Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

BDI

Name: Kartik Jolapara **Branch:** Computer Engineering

SAP ID: 60004200107 **Batch:** B1

EXPERIMENT NO. 3
HDFS Commands

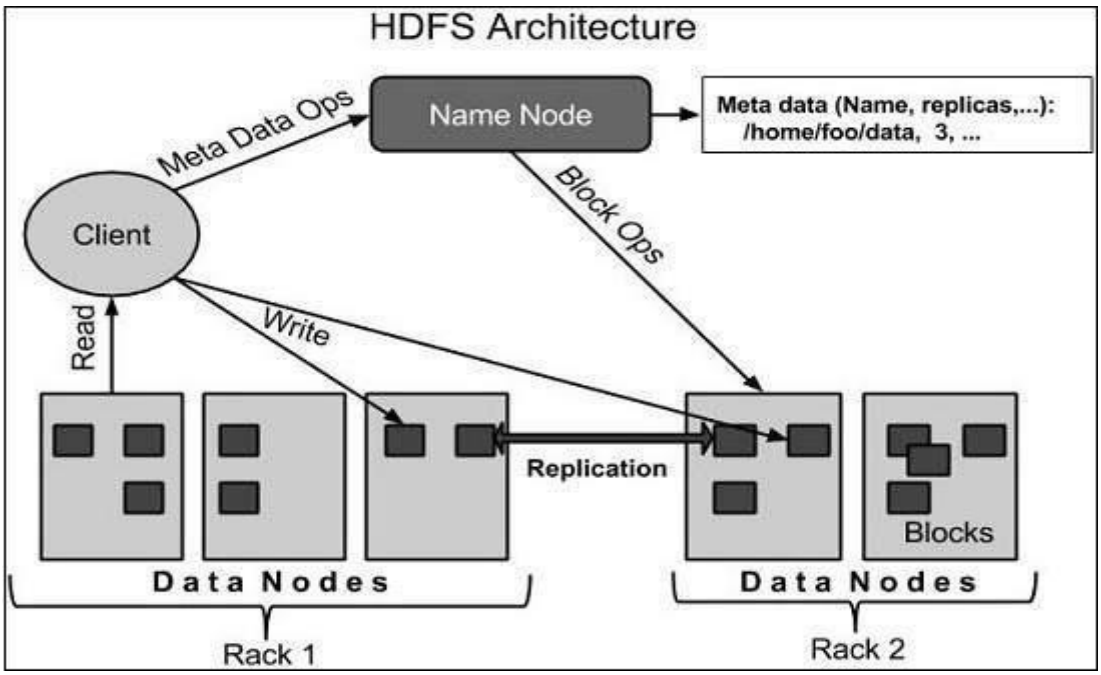
AIM: Execute different HDFS Commands.

THEORY:

Hadoop is a software framework that enables distributed storage and processing of large data sets. It consists of several open source projects, including HDFS, MapReduce, and Yarn. While Hadoop can be used for different purposes, the two most common are Big Data analytics and NoSQL database management. HDFS stands for "Hadoop Distributed File System" and is a decentralized file system that stores data across multiple computers in a cluster. This makes it ideal for large-scale storage as it distributes the load across multiple machines so there's less pressure on each individual machine. MapReduce is a programming model that allows users to write code once and execute it across many servers. When combined with HDFS, MapReduce can be used to process massive data sets in parallel by dividing work up into smaller chunks and executing them simultaneously.

HDFS Architecture

HDFS is an Open source component of the Apache Software Foundation that manages data. HDFS has scalability, availability, and replication as key features. Name nodes, secondary name nodes, data nodes, checkpoint nodes, backup nodes, and blocks all make up the architecture of HDFS. HDFS is fault-tolerant and is replicated. Files are distributed across the cluster systems using the Name node and Data Nodes. The primary difference between Hadoop and Apache HBase is that Apache HBase is a non-relational database and Apache Hadoop is a non-relational data store.



HDFS follows the master-slave architecture and it has the following elements.

Namenode

The namenode is the commodity hardware that contains the GNU/Linux operating system and the namenode software. It is a software that can be run on commodity hardware. The system having the namenode acts as the master server and it does the following tasks −

- Manages the file system namespace.
- Regulates client's access to files.
- It also executes file system operations such as renaming, closing, and opening files and directories.

Datanode

The datanode is a commodity hardware having the GNU/Linux operating system and datanode software. For every node (Commodity hardware/System) in a cluster, there will be a datanode. These nodes manage the data storage of their system.

- Datanodes perform read-write operations on the file systems, as per client request.
- They also perform operations such as block creation, deletion, and replication according to the instructions of the namenode.

Block

Generally the user data is stored in the files of HDFS. The file in a file system will be divided into one or more segments and/or stored in individual data nodes. These file segments are called as blocks. In other words, the minimum amount of data that HDFS can read or write is called a Block. The default block size is 64MB, but it can be increased as per the need to change in HDFS configuration.

HDFS Commands

After successful installation of Hadoop, we execute different HDFS Commands.

Open a new Windows Command Prompt and run below commands. C:\hadoop-2.9.1\hadoop-2.9.1>cd bin

ls: This command is used to list all the files. Use lsr for recursive approach. It is useful when we want a hierarchy of a folder.

C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /

```
Found 2 items
drwxr-xr-x - LENOVO supergroup          0 2023-02-27 20:06 /sampledir
drwxr-xr-x - LENOVO supergroup          0 2023-02-27 19:57 /test
```

mkdir: To create a directory. In Hadoop *dfs* there is no home directory by default.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -mkdir /user C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -mkdir /user/Lenovo
```

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls / Found
3 items
drwxr-xr-x - LENOVO supergroup    0 2023-02-27 20:06 /sampledir
drwxr-xr-x - LENOVO supergroup    0 2023-02-27 19:57 /test
```

```
drwxr-xr-x - LENOVO supergroup          0 2023-02-27 20:23 /user
```

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -lsr /user
lsr: DEPRECATED: Please use 'ls -R' instead.
drwxr-xr-x - LENOVO supergroup          0 2023-02-27 20:23 /user/Lenovo
```

touchz: It creates an empty file.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -touchz /user/myfile.txt C:\hadoop-2.9.1\hadoop-
2.9.1\bin>hdfs dfs -ls -R /user
drwxr-xr-x - LENOVO supergroup          0 2023-02-27 20:23 /user/Lenovo
-rw-r--r-- 1 LENOVO supergroup          0 2023-02-27 20:26 /user/myfile.txt
```

copyFromLocal (or) put: To copy files/folders from local file system to hdfs store. This is the most important command. Local filesystem means the files present on the OS.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -put C:/hadoop-2.9.1/hadoop-2.9.1/Sample.txt /user
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls -R /user drwxr-xr-x - LENOVO supergroup          0
2023-02-27 20:23 /user/Lenovo -rw-r--r-- 1 LENOVO supergroup  12 2023-02-27 20:30
/user/Sample.txt -rw-r--r-- 1 LENOVO supergroup          0 2023-02-27 20:26 /user/myfile.txt
```

cat: To print file contents.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -cat /user/Sample.txt Hello Hadoop
```

copyToLocal (or) get: To copy files/folders from hdfs store to local file system.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -get /user/Sample.txt ../HadoopExamples
```

cp: This command is used to copy files within hdfs.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -mkdir /user_copied C:\hadoop-2.9.1\hadoop-
2.9.1\bin>hdfs dfs -cp /user /user_copied C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls
/user_copied
```

```
Found 1 items
drwxr-xr-x - LENOVO supergroup          0 2023-02-27 20:48 /user_copied/user
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /user Found 3 items
drwxr-xr-x - LENOVO supergroup          0 2023-02-27 20:23 /user/Lenovo
```

```
-rw-r--r-- 1 LENOVO supergroup          12 2023-02-27 20:30 /user/Sample.txt
-rw-r--r-- 1 LENOVO supergroup          0 2023-02-27 20:26 /user/myfile.txt
```

mv: This command is used to move files within hdfs.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -mv /user/myfile.txt /user_copied
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /user Found
2 items
drwxr-xr-x - LENOVO supergroup          0 2023-02-27 20:23 /user/Lenovo
-rw-r--r-- 1 LENOVO supergroup          12 2023-02-27 20:30 /user/Sample.txt
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /user_copied
Found 2 items
-rw-r--r-- 1 LENOVO supergroup       0 2023-02-27 20:26 /user_copied/myfile.txt
drwxr-xr-x - LENOVO supergroup       0 2023-02-27 20:48 /user_copied/user
```

rmr: This command deletes a file from HDFS *recursively*. It is very useful command when you want to delete a *non-empty directory*.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -rmr /user_copied
rmr: DEPRECATED: Please use '-rm -r' instead.
```

```
Deleted /user_copied C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -
ls /user_copied ls: `/user_copied': No such file or directory
```

du: It will give the size of each file in directory.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -du /user 0 /user/Lenovo 12 /user/Sample.txt
```

dus: This command will give the total size of directory/file.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -dus /user dus: DEPRECATED:
Please use 'du -s' instead.
12  /user
```

stat: It will give the last modified time of directory or path. In short it will give stats of the directory or file.

```
C:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -stat /user 2023-02-27 15:20:27
```

setrep: This command is used to change the replication factor of a file/directory in HDFS. By default it is 3 for anything which is stored in HDFS (as set in hdfs *core-site.xml*).

```
c:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -setrep -R 4 /user
Replication 4 set: /user/Sample.txt
c:\hadoop-2.9.1\hadoop-2.9.1\bin>hdfs dfs -ls /user Found
2 items
drwxr-xr-x - LENOVO supergroup          0 2023-02-27 20:23 /user/Lenovo
-rw-r--r-- 4 LENOVO supergroup         12 2023-02-27 20:30 /user/Sample.txt
```

CONCLUSION: We have successfully execute different HDFS Commands.