**HIVE - I**

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

**Date: 02/20/2020**

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

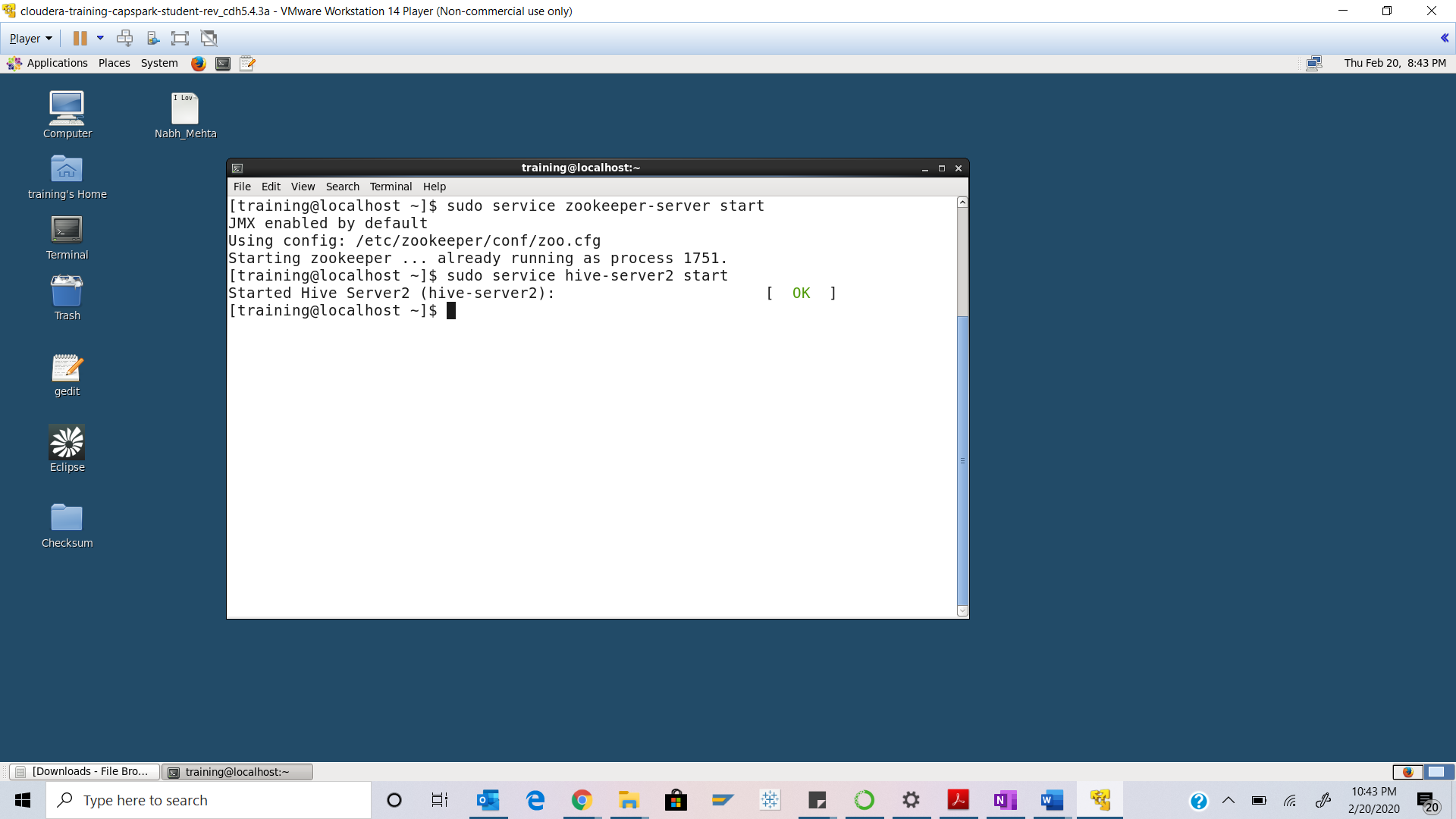
Q. Take a screenshot of the terminal application and paste it below.



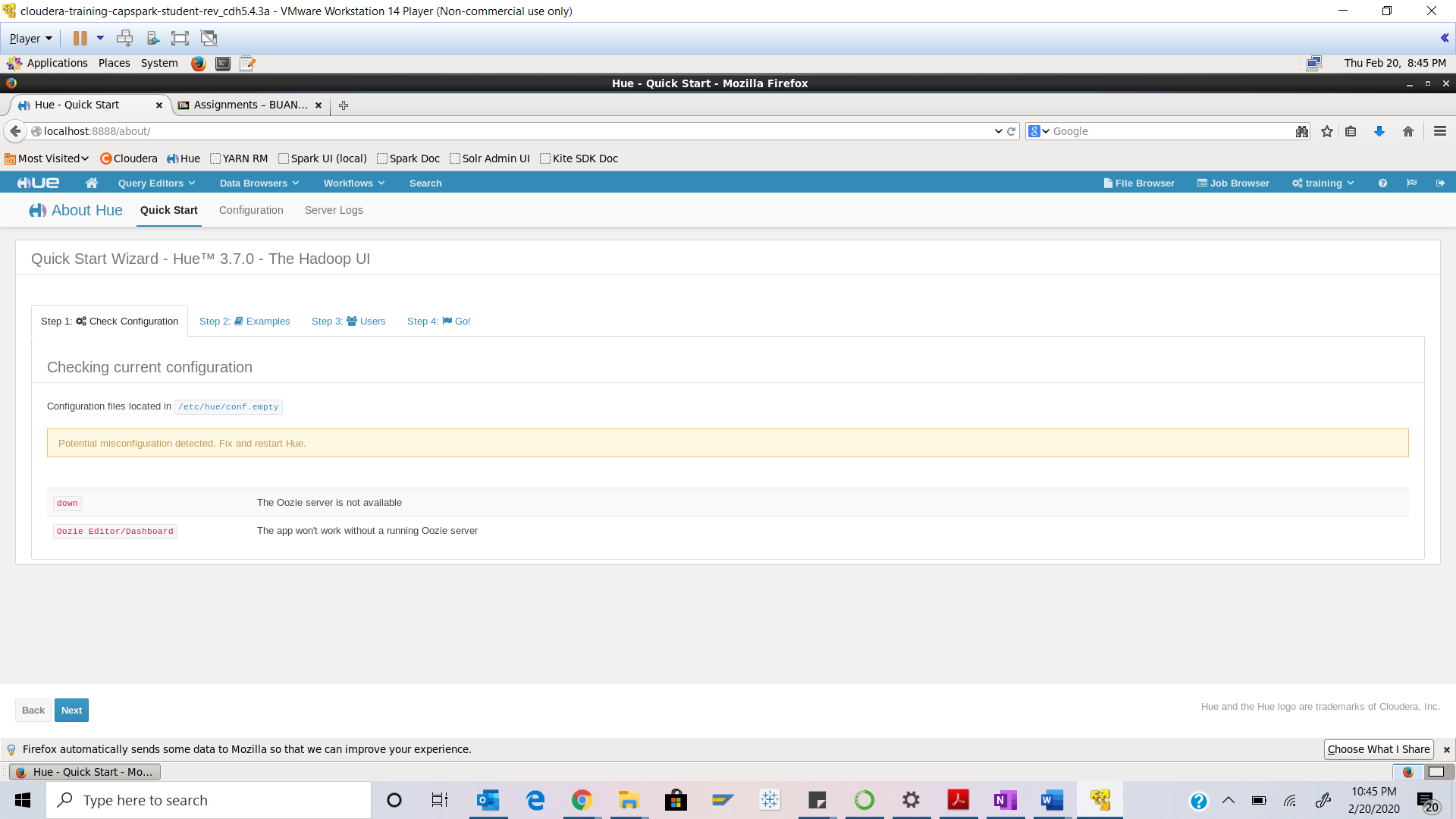
Step-2: Execute the following commands at the shell prompt:



