

## Pyspark en local

Link de la imagen docker : <https://hub.docker.com/r/jupyter/pyspark-notebook>

Ejecutar comando : `docker pull jupyter/pyspark-notebook:2023-04-03`

C:\Users\Seba>docker images				
REPOSITORY	TAG	IMAGE ID	CREATED	SIZE
jupyter/pyspark-notebook	2023-04-03	dfaf51ad7fd4	9 days ago	4.35GB

<https://github.com/jupyter/docker-stacks> (repo con info para correrlo).

Compartir Vista				
Este equipo > Disco local (C:) > Usuarios > Seba > files				
Nombre	Fecha de modificación	Tipo	Tamaño	
yellow_tripdata_2021-01.parquet	7/4/2023 09:57	Archivo PARQUET	21.178 KB	

En volume pasar ruta donde tengo alojados los archivos.

`docker run -it --rm -p 10000:8888 -v C:\Users\Seba\files:/home/jovyan/work jupyter/pyspark-notebook:2023-04-03`

```

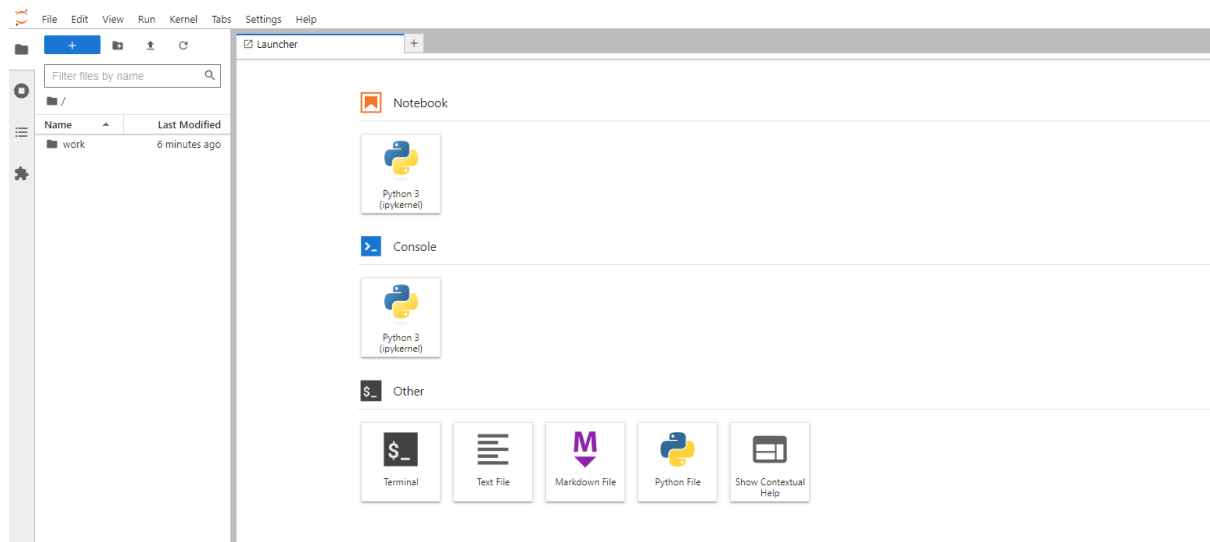
Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions.
https://jupyter-notebook.readthedocs.io/en/latest/migrate_to_notebook7.html
Please note that updating to Notebook 7 might break some of your extensions.

[I 2023-04-12 21:57:18.191 ServerApp] nbclassic | extension was successfully loaded.
[I 2023-04-12 21:57:18.192 ServerApp] Serving notebooks from local directory: /home/jovyan
[I 2023-04-12 21:57:18.192 ServerApp] Jupyter Server 2.5.0 is running at:
[I 2023-04-12 21:57:18.192 ServerApp] http://a9405c0edaee:8888/lab?token=8e55951ad9f0945234d4cf10a40bb5841b0db077fb40033b
[I 2023-04-12 21:57:18.192 ServerApp] http://127.0.0.1:8888/lab?token=8e55951ad9f0945234d4cf10a40bb5841b0db077fb40033b
[I 2023-04-12 21:57:18.192 ServerApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 2023-04-12 21:57:18.203 ServerApp]

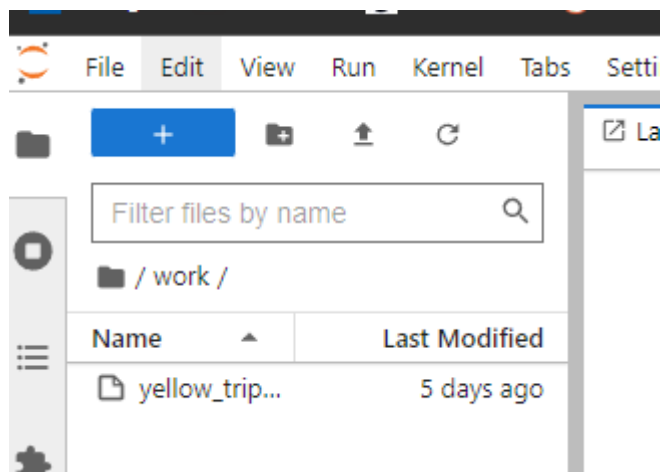
To access the server, open this file in a browser:
file:///home/jovyan/.local/share/jupyter/runtime/jpserver-7-open.html
Or copy and paste one of these URLs:
http://a9405c0edaee:8888/lab?token=8e55951ad9f0945234d4cf10a40bb5841b0db077fb40033b
http://127.0.0.1:8888/lab?token=8e55951ad9f0945234d4cf10a40bb5841b0db077fb40033b
```

