

Université Hassan II de Casablanca –

ENSET Mohammedia

Département : Mathématiques et Informatique

Filière : II-BDCC 2
Module : Virtualisation et Cloud Computing
Année Universitaire : 2025/2026

COMPTE RENDU DU TP N°1

TP 1 : Manipulation du système de fichiers distribué HDFS

Réalisé par :
ELGHOUALI Zakariae
Encadré par :
Abdelmajid BOUSSELHAM
Date de rendu :
24/11/2025

Sommaire

1. Problématique
2. Partie 1 — Création de l'espace de travail HDFS
Partie 2 : Segmentation (VLAN), Propagation (VTP) et Routage
3. Partie 2 — Transfert de données vers HDFS
4. Partie 3 — Consultation et analyse des fichiers HDFS
5. Partie 4 — Gestion des fichiers dans HDFS
6. Partie 5 — Suppression et nettoyage

1. Problématique

Une entreprise dispose d'un cluster Hadoop configuré avec le système de fichiers distribué HDFS afin de stocker et gérer de grands volumes de données de manière fiable et distribuée.

Votre mission consiste à prendre en main HDFS en réalisant les opérations de base de gestion des fichiers à l'aide des commandes HDFS.

L'objectif est de comprendre comment les données sont organisées, stockées et manipulées dans HDFS par rapport à un système de fichiers classique.

2. Partie 1 : EtherChannel & Load Balancing

1. Créer dans HDFS un répertoire principal nommé /tp-hdfs.

2. Créer dans HDFS la structure de répertoires suivante :

```
/tp-hdfs
    └── input
    └── output
        └── archive
```

```
bash-4.2$ hdfs dfs -mkdir /tp-hdfs/input
bash-4.2$ hdfs dfs -mkdir /tp-hdfs/output
bash-4.2$ hdfs dfs -mkdir /tp-hdfs/archive
```

3. Vérifier que la structure de répertoires a été correctement créée.

```
bash-4.2$ hdfs dfs -ls -R /tp-hdfs
drwxr-xr-x  - hadoop supergroup          0 2026-02-09 16:54 /tp-hdfs/archive
drwxr-xr-x  - hadoop supergroup          0 2026-02-09 16:51 /tp-hdfs/input
drwxr-xr-x  - hadoop supergroup          0 2026-02-09 16:51 /tp-hdfs/output
```

3. Partie 2 — Transfert de données vers HDFS

4. Créer localement un fichier texte contenant au minimum dix lignes de texte.

```
bash-4.2$ echo -e "Ligne 1\nLigne 2\nLigne 3\nLigne 4\nLigne 5\nLigne 6\nLigne 7\nLigne
8\nLigne 9\nLigne 10" > donnees.txt
bash-4.2$
```

5. Copier ce fichier depuis le système de fichiers local vers le répertoire /tp-hdfs/input dans HDFS.

```
bash-4.2$ hdfs dfs -put donnees.txt /tp-hdfs/input/
bash-4.2$ ~
```

6. Lister le contenu du répertoire /tp-hdfs/input afin de vérifier la présence du fichier.

```
bash-4.2$ hdfs dfs -ls /tp-hdfs/input
Found 1 items
-rw-r--r-- 3 hadoop supergroup          81 2026-02-09 16:57 /tp-hdfs/input/donnees.txt
bash-4.2$
```

4. Partie 3 — Consultation et analyse des fichiers HDF

7. Afficher le contenu du fichier stocké dans HDFS sans le rapatrier en local.

```
bash-4.2$ hdfs dfs -cat /tp-hdfs/input/donnees.txt
Ligne 1
Ligne 2
Ligne 3
Ligne 4
Ligne 5
Ligne 6
Ligne 7
Ligne 8
Ligne 9
Ligne 10
bash-4.2$
```

