**HDFS Exercise**

1. Create a Directory in HDFS Filesystem with name as 'mydata'.

hdfs dfs -mkdir /user/Sreeram/Hadoop/mydata

3. Create sub directories /mydata/source and /mydata/output in HDFS.

hdfs dfs -mkdir /user/Sreeram/Hadoop/mydata/source

hdfs dfs -mkdir /user/Sreeram/Hadoop/mydata/output

4. Examine the contents of the directory /mydata in HDFS file system.

hdfs dfs -ls /user/Sreeram/Hadoop/mydata

5. Upload “sample.csv” file from local file system to /mydata/source in HDFS file system.

hdfs dfs -put /home/ms83/R23884/sample.csv /user/Sreeram/Hadoop/mydata/source

6. Examine the contents of /mydata/source diretory in HDFS.

hdfs dfs -ls /user/Sreeram/Hadoop/mydata/source

7. Copy the “sample.csv” from HDFS file system to /mydata/output.

hdfs dfs -cp /user/Sreeram/Hadoop/mydata/source/sample.csv /user/Sreeram/Hadoop/mydata/output

8. Examine the contents of /mydata/output diretory in HDFS file system.

hdfs dfs -ls /user/Sreeram/Hadoop/mydata/output

9. Display the contents of “sample.csv” in HDFS.

hdfs dfs -cat /user/Sreeram/Hadoop/mydata/source/sample.csv

or

hdfs dfs -cat /user/Sreeram/Hadoop/mydata/output/sample.csv

10. Display the last few lines of “sample.csv” in HDFS.

hdfs dfs -cat /user/Sreeram/Hadoop/mydata/output/sample.csv | tail -10

11. Download “sample.csv” file in /mydata/output to Local file system.

hdfs dfs -get /user/Sreeram/Hadoop/mydata/output/sample.csv /home/ms83/R23884/samplefromhdfs.csv

12. Delete the output directory in HDFS file system.

hdfs dfs -rm -r /user/Sreeram/Hadoop/mydata/output

13. Copy “example.log” file from Local file system to /mydata/source in HDFS.

hdfs dfs -put /home/ms83/R23884/example.log /user/Sreeram/Hadoop/mydata/source

14. Create a directory “LogFiles” in /mydata> directory in HDFS.

hdfs dfs -mkdir /user/Sreeram/Hadoop/mydata/LogFiles

15. Move “example.log” to /mydata/LogFiles in HDFS.

hdfs dfs -mv /user/Sreeram/Hadoop/mydata/source/elec user/Sreeram/Hadoop/mydata/LogFiles

16. Create an empty file named “summary.txt” in /mydata/source folder in HDFS.

hdfs dfs -touchz /user/Sreeram/Hadoop/mydata/source/summary.txt

17. Check the existence of “summary.txt” in HDFS. Check whether it is zero length file or not.

hdfs dfs -ls /user/Sreeram/Hadoop/mydata/source

18. Display the size of files and directories present in /mydata/source in HDFS.

hdfs dfs -ls /user/Sreeram/Hadoop/mydata/source

19. Merge the files in /mydata/source folder to a destination file. Add new line at the end of each file.

hdfs dfs -getmerge /user/Sreeram/Hadoop/mydata/source /home/ms83/R23884/megedfiles

-HDFS is onetime file system - Write once read many times , cannot be appended or updated to a existing file

20. Count the number of files and directories and their size in HDFS

We cannot count number of files in Directory in HDFS , rather we can count number of records in a file or aggregate total number of records in directory

hdfs dfs -cat /user/Sreeram/Hadoop/MR/part-m\* | wc -l

hdfs dfs -du -s /user/Sreeram/Hadoop/mydata/source