**Q1) Setup Hadoop in Linux Environment**

**Software Prerequisite**

Java 8 installed on the Linux OS

1. **Sudo apt update**

```
priya_hadoop@priya-VB:~$ sudo apt update
[sudo] password for priya_hadoop:
Hit:1 http://in.archive.ubuntu.com/ubuntu focal InRelease
Get:2 http://in.archive.ubuntu.com/ubuntu focal-updates InRelease [114 kB]
Get:3 http://in.archive.ubuntu.com/ubuntu focal-backports InRelease [101 kB]
Get:4 http://security.ubuntu.com/ubuntu focal-security InRelease [114 kB]
Get:5 https://packages.microsoft.com/repos/ms-teams stable InRelease [17.5 kB]
Get:6 http://dl.google.com/linux/chrome/deb stable InRelease [1,811 B]
Get:7 http://in.archive.ubuntu.com/ubuntu focal-updates/main i386 Packages [544 kB]
Get:8 http://in.archive.ubuntu.com/ubuntu focal-updates/main amd64 Packages [1,256 kB]
Get:9 https://packages.microsoft.com/repos/ms-teams stable/main amd64 Packages [8,208 B]
Get:10 http://dl.google.com/linux/chrome/deb stable/main amd64 Packages [1,092 B]
Get:11 http://security.ubuntu.com/ubuntu focal-security/main amd64 DEP-11 Metadata [29.0 kB]
Get:12 http://in.archive.ubuntu.com/ubuntu focal-updates/main amd64 DEP-11 Metadata [283 kB]
Get:13 http://in.archive.ubuntu.com/ubuntu focal-updates/main amd64 c-n-f Metadata [14.4 kB]
Get:14 http://in.archive.ubuntu.com/ubuntu focal-updates/universe amd64 Packages [864 kB]
Get:15 http://in.archive.ubuntu.com/ubuntu focal-updates/universe i386 Packages [641 kB]
Get:16 http://security.ubuntu.com/ubuntu focal-security/universe i386 Packages [510 kB]
Get:17 http://in.archive.ubuntu.com/ubuntu focal-updates/universe Translation-en [184 kB]
Get:18 http://in.archive.ubuntu.com/ubuntu focal-updates/universe amd64 DEP-11 Metadata [361 kB]
Get:19 http://in.archive.ubuntu.com/ubuntu focal-updates/multiverse amd64 DEP-11 Metadata [944 B]
Get:20 http://in.archive.ubuntu.com/ubuntu focal-backports/universe amd64 DEP-11 Metadata [10.4 kB]
Get:21 http://security.ubuntu.com/ubuntu focal-security/universe amd64 Packages [641 kB]
Get:22 http://security.ubuntu.com/ubuntu focal-security/universe Translation-en [102 kB]
Get:23 http://security.ubuntu.com/ubuntu focal-security/universe amd64 DEP-11 Metadata [62.7 kB]
Get:24 http://security.ubuntu.com/ubuntu focal-security/multiverse amd64 DEP-11 Metadata [2,464 B]
Fetched 5,864 kB in 23s (259 kB/s)
Reading package lists... Done
Building dependency tree
Reading state information... Done
86 packages can be upgraded. Run 'apt list --upgradable' to see them.
```

2. **Sudo apt install openssh-server -y**

```
priya_hadoop@priya-VB:~$ sudo apt install openssh-server -y
```

3. **Sudo apt install openssh-client -y**

```
priya_hadoop@priya-VB:~$ sudo apt install openssh-client -y
```

4. **Sudo adduser hadoop**

```
priya_hadoop@priya-VB:~$ sudo adduser hadoop
Adding user `hadoop' ...
Adding new group `hadoop' (1001) ...
Adding new user `hadoop' (1001) with group `hadoop' ...
Creating home directory `/home/hadoop' ...
Copying files from `/etc/skel' ...
New password:
Retype new password:
passwd: password updated successfully
Changing the user information for hadoop
Enter the new value, or press ENTER for the default
        Full Name []: hadoop
        Room Number []:
        Work Phone []:
        Home Phone []:
        Other []:
Is the information correct? [Y/n] y
```

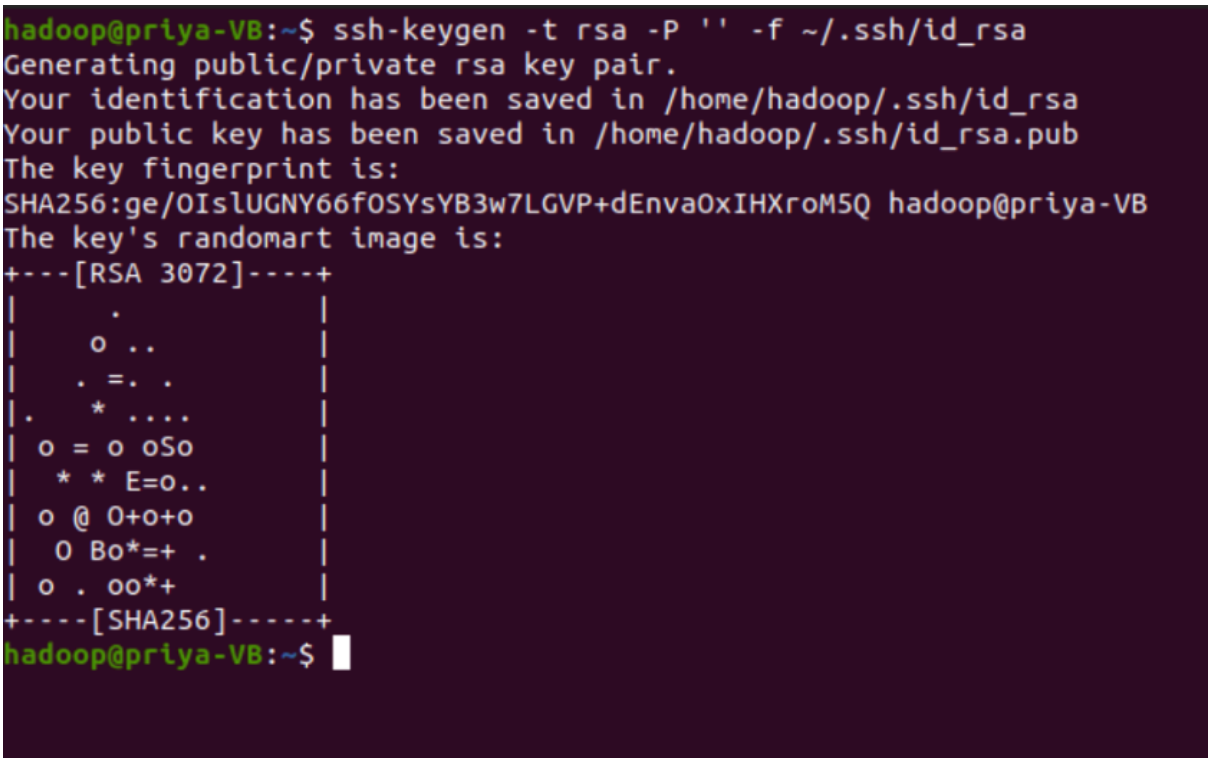
5. **Sudo usermod -aG sudo hadoop**

```
priya_hadoop@priya-VB:~$ sudo usermod -aG sudo hadoop
```

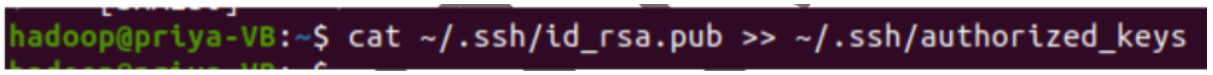
6. **Sudo su - hadoop**

```
priya_hadoop@priya-VB:~$ sudo su - hadoop
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.
```

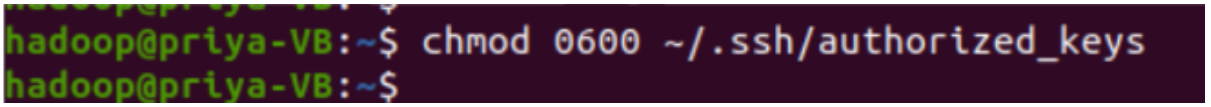
7. Now is you give command ssh localhost, it will as for password
   Type Exit
   To make passwordless, we can do the following steps:
   **Ssh-keygen -t rsa -P ‘’ -f ~/.ssh/id_rsa**




