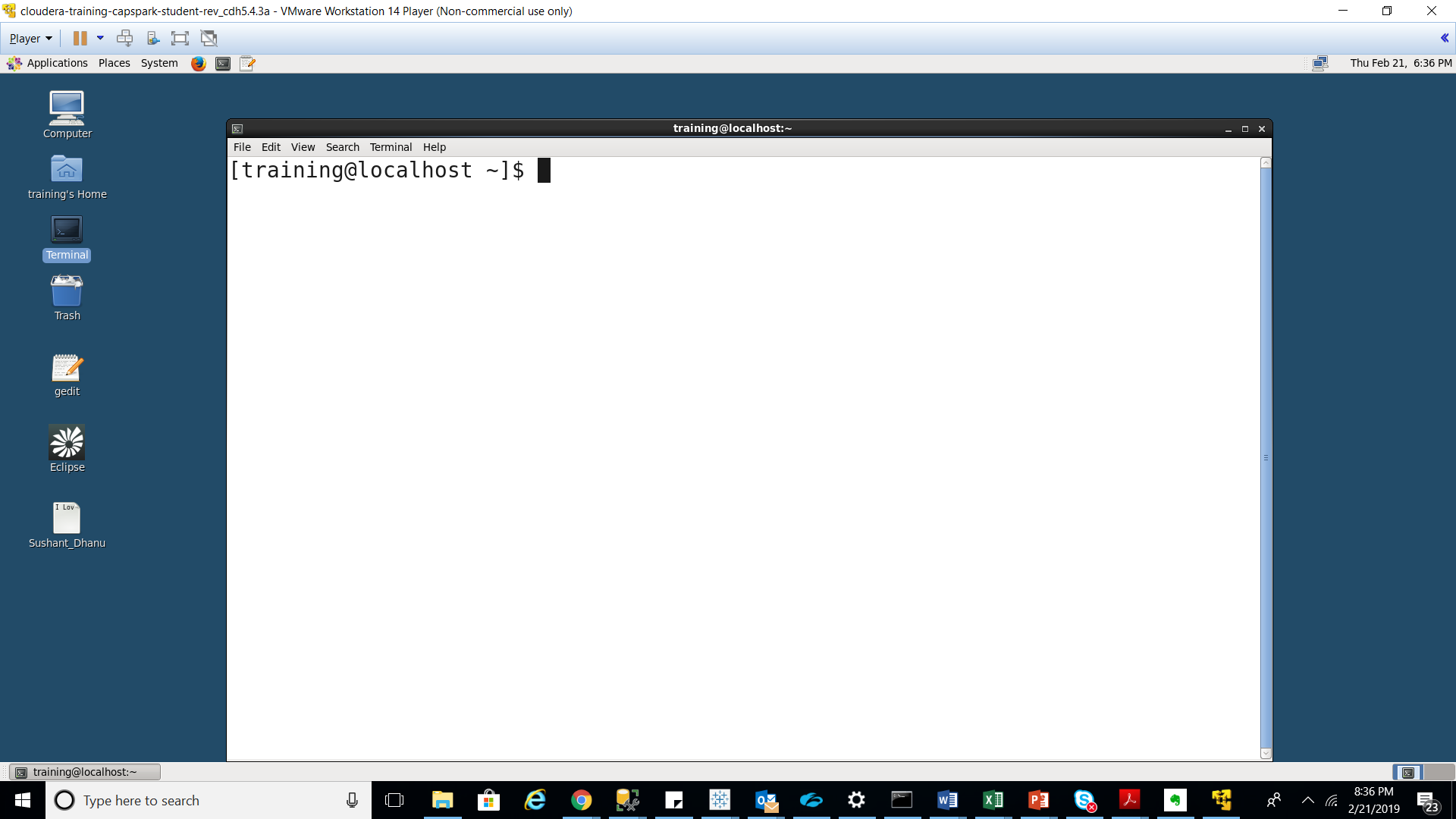
**Sqoop**

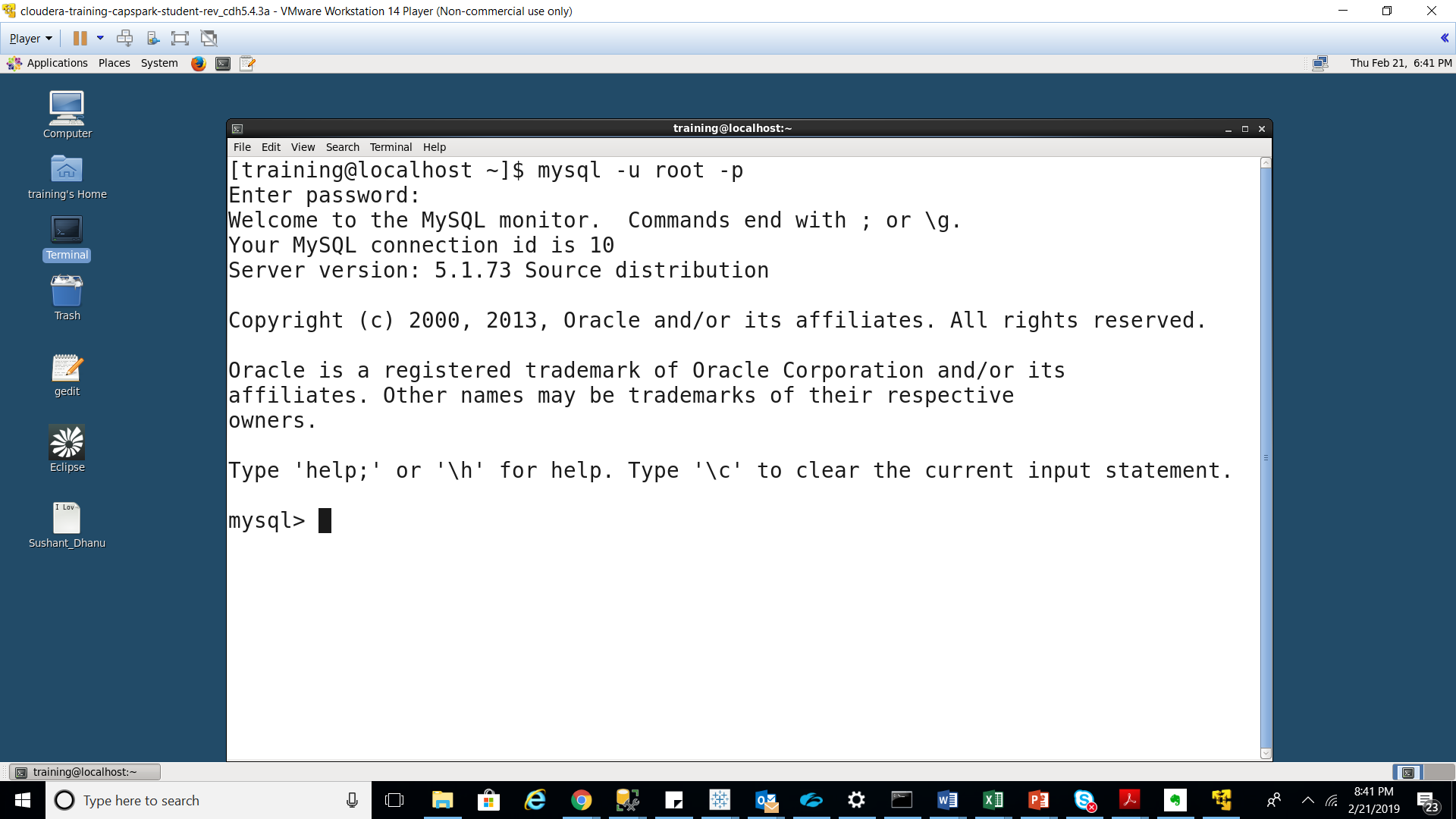
Open the terminal application on the Cloudera VM desktop.