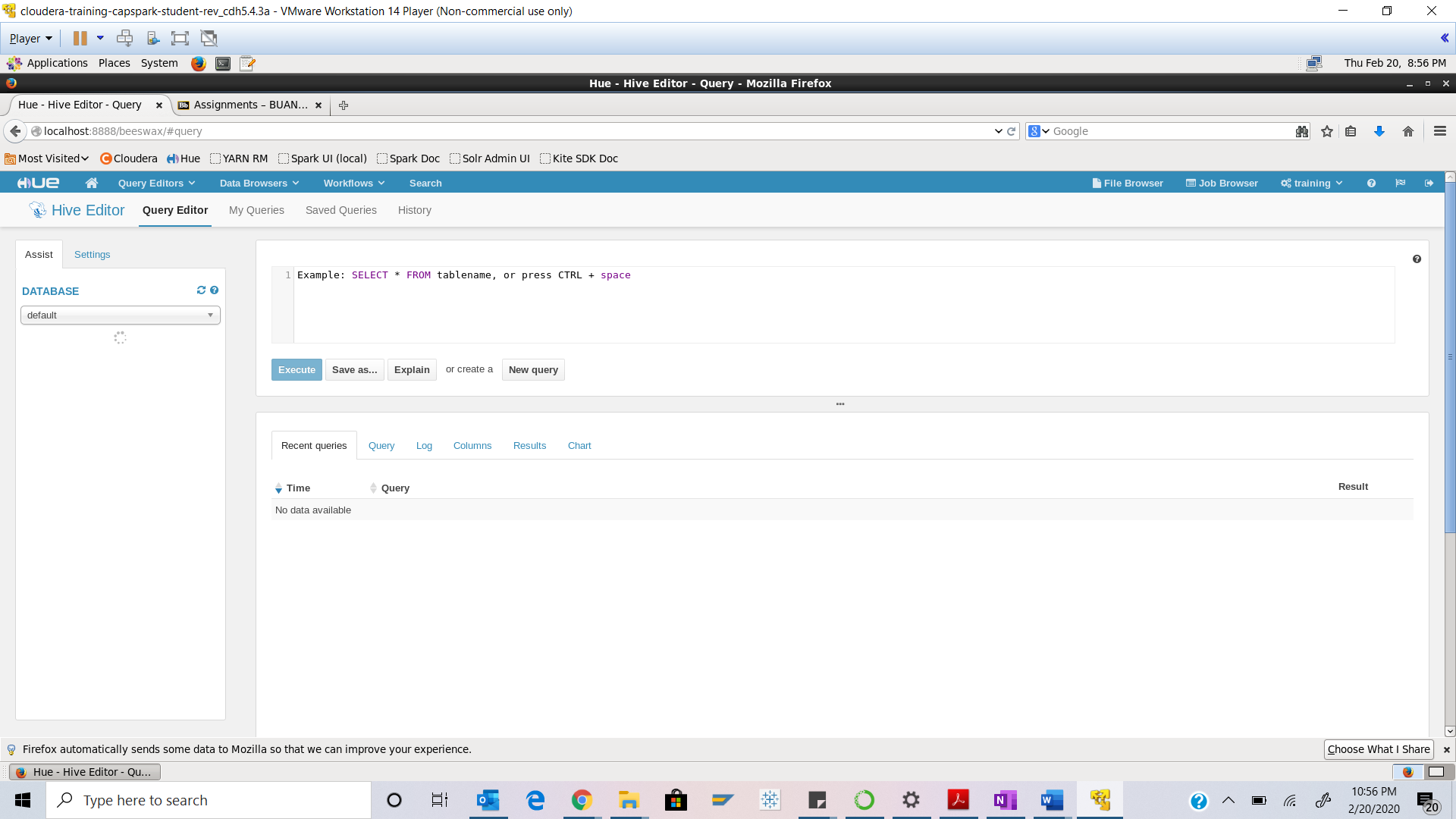
Q. Take a screenshot of the shell output paste it below.

Step-3: Open the Firefox browser and start up HUE from the favorites tab.

Q. Take a screenshot of the HUE splash screen and paste it below.



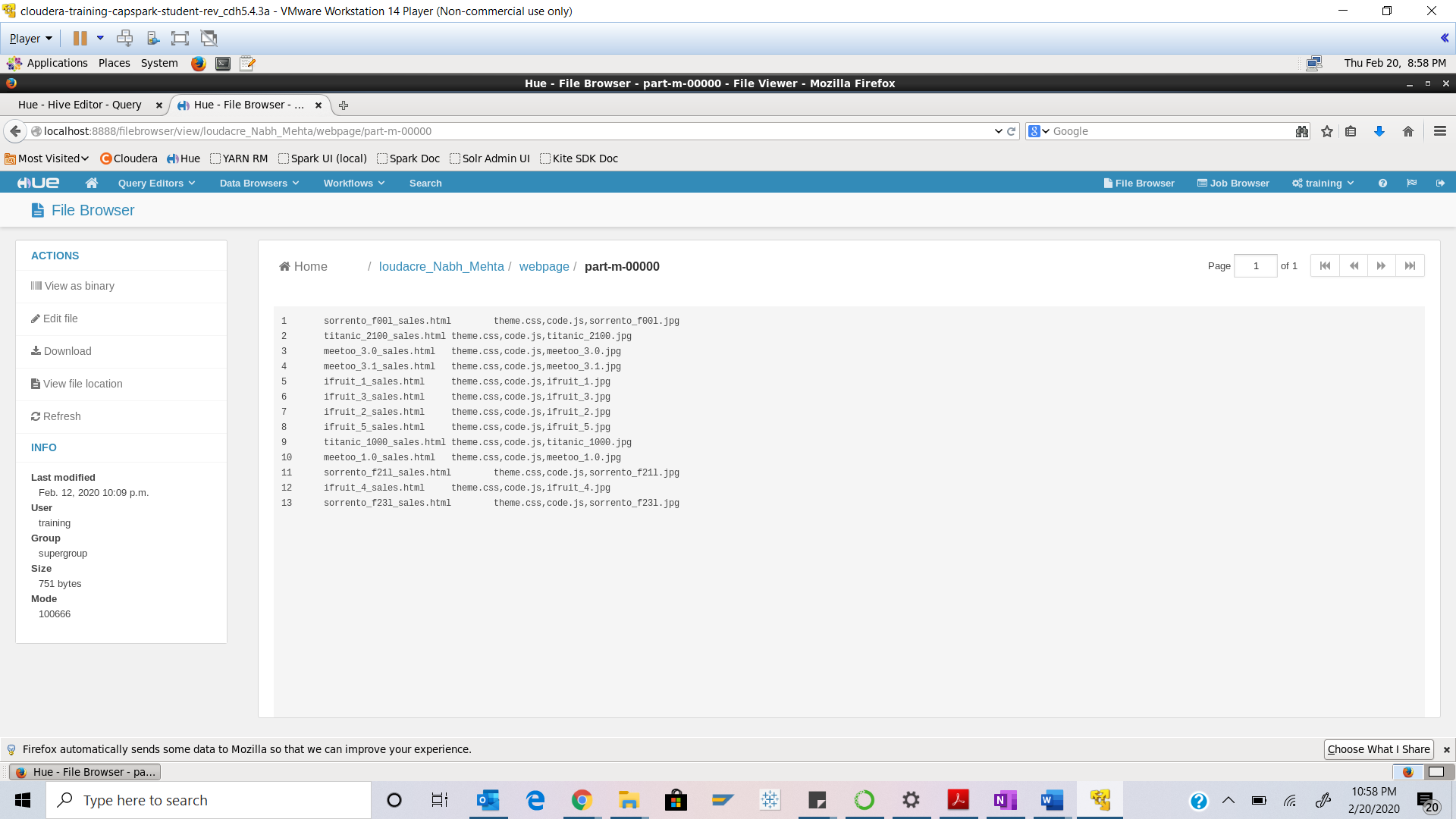
Step-4: Click on the “query editors” button in the HUE menu bar then click on “Hive” as shown in the screenshot below.



Step-5: Open another tab in the Firefox browser and start the HUE file browser. Navigate to the following directory

/loudacre\_FirstName\_LastName/webpage.

Q. Open any of the files in this directory and observe the data in this file. Take a screenshot of the file contents and paste it below. What kind of data formats would you use for the data in this file?



We can store html, .jpg, .css type of data formats as we can see here.

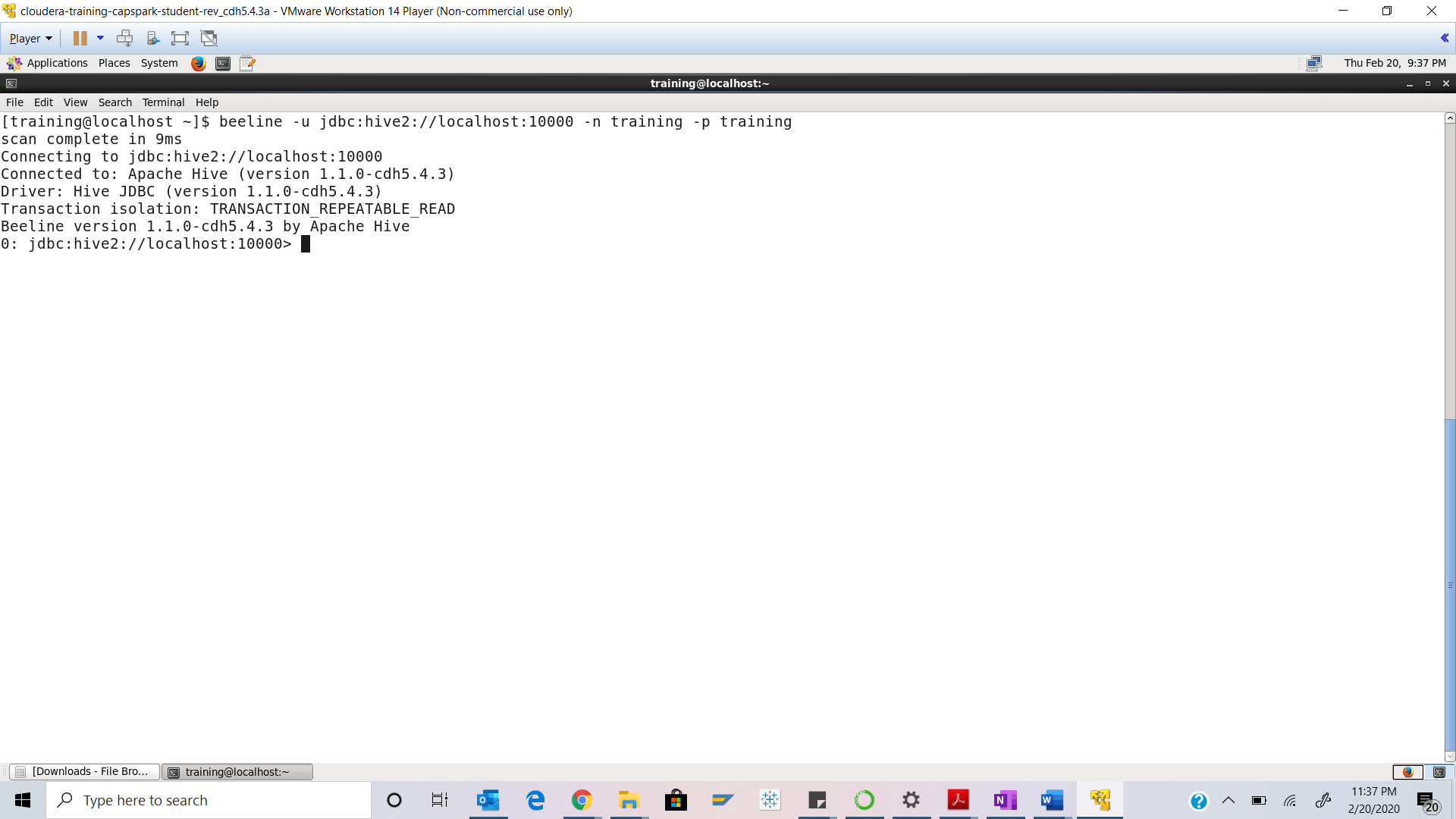
Apart from this Hive supports several file formats:

1. Text File
2. SequenceFile
3. RCFile
4. Avro Files
5. ORC Files
6. Parquet
7. Custom INPUTFORMAT and OUTPUTFORMAT

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



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



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

CREATE EXTERNAL TABLE webpage \

(page\_id SMALLINT, \

name STRING, \

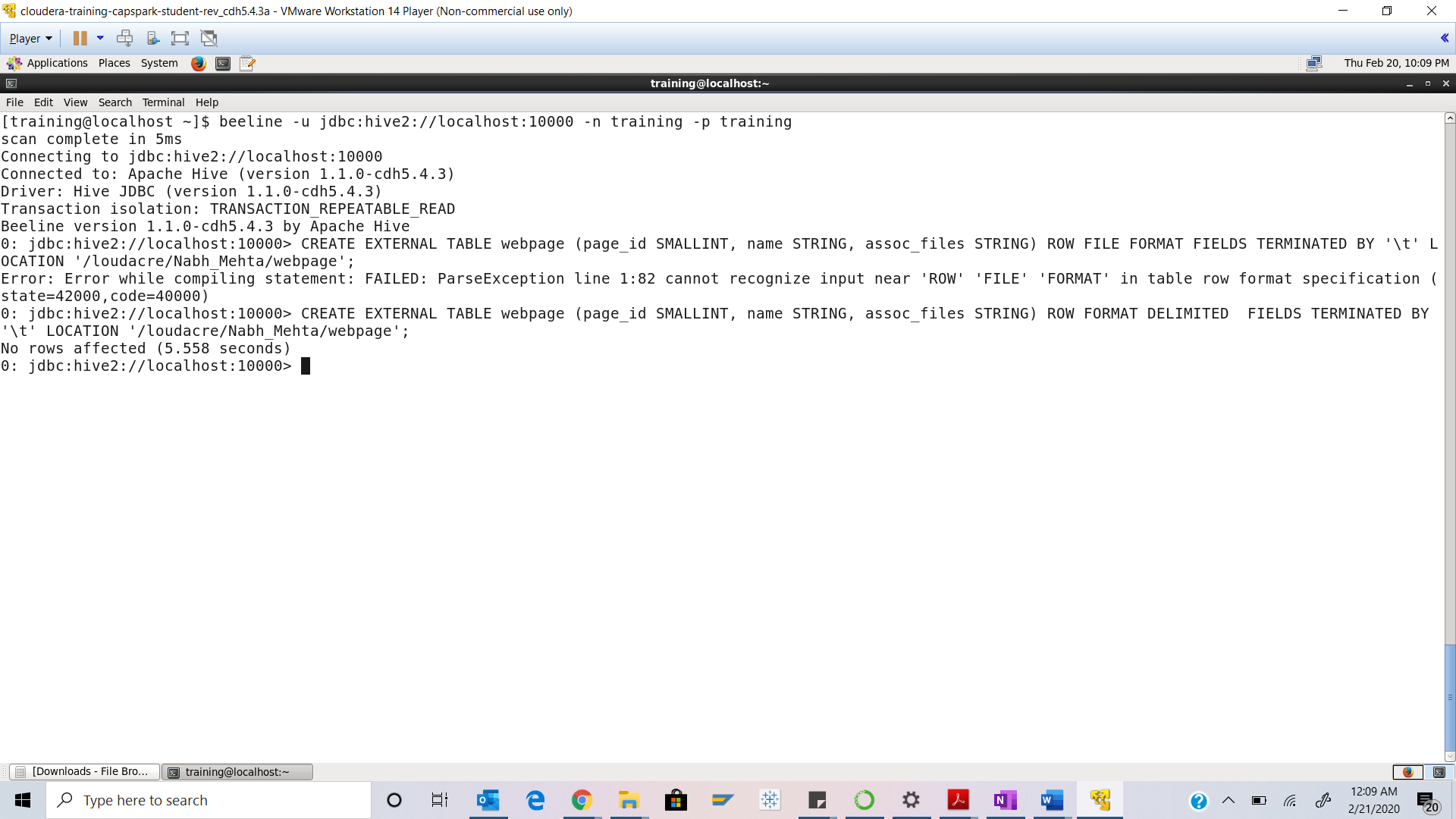
assoc\_files STRING) \

ROW FORMAT DELIMITED \

FIELDS TERMINATED BY ‘\t’ \

LOCATION ‘/loudacre\_FirstName\_LastName/webpage’;

Q. Open the Hive Query Editor and click on the refresh button in the database panel, In the database pane on the left. You should now see the webpage table we just created. Take a screenshot of this screen and paste it below. Highlight the webpage table. Why does the above operation have no results?

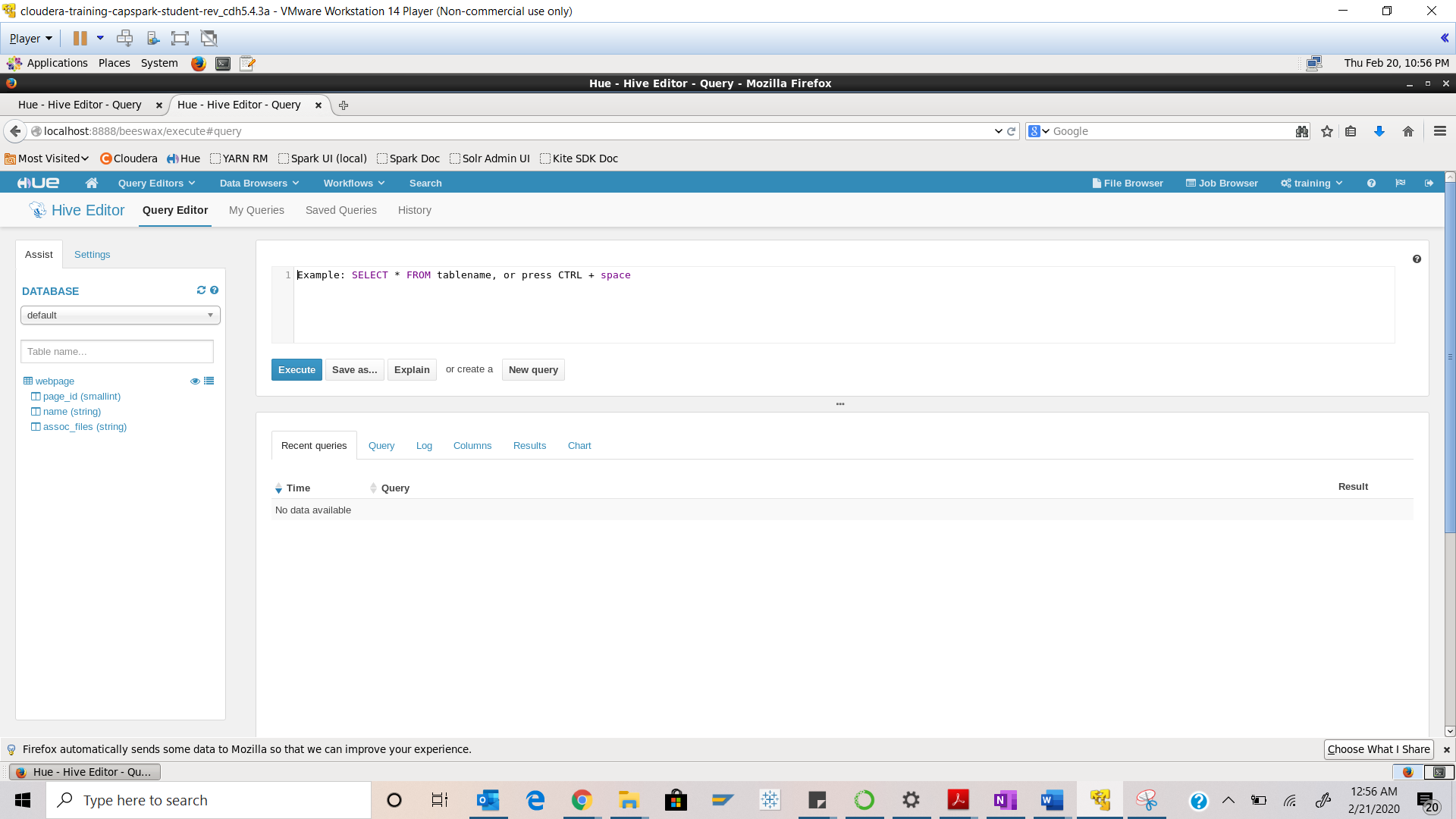




We created an external table in beeline and we have not inserted any data yet.

Step-8: Switch back to the window running the Hive Query Editor and click on the webpage table in the database pane, highlighted in the screenshot in the below.

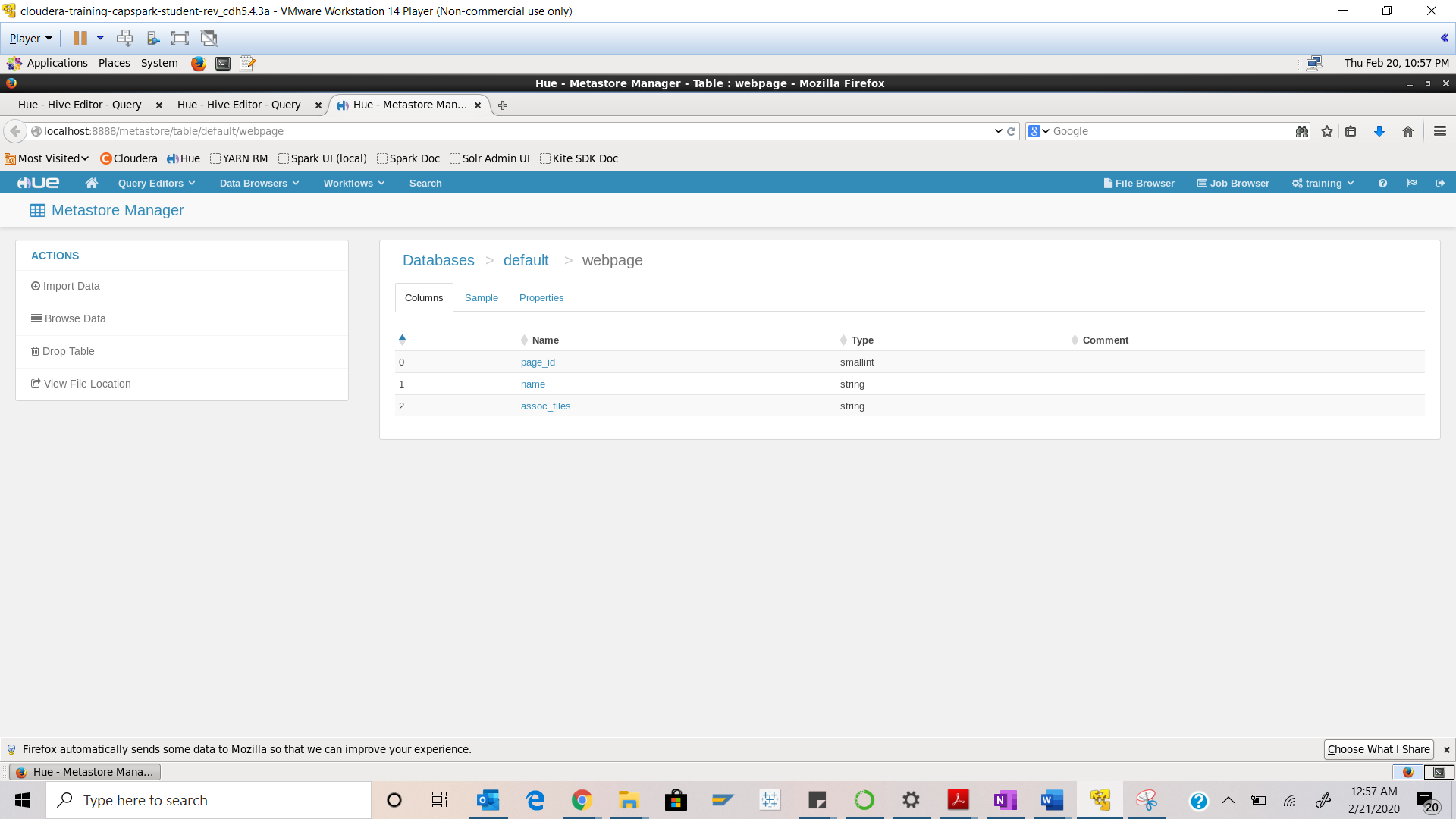
Q. Take a screenshot of this screen and paste it below. Highlight the column definitions of the webpage table.





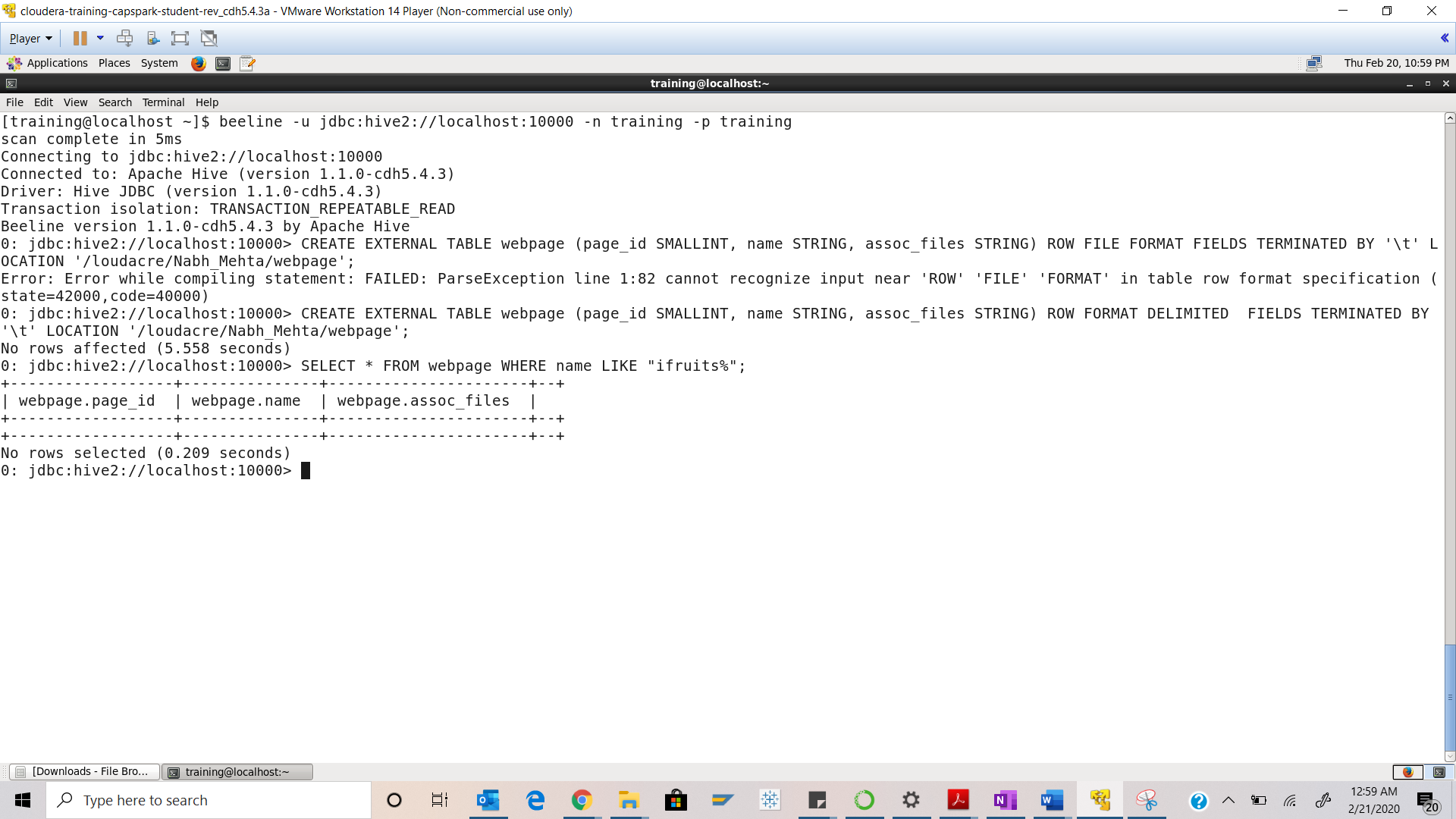
Step-9: Click on the icon highlighted in the screenshot below.

Q. Take a screenshot of the window that pops up and paste it below.

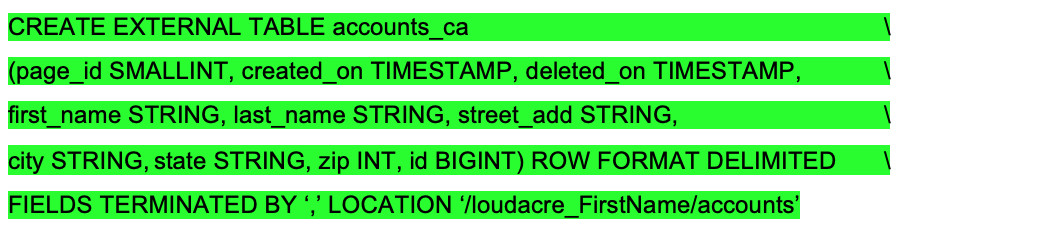


Step-10: Exit the sample data preview, and execute the following query at the beeline shell: 

