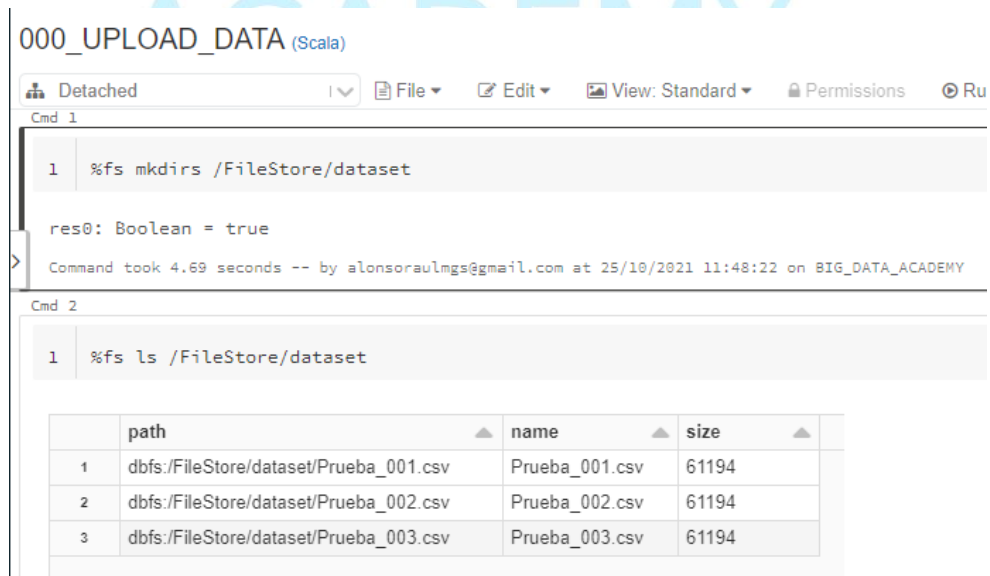
1.1. Seleccionamos la opción “Workspace / Users”, y hacemos clic en el nombre de nuestro usuario.



1.2. Dentro encontraremos los notebooks que creamos, deberemos ver el notebook “000_UPLOAD_DATA” del laboratorio anterior. Para abrirlo, damos clic sobre él.

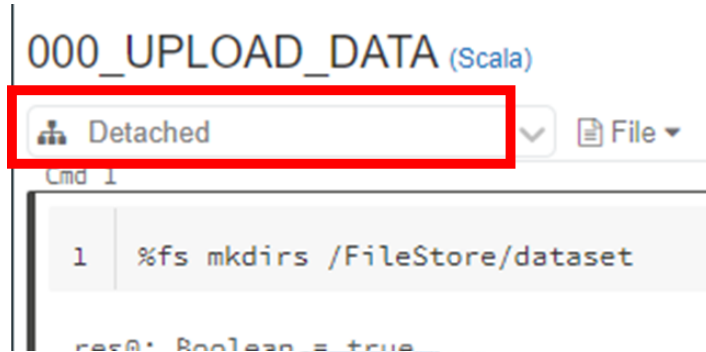


1.3. Veremos el contenido del notebook



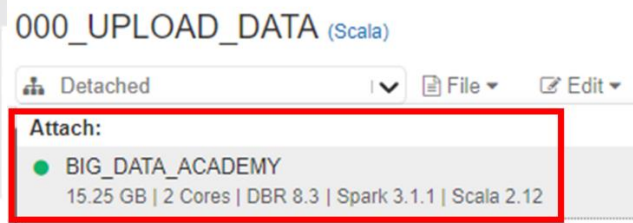
2. ASIGNAR INFRAESTRUCTURA A LOS NOTEBOOKS

- 2.1. En una cuenta gratuita, el cluster se apaga de manera automática luego de dos horas de no haberlo utilizado. Cuando abramos los notebooks, puede ser que el notebook no tenga infraestructura asignada, ya que el cluster está apagado.

