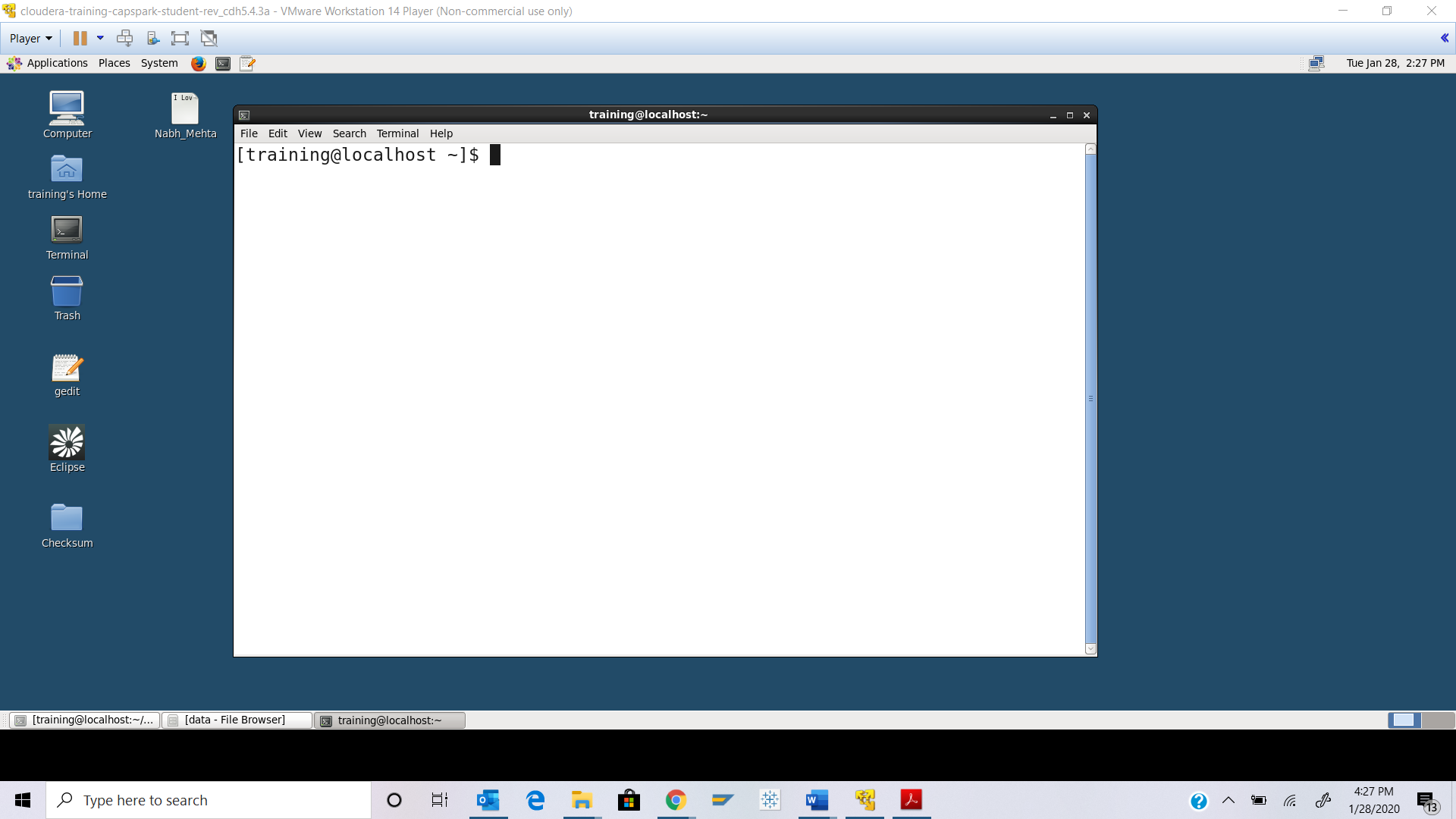
**HDFS**

**Student Name: Nabh Sanjay Mehta Student Id: NSM190002**

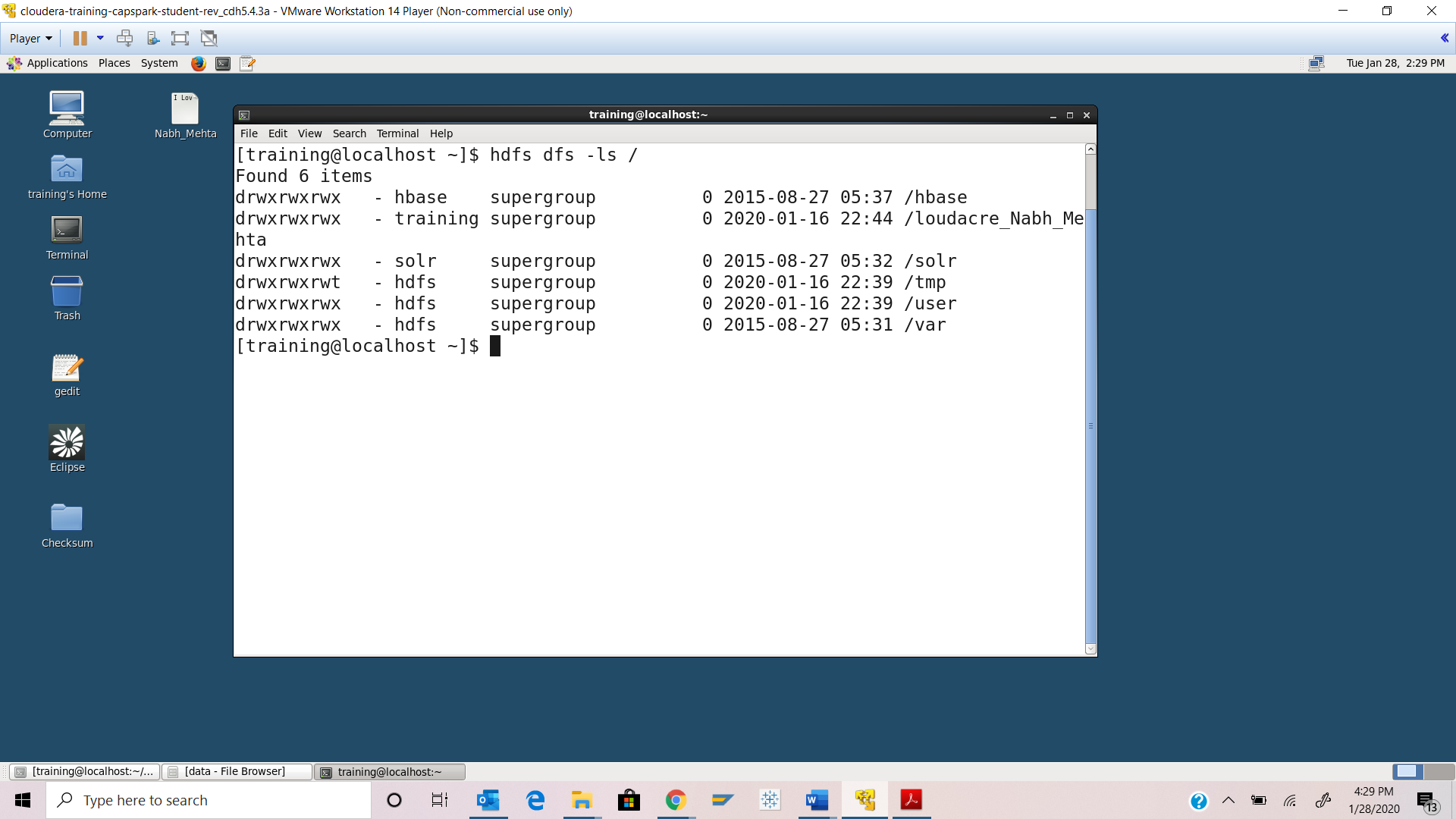
**Date:**

Step-1: Start the Cloudera VM and then launch the “terminal” application

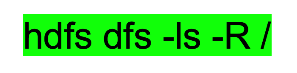


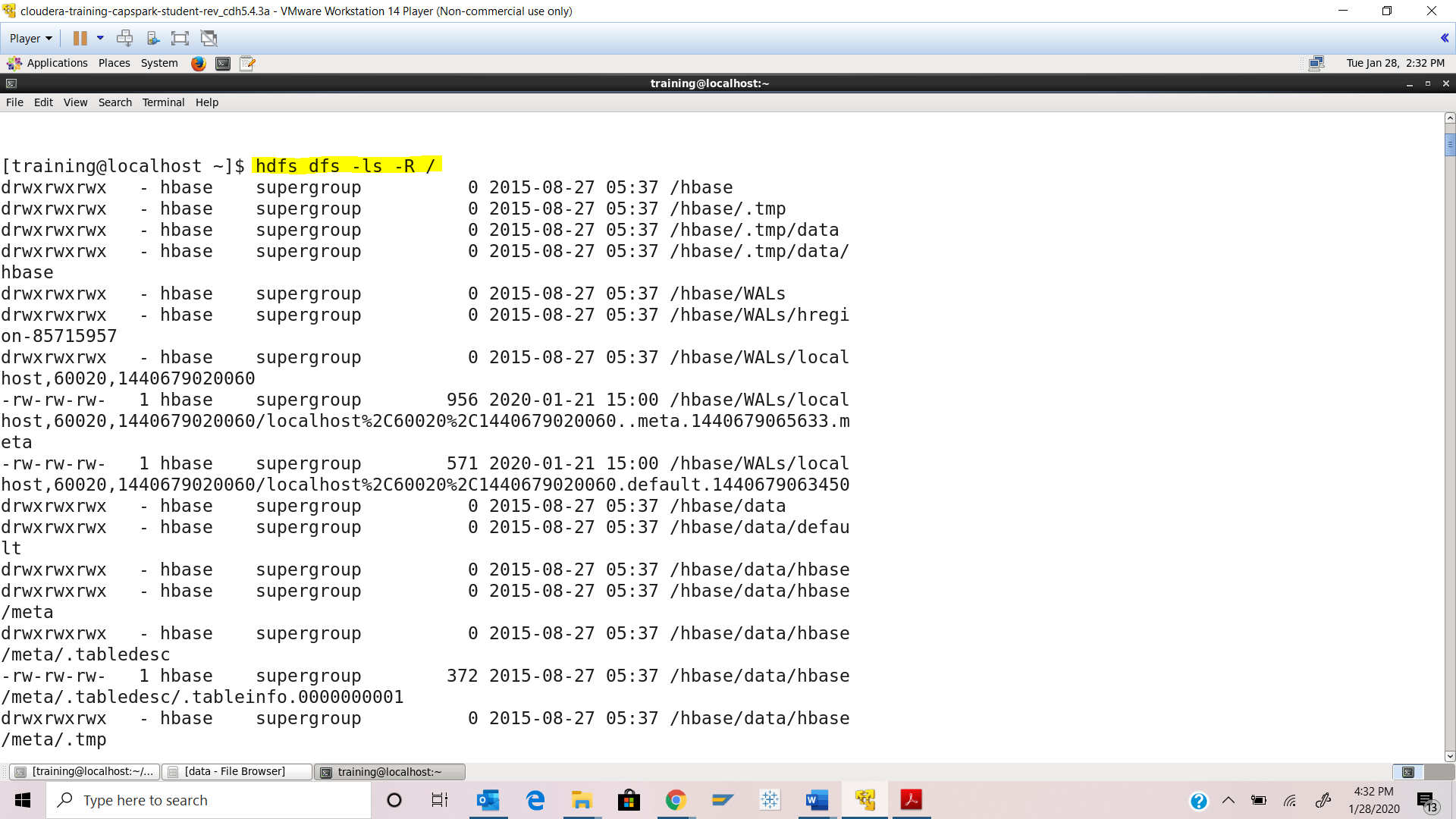
Step-2: Enter the following command at the shell prompt:





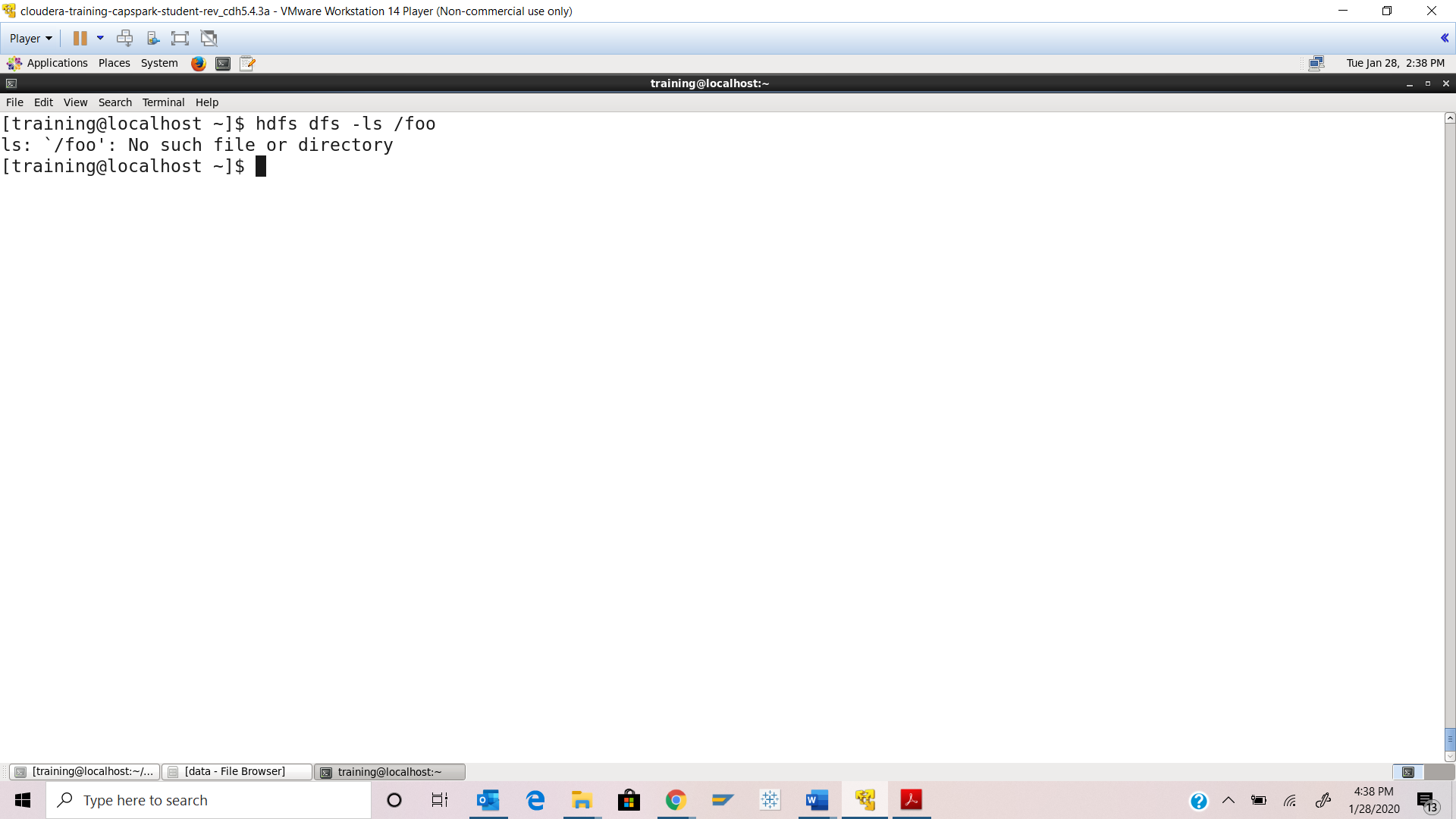
Step-3: Execute the following command at the shell prompt:





Step-4: Execute the following command at the shell prompt:



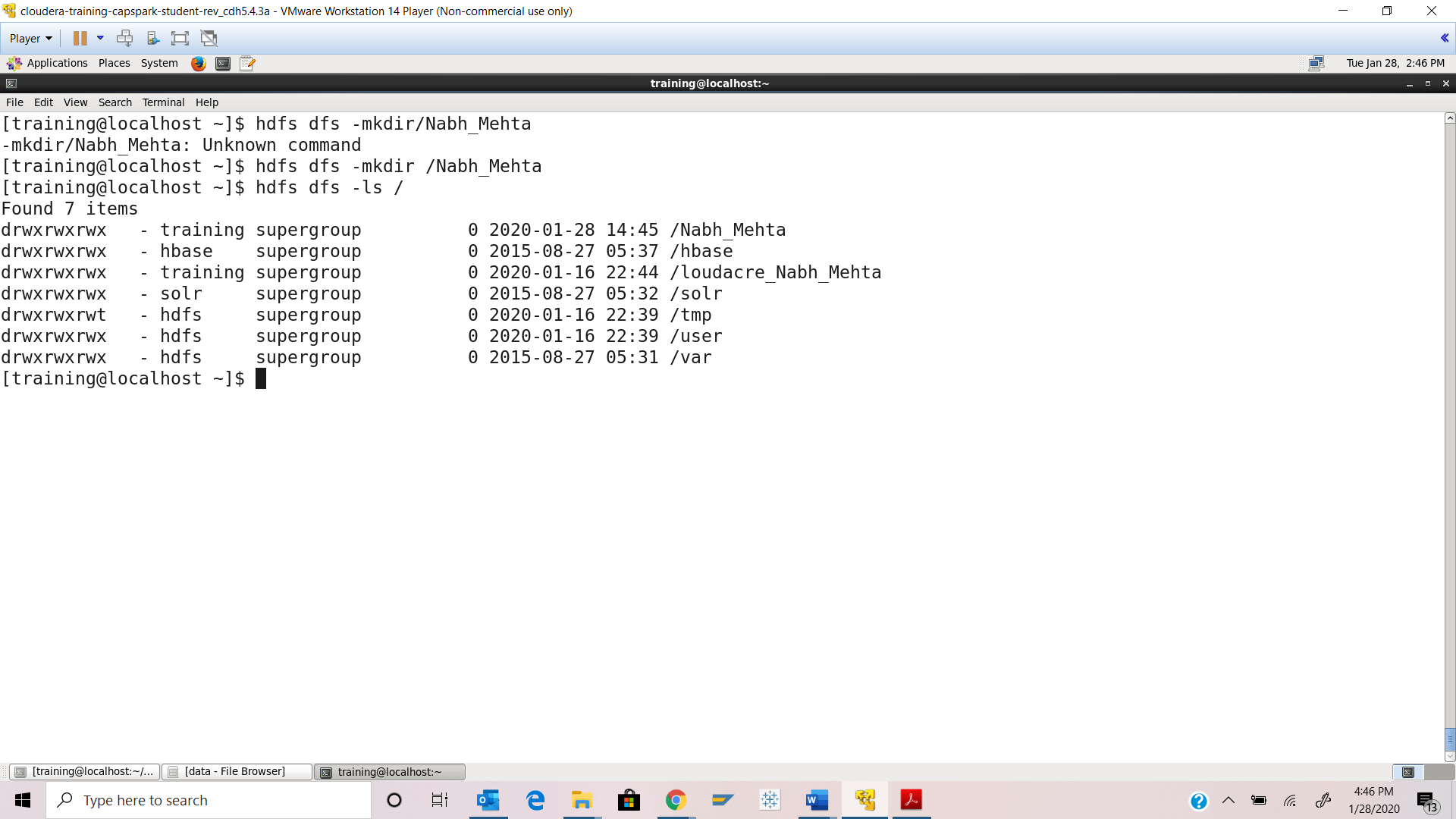


Step-5: Execute the following command at the shell prompt:



Execute the following command at the shell prompt:





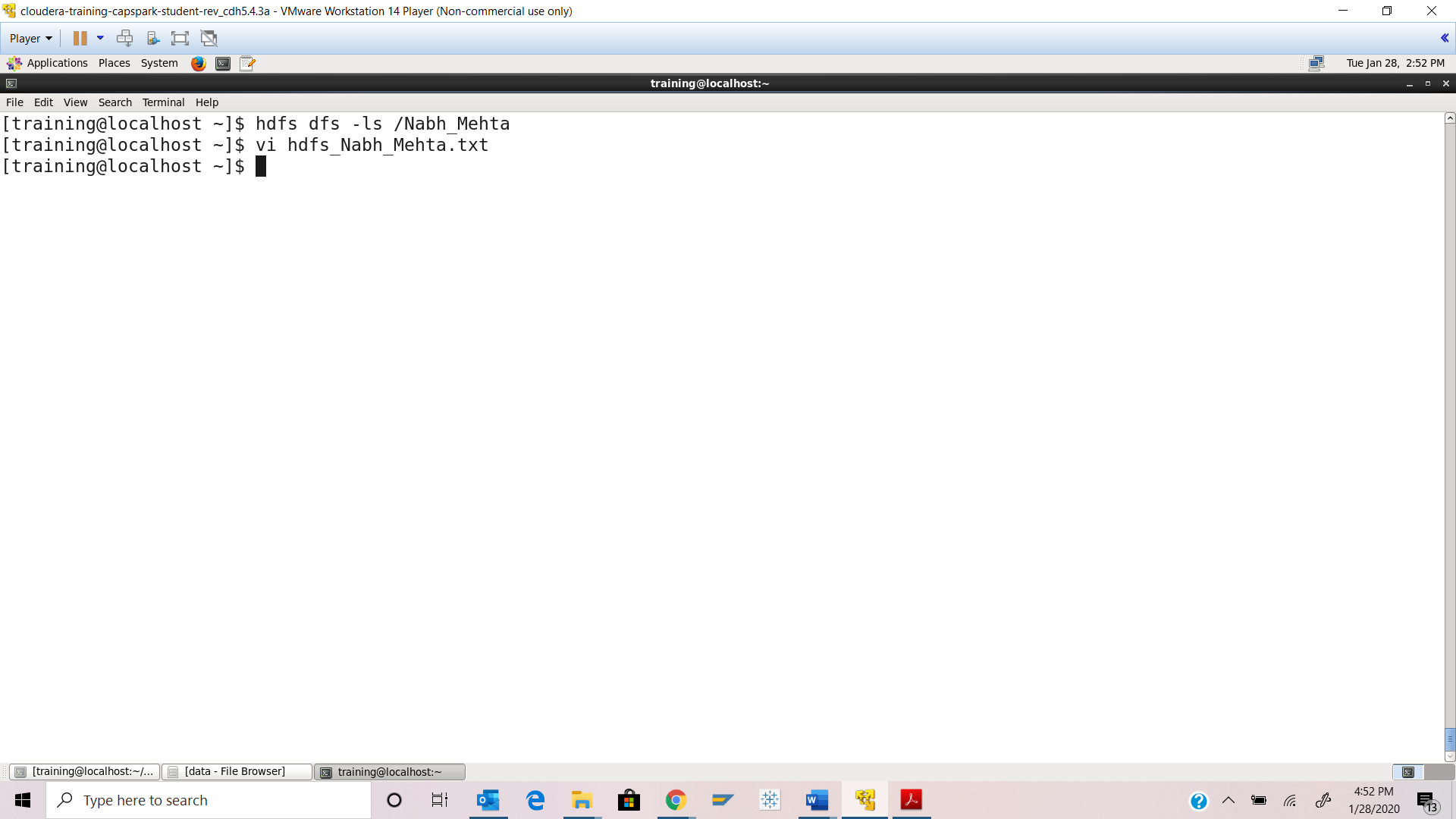
Step-6: Execute the following command at the shell prompt:





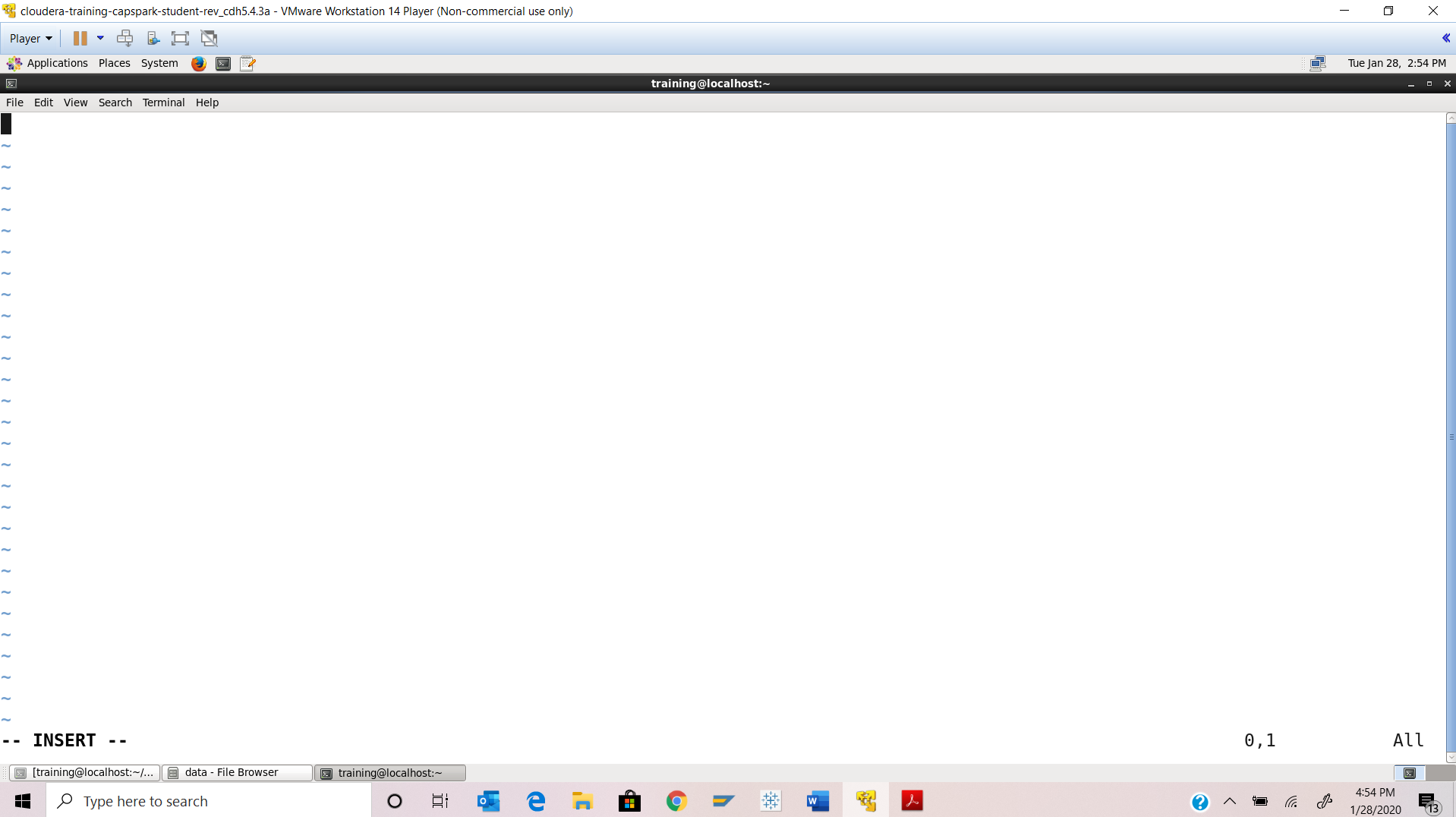
Step-7: Execute the following command at the shell prompt:



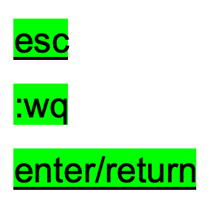


Enter the following command to enter Insert Mode:



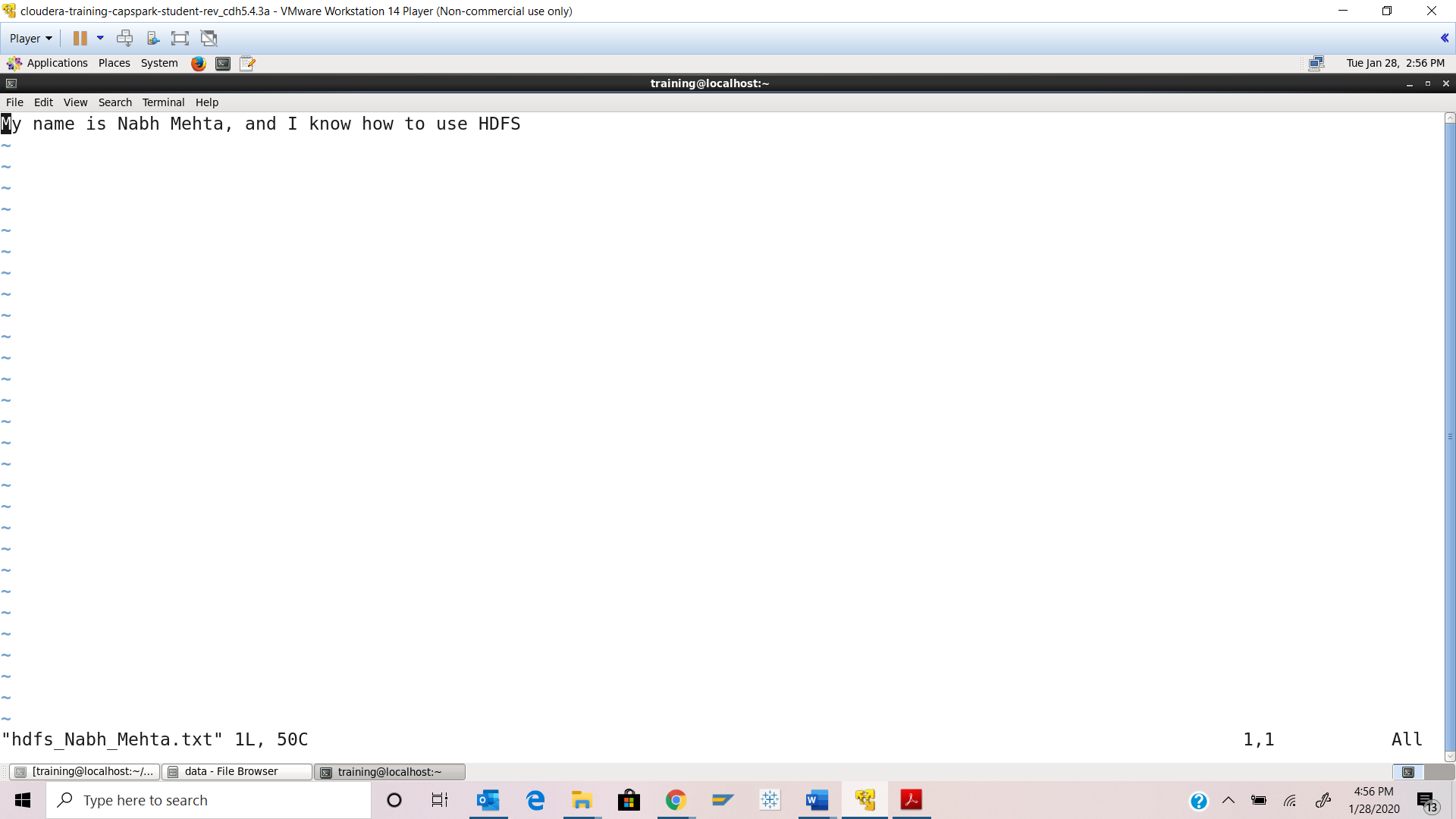


Now, Enter the following commands to save and exit VI editor:



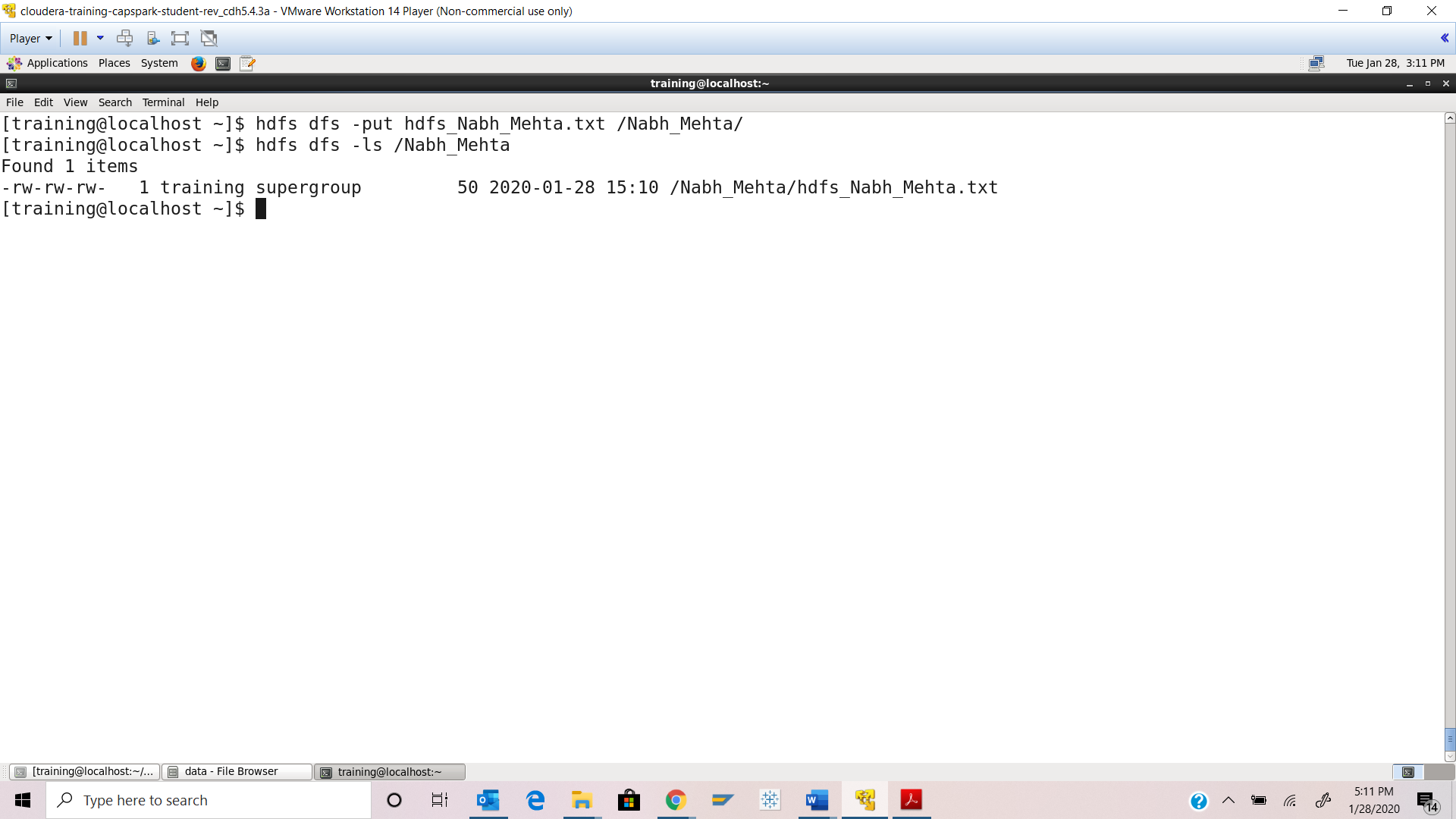
Execute the following command at the shell prompt:





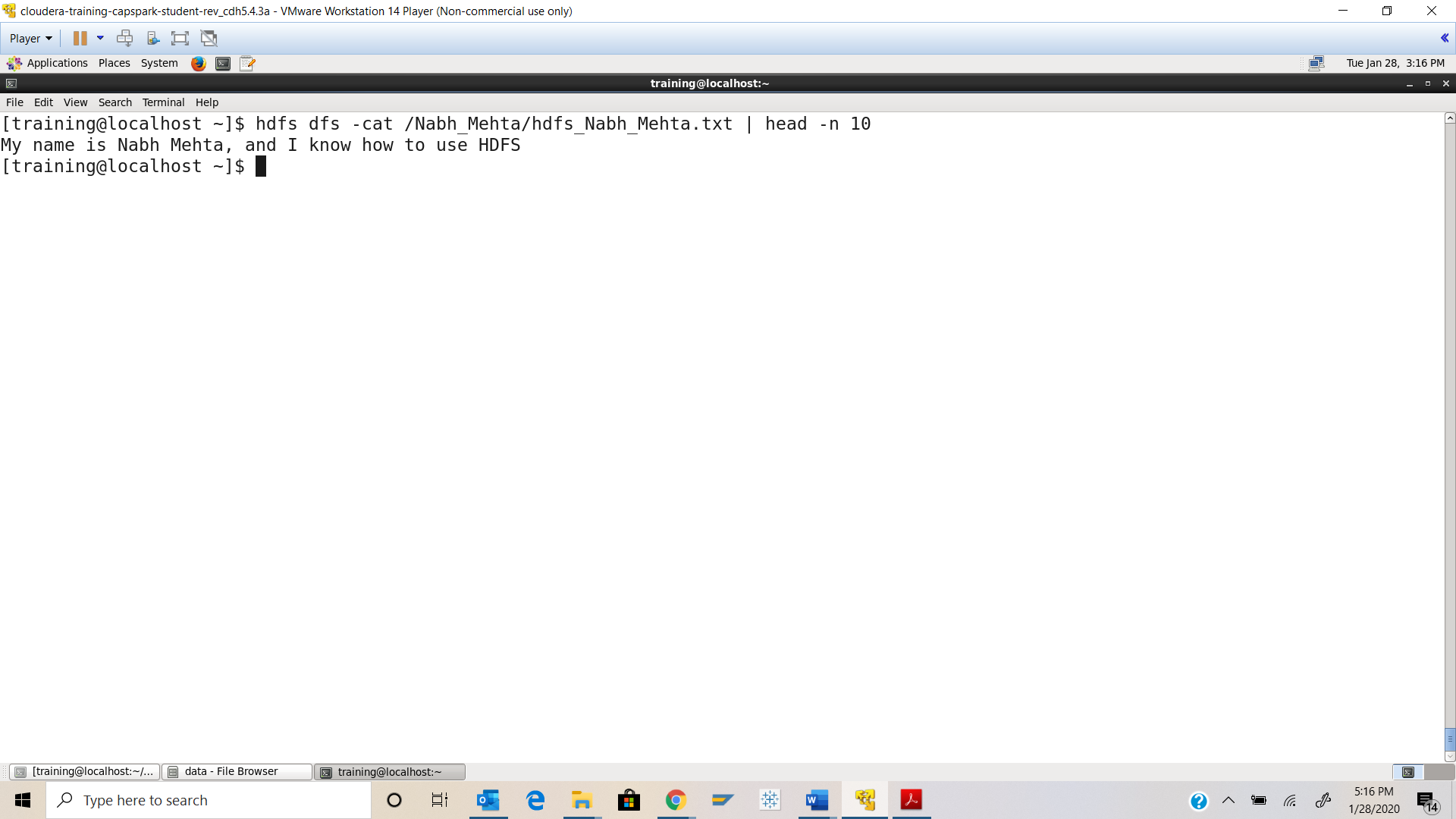
Step-8: Execute the following command:



Execute the following command at the shell prompt:

Step-9: Execute the following command at the shell prompt:



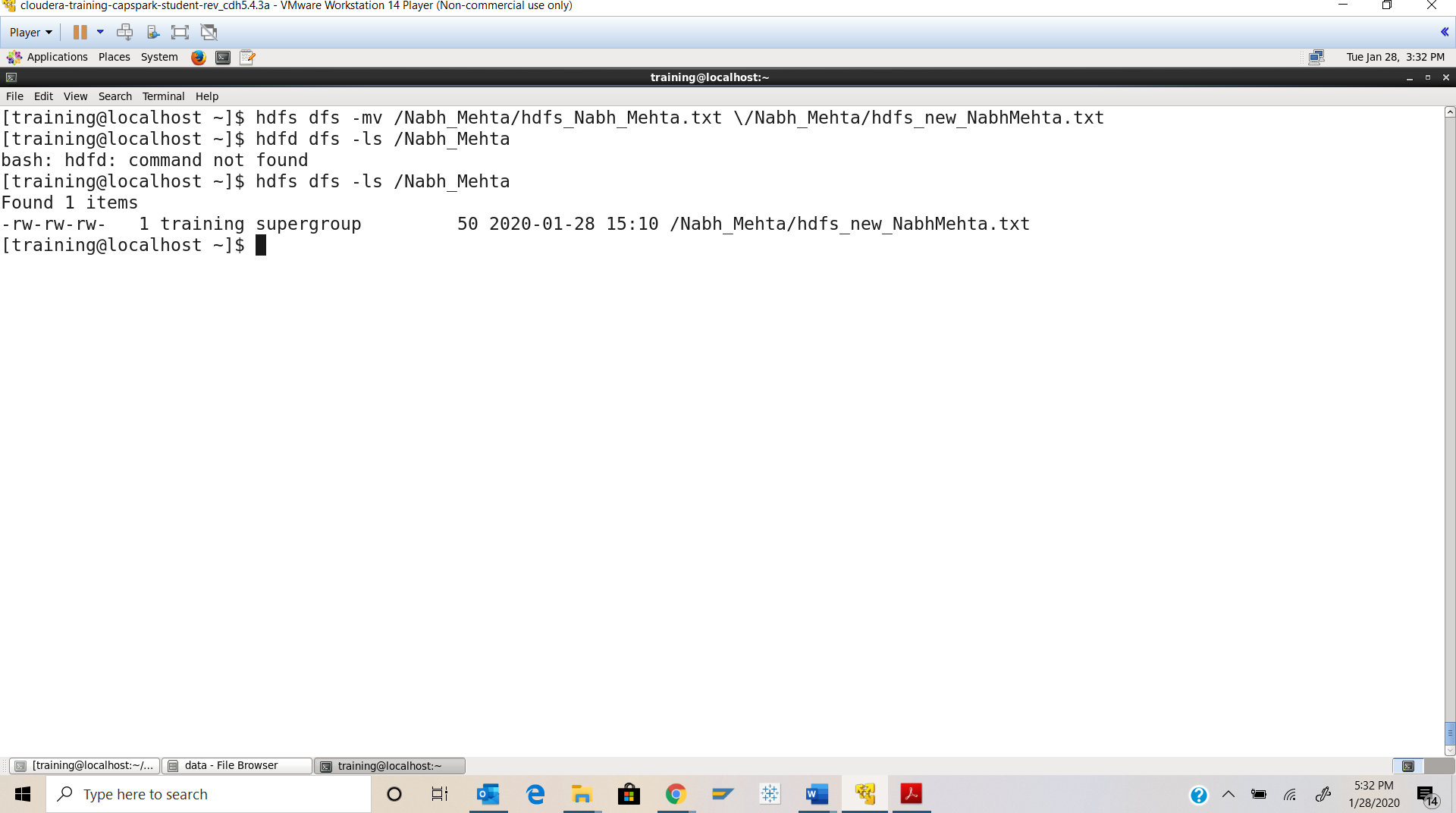


Step-10: Execute the following command at the shell prompt:

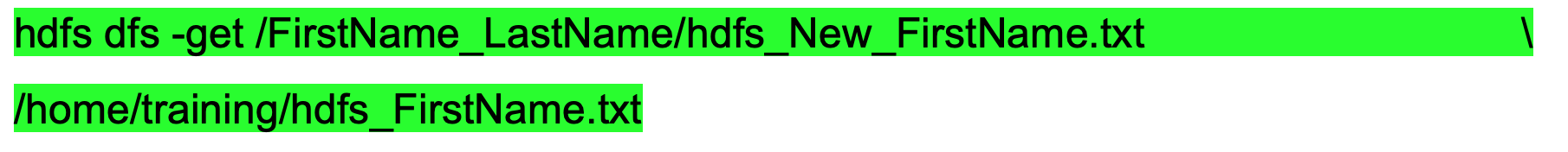


Execute the following command at the shell prompt:





Step-11: Execute the following command at the shell prompt:



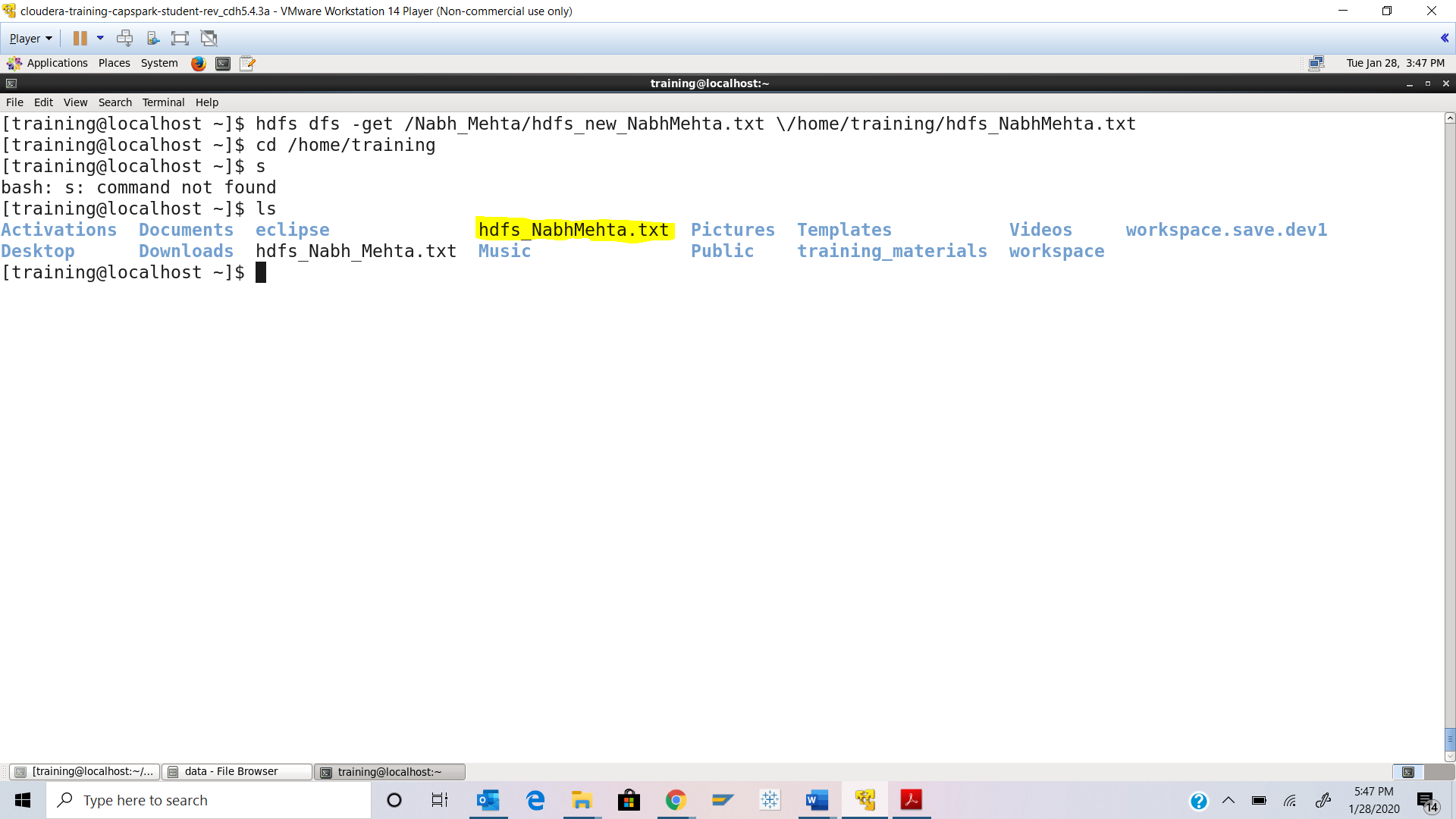
Q. Execute the following at the shell prompt:



Now execute the following command at the shell prompt:



Take a screenshot of the shell output and paste it below. Highlight the file that you copied from HDFS.



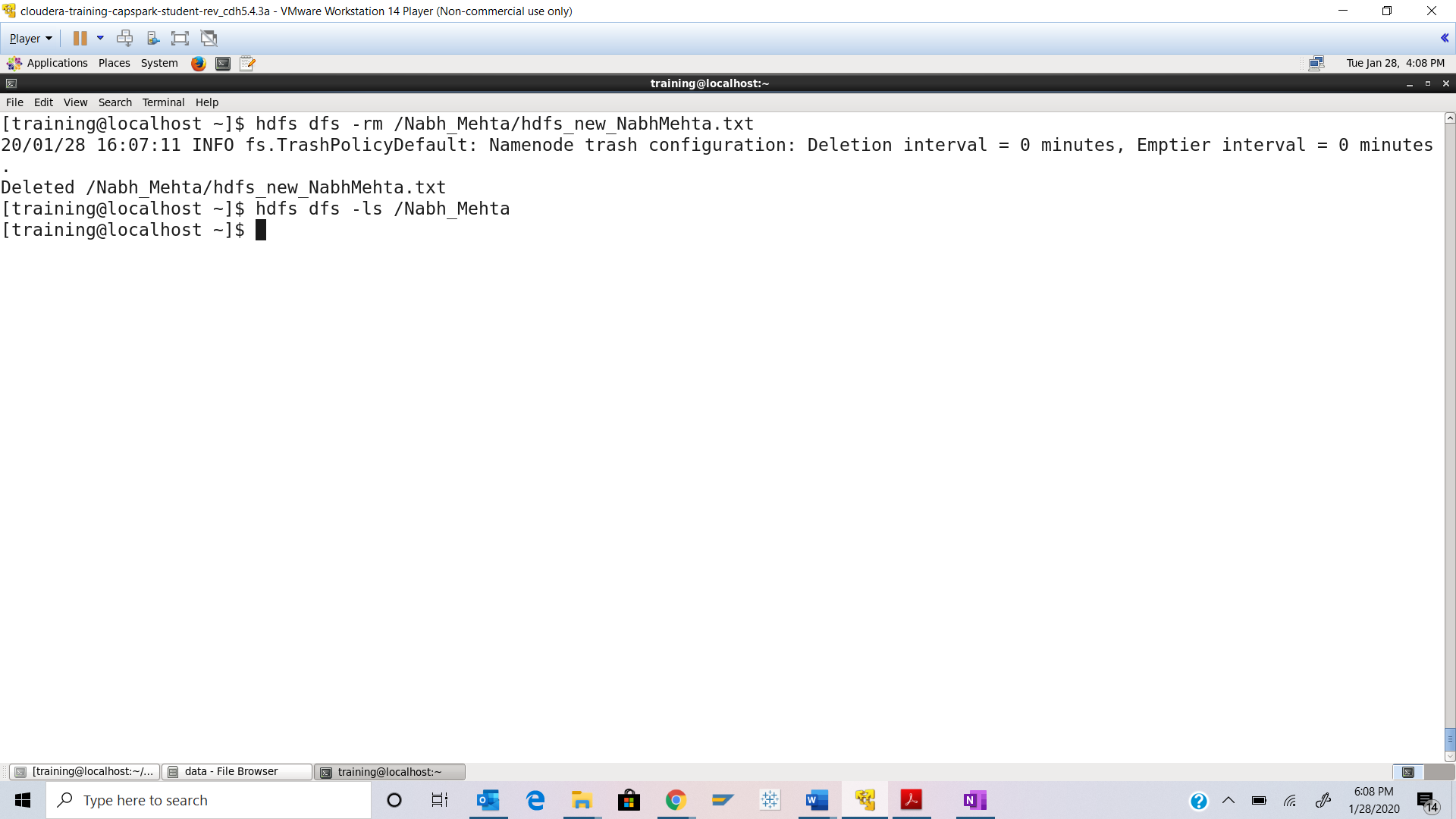
Step-12: Execute the following command at the shell prompt:



Q. Execute the following command at the shell prompt:



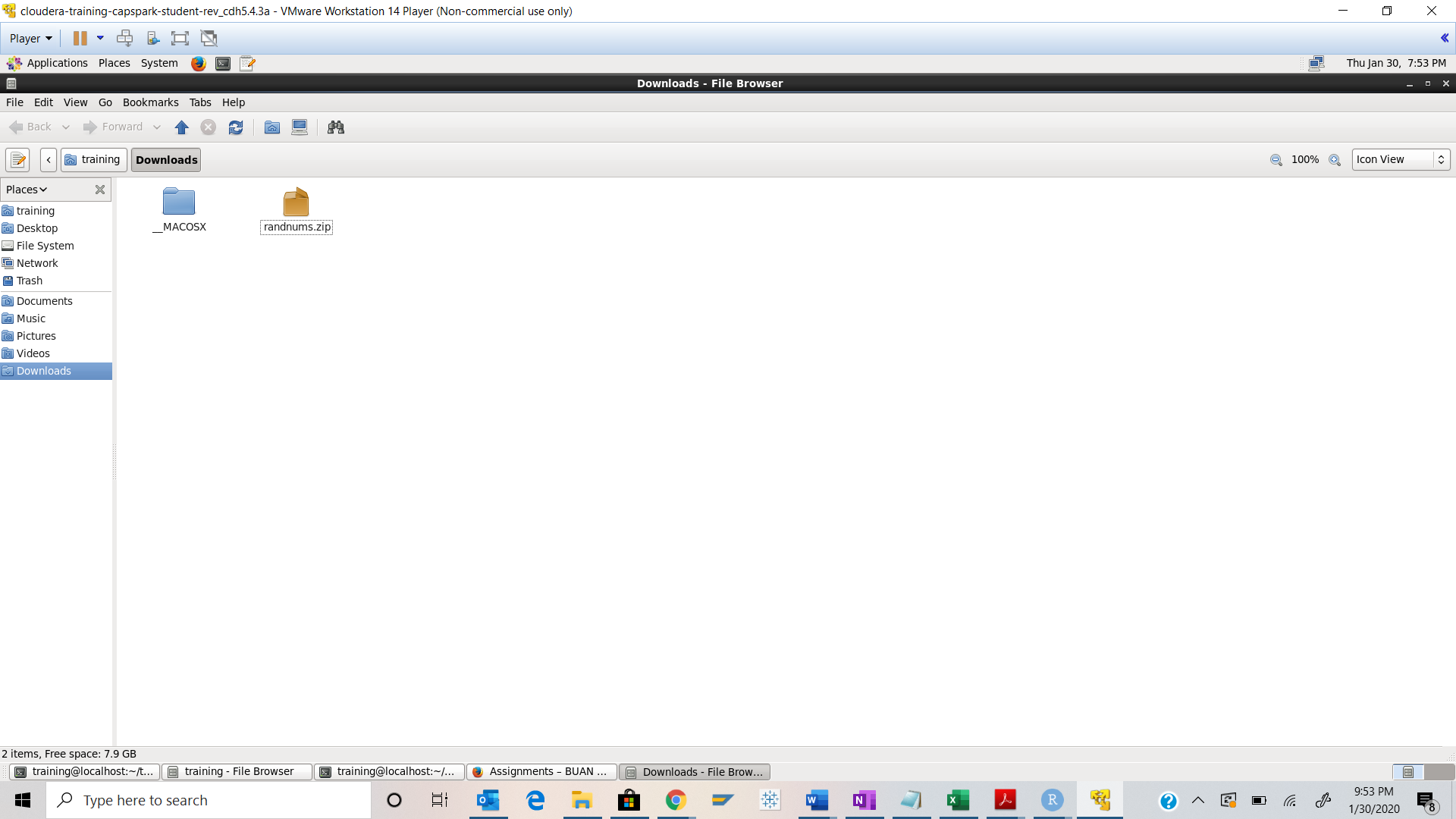
Take a screenshot of the shell output and paste it below. Why is the output empty after you execute the above command?



**Since we deleted the file, the folder is now empty and that is why it does not show anything.**

Part-3: HDFS Storage: In this section we are performing an exercise to get a better understanding of how storage works on HDFS.

Step-13: Open the Firefox browser **within** the Cloudera VM and download the randnums.zip file from eLearning into the downloads folder.



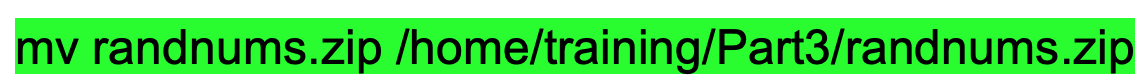
Step-14: Execute the following command:



Now execute the following command to change directories to the download directory:



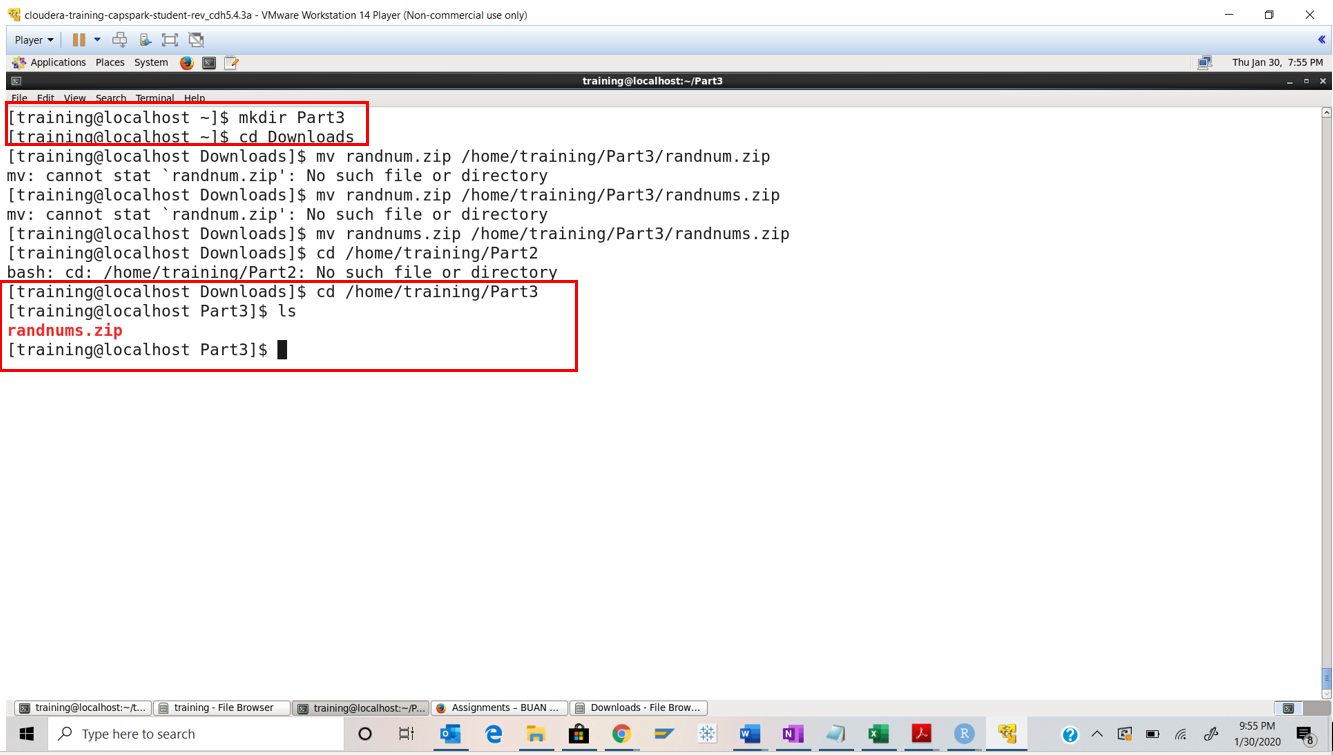
Execute the following command to move the “randnums.zip” folder to the new directory we created:



Now change directories to the directory we just created using the following command:



Q. Execute the command. Take a screenshot of the shell output, highlight the zip file and paste it below.

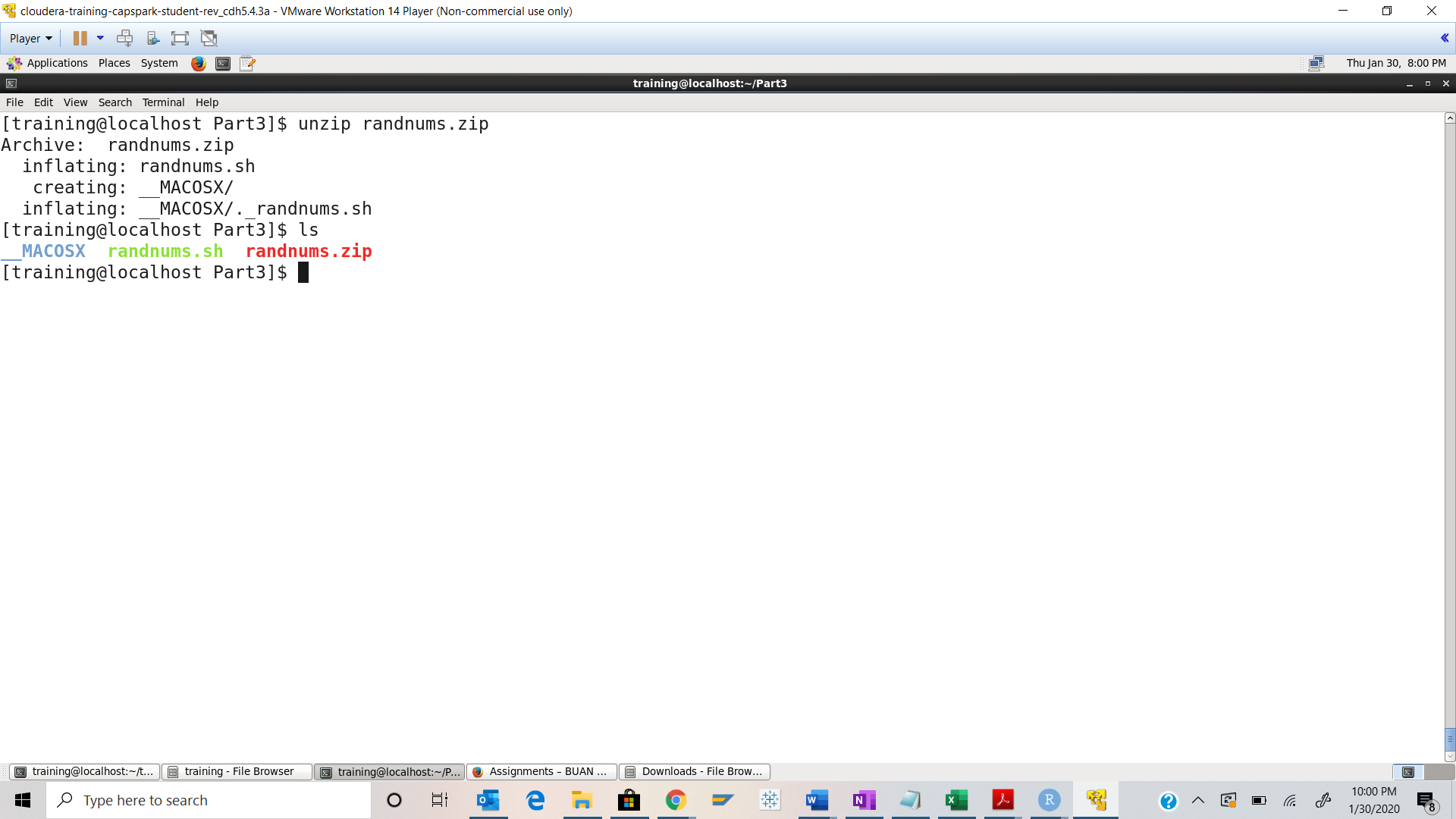


Step-15: Execute the following command at the shell prompt:

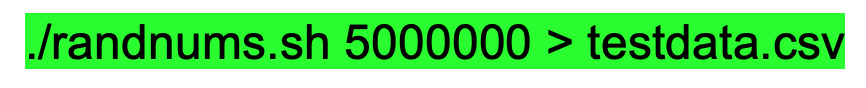




Q. Execute the command. Take a screenshot of the shell output, highlight the contents of the zip file and paste it below.



