### **Hadoop** [https://insatunisia.github.io/TP-BigData/tp1/](https://insatunisia.github.io/TP-BigData/tp1/#hadoop)

#### Exercice 1 : les commandes

Toutes les commandes interagissant avec le système HDFS commencent par **hdfs dfs**. Ensuite, les options ajoutées sont inspirées des commandes Linux.

* Lancer le terminal linux

1. Afficher la liste des fichiers dans un répertoire

hdfs dfs -ls /user/hadoop

hadoop fs- ls : Afficher le contenu du répertoire courant

1. Créer un répertoire

hdfs dfs -mkdir /user/hadoop/input

1. Copier un fichier depuis le système local vers HDFS

hdfs dfs -put data.txt /user/hadoop/input

hadoop fs-put data.txt : Upload un fichier dans hadoop (vers le répertoire courant hdfs)

1. Copier un fichier depuis HDFS vers le système local

hdfs dfs -get /user/hadoop/input/data.txt /home/user/

1. Afficher le contenu d’un fichier

hdfs dfs -cat /user/hadoop/input/data.txt

hadoop fs-cat data.txt | less : afficher le fichier page par page

1. Supprimer un fichier

hdfs dfs -rm /user/hadoop/input/data.txt

1. Supprimer un répertoire

hdfs dfs -rm -r /user/hadoop/input

hadoop fs-rm-f-r /user/hadoop/input : Supprime un répertoire, et son contenu récursivement

1. Afficher les premières lignes d’un fichier

hdfs dfs -cat /user/hadoop/input/data.txt | head -n 10

hadoop fs-tail data.txt : afficher les dernières lignes du fichier

1. Copier un fichier ou un répertoire dans HDFS

hdfs dfs -cp /user/hadoop/input /user/hadoop/backup

1. Déplacer un fichier ou un répertoire dans HDFS

hdfs dfs -mv /user/hadoop/input/data.txt /user/hadoop/output

hadoop fs-mv data.txt data1.txt : Renommer le fichier

1. Afficher les statistiques d’un fichier

hdfs dfs -stat /user/hadoop/input/data.txt

1. Compter le nombre de lignes, mots et caractères

hdfs dfs -cat /user/hadoop/input/data.txt | wc

1. Afficher l’espace utilisé par un répertoire

hdfs dfs -du -s /user/hadoop/input

1. Afficher la taille d’un fichier ou d’un répertoire

hdfs dfs -du -h /user/hadoop/input

#### ****Commandes pour gérer les permissions****

1. Modifier les permissions d’un fichier ou d’un répertoire

hdfs dfs -chmod 755 /user/hadoop/input/data.txt

1. Modifier le propriétaire d’un fichier ou d’un répertoire

hdfs dfs -chown hadoop /user/hadoop/input/data.txt

1. Modifier le groupe d’un fichier ou d’un répertoire

hdfs dfs -chgrp hadoop\_group /user/hadoop/input/data.txt

#### ****Commandes pour diagnostiquer et vérifier****

1. Vérifier si un fichier ou répertoire existe

hdfs dfs -test -e /chemin/vers/lefichier

1. Afficher les informations de santé du NameNode

hdfs dfsadmin -report

1. Afficher les blocs d’un fichier

hdfs fsck /user/hadoop/input/data.txt -files -blocks

#### ****Télécharger des fichiers HDFS compressés (gzip, zip, etc.)****

1. Utilisez cette commande si le fichier est compressé :

hdfs dfs -text /chemin/vers/fichier\_compressé.gz

1. Liste récursive des fichiers

hdfs dfs -rm -r /chemin/vers/lerepertoire/\*

**Exercice 2**

Ecrire un programme java qui permet de lire le contenu d'un fichier texte à partir de hdfs.

java -cp .:/etc/hadoop/conf:/usr/lib/hadoop/lib/*:/usr/lib/hadoop/.//*:/usr/lib/hadoop-hdfs/./:/usr/lib/hadoop-hdfs/lib/*:/usr/lib/hadoop-hdfs/.//*:/usr/lib/hadoop-yarn/lib/*:/usr/lib/hadoop-yarn/.//*:/usr/lib/hadoop-mapreduce/lib/\*:/usr/lib/hadoop-mapreduce/.//\* HDFSWriter /user/cloudera/monfichier.txt "Hello World Hello World Hello World"

Ecrire un programme java qui permet d'écrire un texte dans un fichier sur hdfs.

java -cp .:/etc/hadoop/conf:/usr/lib/hadoop/lib/*:/usr/lib/hadoop/.//*:/usr/lib/hadoop-hdfs/./:/usr/lib/hadoop-hdfs/lib/*:/usr/lib/hadoop-hdfs/.//*:/usr/lib/hadoop-yarn/lib/*:/usr/lib/hadoop-yarn/.//*:/usr/lib/hadoop-mapreduce/lib/\*:/usr/lib/hadoop-mapreduce/.//\* HDFSReader /user/cloudera/monfichier.txt