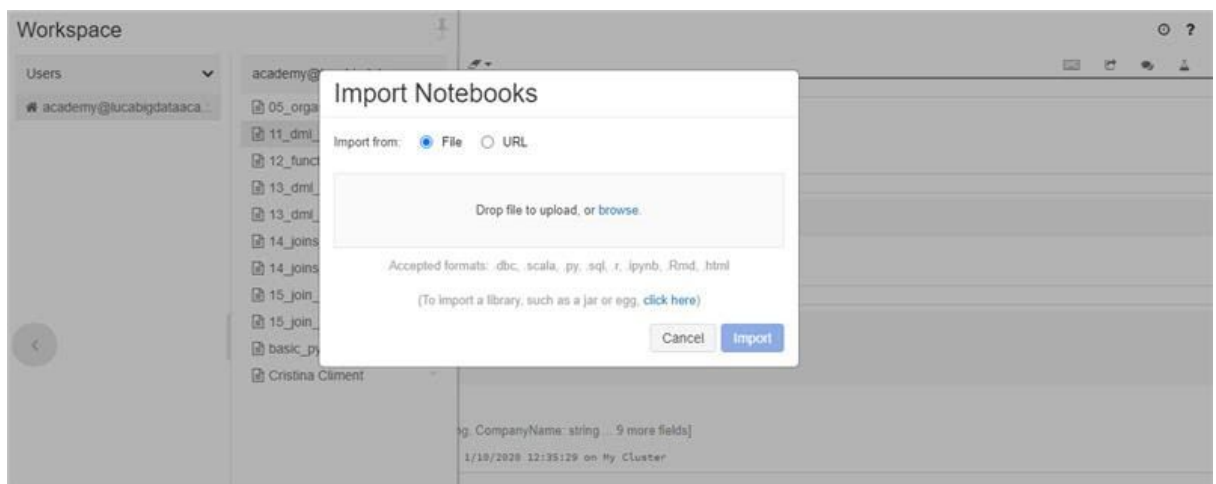
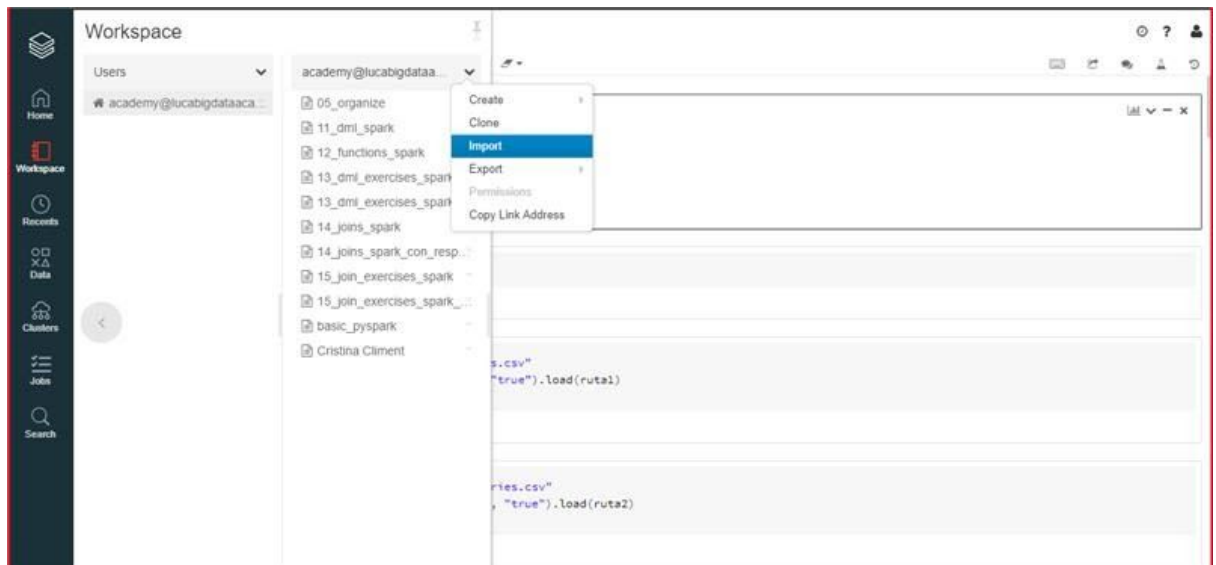