The following command is

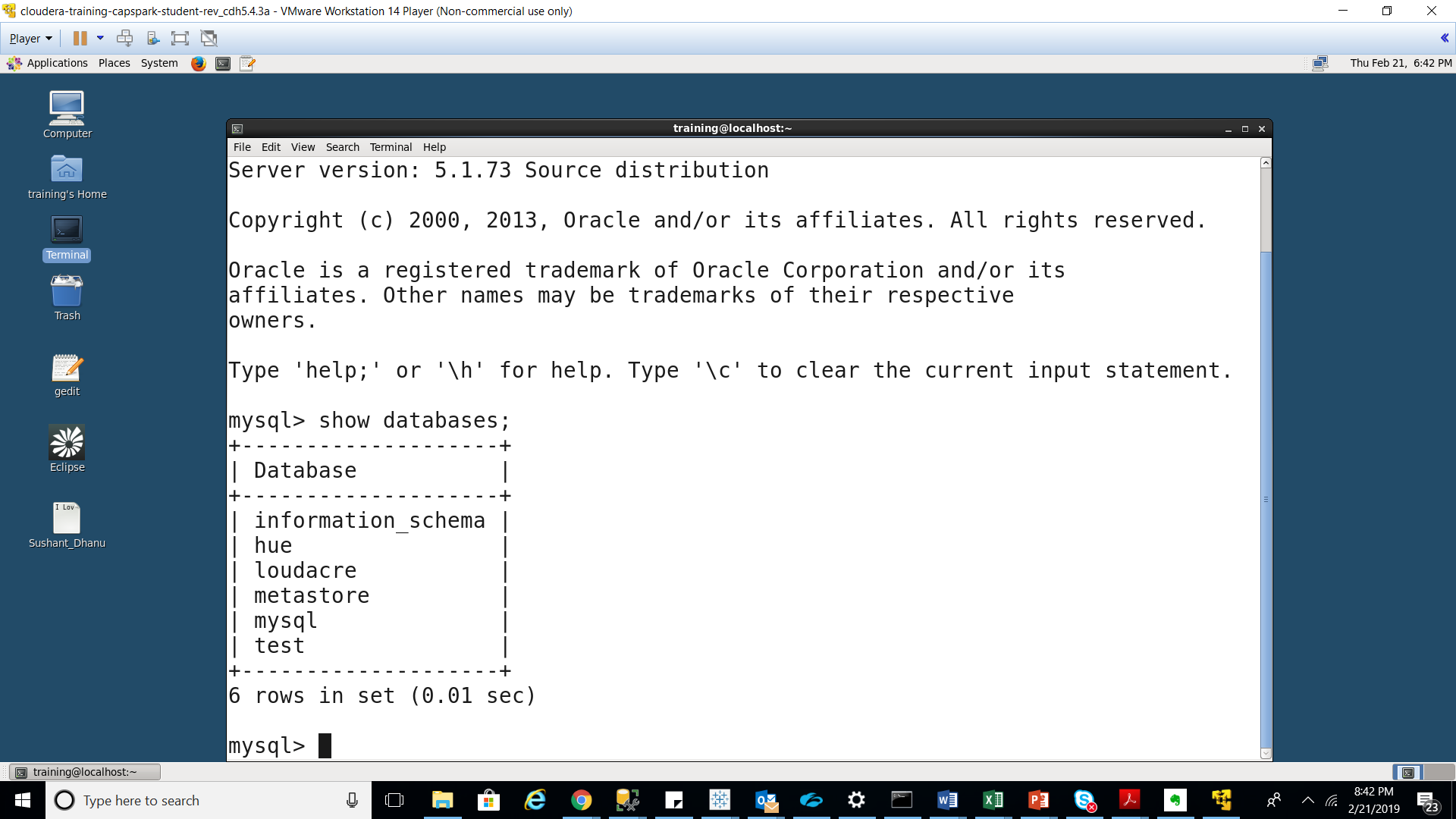
mysql -u root -p

validates that MySQL is installed on our system by connecting us to the MySQL application. The -u is a parameter that specifies the username. The -p is a parameter that specifies the password.



show databases;

