Assignment 5

A. Create a directory /hadoop/hdfs/ in HDFS

- `hdfs dfs -mkdir /hadoop/hdfs`

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -mkdir /hdfs
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /
Found 10 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 18:04 /airlines
drwxr-xr-x   - hadoop supergroup          0 2024-08-31 16:54 /flights
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 18:01 /flights1
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 23:40 /hadoop
drwxr-xr-x   - hadoop supergroup          0 2022-11-21 15:25 /hbase
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 23:45 /hdfs
drwxr-xr-x   - hadoop supergroup          0 2024-09-04 17:54 /testStreaming
drwxrwxrwx   - hadoop supergroup          0 2022-11-21 15:12 /tmp
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 17:07 /user
drwxr-xr-x   - hadoop supergroup          0 2024-08-31 12:42 /wordcount
hadoop@hadoop-VirtualBox:~$
```

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

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -mkdir /temp
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /
Found 11 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 18:04 /airlines
drwxr-xr-x   - hadoop supergroup          0 2024-08-31 16:54 /flights
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 18:01 /flights1
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 23:40 /hadoop
drwxr-xr-x   - hadoop supergroup          0 2022-11-21 15:25 /hbase
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 23:45 /hdfs
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 23:46 /temp
drwxr-xr-x   - hadoop supergroup          0 2024-09-04 17:54 /testStreaming
drwxrwxrwx   - hadoop supergroup          0 2022-11-21 15:12 /tmp
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 17:07 /user
drwxr-xr-x   - hadoop supergroup          0 2024-08-31 12:42 /wordcount
hadoop@hadoop-VirtualBox:~$ hdfs dfs -rm -r /temp
Deleted /temp
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /
Found 10 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 18:04 /airlines
drwxr-xr-x   - hadoop supergroup          0 2024-08-31 16:54 /flights
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 18:01 /flights1
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 23:40 /hadoop
drwxr-xr-x   - hadoop supergroup          0 2022-11-21 15:25 /hbase
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 23:45 /hdfs
drwxr-xr-x   - hadoop supergroup          0 2024-09-04 17:54 /testStreaming
drwxrwxrwx   - hadoop supergroup          0 2022-11-21 15:12 /tmp
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 17:07 /user
drwxr-xr-x   - hadoop supergroup          0 2024-08-31 12:42 /wordcount
hadoop@hadoop-VirtualBox:~$ █
```

C. List all the files/directories for the given HDFS destination path.

- `hdfs dfs -ls /`

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /
Found 10 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 18:04 /airlines
drwxr-xr-x   - hadoop supergroup          0 2024-08-31 16:54 /flights
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 18:01 /flights1
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 23:40 /hadoop
drwxr-xr-x   - hadoop supergroup          0 2022-11-21 15:25 /hbase
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 23:45 /hdfs
drwxr-xr-x   - hadoop supergroup          0 2024-09-04 17:54 /testStreaming
drwxrwxrwx   - hadoop supergroup          0 2022-11-21 15:12 /tmp
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 17:07 /user
drwxr-xr-x   - hadoop supergroup          0 2024-08-31 12:42 /wordcount
hadoop@hadoop-VirtualBox:~$ █
```

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 23:40 /hadoop/issac
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 2024-09-06 00:02 /hadoop/hdfs
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 00:02 /hadoop/hdfs/dir1
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 00:02 /hadoop/hdfs/dir2
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 23:40 /hadoop/issac
hadoop@hadoop-VirtualBox:~$
```

F. List all the directories inside /hadoop/hdfs/ directory which start with 'dir'.

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls -d /hadoop/hdfs/dir*
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 00:02 /hadoop/hdfs/dir1
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 00:02 /hadoop/hdfs/dir2
hadoop@hadoop-VirtualBox:~$
```

G. Create a temp.txt file. Copy this file from the local file system to HDFS.

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

