

Résumé des Commandes - TP2 : Programmation avec l'API HDFS

Démarrage du cluster

- `docker start hadoop-master hadoop-slave1 hadoop-slave2`
- `docker exec -it hadoop-master bash`
- `./start-hadoop.sh`
- `jps`

Installation de l'environnement

- `java -version`
- `mvn clean package`
- Ajouter les JAR : `hadoop-hdfs-3.2.0.jar`, `hadoop-common-3.2.0.jar`, `hadoop-mapreduce-client-core-3.2.0.jar`

HadoopFileStatus (sans paramètres)

- `mvn clean package`
- `cp target/hadoop-hdfs-HadoopFileStatus.jar ~/documents/BigData/hadoop_project/`
- `hadoop jar /shared_volume/hadoop-hdfs-HadoopFileStatus.jar`

HadoopFileStatus (avec paramètres)

- `hadoop jar /shared_volume/hadoop-hdfs-HadoopFileStatusPar.jar /user/root/input/purchases.txt achats.txt`
- `hdfs dfs -ls /user/root/input/`

HDFSInfo (infos fichier)

- `hadoop jar /shared_volume/hadoop-hdfs-HDFSInfo.jar /user/root/input/purchases.txt`

ReadHDFS (lecture)

- `hadoop jar /shared_volume/hadoop-hdfs-ReadHDFS.jar /user/root/input/purchases.txt`

WriteHDFS (écriture)

- `hadoop jar /shared_volume/hadoop-hdfs-WriteHDFS.jar /user/root/input/bonjour.txt 'Mon message'`
- `hdfs dfs -ls /user/root/input/`
- `hadoop jar /shared_volume/hadoop-hdfs-ReadHDFS.jar /user/root/input/bonjour.txt`