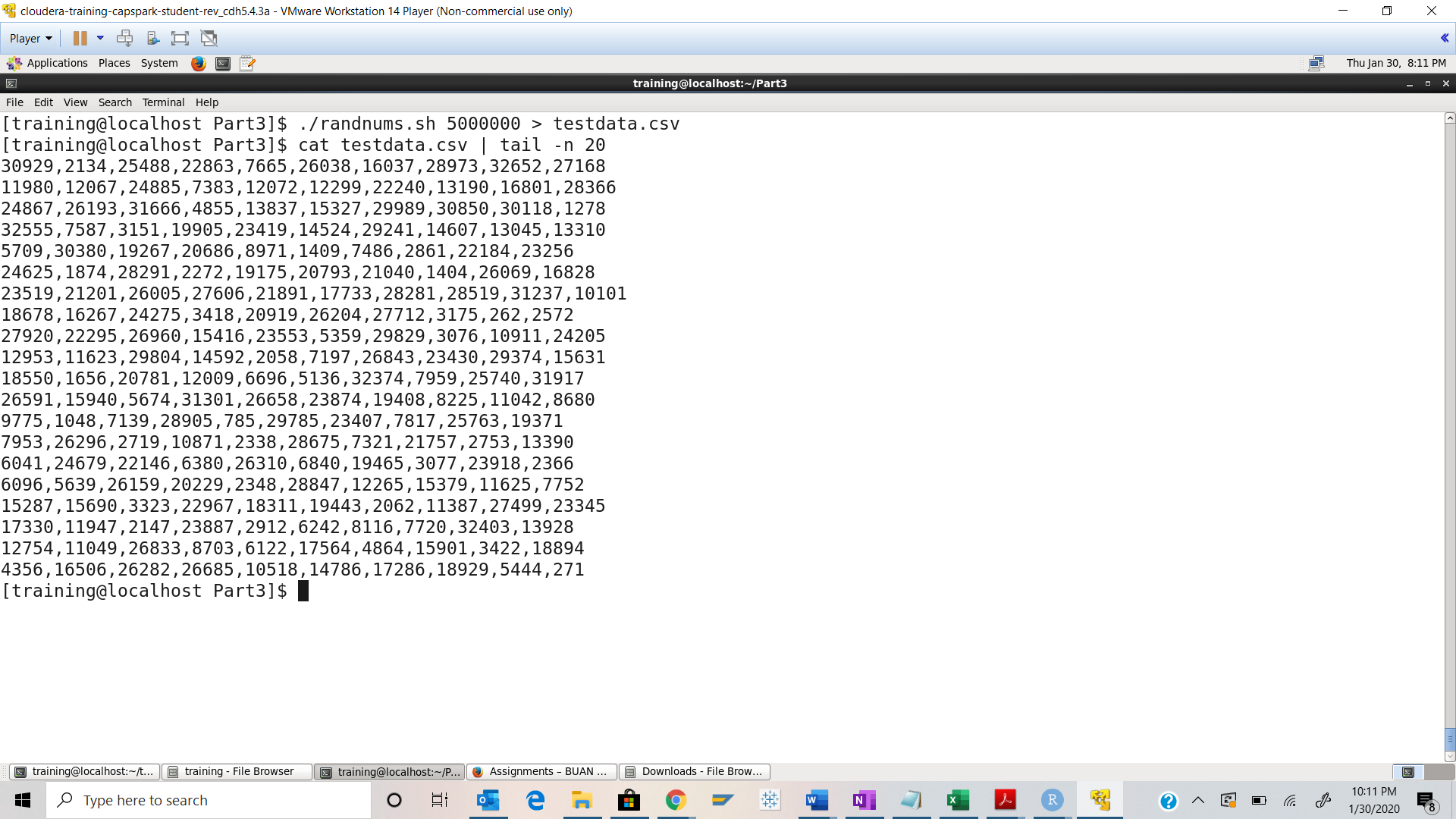
Step-16: Execute the following command at the shell prompt:



Now enter the following command at the shell prompt:



Q. Take a screenshot of the shell output and paste it below.

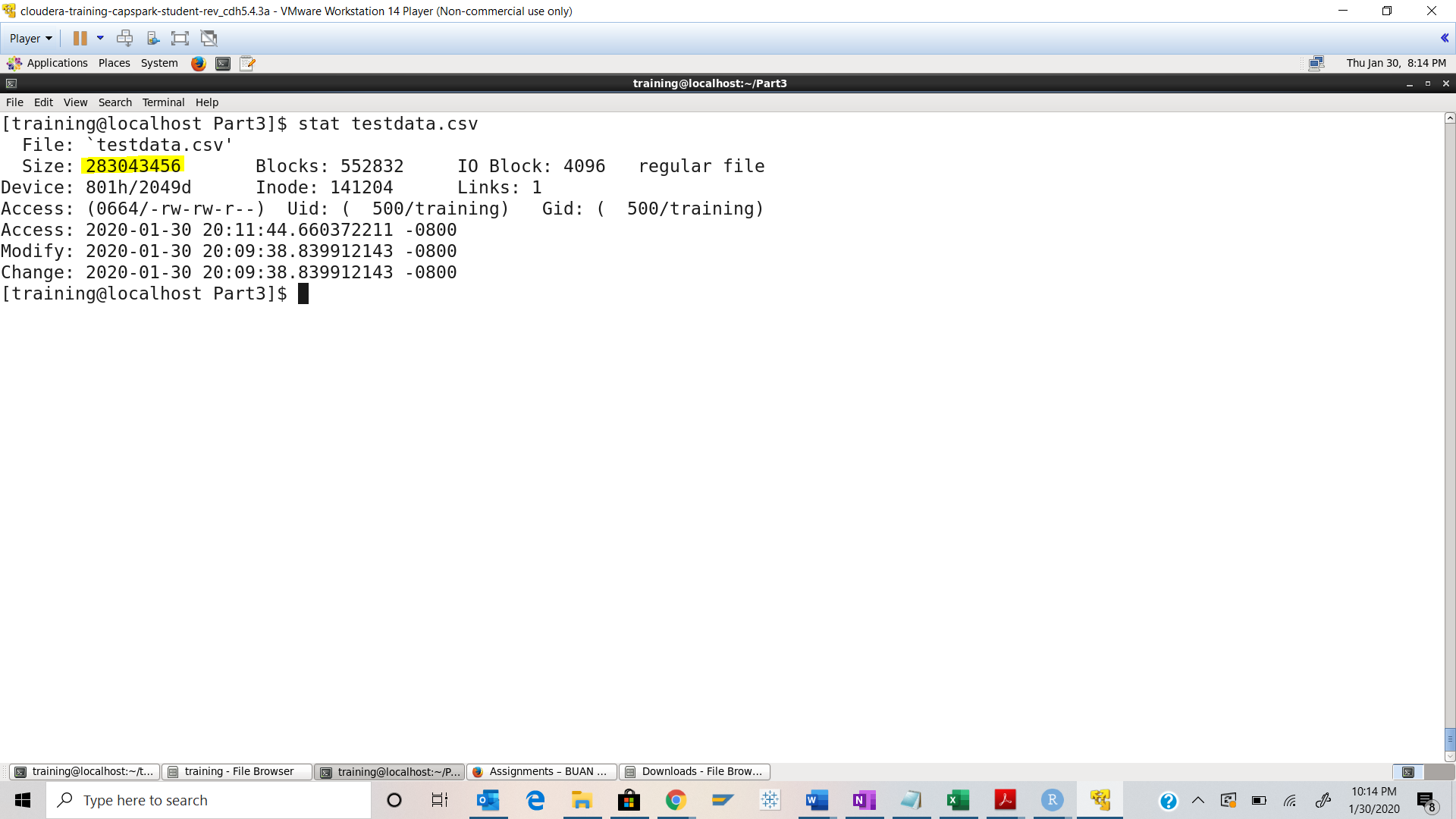


Step-17: Execute the following command at the shell prompt:



Q. Take a screenshot of the shell output and paste it below. What is the size of the testdata.csv file in megabytes?

**Ans 283.043456 MB**



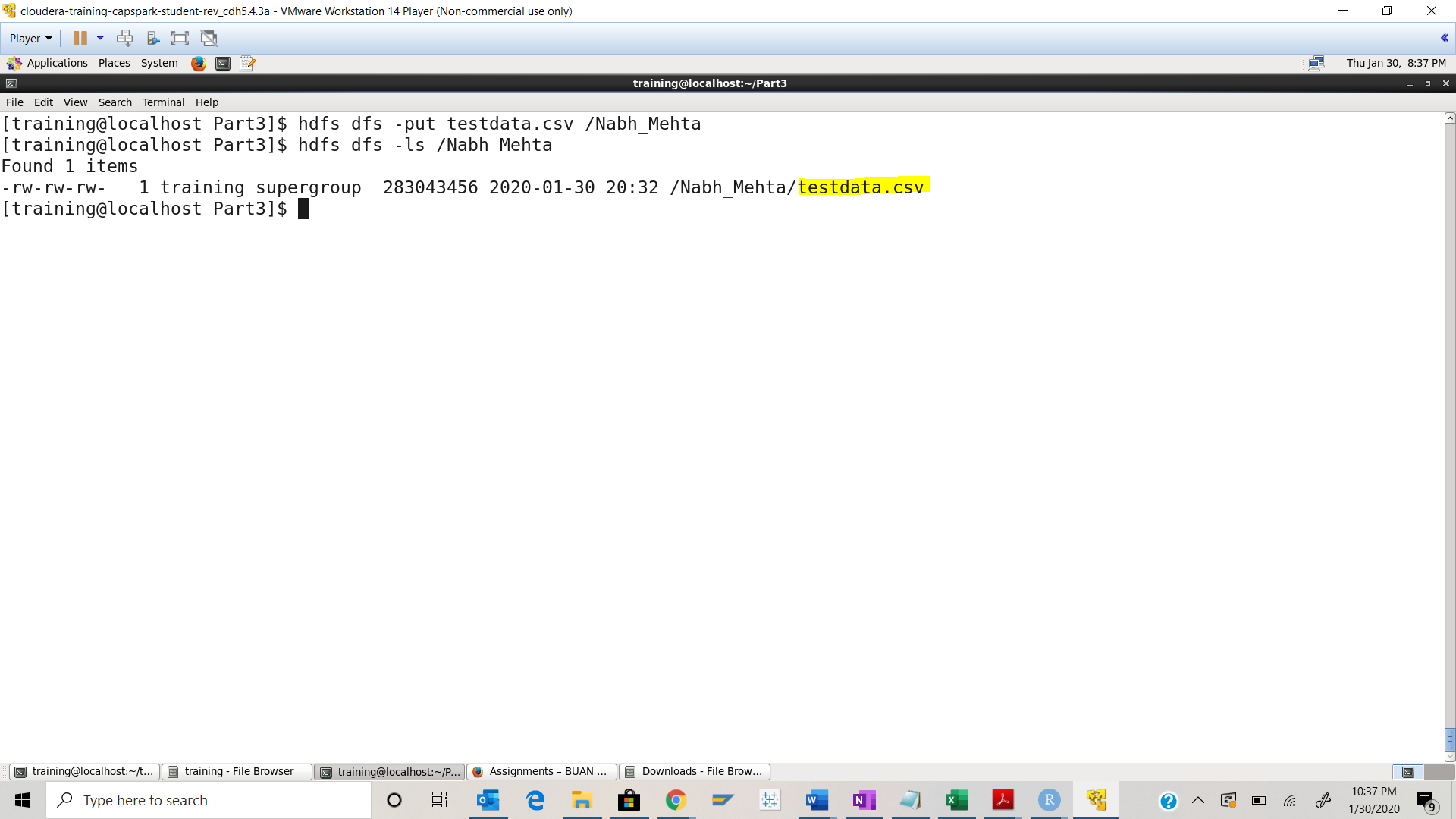
Step-18: Execute the following command at the shell prompt:



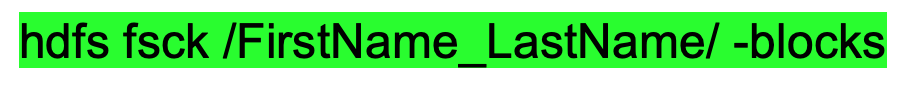
Q. Execute the following command to the home directory on the local system:



Take a screenshot of the shell output and paste it below. Highlight the file that you created in Step-16.

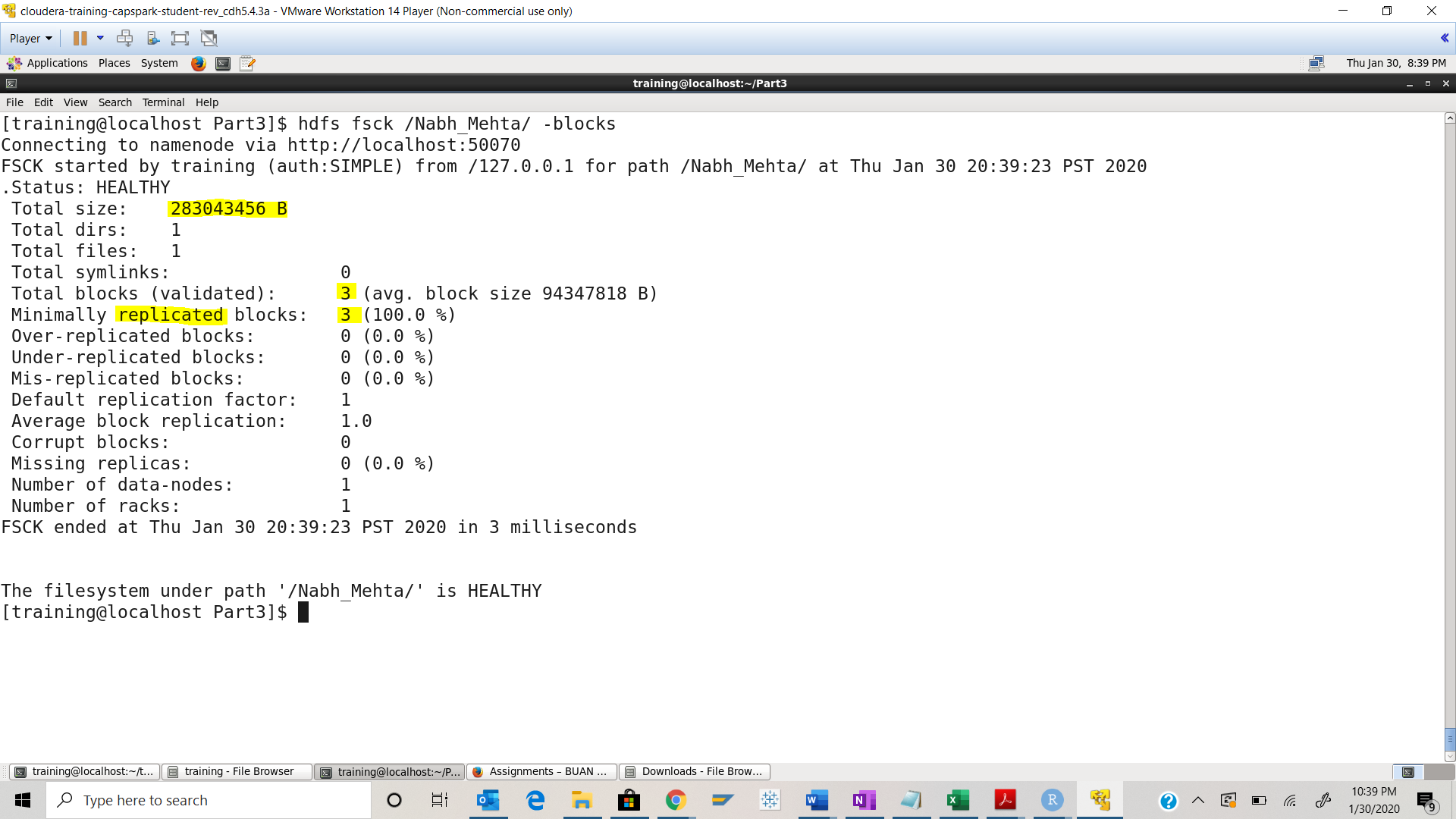


Step-19: Execute the following command at the shell prompt:



Q. Take a screenshot of the shell output and paste it below. How many blocks did our file create? What is the size of these blocks in megabytes? Why did we create these many blocks? Did the blocks replicate?

