Assignment 5

🗓 Date	@September 7, 2024
⊙ Select	hdfs
⁑ Status	Done

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

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

Output

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

```
hdfs dfs -mkdir /temp
hdfs dfs -rmdir /temp
```

Output

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -mkdir /temp
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /temp
hadoop@hadoop-VirtualBox:~$ hdfs dfs -rmdir /temp
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /temp
ls: `/temp': No such file or directory
hadoop@hadoop-VirtualBox:~$
```

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

```
hdfs dfs -ls /
```

Output

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /
Found 12 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 09:49 /HR_Orc
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 09:56 /HR_partition
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 17:46 /airlines
drwxr-xr-x   - hadoop supergroup          0 2024-09-02 09:18 /flights
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 17:44 /flights1
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 12:14 /hadoop
drwxr-xr-x   - hadoop supergroup          0 2022-11-21 15:25 /hbase
drwxr-xr-x   - hadoop supergroup          0 2024-09-04 17:35 /test
-rw-r--r--   1 hadoop supergroup         55 2024-09-04 16:54 /test_streaming
-rw-r--r--   1 hadoop supergroup         55 2024-09-04 16:42 /textfile
drwxrwxrwx   - hadoop supergroup          0 2022-11-21 15:12 /tmp
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 16:50 /user
hadoop@hadoop-VirtualBox:~$
```

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

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

Output

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop
Found 1 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 12:14 /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
```

Output

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls -R /hadoop
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 12:37 /hadoop/hdfs
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 12:37 /hadoop/hdfs/new
-rw-r--r--   1 hadoop supergroup          0 2024-09-06 12:37 /hadoop/hdfs/sample1.txt
-rw-r--r--   1 hadoop supergroup          0 2024-09-06 12:37 /hadoop/hdfs/sample2.txt
hadoop@hadoop-VirtualBox:~$
```

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

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop/hdfs/ | grep "^d" | grep "^.* /hadoop/hdfs/dir"
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 21:34 /hadoop/hdfs/directory
hadoop@hadoop-VirtualBox:~$
```

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

```
touch temp.txt
hdfs dfs -copyFromLocal ~/temp.txt /
```

```
hadoop@hadoop-VirtualBox:~$ touch temp.txt
hadoop@hadoop-VirtualBox:~$ hdfs dfs -copyFromLocal ~/temp.txt /
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /
Found 13 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 09:49 /HR_Orc
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 09:56 /HR_partition
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 17:46 /airlines
drwxr-xr-x   - hadoop supergroup          0 2024-09-02 09:18 /flights
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 17:44 /flights1
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 12:14 /hadoop
drwxr-xr-x   - hadoop supergroup          0 2022-11-21 15:25 /hbase
-rw-r--r--   1 hadoop supergroup          0 2024-09-06 12:46 /temp.txt
drwxr-xr-x   - hadoop supergroup          0 2024-09-04 17:35 /test
-rw-r--r--   1 hadoop supergroup         55 2024-09-04 16:54 /test_streaming
-rw-r--r--   1 hadoop supergroup         55 2024-09-04 16:42 /textfile
drwxrwxrwx   - hadoop supergroup          0 2022-11-21 15:12 /tmp
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 16:50 /user
hadoop@hadoop-VirtualBox:~$
```

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

```
hdfs dfs -copyToLocal /temp.txt ~/
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -copyToLocal /temp.txt ~/
hadoop@hadoop-VirtualBox:~$ ls
derby.log    Downloads                              MapReduceAssignment   Pictures            sample_text2.txt   sample_text.txt   Videos
Desktop      examples.desktop                       metastore_db          Public              sample_text3.txt   Templates
Documents    google-chrome-stable_current_amd64.deb Music                 'Pyspark RDD.ipynb'  sample_text4.txt   temp.txt
hadoop@hadoop-VirtualBox:~$
```

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