```
hadoop@hadoop-VirtualBox:~$ gedit tmp.txt
hadoop@hadoop-VirtualBox:~$ hdfs dfs -put /home/hadoop/tmp.txt /hadoop/hdfs/
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop/hdfs
Found 3 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 00:02 /hadoop/hdfs/dir1
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 00:02 /hadoop/hdfs/dir2
-rw-r--r--   1 hadoop supergroup         25 2024-09-06 00:06 /hadoop/hdfs/tmp.txt
hadoop@hadoop-VirtualBox:~$
```

H. Copy the file from HDFS to the local file system.

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -get /hadoop/hdfs/tmp.txt/ .
hadoop@hadoop-VirtualBox:~$ ls
Assignment4.ipynb  Desktop      Downloads              install-tl-20240830    metastore_db  Pictures  Templates  Videos
derby.log          Documents    examples.desktop  install-tl-unx.tar.gz  Music         Public    tmp.txt
hadoop@hadoop-VirtualBox:~$
```

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

- `hdfs dfs -put /temp1.txt /hadoop/hdfs`

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -put ./temp1.txt /hadoop/hdfs/
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls
Found 3 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-05 23:44 hdfs
drwxrwxrwx   - hadoop supergroup          0 2022-11-21 15:11 warehouse
drwxr-xr-x   - hadoop supergroup          0 2024-08-31 14:32 wordcount
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hdfs
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop/hdfs
Found 4 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 00:02 /hadoop/hdfs/dir1
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 00:02 /hadoop/hdfs/dir2
-rw-r--r--   1 hadoop supergroup         17 2024-09-06 20:58 /hadoop/hdfs/temp1.txt
-rw-r--r--   1 hadoop supergroup         25 2024-09-06 00:06 /hadoop/hdfs/tmp.txt
```

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

- `hdfs dfs -get /hdoop/hdfs/temp.txt ./home`

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -get /hadoop/hdfs/tmp.txt ./home
hadoop@hadoop-VirtualBox:~$ ls
Assignment4.ipynb  Downloads              install-tl-unx.tar.gz  Public      Videos
derby.log          examples.desktop       metastore_db           temp1.txt
Desktop            home                   Music                  Templates
Documents          install-tl-20240830    Pictures               tmp.txt
hadoop@hadoop-VirtualBox:~$
```

K. Command to move from the local directory source to Hadoop directory.

- `hdfs dfs -moveFromLocal /temp2.txt /hadoop/hdfs`

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -moveFromLocal ./temp2.txt /hadoop/hdfs/
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop/hdfs
Found 5 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 00:02 /hadoop/hdfs/dir1
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 00:02 /hadoop/hdfs/dir2
-rw-r--r--   1 hadoop supergroup         17 2024-09-06 20:58 /hadoop/hdfs/temp1.txt
-rw-r--r--   1 hadoop supergroup         17 2024-09-06 21:06 /hadoop/hdfs/temp2.txt
-rw-r--r--   1 hadoop supergroup         25 2024-09-06 00:06 /hadoop/hdfs/tmp.txt
hadoop@hadoop-VirtualBox:~$
```

L. Delete 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 -du -h /wordcount`

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls -h /wordcount
Found 2 items
-rw-r--r--   1 hadoop supergroup    429.4 K 2024-08-31 12:36 /wordcount/HarryPotter_Sorcerers_Stone.txt
-rw-r--r--   1 hadoop supergroup      1.1 M 2024-08-31 12:42 /wordcount/HarryPotterandtheDeathlyHallows.txt
hadoop@hadoop-VirtualBox:~$
```

N. Take a source file and output the file in text format on the terminal.

- hdfs dfs -cat tmp.txt

```
hadoop@hadoop-VirtualBox:~$ cat tmp.txt
Hi, This is a dummy file
hadoop@hadoop-VirtualBox:~$
```

