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
```

4.   Accéder au master et Démarrer hadoop et yarn

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

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


Vérifier les interfaces web :

— NameNode web UI : `localhost:9870`

— Resource Manager UI : `localhost:8088`

— MapReduce JobHistory Server : `localhost:19888`
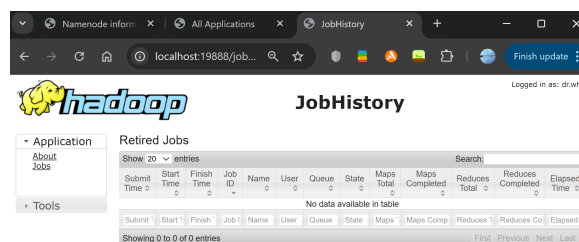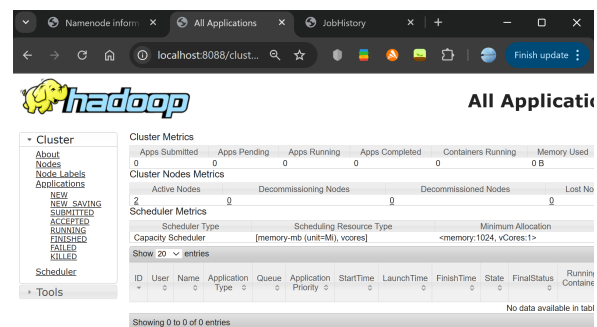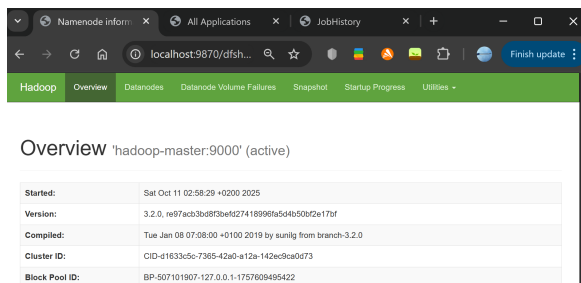
> **N.B :**
>
> Si le JobHistory Server ne démarre pas correctement, vérifier avec la commande `jps` et le démarrer manuellement :
>
> ```
> jps
> mr-jobhistory-daemon.sh start historyserver
> ```

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
```


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.