Big Data (HDFS) Assignment 1

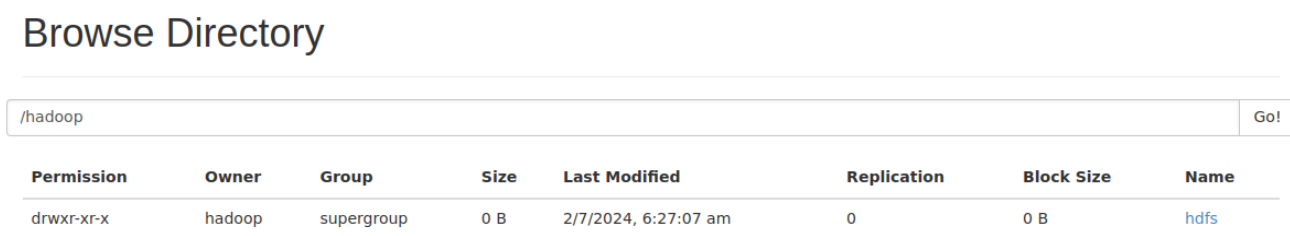
Affan Mohammed N Marikar

281911

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

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

   hdfs dfs -mkdir -p /*hadoop/*hdfs

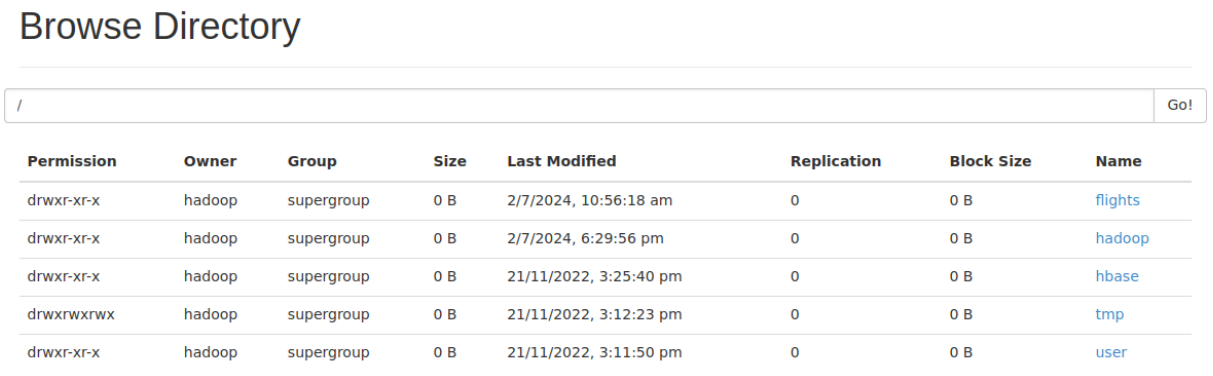


2. Create a temp directory in Hadoop. Run HDFS command to delete "temp" directory.

   hdfs dfs -mkdir -p /*hadoop/*temp

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	hadoop	supergroup	0 B	2/7/2024, 10:56:18 am	0	0 B	flights
drwxr-xr-x	hadoop	supergroup	0 B	2/7/2024, 6:28:24 pm	0	0 B	hadoop
drwxr-xr-x	hadoop	supergroup	0 B	21/11/2022, 3:25:40 pm	0	0 B	hbase
drwxr-xr-x	hadoop	supergroup	0 B	2/7/2024, 6:38:59 am	0	0 B	temp
drwxrwxrwx	hadoop	supergroup	0 B	21/11/2022, 3:12:23 pm	0	0 B	tmp
drwxr-xr-x	hadoop	supergroup	0 B	21/11/2022, 3:11:50 pm	0	0 B	user

   hdfs dfs -rm -r /temp



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

hdfs dfs -ls /

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls /
Found 5 items
drwxr-xr-x   - hadoop supergroup          0 2024-07-02 10:56 /flights
drwxr-xr-x   - hadoop supergroup          0 2024-07-02 18:29 /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 2022-11-21 15:11 /user
hadoop@hadoop-VirtualBox:~$
```

4. 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 2024-07-02 06:27 /hadoop/hdfs
hadoop@hadoop-VirtualBox:~$
```

