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

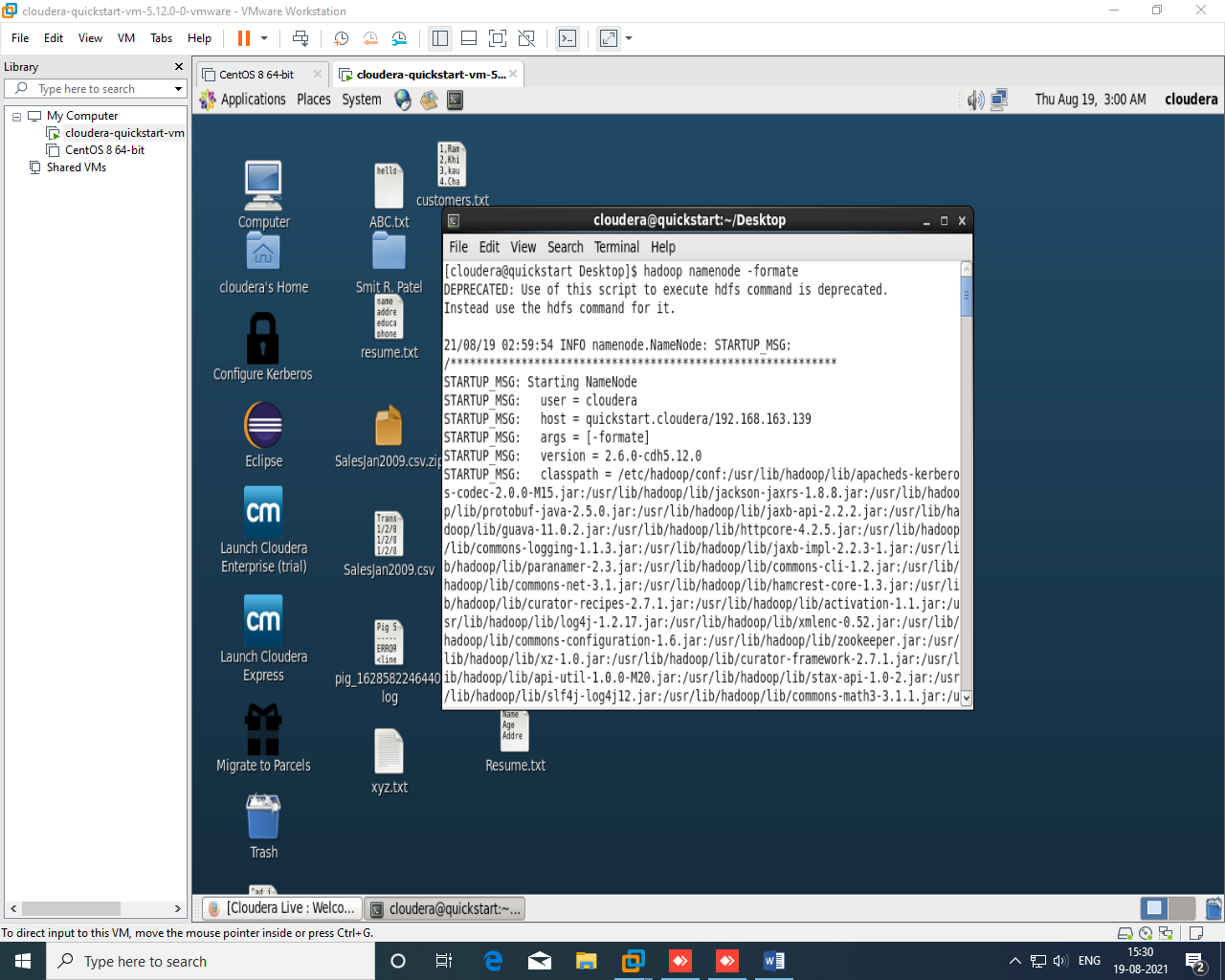
Hadoop system. Hence, perform the following tasks:

1. Starting HDFS:

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

the following command.