Q. Take a screenshot of the query results and paste it below.

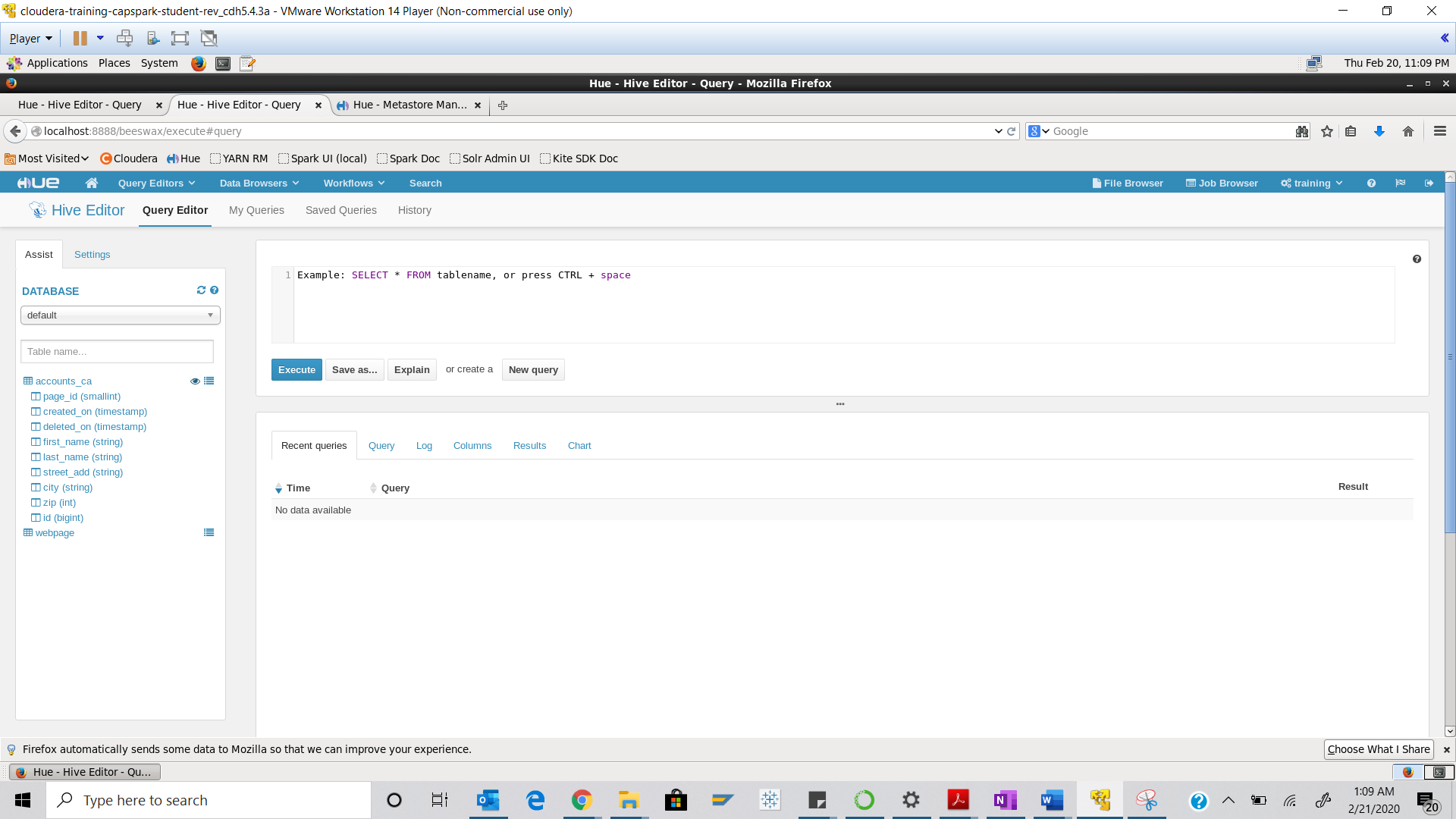


Step-11: Execute the following command at the beeline shell:



Q. Switch back to the window running Hive Editor and refresh the tables in the database pane and take a screenshot of this screen and paste it below. Highlight the accounts\_ca table in the screenshot. Why is the data type for id BIGINT? Instead of INT or SMALLINT.

HINT: Look at the actual data.

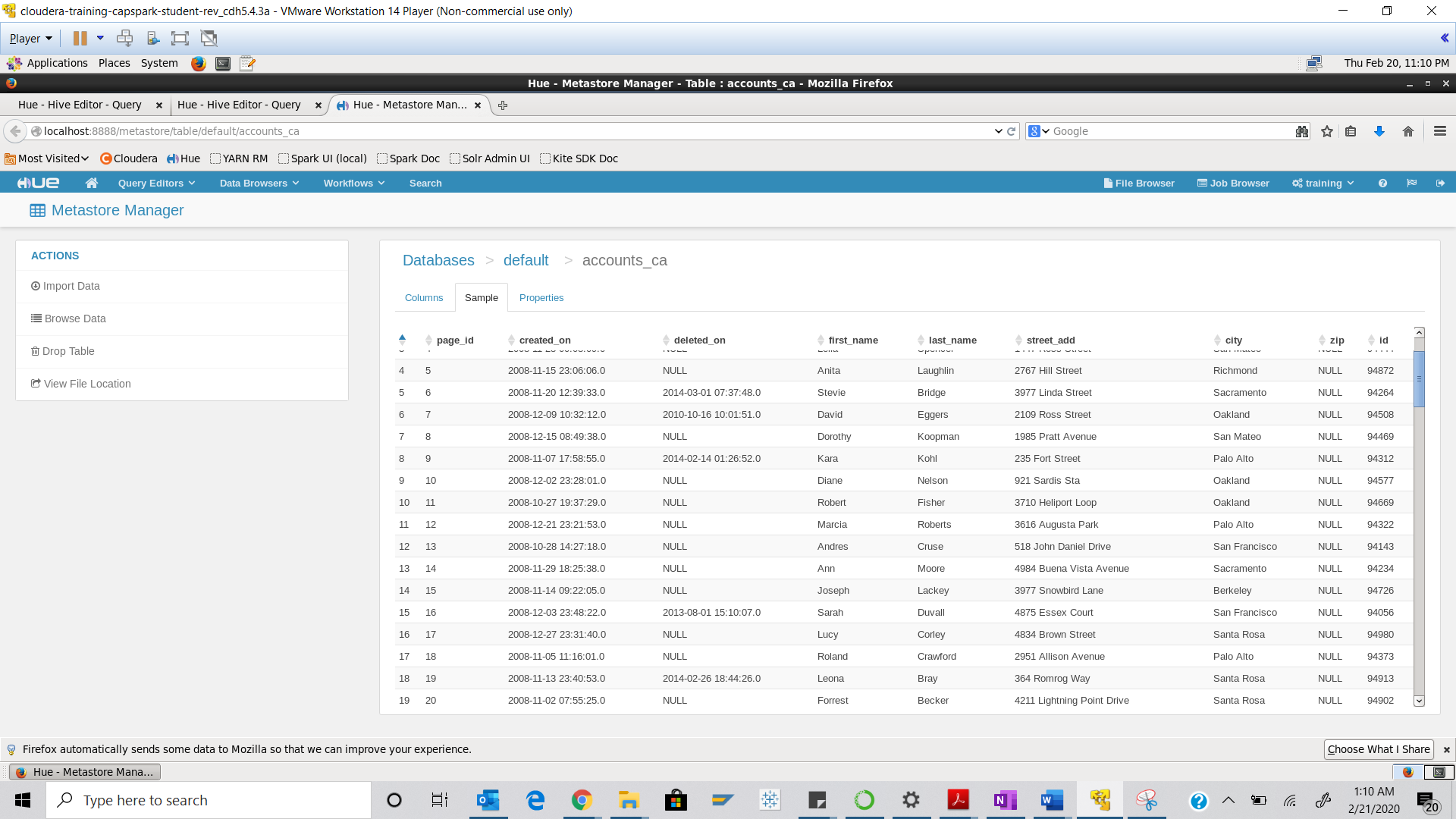




Id can be a bigger value for an integer

Step-12: Click on the ‘Preview Sample Data’ button next to the accounts\_ca table.

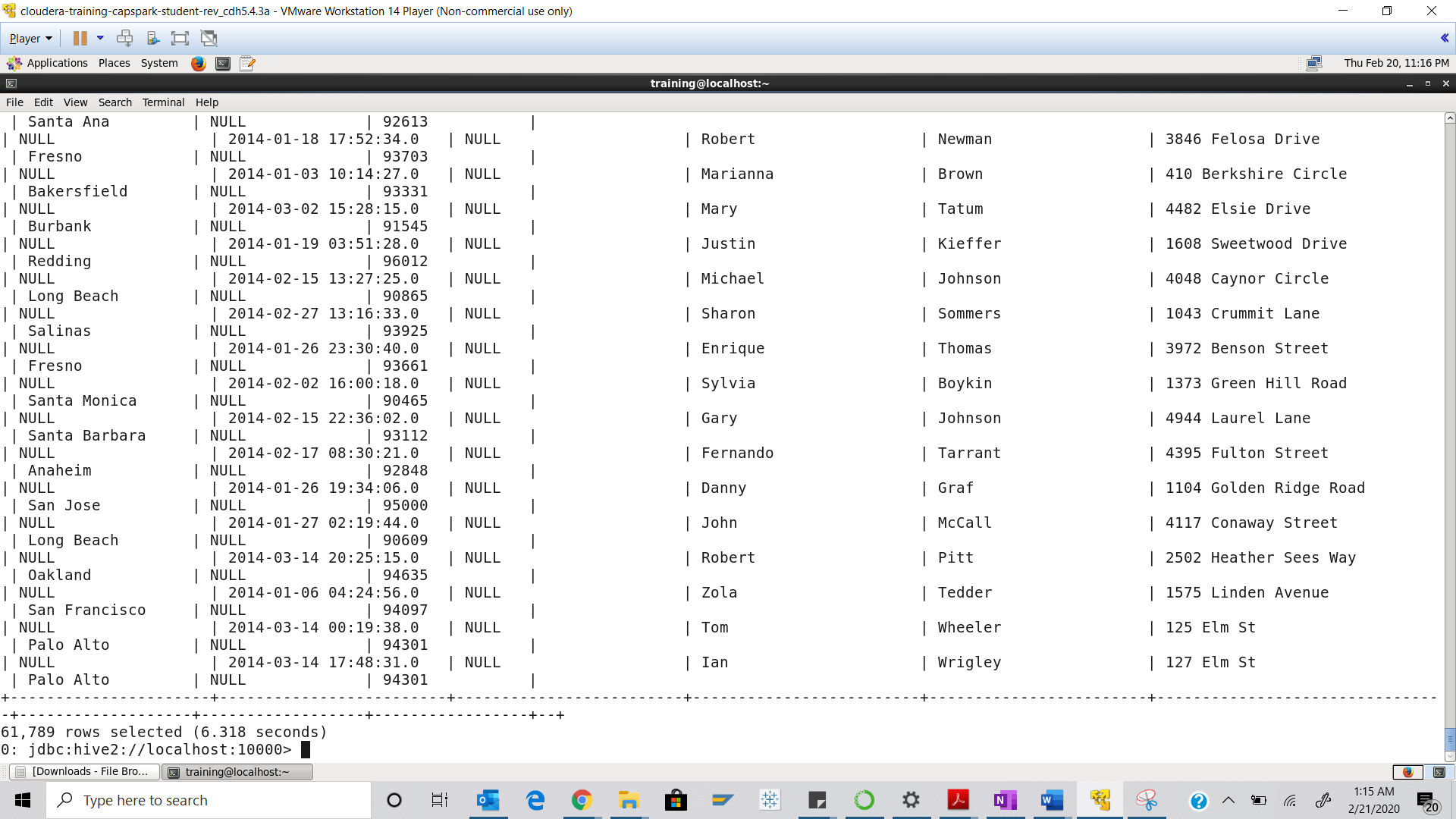
Q. Take a screenshot of the sample data for the accounts\_ca table then paste it below.