   **Cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys**




   **Chmod 0600 ~/.ssh/authorized_keys**




   Now if you write this command **ssh localhost**, it will not ask for password
   Type exit

8. Download hadoop
   Wget
   https://dlcdn.apache.org/hadoop/common/hadoop-3.2.2/hadoop-3.2.2.tar.gz
   Extract the files now:
   Tar xzf hadoop-3.2.2.tar.gz
   Now move to hadoop directory
   **cd home/hadoop/hadoop-3.2.2**

```
hadoop@priya-VB:~$ cd /home/hadoop/hadoop-3.2.2
hadoop@priya-VB:~/hadoop-3.2.2$
```

9. Make changes in bashrc file
   **Sudo nano ~/.bashrc** #open the file in the nano editor

```
hadoop@priya-VB:~/hadoop-3.2.2$ sudo nano  ~/.bashrc
```

   **export HADOOP_HOME=/home/hadoop/hadoop-3.2.2**
   **export HADOOP_INSTALL=$HADOOP_HOME**
   **export HADOOP_MAPRED_HOME=$HADOOP_HOME**
   **export HADOOP_COMMON_HOME=$HADOOP_HOME**
   **export HADOOP_HDFS_HOME=$HADOOP_HOME**
   **export YARN_HOME=$HADOOP_HOME**
   **export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native**
   **export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin**
   **export HADOOP_OPTS=”-Djava.library.path=$HADOOP_HOME/lib/native”**

```
# Hadoop settings
export HADOOP_HOME=/home/hadoop/hadoop-3.2.2
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

Save the above changes in bashrc file and execute the following commands:

**Sudo ~/.bashrc**

```
hadoop@priya-VB:~$ source ~/.bashrc
```

10. Get path of **JAVA_HOME** and set that path in **hadoop-env.sh** file

**Cd /home/hadoop/hadoop-3.2.2**

```
hadoop@priya-VB:~$ cd /home/hadoop/hadoop-3.2.2
hadoop@priya-VB:~/hadoop-3.2.2$
```

Get path of java home directory

**Which java**

```
hadoop@priya-VB:~/hadoop-3.2.2$ which javac
/usr/bin/javac
hadoop@priya-VB:~/hadoop-3.2.2$
```

**Readlink -f /usr/bin/javac**

```
hadoop@priya-VB:~/hadoop-3.2.2$ readlink -f /usr/bin/javac
/usr/lib/jvm/java-8-openjdk-amd64/bin/javac
hadoop@priya-VB:~/hadoop-3.2.2$
```

**/usr/lib/jvm/java-8-openjdk-amd64** will be set as path of **JAVA_HOME** in **hadoop-env.sh** file

```
hadoop@priya-VB:~/hadoop-3.2.2$ sudo nano etc/hadoop/hadoop-env.sh
hadoop@priya-VB:~/hadoop-3.2.2$
```

updated **JAVA_HOME** in **hadoop-env.sh** file

```
# The java implementation to use. By default, this environment
# variable is REQUIRED on ALL platforms except OS X!
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

11. Make changes in **core-site.xml** file

Make a directory

**Mkdir tmpdata**

```
hadoop@priya-VB:~/hadoop-3.2.2$ mkdir tmpdata
hadoop@priya-VB:~/hadoop-3.2.2$
```

Add the following in **core-site.xml** file

**<configuration>**
**<property>**
**<name>hadoop.tmp.dir</name>**
**<value>/home/hadoop/hadoop-3.2.2/tmpdata</value>**
**<description>A base for other temporary directories.</description>**
**</property>**
**<property>**
**<name>fs.default.name</name>**
**<value>hdfs://localhost:9000</value>**
**<description></description>**
**</property>**
**</configuration>**

```
<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
        <name>hadoop.tmp.dir</name>
        <value>/home/hadoop/hadoop-3.2.2/tmpdata</value>
        <description>A base for other temporary directories.</description>
</property>
<property>
        <name>fs.default.name</name>
        <value>hdfs://localhost:9000</value>
        <description></description>
</property>
</configuration>
```

**12.** Make changes in hdfs-site.xml

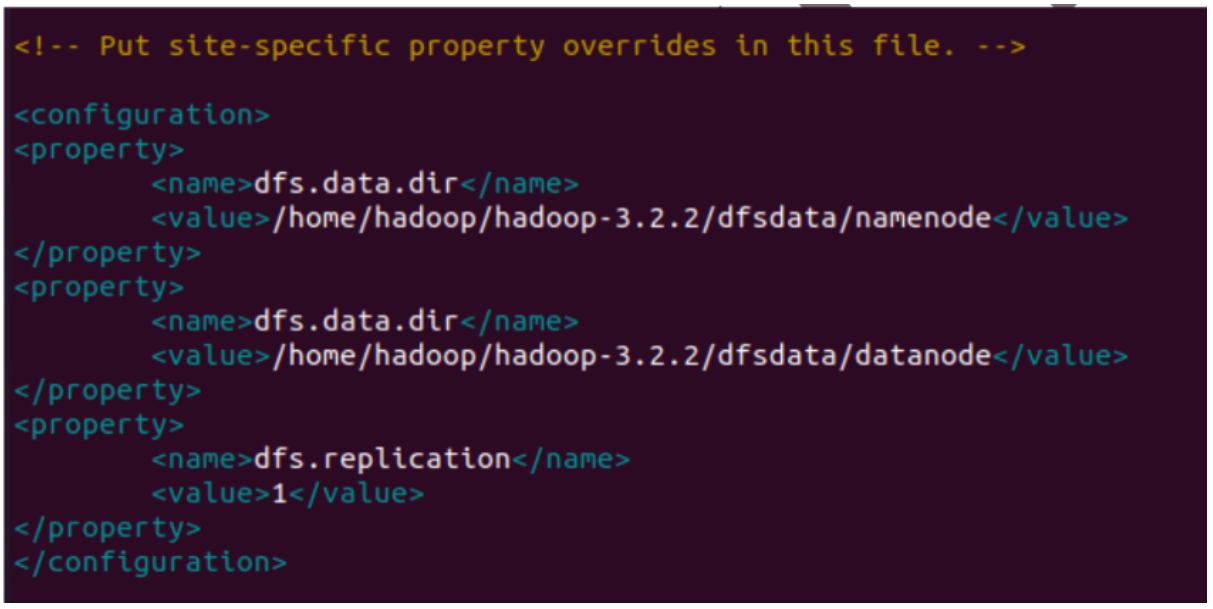
Make two directories

**mkdir -p dfsdata/namenode**
**mkdir -p dfsdata/datanode**

```
hadoop@priya-VB:~/hadoop-3.2.2$ mkdir -p dfsdata/namenode
hadoop@priya-VB:~/hadoop-3.2.2$
hadoop@priya-VB:~/hadoop-3.2.2$ mkdir -p dfsdata/datanode
hadoop@priya-VB:~/hadoop-3.2.2$
```

Add the following in the **hdfs-site.xml** file

**<configuration>**
**<property>**
**<name>dfs.data.dir</name>**
**<value>/home/hadoop/hadoop-3.2.2/dfsdata/namenode</value>**
**</property>**
**<property>**
**<name>dfs.data.dir</name>**
**<value>/home/hadoop/hadoop-3.2.2/dfsdata/datanode</value>**
**</property>**
**<property>**
**<name>dfs.replication</name>**
**<value>1</value>**
**</property>**
**</configuration>**




**13.** Make changes in **mapred-site.xml** file

**<configuration>**
**<property>**
**<name>mapreduce.framework.name</name>**
**<value>yarn</value>**
**</property>**
**</configuration>**

```
<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>mapreduce.framework.name</name>
<value>yarn</value>
</property>
</configuration>
```

**14.** Make changes in **yarn-site.xml** file

**<configuration>**
**<property>**
**<name>yarn.nodemanager.aux-services</name>**
**<value>mapreduce_shuffle</value>**
**</property>**
**<property>**
**<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>**
**<value>org.apache.hadoop.mapred.ShuffleHandler</value>**
**</property>**
**<property>**
**<name>yarn.resourcemanager.hostname</name>**
**<value>127.0.0.1</value>**
**</property>**
**<property>**
**<name>yarn.acl.enable</name>**
**<value>0</value>**
**</property>**
**<property>**
**<name>yarn.nodemanager.env-whitelist</name>**
**<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_HOME,PATH,LANG,TZ,HADOOP_MAPRED_HOME</value>**
**</property>**
**</configuration>**

```
<configuration>

<!-- Site specific YARN configuration properties -->

<property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
</property>
<property>
        <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
        <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
<property>
        <name>yarn.resourcemanager.hostname</name>
        <value>127.0.0.1</value>
</property>
<property>
        <name>yarn.acl.enable</name>
        <value>0</value>
</property>
<property>
        <name>yarn.nodemanager.env-whitelist</name>
        <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_HO>
</property>

</configuration>
```

