

# Atelier : Utiliser les commandes HDFS

## L'objectif :

- Se familiariser avec le stockage des données dans HDFS
- Manipuler, Copier, transférer, supprimer les données dans HDFS
- Dans cet atelier, vous allez effectuer ces commandes pour vous aider à être à l'aise avec HDFS.

**Emplacement du fichier** : /formation/ateliers/hdfs/

**Réalisation** : Vous allez copier, supprimer plusieurs fichiers de données dans HDFS

**Chapitre correspondant** : HadoopDistributed File System (HDFS)

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

- a- Ouvrir le terminal de la ligne de commande de votre VM
- b- Taper : **`hadoop fs`** Que fais cette commande ?

La commande `hadoopfs` permet d'intégrer des options pour l'exécution des commandes HDFS, comme la copie des fichiers à partir d'un dossier local dans HDFS, la récupération d'un fichier à partir de HDFS, copier et déplacer des fichiers, et la création et la suppression des répertoires.

### 2- Créer un répertoire dans HDFS :

- a- Taper la commande `-ls` pour voir le contenu du répertoire courant.

```
hadoop fs -ls
```

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

```
hadoop fs -ls /
```

Important : `ls` affiche le contenu de répertoire de courant de l'utilisateur `/user/coudera` ou `/user/root` selon l'utilisateur avec lequel vous vous êtes connecté

- b- Créer le répertoire test

```
hadoop fs -mkdir test
```

- c- Vérifier si le répertoire test est créé avec succès

```
hadoop fs -ls
```

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

```
hadoop fs -mkdir test/test1
```

```
hadoop fs -mkdir -p test/test2/test3
```

La commande `-p` peut être utilisée pour créer plusieurs répertoires sans que les répertoires parents soient créés

- e- Utiliser la commande `-ls` pour voir le contenu du répertoire courant `/user/root`

```
hadoop fs -ls
```

Taper la commande **`-ls -R`** Que fait la commande `-R` ?

```
hadoop fs -ls-R
```

3- Suppression de répertoire :

- a- Supprimer le répertoire test2 (et tous les sous répertoires)

```
hadoop fs -rm -R test/test2
```

Remarque : Les données supprimées sont dans .Trash Ce dossier .Trash est vidé automatiquement après un laps de temps configuré.

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 mettre dans le répertoire /formation/ateliers/hdfs/

```
cd /root/.../formation/ateliers/hdfs/
```

- b- Vérifier le contenu du fichier data.txt

```
tail data.txt
```

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

```
hadoop fs -put data.txt test/
```

- d- Vérifier que le fichier existe bien dans test

```
hadoopfs -ls test
```

5- Copier les fichiers dans HDFS :

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

```
hadoop fs -cp test/data.txt test/test1/data2.txt
```

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

```
hadoop fs -ls -R test
```

- c- Supprimer le fichier data2.txt

```
hadoop fs -rm test/test1/data2.txt
```

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

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

```
hadoop fs -cat test/data.txt
```

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

```
hadoop fs -tail test/data.txt
```

Cette commande permet de récupérer que les 20 dernières lignes de 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.

```
hadoop fs -get test/data.txt /tmp/
```

```
cd /tmp
```

```
ls data*
```

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 )

```
hadoop fs -put /root/formation/hdfs/small_blocks.txt test/
```

- b- Utiliser la commande getmerge comme suit :

```
hadoop fs -getmerge -nltest /tmp/merged.txt
```

- c- Ouvrez le fichier merged.txt. Que fait la commande getmerge ?

Les deux fichiers du répertoire test dans HDFS ont été fusionnés dans un seul fichier dans la machine locale.

### Résultat :

Vous êtes en mesure d'exécuter les différentes commandes HDFS, comme la création de répertoires, mettre les fichiers dans HDFS, la copie de fichiers sur HDFS, et la suppression des fichiers et des dossiers..