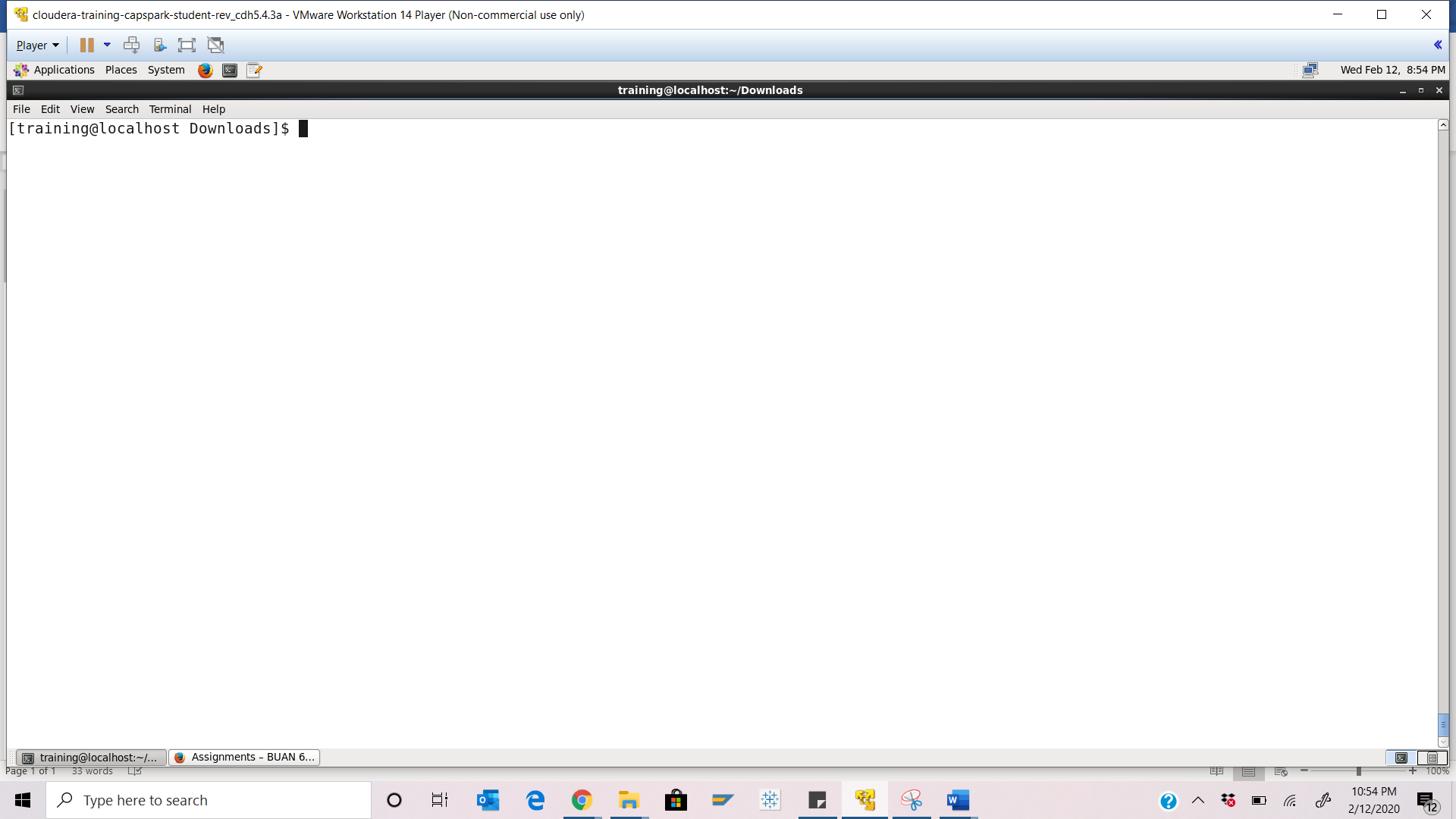
**Sqoop**

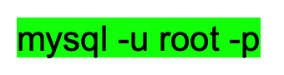
**Student Name: Nabh Sanjay Mehta Student Id: NSM190002**

**Date: 02/12/2020**

Step-1: Open the terminal application on the Cloudera VM desktop.

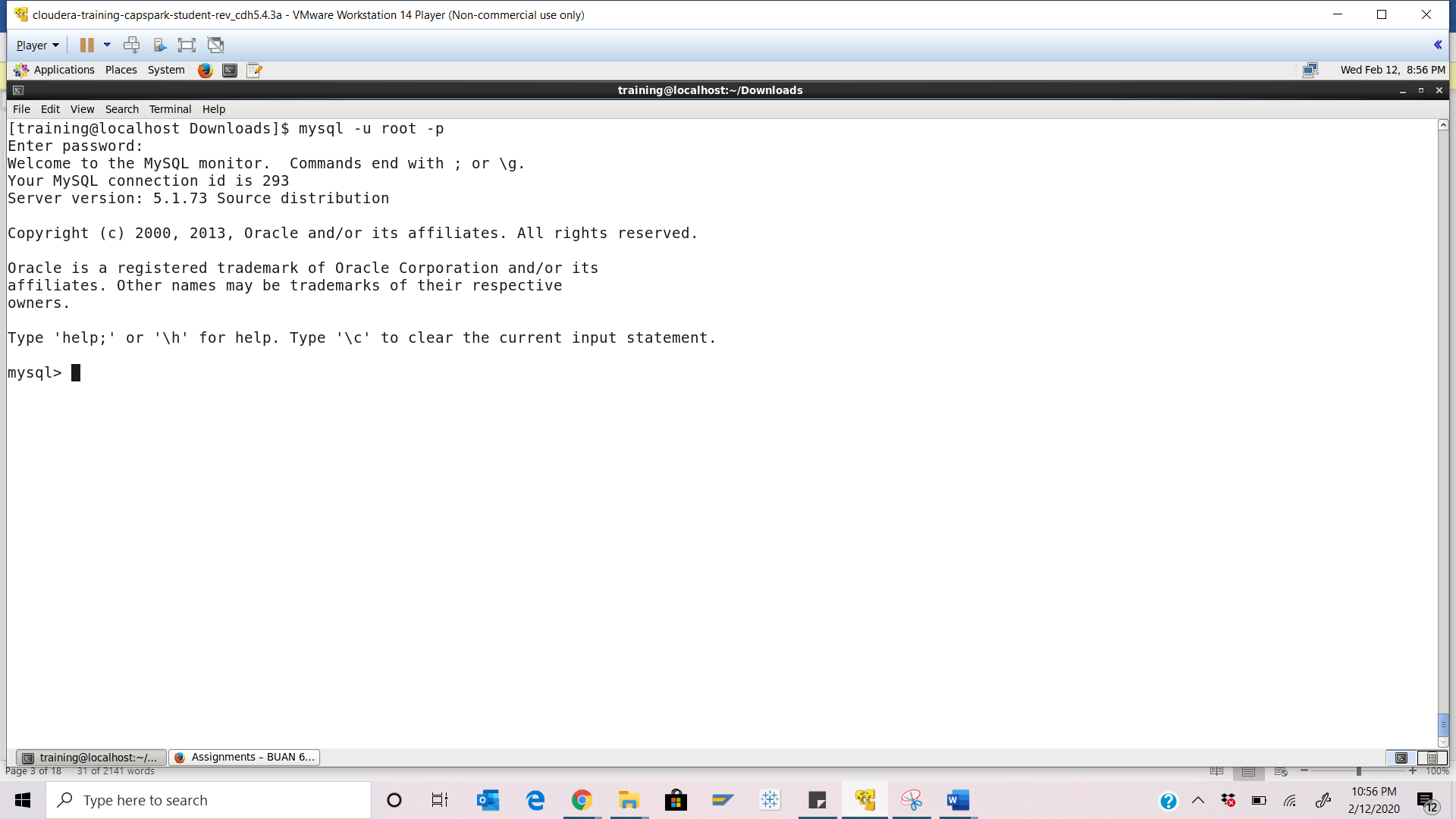
Q. Take a screenshot of the terminal application and paste it below.

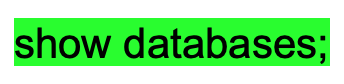
Step-2: Execute the following command at the shell prompt:



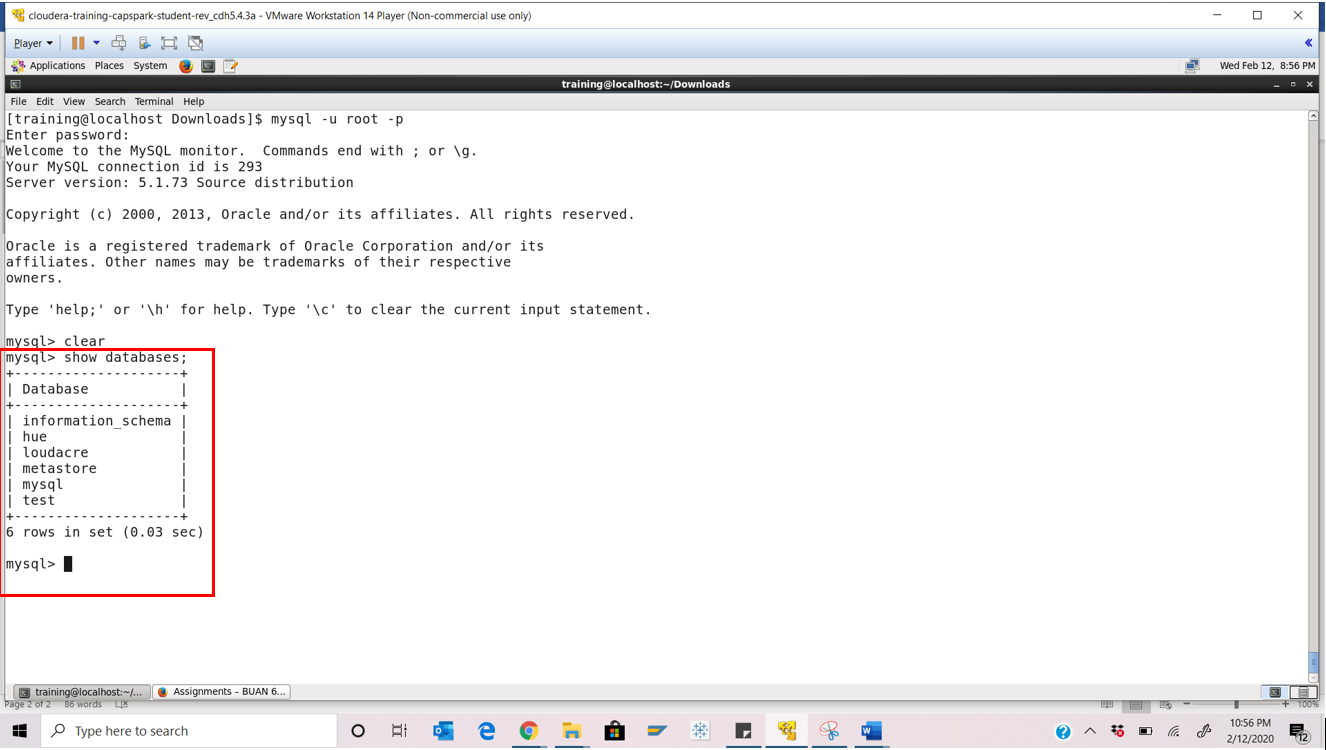
You will then be prompted for a password, hit enter.

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

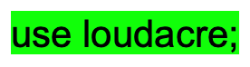
Step-3: Execute the following command at the MySQL shell prompt:



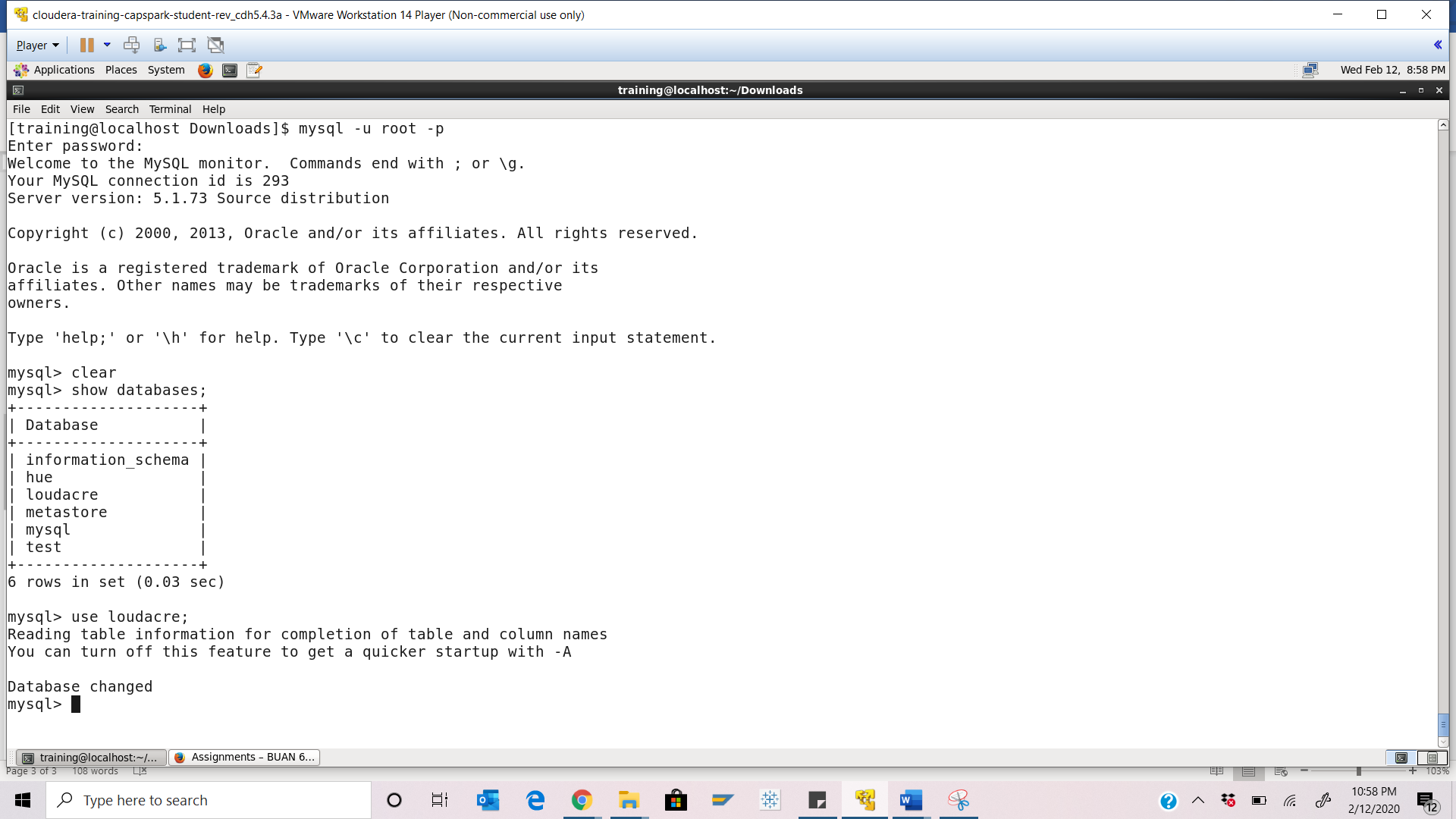
Q. Take a screenshot of the MySQL shell output paste it below.



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



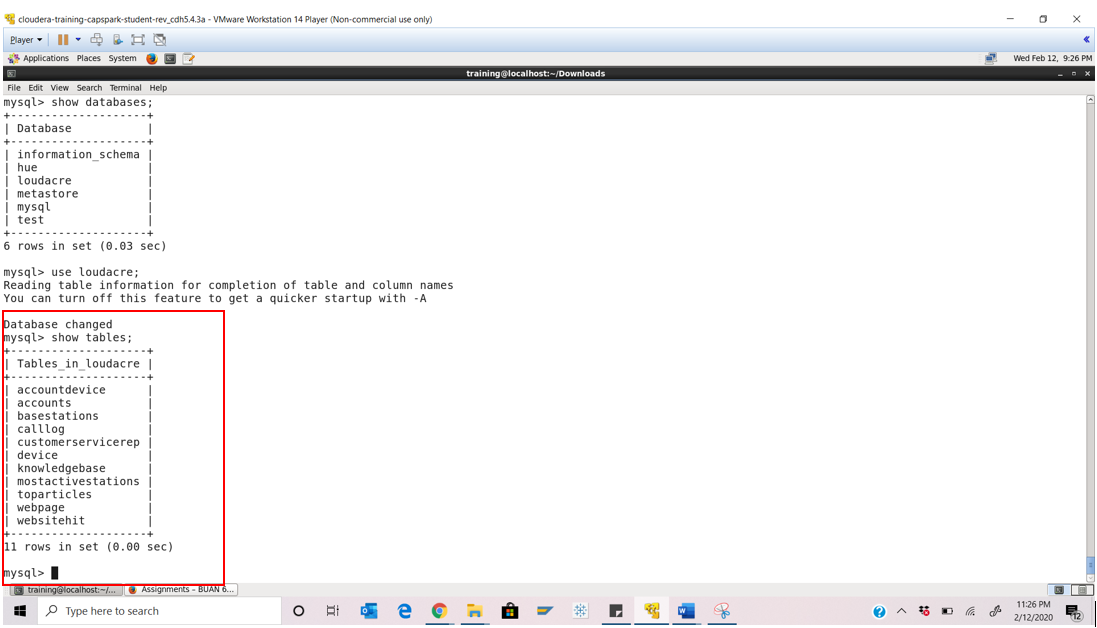
Q. Take a screenshot of the MySQL shell output paste it below.



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

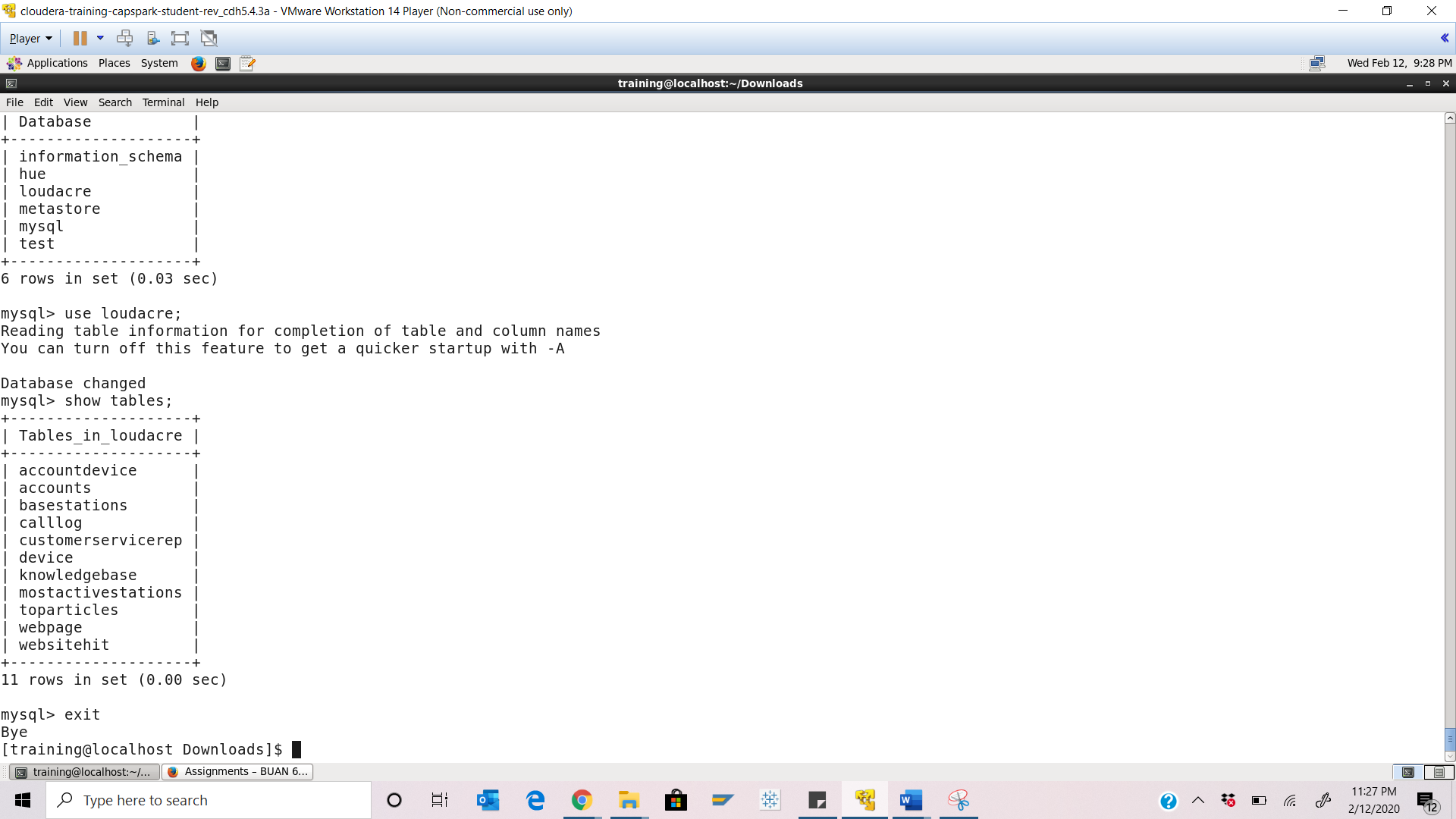


Q. Take a screenshot of the MySQL shell output paste it below.

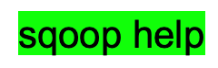
Step-6: Execute the following command at the MySQL shell prompt:



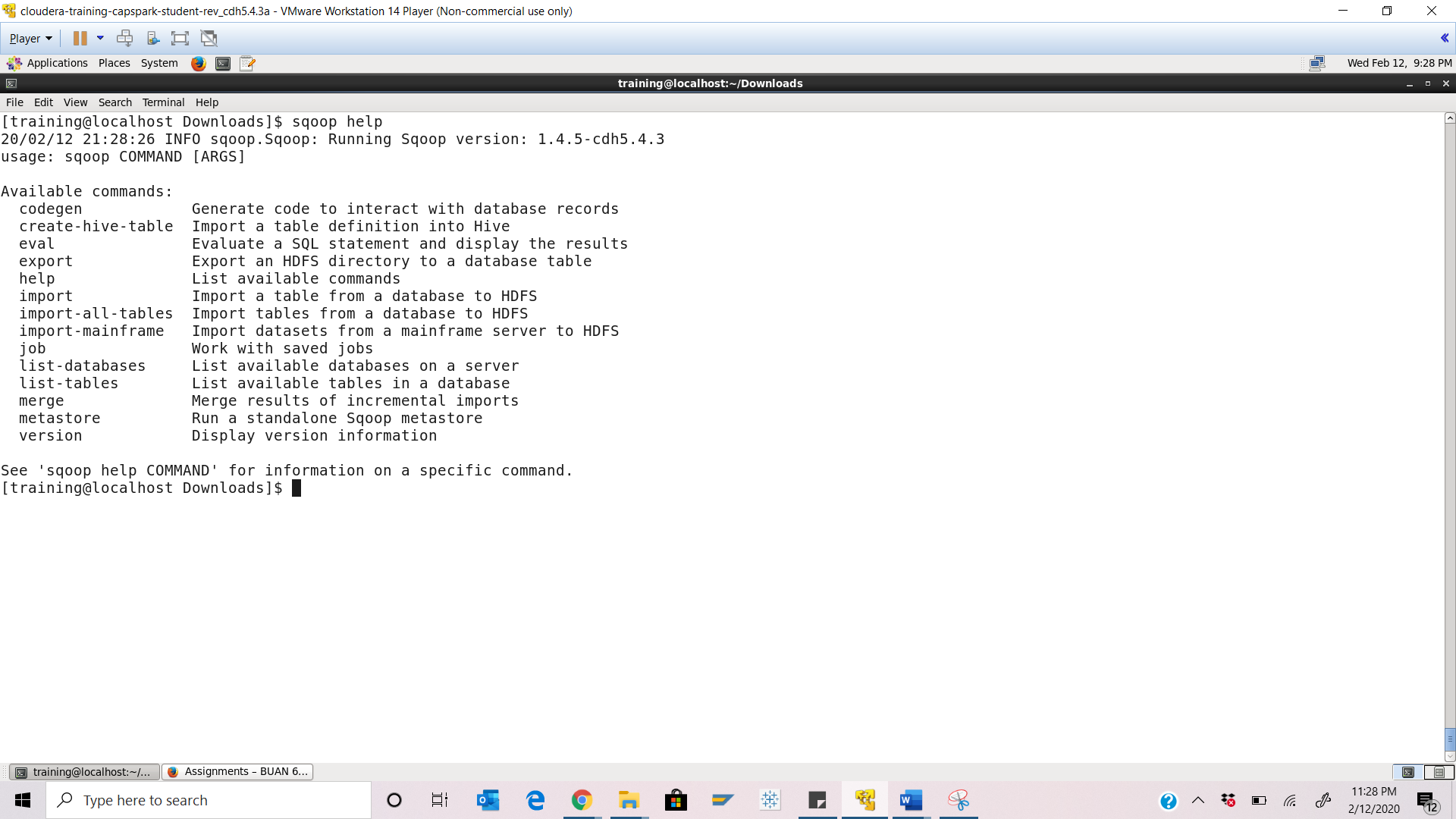
Q. Take a screenshot of the shell output paste it below.

Part-2: Sqoop Help: This section covers how to find help about using Sqoop using the terminal.

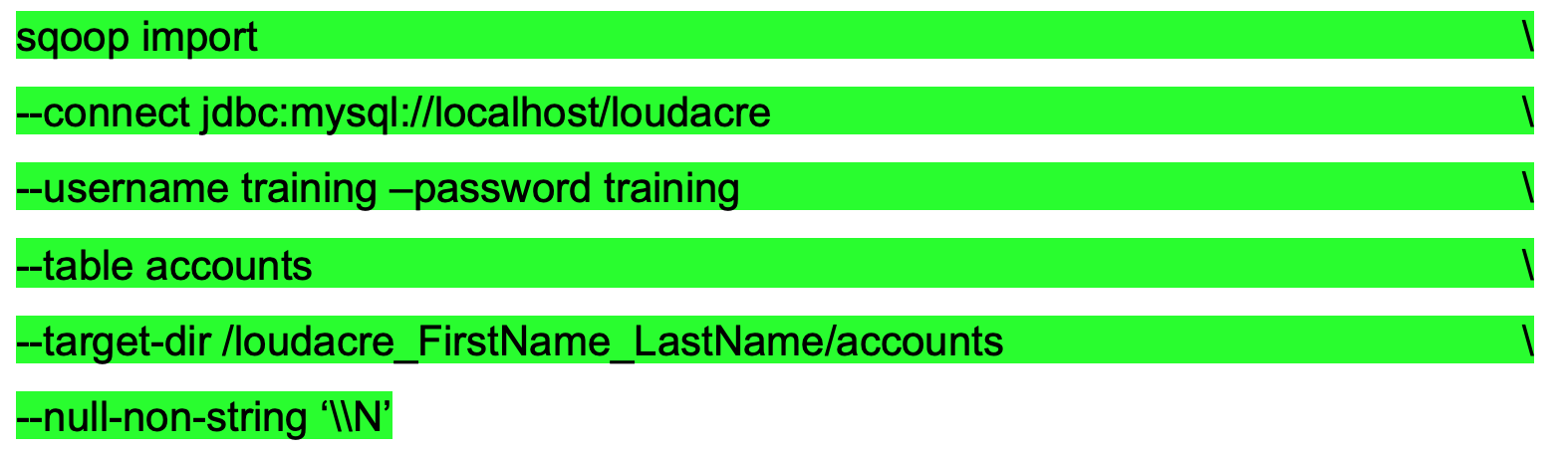
Step-7: Execute the following command at the shell prompt:



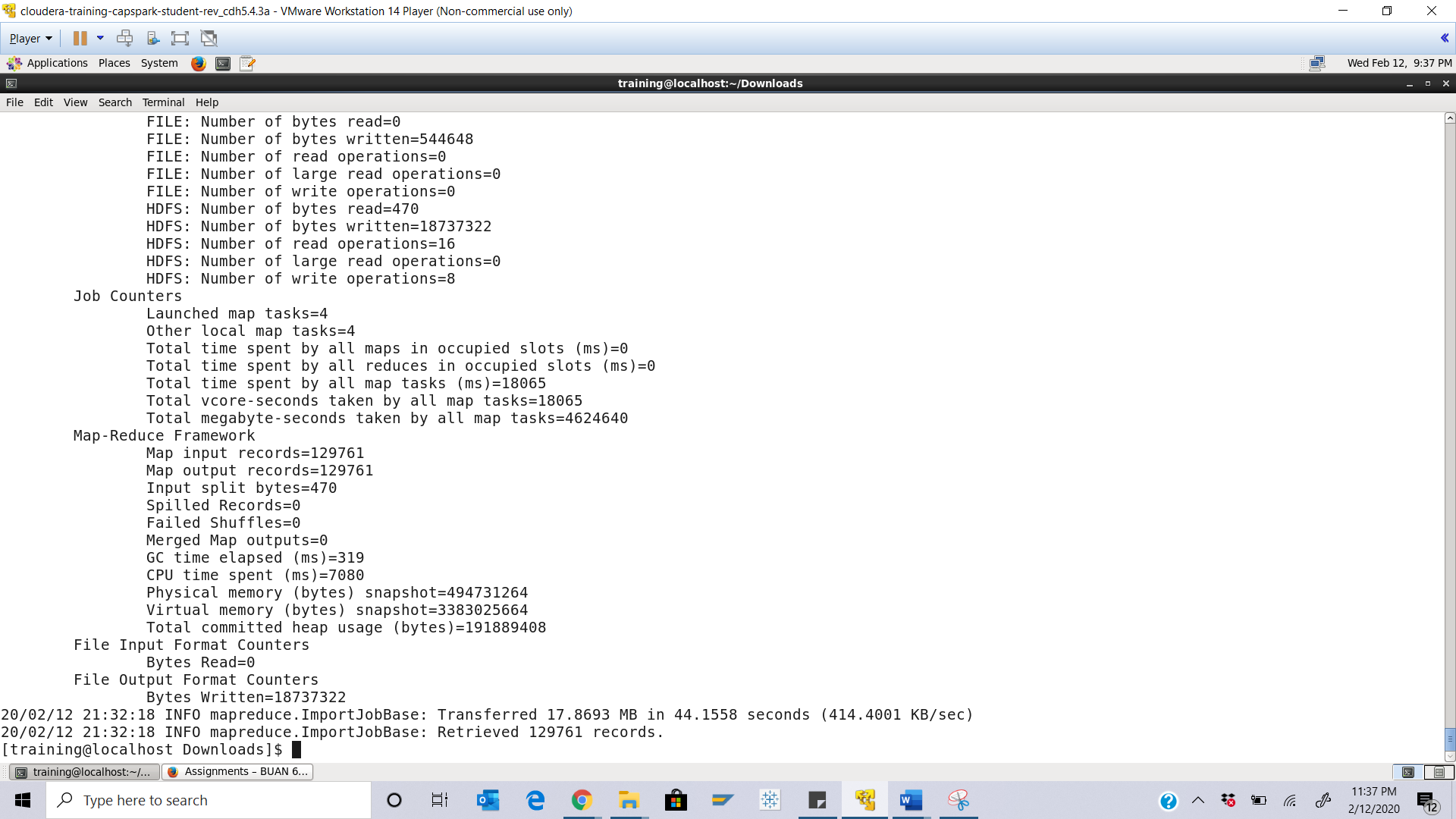
Q. Take a screenshot of the shell output paste it below.

Part-3: Sqoop Import: This section will cover importing database tables into HDFS using Sqoop.

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



Q. Take a screenshot of the shell output paste it below. How many mappers are used by default?

**4 mappers are used by default**

Step-10: Notice the path highlighted in the screenshot below after executing the command in Step-9. Use the computer application on the desktop of the Cloudera VM and navigate to the path is specified on your own system.

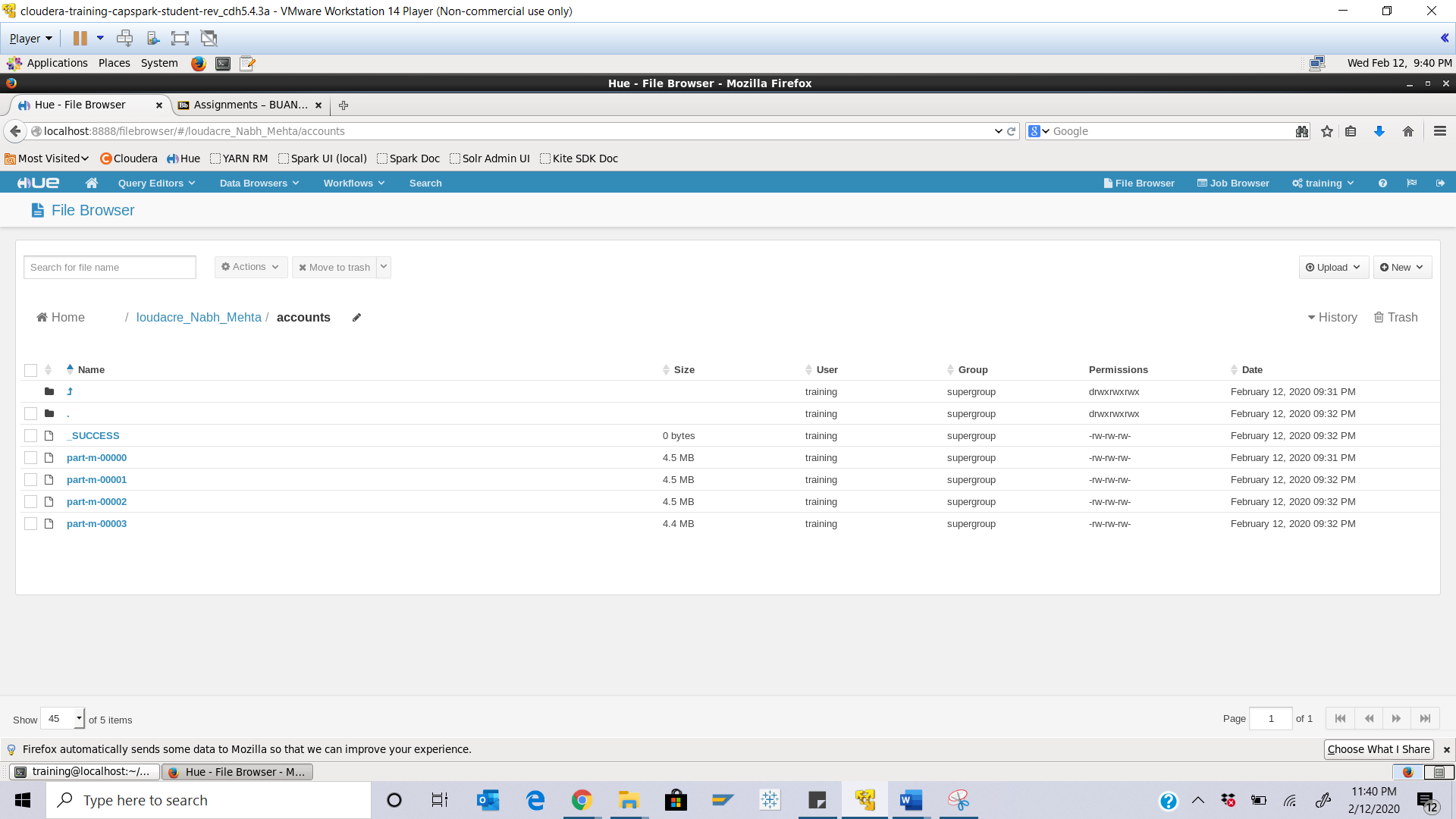
Q. Take a screenshot of the contents at the path and paste it below. Highlight the accounts.jar file.



Step-11: Using the file browser in HUE, navigate to the following path in HDFS:

/loudacre\_FirstName\_LastName/accounts

Q. Take a screenshot of the contents of the folder at the path specified above and paste it below. How many files do you see? Why do you see these many files?

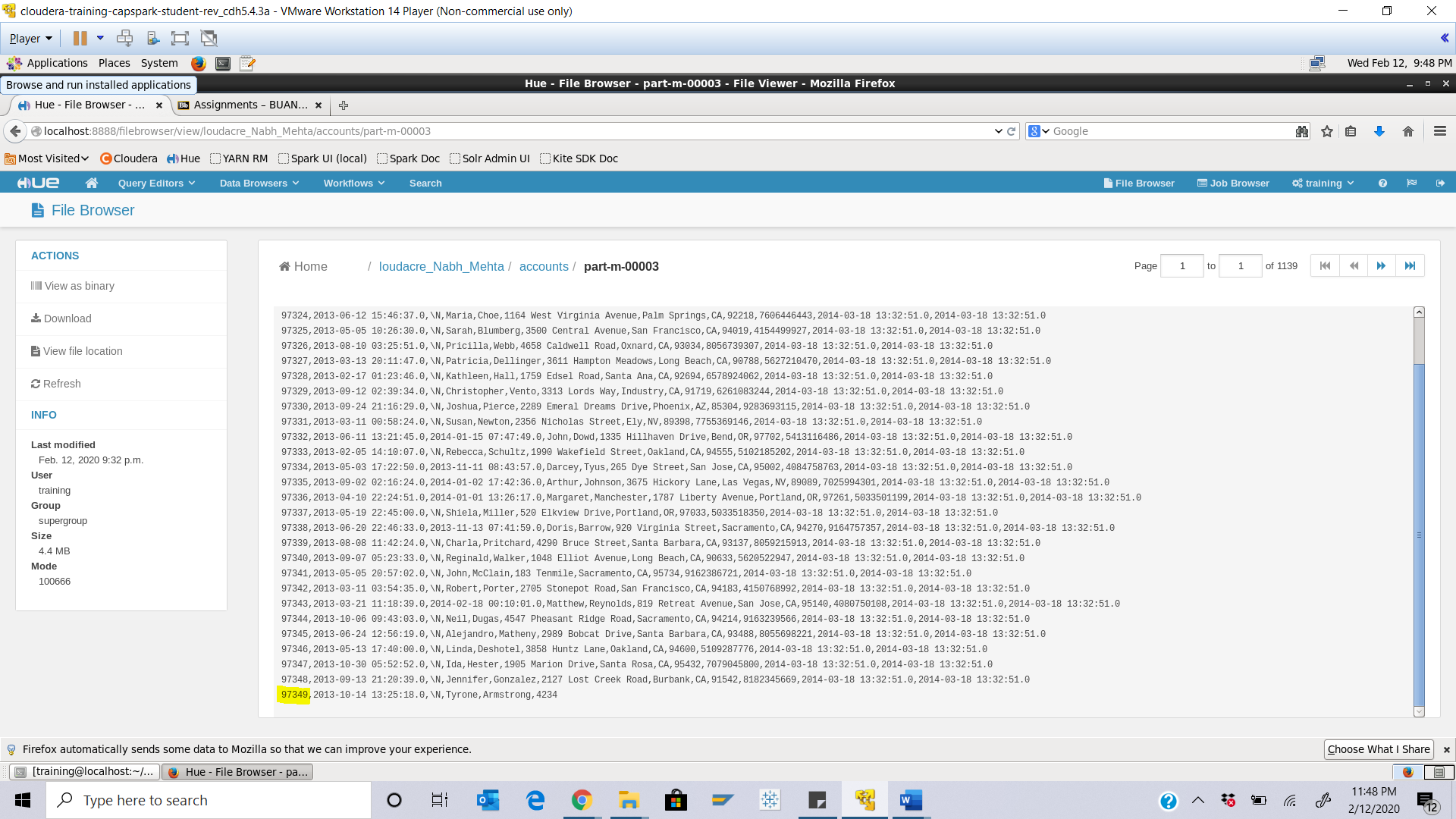


**Ans: No of splits are 4 to achieve higher performance rate by parallel processing and therefore we see 4 files here**

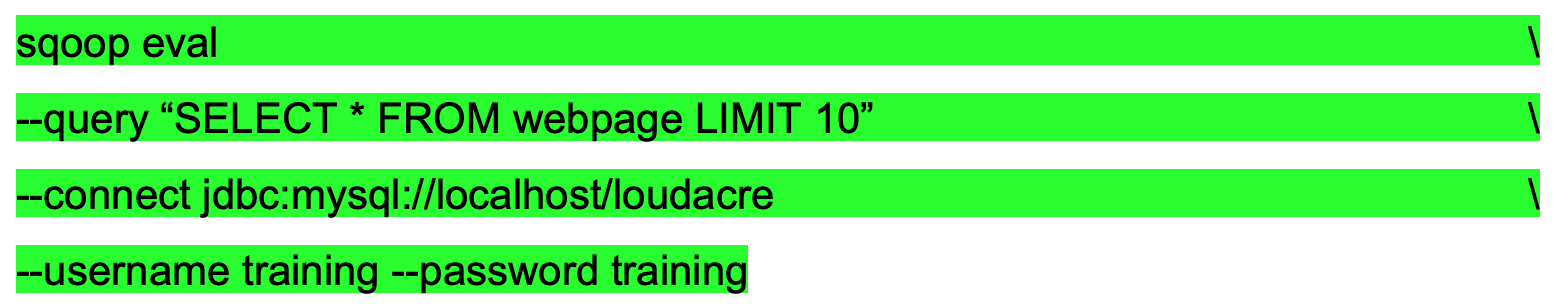
Step-12: Click on the ‘part-m-0003’ file, then click on the button highlighted in the screenshot below

Q. Take a screenshot of this screen. What is the largest account number? The account number is the first value in each line. Highlight the largest account number in the screenshot. Note this account number.

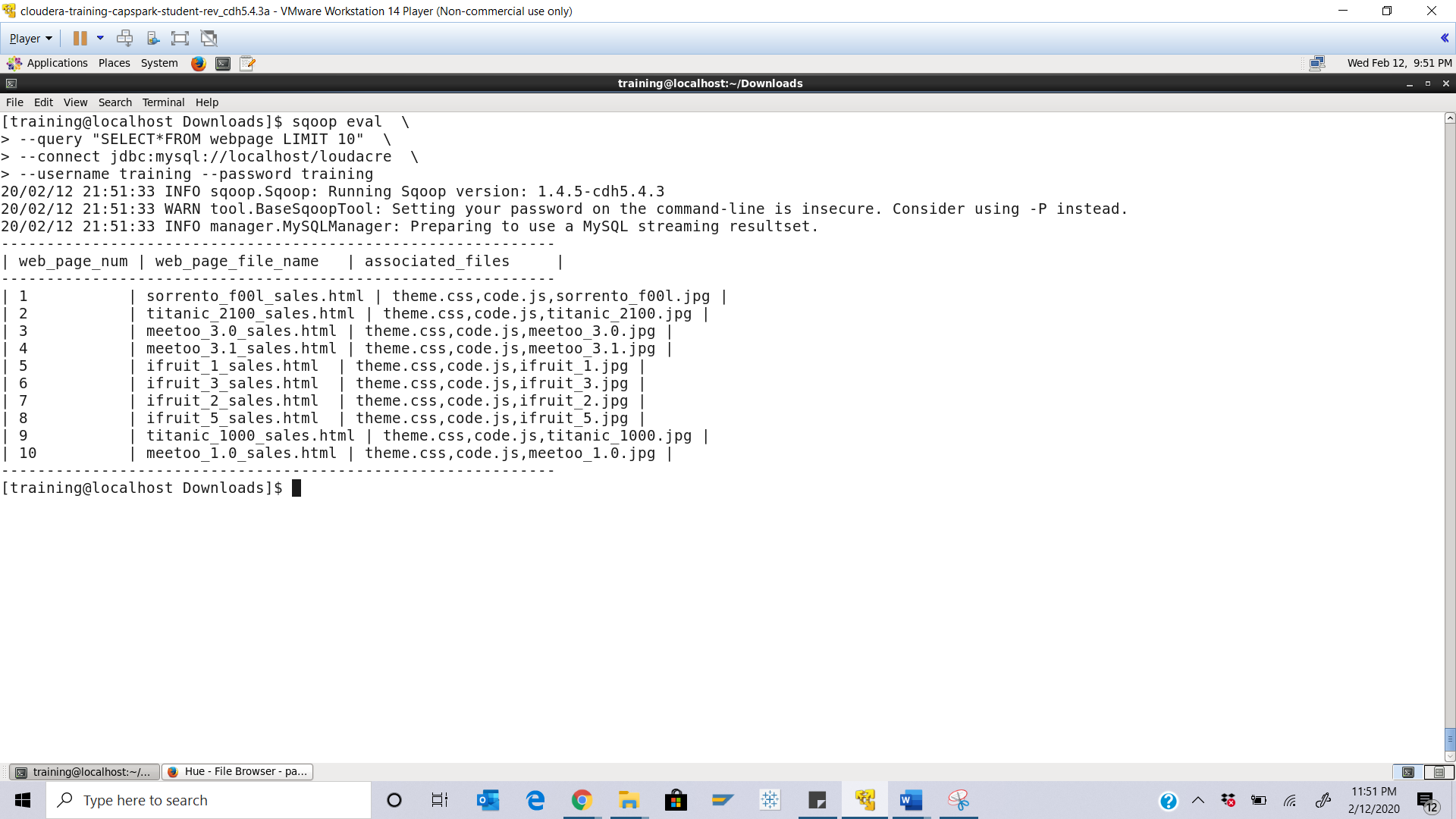
Ans: 97349



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

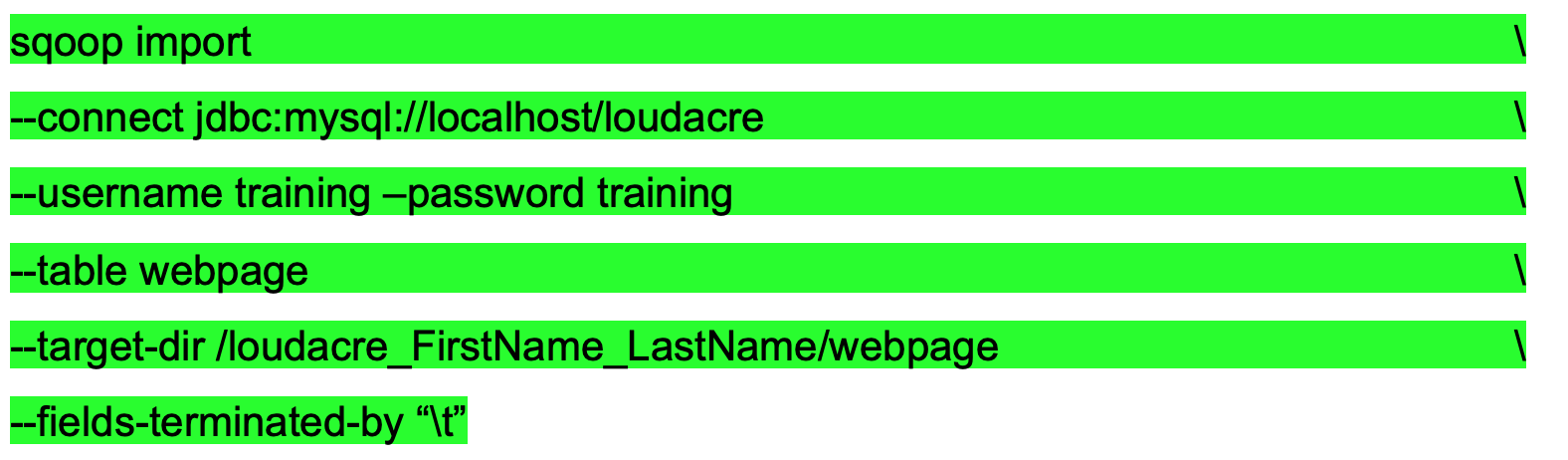


Q. Take a screenshot of the shell output and paste it below. What kind of delimiter is used in the feedback on the screen?

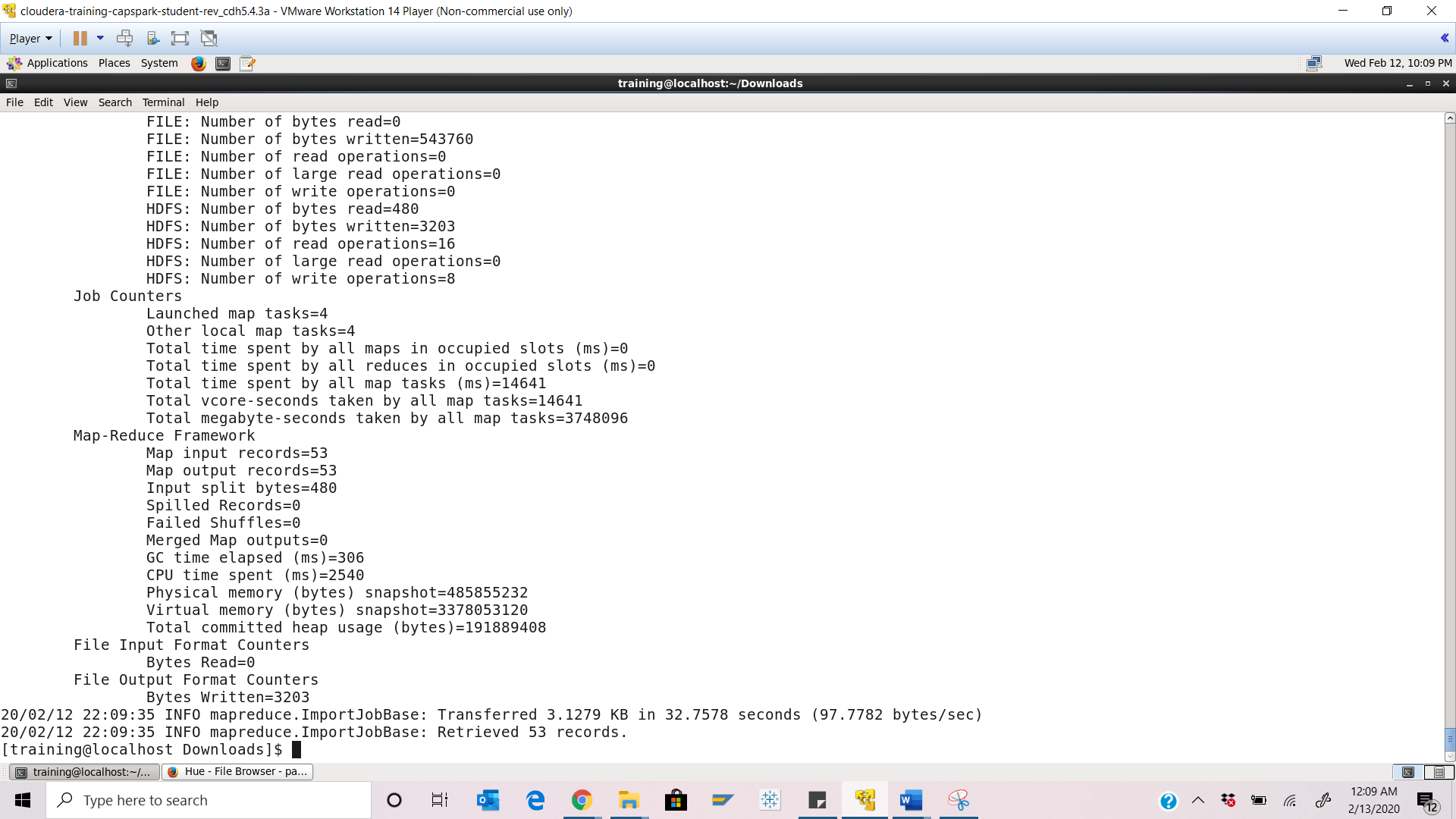


Ans: output is displayed in delimiter as: ‘|’

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



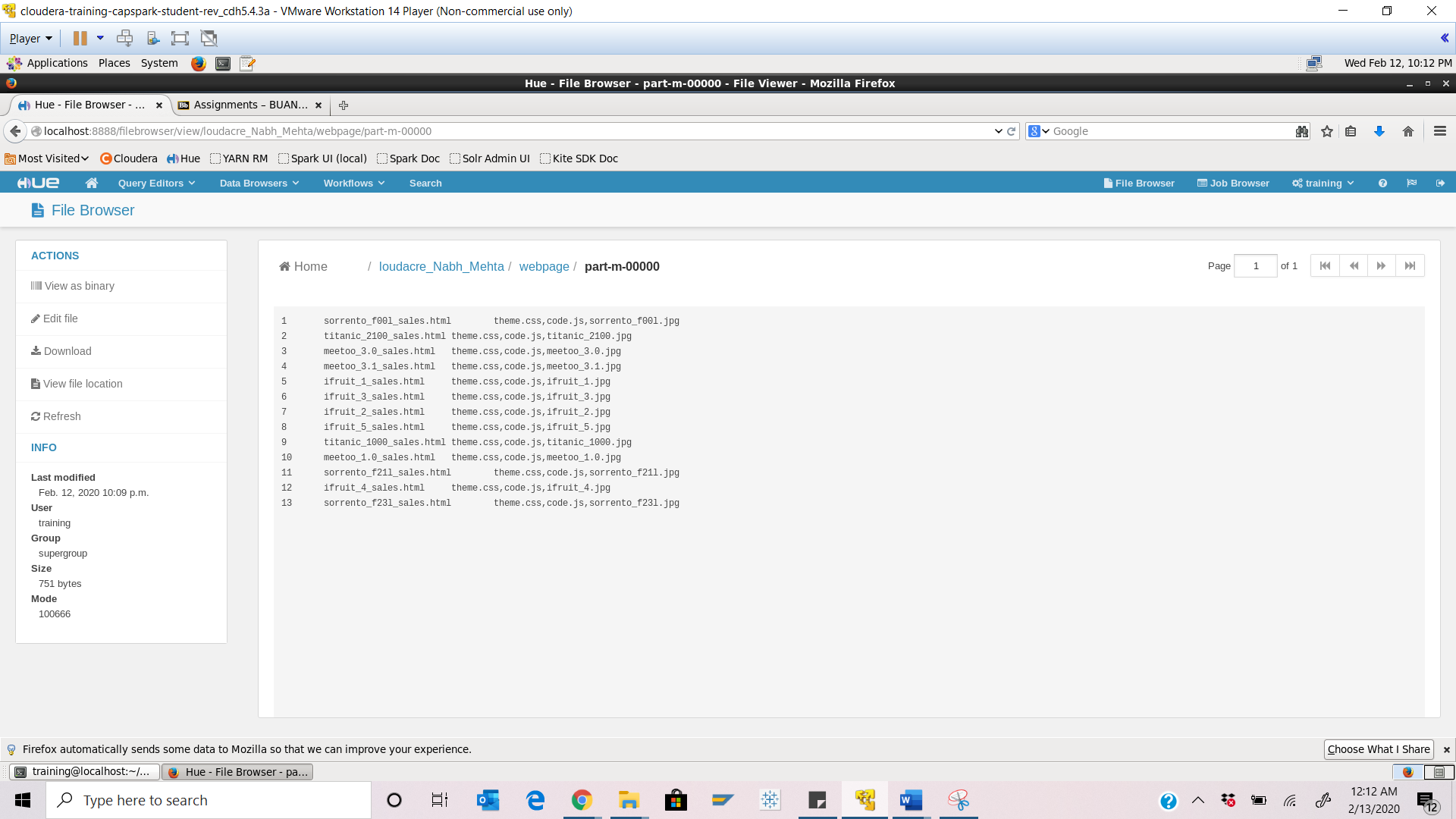
Q. Take a screenshot of the shell output paste it below.



Step-15: Using the file browser in HUE, navigate to the following path in HDFS:

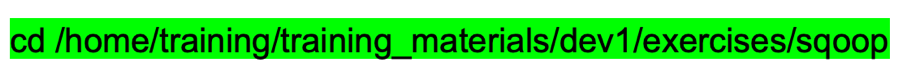
/loudacre\_FirstName\_LastName/webpage

Q. Open any file in this folder and take a screenshot of the contents of the folder at the path specified above and paste it below. What kind of delimiter is used?



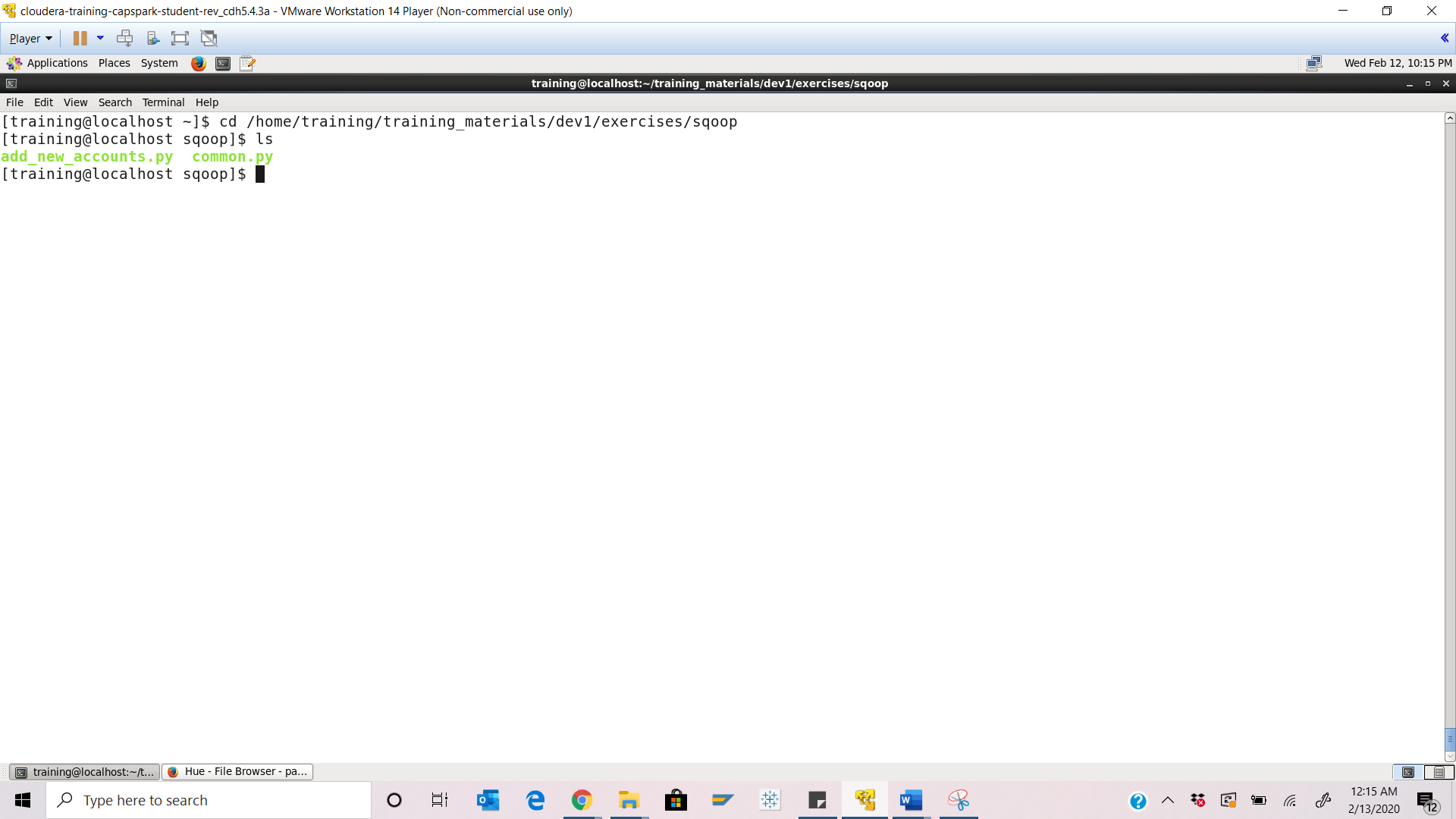
Ans: Tab delimiter is used

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





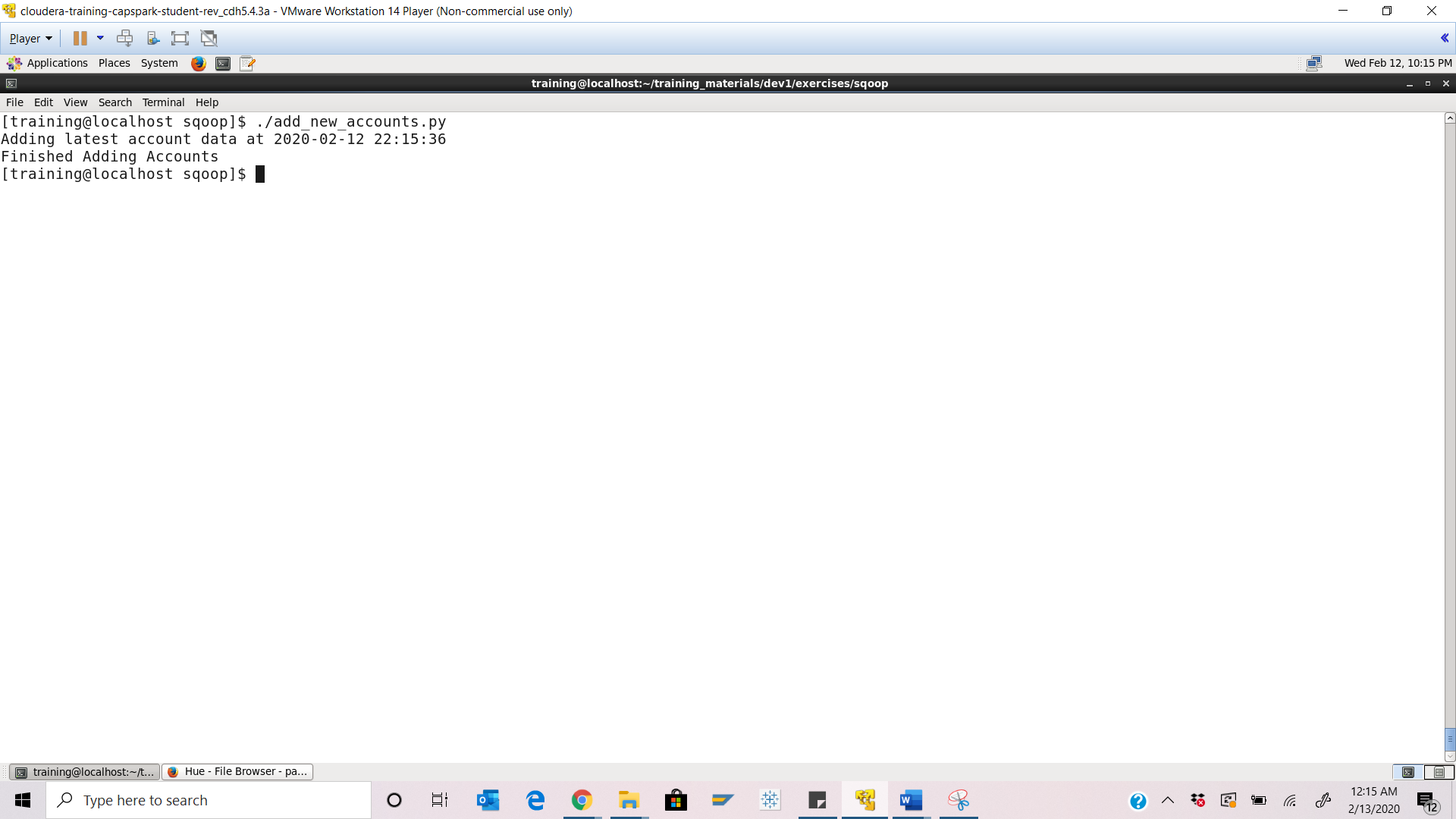
Q. Now type the command and execute. Take a screenshot of the shell output and paste it below.



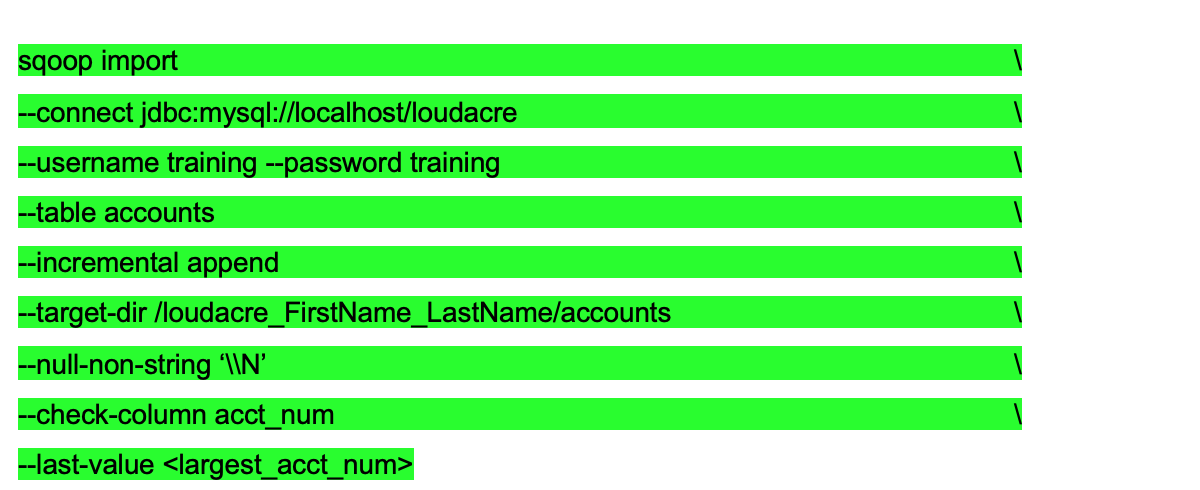
Step-17: Execute the following command at the shell prompt:



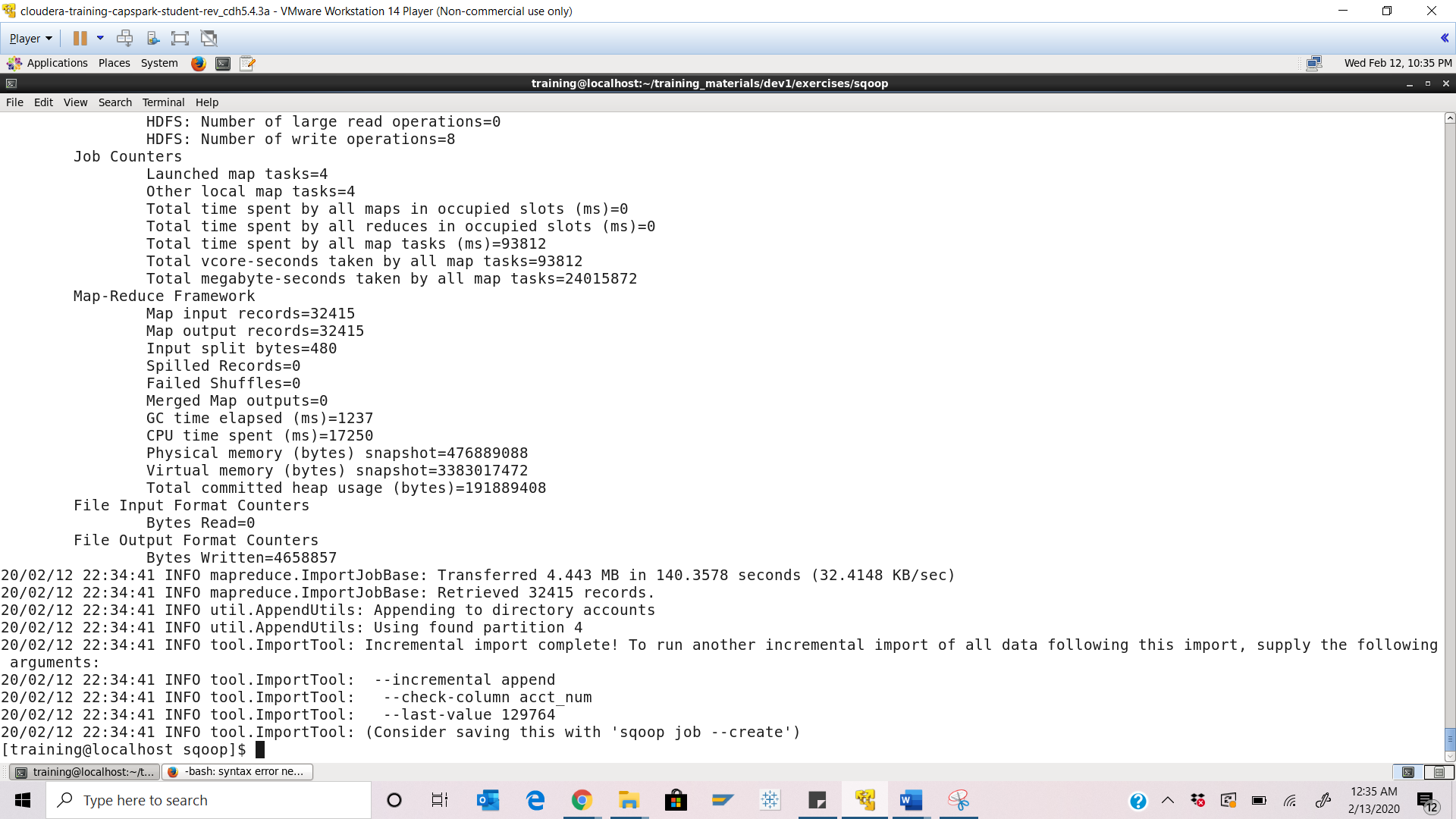
Q. Take a screenshot of the shell output and paste it below.



Step-18: Execute the following command at shell prompt



Q. Take a screenshot of the shell output paste it below. How many mappers are used? Why?

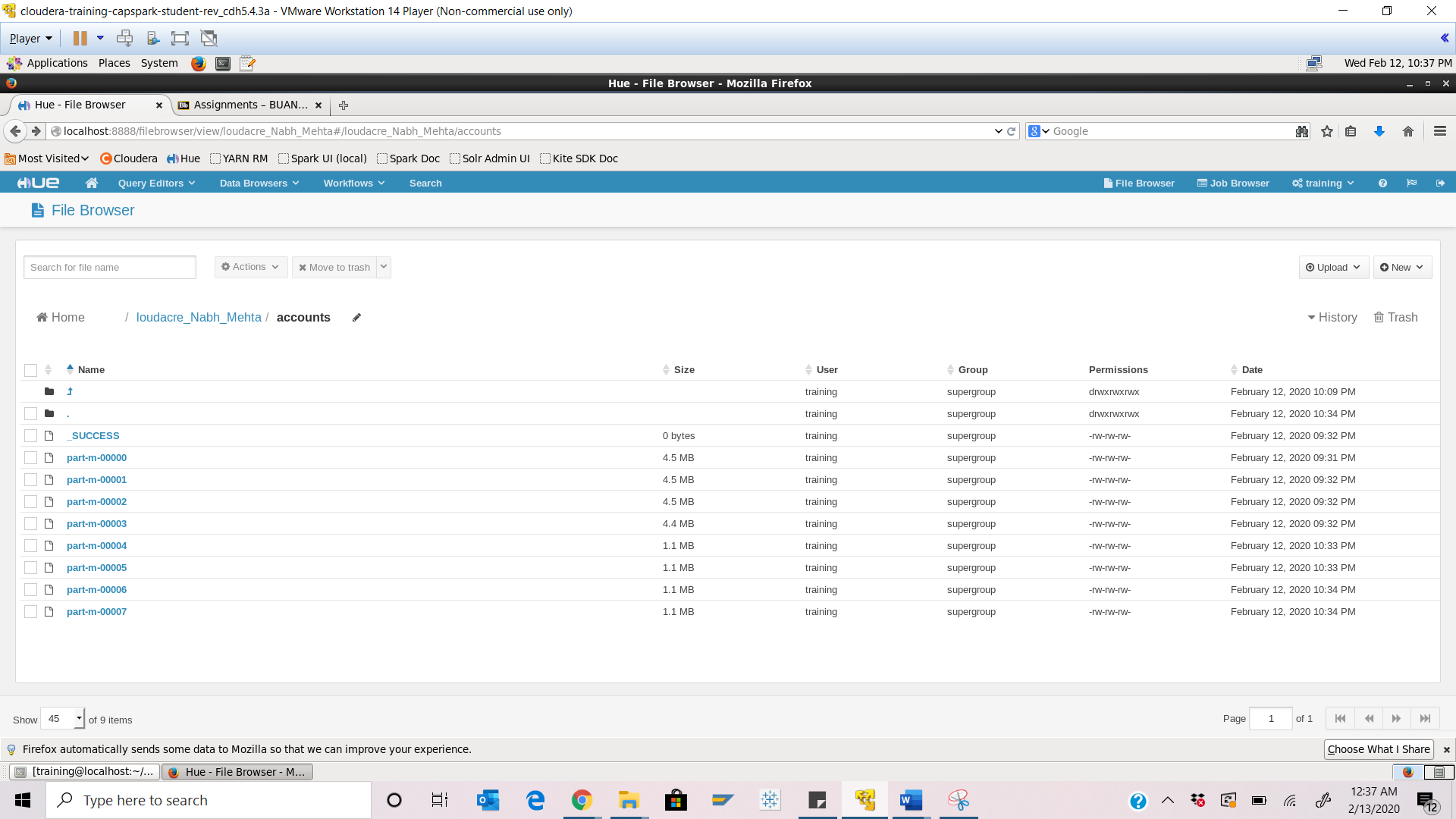


4 Mappers are used here as this is by default feature

Step-19: Using the file browser in HUE, navigate to the following path in HDFS:

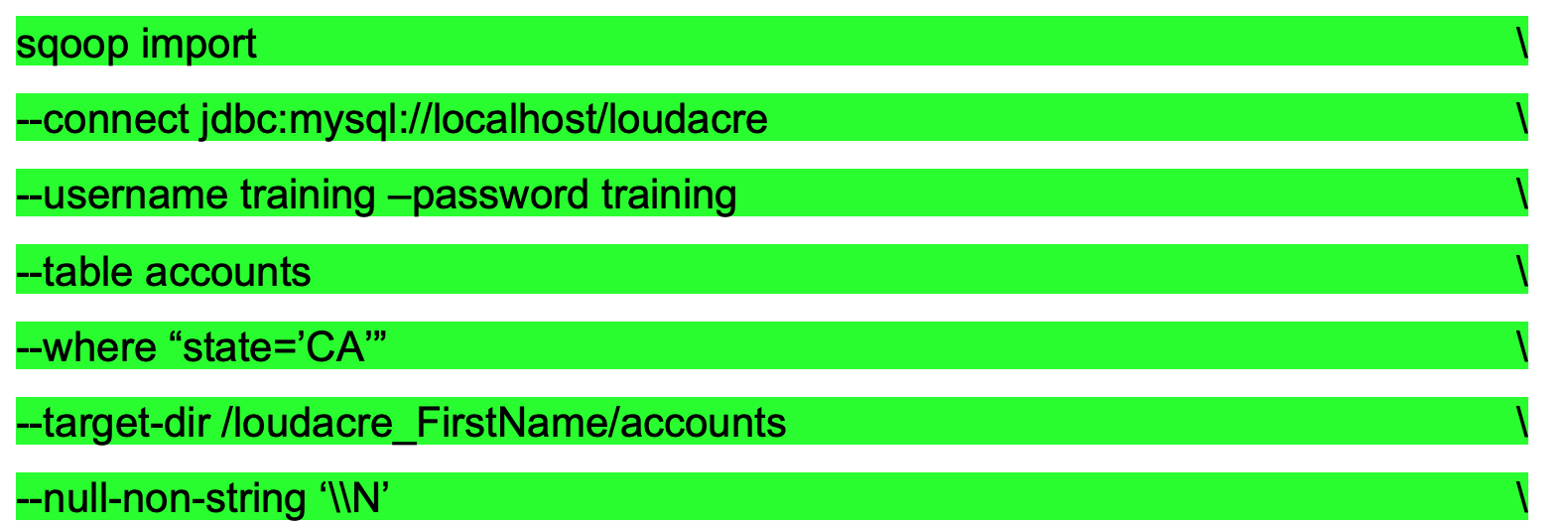
/loudacre\_FirstName\_LastName/accounts

Q. Take a screenshot of the contents of the folder at the path specified above and paste it below. Was our incremental import successful? How can you tell?



Yes it was successful. We can tell this by acknowledging the \_SUCCESS file created at first place

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



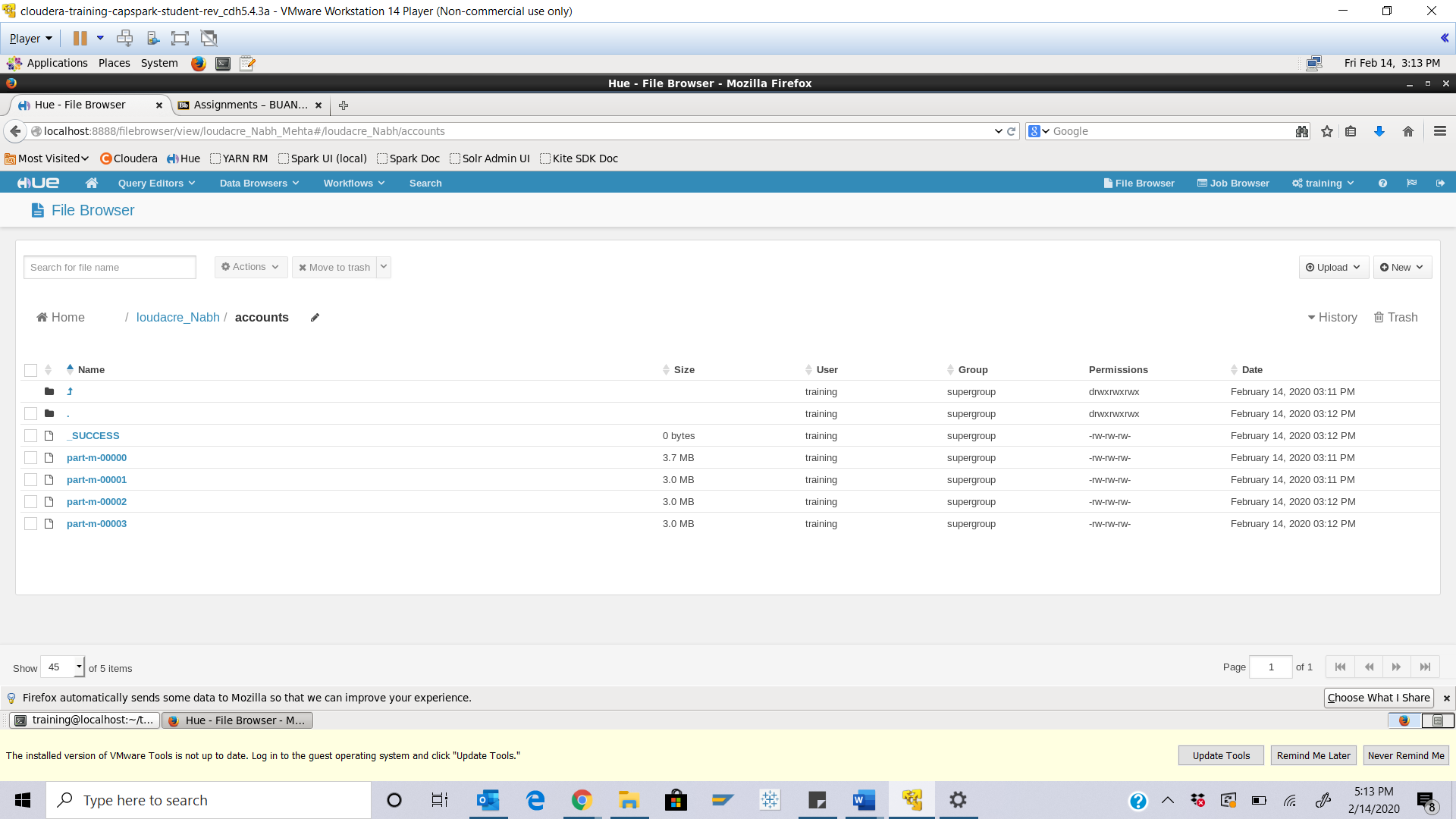
Q. Take a screenshot of the shell output paste it below.