```
hdfs dfs -put test1.txt /hadoop/hdfs/test1.txt
```

Output

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -mkdir -p /hadoop/hdfs/
hadoop@hadoop-VirtualBox:~$ hdfs dfs -put test1.txt /hadoop/hdfs/test1.txt
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop/hdfs
Found 1 items
-rw-r--r--   1 hadoop supergroup         35 2024-09-06 21:19 /hadoop/hdfs/test1.txt
hadoop@hadoop-VirtualBox:~$
```

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

```
hdfs dfs -get /hadoop/hdfs/test1.txt ~/test_to_local.txt
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -get /hadoop/hdfs/test1.txt ~/test_to_local.txt
hadoop@hadoop-VirtualBox:~$ ls
```

```
hadoop@hadoop-VirtualBox:~$ ls -l
total 108356
-rw-r--r-- 1 hadoop hadoop        627 Mar 15  2023  derby.log
drwxr-xr-x 2 hadoop hadoop       4096 Nov 17  2022  Desktop
drwxr-xr-x 2 hadoop hadoop       4096 Nov 17  2022  Documents
drwxr-xr-x 9 hadoop hadoop       4096 Sep  6 15:23  Downloads
-rw-r--r-- 1 hadoop hadoop       8980 Nov 17  2022  examples.desktop
-rw-r--r-- 1 hadoop hadoop  110831940 Aug 28 11:42  google-chrome-stable_current_amd64.deb
-rw-rw-r-- 1 hadoop hadoop          0 Sep  6 18:56  hdfstest.txt
drwxr-xr-x 2 hadoop hadoop       4096 Sep  6 09:21  MapReduceAssignment
drwxr-xr-x 5 hadoop hadoop       4096 Mar 15  2023  metastore_db
drwxr-xr-x 2 hadoop hadoop       4096 Nov 17  2022  Music
drwxr-xr-x 3 hadoop hadoop      12288 Sep  5 17:09  Pictures
drwxr-xr-x 2 hadoop hadoop       4096 Nov 17  2022  Public
-rw-r--r-- 1 hadoop hadoop      21590 Aug 30 17:49 'Pyspark RDD.ipynb'
-rw-r--r-- 1 root   root         31 Sep  4 16:59  sample_text2.txt
-rw-r--r-- 1 root   root         48 Sep  4 17:18  sample_text3.txt
-rw-r--r-- 1 hadoop hadoop       48 Sep  4 17:35  sample_text4.txt
-rw-r--r-- 1 root   root         55 Sep  4 16:40  sample_text.txt
drwxr-xr-x 2 hadoop hadoop       4096 Nov 17  2022  Templates
-rw-r--r-- 1 root   root         35 Sep  6 18:10  test1.txt
-rw-r--r-- 1 root   root         40 Sep  6 18:09  test2.txt
-rw-r--r-- 1 hadoop hadoop       35 Sep  6 21:23  test_to_local.txt
-rw-r--r-- 1 root   root        199 Sep  6 17:40  test.txt
drwxr-xr-x 2 hadoop hadoop       4096 Nov 17  2022  Videos
hadoop@hadoop-VirtualBox:~$
```

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

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -moveFromLocal ~/temp.txt /hadoop/hdfs
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /hadoop/hdfs
Found 4 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 12:37 /hadoop/hdfs/new
-rw-r--r--   1 hadoop supergroup          0 2024-09-06 12:37 /hadoop/hdfs/sample1.txt
-rw-r--r--   1 hadoop supergroup          0 2024-09-06 12:37 /hadoop/hdfs/sample2.txt
-rw-r--r--   1 hadoop supergroup          0 2024-09-06 15:30 /hadoop/hdfs/temp.txt
```

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-VirtualBox:~$ hdfs dfs -ls -h /
Found 13 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 09:49 /HR_Orc
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 09:56 /HR_partition
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 17:46 /airlines
drwxr-xr-x   - hadoop supergroup          0 2024-09-02 09:18 /flights
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 17:44 /flights1
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 15:33 /hadoop
drwxr-xr-x   - hadoop supergroup          0 2022-11-21 15:25 /hbase
-rw-r--r--   1 hadoop supergroup          0 2024-09-06 12:46 /temp.txt
drwxr-xr-x   - hadoop supergroup          0 2024-09-04 17:35 /test
-rw-r--r--   1 hadoop supergroup         55 2024-09-04 16:54 /test_streaming
-rw-r--r--   1 hadoop supergroup         55 2024-09-04 16:42 /textfile
drwxrwxrwx   - hadoop supergroup          0 2022-11-21 15:12 /tmp
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 16:50 /user
```