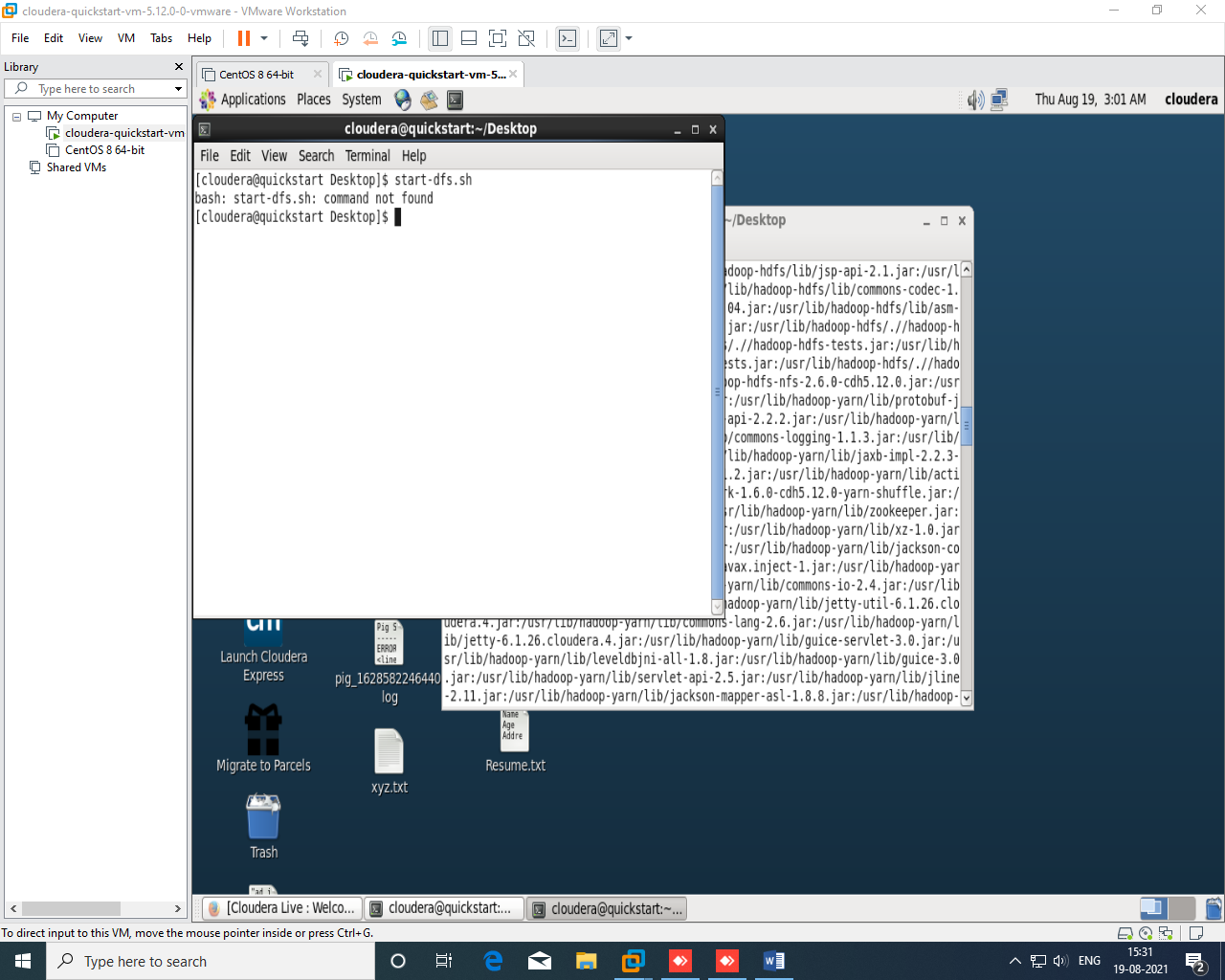
$ hadoop namenode –format



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.

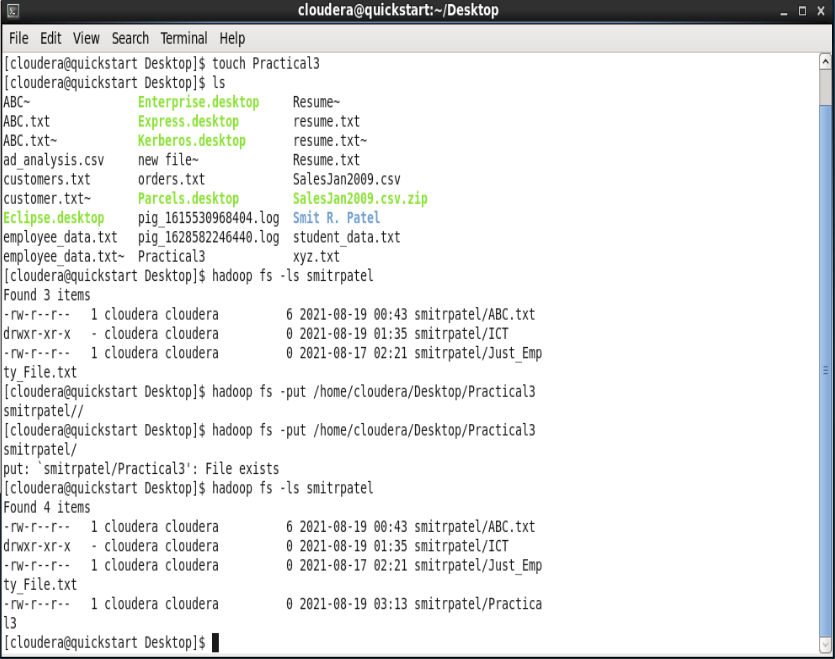
2. Alternatives to the copyFromLocal and copyToLocal commands:

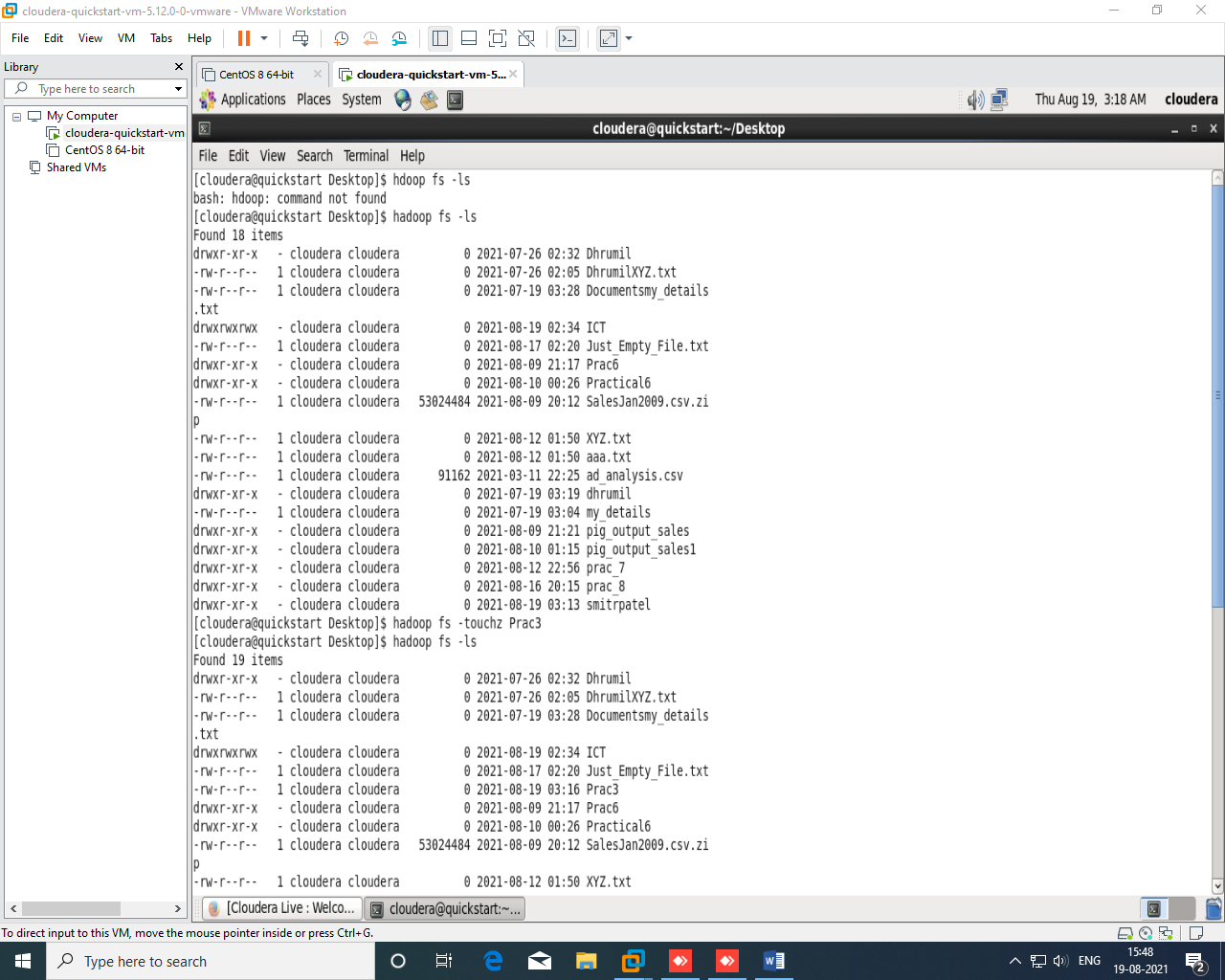
You also have alternatives to the copyFromLocal and copyToLocal commands. Try the GET and PUT