**15.** Format HDFS NAMENODE. This will shutdown namenode

**Hdfs namenode -format**

```
hadoop@priya-VB:~/hadoop-3.2.2$ hdfs namenode -format
WARNING: /home/hadoop/hadoop-3.2.2/logs does not exist. Creating.
2021-10-05 12:55:14,552 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = priya-VB/127.0.1.1
STARTUP_MSG:   args = [-format]
```

```
atted.
2021-10-05 12:55:18,108 INFO namenode.FSImageFormatProtobuf: Saving image file /home/hadoop/hadoop-3.2.2/
image.ckpt_0000000000000000000 using no compression
2021-10-05 12:55:18,389 INFO namenode.FSImageFormatProtobuf: Image file /home/hadoop/hadoop-3.2.2/tmpdata
kpt_0000000000000000000 of size 401 bytes saved in 0 seconds .
2021-10-05 12:55:18,480 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2021-10-05 12:55:18,510 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2021-10-05 12:55:18,512 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at priya-VB/127.0.1.1
************************************************************/
hadoop@priya-VB:~/hadoop-3.2.2$
```

**16.** Start hadoop cluster

**sbin/start-dfs.sh**

```
hadoop@priya-VB:~/hadoop-3.2.2$ sbin/start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [priya-VB]
priya-VB: Warning: Permanently added 'priya-vb' (ECDSA) to the list of known hosts.
hadoop@priya-VB:~/hadoop-3.2.2$
```

**17.** Start YARN resource and node manager

**sbin/start-yarn.sh**

```
hadoop@priya-VB:~/hadoop-3.2.2$ sbin/start-yarn.sh
Starting resourcemanager
Starting nodemanagers
hadoop@priya-VB:~/hadoop-3.2.2$
```

**18.** Check if all daemons are active and running

**jps**

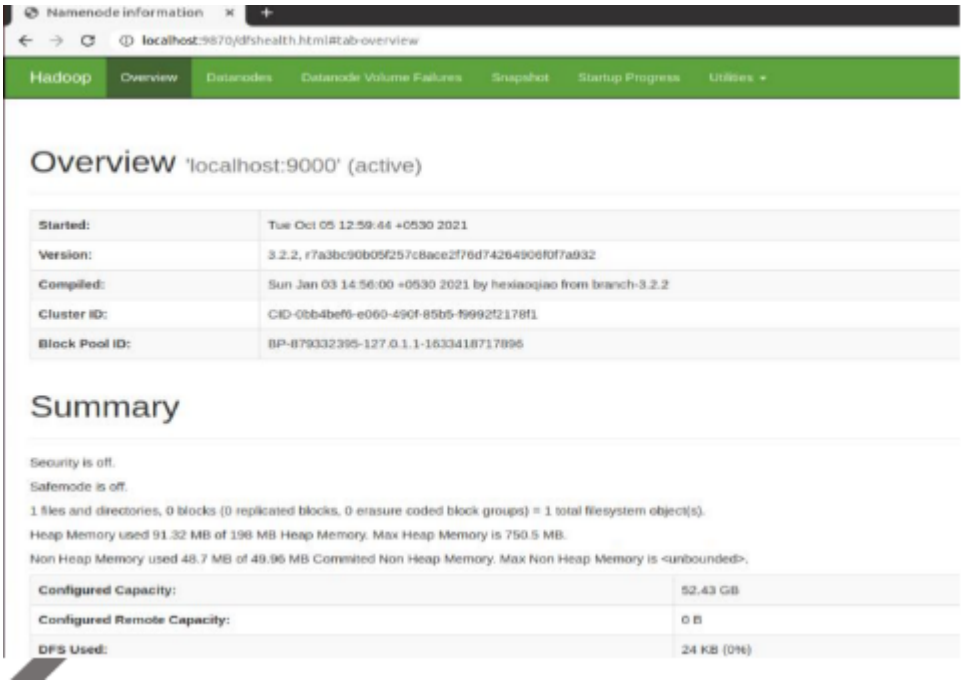
```
hadoop@priya-VB:~/hadoop-3.2.2$ jps
8339 SecondaryNameNode
8166 DataNode
9270 Jps
8039 NameNode
8711 ResourceManager
8841 NodeManager
hadoop@priya-VB:~/hadoop-3.2.2$
```

