

SMIT R PATEL

19162121031

SEM 5

BDA

PRACTICAL 3

AIM- Understanding HDFS commands with respect to Name Nodes, Data Nodes and working of HDFS.

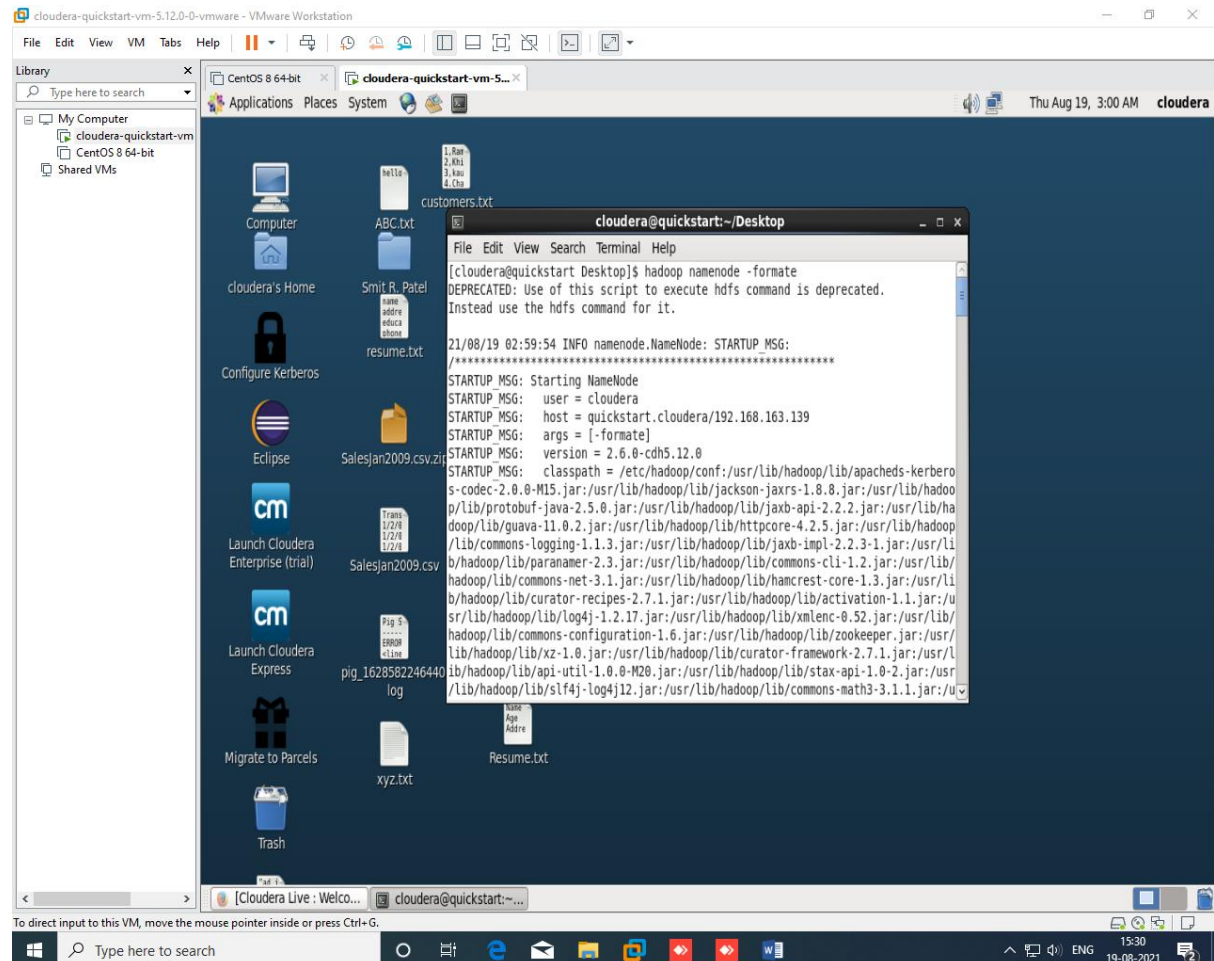
On understanding how HDFS works in reality, you as a student wants to test out a few operations on your

Hadoop system. Hence, perform the following tasks:

1. Starting HDFS:

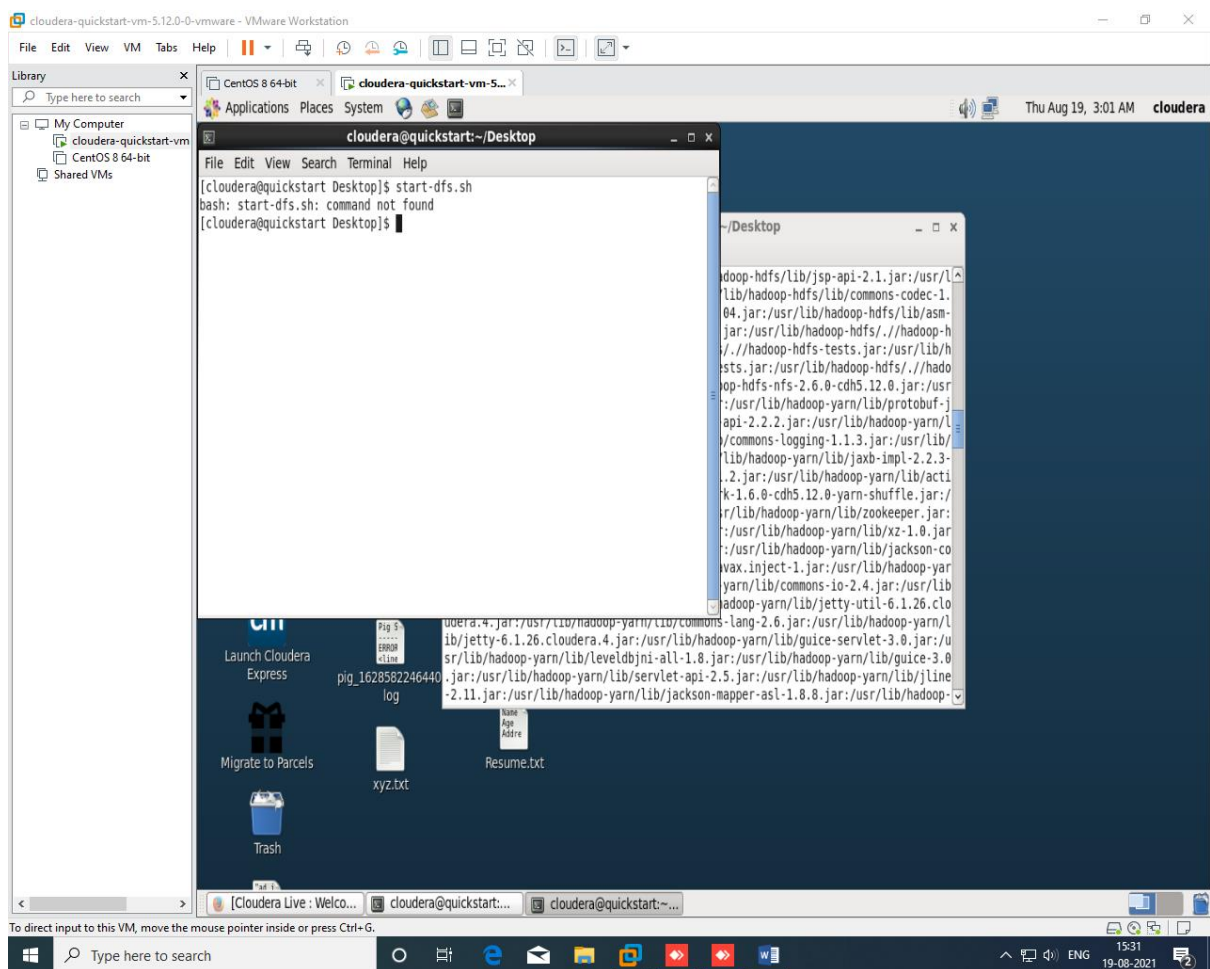
a. Format the configured HDFS file system and then open the namenode (HDFS server) and execute the following command.

`$ hadoop namenode -format`



b. Start the distributed file system and follow the command listed below to start the namenode as well as the data nodes in cluster.

\$ start-dfs.sh

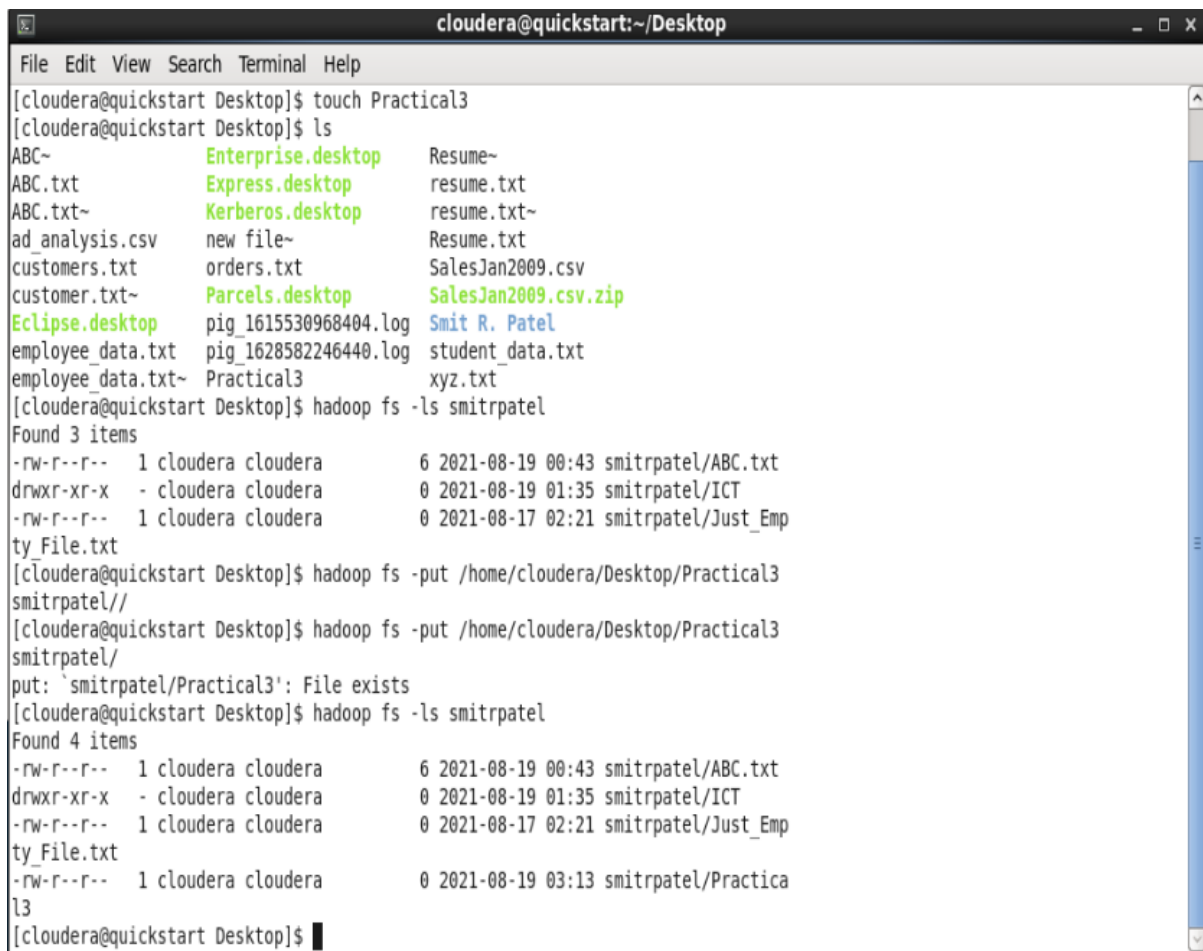


Here, we don't have rights so it will give an error.

