

Big Data (HDFS) Assignment

Problem Statement: Using Command line of HDFS, perform following tasks.

a) Create a directory /hadoop/hdfs/ in HDFS

```
hdfs dfs -mkdir /hadoop
hdfs dfs -mkdir /hadoop/hdfs/
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -mkdir /hadoop
hadoop@hadoop-VirtualBox:~$ hdfs dfs -mkdir /hadoop/hdfs/
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /
Found 8 items
drwxr-xr-x  - hadoop supergroup      0 2024-09-03 17:47 /airlines
drwxr-xr-x  - hadoop supergroup      0 2024-08-31 16:41 /flights
drwxr-xr-x  - hadoop supergroup      0 2024-09-03 17:45 /flights1
drwxr-xr-x  - hadoop supergroup      0 2024-09-04 16:35 /hadoop
drwxr-xr-x  - hadoop supergroup      0 2022-11-21 15:25 /hbase
drwxrwxrwx  - hadoop supergroup      0 2022-11-21 15:12 /tmp
drwxr-xr-x  - hadoop supergroup      0 2024-09-03 16:50 /user
drwxr-xr-x  - hadoop supergroup      0 2024-08-31 14:18 /wordcount
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop/
Found 1 items
drwxr-xr-x  - hadoop supergroup      0 2024-09-04 16:35 /hadoop/hdfs
```

b) Create a temp directory in Hadoop. Run HDFS command to delete "temp" directory.

```
hdfs dfs -mkdir /temp
hdfs dfs -rm -r /temp
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -mkdir /temp
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /
Found 15 items
drwxr-xr-x  - hadoop supergroup      0 2024-09-03 17:47 /airlines
drwxr-xr-x  - hadoop supergroup      0 2024-08-31 16:41 /flights
drwxr-xr-x  - hadoop supergroup      0 2024-09-03 17:45 /flights1
drwxr-xr-x  - hadoop supergroup      0 2024-09-04 16:35 /hadoop
drwxr-xr-x  - hadoop supergroup      0 2022-11-21 15:25 /hbase
drwxr-xr-x  - hadoop supergroup      0 2024-09-04 17:12 /streaming
drwxr-xr-x  - hadoop supergroup      0 2024-09-04 17:14 /streaming2
drwxr-xr-x  - hadoop supergroup      0 2024-09-05 09:25 /temp
drwxrwxrwx  - hadoop supergroup      0 2022-11-21 15:12 /tmp
drwxr-xr-x  - hadoop supergroup      0 2024-09-03 16:50 /user
drwxr-xr-x  - hadoop supergroup      0 2024-08-31 14:18 /wordcount
-rw-r--r--  1 hadoop supergroup 1164038 2024-09-04 17:03 /wordcount2
-rw-r--r--  1 hadoop supergroup 1164038 2024-09-04 17:05 /wordcount3
-rw-r--r--  1 hadoop supergroup   29 2024-09-04 17:09 /wordcount4
drwxr-xr-x  - hadoop supergroup      0 2024-09-04 17:40 /wordcount_top
hadoop@hadoop-VirtualBox:~$ hdfs dfs -rm -r /temp
Deleted /temp
```

c) List all the files/directories for the given hdfs destination path.

```
hdfs dfs -ls <HDFS_PATH>
```

To list the contents of the /user/hadoop directory:

```
hdfs dfs -ls /user/hadoop
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /user/hadoop
Found 3 items
-rw-r--r--  1 hadoop supergroup 1164038 2024-08-31 12:27 /user/hadoop/Harry_Potter_and_the_Deathly_Hallows.txt
drwxrwxrwx  - hadoop supergroup      0 2022-11-21 15:11 /user/hadoop/warehouse
drwxr-xr-x  - hadoop supergroup      0 2024-09-04 17:01 /user/hadoop/wordcount2
```

d) Command that will list the directories in /hadoop folder.

```
hdfs dfs -ls /hadoop | grep '^d'
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop | grep '^d'
drwxr-xr-x  - hadoop supergroup      0 2024-09-04 16:35 /hadoop/hdfs
```

e) Command to list recursively all files in hadoop directory and all subdirectories in hadoop directory

```
hdfs dfs -ls -R /hadoop
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls -R /hadoop
drwxr-xr-x   - hadoop supergroup          0 2024-09-04 16:35 /hadoop/hdfs
```

f) List all the directory inside /hadoop/hdfs/ directory which starts with 'dir'.

```
hdfs dfs -ls /hadoop/hdfs/ | grep '^d.* dir'
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop/hdfs | grep '^d.* dir'
```

