

TP : Hadoop

Iyad LAHSEN-CHERIF
lahsencherif@inpt.ac.ma

INE 2

1 Installation de PySpark

Depuis votre terminal (CMD), installer pyspark en utilisant la commande `docker run -it -p 8888:8888 jupyter/pyspark-notebook`. *docker run* est une commande docker qui permet de créer un conteneur à partir d'une image et de le lancer par la suite.

Une fois le téléchargement terminer,

- Vérifiez sur Docker Desktop que :
 - Une image de pyspark est créée avec le nom `jupyter/pyspark-notebook`
 - Un conteneur de pyspark est lancé.
- Copiez le dernier lien (`127.0.0.1:8888/lab?token xxx`) sur votre navigateur. Ceci va vous permettre de lancer le Jupyter-notebook et pouvoir expérimenter pyspark. Note : vous devez garder votre terminal ouvert avant de quitter votre environnement Jupyter.

Vous avez maintenant un environnement Jupyter Lab avec PySpark pré-configuré, fonctionnant dans un conteneur docker.

Copiez le Jupyter Notebook (ipynb) et Dataset (csv) depuis votre PC local vers le container. Depuis un autre terminal (CMD), utilisez la commande : `docker cp dossier_local_pc nom_container:chemin/dossier/dans/le/container`

Exemple :

```
docker cp Spark-SQL-Questions.ipynb  hardcore_elbakyan:/home/jovyan/work
docker cp drinkquality.csv  hardcore_elbakyan:/home/jovyan/work
```

- Pour trouver le nom du container, ouvrez Docker Desktop, et checkez le nom affecté à votre container pyspark.
- Pour trouver le chemin où mettre les fichiers sur le container, connectez-vous au niveau du container depuis terminal avec la commande `docker exec -it nom_container bash`. Puis, à l'aide de la commande `pwd` depuis le bash du container. Ceci vous donnera le chemin (le nom affecté à votre home). Exemple :

```
docker cp Spark-SQL.ipynb  hardcore_elbakyan:/home/jovyan/work
```

2 A la fin du TP

A la fin du TP, enregistrez le travail effectué dans le container sur votre PC (en local). Pour récupérer, en local sur votre PC, les fichiers que vous avez créés sur votre container, utilisez la commande (depuis CMD).

```
docker cp nom_container:chemin/dossier/dans/le/container  dossier_local_pc
```

Exemple : `docker cp hardcore_elbakyan:/home/jovyan/ TP_spark`