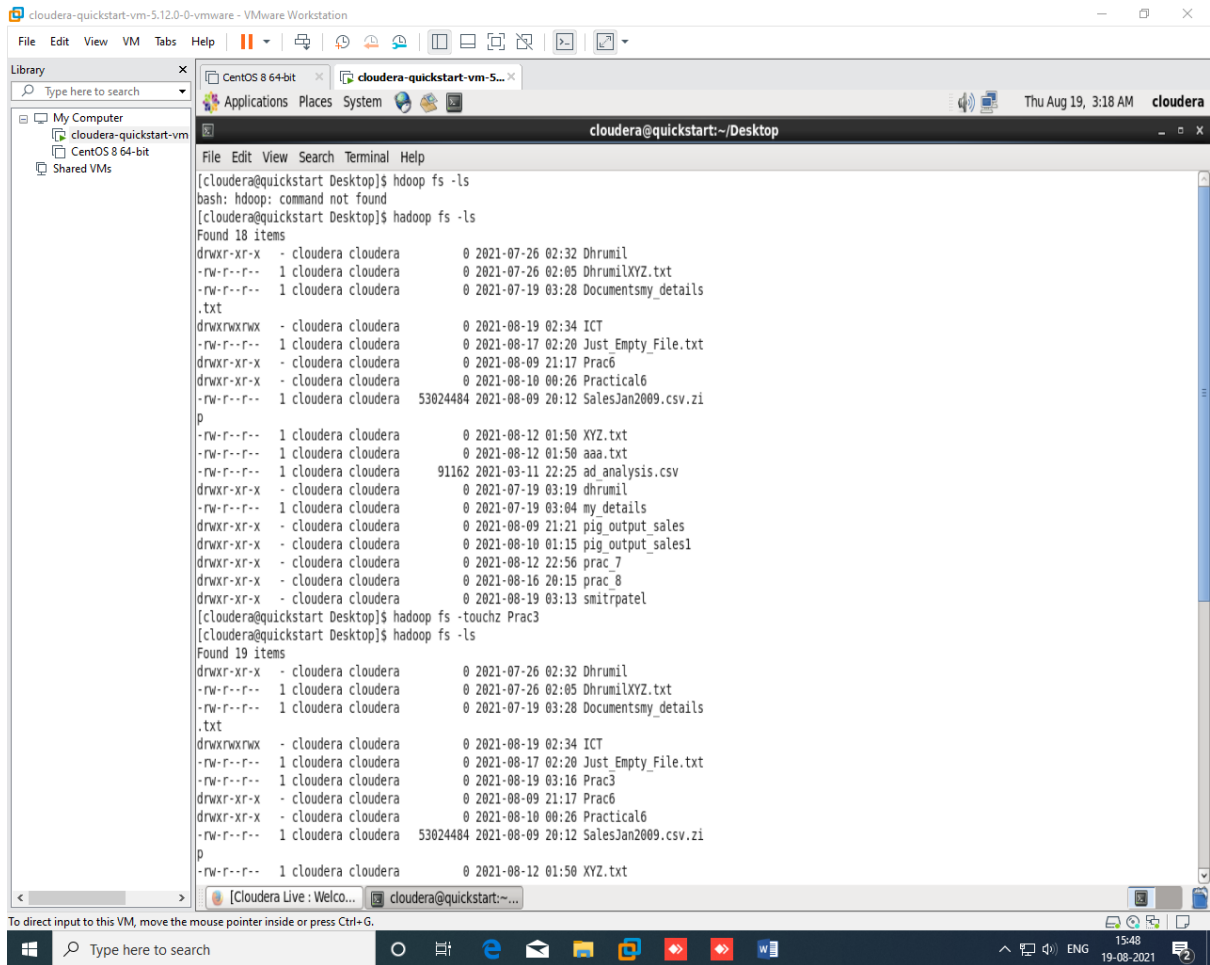
2. Alternatives to the copyFromLocal and copyToLocal commands:

You also have alternatives to the copyFromLocal and copyToLocal commands. Try the GET and PUT command as similar to what you did for copyFromLocal and copyToLocal. Create a file called "Practical

3" in your local system to PUT into the hadoop system, and a file called "Prac3" to GET from the Hadoop system and verify that the files moved into these places.



```
cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
[cloudera@quickstart Desktop]$ touch Practical3
[cloudera@quickstart Desktop]$ ls
ABC~ Enterprise.desktop Resume~
ABC.txt Express.desktop resume.txt
ABC.txt~ Kerberos.desktop resume.txt~
ad_analysis.csv new file~ Resume.txt
customers.txt orders.txt SalesJan2009.csv
customer.txt~ Parcels.desktop SalesJan2009.csv.zip
Eclipse.desktop pig_1615530968404.log Smit R. Patel
employee_data.txt pig_1628582246440.log student_data.txt
employee_data.txt~ Practical3 xyz.txt
[cloudera@quickstart Desktop]$ hadoop fs -ls smitrpatel
Found 3 items
-rw-r--r-- 1 cloudera cloudera 6 2021-08-19 00:43 smitrpatel/ABC.txt
drwxr-xr-x - cloudera cloudera 0 2021-08-19 01:35 smitrpatel/ICT
-rw-r--r-- 1 cloudera cloudera 0 2021-08-17 02:21 smitrpatel/Just_Emp
ty_File.txt
[cloudera@quickstart Desktop]$ hadoop fs -put /home/cloudera/Desktop/Practical3
smitrpatel/
[cloudera@quickstart Desktop]$ hadoop fs -put /home/cloudera/Desktop/Practical3
smitrpatel/
put: 'smitrpatel/Practical3': File exists
[cloudera@quickstart Desktop]$ hadoop fs -ls smitrpatel
Found 4 items
-rw-r--r-- 1 cloudera cloudera 6 2021-08-19 00:43 smitrpatel/ABC.txt
drwxr-xr-x - cloudera cloudera 0 2021-08-19 01:35 smitrpatel/ICT
-rw-r--r-- 1 cloudera cloudera 0 2021-08-17 02:21 smitrpatel/Just_Emp
ty_File.txt
-rw-r--r-- 1 cloudera cloudera 0 2021-08-19 03:13 smitrpatel/Practica
l3
[cloudera@quickstart Desktop]$
```