g) Create a temp.txt file. Copies this file from local file system to HDFS

```
echo "Hi, This is Vismaya" > temp.txt
```

```
hdfs dfs -put temp.txt /hadoop/hdfs/
```

```
hadoop@hadoop-VirtualBox:~$ echo "Hi, This is Vismaya" > temp.txt
hadoop@hadoop-VirtualBox:~$ hdfs dfs -put temp.txt /hadoop/hdfs/
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop/hdfs/
Found 1 items
-rw-r--r--   1 hadoop supergroup          20 2024-09-05 10:42 /hadoop/hdfs/temp.txt
```

h) Copies the file from HDFS to local file system.

```
hdfs dfs -get /hadoop/hdfs/temp.txt /home/hadoop/Downloads/
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -get /hadoop/hdfs/temp.txt /home/hadoop/Downloads/
hadoop@hadoop-VirtualBox:~$ cd /home/hadoop/Downloads/
hadoop@hadoop-VirtualBox:~/Downloads$ ls
airports1.csv      flights              hivexmlserde-1.0.5.3.jar  People.json
airports.csv       Hadoop_Installation_Doc.txt  HR_Employee.csv          raw_flight_data1.csv
bank_edited.json   Harry_Potter_and_the_Deathly_Hallows.txt  json-serde-1.3.8-jar-with-dependencies.jar  raw_flight_data.csv
Employee_Advance.csv  Hbase_Installation.txt      output                   temp.txt
```

i) Command to copy from local directory with the source being restricted to a local file reference.

```
hdfs dfs -put /home/hadoop/Downloads/Harry_Potter_and_the_Deathly_Hallows.txt /hadoop/hdfs
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -put /home/hadoop/Downloads/Harry_Potter_and_the_Deathly_Hallows.txt /hadoop/hdfs/
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop/hdfs/
Found 2 items
-rw-r--r--   1 hadoop supergroup    1164038 2024-09-05 11:15 /hadoop/hdfs/Harry_Potter_and_the_Deathly_Hallows.txt
-rw-r--r--   1 hadoop supergroup      20 2024-09-05 10:42 /hadoop/hdfs/temp.txt
```

j) Command to copy to local directory with the source being restricted to a local file reference.

```
hdfs dfs -get /hadoop/hdfs/Harry_Potter_and_the_Deathly_Hallows.txt
/home/hadoop/Desktop/
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -get /hadoop/hdfs/Harry_Potter_and_the_Deathly_Hallows.txt /home/hadoop/Desktop/
hadoop@hadoop-VirtualBox:~$ cd /home/hadoop/Desktop/
hadoop@hadoop-VirtualBox:~/Desktop$ ls
Harry_Potter_and_the_Deathly_Hallows.txt
```

k) Command to move from local directory source to Hadoop directory.

```
hdfs dfs -moveFromLocal /home/hadoop/Desktop/Harry_Potter_and_the
_Deathly_Hallows.txt /hadoop/
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -moveFromLocal /home/hadoop/Desktop/Harry_Potter_and_the_Deathly_Hallows.txt /hadoop/
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop/
Found 2 items
-rw-r--r-- 1 hadoop supergroup 1164038 2024-09-05 12:26 /hadoop/Harry_Potter_and_the_Deathly_Hallows.txt
drwxr-xr-x - hadoop supergroup 0 2024-09-05 11:15 /hadoop/hdfs
```

l) Deletes the directory and any content under it recursively.

```
hdfs dfs -rm -r /hadoop/hdfs
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -rm -r /hadoop/hdfs
Deleted /hadoop/hdfs
hadoop@hadoop-VirtualBox:~$
```

m) List the files and show Format file sizes in a human-readable fashion.

```
hdfs dfs -ls -h /hadoop/
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls -h /hadoop/
Found 1 items
-rw-r--r-- 1 hadoop supergroup 1.1 M 2024-09-05 12:26 /hadoop/Harry_Potter_and_the_Deathly_Hallows.txt
```

n) Take a source file and outputs the file in text format on the terminal.

```
hdfs dfs -cat /hadoop/sample.txt
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -cat /hadoop/sample.txt
Hi, I am Vismaya. How are you feeling today?
```

o) Display the content of the HDFS file test on your /user/hadoop2 directory.

```
hdfs dfs -cat /user/hadoop/test1.txt
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -cat /user/hadoop/test1.txt
Hi, How are you?
This is a sample test.
```