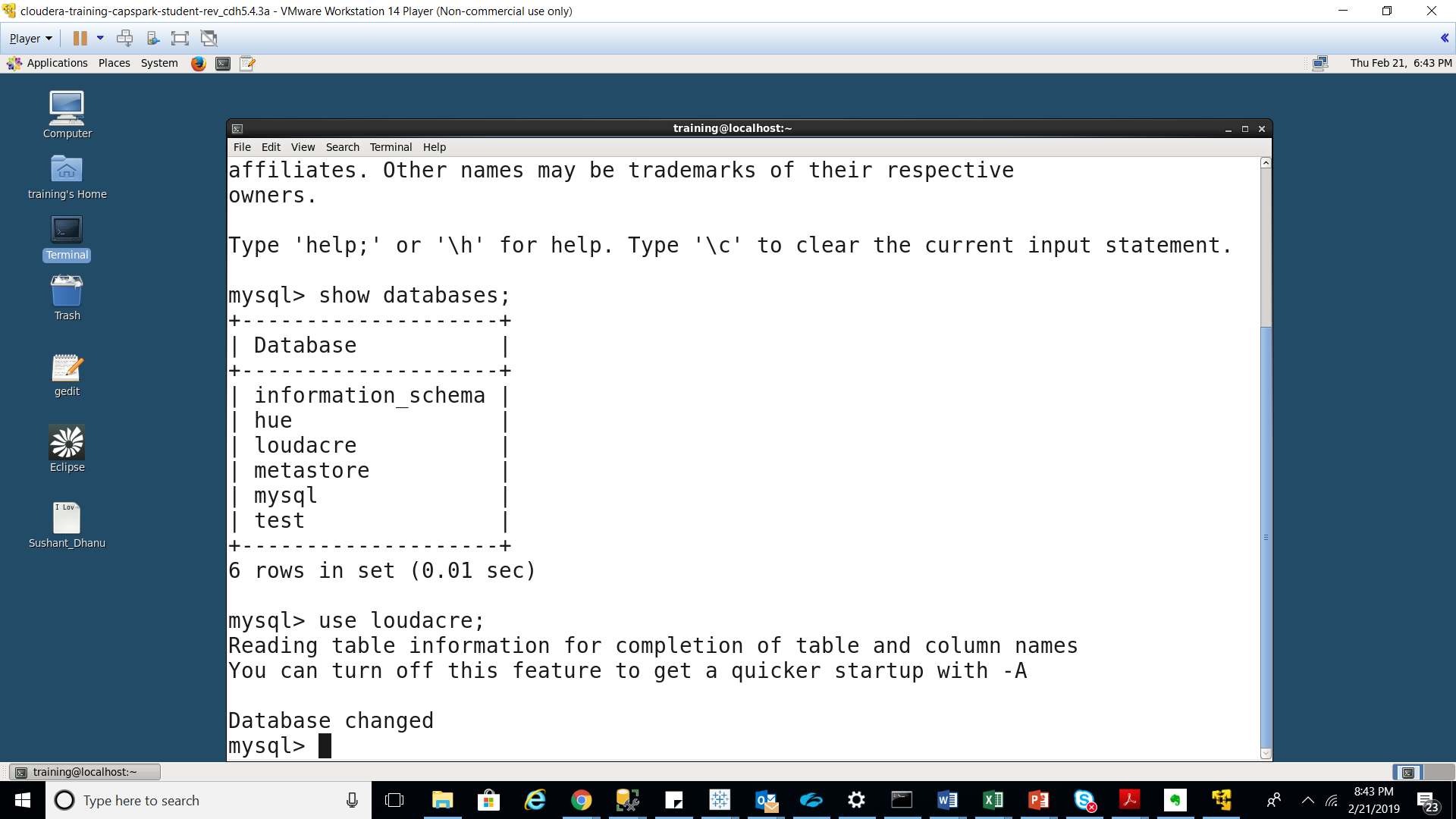
The above command lists all the databases on the system.