cloudera-quickstart-vm-5.12.0-0-vmware - VMware Workstation

File Edit View VM Tabs Help

Library

Type here to search

My Computer

cloudera-quickstart-vm

CentOS 8 64-bit

Shared VMs

Applications Places System

cloudera@quickstart:~/Desktop

File Edit View Search Terminal Help

```
[cloudera@quickstart Desktop]$ hadoop fs -ls
Found 19 items
drwxr-xr-x - cloudera cloudera      0 2021-07-26 02:32 Dhrumil
-rw-r--r-- 1 cloudera cloudera      0 2021-07-26 02:05 DhrumiXYZ.txt
-rw-r--r-- 1 cloudera cloudera      0 2021-07-19 03:28 Documentsmy_details
.txt
drwxrwxrwx - cloudera cloudera      0 2021-08-19 02:34 ICT
-rw-r--r-- 1 cloudera cloudera      0 2021-08-17 02:20 Just_Empty_File.txt
-rw-r--r-- 1 cloudera cloudera      0 2021-08-19 03:16 Prac3
drwxr-xr-x - cloudera cloudera      0 2021-08-09 21:17 Prac6
drwxr-xr-x - cloudera cloudera      0 2021-08-10 00:26 Practical6
-rw-r--r-- 1 cloudera cloudera 53024484 2021-08-09 20:12 SalesJan2009.csv.zip
p
-rw-r--r-- 1 cloudera cloudera      0 2021-08-12 01:50 XYZ.txt
-rw-r--r-- 1 cloudera cloudera      0 2021-08-12 01:50 aaa.txt
-rw-r--r-- 1 cloudera cloudera 91162 2021-03-11 22:25 ad_analysis.csv
drwxr-xr-x - cloudera cloudera      0 2021-07-19 03:19 dhrumil
-rw-r--r-- 1 cloudera cloudera      0 2021-07-19 03:04 my_details
drwxr-xr-x - cloudera cloudera      0 2021-08-09 21:21 pig_output_sales
drwxr-xr-x - cloudera cloudera      0 2021-08-10 01:15 pig_output_sales1
drwxr-xr-x - cloudera cloudera      0 2021-08-12 22:56 prac_7
drwxr-xr-x - cloudera cloudera      0 2021-08-16 20:15 prac_8
drwxr-xr-x - cloudera cloudera      0 2021-08-19 03:13 smitratel
[cloudera@quickstart Desktop]$ hadoop fs -get smitratel/Prac3 /home/cloudera/Desktop/
get: 'smitratel/Prac3': No such file or directory
[cloudera@quickstart Desktop]$ hadoop fs -get Prac3 /home/cloudera/Desktop/
[cloudera@quickstart Desktop]$ ls
ABC~           Express.desktop      resume.txt
ABC.txt        Kerberos.desktop    resume.txt~
ABC.txt~       new file~            Resume.txt
ad_analysis.csv orders.txt           SalesJan2009.csv
customers.txt  Parcels.desktop     SalesJan2009.csv.zip
customer.txt~  pig_1615530968404.log Smit R. Patel
Eclipse.desktop pig_1628582246440.log student_data.txt
employee_data.txt Prac3                xyz.txt
employee_data.txt~ Practical3
Enterprise.desktop Resume~
[cloudera@quickstart Desktop]$
```

To direct input to this VM, move the mouse pointer inside or press Ctrl+G.

