

Compte Rendu
Bases de données NoSQL et Big Data

TP1
Initiation à Hadoop

Réalisé par
SAIDANE Yassine

Partie 2 : Commande de HDFS

Connexion a la centenaire **NameNode**

```
PS C:\Users\ASUS> docker exec -it namenode bash
root@d307457e10fd:/# ls
KEYS      bin      boot    dev      etc      hadoop-data  lib
```

Création du fichier **bonjour.txt**

```
root@d307457e10fd:/# echo "Bonjour Hadoop et HDFS" > bonjour.txt
root@d307457e10fd:/# ls
KEYS      bin      boot    dev      etc      hadoop-data  lib  mapperWC.py  mnt  proc      root  run.sh  srv  tmp  var
bien.txt  bonjour.txt  data  entrypoint.sh  hadoop  home     lib64  media      opt  reducerWC.py  run  sbin    sys  usr
```

Voir le contenu des répertoires **/**, **/user**, **/var**

```
root@d307457e10fd:/# hdfs dfs -ls /
Found 10 items
drwxr-xr-x - root supergroup 0 2023-10-17 11:58 /TPs
drwxrwxrwt - root root 0 2023-10-17 16:34 /app-logs
drwxr-xr-x - root supergroup 0 2023-10-17 14:48 /data
drwxr-xr-x - root supergroup 0 2023-10-17 16:16 /input
drwxr-xr-x - root supergroup 0 2023-10-24 14:48 /output
drwxr-xr-x - root supergroup 0 2023-10-17 11:09 /rmstate
drwxr-xr-x - root supergroup 0 2023-10-17 11:56 /root
drwx----- - root supergroup 0 2023-10-17 16:33 /tmp
drwxr-xr-x - root supergroup 0 2023-10-17 11:40 /user
drwxr-xr-x - root supergroup 0 2023-10-17 11:40 /var
root@d307457e10fd:/# hdfs dfs -ls /user
root@d307457e10fd:/# hdfs dfs -ls -R -h /var
```

Créer le répertoire **/dossier**

```
root@d307457e10fd:/# hdfs dfs -mkdir /dossier
root@d307457e10fd:/# cat bonjour.txt
Bonjour Hadoop et HDFS
```