**19.** Verification

Namenode interface

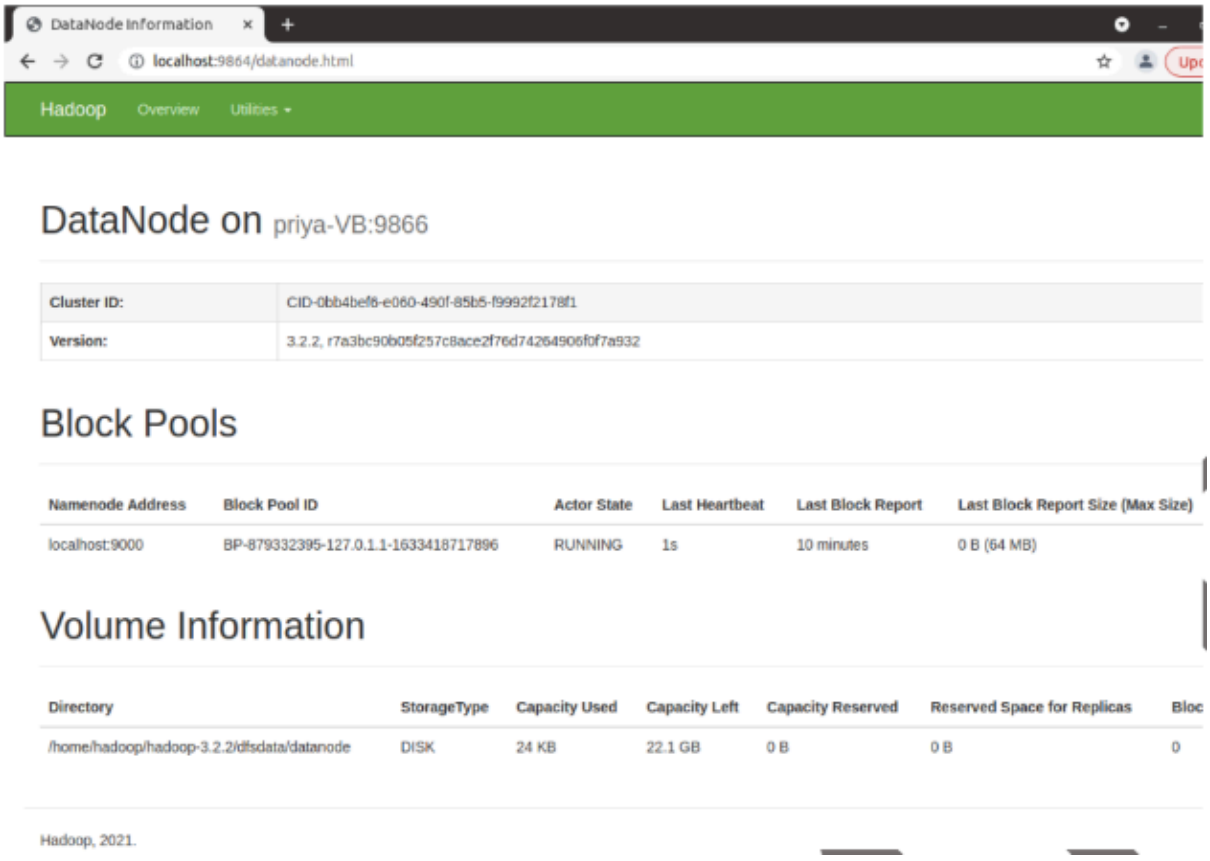
**http://localhost:9870**

Open in UI



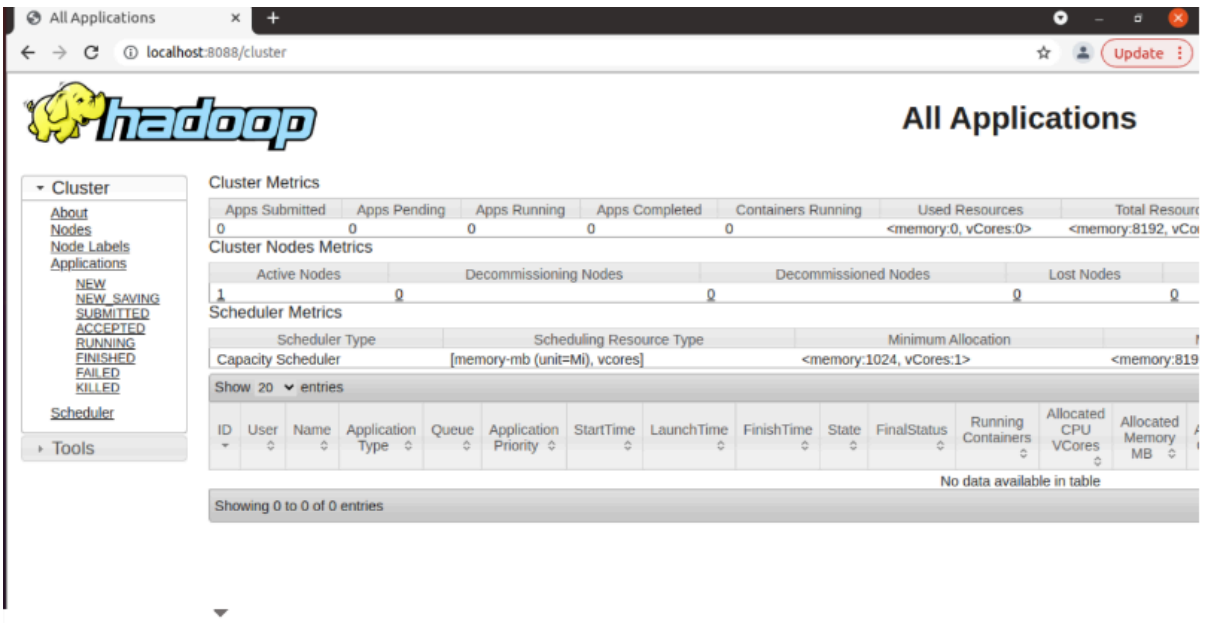
Individual Datanode

**http://localhost:9864**




Yarn Resource Manager

**http://localhost:8088**



Start Hadoop services by using the command
**start-dfs.sh**
**start-yarn.sh**
Now type **jps**

***JPS - Java Virtual Machine Process Status tool is a type of command that is implemented to check out all the Hadoop daemons like DataNode, NodeManager, NameNode, and ResourceManager that are currently running on the machine.JPS command is used to check if a specific daemon is up or not. The command of JPS displays all the processes that are based on Java for a particular user. The command of JPS should run from the root to check all the operating nodes in the host.***

**Q2) Demonstrate HDFS commands in Hadoop**

1. **start-all.sh:** The start-all.sh command is commonly used in Hadoop environments to start all essential services, including HDFS (NameNode, DataNode) and YARN (ResourceManager, NodeManager).

```
kratos@kratos-VirtualBox:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as kratos in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [kratos-VirtualBox]
Starting resourcemanager
Starting nodemanagers
kratos@kratos-VirtualBox:~$
```

2. **Jps:** The jps command is used to check if Java processes (like NameNode, DataNode, ResourceManager, etc.) are running.

```
kratos@kratos-VirtualBox:~$ jps
14882 Jps
13764 NameNode
13910 DataNode
14363 ResourceManager
14157 SecondaryNameNode
14510 NodeManager
kratos@kratos-VirtualBox:~$
```

3. **Hdfs dfs -ls:** The hdfs dfs -ls command is used to list the contents (files and directories) in a specified HDFS directory, similar to the ls command in Linux.

```
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /
Found 6 items
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /demo
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:13 /temp
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /test
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:15 /test1
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:13 /testng
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:13 /usr
kratos@kratos-VirtualBox:~$ S
```

# Browse Directory

/ Go!

Show 25 entries Search:

| Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name |
|---|---|---|---|---|---|---|---|
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:14 | 0 | 0 B | demo |
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:19 | 0 | 0 B | emp123 |
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:13 | 0 | 0 B | temp |
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:14 | 0 | 0 B | test |
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:15 | 0 | 0 B | test1 |
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:13 | 0 | 0 B | testng |
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:13 | 0 | 0 B | usr |

Showing 1 to 7 of 7 entries Previous 1 Next

```
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /test
Found 4 items
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /test/MCA
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /test/employee
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /test/stud
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /test/student123
kratos@kratos-VirtualBox:~$
```

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

# Browse Directory

/test Go!

Show 25 entries Search:

| Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name |
|---|---|---|---|---|---|---|---|
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:14 | 0 | 0 B | MCA |
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:14 | 0 | 0 B | employee |
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:14 | 0 | 0 B | stud |
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:14 | 0 | 0 B | student123 |

Showing 1 to 4 of 4 entries Previous 1 Next

4. **Hdfs dfs -ls -R /:** The command hdfs dfs -ls -R / recursively lists all files and directories in the root directory (/) of the Hadoop Distributed File System (HDFS).

```
kratos@kratos-VirtualBox:~$ hdfs dfs -ls -R /
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /demo
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /demo/faculty
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:13 /temp
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /test
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /test/MCA
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /test/employee
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /test/stud
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /test/student123
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:15 /test1
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:15 /test1/BVIMIT
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:13 /testng
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:13 /usr
kratos@kratos-VirtualBox:~$
```

5. **Hdfs dfs -mkdir:** The hdfs dfs -mkdir command is used to create directories in the Hadoop Distributed File System (HDFS).

```
kratos@kratos-VirtualBox:~$ hdfs dfs -mkdir /emp123
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /
Found 7 items
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /demo
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:18 /emp123
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:13 /temp
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /test
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:15 /test1
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:13 /testng
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:13 /usr
kratos@kratos-VirtualBox:~$
```

# Browse Directory

/ Go!

Show 25 entries Search:

| Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name |
|---|---|---|---|---|---|---|---|
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:14 | 0 | 0 B | demo |
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:19 | 0 | 0 B | emp123 |
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:13 | 0 | 0 B | temp |
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:14 | 0 | 0 B | test |
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:15 | 0 | 0 B | test1 |
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:13 | 0 | 0 B | testng |
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:13 | 0 | 0 B | usr |

Showing 1 to 7 of 7 entries Previous 1 Next

```
kratos@kratos-VirtualBox:~$ hdfs dfs -mkdir /emp123/empdetails
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /emp123
Found 1 items
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:19 /emp123/empdetails
kratos@kratos-VirtualBox:~$
```

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

# Browse Directory

/emp123 Go!

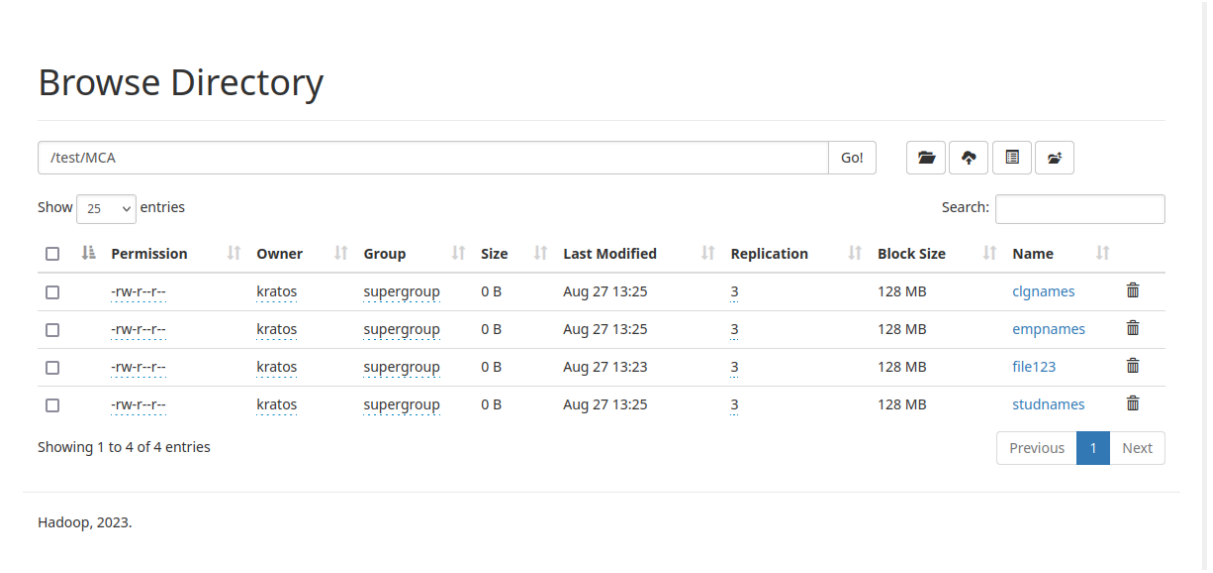
Show 25 entries Search:

| Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name |
|---|---|---|---|---|---|---|---|
| drwxr-xr-x | kratos | supergroup | 0 B | Aug 27 13:19 | 0 | 0 B | empdetails |

