**EXPERIMENT 1**

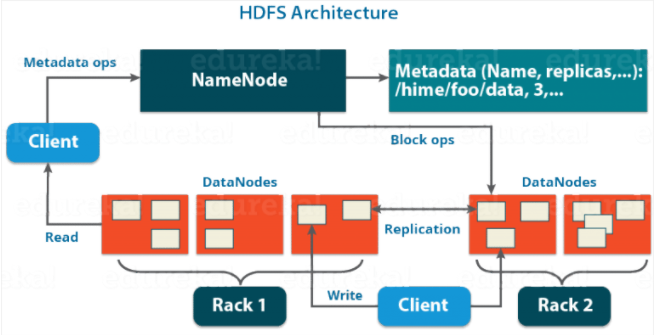
**CLASS: BE CMPN A NAME: Tanmay Hatkar ROLL NO: 33**

**Aim:-** To execute the HDFS Commands

**Theory:-** Write about the basics of HDFS.

Architecture of HDFS

**Apache HDFS** or **Hadoop Distributed File System** is a block-structured file system where each file is divided into blocks of a pre-determined size. These blocks are stored across a cluster of one or several machines. Apache Hadoop HDFS Architecture follows a *Master/Slave Architecture*, where a cluster comprises of a single NameNode (Master node) and all the other nodes are DataNodes (Slave nodes). HDFS can be deployed on a broad spectrum of machines that support Java. Though one can run several DataNodes on a single machine, but in the practical world, these DataNodes are spread across various machines.

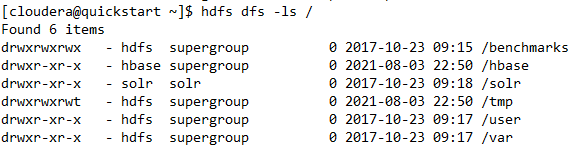


**1. File commands for listing**

Commands to be executed

1. File commands for listing

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



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



c. 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.

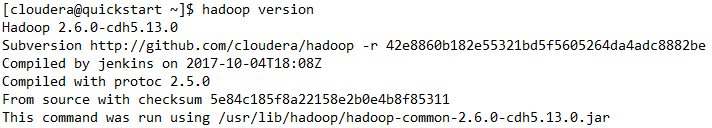


hdfs dfs -tail /user/vagrant/Test1.txt: tail command can be used in the place of cat command as well



**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 sample:- creating a sample directory

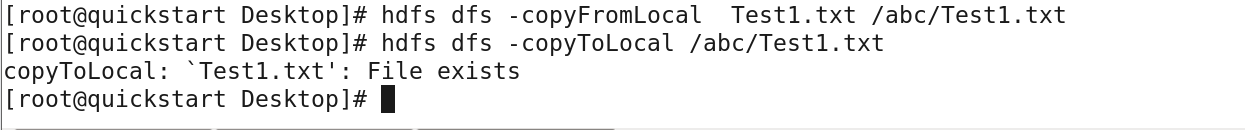


hdfs dfs -ls -R /:- Check if the sample directory is created



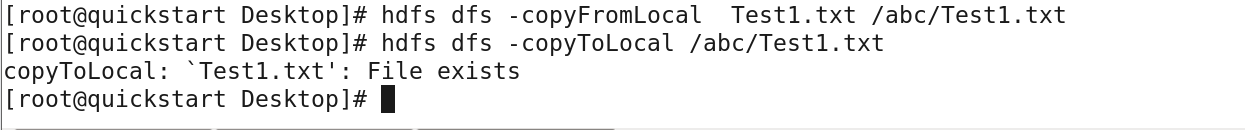
hdfs dfs -copyFromLocal /home/vagrant/abc/ABC/ABC.txt /user/vagrant/test/ABC1.txt - To copy from the remote file of the local system to the hadoop distributed file system.

hdfs dfs -copyFromLocal /home/vagrant/abc/ABC/ /user/vagrant/sample: This command is used for copying from the local directory to a hdfs directory sample



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

hdfs dfs -copyToLocal /user/vagrant/sample/Posts.txt /home/vagrant/abc: This command is used to copy from a hdfs folder sample a file Posts to a local folder abc



**6. Other commands**

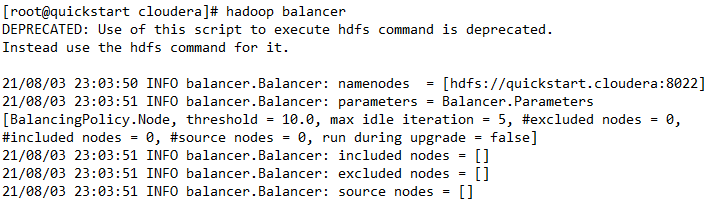
hdfs dfs -df hdfs:/ : Report the amount of space used and available on currently mounted file system



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



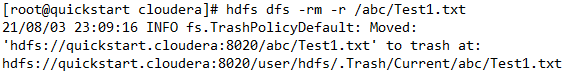
hadoop balancer : Run a cluster balancing utility



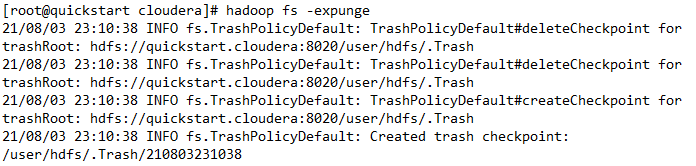
hdfs dfs -du -s -s abc/ABC : To see how much space is occupied in HDFS.



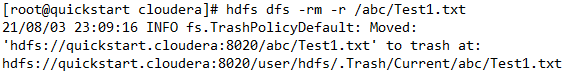
hdfs fs –rm - r abc/Test.txt : To remove an already existing file use the command



vagrant@master:~$ hadoop fs –expunge : To empty a trash the following command is used



hadoop fs -rm /user/vagrant/abc/ABC/\* : To remove files within a directory



**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.