Copier le fichier **bonjour.txt** vers HDFS

```

root@d307457e10fd:/# hdfs dfs -put bonjour.txt /bonjour.txt
2023-10-24 15:02:25,469 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
root@d307457e10fd:/# hdfs dfs -ls -R /
drwxr-xr-x - root supergroup 0 2023-10-17 11:58 /TPs
drwxr-xr-x - root supergroup 0 2023-10-17 11:58 /TPs/codes
drwxr-xr-x - root supergroup 0 2023-10-17 11:57 /TPs/data
drwxrwxrwt - root root 0 2023-10-17 16:34 /app-logs
drwxrwx--- - root root 0 2023-10-17 16:34 /app-logs/root
drwxrwx--- - root root 0 2023-10-24 14:47 /app-logs/root/logs-tfile
drwxrwx--- - root root 0 2023-10-17 16:36 /app-logs/root/logs-tfile/application_1697556801243_0002
-rw-r----- 3 root root 625277 2023-10-17 16:36 /app-logs/root/logs-tfile/application_1697556801243_0002/33fb7e5dc546_33125
drwxrwx--- - root root 0 2023-10-17 16:41 /app-logs/root/logs-tfile/application_1697556801243_0003
-rw-r----- 3 root root 560065 2023-10-17 16:41 /app-logs/root/logs-tfile/application_1697556801243_0003/33fb7e5dc546_33125
drwxrwx--- - root root 0 2023-10-17 16:44 /app-logs/root/logs-tfile/application_1697556801243_0004
-rw-r----- 3 root root 620260 2023-10-17 16:44 /app-logs/root/logs-tfile/application_1697556801243_0004/33fb7e5dc546_33125
drwxrwx--- - root root 0 2023-10-17 16:45 /app-logs/root/logs-tfile/application_1697556801243_0005
-rw-r----- 3 root root 647939 2023-10-17 16:45 /app-logs/root/logs-tfile/application_1697556801243_0005/33fb7e5dc546_33125
drwxrwx--- - root root 0 2023-10-17 17:24 /app-logs/root/logs-tfile/application_1697556801243_0006
-rw-r----- 3 root root 266782 2023-10-17 17:24 /app-logs/root/logs-tfile/application_1697556801243_0006/33fb7e5dc546_33125
drwxrwx--- - root root 0 2023-10-17 17:27 /app-logs/root/logs-tfile/application_1697556801243_0007
-rw-r----- 3 root root 428561 2023-10-17 17:27 /app-logs/root/logs-tfile/application_1697556801243_0007/33fb7e5dc546_33125
drwxrwx--- - root root 0 2023-10-17 17:28 /app-logs/root/logs-tfile/application_1697556801243_0008
-rw-r----- 3 root root 427856 2023-10-17 17:28 /app-logs/root/logs-tfile/application_1697556801243_0008/33fb7e5dc546_33125
drwxrwx--- - root root 0 2023-10-17 17:29 /app-logs/root/logs-tfile/application_1697556801243_0009
-rw-r----- 3 root root 261208 2023-10-17 17:29 /app-logs/root/logs-tfile/application_1697556801243_0009/33fb7e5dc546_33125
drwxrwx--- - root root 0 2023-10-17 18:00 /app-logs/root/logs-tfile/application_1697556801243_0010
-rw-r----- 3 root root 436466 2023-10-17 18:00 /app-logs/root/logs-tfile/application_1697556801243_0010/33fb7e5dc546_33125
drwxrwx--- - root root 0 2023-10-17 20:46 /app-logs/root/logs-tfile/application_1697556801243_0011
-rw-r----- 3 root root 295569 2023-10-17 20:46 /app-logs/root/logs-tfile/application_1697556801243_0011/33fb7e5dc546_46535
drwxrwx--- - root root 0 2023-10-24 15:01 /app-logs/root/logs-tfile/application_1698155995860_0001
-rw-r----- 3 root root 434538 2023-10-24 15:01 /app-logs/root/logs-tfile/application_1698155995860_0001/33fb7e5dc546_42905
-rw-r--r-- 3 root supergroup 23 2023-10-24 15:02 /bonjour.txt
drwxr-xr-x - root supergroup 0 2023-10-17 14:48 /data
-rw-r--r-- 3 root supergroup 3369248 2023-10-17 14:37 /data/pg135.txt

```

Consulter le contenu de bonjour.txt sous HDFS

```

root@d307457e10fd:/# hdfs dfs -cat /bonjour.txt
2023-10-24 15:04:19,483 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
Bonjour Hadoop et HDFS
root@d307457e10fd:/# hdfs dfs -cat /bonjour.txt | more
2023-10-24 15:05:36,974 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
Bonjour Hadoop et HDFS
root@d307457e10fd:/# hdfs dfs -tail /bonjour.txt
2023-10-24 15:06:02,346 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
Bonjour Hadoop et HDFS