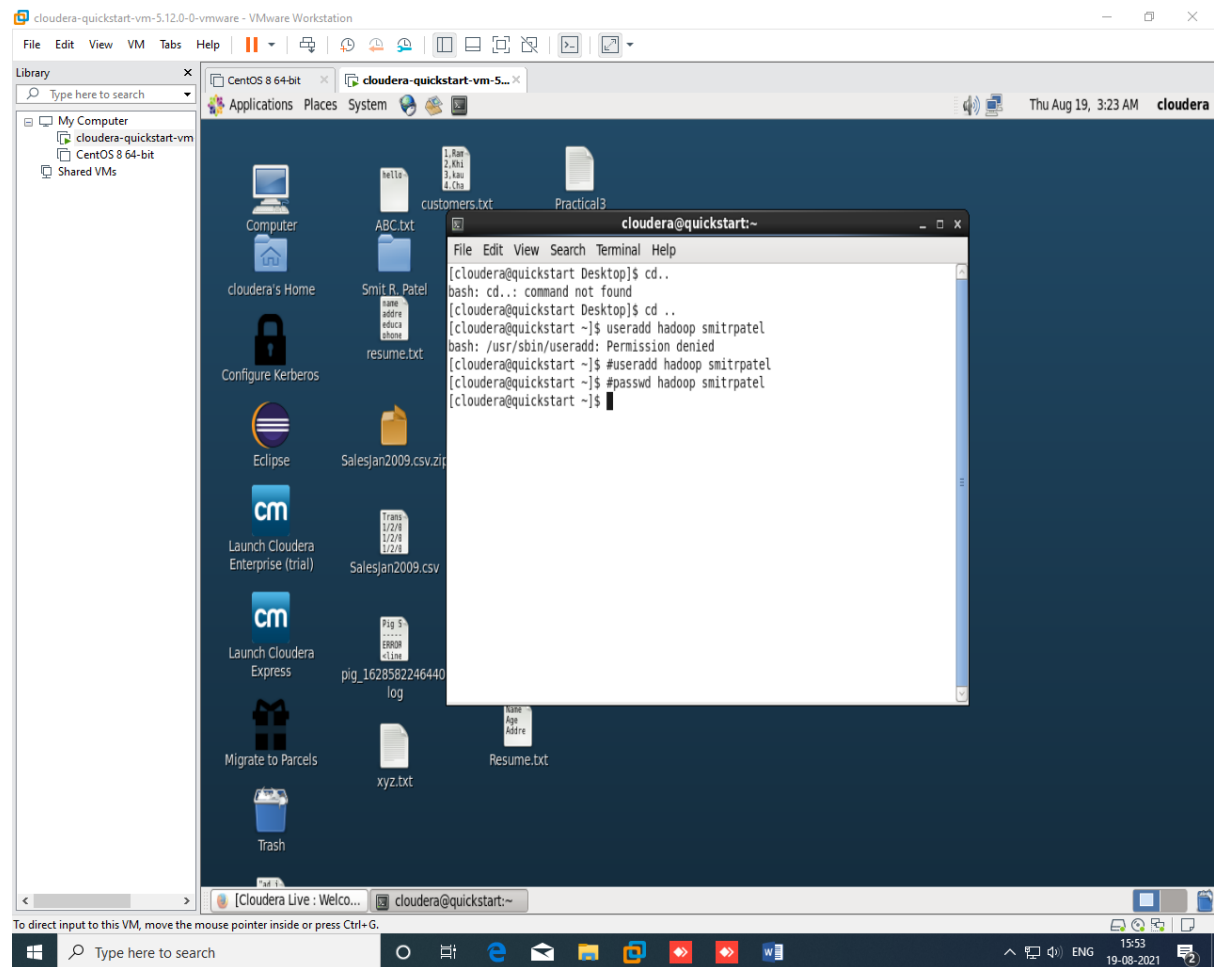
15:48 19-08-2021

3. Create a user on Hadoop:

System user account is used on both master and slave systems for the Hadoop installation. Use:

```
# useradd hadoop
```

```
# passwd Hadoop
```



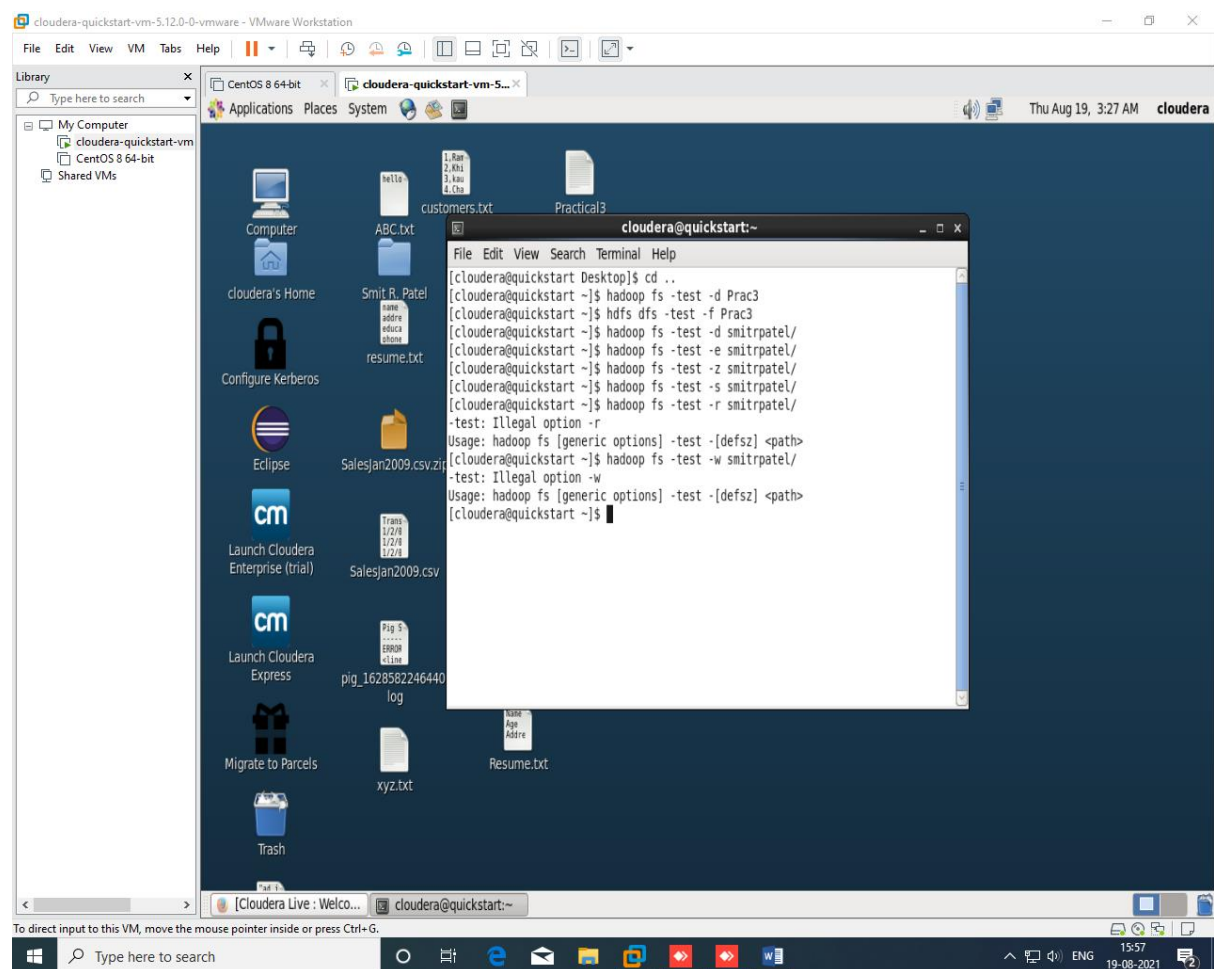
4. Testing the existence of a file:

This particular command fulfills the purpose of testing the existence of a file in the HDFS cluster. The characters from "[defsz]" in the command have to be modified as needed.

`hadoop fs -test -[defsz] <path>`

Here is a brief description of these characters:

