

## Big Data (HDFS) Assignment

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

```
Hdfs dfs -mkdir -p /hadoop/hdfs
```

The screenshot shows the Hadoop web interface at `localhost:50070/explorer.html#/hadoop`. A terminal window is overlaid on the interface, showing the command `hdfs dfs -mkdir -p /hadoop/hdfs` being executed. The terminal output shows the command being run twice. The web interface displays the 'Browse Directory' page for `/hadoop`. The table below shows the directory listing:

| Permission | Owner  | Group      | Size | Last Modified        | Replication | Block Size | Name                 |
|------------|--------|------------|------|----------------------|-------------|------------|----------------------|
| drwxr-xr-x | hadoop | supergroup | 0 B  | 3/7/2024, 6:42:25 am | 0           | 0 B        | <a href="#">hdfs</a> |

Hadoop, 2018.

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

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

The screenshot shows the Hadoop web interface at `localhost:50070/explorer.html#/hadoop`. A terminal window is overlaid on the interface, showing the command `hdfs dfs -mkdir -p /hadoop/hdfs` being executed, followed by `hdfs dfs -mkdir /hadoop/temp`. The web interface displays the 'Browse Directory' page for `/hadoop`. The table below shows the directory listing:

| Permission | Owner  | Group      | Size | Last Modified        | Replication | Block Size | Name                 |
|------------|--------|------------|------|----------------------|-------------|------------|----------------------|
| drwxr-xr-x | hadoop | supergroup | 0 B  | 3/7/2024, 6:42:25 am | 0           | 0 B        | <a href="#">hdfs</a> |
| drwxr-xr-x | hadoop | supergroup | 0 B  | 3/7/2024, 6:51:45 am | 0           | 0 B        | <a href="#">temp</a> |

Hadoop, 2018.

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

The screenshot shows the Hadoop web interface at `localhost:50070/explorer.html#/hadoop`. A terminal window is overlaid on the interface, showing the command `hdfs dfs -rm -r /hadoop/temp` being executed. The web interface displays the 'Browse Directory' page for `/hadoop`. The table below shows the directory listing:

| Permission | Owner  | Group      | Size | Last Modified        | Replication | Block Size | Name                 |
|------------|--------|------------|------|----------------------|-------------|------------|----------------------|
| drwxr-xr-x | hadoop | supergroup | 0 B  | 3/7/2024, 6:42:25 am | 0           | 0 B        | <a href="#">hdfs</a> |

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

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop/
Found 1 items
drwxr-xr-x  - hadoop supergroup          0 2024-07-03 06:42 /hadoop/hdfs
hadoop@hadoop-VirtualBox:~$
```

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

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls -d /hadoop/*
drwxr-xr-x  - hadoop supergroup          0 2024-07-03 06:42 /hadoop/hdfs
hadoop@hadoop-VirtualBox:~$
```

- 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 2024-07-03 06:42 /hadoop/hdfs
```

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

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -mkdir -p /hadoop/hdfs/directory
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls -d /hadoop/hdfs/dir*
drwxr-xr-x  - hadoop supergroup          0 2024-07-03 08:02 /hadoop/hdfs/directory
hadoop@hadoop-VirtualBox:~$
```

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

```
nano temp.txt
```

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

| Permission | Owner  | Group      | Size | hadoop@hadoop-VirtualBox:~\$ nano temp.txt | hadoop@hadoop-VirtualBox:~\$ hdfs dfs -put temp.txt /hadoop/hdfs/ | hadoop@hadoop-VirtualBox:~\$ |
|------------|--------|------------|------|--|---|------------------------------|
| drwxr-xr-x | hadoop | supergroup | 0 B  |  |   |                              |
| -rw-r--r-- | hadoop | supergroup | 15 B | 3/7/2024, 8:09:49 am                       | 1   | 128 MB temp.txt              |

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

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -get /hadoop/hdfs/temp.txt
get: `temp.txt': File exists
hadoop@hadoop-VirtualBox:~$ hdfs dfs -get /hadoop/hdfs/temp.txt
```

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

```
hdfs dfs -copyFromLocal temp.txt /hadoop/hdfs/directory
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -copyFromLocal temp.txt /hadoop/hdfs/
copyFromLocal: '/hadoop/hdfs/temp.txt': File exists
hadoop@hadoop-VirtualBox:~$ hdfs dfs -copyFromLocal temp.txt /hadoop/hdfs/directory
hadoop@hadoop-VirtualBox:~$
```

| Permission | Owner  | Group      | Size | Last Modified        | Replication | Block Size | Name                     |
|------------|--------|------------|------|----------------------|-------------|------------|--------------------------|
| -rw-r--r-- | hadoop | supergroup | 15 B | 3/7/2024, 8:22:23 am | 1           | 128 MB     | <a href="#">temp.txt</a> |

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