```

Supprimer bonjour.txt puis le copier d'une autre manière puis ajouter les droits d'écriture pour le groupe et les autres

```

root@d307457e10fd:/# hdfs dfs -rm /bonjour.txt
Deleted /bonjour.txt
root@d307457e10fd:/# hdfs dfs -copyFromLocal bonjour.txt /bonjour.txt
2023-10-24 15:10:54,278 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
root@d307457e10fd:/# hdfs dfs -ls /
Found 12 items
drwxr-xr-x - root supergroup 0 2023-10-17 11:58 /TPs
drwxrwxrwt - root root 0 2023-10-17 16:34 /app-logs
-rw-r--r-- 3 root supergroup 23 2023-10-24 15:10 /bonjour.txt
drwxr-xr-x - root supergroup 0 2023-10-17 14:48 /data
drwxr-xr-x - root supergroup 0 2023-10-24 14:59 /dossier
drwxr-xr-x - root supergroup 0 2023-10-17 16:16 /input
drwxr-xr-x - root supergroup 0 2023-10-24 15:00 /output
drwxr-xr-x - root supergroup 0 2023-10-17 11:09 /rmstate
drwxr-xr-x - root supergroup 0 2023-10-17 11:56 /root
drwx----- - root supergroup 0 2023-10-17 16:33 /tmp
drwxr-xr-x - root supergroup 0 2023-10-17 11:40 /user
drwxr-xr-x - root supergroup 0 2023-10-17 11:40 /var
root@d307457e10fd:/# hdfs dfs -chmod go+w /bonjour.txt
root@d307457e10fd:/# hdfs dfs -ls /
Found 12 items
drwxr-xr-x - root supergroup 0 2023-10-17 11:58 /TPs
drwxrwxrwt - root root 0 2023-10-17 16:34 /app-logs
-rw-rw-rw- 3 root supergroup 23 2023-10-24 15:10 /bonjour.txt
drwxr-xr-x - root supergroup 0 2023-10-17 14:48 /data
drwxr-xr-x - root supergroup 0 2023-10-24 14:59 /dossier
drwxr-xr-x - root supergroup 0 2023-10-17 16:16 /input
drwxr-xr-x - root supergroup 0 2023-10-24 15:00 /output
drwxr-xr-x - root supergroup 0 2023-10-17 11:09 /rmstate
drwxr-xr-x - root supergroup 0 2023-10-17 11:56 /root
drwx----- - root supergroup 0 2023-10-17 16:33 /tmp
drwxr-xr-x - root supergroup 0 2023-10-17 11:40 /user
drwxr-xr-x - root supergroup 0 2023-10-17 11:40 /var
root@d307457e10fd:/#

```

Enlever les droits de lecture pour le groupe et les autres puis le déplacer vers le répertoire **/dossier**

```
root@d307457e10fd:/# hdfs dfs -chmod go-r /bonjour.txt
root@d307457e10fd:/# hdfs dfs -ls /
Found 12 items
drwxr-xr-x - root supergroup 0 2023-10-17 11:58 /TPs
drwxrwxrwt - root root 0 2023-10-17 16:34 /app-logs
-rw-rw-rw- 3 root supergroup 23 2023-10-24 15:10 /bonjour.txt
drwxr-xr-x - root supergroup 0 2023-10-17 14:48 /data
drwxr-xr-x - root supergroup 0 2023-10-24 14:59 /dossier
drwxr-xr-x - root supergroup 0 2023-10-17 16:16 /input
drwxr-xr-x - root supergroup 0 2023-10-24 15:00 /output
drwxr-xr-x - root supergroup 0 2023-10-17 11:09 /rmstate
drwxr-xr-x - root supergroup 0 2023-10-17 11:56 /root
drwx----- - root supergroup 0 2023-10-17 16:33 /tmp
drwxr-xr-x - root supergroup 0 2023-10-17 11:40 /user
drwxr-xr-x - root supergroup 0 2023-10-17 11:40 /var
root@d307457e10fd:/# hdfs dfs -mv /bonjour.txt /dossier/bonjour.txt
root@d307457e10fd:/# hdfs dfs -ls /dossier
Found 1 items
-rw-rw-rw- 3 root supergroup 23 2023-10-24 15:10 /dossier/bonjour.txt
```

