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 -p /hadoop/hdfs
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -mkdir -p /hadoop/hdfs
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /
Found 9 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 17:48 /airlines
drwxr-xr-x   - hadoop supergroup          0 2024-08-31 16:36 /flights
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 17:47 /flights1
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 09:12 /hadoop
drwxr-xr-x   - hadoop supergroup          0 2022-11-21 15:25 /hbase
drwxr-xr-x   - hadoop supergroup          0 2024-09-02 16:50 /hdfs
drwxrwxrwx   - hadoop supergroup          0 2022-11-21 15:12 /tmp
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 16:52 /user
drwxr-xr-x   - hadoop supergroup          0 2024-09-04 17:16 /wordcount
```

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

```
hdfs dfs -mkdir /hadoop/temp

hdfs dfs -rmdir /hadoop/temp
```

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

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

```
hdfs dfs -ls /
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -rmdir /temp
```

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

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop
Found 1 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 09:12 /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 2024-09-05 09:26 /hadoop/hdfs
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 09:26 /hadoop/hdfs/dir1
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 09:26 /hadoop/hdfs/dir2
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 09:26 /hadoop/hdfs/tmp
```

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

- approach one:

```
hdfs dfs -ls -d /hadoop/hdfs/dir*
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls -d /hadoop/hdfs/dir*
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 09:26 /hadoop/hdfs/dir1
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 09:26 /hadoop/hdfs/dir2
```

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

```
sudo nano test.txt
hdfs dfs -put /home/hadoop/test.txt /hadoop/hdfs/
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -put /home/hadoop/test.txt /hadoop/hdfs/
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop/hdfs/
Found 4 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 09:26 /hadoop/hdfs/dir1
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 09:26 /hadoop/hdfs/dir2
-rw-r--r--   1 hadoop supergroup         23 2024-09-05 09:38 /hadoop/hdfs/test.txt
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 09:26 /hadoop/hdfs/tmp
```

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

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

```
hadoop@hadoop-VirtualBox:~$ cd Downloads
hadoop@hadoop-VirtualBox:~/Downloads$ hdfs dfs -get /hadoop/hdfs/test.txt
```

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

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -put ./file.txt /hadoop/hdfs/
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop/hdfs/
Found 5 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 09:26 /hadoop/hdfs/dir1
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 09:26 /hadoop/hdfs/dir2
-rw-r--r--   1 hadoop supergroup         12 2024-09-05 10:15 /hadoop/hdfs/file.txt
-rw-r--r--   1 hadoop supergroup         23 2024-09-05 09:38 /hadoop/hdfs/test.txt
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 09:26 /hadoop/hdfs/tmp
```

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

```
hdfs dfs -get ./file.txt .
```

```
hadoop@hadoop-VirtualBox:~$ cd Downloads
hadoop@hadoop-VirtualBox:~/Downloads$ hdfs dfs -get /hadoop/hdfs/file.txt
hadoop@hadoop-VirtualBox:~/Downloads$ rm file.txt
hadoop@hadoop-VirtualBox:~/Downloads$ hdfs dfs -get /hadoop/hdfs/file.txt ./
```

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

```
hdfs dfs -moveFromLocal ./file.txt /hadoop/hdfs/
```

```
hadoop@hadoop-VirtualBox:~/Downloads$ hdfs dfs -moveFromLocal ./file.txt /hadoop/hdfs/
```

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

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

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

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

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

```
hadoop@hadoop-VirtualBox:~/Downloads$ hdfs dfs -ls -h /wordcount
Found 2 items
-rw-r--r--   1 hadoop supergroup      1.1 M 2024-09-04 17:16 /wordcount/HarryPotterandtheDeathlyHallows.txt
drwxr-xr-x   - hadoop supergroup          0 2024-08-31 14:09 /wordcount/output
```

Activate Windows

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

```
cat test.txt
```

```
hadoop@hadoop-VirtualBox:~$ cat test.txt
this is a mnormal text
```

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

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -cat /hadoop/test.txt
this is a mnormal text
```

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

```
hdfs dfs -appendToFile /home/hadoop/file.txt /hadoop/test.txt
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -appendToFile /home/hadoop/file.txt /hadoop/test.txt
hadoop@hadoop-VirtualBox:~$ hdfs dfs -cat /hadoop/test.txt
this is a mnormal text
it is from a file
```

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

```
hdfs dfs -df
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -df
Filesystem                  Size        Used      Available  Use%
hdfs://localhost:9000  41954803712  217067520  28412596224    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.

```
hdfs dfs -df -h
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -df -h
Filesystem              Size      Used   Available  Use%
hdfs://localhost:9000  39.1 G   207.0 M     26.5 G    1%
```

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

```
hdfs dfs -du -s /hadoop/hdfs/*txt
```

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -du -s /hadoop/*txt
41  /hadoop/test.txt
```

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.

```
hdfs dfs -du -h /hadoop/hdfs/*txt
```

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -du -h /hadoop/*txt
41  /hadoop/test.txt
```

u) Check the health of the Hadoop file system.

```
hdfs fsck /
```

```
FSCK started by hadoop (auth:SIMPLE) from /127.0.0.1 for path / at Thu Sep 05 11:31:26 IST 2024
.................................................................................................
.................................................................................................
.................................................................................................
.................................................................................................
.................................................................................................
.................................................................................................
.................................................................................................
.................................................................................................
.................................................................................................
.................................................................................................
.................................................................................................
.................................................................................................
.................................................................................................
.................................................................................................
.................................................................................................
........Status: HEALTHY
Total size:    207666314 B
Total dirs:    82
Total files:   1308
Total symlinks:             0
Total blocks (validated):   1301 (avg. block size 159620 B)
Minimally replicated blocks:   1301 (100.0 %)
Over-replicated blocks:     0 (0.0 %)
Under-replicated blocks:    0 (0.0 %)
Mis-replicated blocks:      0 (0.0 %)
Default replication factor:    1
Average block replication:     1.0
Corrupt blocks:             0
Missing replicas:           0 (0.0 %)
Number of data-nodes:          1
Number of racks:               1
FSCK ended at Thu Sep 05 11:31:26 IST 2024 in 69 milliseconds


The filesystem under path '/' is HEALTHY
hadoop@hadoop-VirtualBox:~/assignment$
```

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

```
    hdfs dfsadmin -safemode leave
```

```
hadoop@hadoop-VirtualBox:~/assignment$ 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 -setrep 3 /hadoop/hdfstest.txt
```

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -setrep 3 /hadoop/hdfstest.txt
Replication 3 set: /hadoop/hdfstest.txt
```

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

```
    hdfs dfs -stat %r /hadoop/hdfstest.txt
```

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

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

```
    hdfs dfs -stat /hadoop/hdfstest.txt
```

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -stat /hadoop/hdfstest.txt
2024-09-05 06:21:29
```

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

```
    hdfs dfs -chmod 761 /hadoop/hdfstest.txt
```

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -chmod 761 /hadoop/hdfstest.txt
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -ls /hadoop/
Found 3 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 10:52 /hadoop/hdfs
-rwxrw---x   3 hadoop supergroup          0 2024-09-05 11:51 /hadoop/hdfstest.txt
-rw-r--r--   1 hadoop supergroup         41 2024-09-05 10:56 /hadoop/test.txt
```