

Compte Rendu TP1

Installation d'un Cluster Hadoop avec Docker

Nom et Prénom : AMRANI Hamza

Filière : IDF

I. Installation Cluster Hadoop

1. Installation docker

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

```
docker version
```

```
hamza@Hamza-amr MINGW64 ~/documents/Big Data/hadoop_project
$ docker -v
Docker version 28.0.4, build b8034c0
```

2. Télécharger l'image hadoop-spark-cluster

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

```
docker images
```

```
hamza@Hamza-amr MINGW64 ~/documents/Big Data/hadoop_project
$ docker images
REPOSITORY          TAG         IMAGE ID      CREATED        SIZE
yassern1/hadoop-spark-jupyter  1.0.3      511eec146b04  4 weeks ago   3.67GB
```

3. Création des trois conteneurs

Pour créer les trois conteneurs, nous avons utilisé Docker Compose. (Voir fichier `docker-compose.yml`)

Lancer les conteneurs avec Docker Compose :

```
docker compose up -d
```

```
hamza@Hamza-amr MINGW64 ~/documents/Big Data/hadoop_project
$ docker compose up -d
[+] Running 3/3
  ✓ Container hadoop-slave2  started
  ✓ Container hadoop-master  started
  ✓ Container hadoop-slave1  started
```

hadoop_project	-	-	0%	40 seconds ago	
hadoop-master	25a70a0f1289	yassern1/hadoop-spark:19888:19888	0%	40 seconds ago	
hadoop-slave1	35a0b09fab01	yassern1/hadoop-spark:8040:8042	0%	40 seconds ago	
hadoop-slave2	dd4572b08633	yassern1/hadoop-spark:8041:8042	0%	40 seconds ago	

4. Accéder au master et Démarrer hadoop et yarn

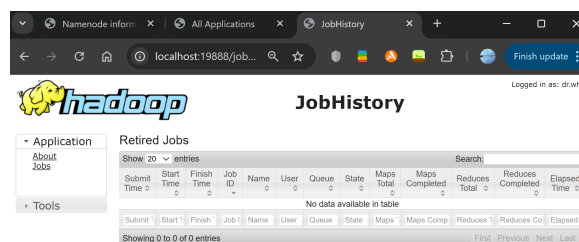
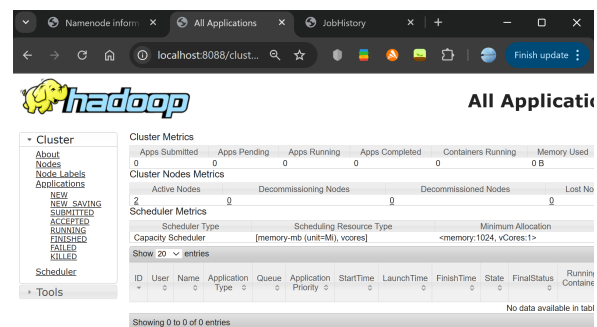
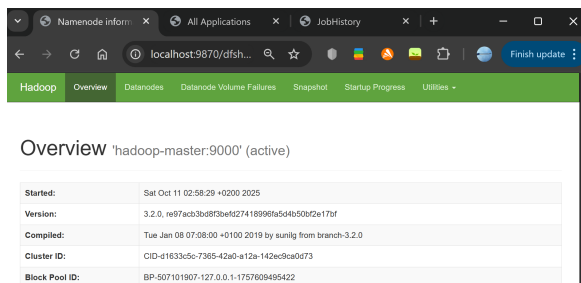
```
docker exec -it hadoop-master bash
```

```
./start-hadoop.sh
```

```
hamza@Hamza-amr MINGW64 ~/documents/Big Data/hadoop_project
$ docker exec -it hadoop-master bash
root@hadoop-master:~# ./start-hadoop.sh
```

Vérifier les interfaces web :

- NameNode web UI : localhost:9870
- Resource Manager UI : localhost:8088
- MapReduce JobHistory Server : localhost:19888



N.B :

Si le JobHistory Server ne démarre pas correctement, vérifier avec la commande `jps` et le démarrer manuellement :

```
jps
mr-jobhistory-daemon.sh start historyserver
```

5. Manipulations sur HDFS

Créer le dossier input et uploader le fichier purchases.txt :

```
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
```

```
root@hadoop-master: ~
root@hadoop-master:~# hadoop fs -mkdir -p /user/root/input
root@hadoop-master:~# hdfs dfs -ls
Found 1 items
drwxr-xr-x  - root supergroup          0 2025-10-11 01:20 input
root@hadoop-master:~# hdfs dfs -put /shared_volume/purchases.txt /user/root/input
root@hadoop-master:~# hdfs dfs -ls -R
drwxr-xr-x  - root supergroup          0 2025-10-11 01:21 input
-rw-r--r--  2 root supergroup        2549 2025-10-11 01:21 input/purchases.txt
root@hadoop-master:~# hdfs dfs -tail input/purchases.txt
ustin Sporting Goods 327.75 Discover
2012-01-01 09:01 Portland CDs 108.69 Amex
2012-01-01 09:01 Riverside Sporting Goods 15.41 Discover
2012-01-01 09:01 Reno Toys 80.46 Visa
2012-01-01 09:01 Anchorage Music 298.86 MasterCard
2012-01-01 09:01 Pittsburgh Sporting Goods 475.26 Amex
2012-01-01 09:01 Spokane Garden 3.85 Amex
2012-01-01 09:01 Spokane Computers 287.65 MasterCard
2012-01-01 09:01 Fresno CDs 466.64 MasterCard
2012-01-01 09:01 Omaha Baby 255.68 MasterCard
2012-01-01 09:01 Anchorage DVDs 6.38 Amex
2012-01-01 09:01 Aurora Consumer Electronics 117.81 MasterCard
2012-01-01 09:01 Philadelphia DVDs 351.31 Cash
2012-01-01 09:01 Fremont Baby 222.61 Cash
2012-01-01 09:01 Anchorage Crafts 22.36 Amex
2012-01-01 09:02 Norfolk Women's Clothing 189.01 Amex
2012-01-01 09:02 Chandler Books 414.08 Cash
2012-01-01 09:02 Minneapolis Computers 182.05 Visa
2012-01-01 09:02 Honolulu Cameras 345.18 Discover
2012-01-01 09:02 Indianapolis Books 135.96 Discover
2012-01-01 09:02 Chandler Books 344.09 Discover
root@hadoop-master:~# hdfs dfs -rm input/purchases.txt
Deleted input/purchases.txt
root@hadoop-master:~# hdfs dfs -ls input/
```

