AIM : To implement HDFS Commands

IMPLEMENTATION:

1. To display the version of hadoop:

```
[cloudera@quickstart ~]$ hadoop version
Hadoop 2.6.0-cdh5.12.0
Subversion http://github.com/cloudera/hadoop -r dba647c5a8bc5e09b572d76a8d29481c78d1a0dd
Compiled by jenkins on 2017-06-29T11:32Z
Compiled with protoc 2.5.0
From source with checksum 7c45ae7a4592ce5af86bc4598c5b4
This command was run using /usr/lib/hadoop/hadoop-common-2.6.0-cdh5.12.0.jar
```

2. List the contents of the root directory in HDFS:

```
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 7 items
drwxrwxrwx   - hdfs     supergroup          0 2017-07-19 05:34 /benchmarks
drwxr-xr-x   - hbase    supergroup          0 2020-11-21 04:24 /hbase
drwxr-xr-x   - solr     solr                0 2017-07-19 05:37 /solr
drwxr-xr-x   - cloudera supergroup          0 2020-11-21 04:34 /system
drwxrwxrwt   - hdfs     supergroup          0 2020-11-21 04:25 /tmp
drwxr-xr-x   - hdfs     supergroup          0 2017-07-19 05:36 /user
drwxr-xr-x   - hdfs     supergroup          0 2017-07-19 05:36 /var
[cloudera@quickstart ~]$
```

3. Report the amount of space used and available on currently mounted file system:

hdfs dfs -df hdfs:/-

```
[cloudera@quickstart ~]$ hdfs dfs -df
Filesystem                              Size       Used    Available  Use%
hdfs://quickstart.cloudera:8020  58531520512  869388288  45895706133    1%
```

4. Count the number of directories, files and bytes under the paths that match the specified file pattern:

```
[cloudera@quickstart ~]$ hdfs dfs -count hdfs://quickstart.cloudera:8020
           1           0                  0 hdfs://quickstart.cloudera:8020/user/cloudera
```

5. Run a cluster balancing utility:

```
[cloudera@quickstart ~]$ hadoop balancer
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

20/11/21 04:34:35 INFO balancer.Balancer: namenodes  = [hdfs://0.0.0.0:8022]
20/11/21 04:34:35 INFO balancer.Balancer: parameters = Balancer.Parameters [BalancingPolicy.Node, threshold = 10.0, max idle iteration = 5, #excluded nodes = 0, #includ
ed nodes = 0, #source nodes = 0, run during upgrade = false]
20/11/21 04:34:35 INFO balancer.Balancer: included nodes = []
20/11/21 04:34:35 INFO balancer.Balancer: excluded nodes = []
20/11/21 04:34:35 INFO balancer.Balancer: source nodes = []
Time Stamp             Iteration#  Bytes Already Moved  Bytes Left To Move  Bytes Being Moved
20/11/21 04:34:43 INFO balancer.Balancer: dfs.balancer.movedWinWidth = 5400000 (default=5400000)
20/11/21 04:34:43 INFO balancer.Balancer: dfs.balancer.moverThreads = 1000 (default=1000)
20/11/21 04:34:43 INFO balancer.Balancer: dfs.balancer.dispatcherThreads = 200 (default=200)
20/11/21 04:34:43 INFO balancer.Balancer: dfs.datanode.balance.max.concurrent.moves = 50 (default=50)
20/11/21 04:34:43 INFO balancer.Balancer: dfs.balancer.max-size-to-move = 10737418240 (default=10737418240)
20/11/21 04:34:43 INFO net.NetworkTopology: Adding a new node: /default-rack/127.0.0.1:50010
20/11/21 04:34:43 INFO balancer.Balancer: 0 over-utilized: []
20/11/21 04:34:43 INFO balancer.Balancer: 0 underutilized: []
The cluster is balanced. Exiting...
Nov 21, 2020 4:34:43 AM         0              0 B             0 B            -1 B
Nov 21, 2020 4:34:43 AM  Balancing took 10.236 seconds
[cloudera@quickstart ~]$
```

