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).

En volume pasar ruta donde tengo alojados los archivos.

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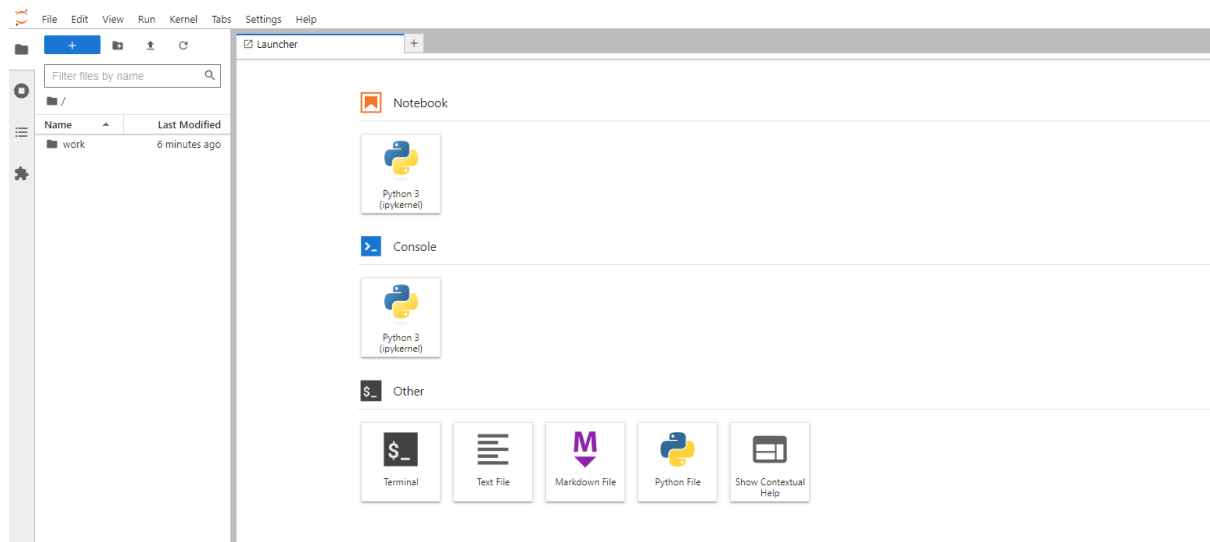

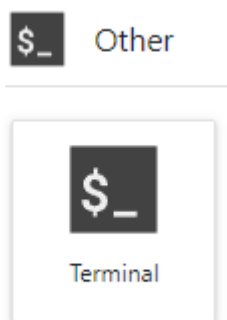Copiar la ruta que está abajo de todo que contiene el token.
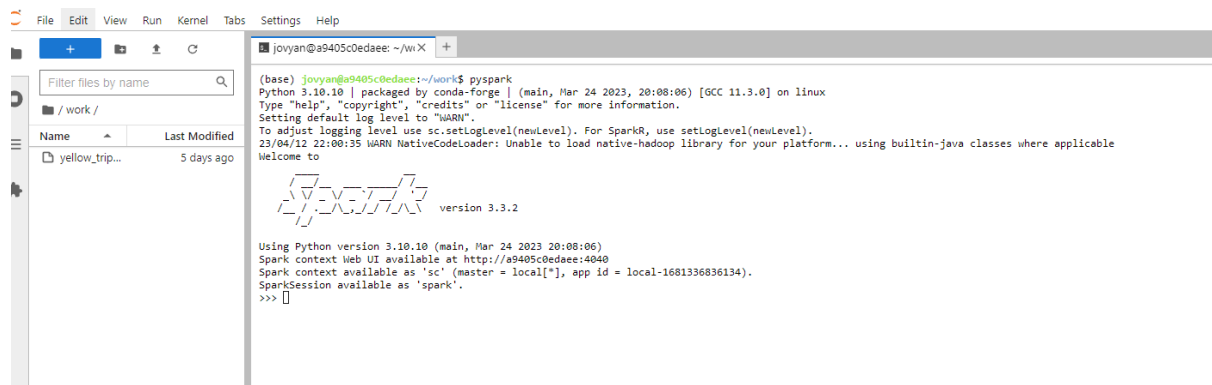Pegar la ruta en el browser cambiando el port de 8888 a 10000.

Dentro de la carpeta work vamos a ver los files de nuestra pc.
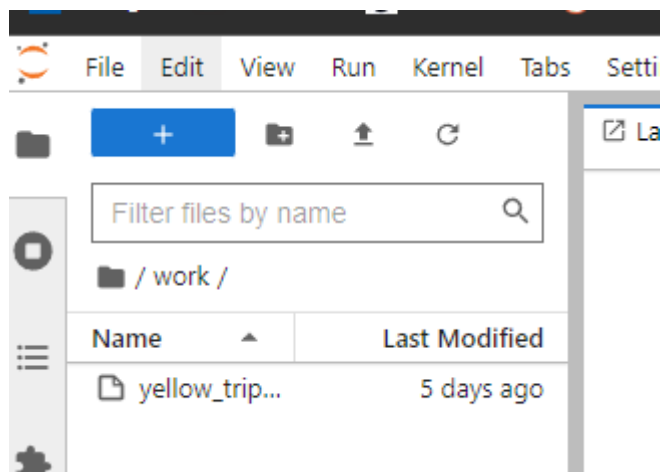
Ingresar en terminal:

Escribir Pyspark y enter:

Ya podemos usar los comandos de spark para practicar.