



MapReduce

1 INSTALLATION DE JUPYTER ET SPARK AVEC DOCKER

Ces instructions vous permettent d'installer une image docker contenant jupyter et Hadoop-Spark utiles pour les TD de MapReduce et le TP de Spark.

2 PRÉPARATION DE L'ENVIRONNEMENT DE TRAVAIL

Afin de préparer environnement de travail, veuillez suivre les instructions suivantes :

2.1 DOCKER

1. Télécharger l'image docker de Jupyter-PySpark avec la commande suivante:

```
docker pull jupyter/pyspark-notebook
```

Si vous avez un problème de permission il faut vous ajouter au groupe docker. Le téléchargement de l'image peut prendre plusieurs minutes en fonction de votre réseau. Cette image sera utilisée pour le TD de mapReduce ainsi que le TP de Spark.

2.2 JUPYTER

Instructions d'installation de l'environnement.

- Dans un terminal lancez la commande suivante :

```
docker system prune
```

et sélectionner l'option y

- Lancer la commande :

```
docker run -p 8888:8888 jupyter/pyspark-notebook
```

et récupérez l'adresse affichée dans le terminal.

- Example :

Or copy and paste one of these URLs:

```
http://81b5cda5e5df:8888/?token=7e126e789bb1b3f03533d6fe6a1c290c08951a0c1727e83c  
or
```

```
http://127.0.0.1:8888/?token=7e126e789bb1b3f03533d6fe6a1c290c08951a0c1727e83c
```

- En alternative ouvrez votre navigateur à l'adresse :

```
http://localhost:8888
```

et copier-collez le token affiché dans le terminal à la demande de Jupyter.

2.3 IMPORTATION DES FICHIERS

Importer les fichiers:

- mapReduce.ipynb
- books.json
- numbers.txt

dans Jupyter à l'aide du bouton de téléchargement de l'interface de Jupyter.