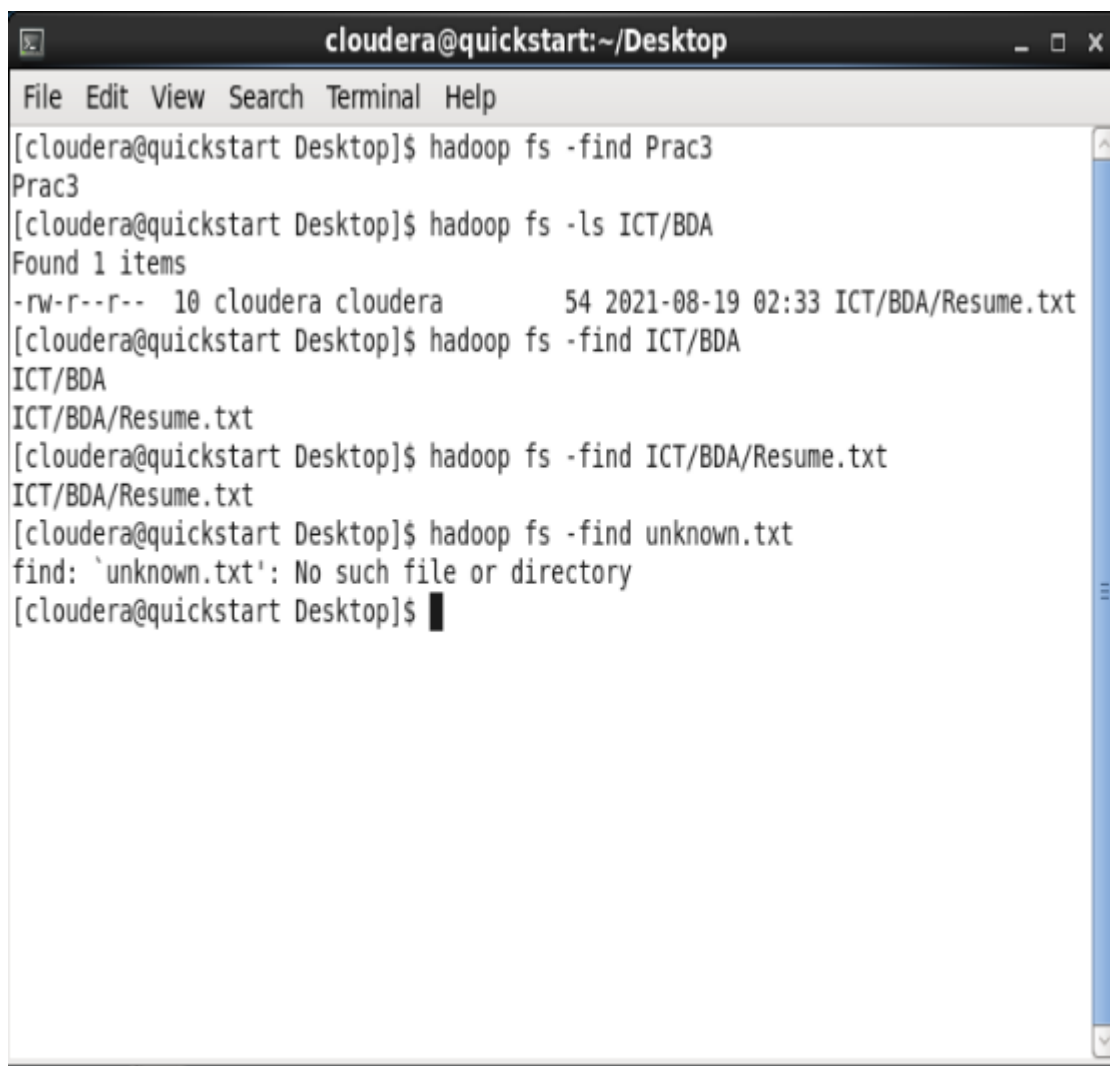
- d -> Checks if it is a directory or not
- e -> Checks if it is a path or not
- f -> Checks if it is a file or not
- s -> Checks if it is an empty path or not
- r -> Checks the path existence and read permission
- w -> Checks the path existence and write permission
- z -> Checks the file size



5. Search for files in the HDFS cluster.

This command is generally used for the purpose to search for files in the HDFS cluster. It scans the given expression in the command with all the files in the cluster, and displays the files that match the defined expression.

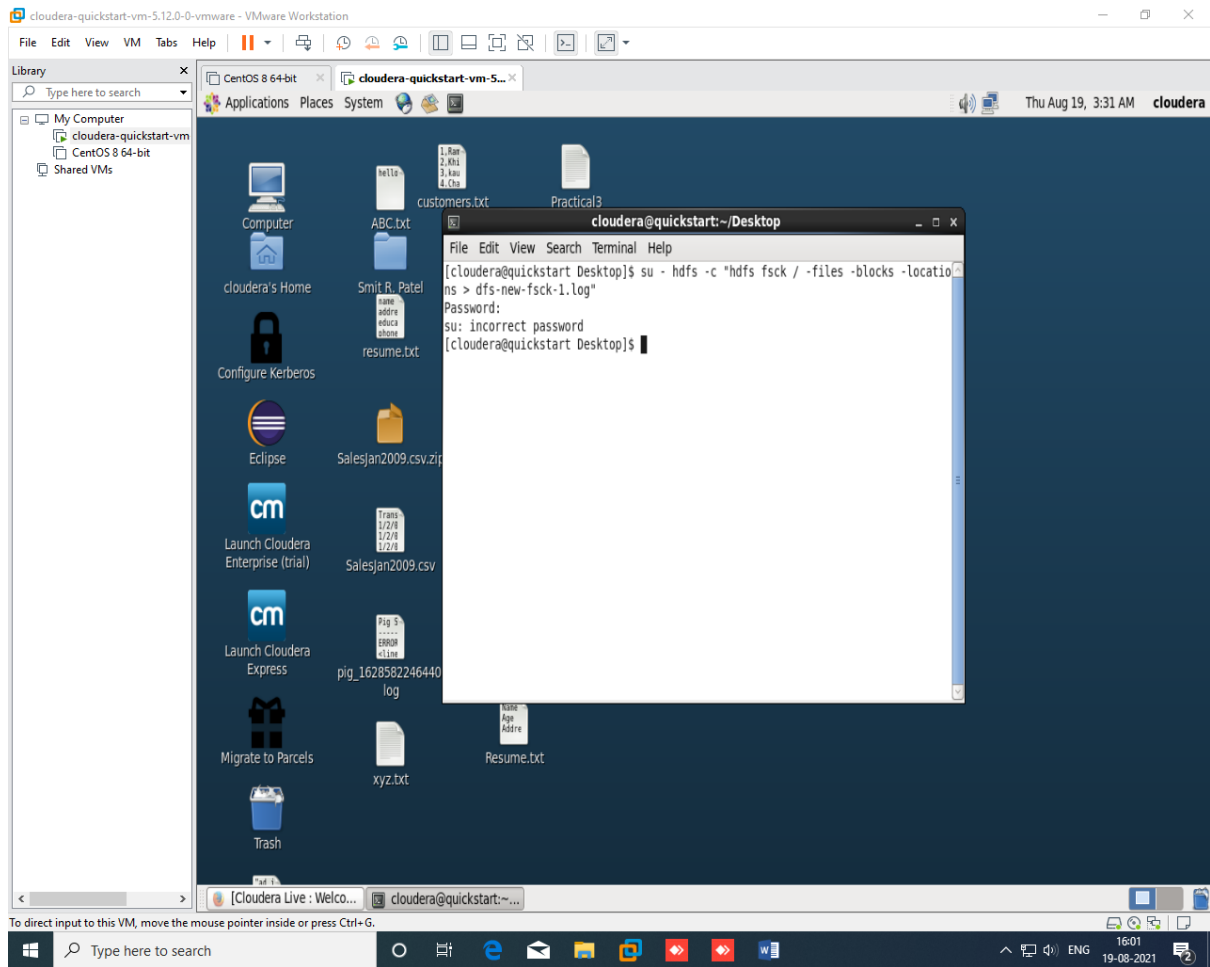
`hadoop fs -find <path> ... <expression>`



```
cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
[cloudera@quickstart Desktop]$ hadoop fs -find Prac3
Prac3
[cloudera@quickstart Desktop]$ hadoop fs -ls ICT/BDA
Found 1 items
-rw-r--r-- 10 cloudera cloudera      54 2021-08-19 02:33 ICT/BDA/Resume.txt
[cloudera@quickstart Desktop]$ hadoop fs -find ICT/BDA
ICT/BDA
ICT/BDA/Resume.txt
[cloudera@quickstart Desktop]$ hadoop fs -find ICT/BDA/Resume.txt
ICT/BDA/Resume.txt
[cloudera@quickstart Desktop]$ hadoop fs -find unknown.txt
find: 'unknown.txt': No such file or directory
[cloudera@quickstart Desktop]$
```

