

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:~/assignment$ hdfs dfs -mkdir -p /hadoop/hdfs/
24/07/02 18:18:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@hadoop-VirtualBox:~/assignment$
```

/hadoop								Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
drwxr-xr-x	hadoop	supergroup	0 B	2/7/2024, 5:18:49 pm	0	0 B	hdfs	

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

hdfs dfs -mkdir /hadoop/temp/

hdfs dfs -rm -r /hadoop/temp/

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -mkdir /hadoop/temp/
24/07/02 18:22:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@hadoop-VirtualBox:~/assignment$
```

Browse Directory

/hadoop								Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
drwxr-xr-x	hadoop	supergroup	0 B	2/7/2024, 5:18:49 pm	0	0 B	hdfs	
drwxr-xr-x	hadoop	supergroup	0 B	2/7/2024, 5:22:13 pm	0	0 B	temp	

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -rm -r /hadoop/temp/
24/07/02 18:23:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
24/07/02 18:23:31 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Emptier interval = 0 minutes.
Deleted /hadoop/temp
hadoop@hadoop-VirtualBox:~/assignment$
```

/hadoop								Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
drwxr-xr-x	hadoop	supergroup	0 B	2/7/2024, 5:18:49 pm	0	0 B	hdfs	

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

hdfs dfs -ls /hadoop/

hdfs dfs -ls /hadoop/hdfs

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -ls /hadoop/
24/07/02 18:25:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x   - hadoop supergroup          0 2024-07-02 18:18 /hadoop/hdfs
hadoop@hadoop-VirtualBox:~/assignment$
```

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -ls /hadoop/hdfs
24/07/02 18:26:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@hadoop-VirtualBox:~/assignment$
```

/hadoop							Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	hadoop	supergroup	0 B	2/7/2024, 5:18:49 pm	0	0 B	hdfs

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

hdfs dfs -ls /hadoop/

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -ls /hadoop/
24/07/02 18:25:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x   - hadoop supergroup          0 2024-07-02 18:18 /hadoop/hdfs
hadoop@hadoop-VirtualBox:~/assignment$
```

/hadoop							Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	hadoop	supergroup	0 B	2/7/2024, 5:18:49 pm	0	0 B	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:~/assignment$ hdfs dfs -mkdir /hadoop/hdfs/direct
24/07/02 18:54:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -mkdir /hadoop/hdfs/work
24/07/02 18:54:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -ls -R /hadoop/
24/07/02 18:54:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
drwxr-xr-x   - hadoop supergroup          0 2024-07-02 18:54 /hadoop/hdfs
drwxr-xr-x   - hadoop supergroup          0 2024-07-02 18:54 /hadoop/hdfs/direct
drwxr-xr-x   - hadoop supergroup          0 2024-07-02 18:54 /hadoop/hdfs/work
hadoop@hadoop-VirtualBox:~/assignment$
```


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

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

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -ls -d /hadoop/hdfs/dir*
24/07/02 18:56:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
drwxr-xr-x   - hadoop supergroup          0 2024-07-02 18:54 /hadoop/hdfs/direct
hadoop@hadoop-VirtualBox:~/assignment$
```

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

sudo nano temp.txt

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

```
hadoop@hadoop-VirtualBox:~/assignment$ sudo nano temp.txt
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -put temp.txt /hadoop/hdfs
24/07/02 19:01:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@hadoop-VirtualBox:~/assignment$
```

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

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

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -get /hadoop/hdfs/temp.txt /home/hadoop/Downloads/
24/07/02 19:05:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@hadoop-VirtualBox:~/assignment$
```

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

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

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -copyFromLocal /home/hadoop/Downloads/temp.txt /hadoop/hdfs/
24/07/02 19:14:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
copyFromLocal: `/hadoop/hdfs/temp.txt': File exists
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -copyFromLocal /home/hadoop/Downloads/temp.txt /hadoop/hdfs/work
24/07/02 19:14:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@hadoop-VirtualBox:~/assignment$
```

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

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

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -copyToLocal /hadoop/hdfs/temp.txt /home/hadoop/Downloads/
24/07/02 19:40:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
copyToLocal: `/home/hadoop/Downloads/temp.txt': File exists
hadoop@hadoop-VirtualBox:~/assignment$
```

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

```
hdfs dfs -moveFromLocal /home/hadoop/Downloads/temp.txt /hadoop/hdfs/
```

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -moveFromLocal /home/hadoop/Downloads/temp.txt /hadoop/hdfs/
24/07/02 19:43:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
moveFromLocal: `/hadoop/hdfs/temp.txt': File exists
hadoop@hadoop-VirtualBox:~/assignment$
```

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

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

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -rm -r /hadoop/hdfs
24/07/02 19:47:53 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
24/07/02 19:47:54 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval = 0 minutes, Empty interval = 0 minutes.
Deleted /hadoop/hdfs
hadoop@hadoop-VirtualBox:~/assignment$
```

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

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

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -ls -h /hadoop
24/07/02 19:50:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@hadoop-VirtualBox:~/assignment$
```