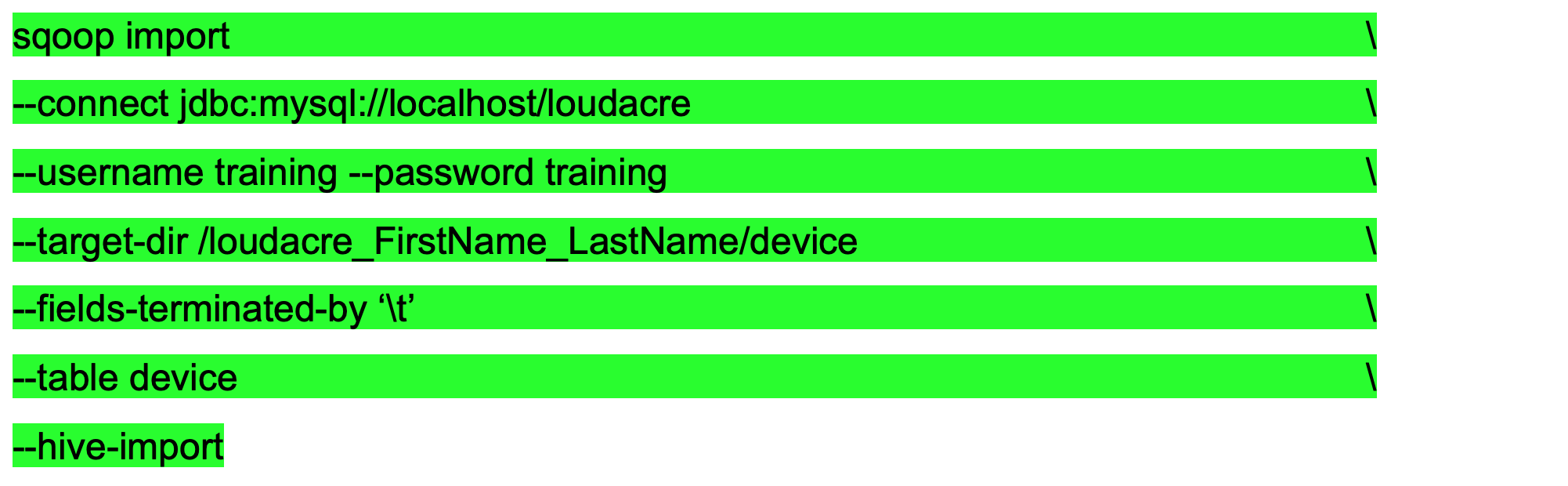
Step-13: Execute the following query in the beeline shell:



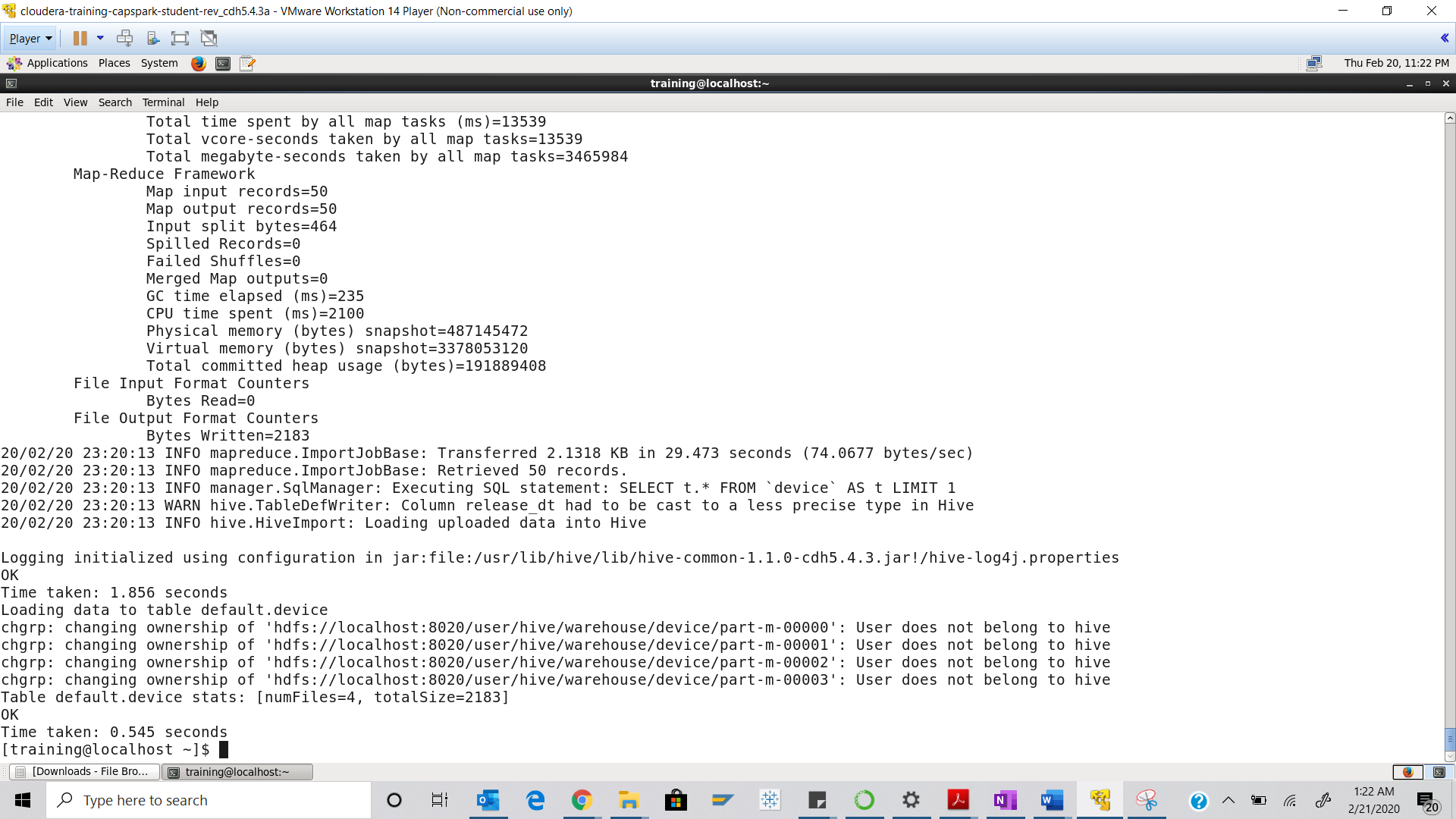
Q. Take a screenshot of the query results and paste it below.



Step-14: Switch back to the terminal window and execute the following command at the shell prompt:



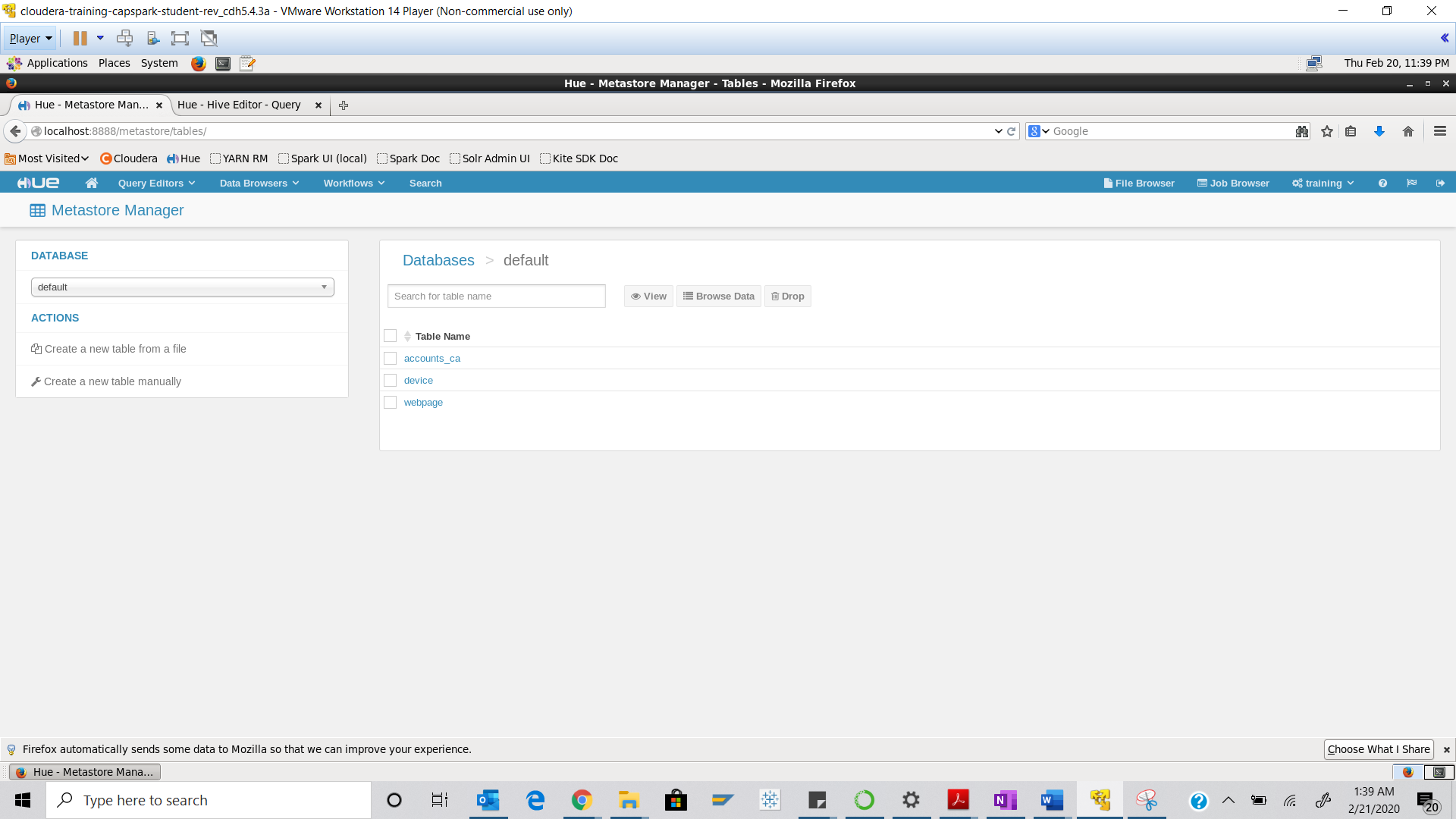
Q. Take a screenshot of the shell output paste it below. Which sqoop tool are we using in the above command?



We are using scoop import tool

Step-15: Switch back to the Firefox tab running the hive query editor. Click on the ‘Data Browsers” button then the ‘Metastore Tables’ button in the HUE menu bar highlighted in the screenshot below.

Q. Do you see the device table in the Metastore? Take a screenshot of this screen and paste it below. Highlight the device table.

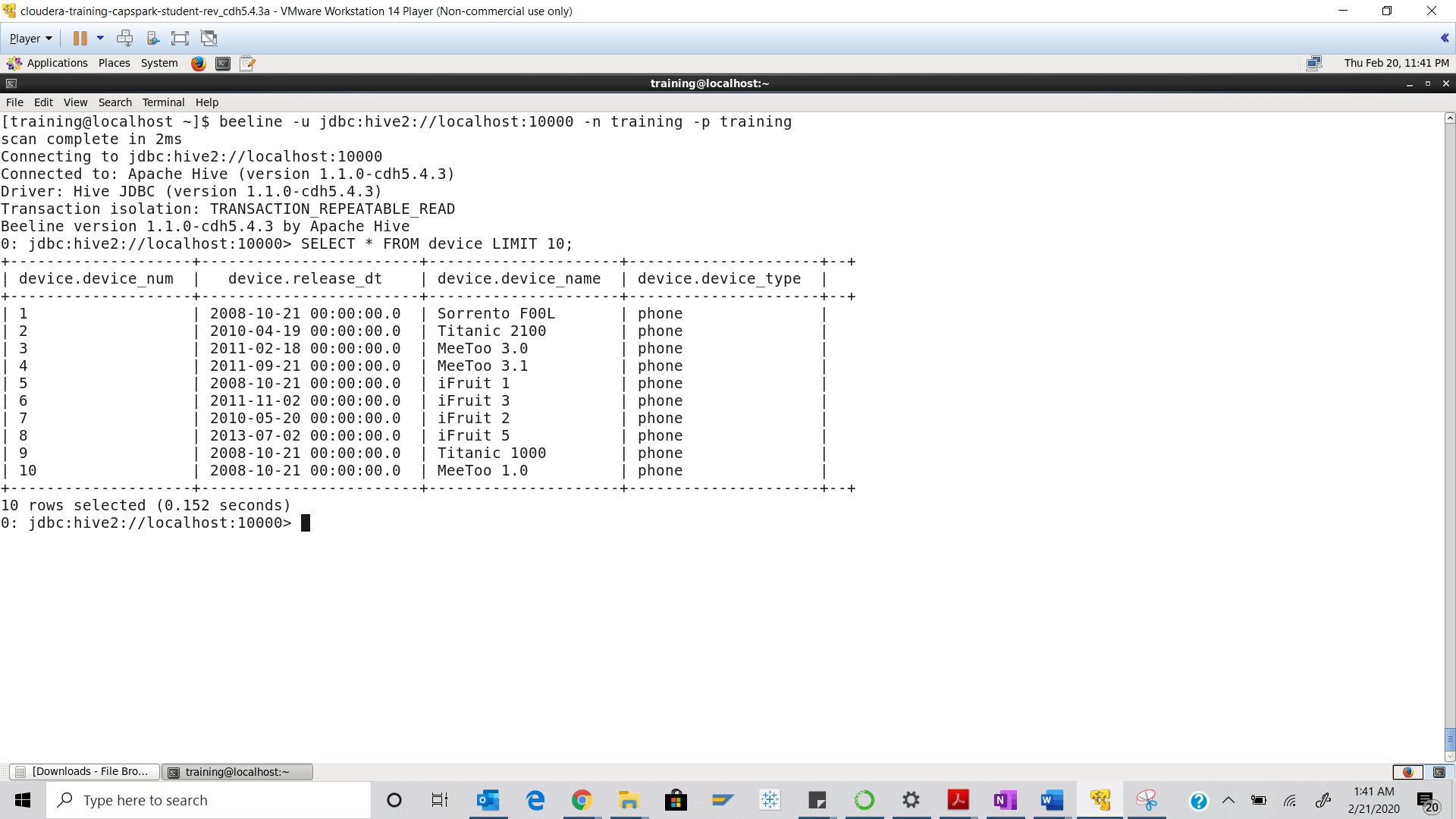




Step-16: Switch back to the beeline shell and run the following query:



Q. Take a screenshot of the query results and paste it below.



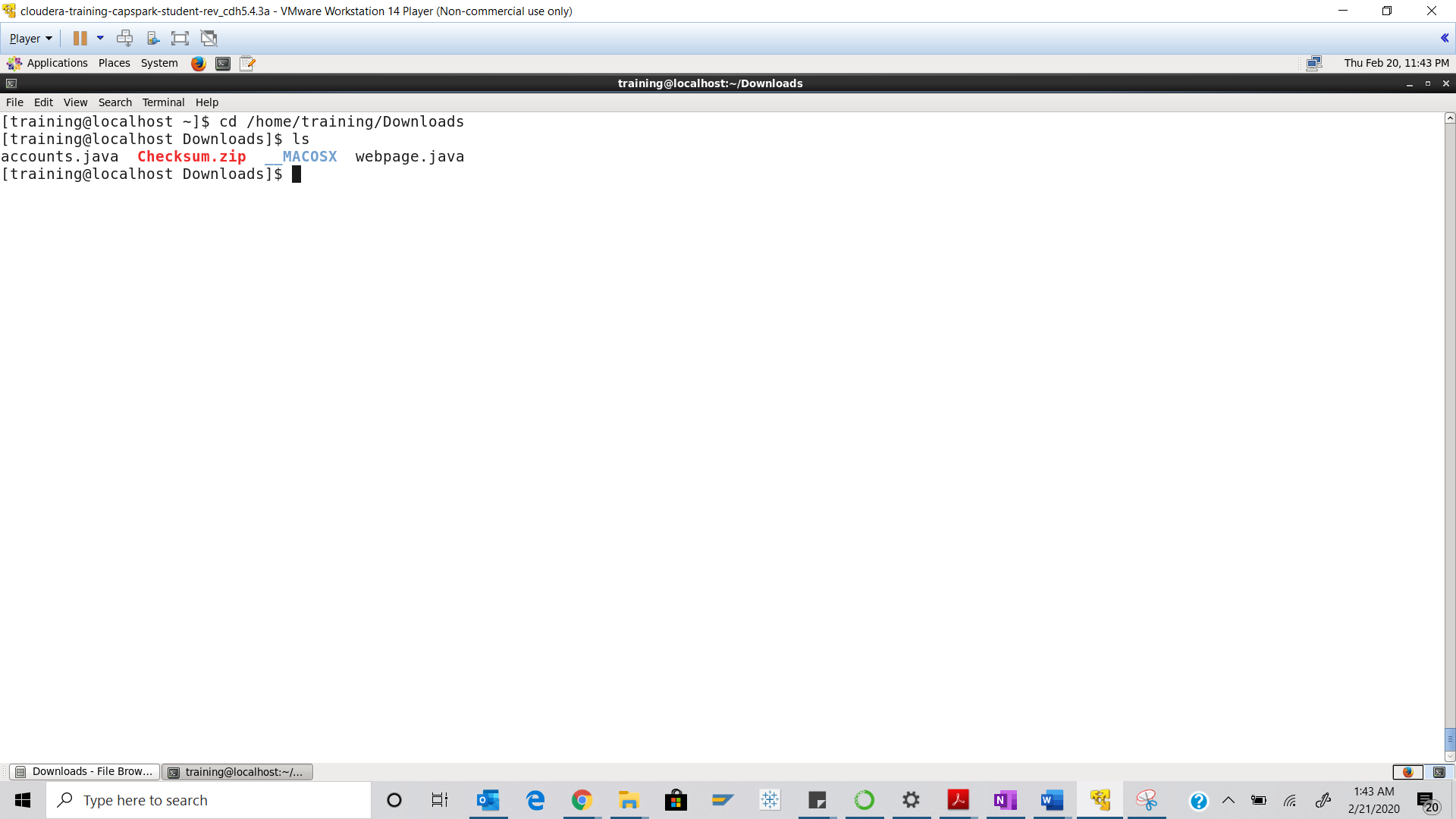
Part-3: Checksum: In this section we will perform a checksum