p) Append the content of a local file test1 to a hdfs file test2.

```
hdfs dfs -appendToFile test2.txt /user/hadoop/test1.txt
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -cat /user/hadoop/test1.txt
Hi, How are you?
This is a sample test.

hadoop@hadoop-VirtualBox:~$ hdfs dfs -appendToFile test2.txt /user/hadoop/test1.txt
hadoop@hadoop-VirtualBox:~$ hdfs dfs -cat /user/hadoop/test1.txt
Hi, How are you?
This is a sample test.
This is added text.
This is another sample text.
```

q) Show the capacity, free and used space of the filesystem

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -df
Filesystem                Size      Used    Available  Use%
hdfs://localhost:9000  42005135360  223338496  27894054912    1%
hadoop@hadoop-VirtualBox:~$
```

r) Shows the capacity, free and used space of the filesystem. Add parameter Formats the sizes of files in a human-readable fashion.

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -df -h
Filesystem                Size      Used    Available  Use%
hdfs://localhost:9000   39.1 G   213.0 M    26.0 G     1%
hadoop@hadoop-VirtualBox:~$
```

s) Show the amount of space, in bytes, used by the files that match the specified file pattern.

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -du -s /user/hadoop/*.txt
1164038 /user/hadoop/Harry_Potter_and_the_Deathly_Hallows.txt
93      /user/hadoop/test1.txt
51      /user/hadoop/test2.txt
hadoop@hadoop-VirtualBox:~$
```

t) Show the amount of space, in bytes, used by the files that match the specified file pattern. Formats the sizes of files in a human-readable fashion.

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -du -s -h /user/hadoop/*.txt
1.1 M /user/hadoop/Harry_Potter_and_the_Deathly_Hallows.txt
93 /user/hadoop/test1.txt
51 /user/hadoop/test2.txt
hadoop@hadoop-VirtualBox:~$
```

u) Check the health of the Hadoop file system.

[illegible]

v) Command to turn off the safemode of Name Node.

```
hdfs dfsadmin -safemode leave
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfsadmin -safemode leave
Safe mode is OFF
```

w) HDFS command to format NameNode.

```
hdfs namenode -format
```

x) Create a file named `hdfstest.txt` and change it number of replications to 3.


```
hdfs dfs -put hdfsatest.txt /user/hadoop/
hdfs dfs -setrep 3 /user/hadoop/hdfsatest.txt
```

```
hadoop@hadoop-VirtualBox:~$ sudo nano hdfsatest.txt
[sudo] password for hadoop:
hadoop@hadoop-VirtualBox:~$ hdfs dfs -put hdfsatest.txt /user/hadoop/
hadoop@hadoop-VirtualBox:~$ hdfs dfs -setrep 3 /user/hadoop/hdfsatest.txt
Replication 3 set: /user/hadoop/hdfsatest.txt
hadoop@hadoop-VirtualBox:~$
```

y) Write command to display number of replicas for hdfsatest.txt file.

```
hdfs dfs -stat %r /user/hadoop/hdfsatest.txt
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -stat %r /user/hadoop/hdfsatest.txt
3
```

z) Write command to Display the status of file "hdfsatest.txt" like block size, filesize in bytes.

```
hdfs dfs -stat "%b %o" /user/hadoop/hdfsatest.txt
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -stat "%b %o" /user/hadoop/hdfsatest.txt
12 134217728
```

aa) Write HDFS command to change file permission from rw – r – r to rwx-rw-x for hdfsatest.txt.

```
hdfs dfs -chmod 761 /user/hadoop/hdfsatest.txt
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -chmod 761 /user/hadoop/hdfsatest.txt
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /user/hadoop/
Found 6 items
-rw-r--r-- 3 hadoop supergroup 1164038 2024-08-31 12:27 /user/hadoop/Harry_Potter_and_the_Deathly_Hallows.txt
-rwxrw--x 3 hadoop supergroup 12 2024-09-06 17:29 /user/hadoop/hdfsatest.txt
-rw-r--r-- 3 hadoop supergroup 93 2024-09-06 15:54 /user/hadoop/test1.txt
-rw-r--r-- 3 hadoop supergroup 51 2024-09-06 15:50 /user/hadoop/test2.txt
drwxrwxrwx - hadoop supergroup 0 2022-11-21 15:11 /user/hadoop/warehouse
drwxr-xr-x - hadoop supergroup 0 2024-09-04 17:01 /user/hadoop/wordcount2
```