8. Afficher les informations détaillées du fichier HDFS (taille, propriétaire, date, facteur de réPLICATION).

```
bash-4.2$ hdfs dfs -ls /tp-hdfs/input/donnees.txt
-rw-r--r-- 3 hadoop supergroup          81 2026-02-09 16:57 /tp-hdfs/input/donnees.txt
bash-4.2$
```

Partie 4 — Gestion des fichiers dans HDFS

10. Copier le fichier présent dans /tp-hdfs/input vers le répertoire /tp-hdfs/archive.

```
bash-4.2$ hdfs dfs -cp /tp-hdfs/input/donnees.txt /tp-hdfs/archive/  
bash-4.2$
```

11. Renommer le fichier copié dans le répertoire /tp-hdfs/archive.

```
bash-4.2$ hdfs dfs -mv /tp-hdfs/archive/donnees.txt /tp-hdfs/archive/donnees_old.txt  
bash-4.2$
```

12. Déplacer le fichier original du répertoire /tp-hdfs/input vers le répertoire /tp-hdfs/output.

13. Vérifier la présence des fichiers dans les répertoires /tp-hdfs/output et /tp-hdfs/archive.

```
bash-4.2$ hdfs dfs -ls -R /tp-hdfs  
drwxr-xr-x  - hadoop supergroup          0 2026-02-09 16:59 /tp-hdfs/archive  
-rw-r--r--  3 hadoop supergroup         81 2026-02-09 16:58 /tp-hdfs/archive/donnees_o  
ld.txt  
drwxr-xr-x  - hadoop supergroup          0 2026-02-09 16:57 /tp-hdfs/input  
-rw-r--r--  3 hadoop supergroup         81 2026-02-09 16:57 /tp-hdfs/input/donnees.txt  
drwxr-xr-x  - hadoop supergroup          0 2026-02-09 16:51 /tp-hdfs/output  
bash-4.2$
```

Partie 5 — Suppression et nettoyage

14. Supprimer un fichier présent dans HDFS.

```
bash-4.2$ hdfs dfs -rm /tp-hdfs/output/donnees.txt  
rm: '/tp-hdfs/output/donnees.txt': No such file or directory  
bash-4.2$
```

15. Supprimer le répertoire /tp-hdfs/archive ainsi que son contenu.

16. Vérifier que les éléments supprimés ne sont plus accessibles dans HDFS. 16. Vérifier que les éléments supprimés ne sont plus accessibles dans HDFS.

```
bash-4.2$ hdfs dfs -ls -R /tp-hdfs
drwxr-xr-x  - hadoop supergroup          0 2026-02-09 16:59 /tp-hdfs/archive
-rw-r--r--  3 hadoop supergroup         81 2026-02-09 16:58 /tp-hdfs/archive/donnees_o
ld.txt
drwxr-xr-x  - hadoop supergroup          0 2026-02-09 16:57 /tp-hdfs/input
-rw-r--r--  3 hadoop supergroup         81 2026-02-09 16:57 /tp-hdfs/input/donnees.txt
drwxr-xr-x  - hadoop supergroup          0 2026-02-09 16:51 /tp-hdfs/output
bash-4.2$
```

```
bash-4.2$ exit
exit
PS C:\Users\Administrator\Desktop\BigData\TD1>
```

```
PS C:\Users\Administrator\Desktop\BigData\TD1> docker-compose down
[+] down 9/9
✓Container td1-resourcemanager-1 Removed           1.6s
✓Container td1-namenode-1     Removed             0.9s
✓Container td1-datanode4-1   Removed             1.3s
✓Container td1-datanode2-1   Removed             0.5s
✓Container td1-datanode5-1   Removed             1.2s
✓Container td1-datanode1-1   Removed             1.5s
✓Container td1-nodemanager-1 Removed             1.3s
✓Container td1-datanode3-1   Removed             1.4s
✓Network td1_default        Removed             0.3s
PS C:\Users\Administrator\Desktop\BigData\TD1>
```