Como instalar Jupyter Notebook en sus ambientes de Hadoop

Ejecutar el docker run pero abriendo el puerto **8889** si es que aún no lo han ejecutado nunca o si ya está creado el contenedor verifiquen que el puerto 8889 esté abierto (docker inspect edvai_hadoop)

```
docker run --name edvai_hadoop -p 8081:8081 -p 8080:8080 -p 8088:8088 -p
8889:8889 -p 9870:9870 -p 9868:9868 -p 9864:9864 -p 1527:1527 -p
10000:10000 -p 10002:10002 -p 8010:8010 -p 9093:9093 -p 2181:2182 -it
--restart unless-stopped fedepineyro/edvai_ubuntu:v6 /bin/bash -c
"/home/hadoop/scripts/start-services.sh"
```

En edvai_hadoop con el usuario hadoop instalar jupyter:

```
pip install notebook
```

Luego crear en edvai_hadoop el siguiente script en /home/hadoop/scripts:

```
cat > pyspark_jupyter.sh
```

```
#! /bin/bash

# Define env Python to use
export PYSPARK_PYTHON=/usr/bin/python3

# Define IPython driver
export PYSPARK_DRIVER_PYTHON='jupyter'

# Define Spark conf to use
export PYSPARK_DRIVER_PYTHON_OPTS='notebook --ip=172.17.0.2 --port=8889'

/home/hadoop/spark/bin/pyspark
```

Cambiar los permisos del script:

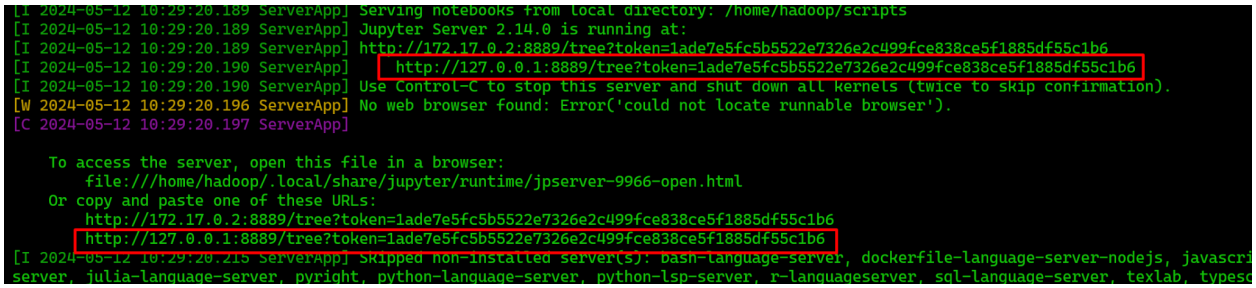
```
chmod 777 /home/hadoop/scripts/pyspark_jupyter.sh
```

Ingresar al directorio /home/hadoop y ejecutar el script:

```
/home/hadoop/scripts/pyspark_jupyter.sh
```

NOTA: Es importante que antes de ejecutar el script entren al directorio /home/hadoop

Al ejecutar el script nos dara la ruta al cual debemos acceder:

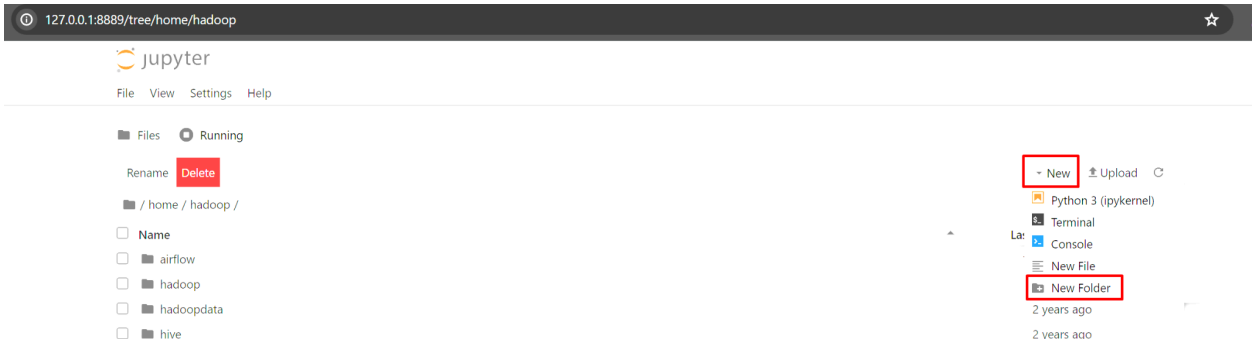


Ejemplo:
http://127.0.0.1:8889/tree?token=1ade7e5fc5b5522e7326e2c499fce838ce5f1885df55c1b6

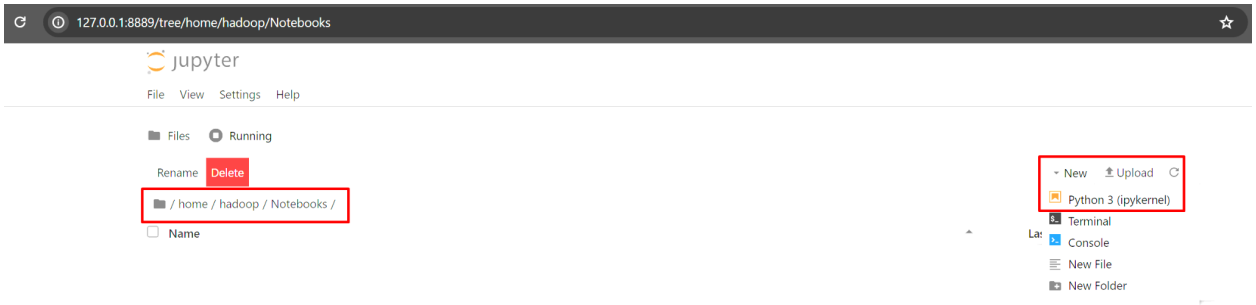
Debemos ingresar siempre a la ruta que comienza con http://127.0.0.1:8889/tree.....

Podemos acceder con Ctrl y seleccionando esa ruta o la podemos copiar y pegar en el navegador.

Allí crearemos una carpeta llamada Notebooks, haciendo clic en nuevo y luego -> Nueva carpeta:



Una vez dentro de esa carpeta que hemos creado (/home/hadoop/notebooks)Debemos seleccionar New y luego Python 3



Al iniciar una nueva notebook debemos crear una SparkSession de la siguiente manera:

```
from pyspark.sql import SparkSession
spark = SparkSession.builder \
.master("spark://localhost:7077") \
.getOrCreate()
```

Y luego continuamos trabajando en la notebook con Pyspark: