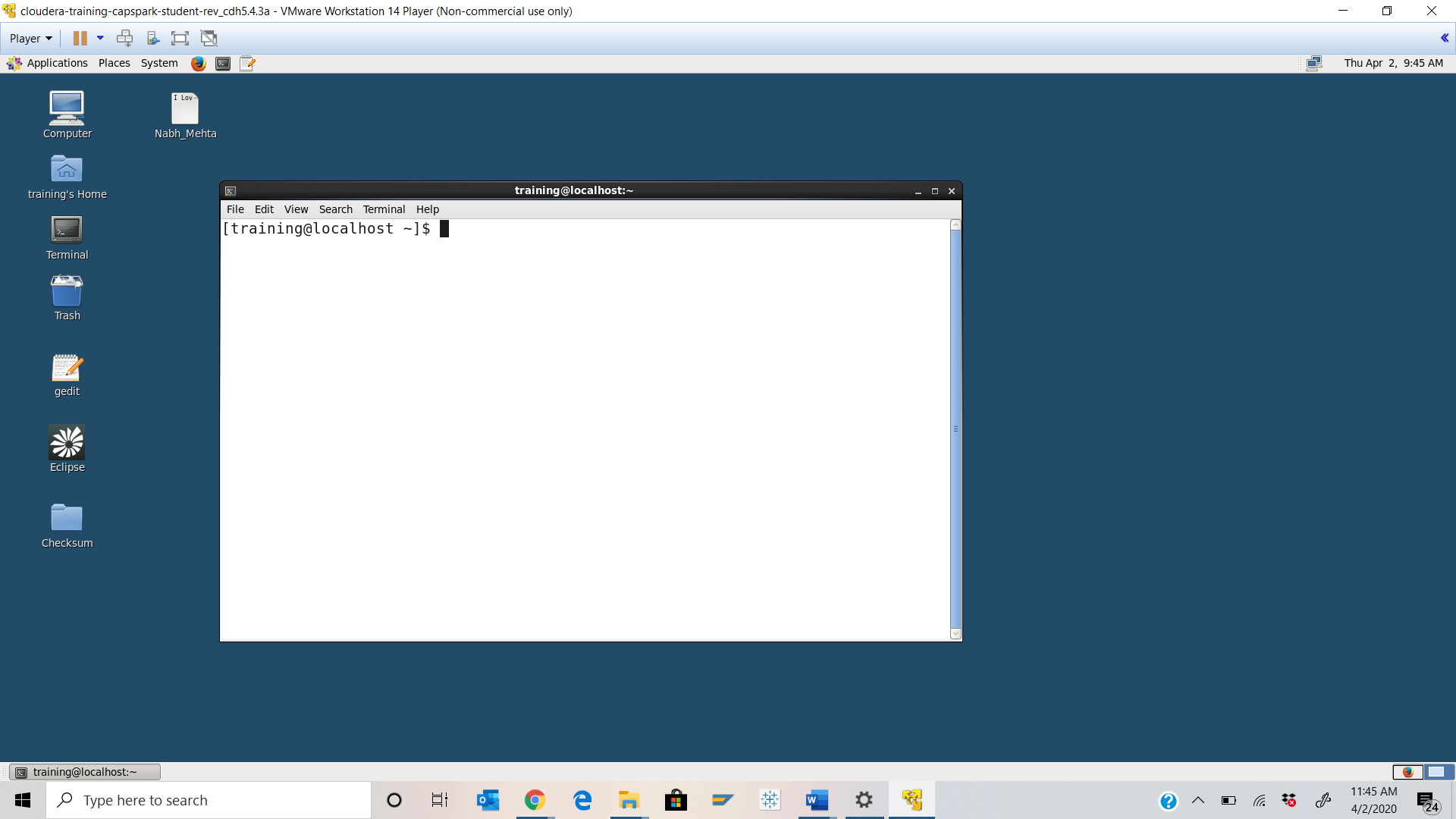
**Pig - II**

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

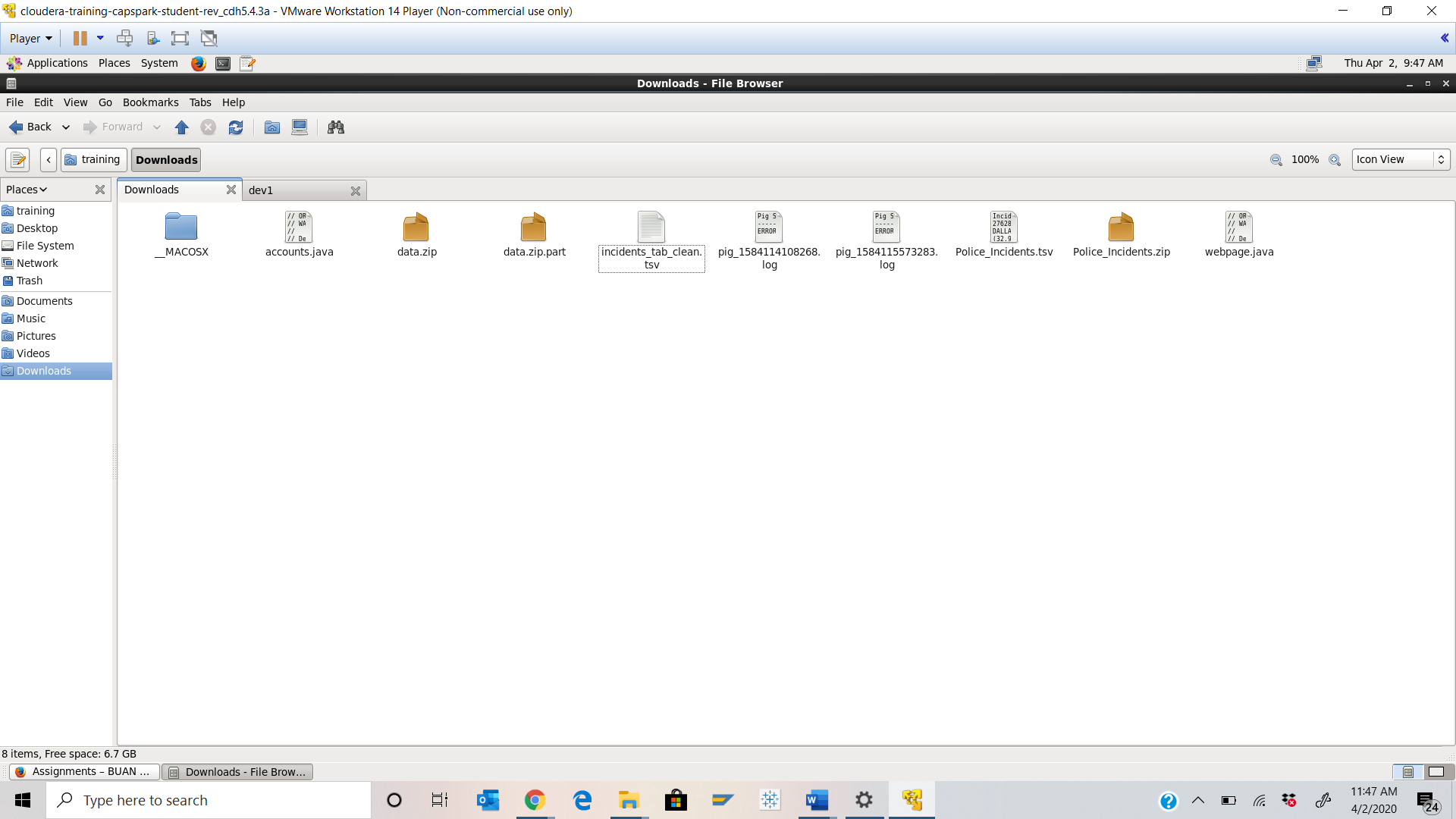
**Date: 04/02/2020**

Part-1: Pig Latin: In this section we will setup the necessary data files for this assignment.

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



Step-2: Download the data.zip file from eLearning into the downloads folder IN the Cloudera VM using Mozilla Firefox.



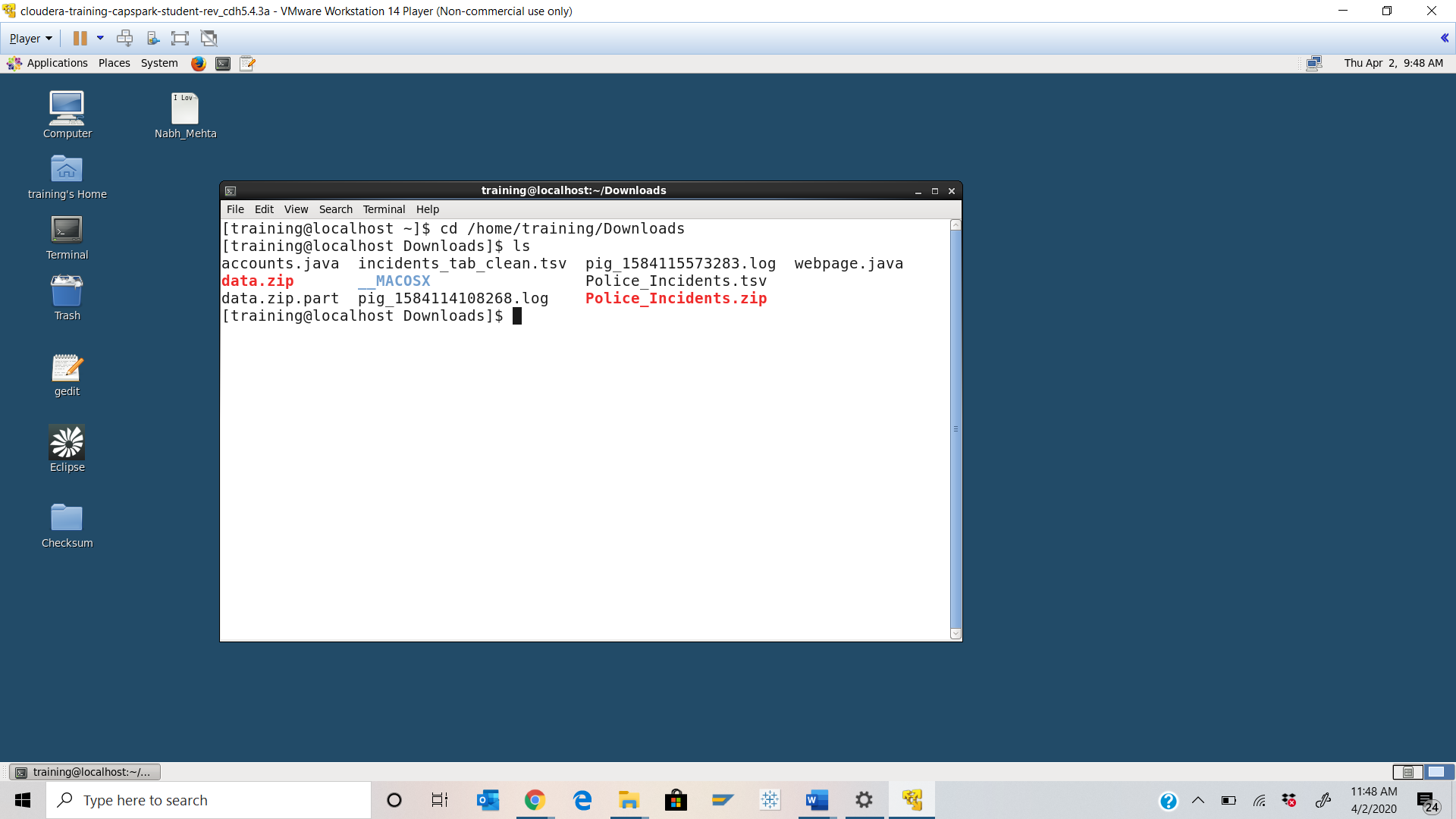


Step-3: Execute the following command:





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

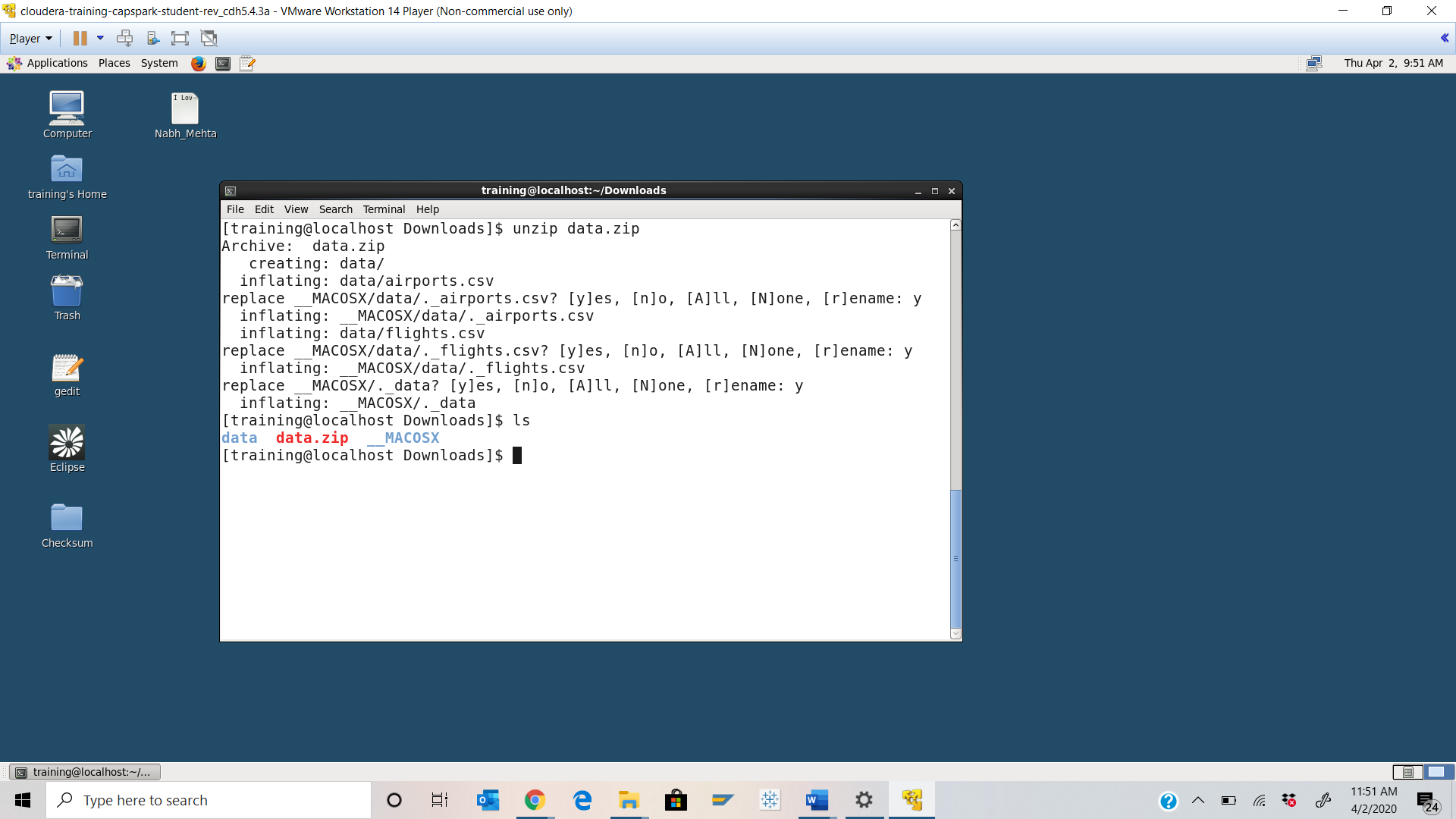




Step-4: Execute the following command at the shell prompt:



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.

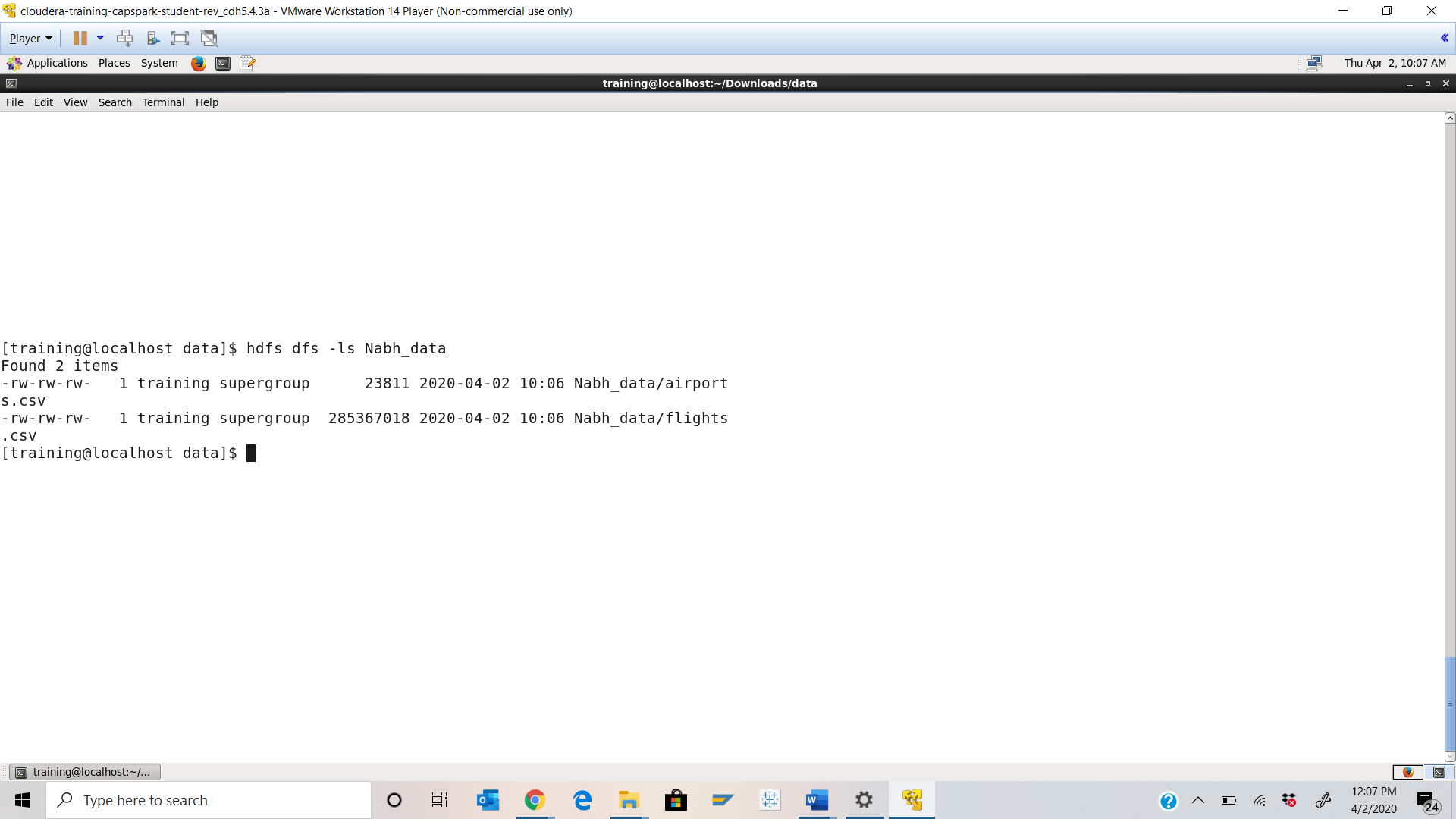




Step-5: Next place the data into Hadoop by executing below 3 commands at the shell prompt



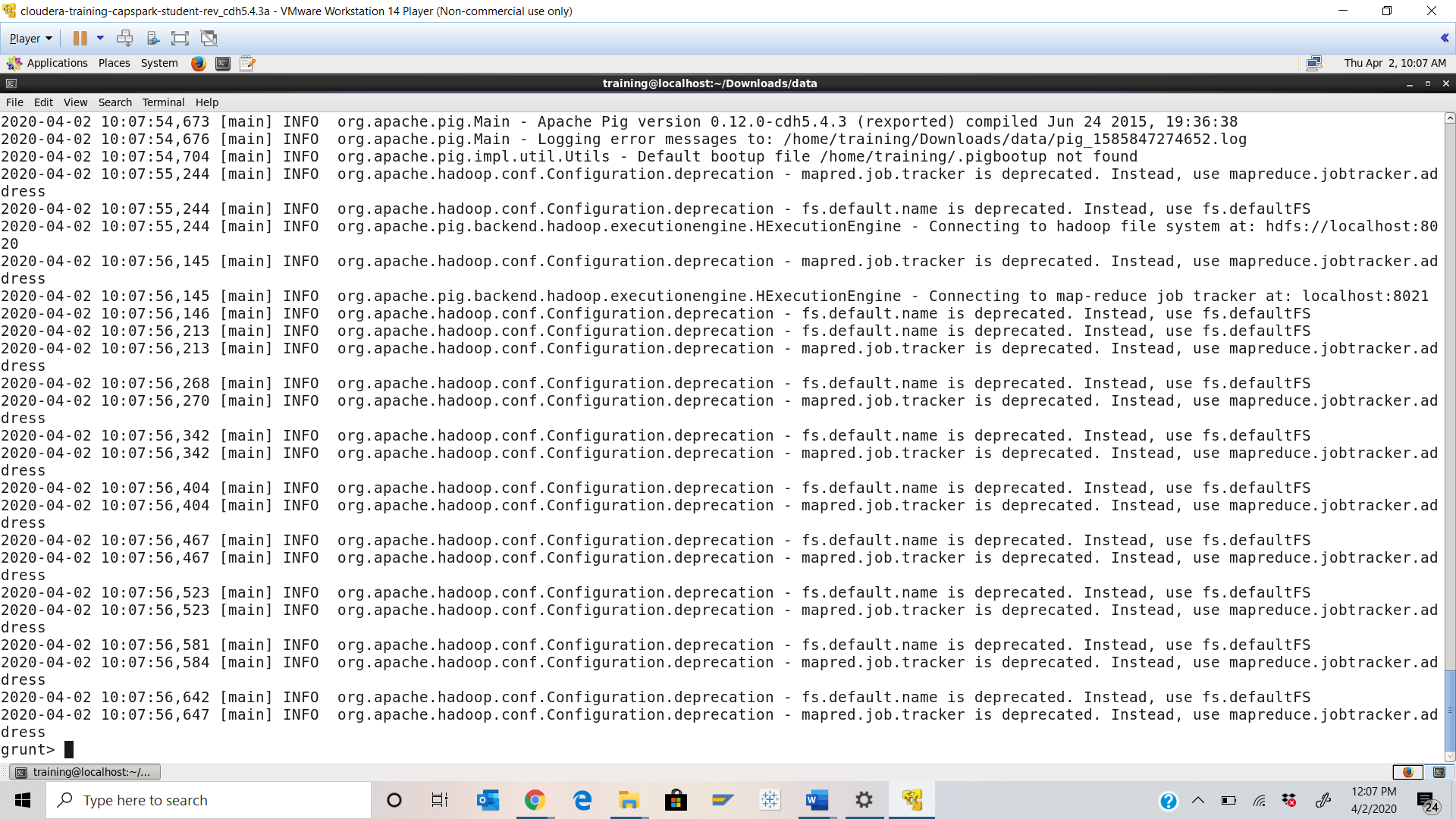
Q. Now type the below command and execute. Take a screenshot of the shell output and paste it below. 



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



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

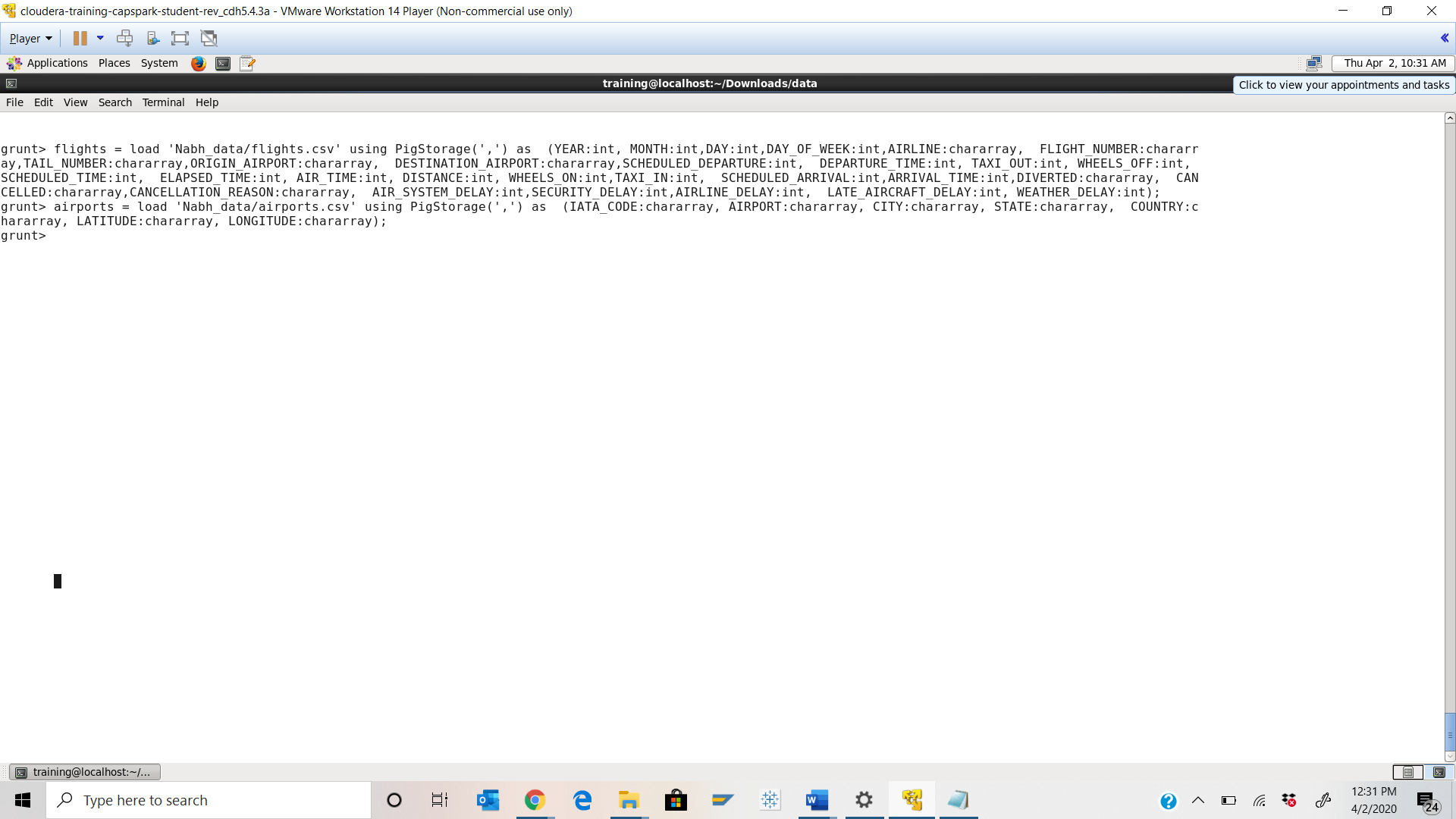


Step-7: Execute the following 2 command at the shell prompt:





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

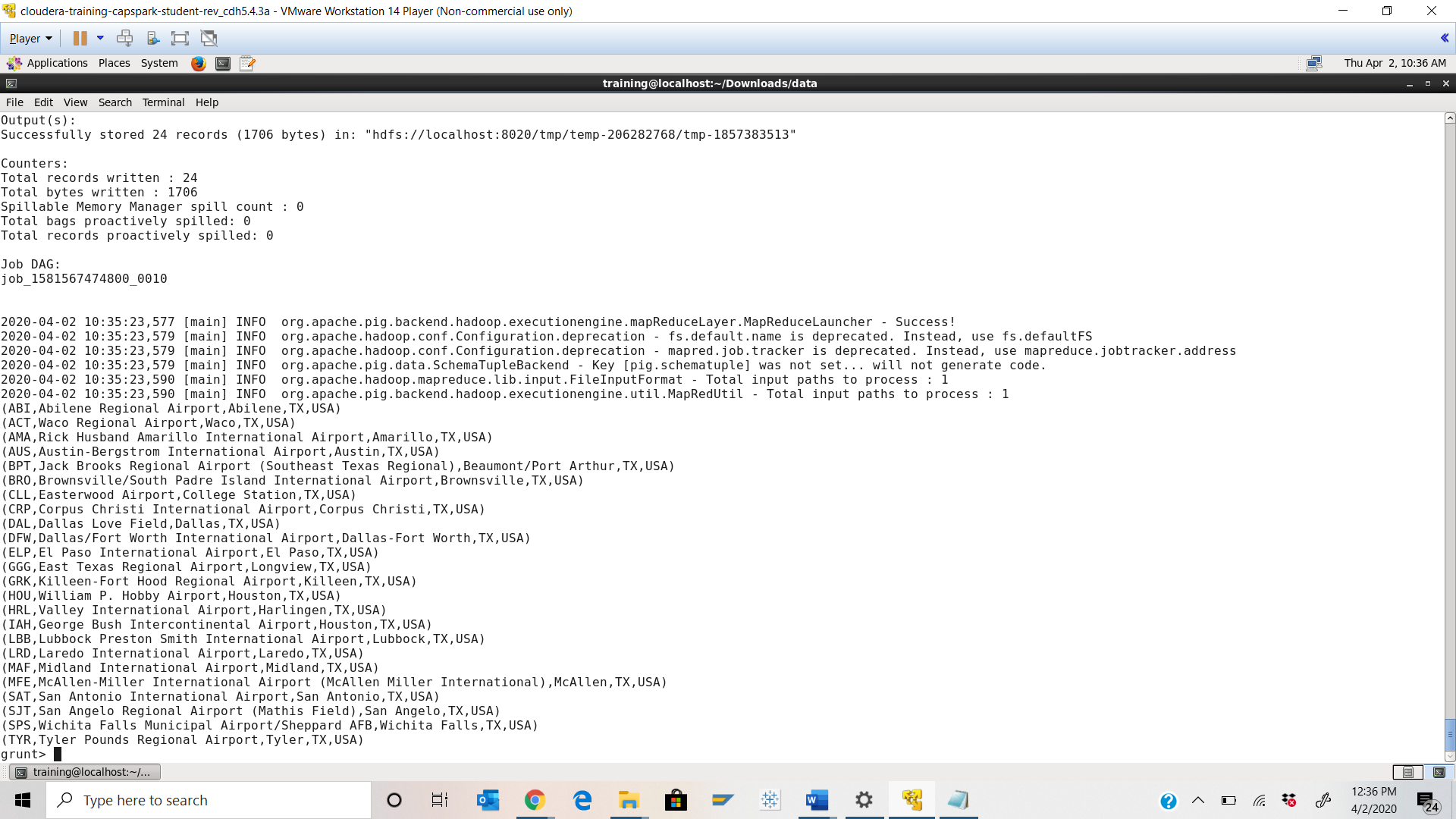


Step-8: Execute the following command at the shell prompt:



Q. Now type the below 2 command and execute. Take a screenshot of the shell output and paste it below. How many numbers of airports are there in Texas?





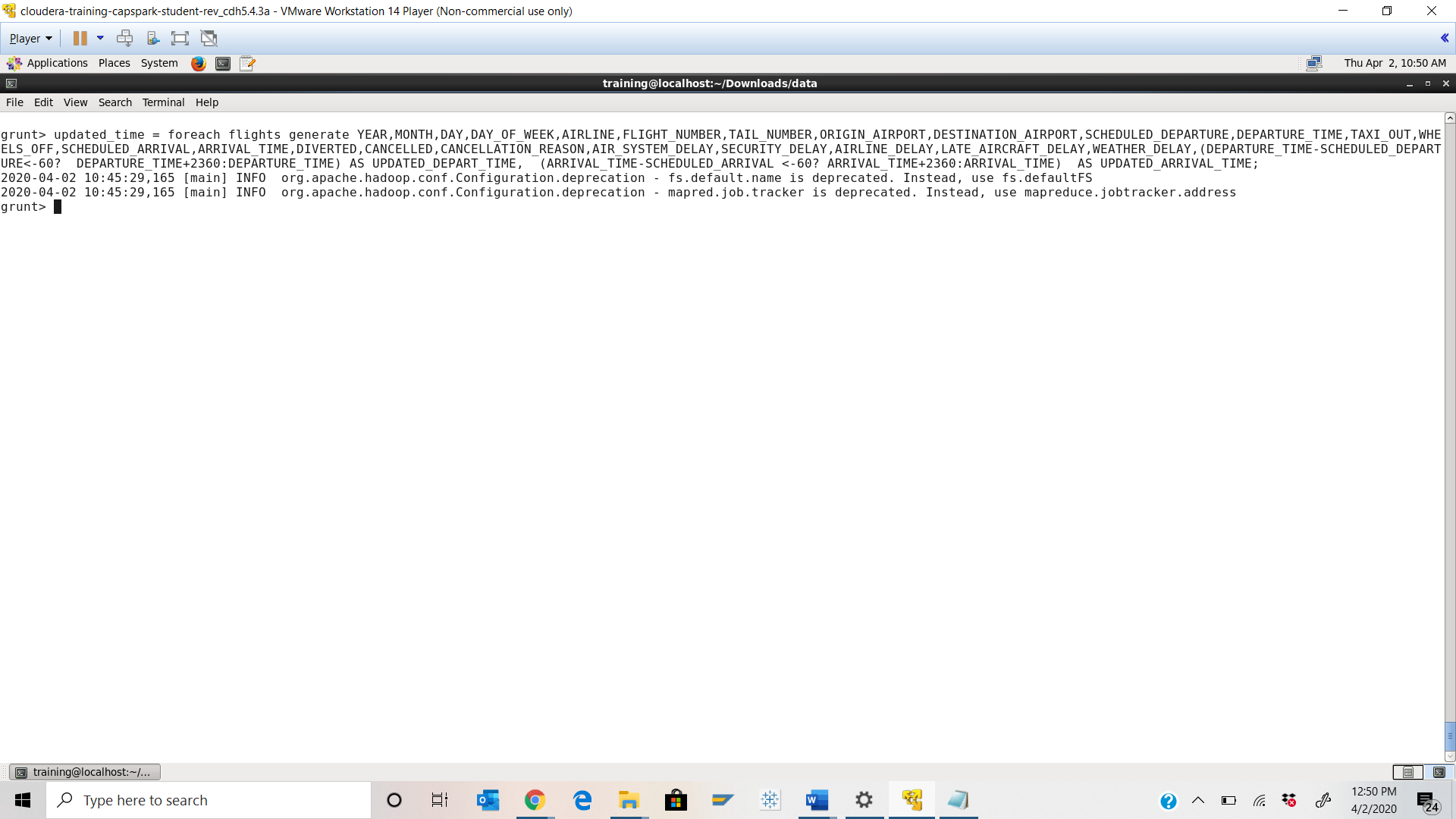


Ans: **24 airports**

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



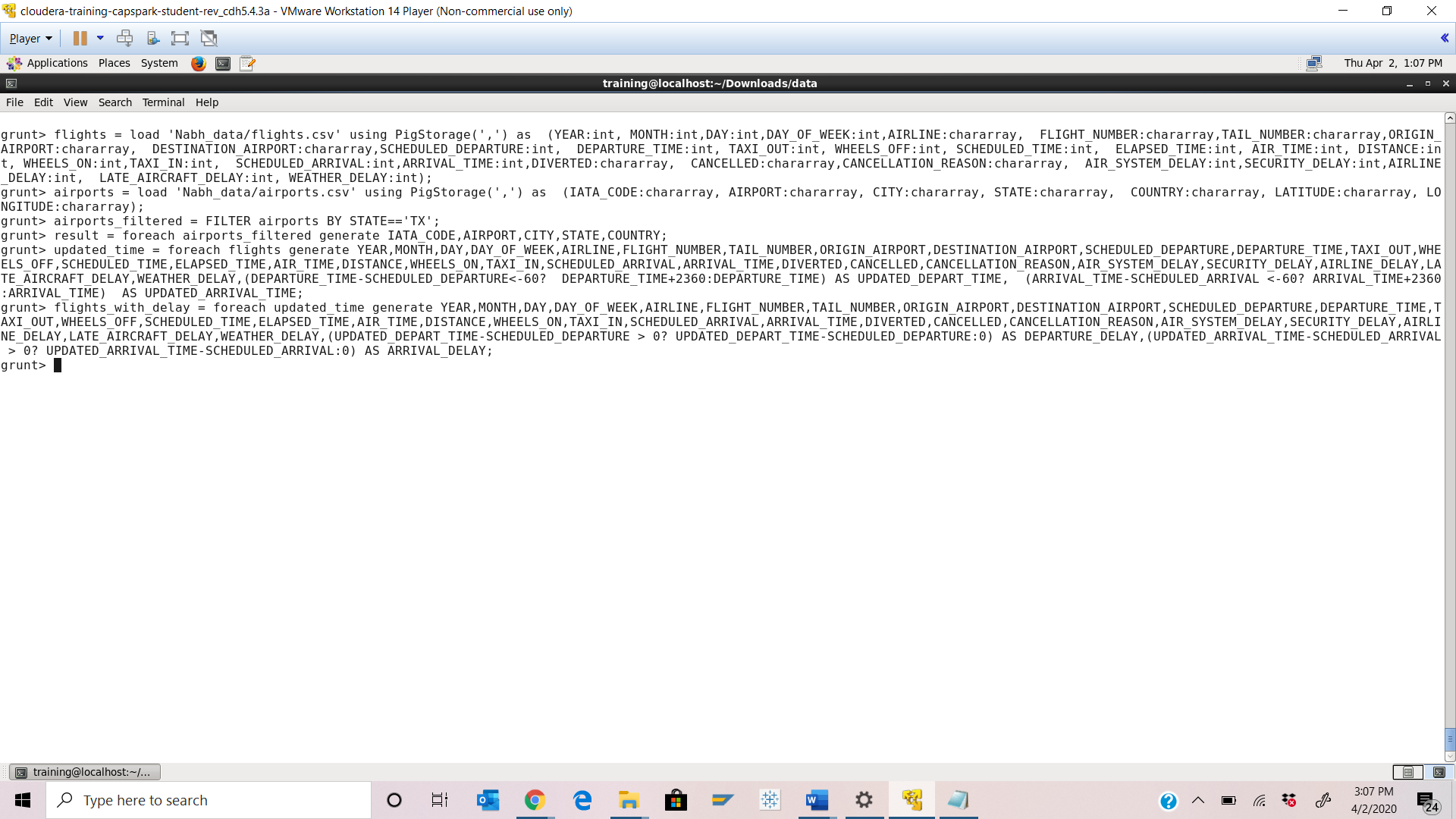
Q. Take a screenshot of the shell output and paste it below.



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



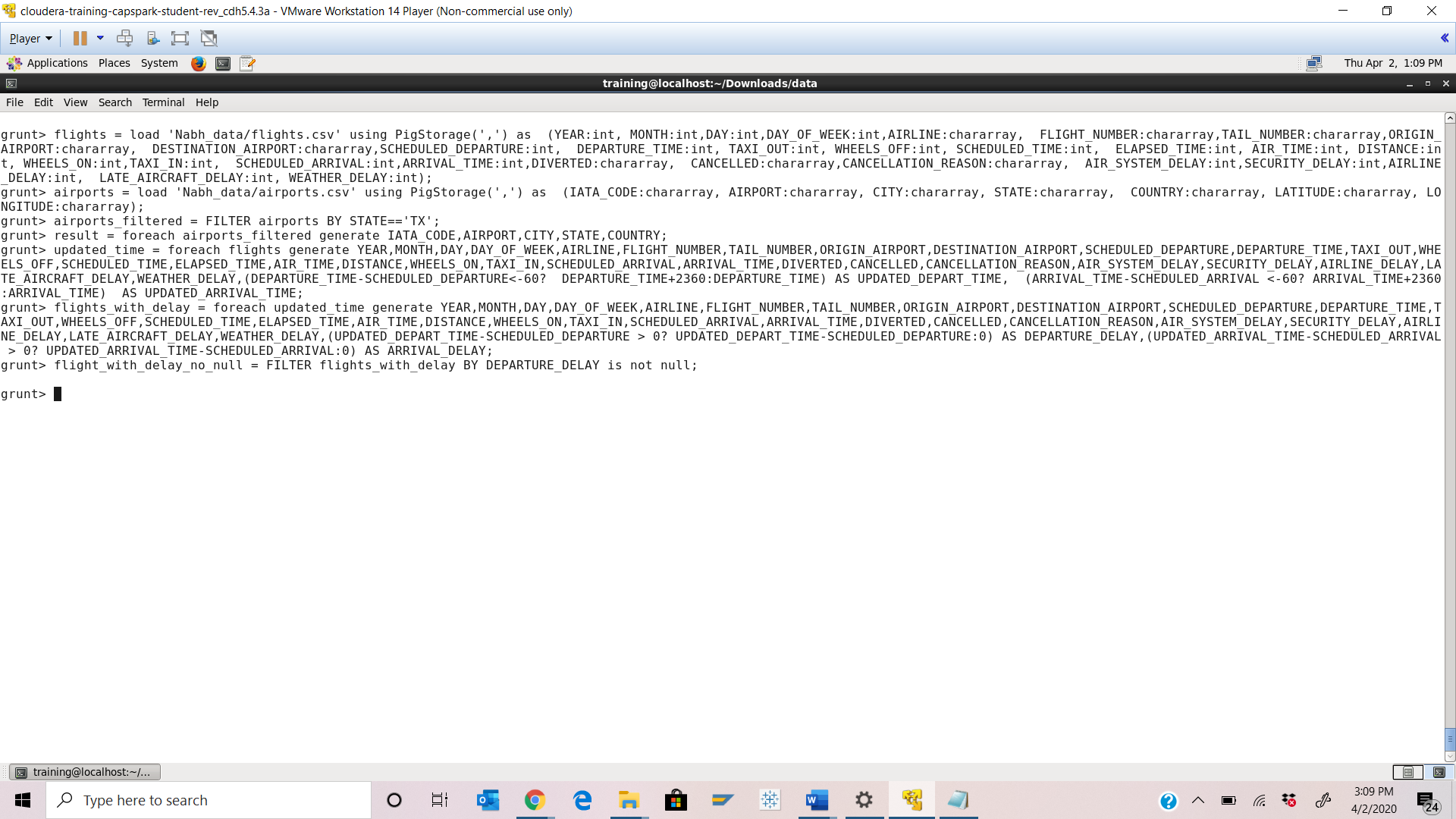
Q. Take a screenshot of the shell output and paste it below.



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



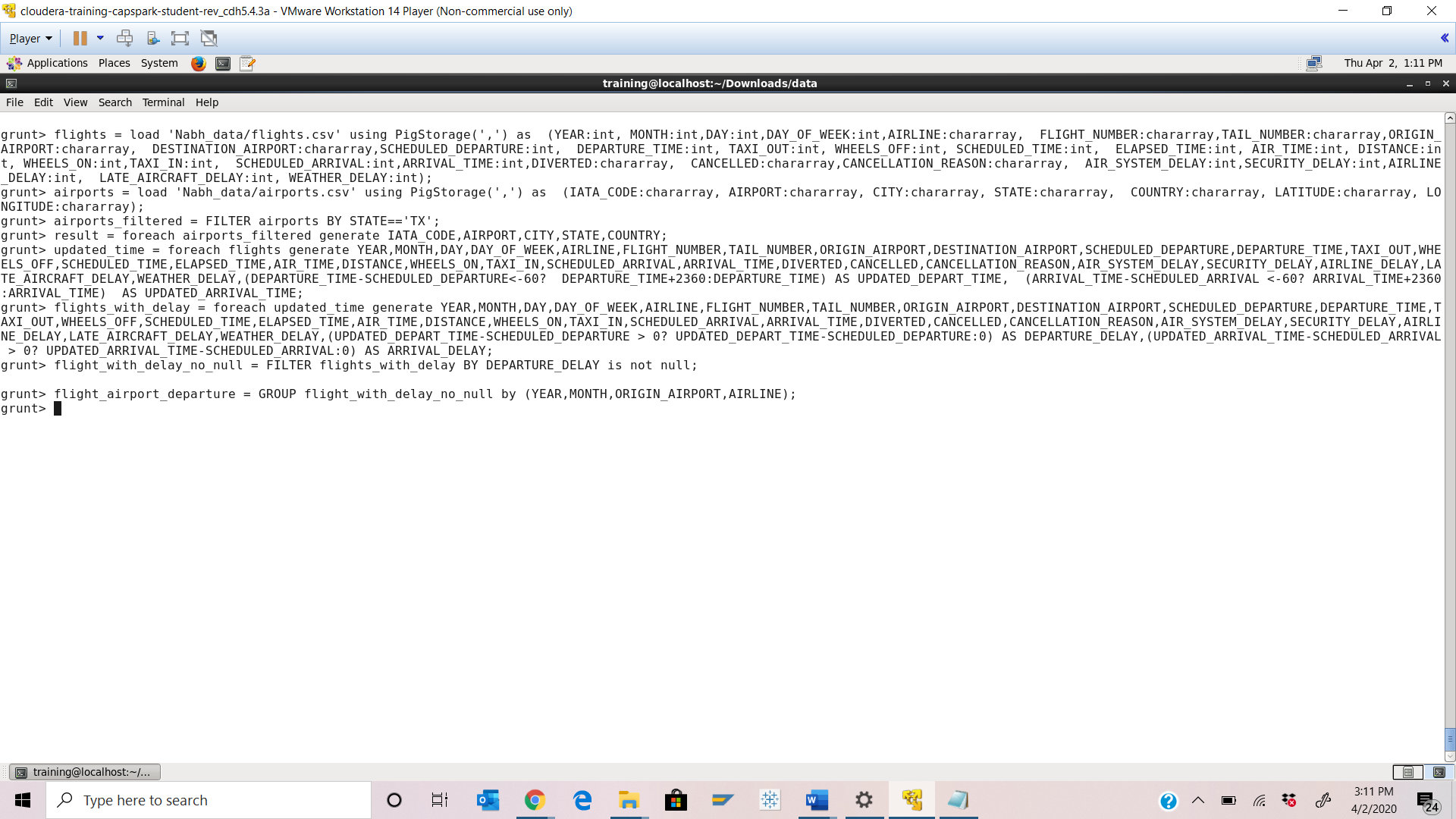
Q. Take a screenshot of the shell output and paste it below.



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



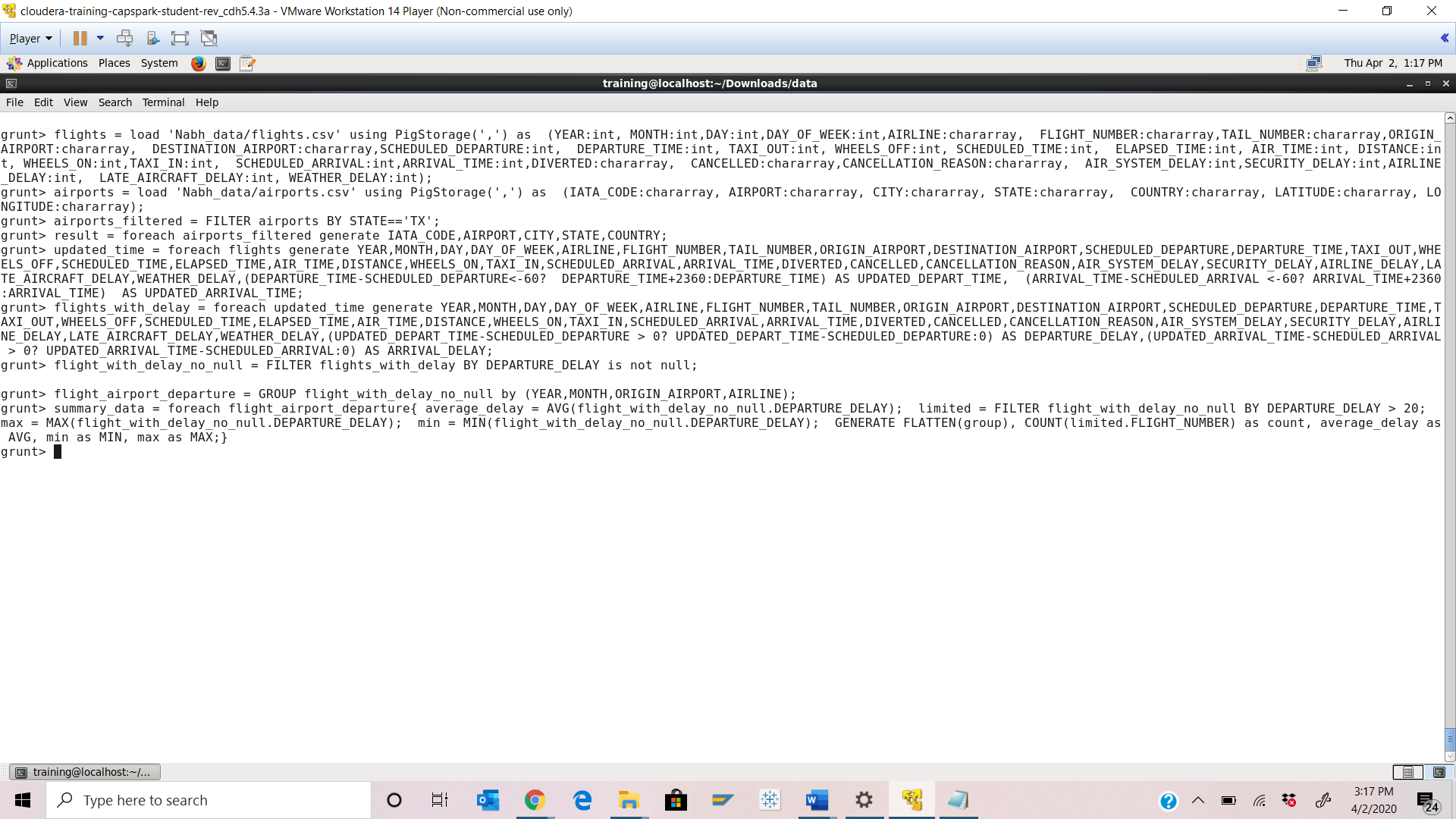
Q. Take a screenshot of the shell output and paste it below.



Step-13: Execute the following command at the shell prompt:



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

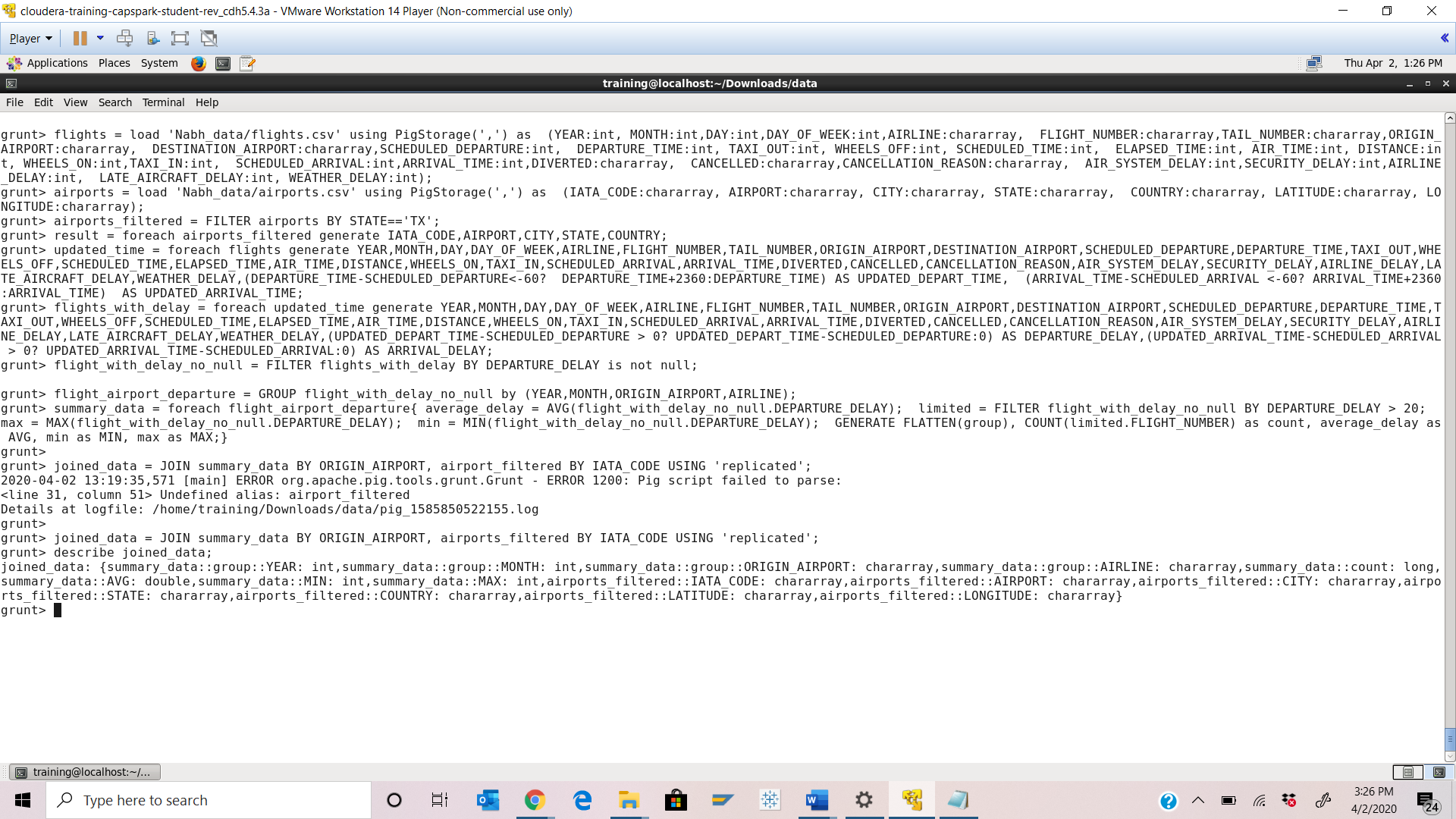


Step-14: Execute the following command at the shell prompt:



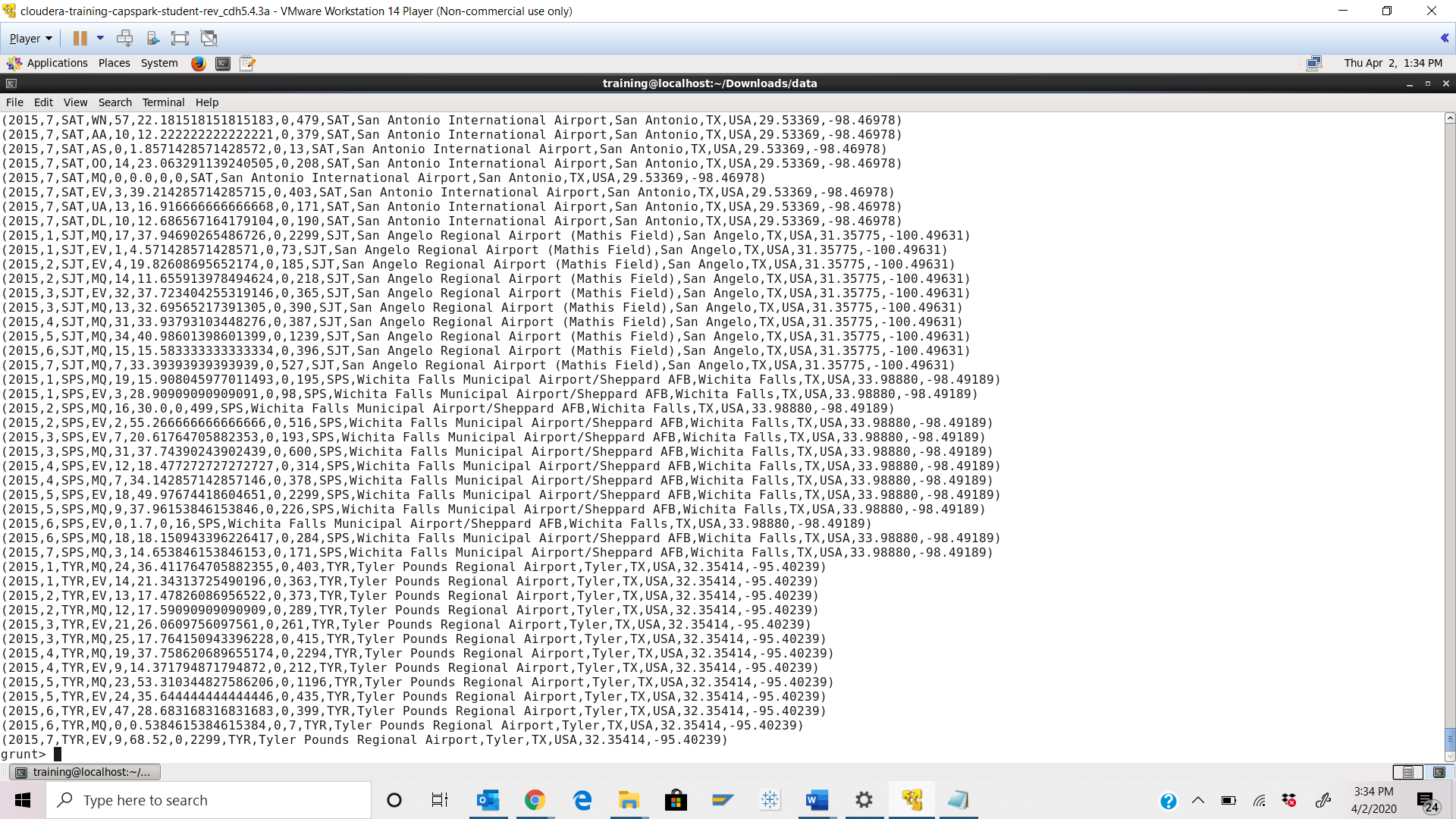


Q. Take a screenshot of the shell output and paste it below.



Step-15: Execute the following 2 command at the shell prompt:

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

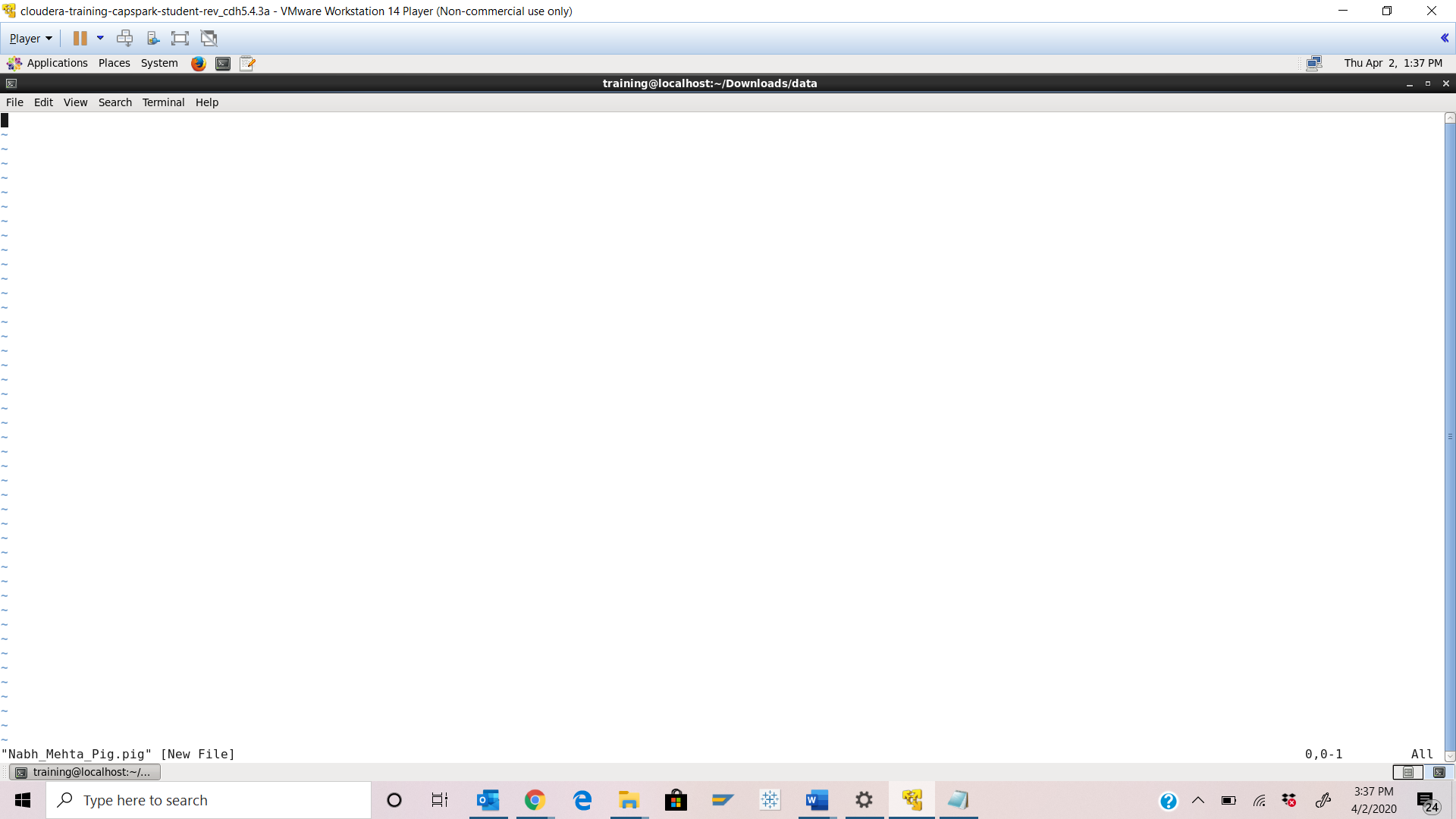


Part-2: Pig Script: In this section we are going to create a pig script

Step-16:Create a new pig file in downloads folder. Press <Ctrl+d>. Then type below command on shell prompt:



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



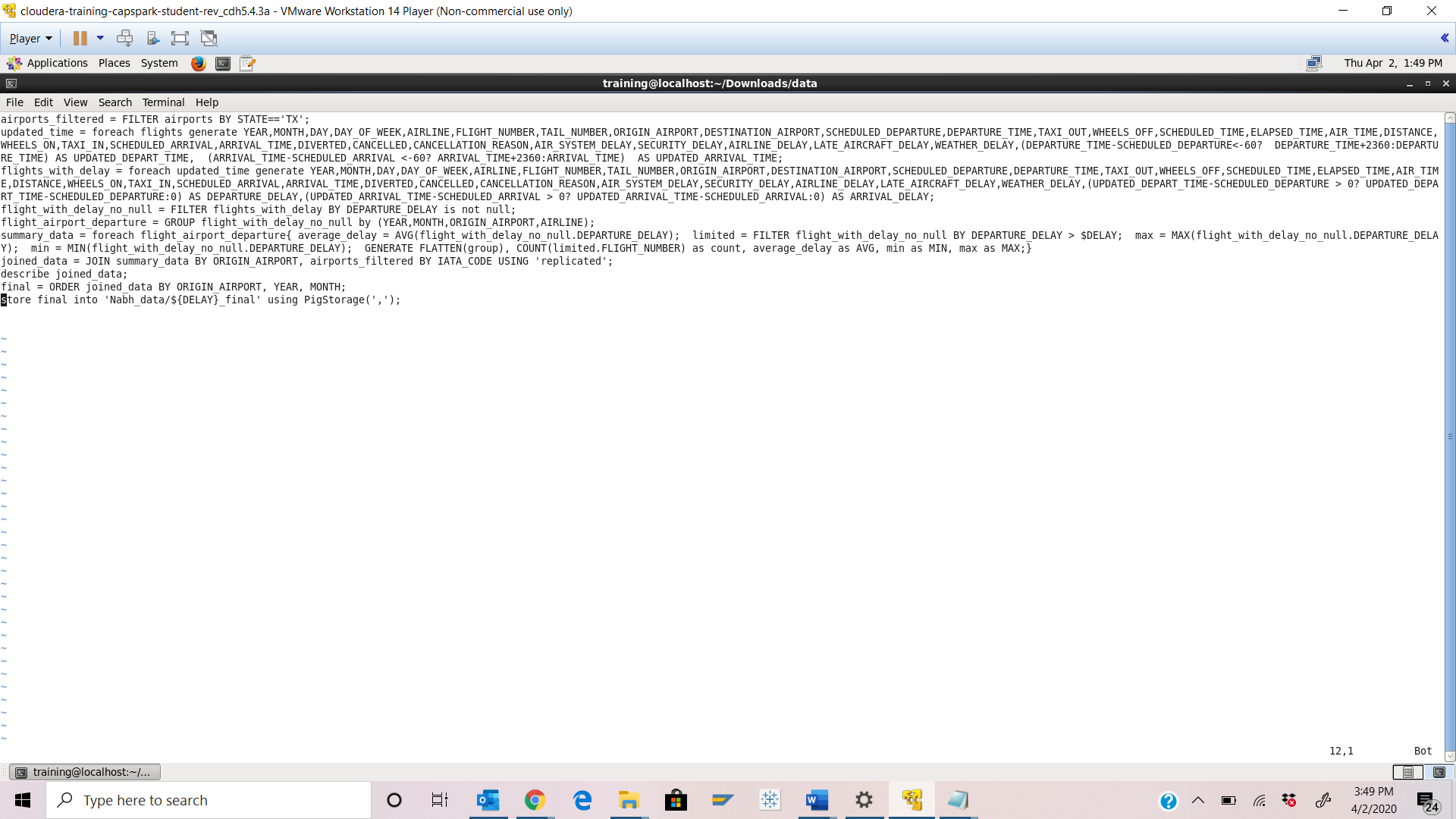
Q. Take a screenshot of the shell output and paste it below.

