EXPERIMENT 01

Part 3

CLASS: BE CMPN A ROLL NO. : 19

NAME: REBECCA DIAS PID: 182027

Aim:- To execute the HDFS Commands.

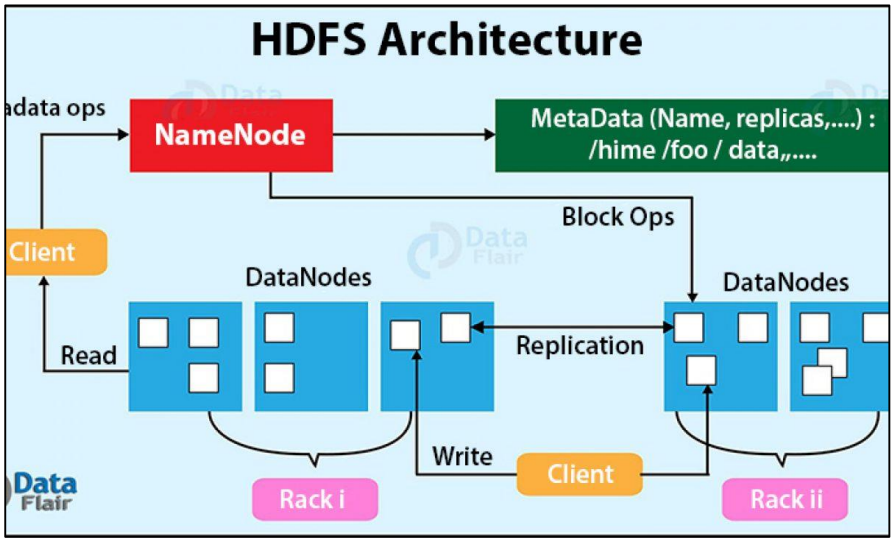
Theory:- Write about the basics of HDFS.

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

```
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 6 items
drwxrwxrwx   - hdfs  supergroup          0 2017-10-23 09:15 /benchmarks
drwxr-xr-x   - hbase supergroup          0 2021-07-27 23:54 /hbase
drwxr-xr-x   - solr  solr                0 2017-10-23 09:18 /solr
drwxrwxrwt   - hdfs  supergroup          0 2021-07-27 23:54 /tmp
drwxr-xr-x   - hdfs  supergroup          0 2017-10-23 09:17 /user
drwxr-xr-x   - hdfs  supergroup          0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$
```

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

```
[root@quickstart cloudera]# hdfs dfs -ls -d /user
drwxr-xr-x   - hdfs supergroup          0 2021-08-03 23:06 /user
```

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

```
[root@quickstart cloudera]# hdfs dfs -ls -d /abc
drwxr-xr-x   - hdfs supergroup          0 2021-08-03 23:09 /abc
```

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

```
[root@quickstart cloudera]# hdfs dfs -get /reb/Test1.txt
[root@quickstart cloudera]# vi Test1.txt
[root@quickstart cloudera]# hdfs dfs -put -f Test1.txt /reb/Test1.txt
[root@quickstart cloudera]#
```

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

```
[root@quickstart cloudera]# hdfs dfs -put -f Test1.txt /reb/Test1.txt
[root@quickstart cloudera]# hdfs dfs -cat /reb/Test1.txt
Hi my name is Rebecca
[root@quickstart cloudera]#
```

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

```
[root@quickstart cloudera]# hdfs dfs -tail /reb/Test1.txt
Hi my name is Rebecca
[root@quickstart cloudera]#
```

3. File commands for displaying the version of hadoop

Hadoop version:- Used to display the version of hadoop

```
[cloudera@quickstart ~]$ hadoop version
Hadoop 2.6.0-cdh5.13.0
Subversion http://github.com/cloudera/hadoop -r 42e8860b182e55321bd5f5605264da4a
dc8882be
Compiled by jenkins on 2017-10-04T18:08Z
Compiled with protoc 2.5.0
From source with checksum 5e84c185f8a22158e2b0e4b8f85311
This command was run using /usr/lib/hadoop/hadoop-common-2.6.0-cdh5.13.0.jar
[cloudera@quickstart ~]$
```

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

hdfs dfs -mkdir rebecca2 :- creating a sample directory

```
[root@quickstart cloudera]# hdfs dfs -mkdir rebecca2
[root@quickstart cloudera]#
```

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

```
[root@quickstart cloudera]# hdfs dfs -ls
Found 1 items
drwxr-xr-x   - root supergroup          0 2021-08-04 04:10 rebecca2
[root@quickstart cloudera]#
```

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

```
[root@quickstart cloudera]# hdfs dfs -ls -d /reb
drwxr-xr-x   - root supergroup          0 2021-08-04 03:23 /reb
[root@quickstart cloudera]#
```

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

```
[root@quickstart cloudera]# hdfs dfs -copyToLocal /reb/Test1.txt
copyToLocal: `Test1.txt': File exists
[root@quickstart cloudera]# hdfs dfs -copyFromLocal Test1.txt/reb/Test1.txt
```

6. Other commands

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

```
[cloudera@quickstart ~]$  hdfs dfs -df hdfs:/
Filesystem                                Size       Used    Available  Use%
hdfs://quickstart.cloudera:8020  58479091712  872557143  46088228864    1%
```

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

```
[cloudera@quickstart ~]$ hdfs dfs -count hdfs:/
          75          934          861289410 hdfs:///
```

hadoop balancer : Run a cluster balancing utility

```
[cloudera@quickstart ~]$ hadoop balancer
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

21/08/04 03:14:49 INFO balancer.Balancer: namenodes  = [hdfs://0.0.0.0:8022]
21/08/04 03:14:49 INFO balancer.Balancer: parameters = Balancer.Parameters [BalancingPolicy.Node, threshold = 10.0,
21/08/04 03:14:49 INFO balancer.Balancer: included nodes = []
21/08/04 03:14:49 INFO balancer.Balancer: excluded nodes = []
21/08/04 03:14:49 INFO balancer.Balancer: source nodes = []
Time Stamp               Iteration#  Bytes Already Moved  Bytes Left To Move  Bytes Being Moved
21/08/04 03:14:52 INFO balancer.Balancer: dfs.balancer.movedWinWidth = 5400000 (default=5400000)
21/08/04 03:14:52 INFO balancer.Balancer: dfs.balancer.moverThreads = 1000 (default=1000)
21/08/04 03:14:52 INFO balancer.Balancer: dfs.balancer.dispatcherThreads = 200 (default=200)
21/08/04 03:14:52 INFO balancer.Balancer: dfs.datanode.balance.max.concurrent.moves = 50 (default=50)
21/08/04 03:14:52 INFO balancer.Balancer: dfs.balancer.max-size-to-move = 10737418240 (default=10737418240)
21/08/04 03:14:52 INFO net.NetworkTopology: Adding a new node: /default-rack/10.0.2.15:50010
21/08/04 03:14:52 INFO balancer.Balancer: 0 over-utilized: []
21/08/04 03:14:52 INFO balancer.Balancer: 0 underutilized: []
The cluster is balanced. Exiting...
Aug 4, 2021 3:14:52 AM           0                 0 B                 0 B              -1 B
Aug 4, 2021 3:14:52 AM   Balancing took 4.615 seconds
[cloudera@quickstart ~]$
```

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

```
[root@quickstart cloudera]# hdfs dfs -du -s -s /abc/Test1.txt
26  26  /abc/Test1.txt
```

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

```
[root@quickstart cloudera]# hdfs dfs -rm -r /abc/Test1.txt
21/08/03 23:09:16 INFO fs.TrashPolicyDefault: Moved:
'hdfs://quickstart.cloudera:8020/abc/Test1.txt' to trash at:
hdfs://quickstart.cloudera:8020/user/hdfs/.Trash/Current/abc/Test1.txt
```

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

```
[root@quickstart cloudera]# hadoop fs -expunge
21/08/03 23:10:38 INFO fs.TrashPolicyDefault: TrashPolicyDefault#deleteCheckpoint for
trashRoot: hdfs://quickstart.cloudera:8020/user/hdfs/.Trash
21/08/03 23:10:38 INFO fs.TrashPolicyDefault: TrashPolicyDefault#deleteCheckpoint for
trashRoot: hdfs://quickstart.cloudera:8020/user/hdfs/.Trash
21/08/03 23:10:38 INFO fs.TrashPolicyDefault: TrashPolicyDefault#createCheckpoint for
trashRoot: hdfs://quickstart.cloudera:8020/user/hdfs/.Trash
21/08/03 23:10:38 INFO fs.TrashPolicyDefault: Created trash checkpoint:
/user/hdfs/.Trash/210803231038
```

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.