5. 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 2024-07-02 06:27 /hadoop/hdfs
```

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

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

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -ls -d /hadoop/hdfs/dir*
ls: `/hadoop/hdfs/dir*': No such file or directory
```

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

   sudo nano temp.txt

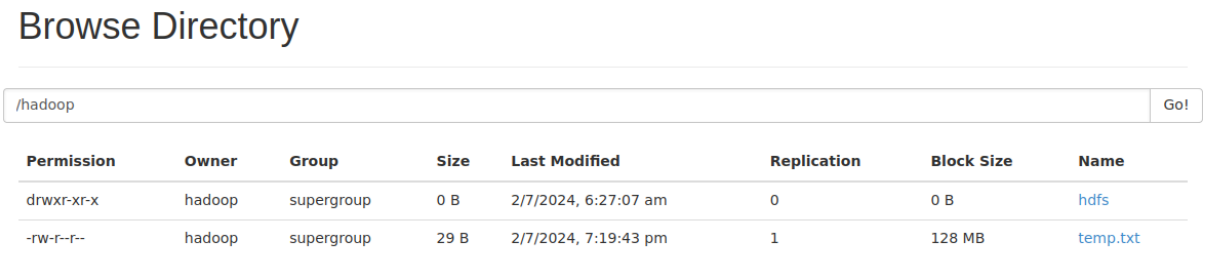
```
#hi there
#welcome to hadoop
```

hdfs dfs -put temp.txt /

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	hadoop	supergroup	0 B	2/7/2024, 10:56:18 am	0	0 B	flights
drwxr-xr-x	hadoop	supergroup	0 B	2/7/2024, 6:29:56 pm	0	0 B	hadoop
drwxr-xr-x	hadoop	supergroup	0 B	21/11/2022, 3:25:40 pm	0	0 B	hbase
-rw-r--r--	hadoop	supergroup	29 B	2/7/2024, 6:53:40 pm	1	128 MB	temp.txt
drwxrwxrwx	hadoop	supergroup	0 B	21/11/2022, 3:12:23 pm	0	0 B	tmp
drwxr-xr-x	hadoop	supergroup	0 B	21/11/2022, 3:11:50 pm	0	0 B	user

8. Copies the file from HDFS to local file system.

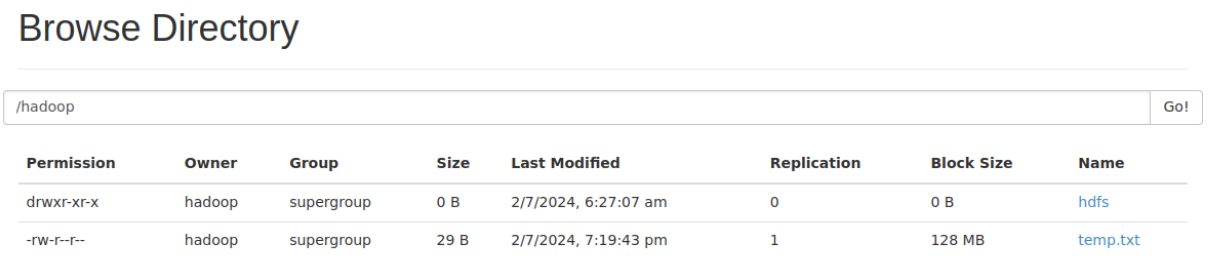
   hdfs dfs -get /temp.txt /newFolder



Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	hadoop	supergroup	0 B	2/7/2024, 6:27:07 am	0	0 B	hdfs
-rw-r--r--	hadoop	supergroup	29 B	2/7/2024, 7:19:43 pm	1	128 MB	temp.txt