Télécharger le fichier `bonjour.txt` vers le système de fichier locale du **NameNode** sous le nom de `bien.txt` puis copier le fichier `bonjour.txt` dans HDFS et le coller sous le nom `bien.txt` sous le répertoire **/dossier** dans HDFS puis on affiche le nombre des sous-dossiers, des fichiers et octets occupés dans **/dossier** puis on supprime le répertoire **/dossier** avec tout son contenu

```
root@d307457e10fd:/# hdfs dfs -get /dossier/bonjour.txt bien.txt
2023-10-25 05:56:12,648 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
root@d307457e10fd:/# ls
KEYS bin boot dev etc hadoop-data input.txt lib64 media opt reducerWC.py run sbin sys usr
bien.txt bonjour.txt data entrypoint.sh hadoop home lib mapperWC.py mnt proc root run.sh srv tmp var
root@d307457e10fd:/# hdfs dfs -cp /dossier/bonjour.txt /dossier/bien.txt
2023-10-25 05:57:29,555 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-25 05:57:30,531 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
root@d307457e10fd:/# hdfs dfs -ls /dossier
Found 2 items
-rw-r--r-- 3 root supergroup 23 2023-10-25 05:57 /dossier/bien.txt
-rw-rw-rw- 3 root supergroup 23 2023-10-24 15:10 /dossier/bonjour.txt
root@d307457e10fd:/# hdfs dfs -count -h /dossier
1 2 46 /dossier
root@d307457e10fd:/# hdfs dfs -rm /dossier/bonjour.txt
Deleted /dossier/bonjour.txt
root@d307457e10fd:/# hdfs dfs -ls /dossier
Found 1 items
-rw-r--r-- 3 root supergroup 23 2023-10-25 05:57 /dossier/bien.txt
root@d307457e10fd:/# hdfs dfs -rm -r /dossier
Deleted /dossier
root@d307457e10fd:/# hdfs dfs -ls /
Found 10 items
drwxr-xr-x - root supergroup 0 2023-10-17 11:58 /TPs
drwxrwxrwt - root root 0 2023-10-17 16:34 /app-logs
drwxr-xr-x - root supergroup 0 2023-10-17 14:48 /data
drwxr-xr-x - root supergroup 0 2023-10-17 16:16 /input
drwxr-xr-x - root supergroup 0 2023-10-24 15:00 /output
drwxr-xr-x - root supergroup 0 2023-10-17 11:09 /rmstate
drwxr-xr-x - root supergroup 0 2023-10-17 11:56 /root
drwx----- - root supergroup 0 2023-10-17 16:33 /tmp
drwxr-xr-x - root supergroup 0 2023-10-17 11:40 /user
drwxr-xr-x - root supergroup 0 2023-10-17 11:40 /var
root@d307457e10fd:/#
```

Partie 3 : Création d'une arborescence et téléchargement de fichier

Copier les fichiers de différentes manières

```
ASUS@DESKTOP-T96Q426 MINGW64 /e/Documents/Academic/ISIMM/Semestre 7/Big Data/TP1
$ docker exec namenode sh -c 'cat > /data/pg4300.txt' < pg4300.txt
```

```
ASUS@DESKTOP-T96Q426 MINGW64 /e/Documents/Academic/ISIMM/Semestre 7/Big Data/TP1
$ docker cp purchases.txt namenode:/data/purchases.txt
```

Téléchargement des fichiers en utilisant la commande **curl**

```
root@d307457e10fd:/data# curl -o pg135.txt https://www.gutenberg.org/cache/epub/135/pg135.txt
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
           Dload  Upload   Total   Spent    Left   Speed
100 3290k  100 3290k    0     0  428k      0  0:00:07  0:00:07 --:--:-- 797k
root@d307457e10fd:/data# ls
pg135.txt pg4300.txt purchases.txt
root@d307457e10fd:/data#
```

Créer l'arborescence sous HDFS

```
root@d307457e10fd:/# hdfs dfs -mkdir /TPs
root@d307457e10fd:/# hdfs dfs -mkdir /TPs/data
root@d307457e10fd:/# hdfs dfs -mkdir /TPs/codes
```

Télécharger les fichiers sous HDFS

```
root@d307457e10fd:/# hdfs dfs -put data/purchases.txt /TPs/data/
2023-10-25 10:45:38,817 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2023-10-25 10:46:03,466 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
root@d307457e10fd:/# hdfs dfs -put data/pg4300.txt /TPs/data/
2023-10-25 10:46:21,082 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
root@d307457e10fd:/# hdfs dfs -put data/pg135.txt /TPs/data/
2023-10-25 10:46:34,937 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
```