6. Verify HDFS file system health

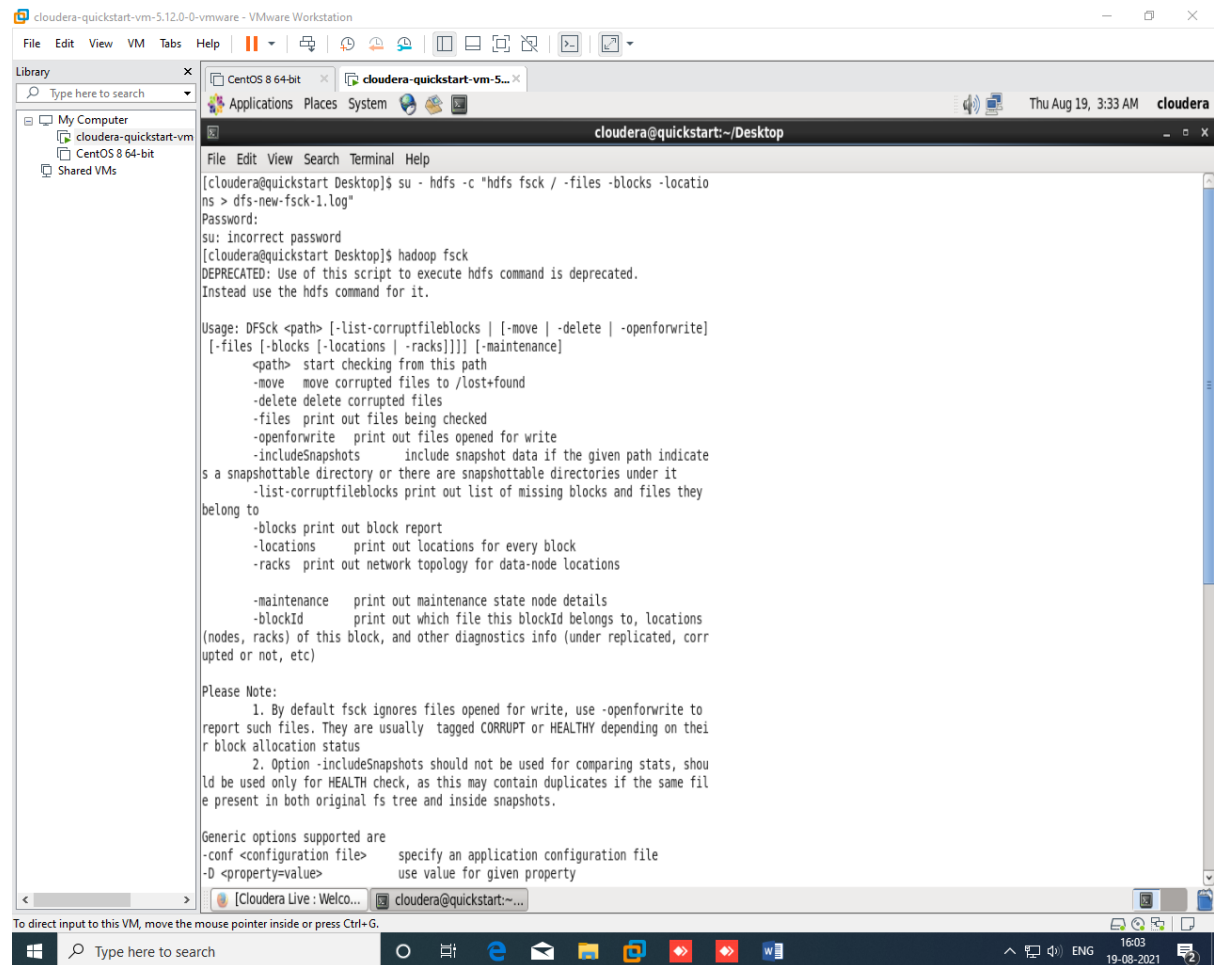
```
su - hdfs -c "hdfs fsck / -files -blocks -locations > dfs-new-fsck-1.log"
```



Here, We don't have rights so it will give an error.

Alternatively, We can check health my following commands without any rights.