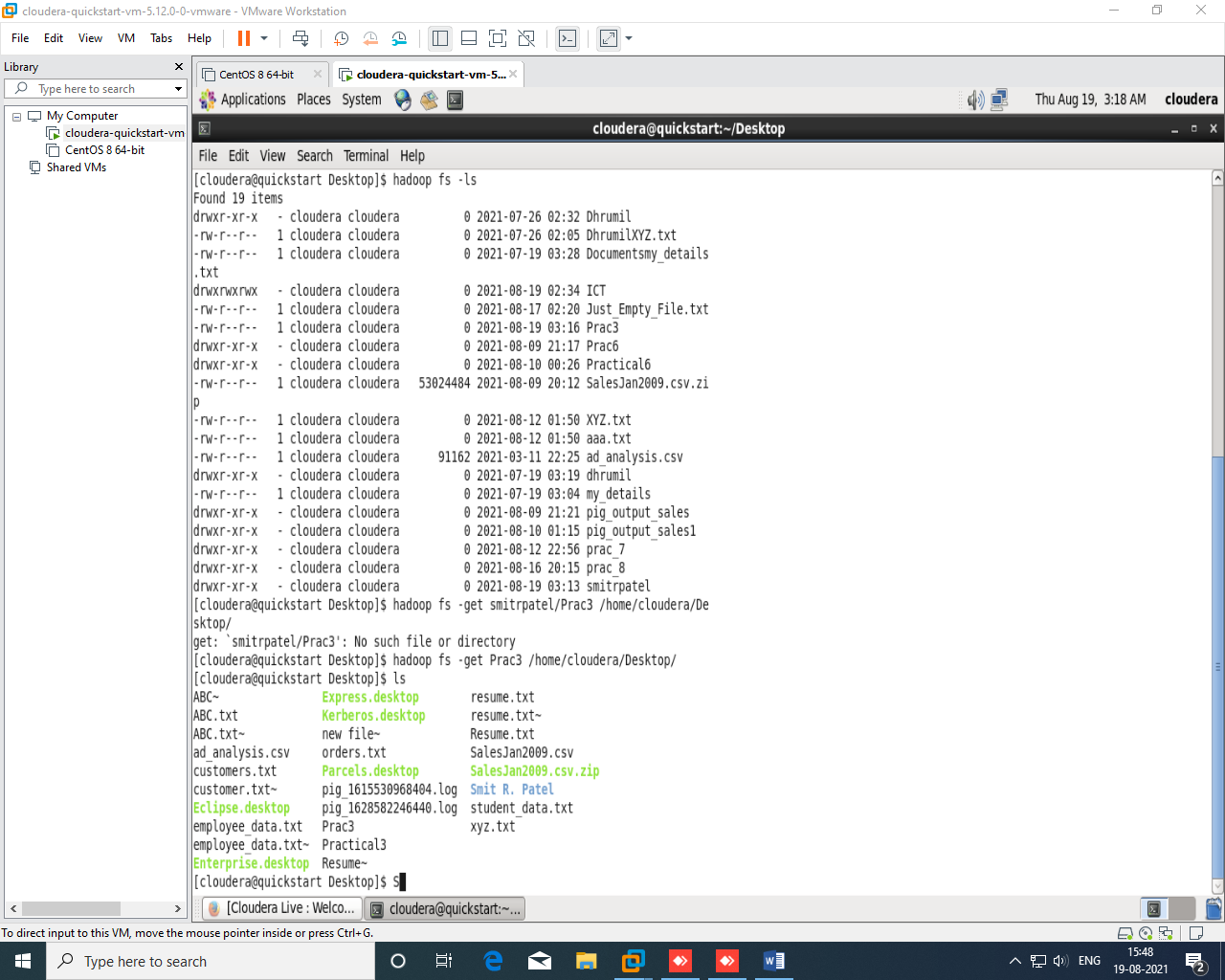
command as similar to what you did for copyFromLocal and copyToLocal. Create a file called “Practical

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.