## DATOS SOBRE DATABRIKS:

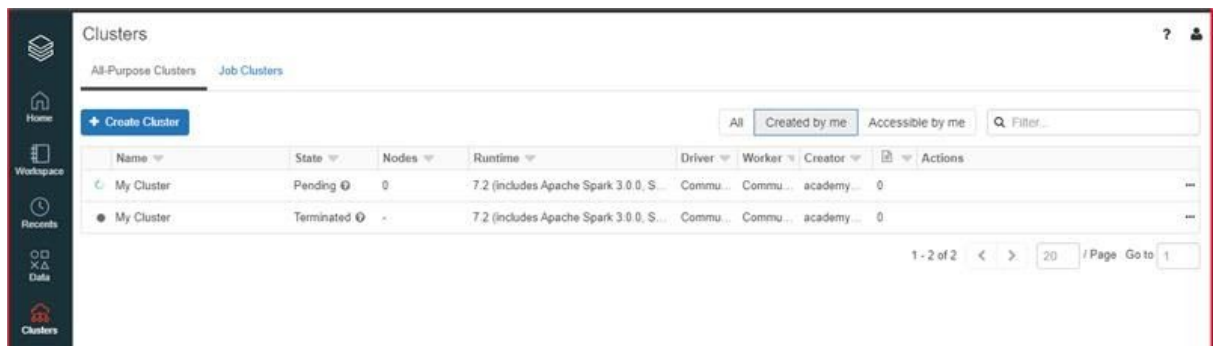
- Está comprobado su buen uso con Spark, pero de momento no con Python
- Hay una opción gratuita: Tiene un límite de recursos de 6GB. Creemos, que como los datos son pequeños, no debería dar problemas

## PRIMEROS PASOS

- Los alumnos tendrían que crearse una cuenta nueva y elegir la opción Community Edition en: <https://databricks.com/try-databricks>
- Los notebooks de spark funcionan bien, y se pueden importar directamente desde el Workspace.

