

24
25
(Impact)

Arzay Shah
60004210073
C12-TE Comp

Experiment No. 3

Aim: To execute HDFS commands in Hadoop Hortonworks Sandbox

Theory:

- i) Hadoop File System was developed using distributed file system design
- ii) It is run on commodity hardware
- iii) Unlike other distributed systems, HDFS is highly fault-tolerant and designed using low-cost hardware.
- iv) HDFS holds a very large amount of data & provides easier access.
- v) To store such huge data, the files are stored across multiple machines
- vi) Those files are stored in a redundant fashion to rescue the system from possible data losses in case of failure.
- vii) HDFS also makes applications available for parallel processing.

HDFS Architecture

- i) Name Node
 - a) NameNode is the master node that contains the metadata
 - b) The NameNode is responsible for the workings of the data nodes.
 - c) NameNode is the primary server that manages the file system namespace and controls client access to files.
 - d) The NameNode performs file system namespace operations, including opening, closing & renaming files and directories.
 - e) The NameNode also governs the mapping of blocks to the Datanode.

ii) DataNode

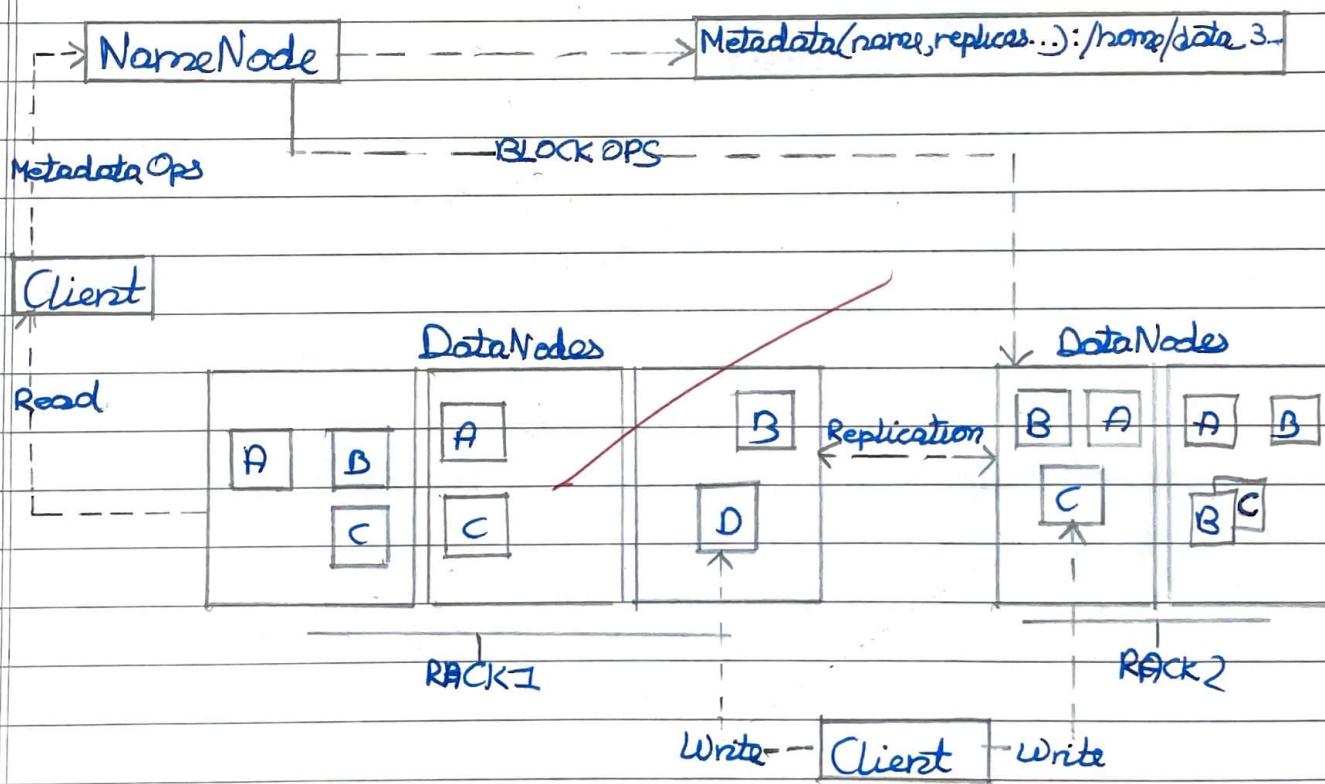
- a) The DataNodes are called the slaves.
- b) The DataNodes read, write, process, and replicate the data.
- c) They also send signals, known as heartbeats, to the NameNode. These heartbeats show the status of the DataNode.
- d) While there is only one NameNode, there can be multiple DataNodes.

HDFS Commands

- i) ls - This command is used to list all the files
eg - hdfs dfs -ls
- ii) mkdir - To create a directory
eg - hdfs dfs -mkdir /user
- iii) touchz - It creates an empty file
eg - hdfs dfs -touchz /user/nomyfile.txt
- iv) copyFromLocal or put - To copy files/folders from local file system to hdfs store
eg - hdfs dfs -put C:/hadoop-2.9.1/Sample.txt /user
- v) cat - To print file contents
eg - hdfs dfs -cat /user/Sample.txt
- vi) copyToLocal or get - To copy files/folders from hdfs store to local file system
eg - hdfs dfs -get /user/Sample.txt /HadoopExamples
- vii) cp - This command is used to copy files within hdfs
eg - hdfs dfs -cp /user/user_copied
- viii) mv - This command is used to move files within hdfs
eg - hdfs dfs -mv /user/nomyfile.txt /user_copied
- ix) rmr - This command deletes a file from HDFS recursively.
eg - hdfs dfs -rmr /user_copied
- x) du - It will give the size of each file in directory
eg - hdfs dfs -du /user

- x) du - This command will give the total size of directory/file
 eg - hdfs dfs -du /user
- xii) stat - It will give the last modified time of directory or path
 eg - hdfs dfs -stat
- xiii) setrep - This command is used to change the replication factor of a file/directory in HDFS.
 eg - hdfs dfs -setrep

Architecture:



Conclusion: In conclusion we have performed several HDFS commands and learnt about the basics and architecture of HDFS.

Q&A