**EXPERIMENT 01**

**Part 3**

CLASS: BE CMPN A ROLL NO. : 19

NAME: REBECCA DIAS PID: 182027

Aim:- To execute the HDFS Commands.

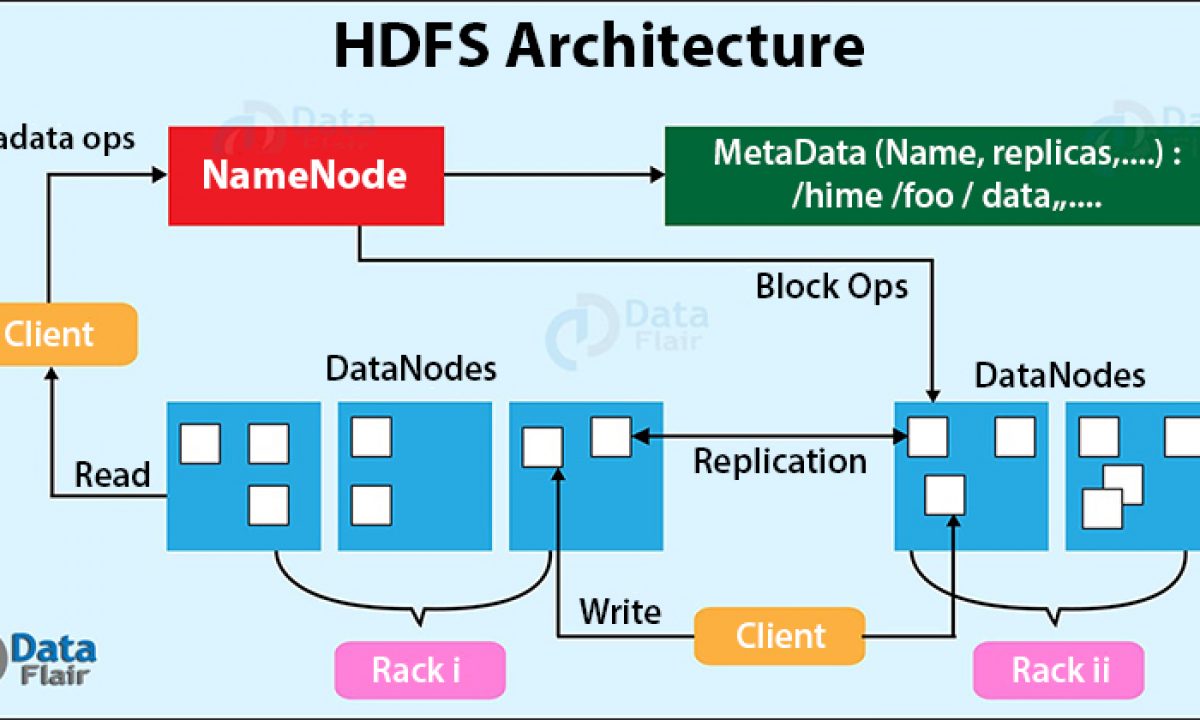
Theory:- Write about the basics of HDFS.

Architecture of HDFS –

The Hadoop architecture is a package of the file system, MapReduce engine and the HDFS (Hadoop Distributed File System). The MapReduce engine can be MapReduce/MR1 or YARN/MR2.

A Hadoop cluster consists of a single master and multiple slave nodes. The master node includes Job Tracker, Task Tracker, NameNode, and DataNode whereas the slave node includes DataNode and TaskTracker. The Hadoop Distributed File System (HDFS) is a distributed file system for Hadoop. It contains a master/slave architecture. This architecture consist of a single NameNode performs the role of master, and multiple DataNodes performs the role of a slave.

Both NameNode and DataNode are capable enough to run on commodity machines. The Java language is used to develop HDFS. So any machine that supports Java language can easily run the NameNode and DataNode software.

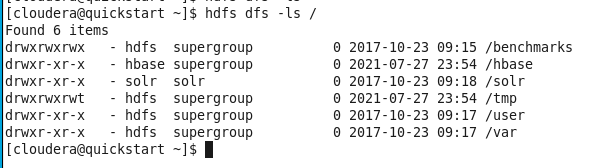


Commands:-

**1. File commands for listing**

1. File commands for listing

hdfs dfs –ls / :List all the files/directories for the given hdfs destination path



hdfs dfs -ls -d /user : Directories are listed as plain files. In this case, this command will list the details of hadoop folder.

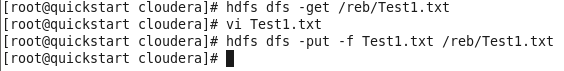


hdfs dfs -ls -R /home:- Recursively list all files in hadoop directory and all subdirectories in hadoop directory.

