Experiment 3

Shashwat Shah
60004220126
TY BTECH B2

Aim: Execute different HDFS commands

Theory: The hadoop Distributed File System (HDFS) is a distributed file system designed to run on commodity hardware.
It has many similarities with existing distributed file system. However, the difference from other distributed file systems are significant.
HDFS is highly fault-tolerant and is designed to be low cost hardware.
HDFS provides high throughput access to application data and is suitable for the application that have large dataset
HDFS relaxes a few POSTFIX requirements to enable streamy access to a file system data.
HDFS was originally built as infrastructure for the Apache Nutch web search engine project.

Some of the HDFS commands are:-

1) ls - Mainly lists all the files. Use lsr to list in a recursive approach.
→ hdfs dfs -ls /

2) mkdir - mainly to create a new directory
→ hdfs dfs -mkdir /user

3) touchz : creates an empty file.
→ hdfs dfs -touchz /user/myfile.txt

4) cat - To print the file content
→ hdfs dfs - cat /user/sample.txt

5) cp - Mainly to copy files within hdfs.
→ hdfs dfs - cp /user/user-copied

6) mv - move files within hdfs.
→ hdfs dfs -mv /user/myfile.txt /user-copied

7) rmr - delete files recursively
→ hdfs dfs - rmr /user-copied

8) du - gives the size for each file
→ hdfs dfs - du /user.

9) dus - total size of the directory/file
→ hdfs dfs - dus /user.

10) Stat - Gives the last modified time
→ hdfs dfs -stat /user.

11) setrep - mainly to change the replication factor.
→ hdfs dfs. -setrep -R 4/user.

Conclusion - Thus we have successfully executed different HDFS commands

Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

HDFS Commands

After successful installation of Hadoop, we execute different HDFS Commands.

Open a new Windows Command Prompt and run below commands. C:\hadoop-2.9.1\hadoop-2.9.1>cd bin ls: This command is used to list all the files. Use lsr for recursive approach. It is useful when we want a

hierarchy of a folder.

C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /

```
Found 2 items
drwxr-xr-x - LENOVO supergroup          0 2023-02-27 20:06 /sampledir
drwxr-xr-x - LENOVO supergroup          0 2023-02-27 19:57 /test
```

mkdir: To create a directory. In Hadoop *dfs* there is no home directory by default.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -mkdir /user C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -mkdir /user/Lenovo
```

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls / Found
3 items
drwxr-xr-x - LENOVO supergroup  0 2023-02-27 20:06 /sampledir drwxr-xr-x - LENOVO supergroup  0 2023-02-27 19:57 /test
```

```
drwxr-xr-x - LENOVO supergroup          0 2023-02-27 20:23 /user
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -lsr /user lsr:
DEPRECATED: Please use 'ls -R' instead.
drwxr-xr-x - LENOVO supergroup          0 2023-02-27 20:23 /user/Lenovo
```

touchz: It creates an empty file.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -touchz /user/myfile.txt C:\hadoop-2.9.1\hadoop-
2.9.1\bin>hdfs dfs -ls -R /user drwxr-xr-x - LENOVO supergroup    0 2023-02-27
20:23 /user/Lenovo -rw-r--r-- 1 LENOVO supergroup          0 2023-02-27 20:26
/user/myfile.txt
```

copyFromLocal (or) put: To copy files/folders from local file system to hdfs store. This is the most

important command. Local filesystem means the files present on the OS.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -put C:/hadoop-2.9.1/hadoop-2.9.1/Sample.txt /user     0
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls -R /user drwxr-xr-x - LENOVO supergroup
2023-02-27 20:23 /user/Lenovo -rw-r--r-- 1 LENOVO supergroup  12 2023-02-27 20:30
/user/Sample.txt -rw-r--r-- 1 LENOVO supergroup             0 2023-02-27 20:26 /user/myfile.txt
```

cat: To print file contents.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -cat /user/Sample.txt Hello Hadoop
```

copyToLocal (or) get: To copy files/folders from hdfs store to local file system.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -get /user/Sample.txt ../HadoopExamples
```

cp: This command is used to copy files within hdfs.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -mkdir /user_copied C:\hadoop-2.9.1\hadoop-
2.9.1\bin>hdfs dfs -cp /user /user_copied C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls
/user_copied
```

```
Found 1 items
drwxr-xr-x - LENOVO supergroup   0 2023-02-27 20:48 /user_copied/user
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /user Found 3 items drwxr-xr-x -
LENOVO supergroup           0 2023-02-27 20:23 /user/Lenovo
```

```
-rw-r--r-- 1 LENOVO supergroup          12 2023-02-27 20:30 /user/Sample.txt
-rw-r--r-- 1 LENOVO supergroup           0 2023-02-27 20:26 /user/myfile.txt
```

mv: This command is used to move files within hdfs.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -mv /user/myfile.txt /user_copied
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /user Found
2 items
drwxr-xr-x - LENOVO supergroup           0 2023-02-27 20:23 /user/Lenovo
-rw-r--r-- 1 LENOVO supergroup          12 2023-02-27 20:30 /user/Sample.txt
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /user_copied
Found 2 items
-rw-r--r-- 1 LENOVO supergroup    0 2023-02-27 20:26 /user_copied/myfile.txt drwxr-xr-x -
LENOVO supergroup           0 2023-02-27 20:48 /user_copied/user
```

rmr: This command deletes a file from HDFS *recursively*. It is very useful command when you want to delete a *non-empty directory*.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -rmr /user_copied
rmr: DEPRECATED: Please use '-rm -r' instead.

Deleted /user_copied C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs
ls /user_copied ls: `/user_copied': No such file or directory
```

du: It will give the size of each file in directory.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -du /user 0 /user/Lenovo 12 /user/Sample.txt
```

dus: This command will give the total size of directory/file.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -dus /user dus: DEPRECATED:

Please use 'du -s' instead.

12 /user
```

stat: It will give the last modified time of directory or path. In short it will give stats of the directory or file.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -stat /user 2023-02-27 15:20:27
```

setrep: This command is used to change the replication factor of a file/directory in HDFS. By default it is 3 for anything which is stored in HDFS (as set in hdfs *core-site.xml*).

```
c:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -setrep -R 4 /user Replication

4 set: /user/Sample.txt

c:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /user Found

2 items

drwxr-xr-x - LENOVO supergroup          0 2023-02-27 20:23 /user/Lenovo

-rw-r--r-- 4 LENOVO supergroup          12 2023-02-27 20:30 /user/Sample.txt
```

CONCLUSION: We have successfully execute different HDFS Commands.