Step-21: Using the file browser in HUE, navigate to the following path in HDFS:

/loudacre\_FirstName/accounts

Q. Click on any file in this folder, take a screenshot of the contents and paste it below.



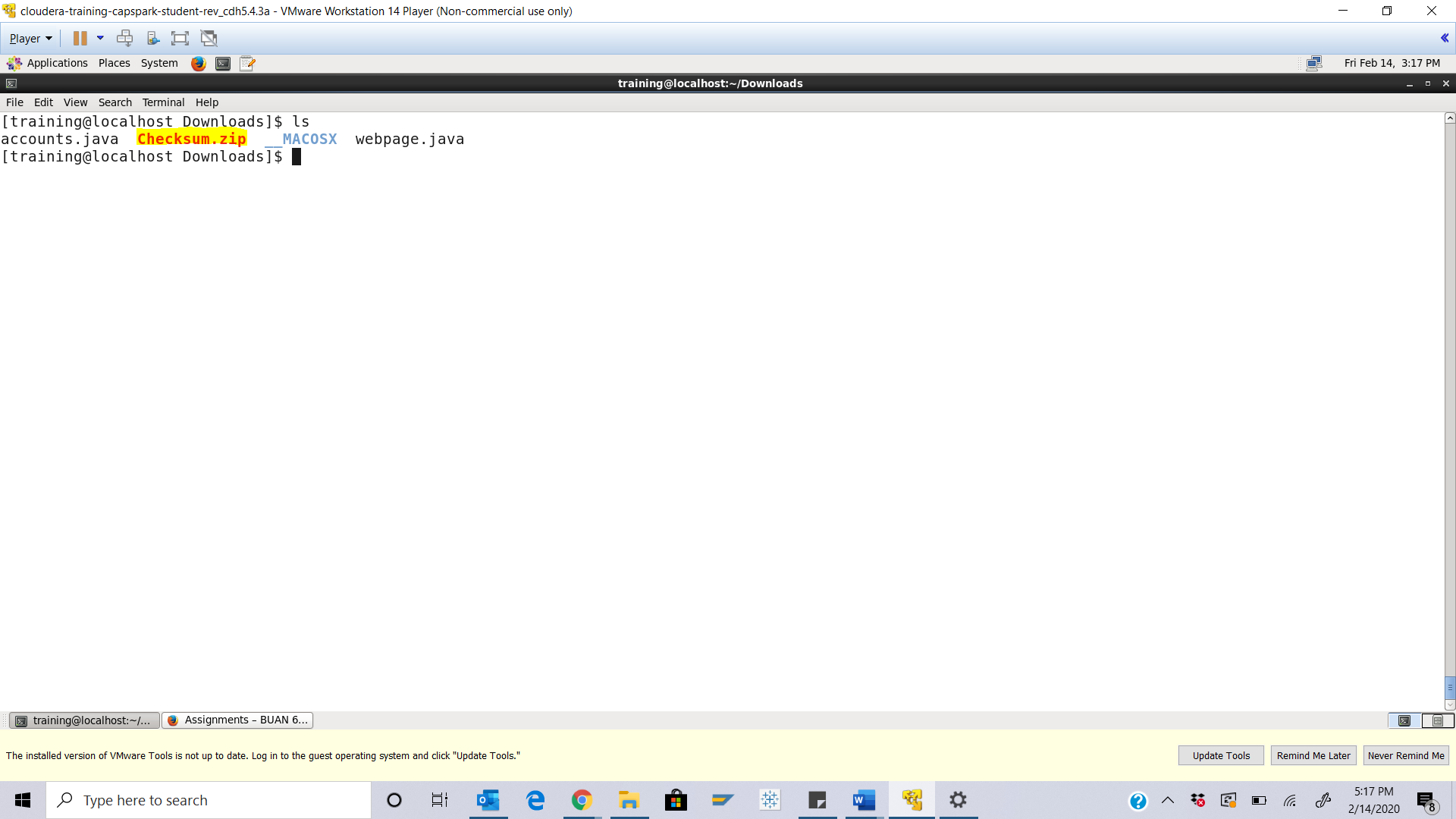
Part-6: Checksum: In this section we will perform a checksum

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:



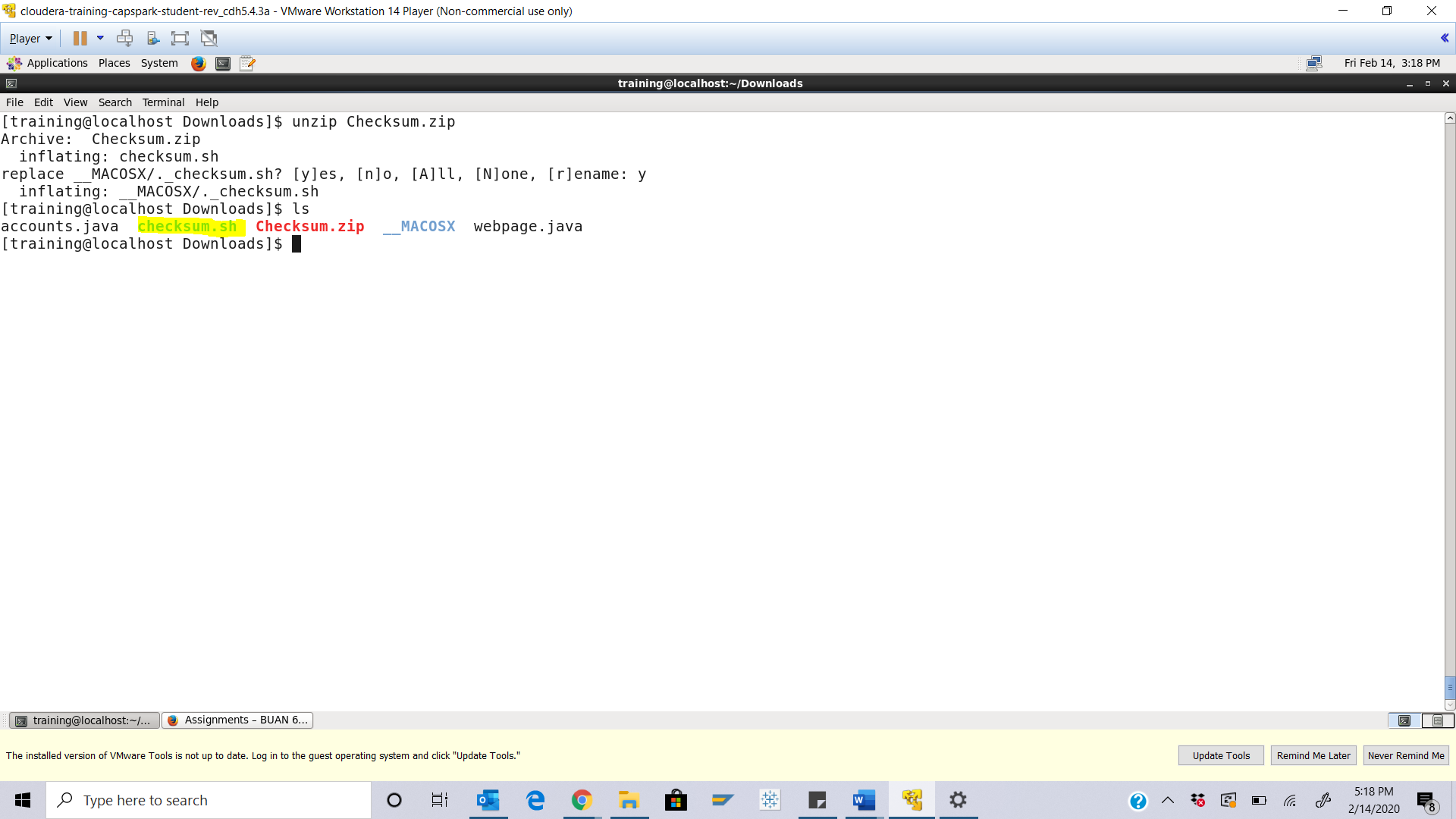
Q. Now type the command and execute. Take a screenshot of the shell output, highlight the zip file and paste it below.



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.

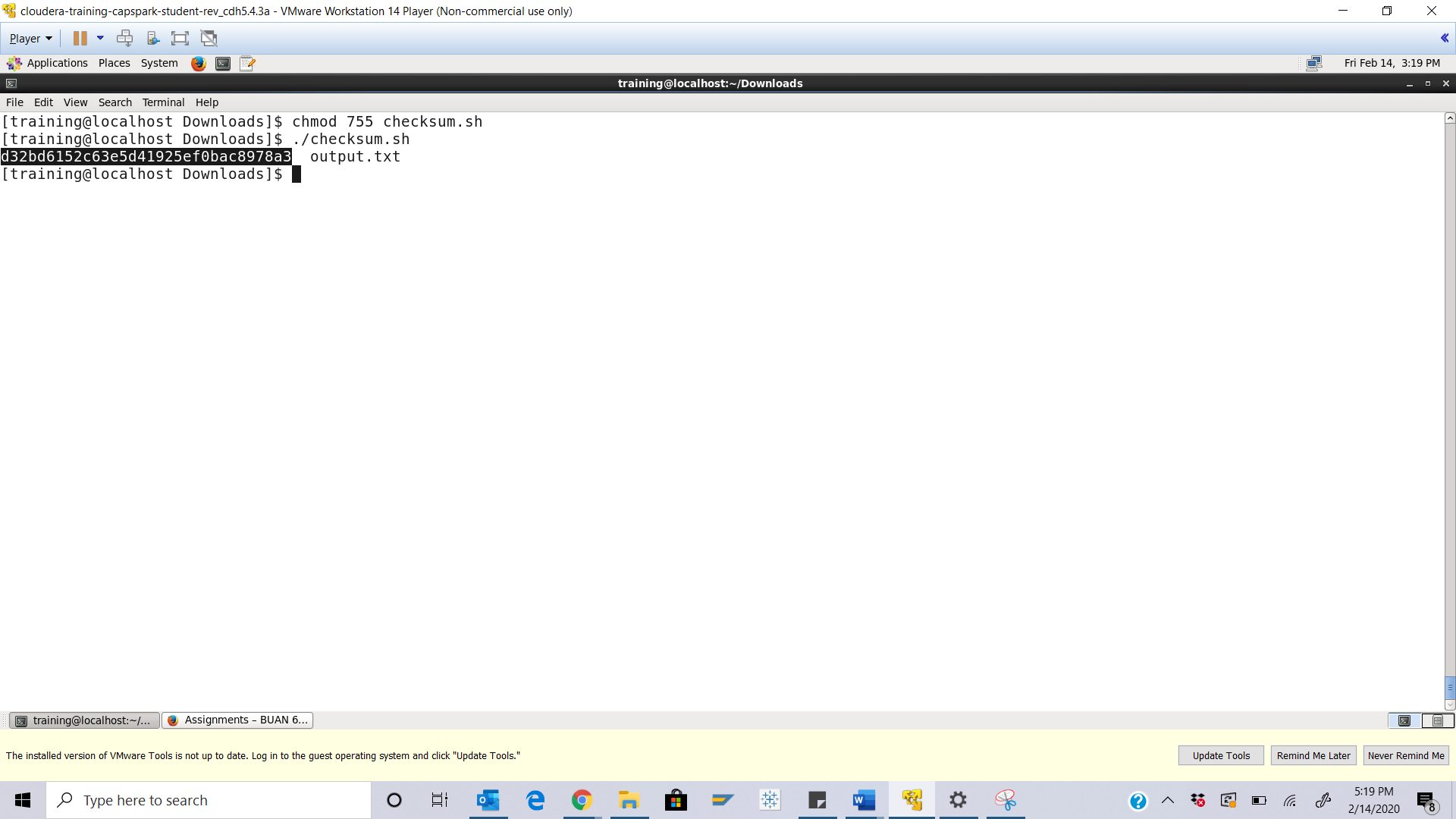


Step-25: Execute the following command at the shell prompt:

Then execute the following command:

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

ChecksumValue - d32bd6152c63e5d41925ef0bac8978a3



Step-26: 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.