Execute the following command next

use loudacre;

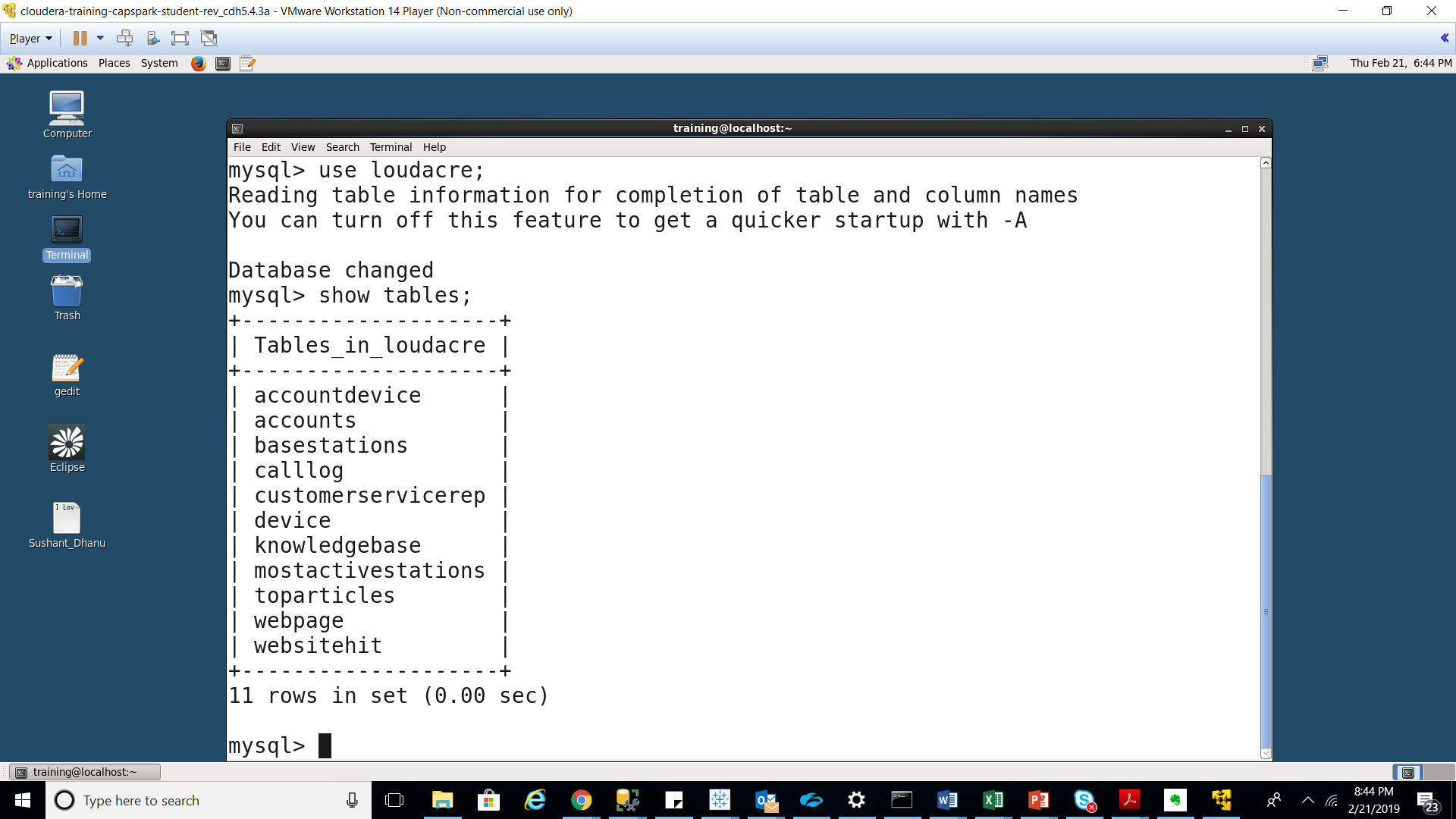
There may be more than one database, the “use” command specifies which table we want to perform our operations on next



Execute the following command

show tables;