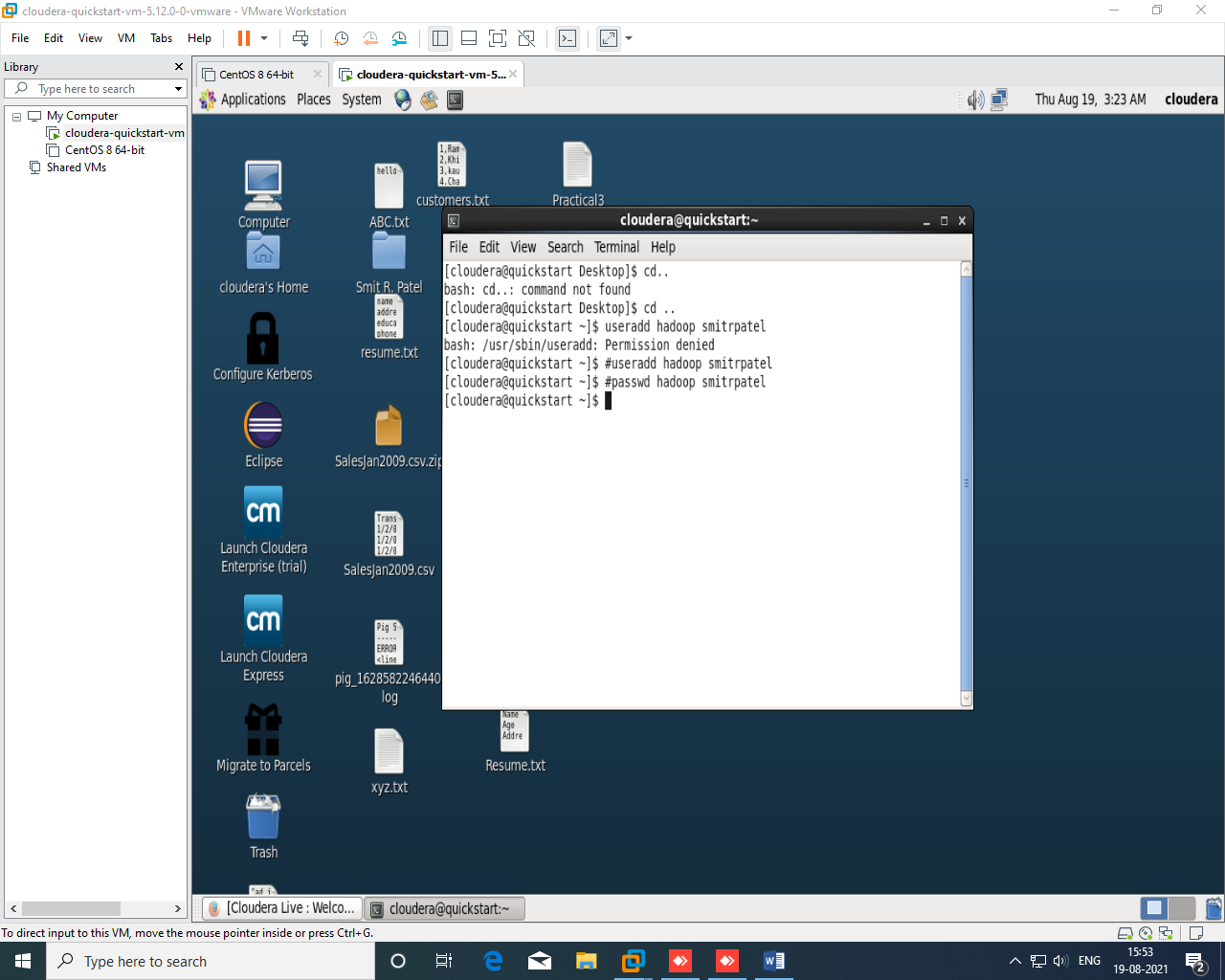


3. Create a user on Hadoop:

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

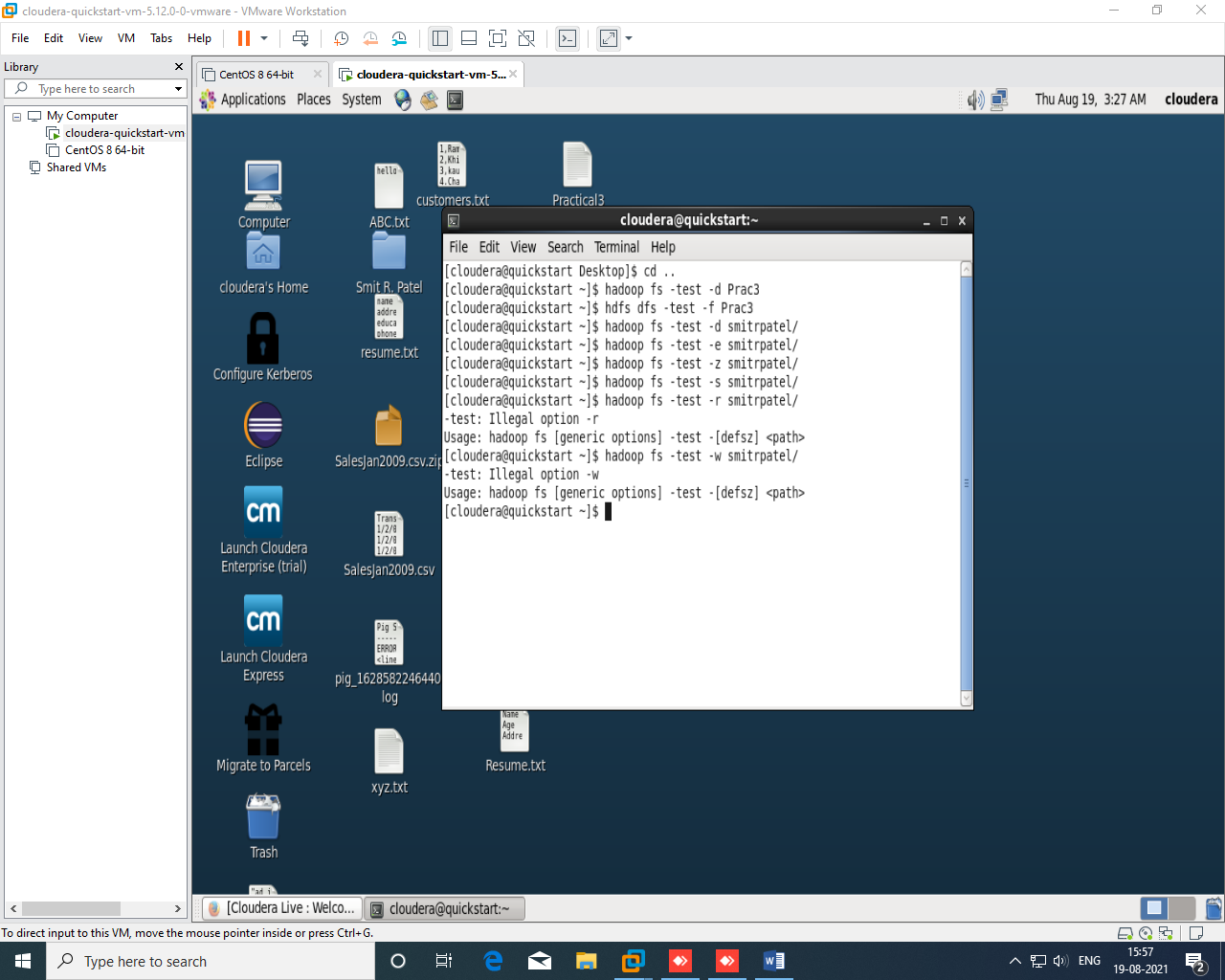
• f -> Checks if it is a file or not

• 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



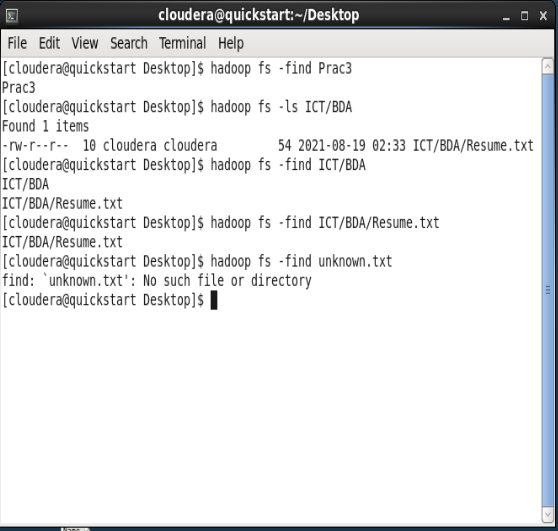
5. Search for files in the HDFS cluster.