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

```
cat test.txt
```

```
hadoop@hadoop-VirtualBox:~$ sudo nano test.txt
[sudo] password for hadoop:
hadoop@hadoop-VirtualBox:~$ cat test.txt
India, officially the Republic of India (ISO: Bhārat Gaṇarājya),
[21] is a country in South Asia. It is the seventh-largest country by area;
the most populous country with effect from June 2023
hadoop@hadoop-VirtualBox:~$
```

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

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -cat /user/hadoop2/
India, officially the Republic of India (ISO: Bhārat Gaṇarājya),
[21] is a country in South Asia. It is the seventh-largest country by area;
the most populous country with effect from June 2023
hadoop@hadoop-VirtualBox:~$
```

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

```
hdfs dfs -appendToFile ~/test1.txt /test2.txt
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -appendToFile ~/test1.txt /test2.txt
hadoop@hadoop-VirtualBox:~$ hdfs dfs -cat /test2.txt
This is the native text file from test2
This text is from test1 from local
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   214269952   27547213824    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   204.3 M      25.7 G    1%
```

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

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -du /*txt
0    /hdfstest.txt
0    /temp.txt
199  /test.txt
75   /test2.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 /*txt
```

```
                    /LESLE.LAL
hadoop@hadoop-VirtualBox:~$ hdfs dfs -du -h /*txt
0   /temp.txt
199  /test.txt
75  /test2.txt
```

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 18:38:20 IST 2024
................................................................................................
................................................................................................
.....
........Status: HEALTHY
Total size:    204513266 B
Total dirs:    82
Total files:   1408
Total symlinks:                0
Total blocks (validated):      1399 (avg. block size 146185 B)
Minimally replicated blocks:   1399 (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 18:38:20 IST 2024 in 171 milliseconds
```

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

```
hdfs dfsadmin -safemode leave
```

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

w) HDFS command to format NameNode.

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

```
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 "%o %b" /hdfstest.txt
```

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -stat "%o %b" /hdfstest.txt
134217728 40
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 761 /hdfstest.txt
```

```
2024-09-06 15:20:22
hadoop@hadoop-VirtualBox:~$ hdfs dfs -chmod 761 /hdfstest.txt
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /
```

```
Usage: hadoop fs [generic options] -ls [-d] [-h] [-R] [<path> ...]
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /
Found 16 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 09:49 /HR_Orc
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 09:56 /HR_partition
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 17:46 /airlines
drwxr-xr-x   - hadoop supergroup          0 2024-09-02 09:18 /flights
drwxr-xr-x   - hadoop supergroup          0 2024-09-03 17:44 /flights1
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 15:33 /hadoop
drwxr-xr-x   - hadoop supergroup          0 2022-11-21 15:25 /hbase
-rwxrw---x   3 hadoop supergroup          0 2024-09-06 18:56 /hdfstest.txt
-rw-r--r--   1 hadoop supergroup          0 2024-09-06 12:46 /temp.txt
drwxr-xr-x   - hadoop supergroup          0 2024-09-04 17:35 /test
-rw-r--r--   1 hadoop supergroup        199 2024-09-06 18:04 /test.txt
-rw-r--r--   1 hadoop supergroup         75 2024-09-06 18:11 /test2.txt
-rw-r--r--   1 hadoop supergroup         55 2024-09-04 16:54 /test_streaming
-rw-r--r--   1 hadoop supergroup         55 2024-09-04 16:42 /textfile
drwxrwxrwx   - hadoop supergroup          0 2022-11-21 15:12 /tmp
drwxr-xr-x   - hadoop supergroup          0 2024-09-06 18:06 /user
hadoop@hadoop-VirtualBox:~$
```