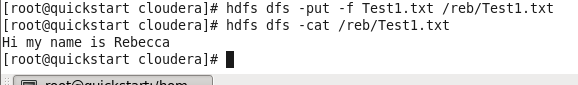


**2. File commands for reading and writing files**

hdfs dfs -put Test1.txt:- The put command is used to copy a file from a local file system to HDFS



hdfs dfs -cat Test1.txt:- To see the contents of the file Test1.txt.

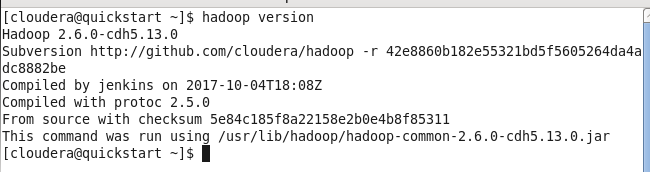


Hdfs dfs –tail/reb/Test1.txt :- tail command can be used in the place of cat command as well



**3. File commands for displaying the version of hadoop**

Hadoop version:- Used to display the version of hadoop

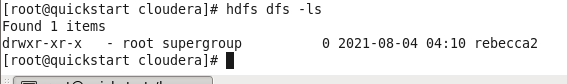


**4. File commands for making a directory and copying contents inside that directory**

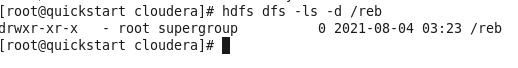
hdfs dfs -mkdir rebecca2 :- creating a sample directory



hdfs dfs -ls -R /rebecca2 :- Check if the sample directory is created



hdfs dfs -ls -R /home:- Recursively list all files in hadoop directory and all subdirectories in hadoop directory.



**5. File commands for copying to/from the local directory**

hdfs dfs –copyFromLocal /reb/Test1.txt - To copy from the remote file of the local system to the hadoop distributed file system.

hdfs dfs –copyFromLocal Test1.txt/reb/Test1.txt: This command is used for copying from the local directory to a hdfs directory sample



**6. Other commands**

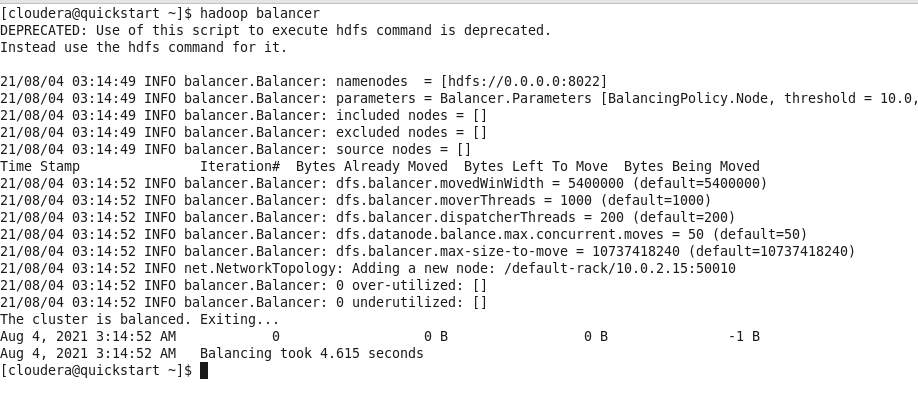
hdfs dfs -df hdfs:/ : Report the amount of space used and available on currently mounted file system



hdfs dfs -count hdfs:/ : Count the number of directories, files and bytes under the paths that match the specified file pattern



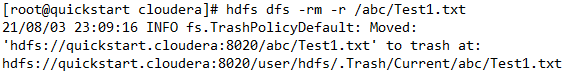
hadoop balancer : Run a cluster balancing utility



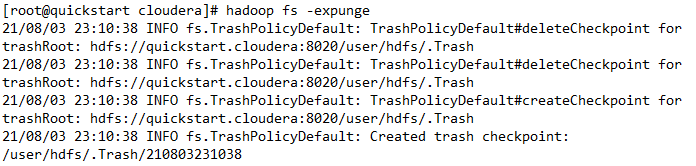
hdfs dfs -du -s –s /reb/Test1.txt: To see how much space is occupied in HDFS.



hdfs fs –rm - r reb/Test1.txt : To remove an already existing file use the command



hadoop fs –expunge : To empty a trash the following command is used



**Conclusion:**

In this experiment we learned how to install and use Hadoop and Cloudera. We learnt the use of different commands for file listing, commands for reading and writing files, commands for making a directory and copying contents inside that directory and commands for copying from hdfs to the local directory.