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

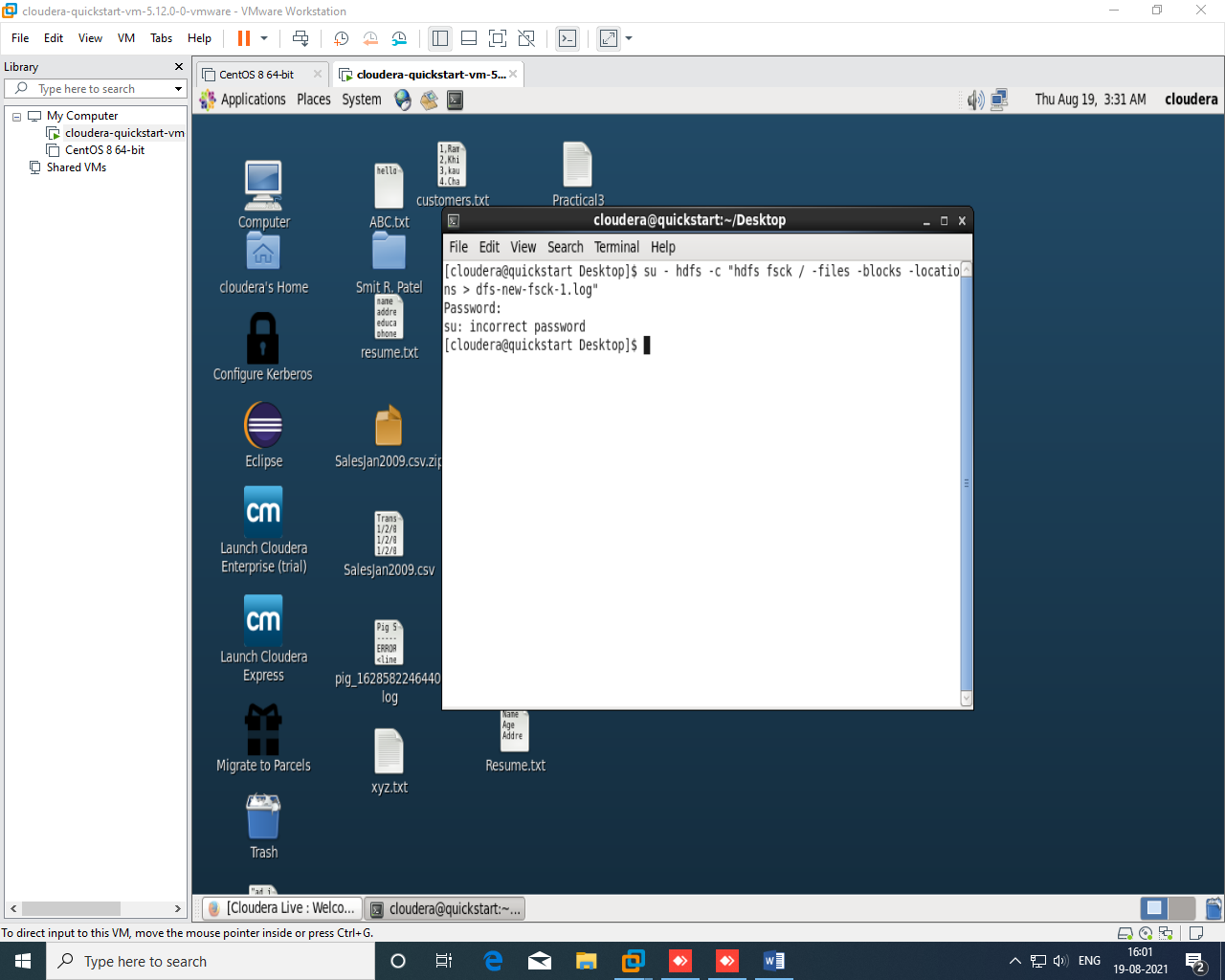
defined expression.