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

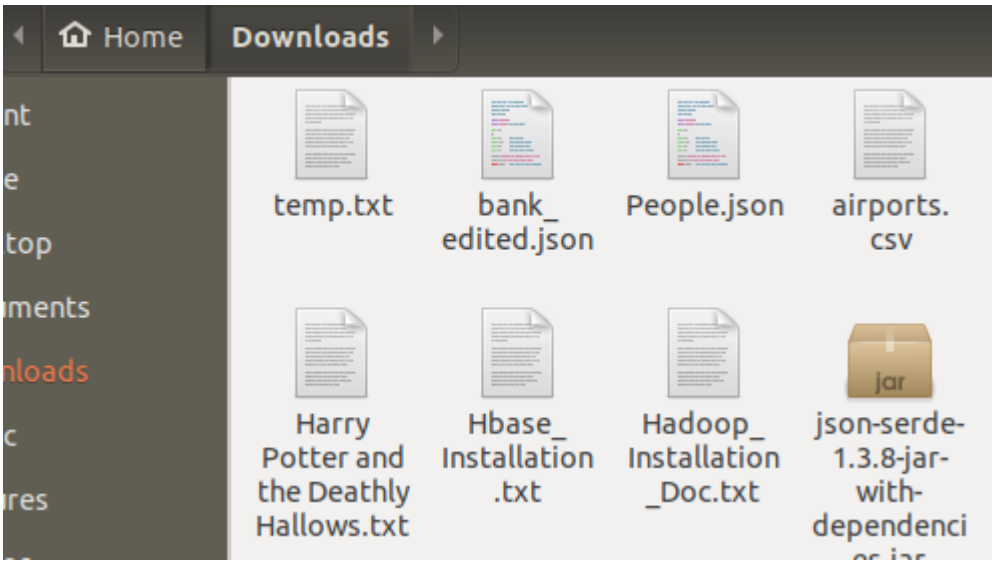
   hdfs dfs -copyFromLocal temp.txt /hadoop



Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	hadoop	supergroup	0 B	2/7/2024, 6:27:07 am	0	0 B	hdfs
-rw-r--r--	hadoop	supergroup	29 B	2/7/2024, 7:19:43 pm	1	128 MB	temp.txt

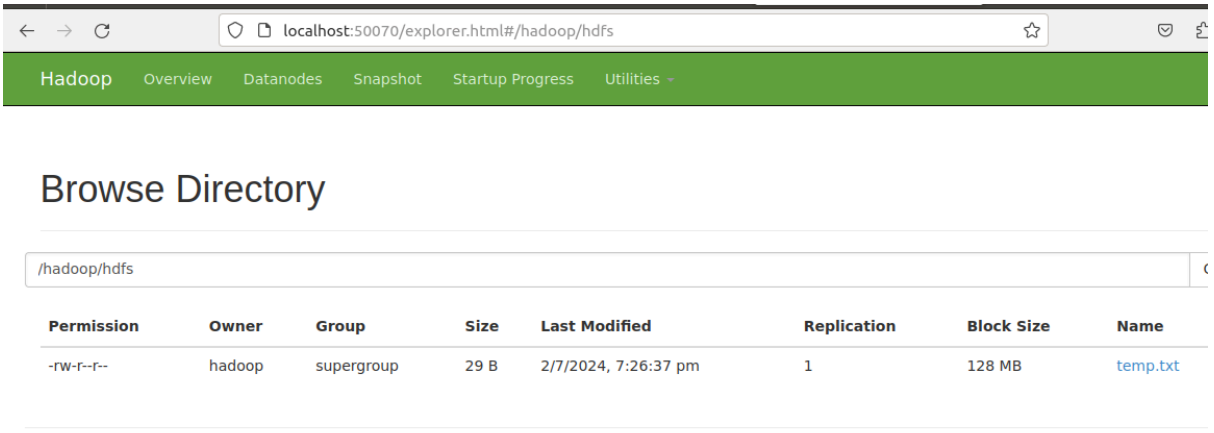
10. Command to copies to local directory with the source being restricted to a local file reference.