Command : Hadoop fsck



```
cloudera-quickstart-vm-5.12.0-0-vmware - VMware Workstation
File Edit View VM Tabs Help
Library
Type here to search
My Computer
cloudera-quickstart-vm
CentOS 8 64-bit
Shared VMs
cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
[cloudera@quickstart Desktop]$ su - hdfs -c "hdfs fsck / -files -blocks -locations"
ns > dfs-new-fsck-1.log"
Password:
su: incorrect password
[cloudera@quickstart Desktop]$ hadoop fsck
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

Usage: DFSck <path> [-list-corruptfileblocks | [-move | -delete | -openforwrite]
[-files [-blocks [-locations | -racks]]]] [-maintenance]
<path> start checking from this path
-move move corrupted files to /lost+found
-delete delete corrupted files
-files print out files being checked
-openforwrite print out files opened for write
-includeSnapshots include snapshot data if the given path indicate
s a snapshottable directory or there are snapshottable directories under it
-list-corruptfileblocks print out list of missing blocks and files they
belong to
-blocks print out block report
-locations print out locations for every block
-racks print out network topology for data-node locations

-maintenance print out maintenance state node details
-blockId print out which file this blockId belongs to, locations
(nodes, racks) of this block, and other diagnostics info (under replicated, corr
upted or not, etc)

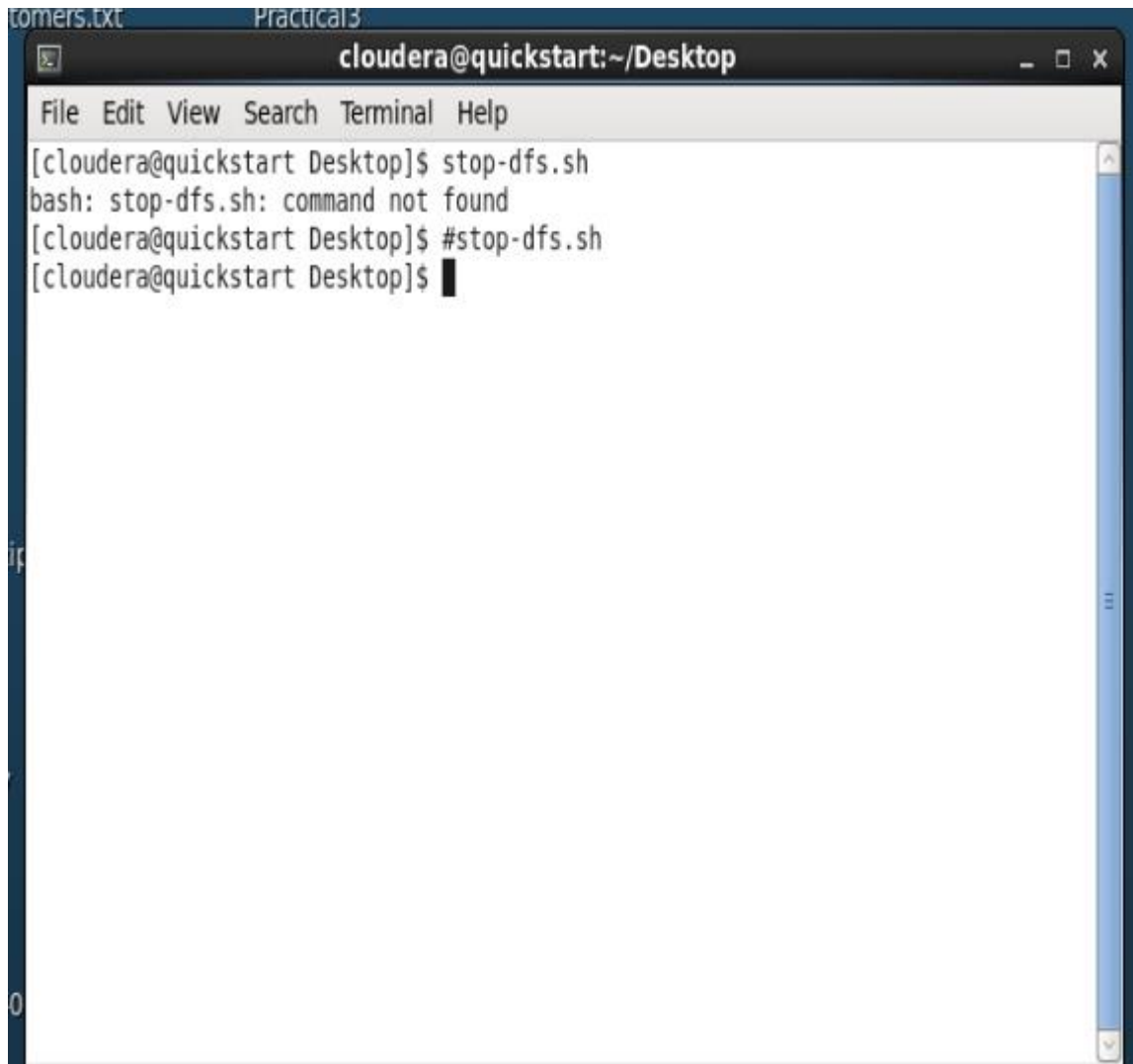
Please Note:
1. By default fsck ignores files opened for write, use -openforwrite to
report such files. They are usually tagged CORRUPT or HEALTHY depending on thei
r block allocation status
2. Option -includeSnapshots should not be used for comparing stats, shou
ld be used only for HEALTH check, as this may contain duplicates if the same fil
e present in both original fs tree and inside snapshots.

Generic options supported are
-conf <configuration file> specify an application configuration file
-D <property=value> use value for given property
```

7. Shutting Down HDFS

Shut down the HDFS files by following the below command:

```
$ stop-dfs.sh
```

A screenshot of a terminal window titled 'cloudera@quickstart:~/Desktop'. The window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The terminal shows the following commands and output:

```
[cloudera@quickstart Desktop]$ stop-dfs.sh
bash: stop-dfs.sh: command not found
[cloudera@quickstart Desktop]$ #stop-dfs.sh
[cloudera@quickstart Desktop]$
```

The cursor is at the end of the last line. The terminal window is part of a larger application with tabs labeled 'tomers.txt' and 'Practical3'.

Here, We don't have rights so it will give an error and nothing we can do now.