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