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

- hdfs dfs -cat /wordcount/testHP.txt

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -put ./testHP.txt /wordcount
hadoop@hadoop-VirtualBox:~$ hdfs dfs -cat /wordcount/testHP.txt
Mr. and Mrs. Dursley, of number four, Privet Drive, were proud to say
that they were perfectly normal, thank you very much. They were the last
people you'd expect to be involved in anything strange or mysterious,
because they just didn't hold with such nonsense.
hadoop@hadoop-VirtualBox:~$
```

P. Append the content of a local file test1 to an HDFS file test2.

- hdfs dfs -appendToFile /tmp.txt /wordcount/testHP.txt

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -appendToFile ./tmp.txt /wordcount/testHP.txt
hadoop@hadoop-VirtualBox:~$ hdfs dfs -cat /wordcount/testHP.txt
Mr. and Mrs. Dursley, of number four, Privet Drive, were proud to say
that they were perfectly normal, thank you very much. They were the last
people you'd expect to be involved in anything strange or mysterious,
because they just didn't hold with such nonsense.
Hi, This is a dummy file
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  41954803712  303247659  24193789952   1%
hadoop@hadoop-VirtualBox:~$
```

R. Show 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   289.2 M    22.5 G    1%
hadoop@hadoop-VirtualBox:~$
```

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

- `hdfs dfs -du -s /wordcount/testHP.txt`

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -du -s /wordcount/HarryPotterandtheDeathlyHallows.txt
1164038  /wordcount/HarryPotterandtheDeathlyHallows.txt
hadoop@hadoop-VirtualBox:~$
```

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

- `hdfs dfs -du -h /wordcount/testHP.txt`

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -du -h /wordcount/HarryPotterandtheDeathlyHallows.txt
1.1 M  /wordcount/HarryPotterandtheDeathlyHallows.txt
hadoop@hadoop-VirtualBox:~$
```

U. Check the health of the Hadoop file system.

- `hdfs fsck /`

```
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 Fri Sep 06 21:21:20 IST 2024
...........................................................................................
...........................................................................................
...........................................................................................
...........................................................................................
...........................................................................................
...........................................................................................
...........................................................................................
...........................................................................................
...........................................................................................
...........................................................................................
...........................................................................................
...........................................................................................
..........................................Status: HEALTHY
 Total size:    293445089 B
 Total dirs:    65
 Total files:   1247
 Total symlinks:            0
 Total blocks (validated):    1242 (avg. block size 236268 B)
 Minimally replicated blocks:   1242 (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 Fri Sep 06 21:21:20 IST 2024 in 135 milliseconds


The filesystem under path '/' is HEALTHY
```

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
hadoop@hadoop-VirtualBox:~$
```

W. HDFS command to format NameNode.

- hdfs namenode -format

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

- hdfs dfs -setrep 3 /wordcount/testHP.txt

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

Y. Write command to display the number of replicas for hdfstest.txt file.

- hdfs dfs -stat %r /wordcount/testHP.txt

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

Z. Write command to display the status of file "hdfstest.txt" like block size, file size in bytes.

- hdfs dfs -stat /wordcount/testHP.txt

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -stat /wordcount/testHP.txt
2024-09-06 15:47:28
hadoop@hadoop-VirtualBox:~$
```

AA. Write HDFS command to change file permission from rw-r--r-- to rwxrwxr-x for hdfstest.txt.

- hdfs dfs -chmod 761 /wordcount/testHP.txt

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -chmod 761 /wordcount/testHP.txt
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /wordcount
Found 3 items
-rw-r--r--   1 hadoop supergroup     439742 2024-08-31 12:36 /wordcount/HarryPotter_Sorcerers_Stone.txt
-rw-r--r--   1 hadoop supergroup    1164038 2024-08-31 12:42 /wordcount/HarryPotterandtheDeathlyHallows.txt
-rwxrw---x   3 hadoop supergroup        288 2024-09-06 21:17 /wordcount/testHP.txt
hadoop@hadoop-VirtualBox:~$
```