

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.

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

Chapitre correspondant : Hadoop Distributed File System (HDFS)

1- Voir la commande hadoopfs :

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

2- Créer un répertoire dans HDFS :

- a- Taper la commande **-ls** pour voir le contenu du répertoire courant.
Exécutez la commande **-ls** à nouveau, mais cette fois spécifier le dossier HDFS racine
- b- Créer le répertoire test
- c- Vérifier si le répertoire test est créé avec succès
- d- Créer les sous répertoires test/test1 et test/test2/test3
- e- Utiliser la commande **-ls** pour voir le contenu du répertoire courant /user/root
Taper la commande **-ls -R** Que fait la commande **-R** ?

3- Suppression de répertoire :

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

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/
- b- Vérifier le contenu du fichier data.txt
- c- Copier le fichier data.txt dans le fichier test de HDFS
- d- Vérifier que le fichier existe bien dans test

5- Copier les fichiers dans HDFS :

- a- Copier le fichier data.txt qui est dans test à test/test1/data2.txt
- b- Vérifier l'existence de fichier à la fois dans test et test2 en une seule commande

- c- Supprimer le fichier data2.txt
- 6- Visualiser le contenu d'un fichier dans HDFS :
 - a- Utiliser la commande `-cat` pour voir le contenu de test/data.txt
 - b- Utiliser la commande `-tail` pour visualiser la fin du fichier 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.
- 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)
 - b- Utiliser la commande getmerge comme suit :
 - c- Ouvrez le fichier merged.txt. Que fait la commande getmerge ?

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..