HDFS Commands

ls

- This command is used to list all files on the current working directory
- Example
    - hdfs dfs -ls /

mkdir

- this command is used to create new directory. In Hadoop there is no home directory by default.
- example
    - hdfs dfs -mkdir /user
    - hdfs dfs -mkdir /user/youssef

touchz

- used to create an empty file
- example
    - hdfs dfs -touchz /user/Youssef/file.txt

put

- used to copy files/directories from local system to hdfs store. This is the most important command. Local file system means the files present on the OS
- example
    - hdfs dfs -put /home/ubuntu/Desktop/file.txt /user/youssef/
        - this will copy file local file system into to the HDFS

copyFromLocal

- used to copy files/directories from local system to hdfs store. This is the most important command. Local file system means the files present on the OS
- example
    - hdfs dfs -copyFromLocal /home/ubuntu/Desktop/file.txt /user/youssef/
        - this will copy file local file system into to the HDFS

cat

- used to print the file content
- example
    - hdfs dfs -cat /user/youssef/file.txt

copyToLocal

- used to copy files/directories from the HDFS to the local file system
- example
    - hdfs dfs -copyToLocal /user/youssef/file.txt /home/ubuntu/Desktop
        - this will copy file from the HDFS to the Desktop of Ubuntu system

get

- used to copy files/directories from the HDFS to the local file system
- example
    - hdfs dfs -get /user/youssef/file.txt /home/ubuntu/Desktop
        - this will copy file from the HDFS to the Desktop of Ubuntu system

moveFromLocal

- used to move files from the local file system to the HDFS
- example
    - hdfs dfs -moveFromLocal /home/ubuntu/Desktop/file.txt /user/youssef
        - this will move the file.txt to the HDFS → to /user/youssef

cp

- used to copy files inside the HDFS
- example
    - hdfs dfs -cp /user /user_copied

mv

- used to move files inside the HDFS
- example
    - hdfs dfs -mv /user/file.txt /user_copied/
        - this will move the file.txt inside the user_copied directory

rmr

- used to deletes a file from HDFS recursively. It is very useful command when you want to delete a non-empty directory
- example
    - hdfs dfs -rmr /user_copied
        - this will delete the content inside the directory then delete the directory itself

du

- it will give the size of each file in directory
- example
    - hdfs dfs -du /user/youssef
        - get the directory size
        - displays the size in bytes, disk space consumed, and fie or directory name
    - hdfs dfs -du -s /labb.txt
        - this will only give the size of the **/lab.txt** file

stat

- used to give the last modified time of the directory or path.
- It will give the states of the directory or file
- Example
    - hdfs dfs -stat /user
        - the output will like the following → 2020-11-17 14:03:58

you can use formatting with **stat** command as following

Format	Meaning
%b	Size in bytes
%n	File name
%o	Block size
%r	Replication
%y	Modification date
%F	File type

- example using formatting
    - hdfs dfs -stat "%n %b bytes" /labb.txt
        - labb.txt 216 bytes
    - hdfs dfs -stat "%n %y" /labb.txt
        - labb.txt 2020-11-17 14:03:58

jps

- stands for java virtual machine process status tool
- it lists all java processes running on your machine, which is very useful in Hadoop because all Hadoop daemons are java processes
- the output like the following
    - 12345 NameNode
    - 12356 DataNode
    - 12367 SecondaryNameNode
    - 12378 ResourceManager
    - 12389 NodeManager
    - 12400 Jps

count

- Quicky see the number of files, directories, and total size in a path
- Example
    - hdfs dfs -count /user

appendToFile

- used for adding content to an existing file without overwriting it
- hdfs dfs -appendToFile /home/ubuntu/Desktop/new_data.txt /user/youssef/file.txt
    - contents of new_data.txt are added at the end of file.txt

cat

- displays the contents of the file stored on the hdfs, contents appears on the terminal.
- hdfs dfs -cat /user/lab.txt

head

- displays the first 1KB of a file stored in HDFS by default
- useful for previewing large files without flooding your terminal
- example
    - hdfs dfs -cat /labb.txt | head -n 2

tail

- show last 1KB of a file
- example
    - hdfs dfs -cat /labb.txt | tail

    - hdfs dfs -tail /labb.txt

dfsadmin -report

- show the HDFS cluster storage usage
- example
    - hdfs dfsadmin -report

df

- show the HDFS storage usage and free space
- example
    - hdfs dfs -df -h

setrep

- used to change the replication factor of a file/directory in HDFS. By default it is 3 for anything which is stored in HDFS
- example
    - hdfs dfs -setrep -R -w 6 /user/file.txt
        - -setrep: change the replication factor of the file/directory
        - -R → recursively apply to the directories and their contents
        - -w → wait for replication to complete before returning
        - 6 → the new replication factor
        - /user/file.txt → the target file or directory

Enable and disable the safe mode

- hdfs dfsadmin -safemode get
    - check the status of the safe mode
- hdfs dfsadmin -safemode get
    - enable the safe mode
- hdfs dfsadmin -safemode leave
    - disable the safe mode