hdfs dfs -copyToLocal /hadoop/temp.txt /home/hadoop/Downloads

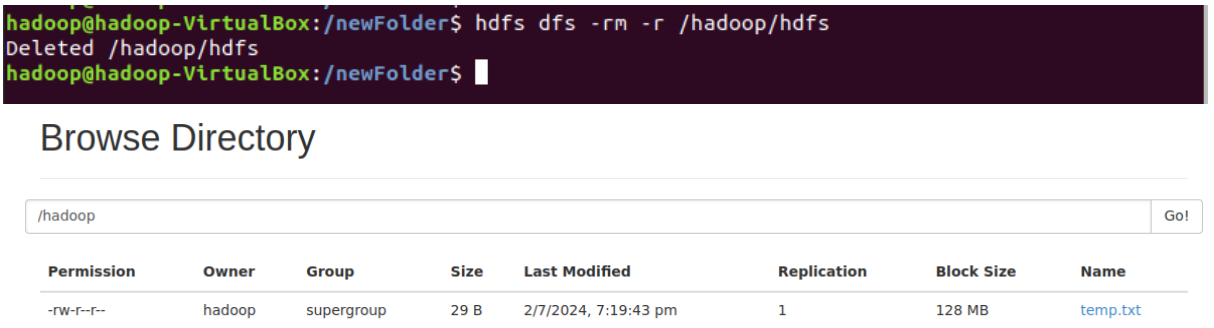


11. Command to move from local directory source to Hadoop directory.

hdfs dfs -moveFromLocal temp.txt /hadoop/hdfs



12. Deletes the directory and any content under it recursively.
hdfs dfs -rm -r /hadoop/hdfs



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

hdfs dfs -ls -h /hadoop

```
hadoop@hadoop-VirtualBox:/newFolder$ hdfs dfs -ls -h /hadoop
Found 1 items
-rw-r--r--   1 hadoop supergroup         29 2024-07-02 19:19 /hadoop/temp.txt
hadoop@hadoop-VirtualBox:/newFolder$
```

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

hdfs dfs -text /hadoop/temp.txt

```
hadoop@hadoop-VirtualBox:/newFolder$ hdfs dfs -text /hadoop/temp.txt
#hi there
#welcome to hadoop
hadoop@hadoop-VirtualBox:/newFolder$
```

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

hdfs dfs -cat /usr/hadoop2/test.txt

```
hadoop@hadoop-VirtualBox:/$ hdfs dfs -put test.txt /usr/hadoop2
hadoop@hadoop-VirtualBox:/$ hdfs dfs -cat  /usr/hadoop2/test.txt
#hi
#this is test 2
hadoop@hadoop-VirtualBox:/$
```

16. Append the content of a local file test1 to a hdfs file test2.
hdfs dfs -appendToFile test2.txt /hadoop/temp.txt

```
hadoop@hadoop-VirtualBox:/$ hdfs dfs -appendToFile test2.txt /hadoop/temp.txt
hadoop@hadoop-VirtualBox:/$ hdfs dfs -cat /hadoop/temp.txt
#hi there
#welcome to hadoop
#hi this is new test file.
# hello world
hadoop@hadoop-VirtualBox:/$
```

17. Show the capacity, free and used space of the filesystem

hdfs dfs -df /

hdfs dfs -du /

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -df /
Filesystem                     Size      Used    Available  Use%
hdfs://localhost:9000  41954803712  73232384  28850573312    0%
hadoop@hadoop-VirtualBox:~$ hdfs dfs -du /
72491854  /flights
71        /hadoop
14094     /hbase
29        /temp.txt
0         /tmp
0         /user
21        /usr
hadoop@hadoop-VirtualBox:~$
```

18. 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 /

    hdfs dfs -du -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%
hadoop@hadoop-VirtualBox:~$ hdfs dfs -du -h /
69.1 M  /flights
71      /hadoop
13.8 K  /hbase
29      /temp.txt
0       /tmp
0       /user
21      /usr
hadoop@hadoop-VirtualBox:~$
```

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

    hdfs dfs -du -s /usr/hadoop2/test*

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -du -s /usr/hadoop2/test*
21  /usr/hadoop2/test.txt
hadoop@hadoop-VirtualBox:~$
```

20. 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 -s -h /usr/hadoop2/test*

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -du -s -h /usr/hadoop2/test*
21  /usr/hadoop2/test.txt
hadoop@hadoop-VirtualBox:~$
```

21. Check the health of the Hadoop file system.

hdfs dfsadmin -report

```
Configured Capacity: 41954803712 (39.07 GB)
Present Capacity: 28923265024 (26.94 GB)
DFS Remaining: 28850032640 (26.87 GB)
DFS Used: 73232384 (69.84 MB)
DFS Used%: 0.25%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0
Missing blocks (with replication factor 1): 0

-------------------------------------------------
Live datanodes (1):

Name: 127.0.0.1:50010 (localhost)
Hostname: hadoop-VirtualBox
Decommission Status : Normal
Configured Capacity: 41954803712 (39.07 GB)
DFS Used: 73232384 (69.84 MB)
Non DFS Used: 10867384320 (10.12 GB)
DFS Remaining: 28850032640 (26.87 GB)
DFS Used%: 0.17%
DFS Remaining%: 68.76%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 1
Last contact: Tue Jul 02 22:17:45 IST 2024
```