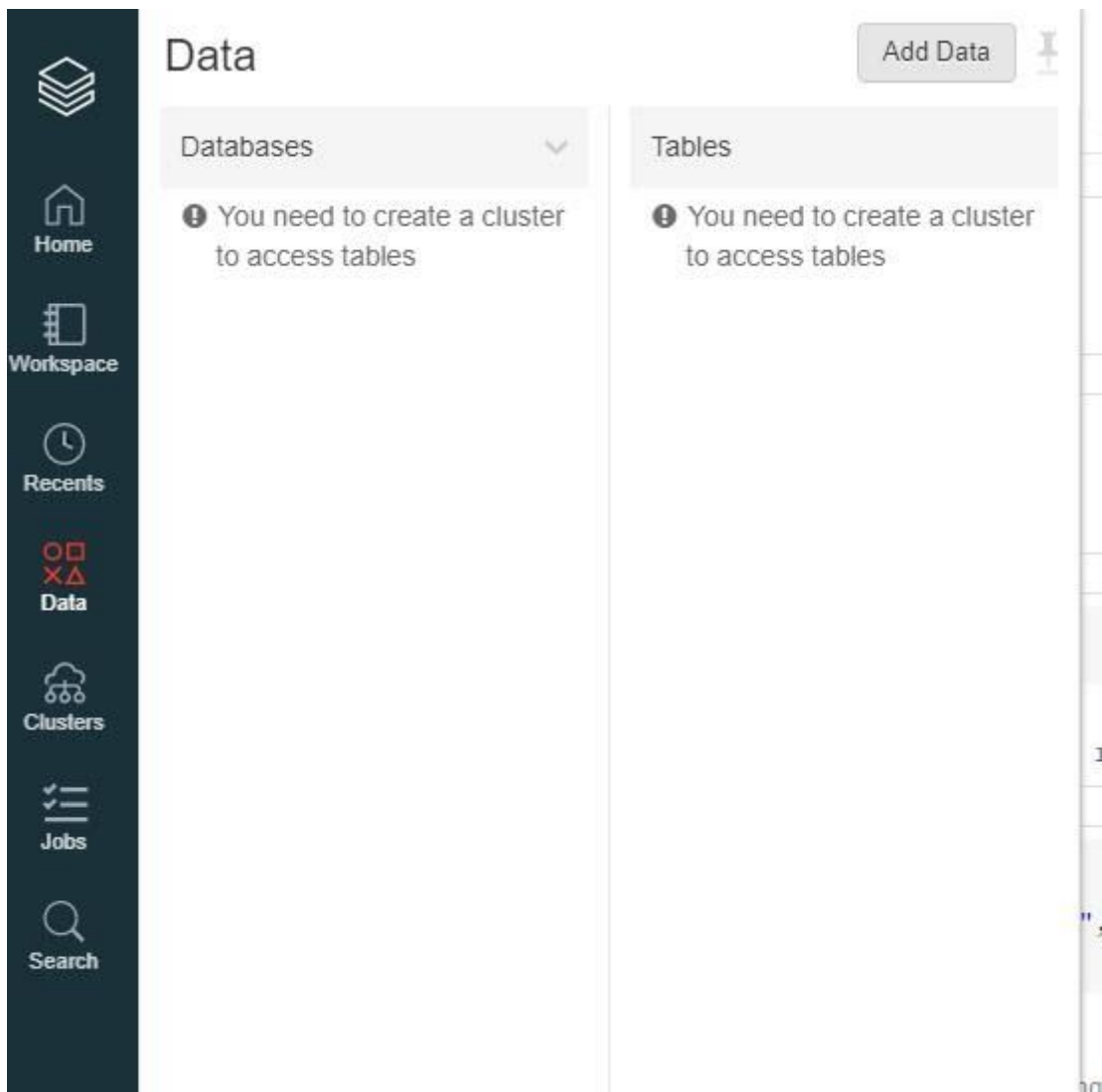


- Antes de importar los datos, debemos encender un cluster o crearlo. Para ello, vamos a la sección “Clusters”



Si ya estaba creado, le podremos dar al play, y si no, podremos crear, recomendación: elegir la última opción que será la más actualizada.

- Importar los datos: En Databricks, tenemos que ir a DATA y “Add data” y subir los datos que necesitemos.



Al importar los datos tenemos dos opciones para crear la tabla:

1. Table with UI: (OPCIÓN RECOMENDADA) Importar los datos desde local y seleccionar el cluster que previamente habíamos encendido.

**Create New Table**

Upload File S3 DBFS Other Data Sources

DBFS Target Directory ?

/FileStore/tables/ (optional) Select

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

Files ?

employees.csv  
0.4 KB  
[Remove file](#)

✓ File uploaded to /FileStore/tables/employees-2.csv

Create Table with UI Create Table in Notebook ?

**Select a Cluster to Preview the Table**

Choose a cluster with which you will read and preview the data.

Cluster ?

My Cluster

Preview Table

Una vez cargada la preview table, podremos crear la tabla.

Si da error de ejecución en algún punto, actualiza la página.

2. Create table in notebook: Te crea un nuevo notebook. Hay que ejecutar las celdas que vienen pre hechas. **IMPORTANTE: NO PONER GUIONES** en el temp\_table\_name. Los guiones bajos no le molestan, los medios producen error. Si queremos guardar la tabla en el cluster para poder utilizarla en otros notebooks, tenemos que habilitar la última sentencia:  
`df.write.format("parquet").saveAsTable(permanent_table_name)`

- Una vez que tenemos ya los datos, podemos ir a nuestro notebook, y hay que cambiar la ruta de la lectura de los datos, tiene que ser esta: `/FileStore/tables/`