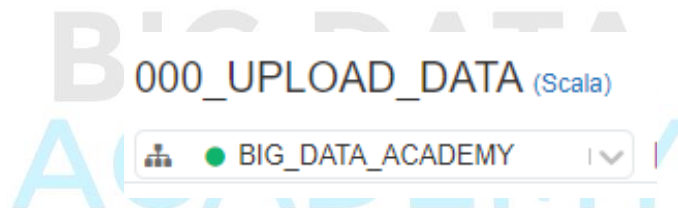


Para asignarle infraestructura, hacemos clic en “Detached”

- 2.2. Seleccionamos el clúster de Big Data.



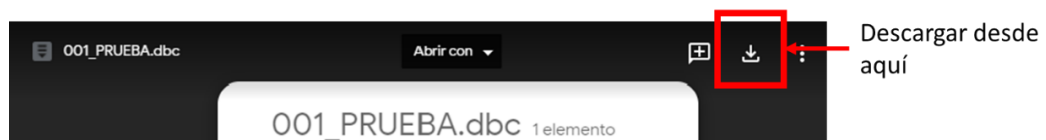
- 2.3. El Notebook tendrá un clúster asignado para ejecutar los códigos



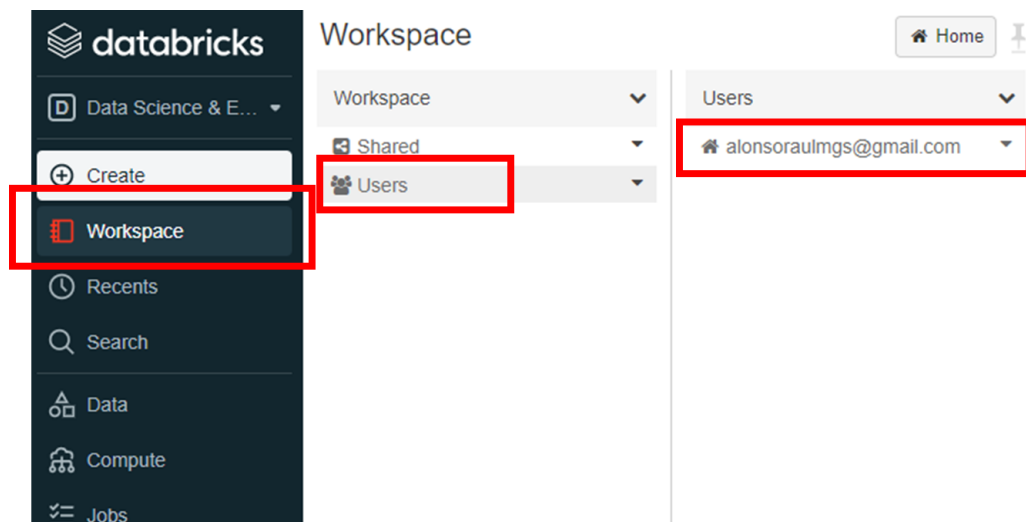
3. SUBIR NOTEBOOKS

- 3.1. Podemos subir notebooks que nos compartan a nuestra cuenta. Los archivos de notebooks tendrán la extensión “.dbc”. Descargar el notebook de ejemplo desde el siguiente enlace