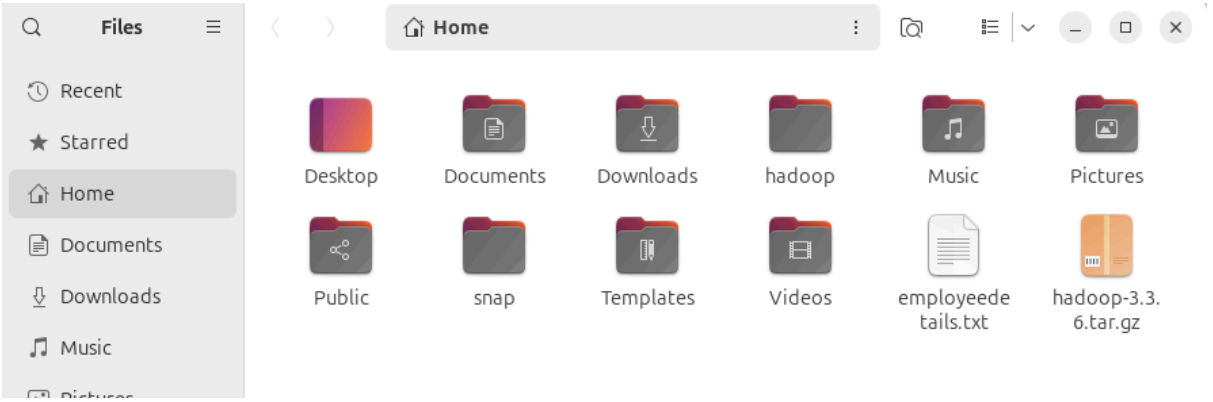
Showing 1 to 1 of 1 entries Previous 1 Next

6. **Hdfs dfs -touchz:** The hdfs dfs -touchz command is used to create an empty file in the Hadoop Distributed File System (HDFS). The file created has a size of 0 bytes.

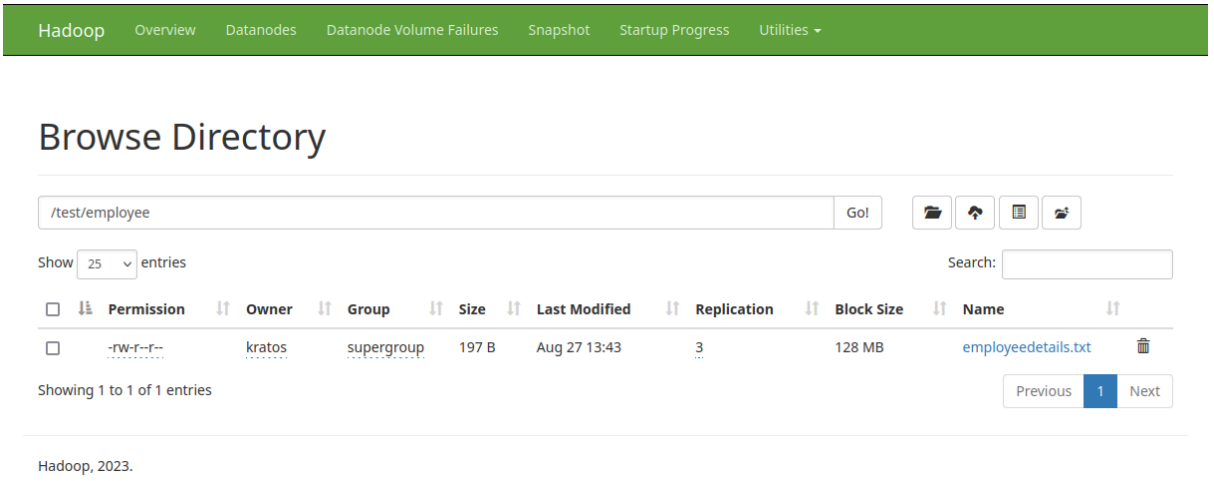
```
kratos@kratos-VirtualBox:~$ hdfs dfs -touchz /test/MCA/file123
kratos@kratos-VirtualBox:~$ hdfs dfs -touchz /test/MCA/empnames
kratos@kratos-VirtualBox:~$ hdfs dfs -touchz /test/MCA/studnames
kratos@kratos-VirtualBox:~$ hdfs dfs -touchz /test/MCA/clgnames
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /test/MCA/
Found 4 items
-rw-r--r--   3 kratos supergroup          0 2024-08-27 13:25 /test/MCA/clgnames
-rw-r--r--   3 kratos supergroup          0 2024-08-27 13:25 /test/MCA/empnames
-rw-r--r--   3 kratos supergroup          0 2024-08-27 13:23 /test/MCA/file123
-rw-r--r--   3 kratos supergroup          0 2024-08-27 13:25 /test/MCA/studnames
kratos@kratos-VirtualBox:~$
```




7. **Hdfs dfs -put:** The hdfs dfs -put command is used to copy files from the local file system to the Hadoop Distributed File System (HDFS).



```
kratos@kratos-VirtualBox:~$ hdfs dfs -put employeedetails.txt /test/employee
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /test/employee
Found 1 items
-rw-r--r--   3 kratos supergroup        197 2024-08-27 13:43 /test/employee/employeedetails.txt
kratos@kratos-VirtualBox:~$
```



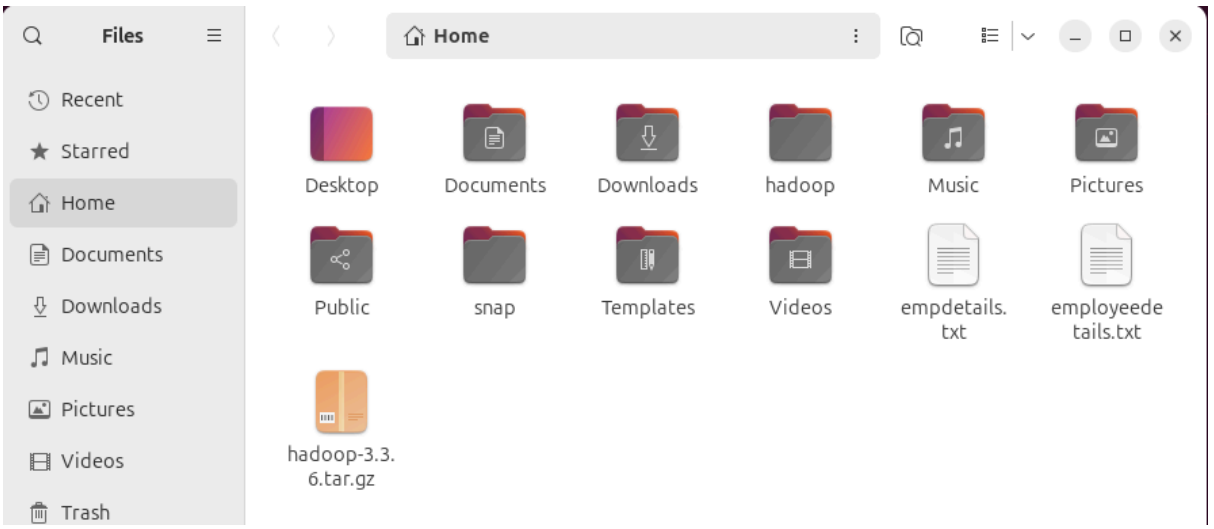

8. **Hdfs dfs -cat:** The hdfs dfs -cat command is used to display the contents of a file stored in the Hadoop Distributed File System (HDFS). It works similarly to the cat command in Linux for reading and outputting file content to the console.

```
kratos@kratos-VirtualBox:~$ hdfs dfs -cat /test/employee/employeedetails.txt
Name: Vinayak, Salary: 19000
Name: Divya, Salary: 23000
Name: Deepika, Salary: 21000
Name: Purva, Salary: 12000
Name: Sairaj, Salary: 4000
Name: Kavya, Salary: 53000
Name: Samiksha, Salary: 18000

kratos@kratos-VirtualBox:~$
```

9. **Hdfs dfs -get:** The hdfs dfs -get command is used to copy files or directories from the Hadoop Distributed File System (HDFS) to the local file system. It's similar to the -put command, but in the reverse direction (from HDFS to local).