Supprimer le fichier, le remettre avec copyFromLocal, et modifier les permissions :

```
hdfs dfs -rm input/purchases.txt
hdfs dfs -ls input/
hdfs dfs -copyFromLocal /shared_volume/purchases.txt ./input
hdfs dfs -ls input/
hdfs dfs -chmod go+w ./input/purchases.txt
hdfs dfs -ls input/
hdfs dfs -chmod go-r ./input/purchases.txt
hdfs dfs -ls input/
```

Déplacer, télécharger et copier le fichier :

```
hdfs dfs -mv ./input/purchases.txt /user/root/purchases.txt
hdfs dfs -ls -R
hdfs dfs -get /user/root/purchases.txt /shared_volume/achat.txt
hdfs dfs -cp /user/root/purchases.txt ./purchases_copy.txt
hdfs dfs -ls
```

```

root@hadoop-master: ~
root@hadoop-master:~# hdfs dfs -rm input/purchases.txt
Deleted input/purchases.txt
root@hadoop-master:~# hdfs dfs -ls input/
root@hadoop-master:~# hdfs dfs -copyFromLocal /shared_volume/purchases.txt /user/root/input
root@hadoop-master:~# hdfs dfs -ls input/
Found 1 items
-rw-r--r--  2 root supergroup      2549 2025-10-11 01:24 input/purchases.txt
root@hadoop-master:~# hdfs dfs -chmod go+w ./input/purchases.txt
root@hadoop-master:~# hdfs dfs -ls input/
Found 1 items
-rw-rw-rw-  2 root supergroup      2549 2025-10-11 01:24 input/purchases.txt
root@hadoop-master:~# hdfs dfs -chmod go-r ./input/purchases.txt
root@hadoop-master:~# hdfs dfs -chmod go-r ./input/purchases.txt
root@hadoop-master:~# hdfs dfs -ls input/
Found 1 items
-rw--w--w-  2 root supergroup      2549 2025-10-11 01:24 input/purchases.txt
root@hadoop-master:~# hdfs dfs -mv ./input/purchases.txt /user/root/purchases.txt
root@hadoop-master:~# hdfs dfs -ls -R
drwxr-xr-x  - root supergroup          0 2025-10-11 01:27 input
-rw--w--w-  2 root supergroup      2549 2025-10-11 01:24 purchases.txt
root@hadoop-master:~# hdfs dfs -get /user/root/purchases.txt /shared_volume/achat.txt
get: '/shared_volume/achat.txt': File exists
root@hadoop-master:~# hdfs dfs -cp /user/root/purchases.txt ./purchases_copy.txt
root@hadoop-master:~# hdfs dfs -ls
Found 3 items
drwxr-xr-x  - root supergroup          0 2025-10-11 01:27 input
-rw--w--w-  2 root supergroup      2549 2025-10-11 01:24 purchases.txt
-rw-r--r--  2 root supergroup      2549 2025-10-11 01:28 purchases_copy.txt

```

6. Télécharger un fichier sur HDFS

Créer le dossier web_input, télécharger alice.txt et l'uploader vers 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

```

Arrêter les conteneurs :

```

docker stop hadoop-master hadoop-slave1 hadoop-slave2

```

```

MINGW64/c/Users/hamza/documents/Big Data/hadoop_project
root@hadoop-master:~# hdfs dfs -mkdir web_input
root@hadoop-master:~# wget http://www.textfiles.com/etext/FICTION/alice.txt
--2025-10-11 01:30:00-- http://www.textfiles.com/etext/FICTION/alice.txt
Resolving www.textfiles.com (www.textfiles.com)... 208.86.224.90
Connecting to www.textfiles.com (www.textfiles.com)|208.86.224.90|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 150886 (147K) [text/plain]
Saving to: 'alice.txt'

alice.txt                                100%[=====]
2025-10-11 01:30:03 (32.4 MB/s) - 'alice.txt' saved [150886/150886]

root@hadoop-master:~# cp alice.txt /shared_volume/
root@hadoop-master:~# hdfs dfs -put /shared_volume/alice.txt web_input
root@hadoop-master:~# hdfs dfs -ls web_input
Found 1 items
-rw-r--r--  2 root supergroup    150886 2025-10-11 01:30 web_input/alice.txt
root@hadoop-master:~# exit
exit

hamza@Hamza-amr MINGW64 ~/documents/Big Data/hadoop_project
$ docker stop hadoop-master hadoop-slave1 hadoop-slave2
hadoop-master
hadoop-slave1
hadoop-slave2

```

II. Conclusion

Dans ce TP, nous avons réussi à installer Docker, créer un cluster Hadoop avec 3 conteneurs, et manipuler le système de fichiers HDFS avec différentes commandes.