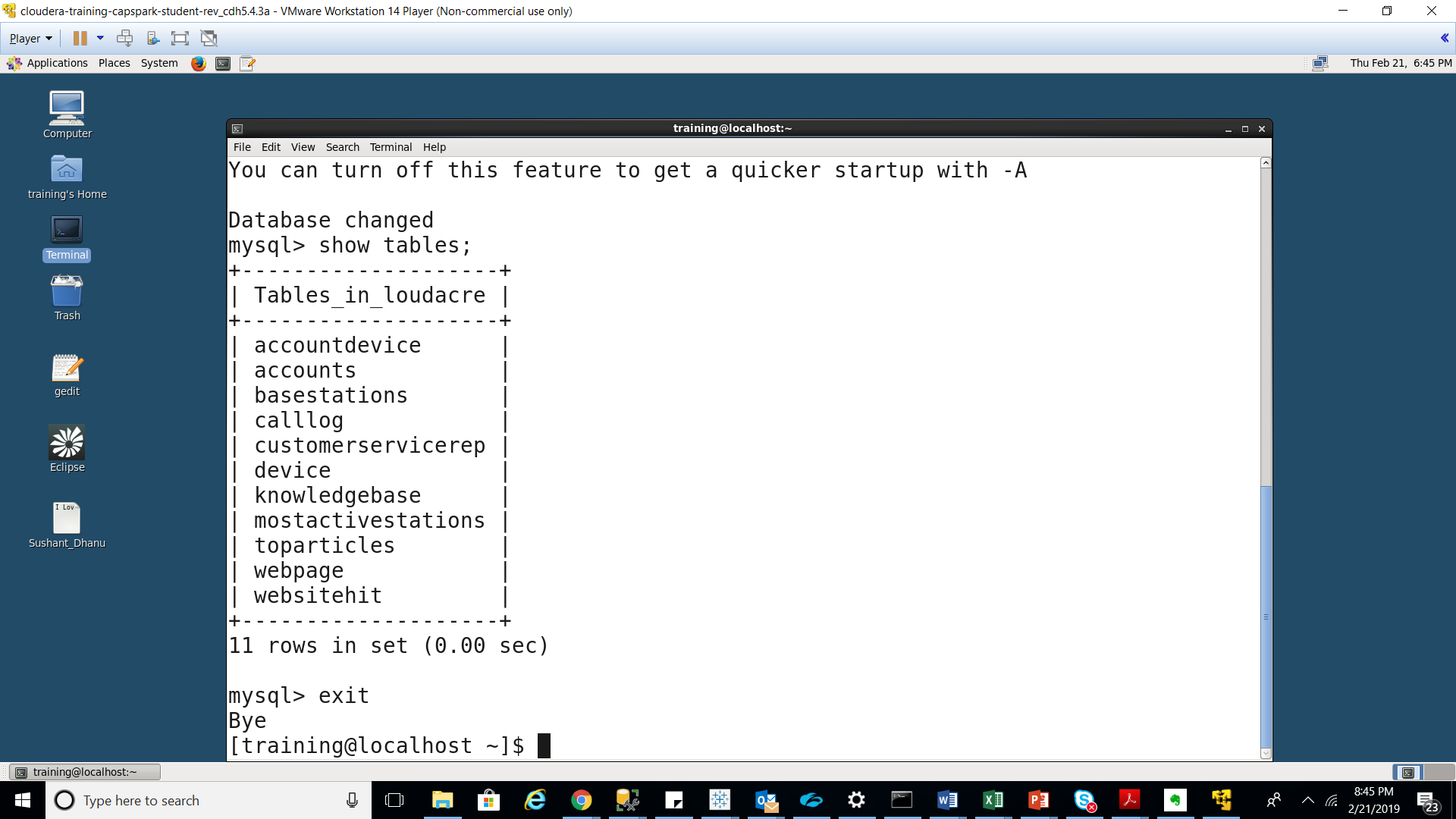
The above command lists all the tables in the selected database.



Use the following command

“Exit”

The command exits the MySQL application.



Excecute the following command

Sqoop import \

--connect jdbc:mysql://localhost/loudacre \

--username training –password training \

--table accounts \

--target-dir /loudacre\_Sushant\_Dhanu/accounts \

--null-non-string ‘\\N’

Explanation: We are importing the accounts table from the loudacre database into HDFS using the Sqoop import command.