Copiar la ruta que está abajo de todo que contiene el token.

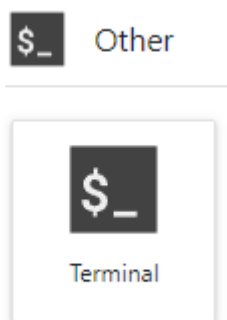
Pegar la ruta en el browser cambiando el port de 8888 a 10000.



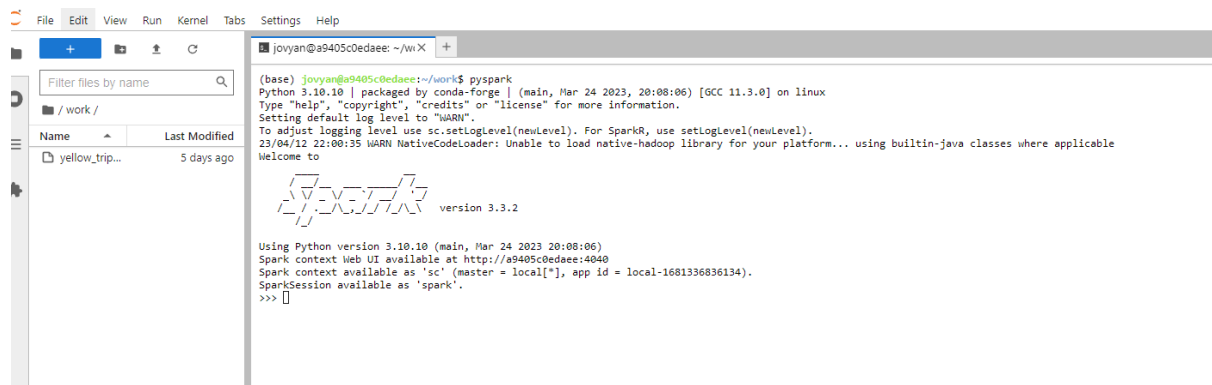
Dentro de la carpeta work vamos a ver los files de nuestra pc.



Ingresar en terminal:



Escribir Pyspark y enter:



Ya podemos usar los comandos de spark para practicar.

[illegible]