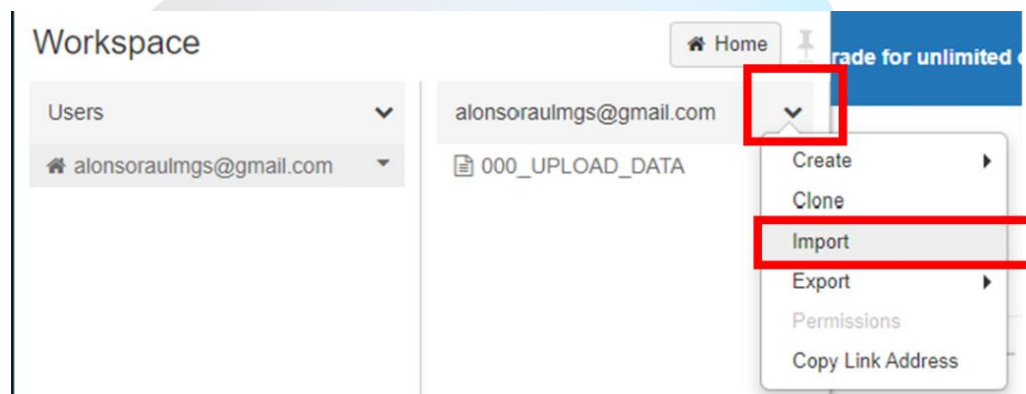
https://drive.google.com/file/d/1t5W_ruiilZf6U4a_HzPSdACT66k5uEju/view?usp=sharing



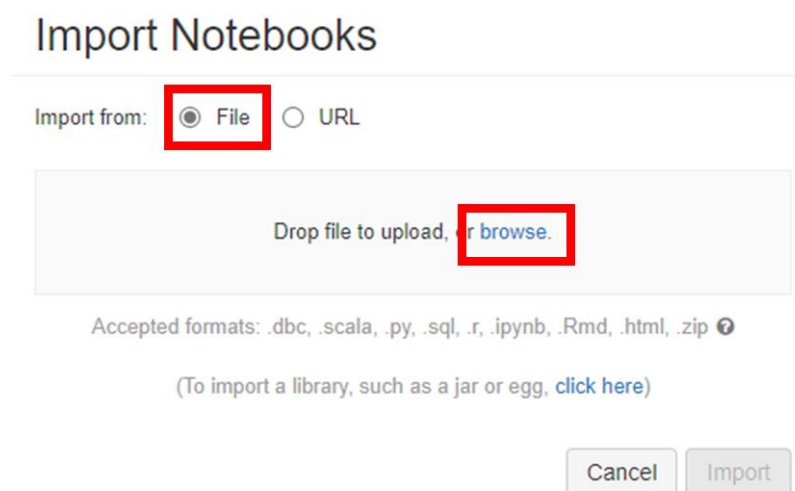
- 3.2. Seleccionamos la opción “Workspace / Users”, y hacemos clic en el nombre de nuestro usuario.



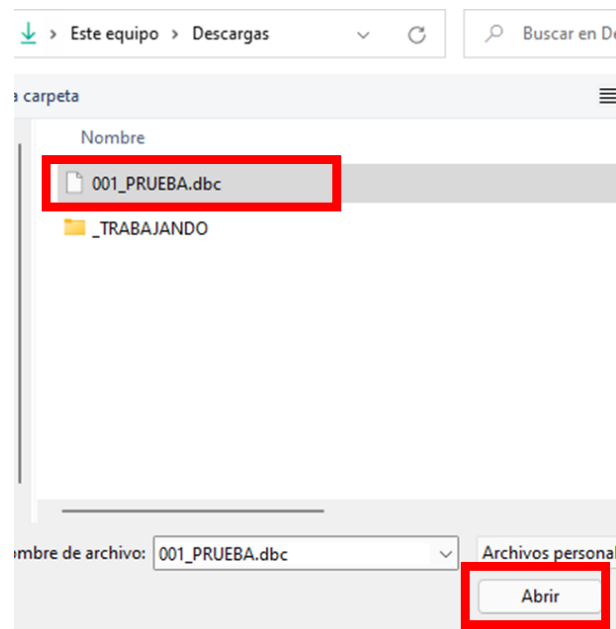
- 3.3. Hacemos clic al lado de la flecha en el nombre de nuestro usuario y seleccionamos la opción “Import”.