6. Creating a new directory:

```
[cloudera@quickstart ~]$ hdfs dfs -mkdir /abcHarsh
[cloudera@quickstart ~]$
```

7. To view the created directory:

```
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 8 items
drwxr-xr-x   - cloudera supergroup          0 2020-11-21 04:46 /abcHarsh
drwxrwxrwx   - hdfs     supergroup          0 2017-07-19 05:34 /benchmarks
drwxr-xr-x   - hbase    supergroup          0 2020-11-21 04:24 /hbase
drwxr-xr-x   - solr     solr                0 2017-07-19 05:37 /solr
drwxr-xr-x   - cloudera supergroup          0 2020-11-21 04:34 /system
drwxrwxrwt   - hdfs     supergroup          0 2020-11-21 04:25 /tmp
drwxr-xr-x   - hdfs     supergroup          0 2017-07-19 05:36 /user
drwxr-xr-x   - hdfs     supergroup          0 2017-07-19 05:36 /var
[cloudera@quickstart ~]$
```

8. To remove an already existing file use the command:

```
[cloudera@quickstart ~]$ hdfs dfs -rm -r abcHarsh/Test.txt
rm: `abcHarsh/Test.txt': No such file or directory
```

9. To see the content of the file:

```
[cloudera@quickstart ~]$ hadoop fs -cat Test1.txt
cat: `Test1.txt': No such file or directory
```

10. To copy from the remote file of the local system to the hadoop distributed file system:

Copy the file ABC in the folder abc/ABC to the test folder in hadoop with a file name ABC!

- Creating a file locally

```
[cloudera@quickstart ~]$ mkdir forbdaprac
[cloudera@quickstart ~]$ cd forbdaprac/
[cloudera@quickstart forbdaprac]$ echo "Hi this is harsh" >> Testh.txt
[cloudera@quickstart forbdaprac]$ cat Testh.txt
Hi this is harsh
```

- Now applying the command and using cat for checking the file.

```
[cloudera@quickstart forbdaprac]$ hdfs dfs -copyFromLocal /home/cloudera/forbdaprac/Testh.txt /abcHarsh
[cloudera@quickstart forbdaprac]$ hdfs dfs -cat /abcHarsh/Testh.txt
Hi this is harsh
```

11. To see how much space is occupied in HDFS:

```
[cloudera@quickstart forbdaprac]$ hdfs dfs -du -s -s /abcHarsh
17  17  /abcHarsh
[cloudera@quickstart forbdaprac]$
```

12. To delete a specific file in a specific folder:

```
[cloudera@quickstart forbdaprac]$ hdfs dfs -ls /abcHarsh
Found 1 items
-rw-r--r--   1 cloudera supergroup         17 2020-11-21 04:56 /abcHarsh/Testh.txt
[cloudera@quickstart forbdaprac]$ hadoop fs -rm -r /abcHarsh/Testh.txt
Deleted /abcHarsh/Testh.txt
[cloudera@quickstart forbdaprac]$ hdfs dfs -ls /abcHarsh
[cloudera@quickstart forbdaprac]$
```

13. To empty a trash :

```
[cloudera@quickstart forbdaprac]$ hadoop fs -expunge
[cloudera@quickstart forbdaprac]$
```

14. To remove files within a directory:

- Adding some files into the folder and viewing them using ls

```
[cloudera@quickstart forbdaprac]$ hdfs dfs -copyFromLocal /home/cloudera/forbdaprac/Testh.txt /abcHarsh
[cloudera@quickstart forbdaprac]$ echo "Test file 2" >> Testh2.txt
[cloudera@quickstart forbdaprac]$ hdfs dfs -copyFromLocal /home/cloudera/forbdaprac/Testh2.txt /abcHarsh
[cloudera@quickstart forbdaprac]$ hdfs dfs -ls /abcHarsh
Found 2 items
-rw-r--r--   1 cloudera supergroup         17 2020-11-21 05:06 /abcHarsh/Testh.txt
-rw-r--r--   1 cloudera supergroup         12 2020-11-21 05:08 /abcHarsh/Testh2.txt
[cloudera@quickstart forbdaprac]$
```

- Now using the command , removing all files

```
[cloudera@quickstart forbdaprac]$ hadoop fs -rm /abcHarsh/*
Deleted /abcHarsh/Testh.txt
Deleted /abcHarsh/Testh2.txt
[cloudera@quickstart forbdaprac]$ hdfs dfs -ls /abcHarsh
[cloudera@quickstart forbdaprac]$
```