22. 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:~$
```

23. HDFS command to format NameNode.

hdfs namenode -format

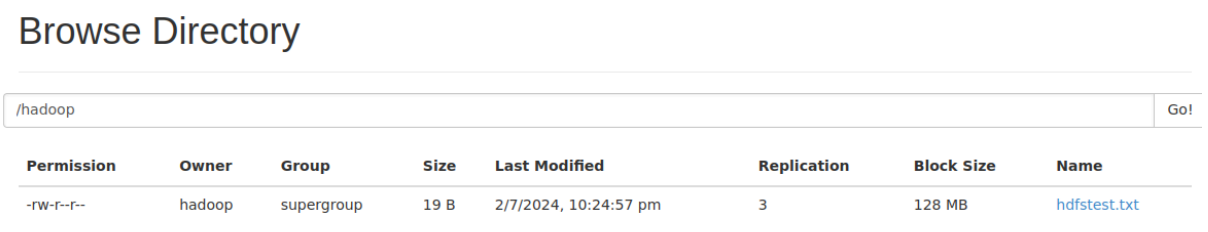
```
hadoop@hadoop-VirtualBox:~$ hdfs namenode -format
Formatting using clusterid: CID-847f2efc-2bae-4b1e-a015-0ea48ab76447
Re-format filesystem in Storage Directory /usr/local/hadoop/yarn_data/hdfs/namenode ? (Y or N
) n
Format aborted in Storage Directory /usr/local/hadoop/yarn_data/hdfs/namenode
hadoop@hadoop-VirtualBox:~$
```

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

    sudo nano hdfstest.txt

    hdfs dfs -put hdfstest.txt /hadoop

    hdfs dfs -setrep 3 /hadoop/hdfstest.txt



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

    hdfs fsck /hadoop/hdfstest.txt -files -blocks -locations

```
hadoop@hadoop-VirtualBox:~$ hdfs fsck /hadoop/hdfstest.txt -files -blocks -locations
Connecting to namenode via http://localhost:50070/fsck?ugi=hadoop&files=1&blocks=1&locations=1&path=%2Fhadoop%2Fhdfstest.txt
FSCK started by hadoop (auth:SIMPLE) from /127.0.0.1 for path /hadoop/hdfstest.txt at Tue Jul 02 22:29:28 IST 2024
/hadoop/hdfstest.txt 19 bytes, 1 block(s):  Under replicated BP-1998725987-127.0.1.1-1668676850545:blk_1073741843_1021. Target Replic
as is 3 but found 1 replica(s).
0. BP-1998725987-127.0.1.1-1668676850545:blk_1073741843_1021 len=19 repl=1 [DatanodeInfoWithStorage[127.0.0.1:50010,DS-3194ee54-44ec-
4aa6-a64f-a72b31691456,DISK]]

Status: HEALTHY
 Total size:    19 B
 Total dirs:    0
 Total files:   1
 Total symlinks:                0
 Total blocks (validated):      1 (avg. block size 19 B)
 Minimally replicated blocks:   1 (100.0 %)
 Over-replicated blocks:        0 (0.0 %)
 Under-replicated blocks:       1 (100.0 %)
 Mis-replicated blocks:         0 (0.0 %)
 Default replication factor:    1
 Average block replication:     1.0
 Corrupt blocks:                0
 Missing replicas:              2 (66.666664 %)
 Number of data-nodes:          1
 Number of racks:               1
FSCK ended at Tue Jul 02 22:29:28 IST 2024 in 12 milliseconds


The filesystem under path '/hadoop/hdfstest.txt' is HEALTHY
```


26. Write command to Display the status of file “hdfstest.txt” like block size, filesize in bytes.

    hdfs dfs -stat "%o %b %s" /hadoop//hdfstest.txt

```
hadoop@hadoop-VirtualBox:~$ hdfs dfs -stat "%o %b %s" /hadoop//hdfstest.txt
134217728 19 s
```

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

hdfs dfs -chmod 765 /hadoop/hdfstest.txt