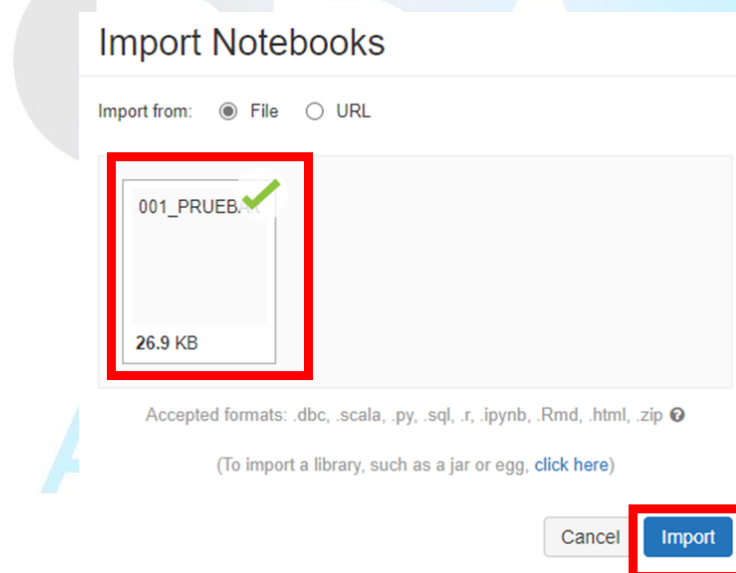
- 3.4. Seleccionamos la opción “File” y damos clic en “browse”.



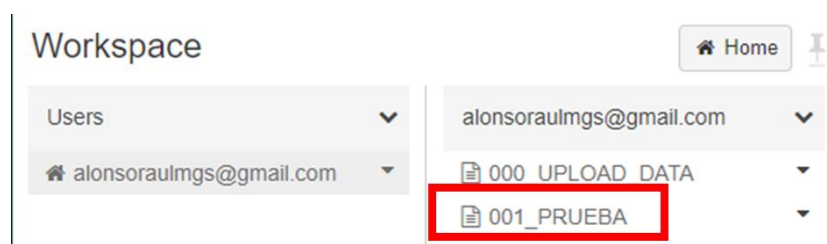
3.5. Seleccionamos el notebook descargado y hacemos clic en “Abrir”.