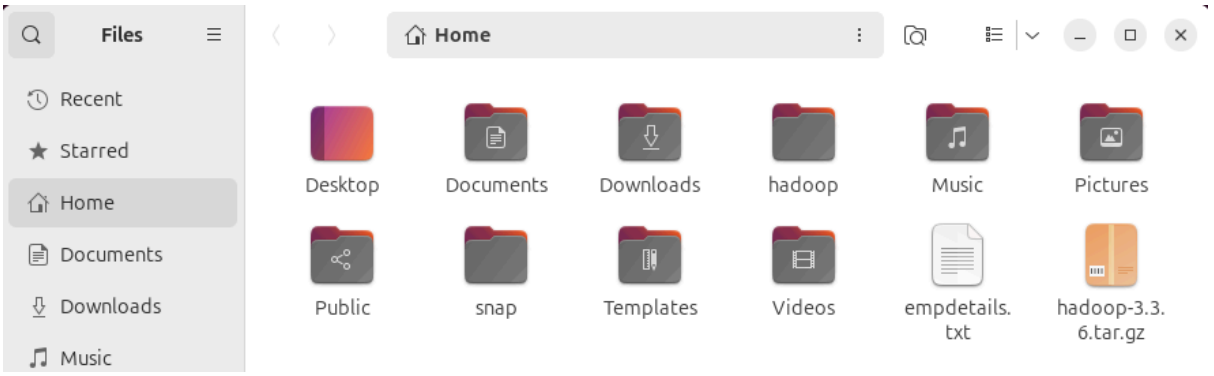
```
kratos@kratos-VirtualBox:~$ hdfs dfs -get /test/employee/employeedetails.txt empdetails.txt
kratos@kratos-VirtualBox:~$ ls
Desktop    Downloads       employeedetails.txt  hadoop-3.3.6.tar.gz  Pictures  snap       Videos
Documents  empdetails.txt  hadoop               Music                Public    Templates
kratos@kratos-VirtualBox:~$
```

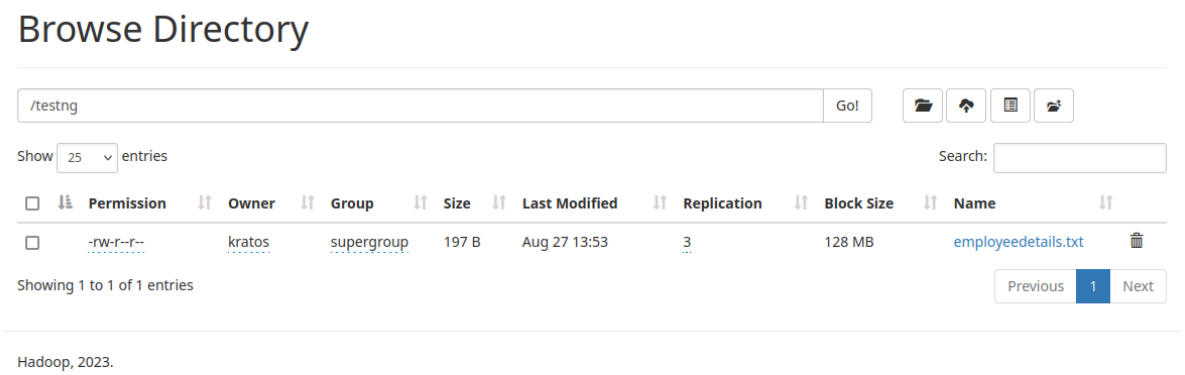



10. **Hdfs dfs -moveFromLocal:** The hdfs dfs -moveFromLocal command is used to move files or directories from the local file system to the Hadoop Distributed File System (HDFS). It's similar to the -put command but with one key difference: after copying the file to HDFS, the original file in the local file system is deleted.

```
kratos@kratos-VirtualBox:~$ hdfs dfs -moveFromLocal employeedetails.txt /testng employee.txt
```

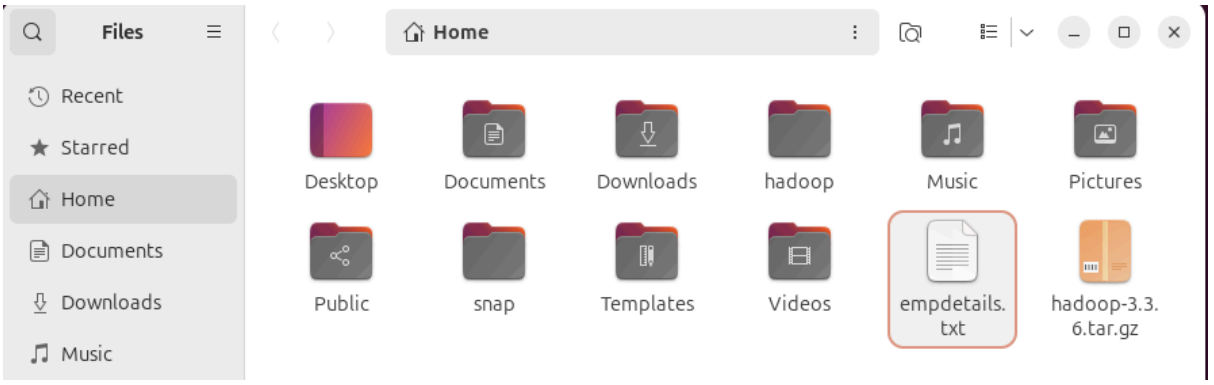
```
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /testng
Found 1 items
-rw-r--r--   3 kratos supergroup        197 2024-08-27 13:53 /testng/employeedetails.txt
kratos@kratos-VirtualBox:~$
```

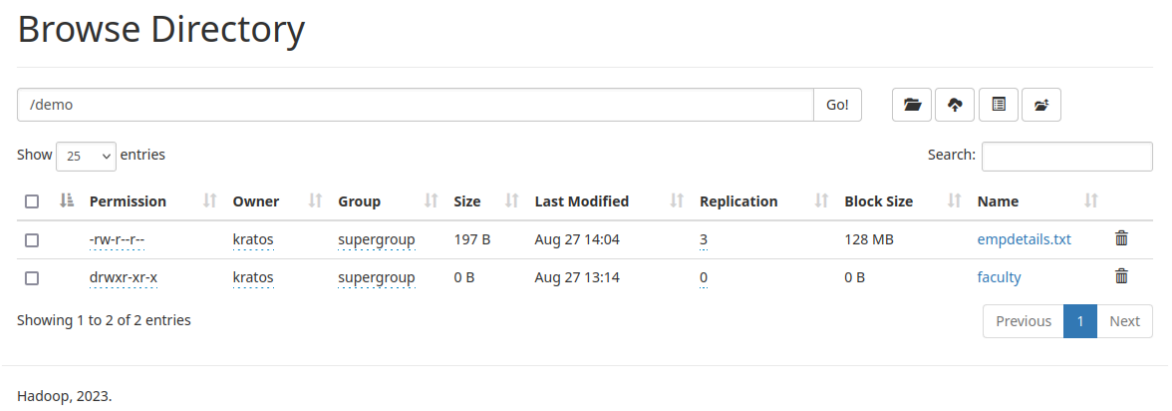





11. **Hdfs dfs -copyFromLocal:** The hdfs dfs -copyFromLocal command is used to copy files or directories from the local file system to the Hadoop Distributed File System (HDFS). Unlike -put, -copyFromLocal does not delete the original files from the local file system after copying them.