Step-17: Press i to insert. Copy the code from Step 7 – 14 into the file. Replace value of 20 in nested foreach by $DELAY.

Then replace the dump line by below code.



Q. Take a screenshot of the shell output and paste it below. The entire script should be visible in snapshot.

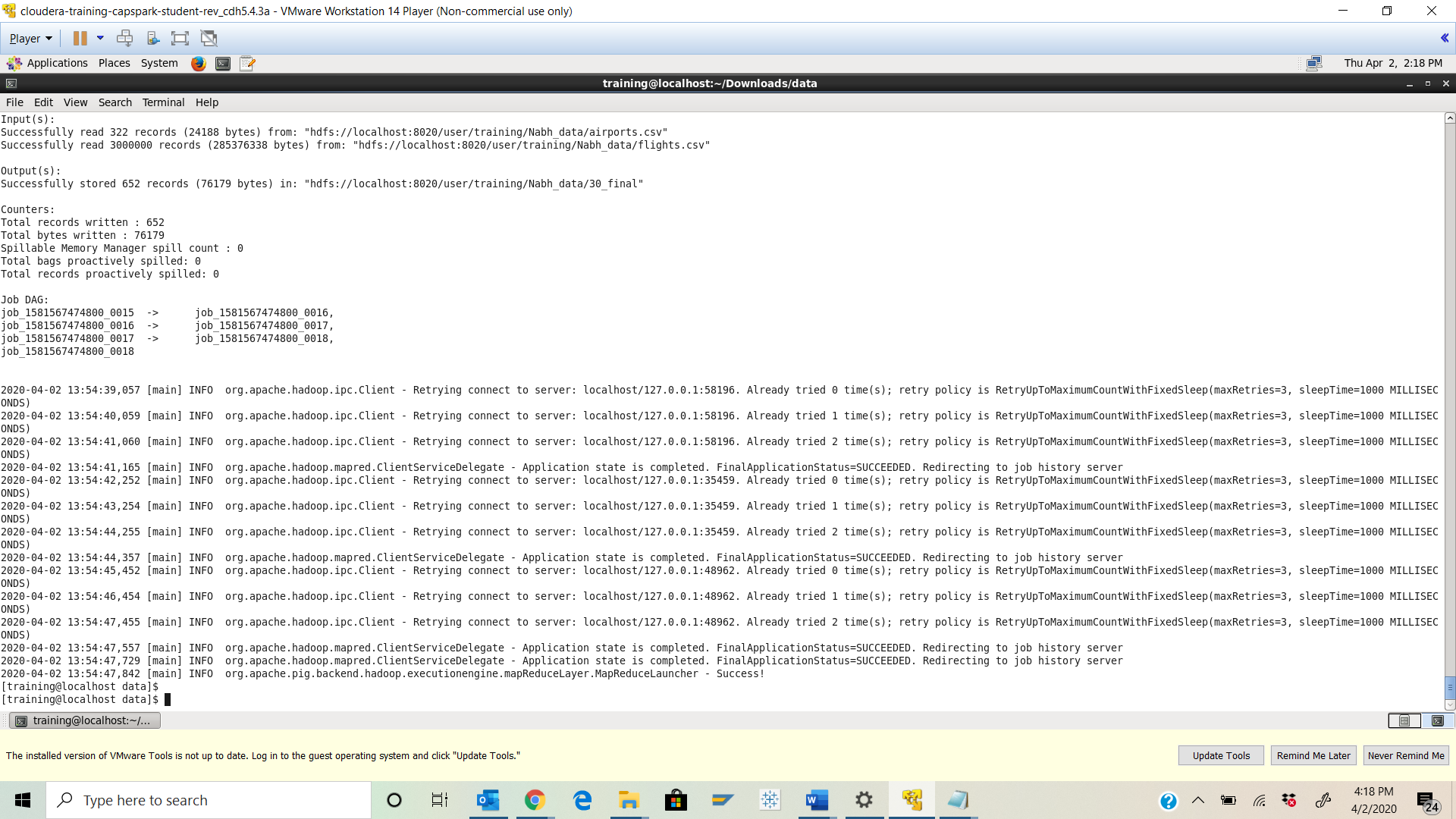


Step-18: Execute the following command at the shell prompt:





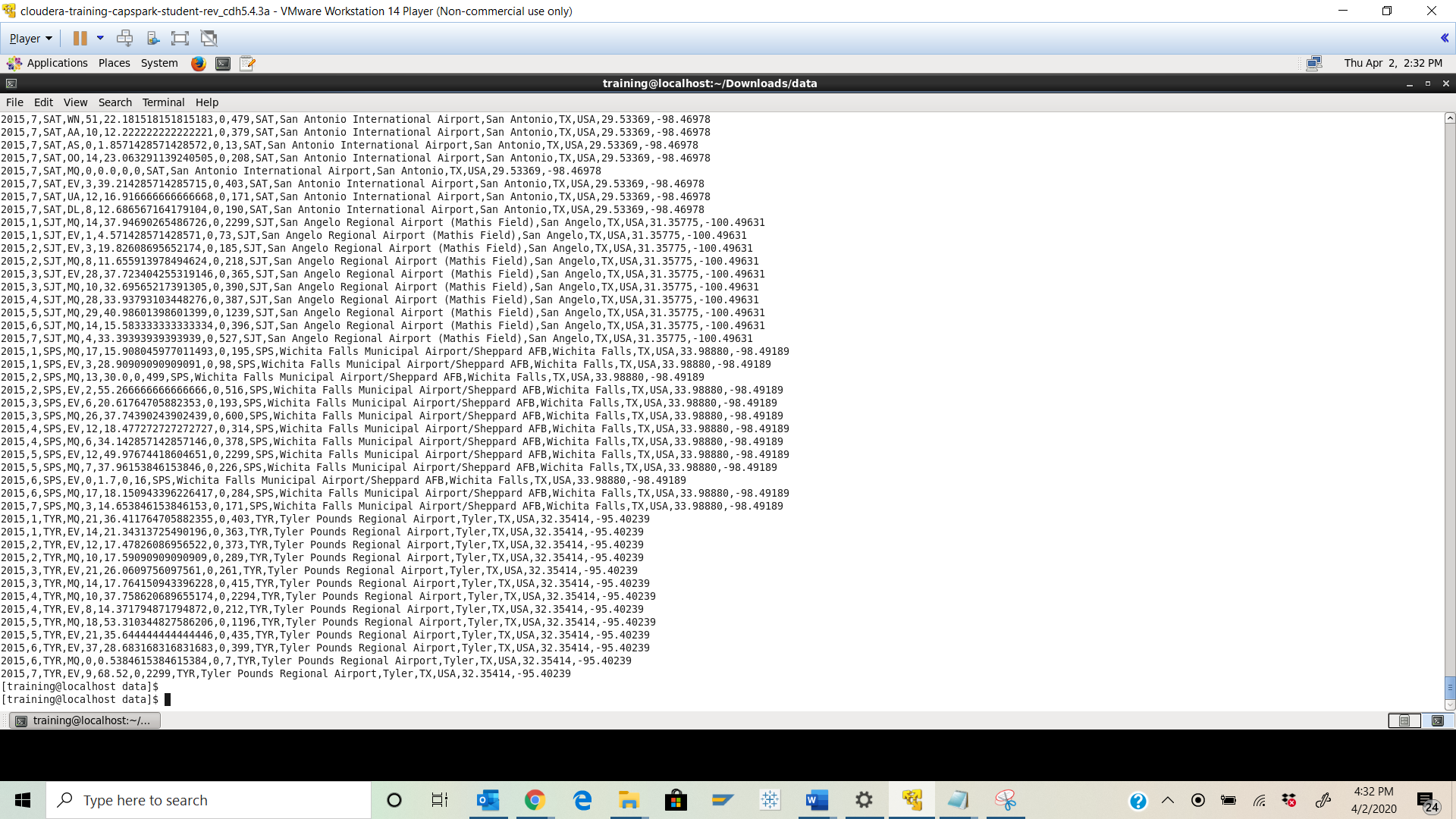
Q. Take a screenshot of the shell output and paste it below. Describe the meaning of the output that you see ?



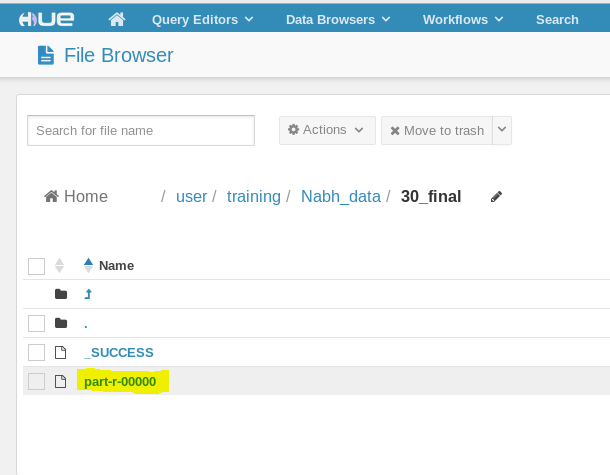
Step-19: Execute the following command at the shell prompt:



Q. Take a screenshot of the shell output and paste it below. Describe the meaning of the output that you see ?



Explanation: As the output contains flight departures greater than 30 mins, using -cat command we have copied desired output to a new file with initials ‘part’. This can also be confirmed in HUE file browser.



