

**NOM : ACHIR**  
**Prénom : Ounissa**

## **Chapitre : Hadoop Distributed File System (HDFS)**

### 1- Voir la commande `hadoopfs` :

Après ouverture du terminal de la VM

**b- Hdfs dfs : liste les commandes de hadoop hdfs liées à la gestion des fichiers et répertoires.**

### 2- Créer un répertoire dans HDFS :

**a- La commande pour voir le contenu du répertoire courant : `hdfs dfs -ls`**

Exécutez la commande `-ls` à nouveau, mais cette fois spécifier le dossier HDFS racine : **`hdfs dfs -ls -d`**

**b- Créer le répertoire test : `hdfs dfs -mkdir test`**

**c- Vérifier si le répertoire test est créé avec succès : `hdfs dfs -ls`**

```
[cloudera@quickstart ~]$ hdfs dfs -mkdir test
mkdir: `test': File exists
[cloudera@quickstart ~]$ hdfs dfs -mkdir test/test1
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x  - cloudera cloudera          0 2023-10-20 09:55 test
```

**d- Créer les sous répertoires test/test1 et test/test2/test3 :**

- **`hdfs dfs _mkdir test/test1`**
- **`hdfs dfs -mkdir test/test2/test3` ou `hdfs dfs -mkdir -p test/test2/test3` pour créer récursivement les sous répertoires.**

```
[cloudera@quickstart ~]$ hdfs dfs -mkdir test/test2/test3
mkdir: `test/test2/test3': File exists
```

**e- Utiliser la commande `-ls` pour voir le contenu du répertoire courant /user/root**  
Taper la commande `-ls -R` Que fait la commande `-R` ?

**`hdfs dfs -ls -R`**

```
drwxr-xr-x  - cloudera cloudera          0 2023-10-20 09:55 test/test1
drwxr-xr-x  - cloudera cloudera          0 2023-10-20 10:02 test/test2
drwxr-xr-x  - cloudera cloudera          0 2023-10-20 10:02 test/test2/test3
```

### 3- Suppression de répertoire :

**a- Supprimer le répertoire test2 (ainsi que tous les sous répertoires)**

- **`hdfs dfs -rm -r test/test2`**
- **`hdfs dfs -ls -R` pour s'assurer de la suppression.**

```
[cloudera@quickstart ~]$ hdfs dfs -rm -r test/test2
23/10/20 10:16:55 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Empty interval = 0 minutes.
Deleted test/test2
[cloudera@quickstart ~]$ hdfs dfs -ls -R
drwxr-xr-x  - cloudera cloudera          0 2023-10-20 10:20 test
drwxr-xr-x  - cloudera cloudera          0 2023-10-20 09:55 test/test1
[cloudera@quickstart ~]$
```

#### 4- Charger un fichier dans HDFS :

a- L'objectif est de mettre le fichier data.txt dans le répertoire test dans HDFS, pour cela, il se mettra dans le répertoire /formation/ateliers/hdfs

- **hdfs dfs -mkdir formation** : création du répertoire
- **hdfs dfs -mkdir formation/ateliers/hdfs** : création des sous répertoires
- **hdfs dfs -copyFromLocal /home/cloudera/Desktop/data.txt formation/ateliers/hdfs** : mettre le fichier data.txt dans le répertoire /formation/ateliers/hdfs
- **hdfs dfs -ls formation/ateliers/hdfs** : lister les fichiers dans ce répertoire pour s'assurer de l'existence du fichier data.txt.

b- Vérifier le contenu du fichier data.txt :

**hdfs dfs -cat formation/ateliers/hdfs/data.txt**

c- Copier le fichier data.txt dans le fichier test de HDFS

**hdfs dfs -cp formation/ateliers/hdfs/data.txt test**

d- Vérifier que le fichier existe bien dans test :

**hdfs dfs -ls test**

#### 5- Copier les fichiers dans HDFS :

a- Copier le fichier data.txt qui est dans test à test/test1/data2.txt

**hdfs dfs -cp test/data.txt test/test1/data2.txt** : on copie le fichier en renommant data.txt à data2.txt.

b- Vérifier l'existence de fichier à la fois dans test et test2 en une seule commande

**hdfs dfs -test -e test/data.txt && test/test1/data2.txt** retourne 0 si existe dans les deux.

ou

**hdfs dfs -ls test/data.txt test/test1/data2.txt** : liste les fichiers si ils existent.

c- Supprimer le fichier data2.txt

**hdfs dfs -rm test/test2/data2.txt**

#### 6- Visualiser le contenu d'un fichier dans HDFS :

a- Utiliser la commande -cat pour voir le contenu de test/data.txt

**hdfs dfs -cat test/data.txt**

b- Utiliser la commande -tail pour visualiser la fin du fichier test/data.txt

**hdfs dfs -tail test/data.txt**

#### 7- Récupérer un fichier de HDFS et le copier en local :

a- Copier le fichier test/data.txt de HDFS en local dans le répertoire /tmp et vérifier ensuite qu'il est bien copié.

- **hdfs dfs -mkdir tmp** : création du répertoire tmp.
- **hdfs dfs -copyToLocal test/data.txt tmp** : copié en local.
- **hdfs dfs -ls tmp** : lister les fichiers dans tmp pour vérifier si data existe.

#### 8- La commande getmerge :

**a-** Copier le fichier /formation/ateliers/hdfs/small\_blocks.txt dans hdfs dans le répertoire test, vous devriez donc avoir deux fichiers dans test ( data.txt et small\_blocks.txt )

- **hdfs dfs -copyFromLocal /home/cloudera/Desktop/small\_blocks.txt formation/ateliers/hdfs** : mettre le fichier small\_blocks.txt dans le répertoire /formation/ateliers/hdfs
- **hdfs dfs -ls formation/ateliers/hdfs** : lister les fichiers dans ce répertoire pour s'assurer de l'existence du fichier small\_blocks.txt.
- **hdfs dfs -copyFromLocal formation/ateliers/hdfs/small\_blocks.txt test** : mettre le fichier small\_blocks.txt dans le répertoire test.
- **hdfs dfs -ls test** : lister les fichiers dans ce répertoire pour s'assurer de l'existence du fichier small\_blocks.txt.

**b-** Utiliser la commande getmerge comme suit :

getmerge est utilisé pour fusionner tous les fichiers du répertoire mentionné dans un seul fichier local.

Puisque le répertoire test a deux fichiers data.txt et small\_blocks.txt on applique cette fusion sur test dans le fichier fichier\_fusion.txt .

**hdfs dfs -getmerge test fichier\_fusion.txt**