Lister le contenu de /TPs/data




```
root@d307457e10fd:/# hdfs dfs -ls /TPs/data
Found 3 items
-rw-r--r--  3 root supergroup    3369248 2023-10-25 10:46 /TPs/data/pg135.txt
-rw-r--r--  3 root supergroup    1586410 2023-10-25 10:46 /TPs/data/pg4300.txt
-rw-r--r--  3 root supergroup   211312924 2023-10-25 10:46 /TPs/data/purchases.txt
root@d307457e10fd:/#
```

Partie 4 : Etat du cluster

Connecter au **NameNode** via le navigateur

Browse Directory

Go!



Hadoop, 2019.

Consulter la page *Overview*

Summary

Security is off.

Safemode is off.

133 files and directories, 85 blocks (85 replicated blocks, 0 erasure coded block groups) = 218 total filesystem object(s).

Heap Memory used 120.33 MB of 295 MB Heap Memory. Max Heap Memory is 1.72 GB.

Non Heap Memory used 59.42 MB of 60.69 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	250.98 GB
Configured Remote Capacity:	0 B
DFS Used:	214.57 MB (0.08%)
Non DFS Used:	6.1 GB
DFS Remaining:	231.86 GB (92.38%)
Block Pool Used:	214.57 MB (0.08%)
DataNodes usages% (Min/Median/Max/stdDev):	0.08% / 0.08% / 0.08% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0

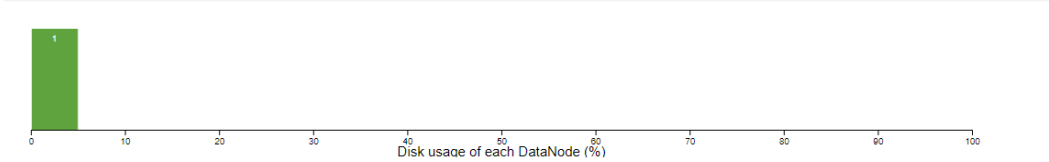
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	85
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Tue Oct 24 14:58:46 +0100 2023
Last Checkpoint Time	Tue Oct 24 14:58:52 +0100 2023
Enabled Erasure Coding Policies	RS-6-3-1024k

NameNode Journal Status

[Hadoop](#)
[Overview](#)
[Datanodes](#)
[Datanode Volume Failures](#)
[Snapshot](#)
[Startup Progress](#)
[Utilities ▾](#)

✔ In service
🔴 Down
🟢 Decommissioning
🟠 Decommissioned
🔴 Decommissioned & dead










🟢 Entering Maintenance
🟠 In Maintenance
🔴 In Maintenance & dead



L'arbre de fichiers sous Hadoop

Show25entries

Search:

<input type="checkbox"/>		Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>		drwxr-xr-x	root	supergroup	0 B	Oct 17 12:58	0	0 B	TPs	
<input type="checkbox"/>		drwxrwxrwt	root	root	0 B	Oct 17 17:34	0	0 B	app-logs	
<input type="checkbox"/>		drwxr-xr-x	root	supergroup	0 B	Oct 17 15:48	0	0 B	data	
<input type="checkbox"/>		drwxr-xr-x	root	supergroup	0 B	Oct 17 17:16	0	0 B	input	
<input type="checkbox"/>		drwxr-xr-x	root	supergroup	0 B	Oct 24 16:00	0	0 B	output	
<input type="checkbox"/>		drwxr-xr-x	root	supergroup	0 B	Oct 17 12:09	0	0 B	rmstate	
<input type="checkbox"/>		drwxr-xr-x	root	supergroup	0 B	Oct 17 12:56	0	0 B	root	
<input type="checkbox"/>		drwxr-xr-x	root	supergroup	0 B	Oct 17 17:00	0	0 B	root	