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

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

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -text /hadoop/temp.txt
24/07/02 19:54:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Hello world.
hadoop@hadoop-VirtualBox:~/assignment$
```

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

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

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -mkdir -p /user/hadoop2
24/07/02 19:58:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -put temp.txt /user/hadoop2/
24/07/02 19:58:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -cat /user/hadoop2/temp.txt
24/07/02 19:59:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Hello world.
hadoop@hadoop-VirtualBox:~/assignment$
```


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

```
hdfs dfs -appendToFile test.txt /user/hadoop2/temp.txt
```

```
hadoop@hadoop-VirtualBox:~/assignment$ sudo nano test.txt
[sudo] password for hadoop:
hadoop@hadoop-VirtualBox:~/assignment$
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -appendToFile test.txt /user/had
oop2/temp.txt
24/07/02 20:01:58 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
hadoop@hadoop-VirtualBox:~/assignment$
```

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

```
hdfs dfs -df /
```

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -df /
24/07/02 20:07:32 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Filesystem                Size      Used Available Use%
hdfs://localhost:9000    21001486336  233552   11576410112    0%
hadoop@hadoop-VirtualBox:~/assignment$
```

- 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:~/assignment$ hdfs dfs -df -h /
24/07/02 20:08:29 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Filesystem                Size      Used Available Use%
hdfs://localhost:9000    19.6 G    236 K    10.8 G    0%
hadoop@hadoop-VirtualBox:~/assignment$
```

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

```
hdfs dfs -du -s /user/hadoop2/temp*
```

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -du -s /user/hadoop2/temp*
24/07/02 20:10:27 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
45 /user/hadoop2/temp.txt
hadoop@hadoop-VirtualBox:~/assignment$
```

- 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 -s -h /user/hadoop2/temp*
```

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -du -s -h /user/hadoop2/temp*
24/07/02 20:11:29 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
45 /user/hadoop2/temp.txt
```

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

`hdfs dfsadmin -report`

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfsadmin -report
24/07/02 20:16:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Configured Capacity: 21001486336 (19.56 GB)
Present Capacity: 11576651776 (10.78 GB)
DFS Remaining: 11576410112 (10.78 GB)
DFS Used: 241664 (236 KB)
DFS Used%: 0.00%
Under replicated blocks: 1
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: 21001486336 (19.56 GB)
DFS Used: 241664 (236 KB)
Non DFS Used: 8334422016 (7.76 GB)
DFS Remaining: 11576410112 (10.78 GB)
DFS Used%: 0.00%
DFS Remaining%: 55.12%
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 20:15:59 EAT 2024

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
24/07/02 20:18:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Safe mode is OFF
hadoop@hadoop-VirtualBox:~/assignment$
```

- w) HDFS command to format NameNode.

`hdfs namenode -format`

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs namenode -format
24/07/02 20:18:58 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = hadoop-VirtualBox/127.0.1.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 2.7.6
STARTUP_MSG:   classpath = /usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share/
hadoop/common/lib/commons-beanutils-1.7.0.jar:/usr/local/hadoop/share/hadoop/com
mon/lib/jets3t-0.9.0.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-beanu
tils-core-1.8.0.jar:/usr/local/hadoop/share/hadoop/common/lib/jackson-xc-1.9.13.
jar:/usr/local/hadoop/share/hadoop/common/lib/jackson-core-asl-1.9.13.jar:/usr/l
```


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

```
sudo nano hdfs-test.txt
```

```
hdfs dfs -put hdfs-test.txt /
```

```
hdfs dfs -setrep 3 /hdfs-test.txt
```

```
hadoop@hadoop-VirtualBox:~/assignment$ sudo nano hdfs-test.txt
[sudo] password for hadoop:
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -put hdfs-test.txt /
24/07/02 20:21:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -setrep 3 /hdfs-test.txt
24/07/02 20:22:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Replication 3 set: /hdfs-test.txt
hadoop@hadoop-VirtualBox:~/assignment$
```

- y) Write command to display number of replicas for hdfs-test.txt file.

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

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -stat %r /hdfs-test.txt
24/07/02 20:27:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
3
hadoop@hadoop-VirtualBox:~/assignment$
```

- z) Write command to Display the status of file “hdfs-test.txt” like block size, filesize in bytes.

```
hdfs dfs -stat "%b %o %r %n %y %u %g" /hdfs-test.txt
```

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -stat "%b %o %r %n %y %u %g" /hdfs-test.txt
24/07/02 20:28:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
8 134217728 3 hdfs-test.txt 2024-07-02 17:21:16 hadoop supergroup
hadoop@hadoop-VirtualBox:~/assignment$
```

- aa) Write HDFS command to change file permission from

```
rw - r - r to rwx-rw-x for hdfs-test.txt.
```

```
hdfs dfs -chmod 777 hdfs-test.txt
```

```
hadoop@hadoop-VirtualBox:~/assignment$ hdfs dfs -chmod 777 /hdfs-test.txt
24/07/02 20:38:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@hadoop-VirtualBox:~/assignment$
```

-rwxrwxrwx	hadoop	supergroup	8 B	2/7/2024, 7:21:16 pm	3	128 MB	hdfs-test.txt
------------	--------	------------	-----	----------------------	---	--------	-------------------------------