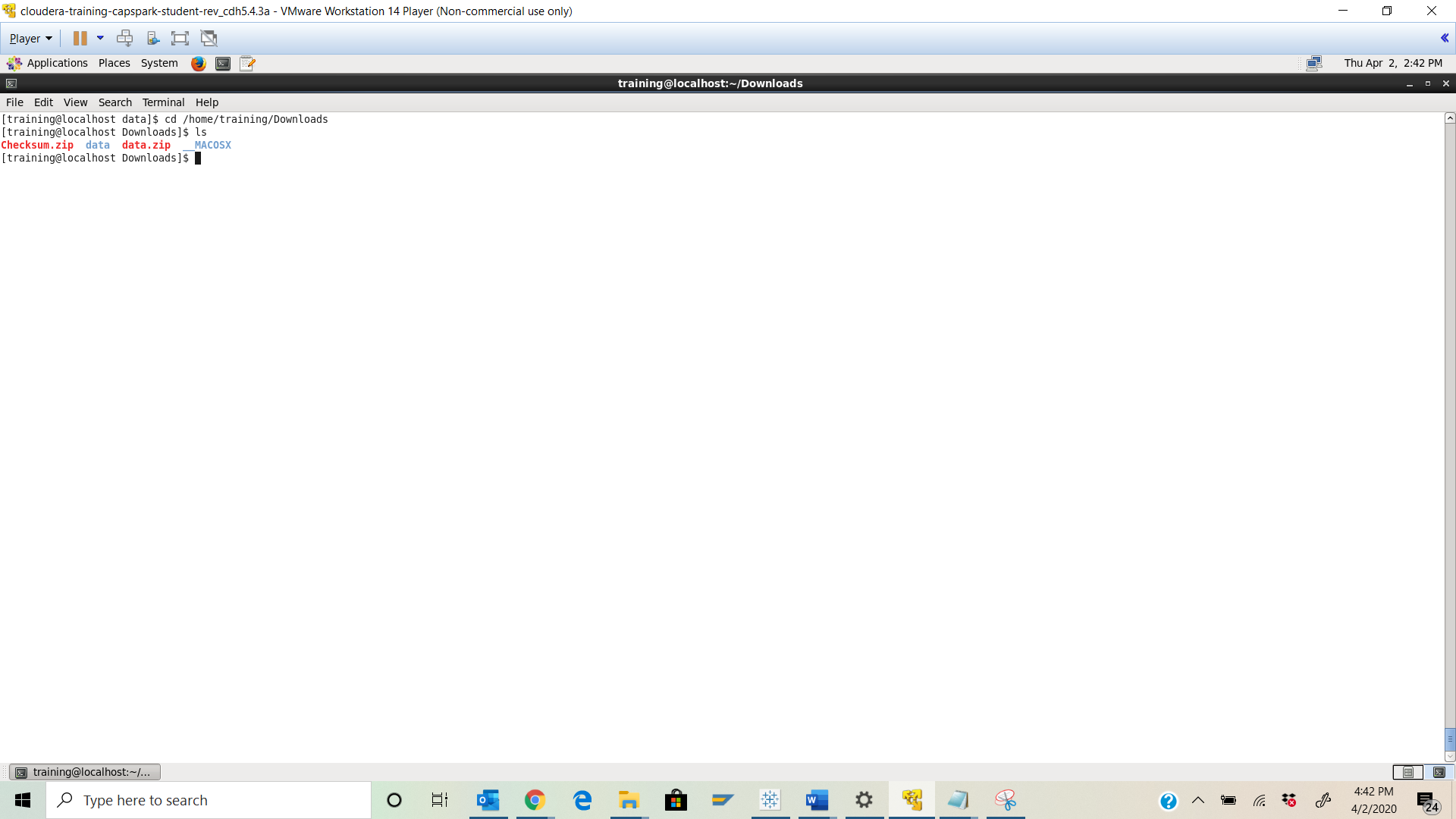
Part-3: Checksum: In this section we will perform a checksum

Step-20: Download the Checksum.zip file from eLearning in to the downloads folder **IN** the Cloudera VM using Mozilla Firefox.

Step-21: 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.

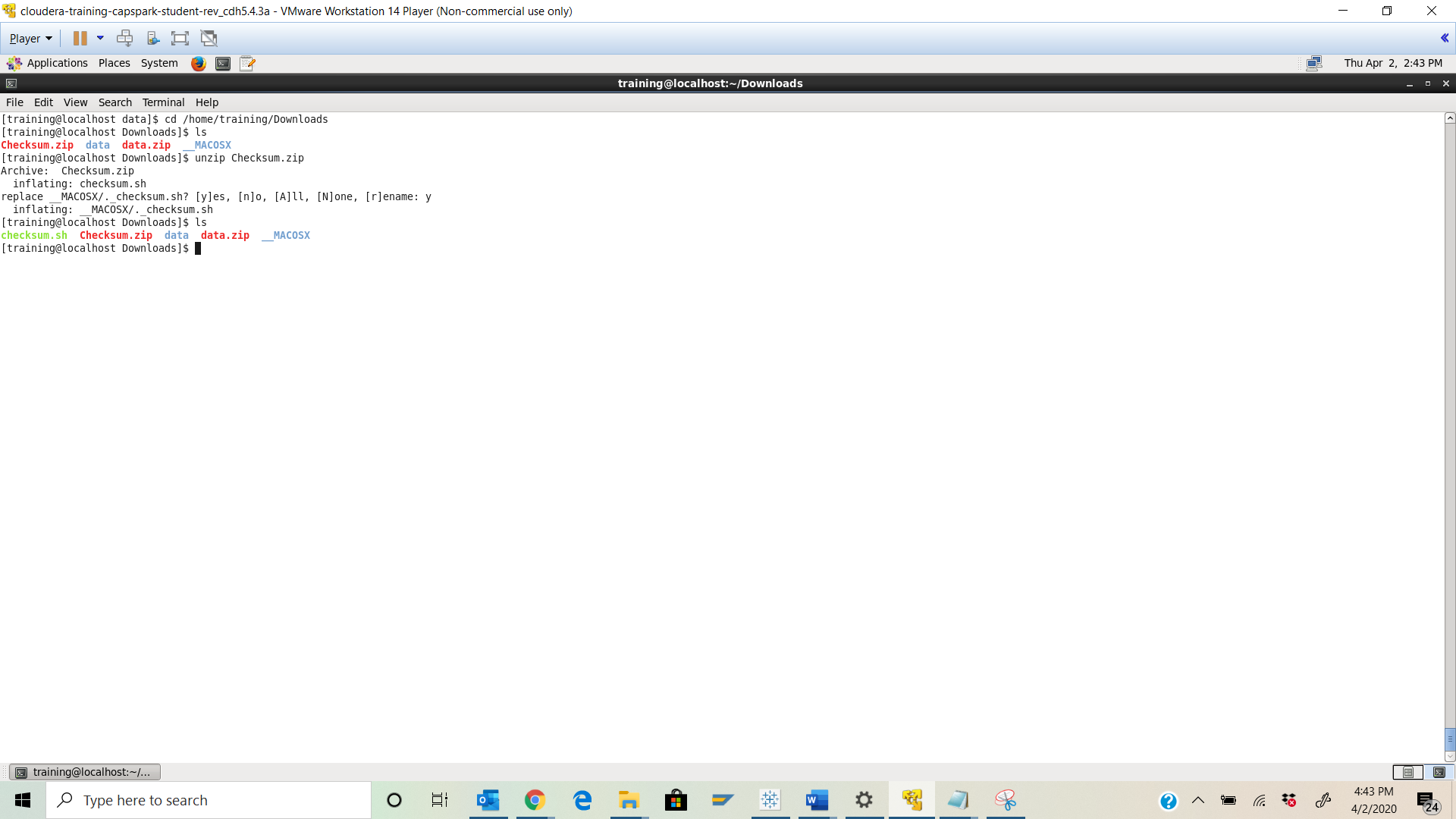




Step-22: Execute the following command at the shell prompt:



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.



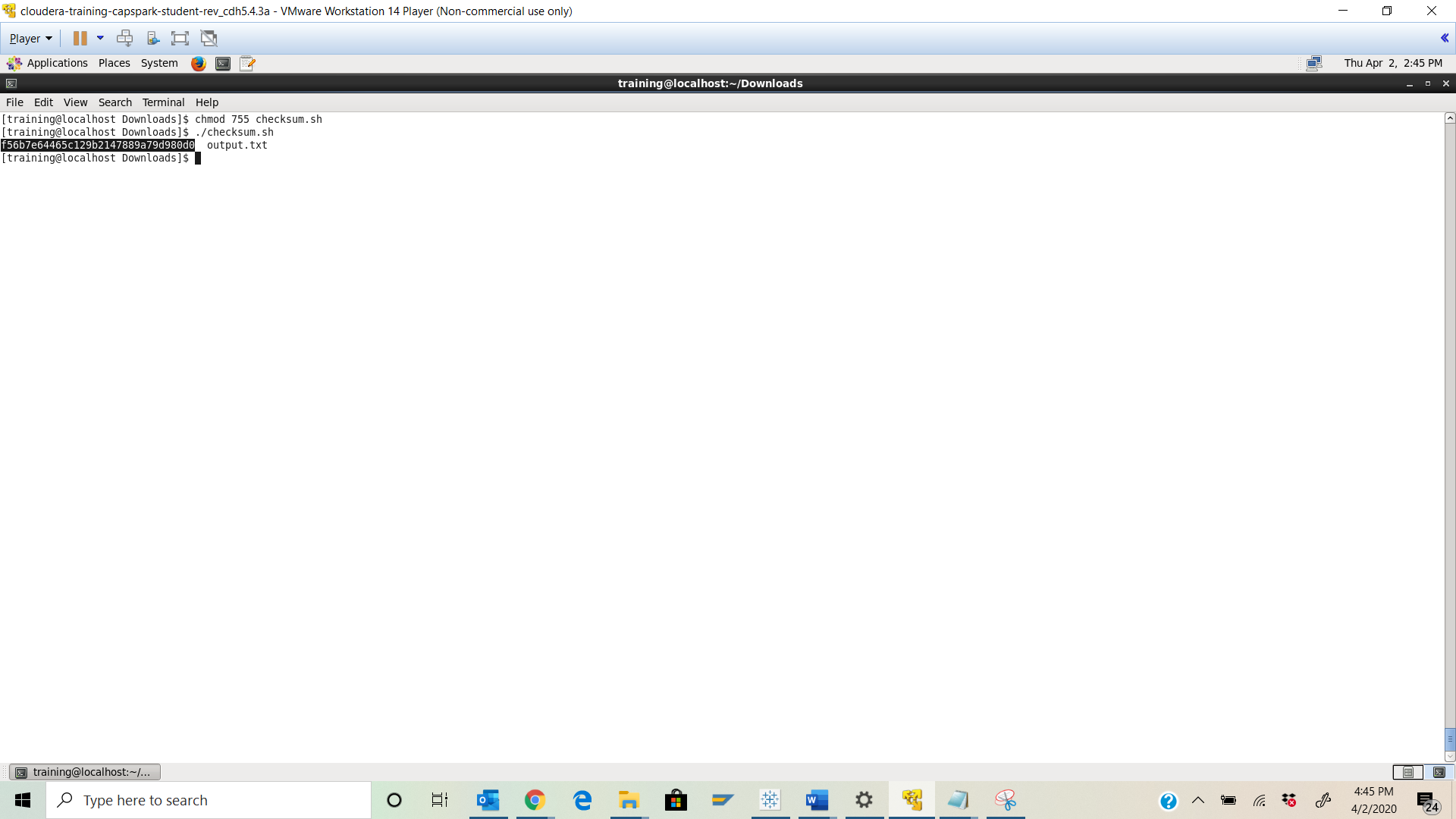


Step-23: Execute the following command at the shell prompt:

Then execute the following command:

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

Checksum value - f56b7e64465c129b2147889a79d980d0

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

rm Checksum.zip checksum.sh output.txt



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