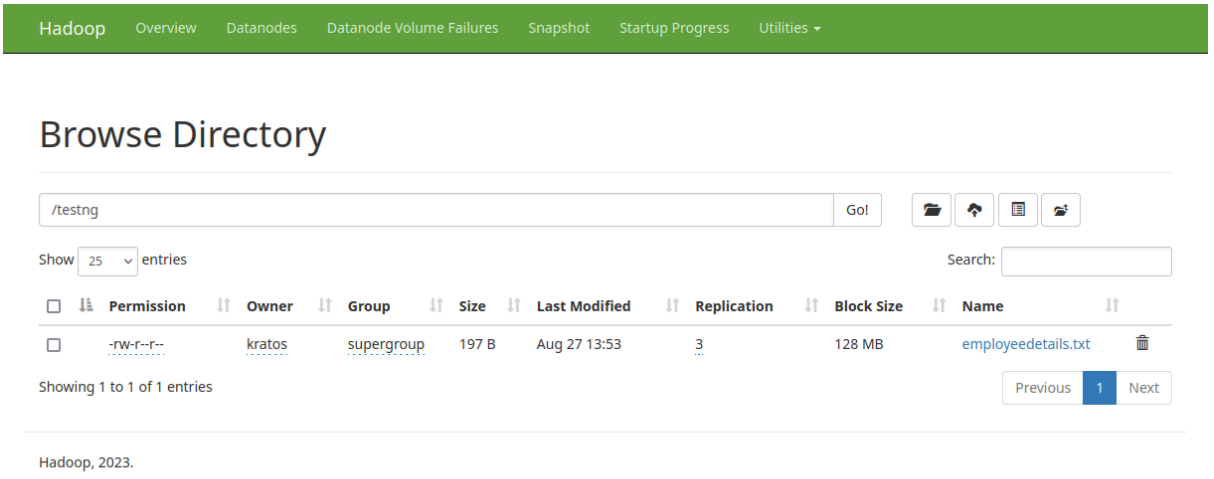
```
kratos@kratos-VirtualBox:~$ hdfs dfs -copyFromLocal empdetails.txt /demo
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /demo
Found 2 items
-rw-r--r--   3 kratos supergroup        197 2024-08-27 14:04 /demo/empdetails.txt
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /demo/faculty
kratos@kratos-VirtualBox:~$
```

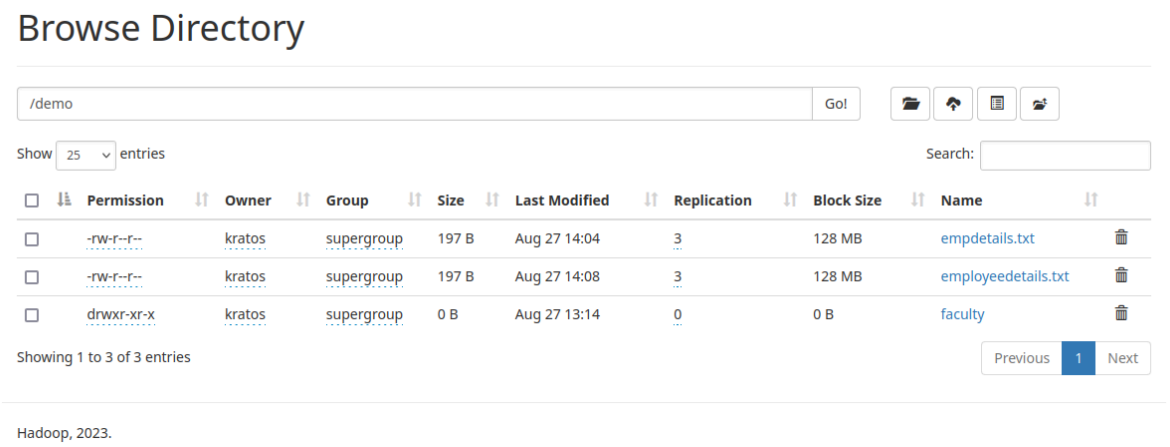





12. **Hdfs dfs -cp:** The hdfs dfs -cp command is used to copy files or directories within the Hadoop Distributed File System (HDFS). This command allows you to create a duplicate of a file or directory at a different location within HDFS.

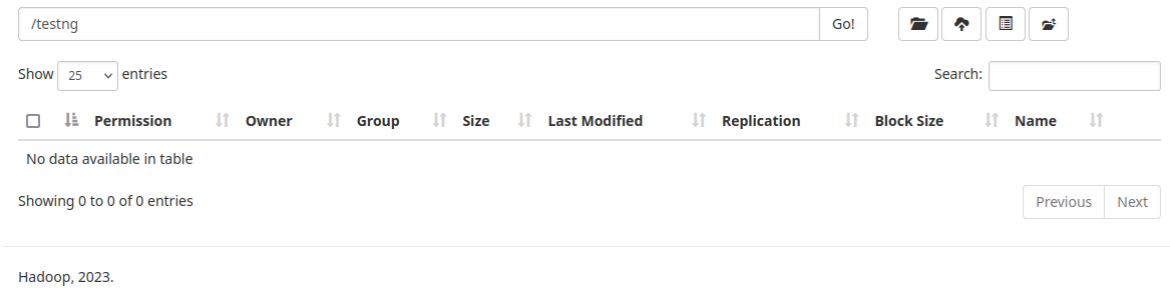
```
kratos@kratos-VirtualBox:~$ hdfs dfs -cp /testng/employeedetails.txt /demo
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /testng
Found 1 items
-rw-r--r--   3 kratos supergroup        197 2024-08-27 13:53 /testng/employeedetails.txt
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /demo
Found 3 items
-rw-r--r--   3 kratos supergroup        197 2024-08-27 14:04 /demo/empdetails.txt
-rw-r--r--   3 kratos supergroup        197 2024-08-27 14:08 /demo/employeedetails.txt
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /demo/faculty
kratos@kratos-VirtualBox:~$
```

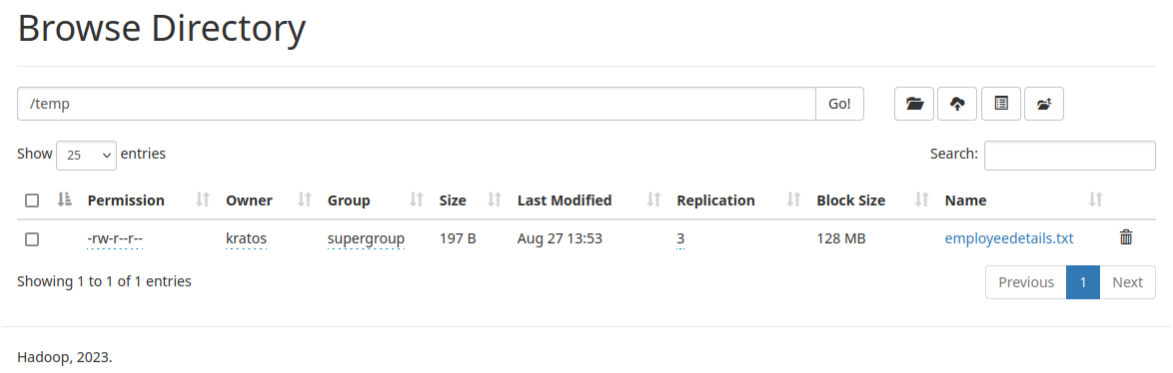





13. **Hdfs dfs -mv:** The hdfs dfs -mv command is used to move or rename files and directories within the Hadoop Distributed File System (HDFS). This command allows you to change the location or name of a file or directory within HDFS.

```
kratos@kratos-VirtualBox:~$ hdfs dfs -mv /testng/employeedetails.txt /temp
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /testng
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /temp
Found 1 items
-rw-r--r--   3 kratos supergroup        197 2024-08-27 13:53 /temp/employeedetails.txt
kratos@kratos-VirtualBox:~$
```

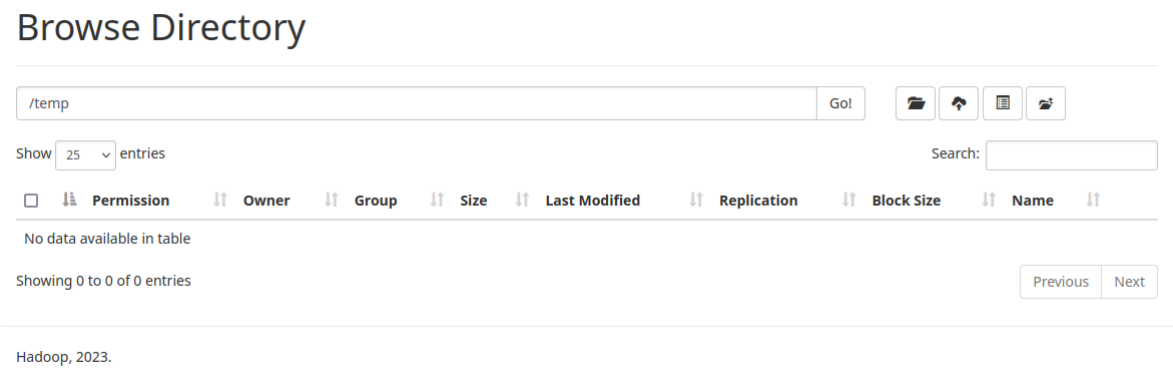





14. **Hdfs dfs -rm:** The hdfs dfs -rm command is used to delete files or directories from the Hadoop Distributed File System (HDFS). This command removes the specified files or directories, which can be useful for managing and cleaning up HDFS storage