1. No of Blocks: 3
2. Average block size: 94.347818 MB
3. HDFS files system has default redundancy of 3. HDFS is setup for high availability with active and standby NameNodes
4. Yes, blocks are replicated as shown below in snapshot highlighted

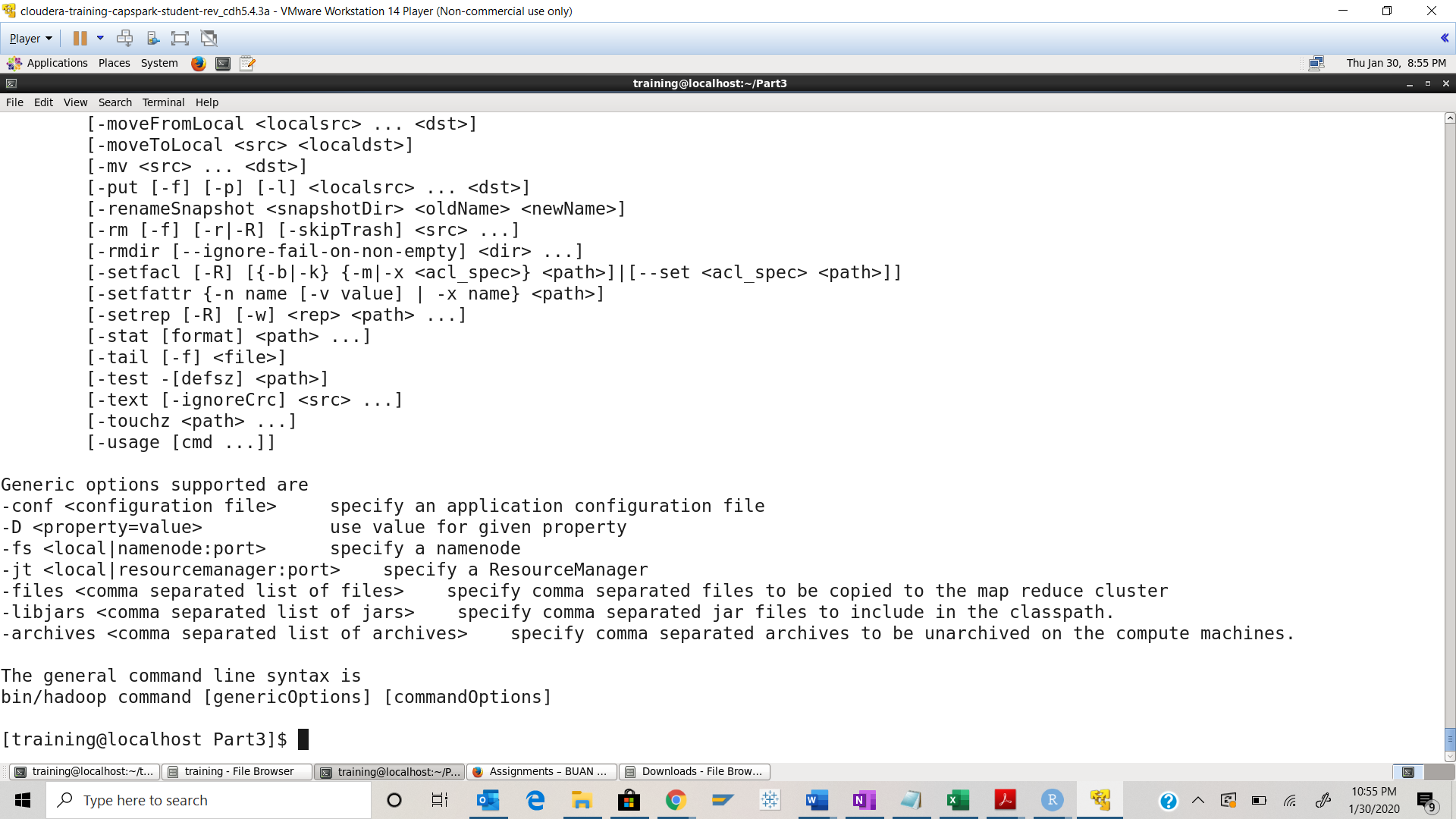


Part-4: Finding Help: This section will cover how to get additional information about commands using the shell.

Step-20: Execute the following command at the shell prompt:



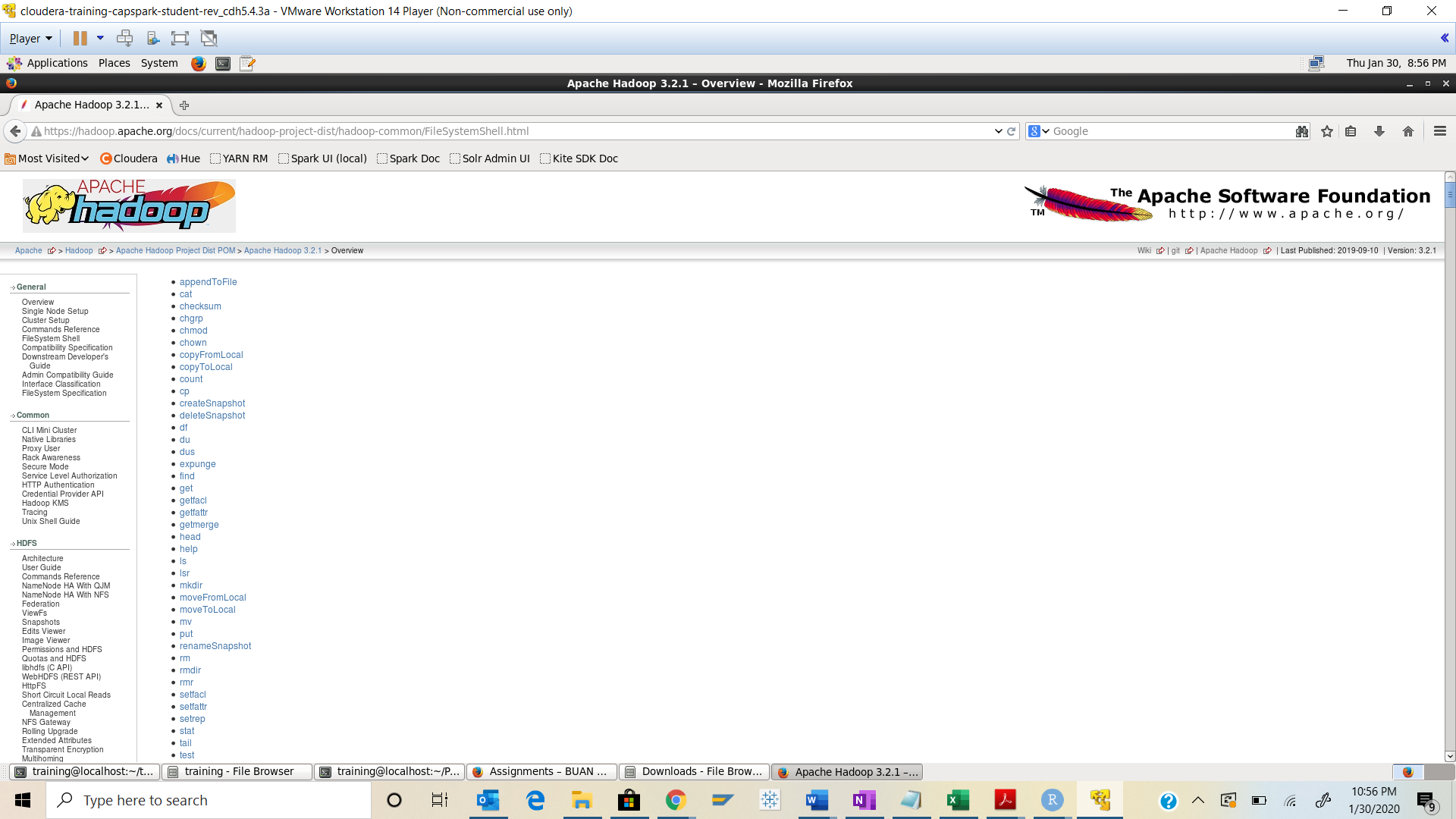
Q. Take a screenshot of the shell output and paste it below.



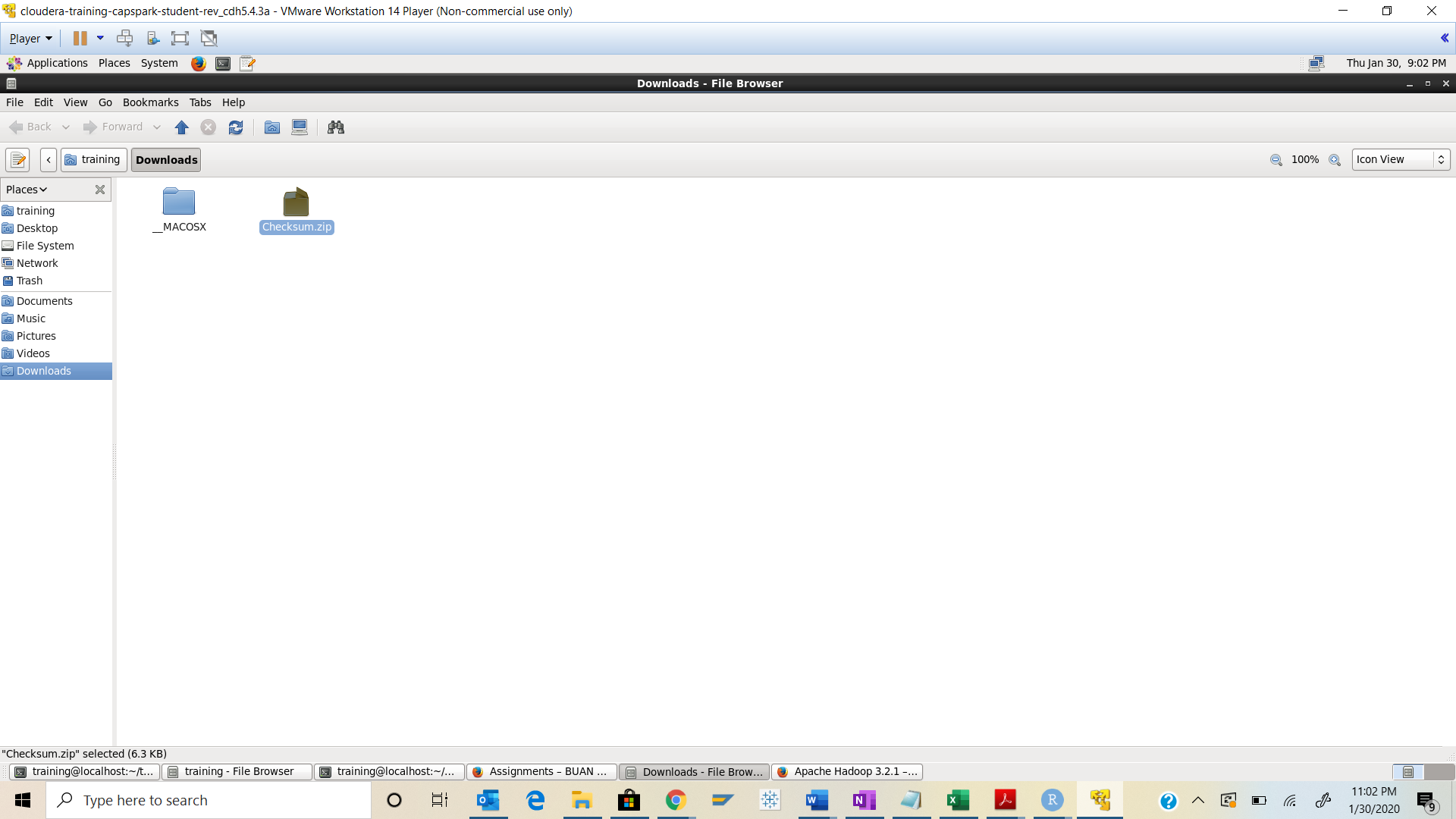
Step-21: The following link also provides additional documentation about HDFS filesystem commands.