Step-17: Download the Checksum.zip file from eLearning in to the downloads folder **IN** the Cloudera VM using Mozilla Firefox.

Step-18: 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.

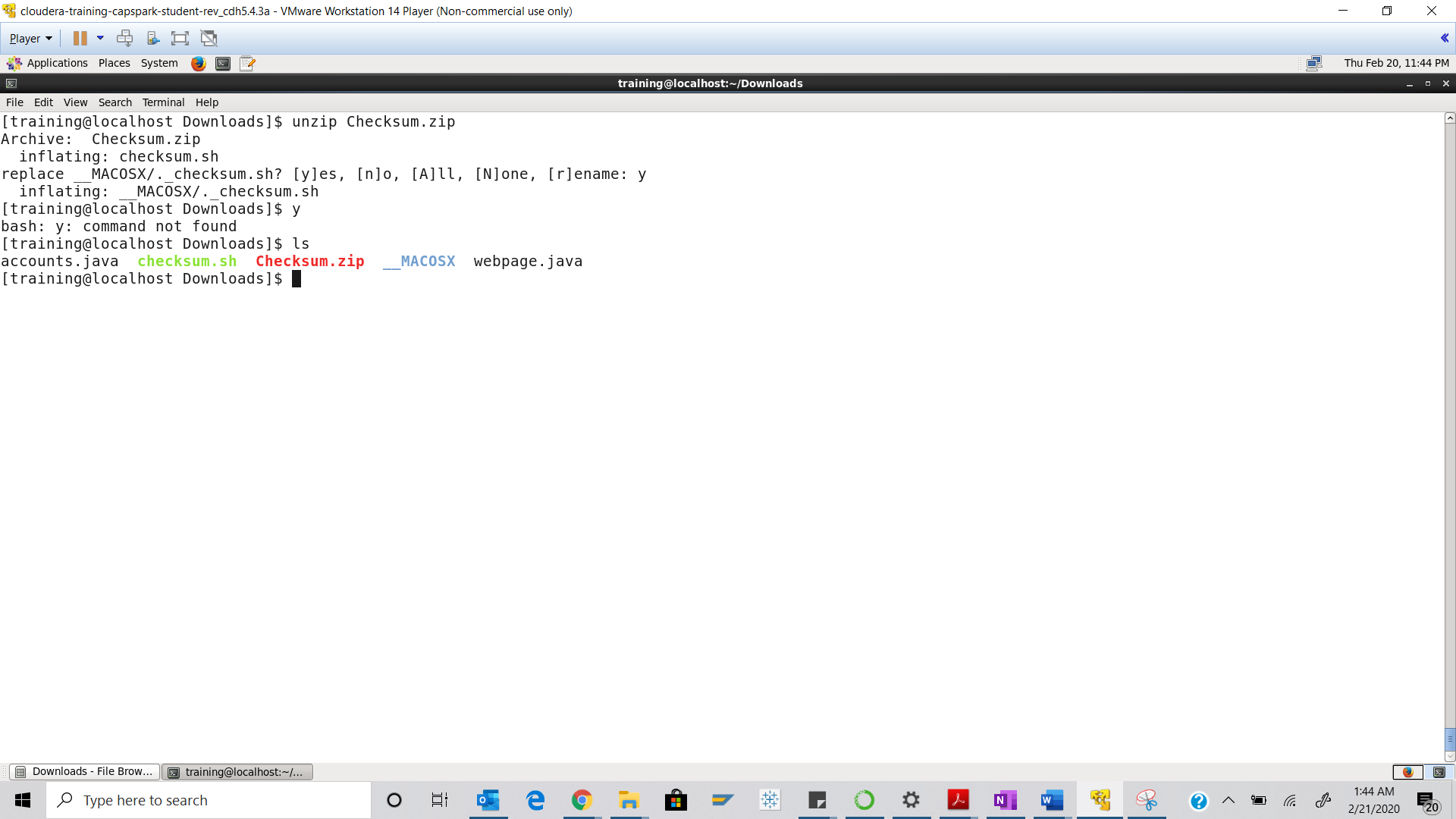




Step-19: Execute the following command at the shell prompt:



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.



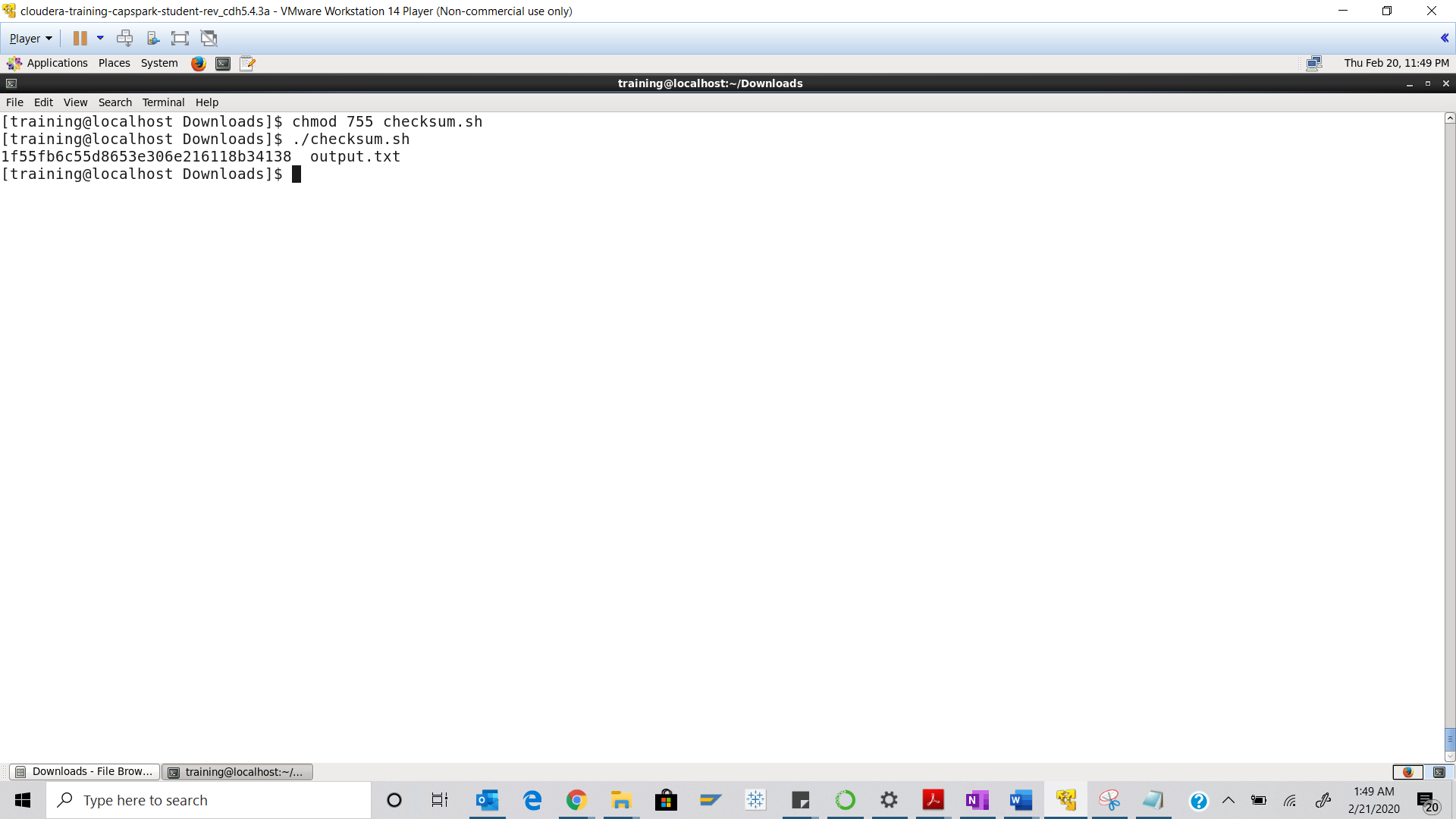


Step-20: Execute the following command at the shell prompt:

Then execute the following command:

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

Check sum value - 1f55fb6c55d8653e306e216118b34138



Step-21: 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.

