Résumé des Commandes - Compte Rendu TP1 Hadoop avec Docker

1. Installation de Docker

- sudo apt-get install docker.io
- docker version

2. Télécharger l'image Hadoop-Spark-Cluster

- docker pull yassern1/hadoop-spark-jupyter:1.0.3
- docker images

3. Création des trois conteneurs

- docker compose up -d

4. Démarrage du cluster Hadoop

- docker exec -it hadoop-master bash
- ./start-hadoop.sh
- jps
- mr-jobhistory-daemon.sh start historyserver
- Interfaces Web :
- - NameNode UI : localhost:9870
- - Resource Manager UI : localhost:8088
- - JobHistory Server : localhost:19888

5. Manipulations sur HDFS

- hadoop fs -mkdir -p /user/root/input
- hdfs dfs -put /shared_volume/purchases.txt /user/root/input
- hdfs dfs -ls -R
- hdfs dfs -tail input/purchases.txt
- hdfs dfs -rm input/purchases.txt
- hdfs dfs -copyFromLocal /shared_volume/purchases.txt ./input
- hdfs dfs -chmod go+w ./input/purchases.txt
- hdfs dfs -chmod go-r ./input/purchases.txt
- hdfs dfs -mv ./input/purchases.txt /user/root/purchases.txt
- hdfs dfs -get /user/root/purchases.txt /shared_volume/achat.txt
- hdfs dfs -cp /user/root/purchases.txt ./purchases_copy.txt

6. Télécharger un fichier sur HDFS

- hdfs dfs -mkdir web_input
- wget http://www.textfiles.com/etext/FICTION/alice.txt
- cp alice.txt /shared_volume/
- hdfs dfs -put /shared_volume/alice.txt web_input

- hdfs dfs -ls web_input
- exit
- docker stop hadoop-master hadoop-slave1 hadoop-slave2