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



6. Verify HDFS file system health

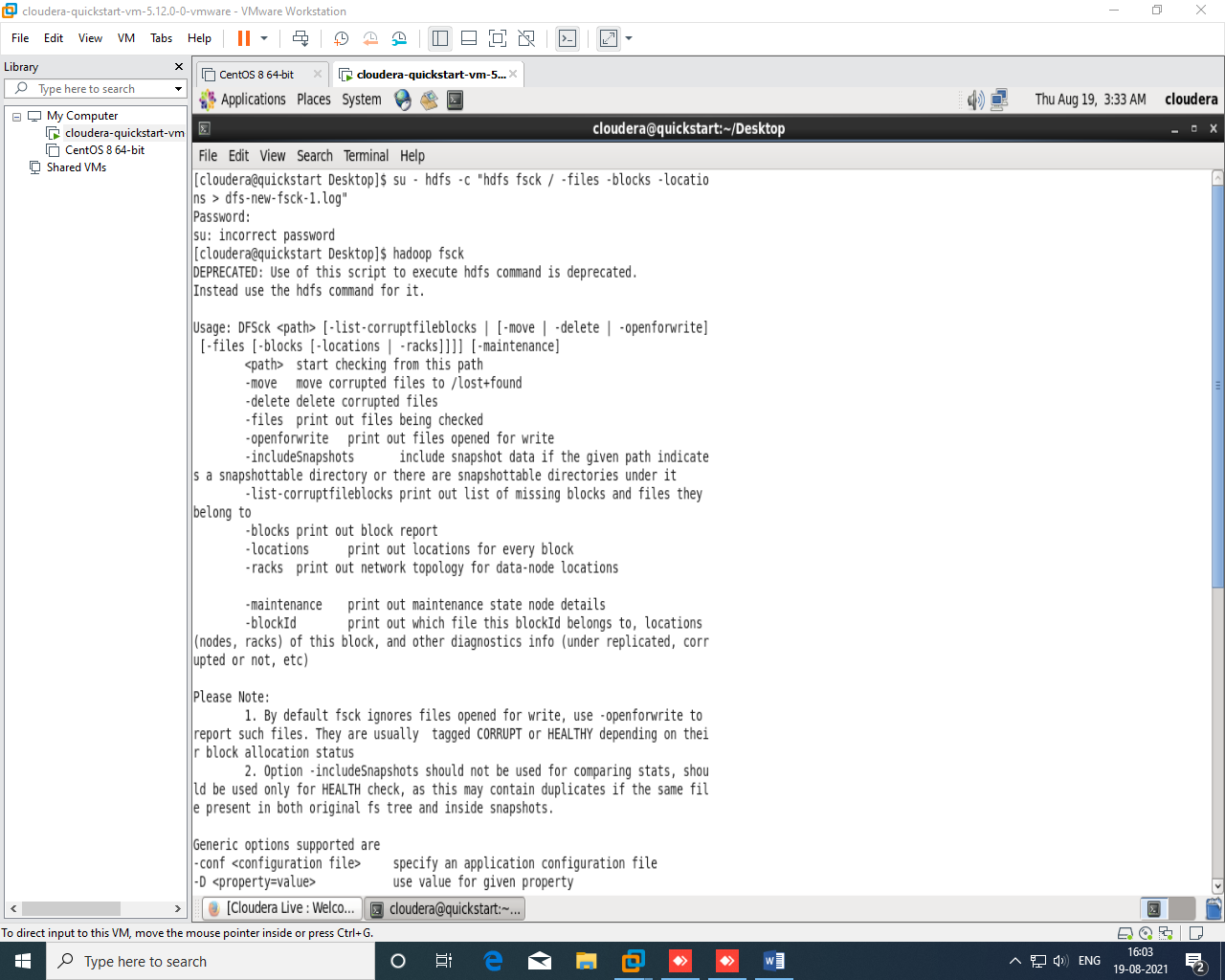
su - hdfs -c "hdfs fsck / -files -blocks -locations > dfs-new-fsck-1.log"



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

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

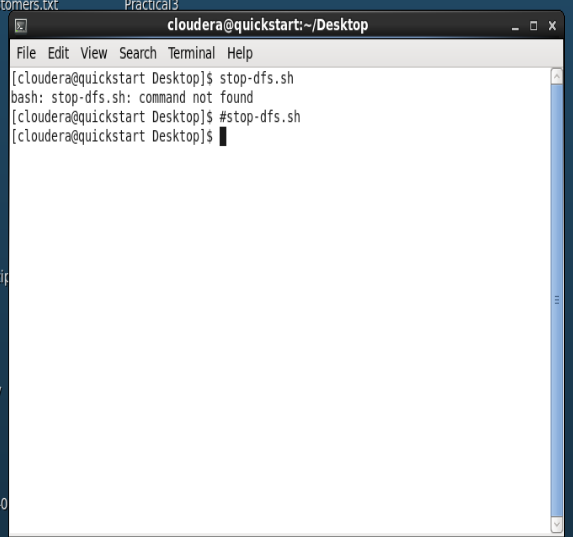
Command : Hadoop fsck



7. Shutting Down HDFS

Shut down the HDFS files by following the below command:

$ stop-dfs.sh



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