<https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/FileSystemShell.html>

Q. Navigate to the webpage from the link provided above, take a screenshot of the landing page and paste it below.



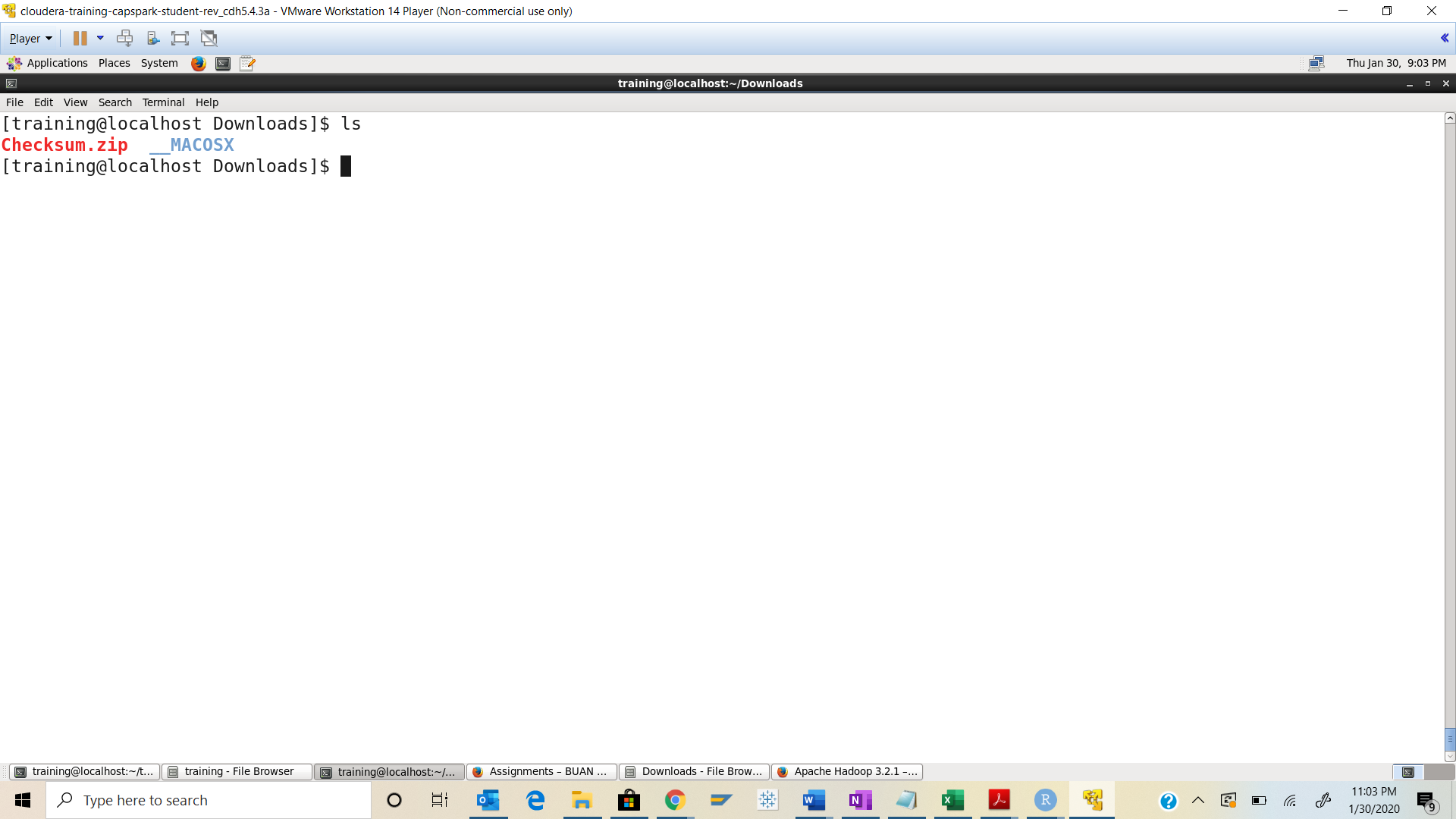
Step-22: Download the Checksum.zip file from eLearning in to the downloads folder **IN** the Cloudera VM using Mozilla Firefox.



Step-23: Execute the following command:



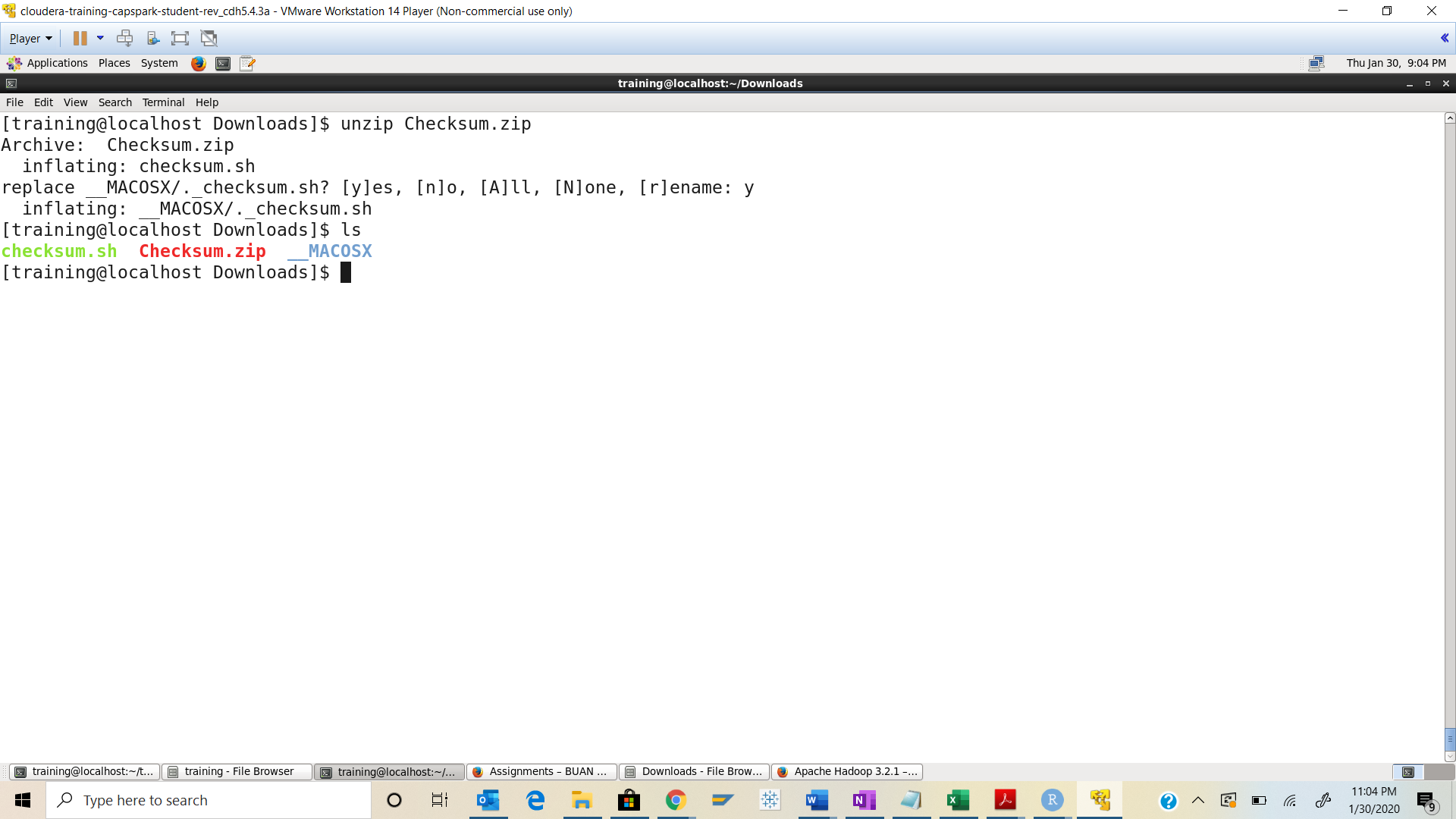
Q. Now type the command and execute. Take a screenshot of the shell output, highlight the zip file and paste it below.



Step-24: Execute the following command at the shell prompt:



Q. Now type the command and execute. Take a screenshot of the shell output, highlight the contents of the zip file and paste it below.

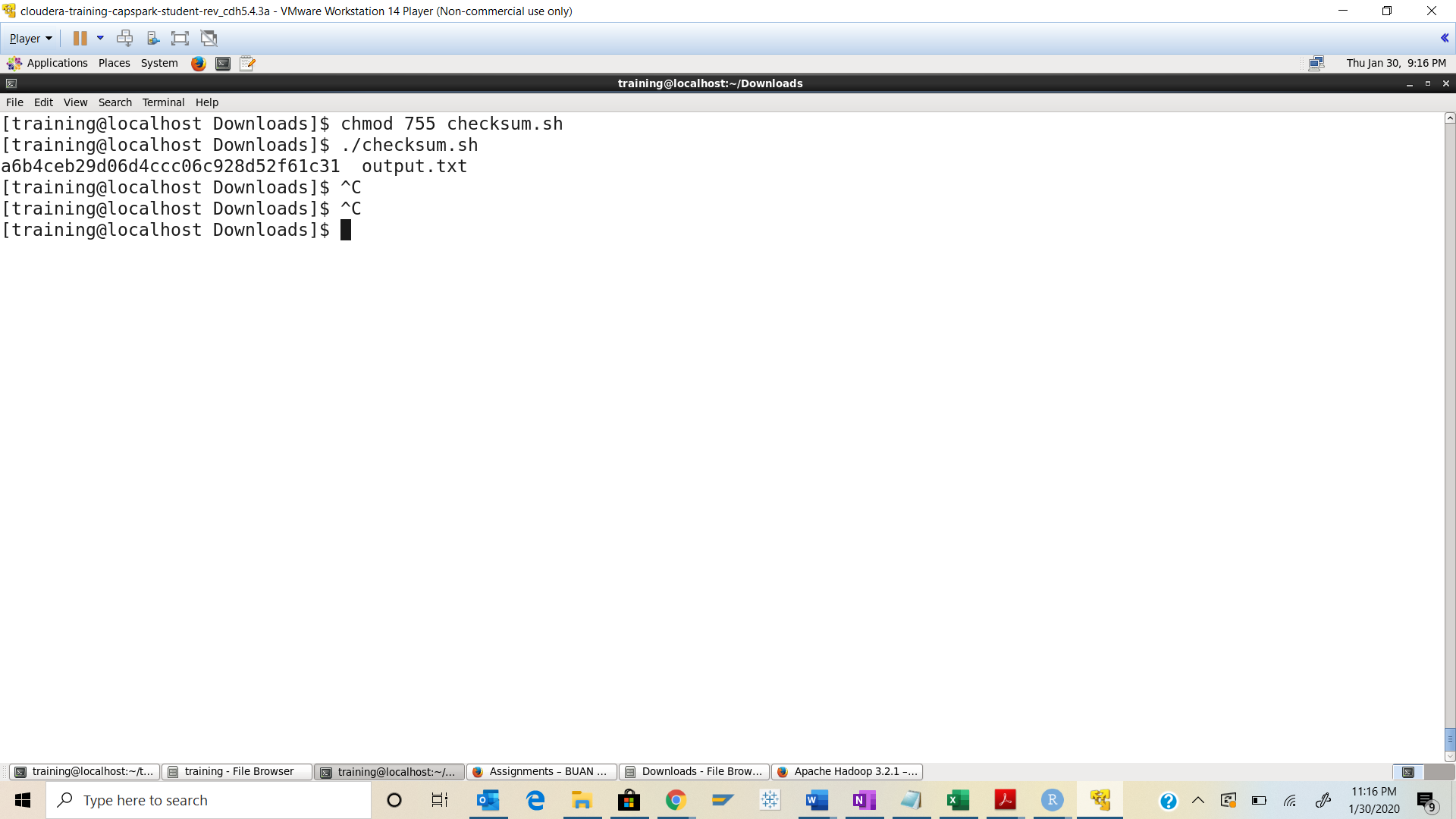


Step-25: Execute the following command at the shell prompt:

Then execute the following command:

Q. Take a screenshot of the shell output and paste it below.

Checksum value - a6b4ceb29d06d4ccc06c928d52f61c31



Step-26: Execute the following command at the shell prompt:





Q. Now type the command and execute. Take a screenshot of the shell output and paste it below.