```
hdfs dfs -copyToLocal /hadoop/hdfs/temp.txt /directory/temp.txt
```

```
hadoop-VirtualBox:~$ sudo chmod 755 /directory
hadoop-VirtualBox:~$ hdfs dfs -copyToLocal /hadoop/hdfs/directory/temp.txt /directory/temp.txt
hadoop-VirtualBox:~$
```

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

```
hdfs dfs -moveFromLocal /directory/temp.txt /hadoop/
```

```
/hadoop/
hadoop@hadoop-VirtualBox:~$ hdfs dfs -moveFromLocal /directory/temp.txt /hadoop/
hadoop@hadoop-VirtualBox:~$
```

| Permission | Owner  | Group      | Size | Last Modified        | Replication | Block Size | Name                     |
|------------|--------|------------|------|----------------------|-------------|------------|--------------------------|
| drwxr-xr-x | hadoop | supergroup | 0 B  | 3/7/2024, 8:09:49 am | 0           | 0 B        | <a href="#">hdfs</a>     |
| -rw-r--r-- | hadoop | supergroup | 15 B | 3/7/2024, 9:05:14 am | 1           | 128 MB     | <a href="#">temp.txt</a> |

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

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

The image shows a terminal window with a dark background. The prompt is 'hadoop@hadoop-VirtualBox:~\$'. The command entered is 'hdfs dfs -rm -r /hadoop/hdfs/'. The output shows 'Deleted /hadoop/hdfs' followed by a new prompt. Below the terminal, a light red error message box is visible, stating 'Path does not exist on HDFS or WebHDFS is disabled. Please check your path or enable WebHDFS'. At the bottom, a web browser address bar shows the path '/hadoop/hdfs/directory' and a 'Go!' button.

- 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 15 2024-07-03 09:05 /hadoop/temp.txt
hadoop@hadoop-VirtualBox:~$
```

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

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -cat /hadoop/temp.txt
Sample content
hadoop@hadoop-VirtualBox:~$
```

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

```
hdfs dfs -text /user/hadoop2/test.txt
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -put test.txt /user/hadoop2/
hadoop@hadoop-VirtualBox:~$ hdfs dfs -text /user/hadoop2/test.txt
sample test file 'test'
--content--
hadoop@hadoop-VirtualBox:~$
```

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

```
hdfs dfs -appendToFile test1.txt /user/hadoop2/test2.txt
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -put test2.txt /user/hadoop2/
hadoop@hadoop-VirtualBox:~$ hdfs dfs -appendToFile test1.txt /user/hadoop2/test2.txt
hadoop@hadoop-VirtualBox:~$ hdfs dfs -text /user/hadoop2/test2.txt
test2 file
content

test1 file
content
hadoop@hadoop-VirtualBox:~$
```

- 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 42005135360 73232384 28900110336    0%
hadoop@hadoop-VirtualBox:~$ hdfs dfs -df -h /
```

- 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    69.8 M    26.9 G     0%
```

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

```
hdfs dfs -du /hadoop/
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -du /hadoop/
15 /hadoop/temp.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.

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -du -h /hadoop/
15 /hadoop/temp.txt
hadoop@hadoop-VirtualBox:~$ hdfs dfs -du /hadoop/
```

- u) Check the health of the Hadoop file system.

```
hdfs fsck /
```

```
is /hadoop/temp.txt
hadoop@hadoop-VirtualBox:~$ hdfs fsck /
Connecting to namenode via http://localhost:50070/fsck?ugi=hadoop&path=%2F
FSCK started by hadoop (auth:SIMPLE) from /127.0.0.1 for path / at Wed Jul 03 10:18:10 IST 2024
.....Status: HEALTHY
Total size:      72506040 B
Total dirs:      37
Total files:     17
Total symlinks:   0
Total blocks (validated): 15 (avg. block size 4833736 B)
Minimally replicated blocks: 15 (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 Wed Jul 03 10:18:10 IST 2024 in 11 milliseconds
```

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

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

```
hdfs dfsadmin -safemode leave
```

```
The filesystem under path '/' is HEALTHY
hadoop@hadoop-VirtualBox:~$ hdfs dfsadmin -safemode leave
Safe mode is OFF
```

w) HDFS command to format NameNode.

```
hdfs namenode -format
```

```
Format aborted in Storage Directory /usr/local/hadoop/yarn_data/hdfs/namenode
hadoop@hadoop-VirtualBox:~$ hdfs namenode -format
Formatting using clusterid: CID-68bed81b-a4d5-4c13-a605-372a0385d182
Re-format filesystem in Storage Directory /usr/local/hadoop/yarn_data/hdfs/namenode ? (Y or N) Y
hadoop@hadoop-VirtualBox:~$
```

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

```
nano hdfstest.txt
```

```
hdfs dfs -setrep 3 hdfstest.txt
```

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

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

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

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

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

```
hdfs dfs -stat '%b %o' /hdfstest.txt
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -stat '%b %o' /hdfstest.txt
23 134217728
hadoop@hadoop-VirtualBox:~$
```

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

```
hdfs dfs -chmod 765 /hdfstest.txt
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -chmod 765 /hdfstest.txt
```