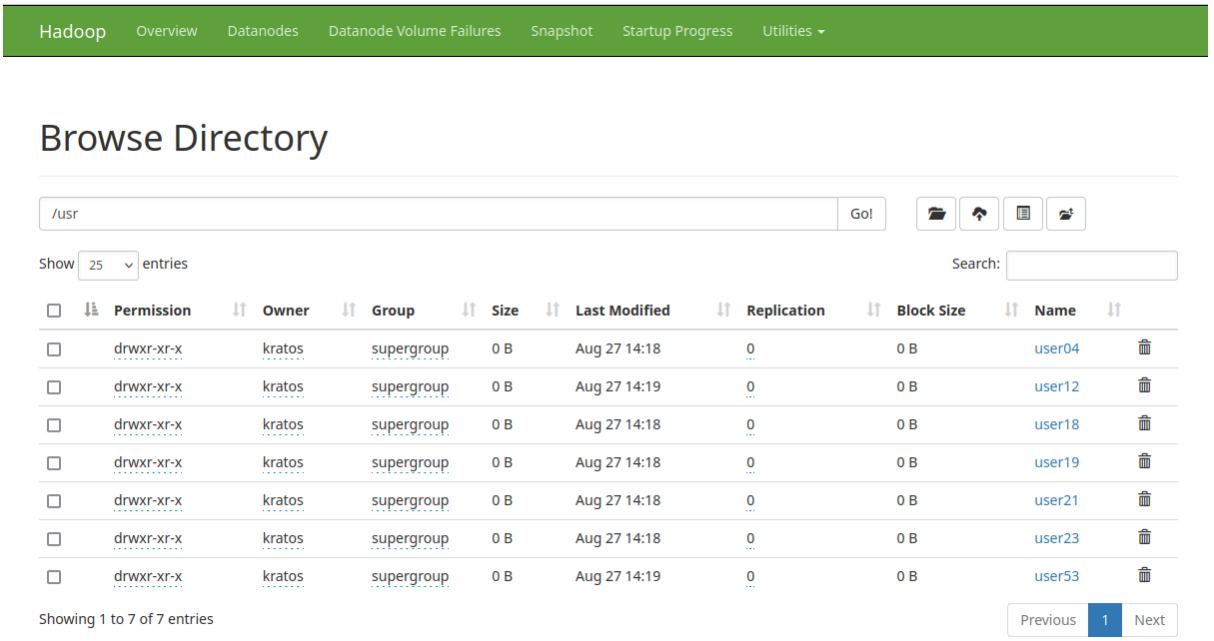
```
kratos@kratos-VirtualBox:~$ hdfs dfs -rm /temp/employeedetails.txt
Deleted /temp/employeedetails.txt
```

```
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /temp
kratos@kratos-VirtualBox:~$
```

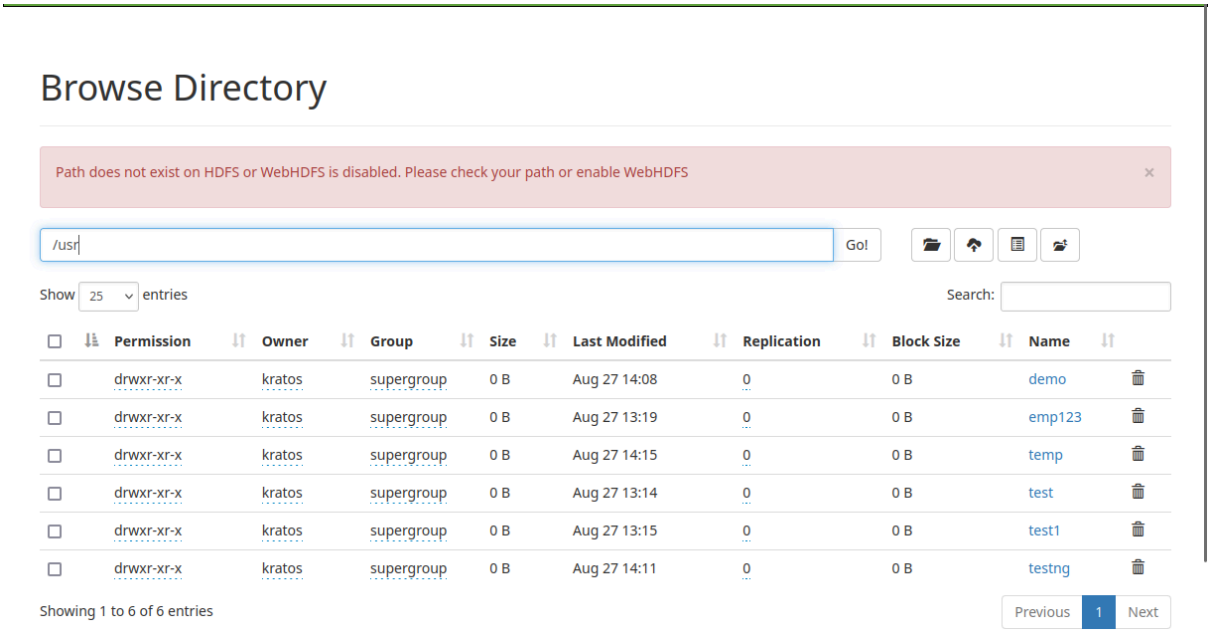


15. **Hdfs dfs -rm -r:** The hdfs dfs -rm -r command is used to recursively delete directories and their contents from the Hadoop Distributed File System (HDFS). This command is useful for removing entire directories, including all files and subdirectories within them.

```
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /usr
Found 7 items
drwxr-xr-x   - kratos supergroup          0 2024-08-27 14:18 /usr/user04
drwxr-xr-x   - kratos supergroup          0 2024-08-27 14:19 /usr/user12
drwxr-xr-x   - kratos supergroup          0 2024-08-27 14:18 /usr/user18
drwxr-xr-x   - kratos supergroup          0 2024-08-27 14:18 /usr/user19
drwxr-xr-x   - kratos supergroup          0 2024-08-27 14:18 /usr/user21
drwxr-xr-x   - kratos supergroup          0 2024-08-27 14:18 /usr/user23
drwxr-xr-x   - kratos supergroup          0 2024-08-27 14:19 /usr/user53
kratos@kratos-VirtualBox:~$
```




```
kratos@kratos-VirtualBox:~$ hdfs dfs -rm -r /usr
Deleted /usr
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /
Found 6 items
drwxr-xr-x   - kratos supergroup          0 2024-08-27 14:08 /demo
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:19 /emp123
drwxr-xr-x   - kratos supergroup          0 2024-08-27 14:15 /temp
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:14 /test
drwxr-xr-x   - kratos supergroup          0 2024-08-27 13:15 /test1
drwxr-xr-x   - kratos supergroup          0 2024-08-27 14:11 /testng
kratos@kratos-VirtualBox:~$ hdfs dfs -ls /usr
ls: `/usr': No such file or directory
kratos@kratos-VirtualBox:~$
```




16. **Hdfs dfs -du:** The hdfs dfs -du command is used to display the disk usage of files and directories within the Hadoop Distributed File System (HDFS). It provides information on the amount of space used by files and directories in HDFS.

```
kratos@kratos-VirtualBox:~$ hdfs dfs -du /demo
197  591  /demo/empdetails.txt
197  591  /demo/employeedetails.txt
0    0    /demo/faculty
kratos@kratos-VirtualBox:~$
```

17. **Hdfs dfs -du -s:** The hdfs dfs -du -s command is used to display the summarized disk usage of files and directories within the Hadoop Distributed File System (HDFS). Unlike the basic -du command, which shows the disk usage of each file and directory individually, the -s (summary) option provides a single total size for each specified path.

```
kratos@kratos-VirtualBox:~$ hdfs dfs -du -s /demo
394  1182  /demo
kratos@kratos-VirtualBox:~$
```

18. **Hdfs dfs stat:** The hdfs dfs -stat command provides detailed information about files and directories in the Hadoop Distributed File System (HDFS). It allows you to retrieve various metadata attributes for HDFS files or directories.