3.6. Esperamos que el notebook se suba y damos clic en “Import”.

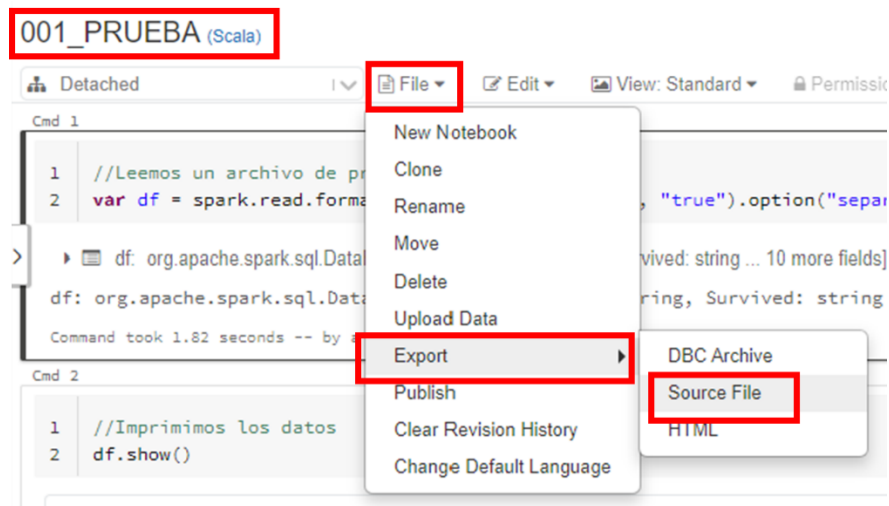


3.7. El notebook se subirá en la carpeta de nuestro usuario. Hacemos clic sobre él para abrirlo.



4. DESCARGAR CÓDIGOS FUENTE DE NOTEBOOKS

4.1. Abrimos el notebook “001_PRUEBA” y damos clic en “File / Export / Source File”



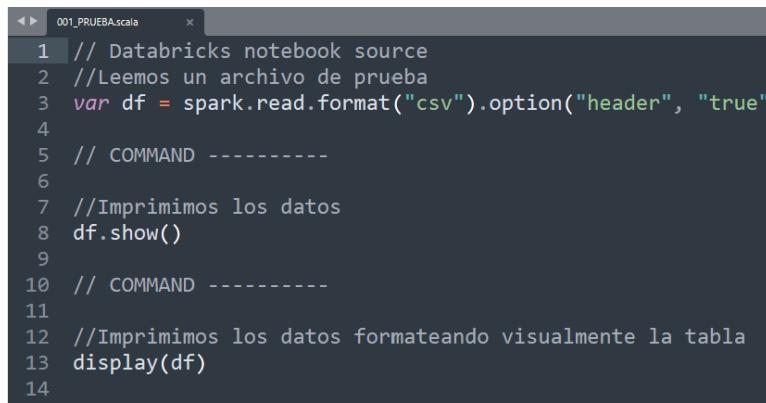
4.2. Se descargará el código fuente del Notebook



4.3. Para visualizar los códigos fuentes, en las clases se usará “Sublime Text”, puedes descargarlo e instalarlo desde:

<https://www.sublimetext.com/download>

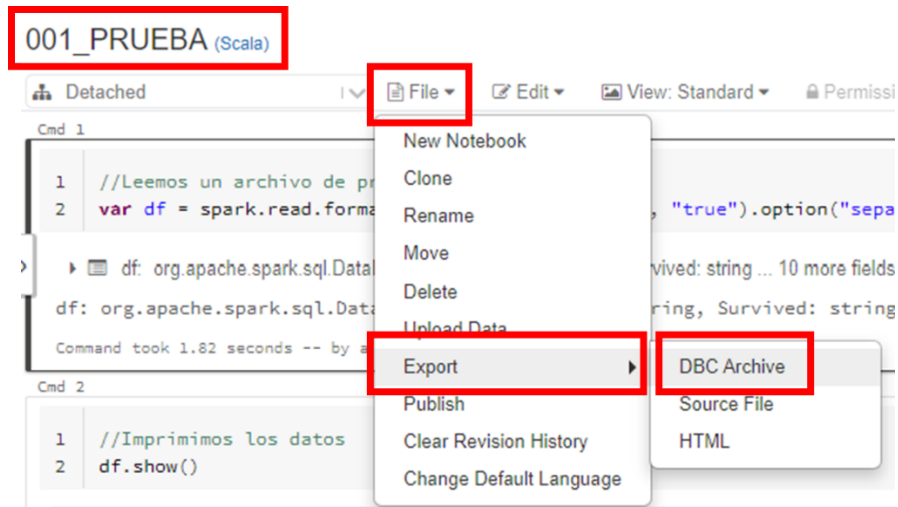
4.4. Para abrir el código fuente abrimos Sublime y arrastramos y soltamos el archivo de código

A screenshot of the Sublime Text editor with a file named '001_PRUEBA.scala' open. The code is as follows:

```
1 // Databricks notebook source
2 //Leemos un archivo de prueba
3 var df = spark.read.format("csv").option("header", "true")
4
5 // COMMAND -----
6
7 //Imprimimos los datos
8 df.show()
9
10 // COMMAND -----
11
12 //Imprimimos los datos formateando visualmente la tabla
13 display(df)
14
```

5. DESCARGAR NOTEBOOKS

5.1. Para descargar los notebooks, damos clic en “File / Export” y seleccionamos la opción “DBC Archive”.



5.2. El notebook se descargará un archivo con extensión “.dbc” que podemos compartir para que pueda ser subido a Databricks.