Line 1: This line specifies we will be using the Sqoop import tool.

Line 2: This line connects to the loudacre database using the JDBC MySQL connector.

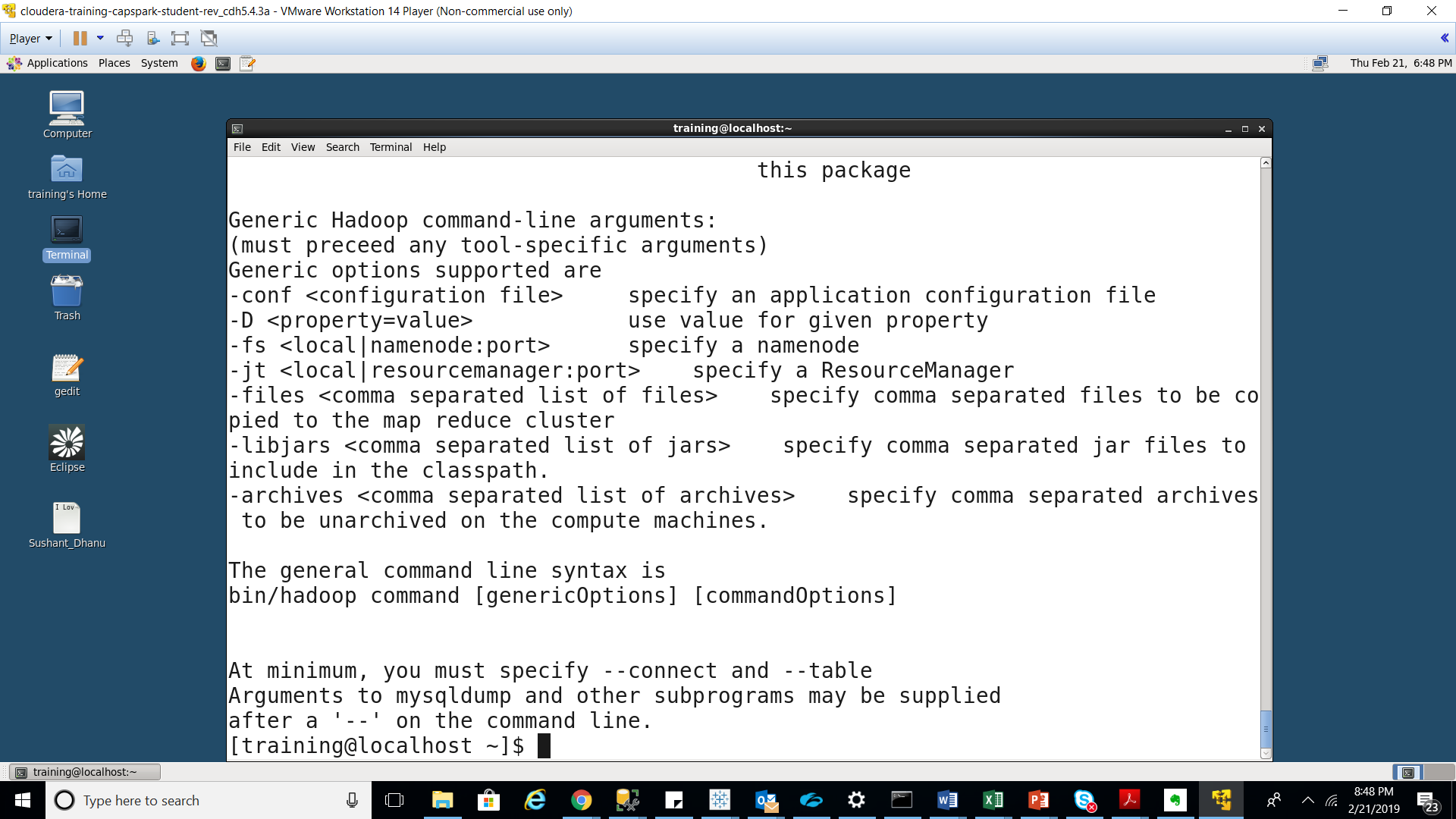
Line 3: This line specifies the username and the password for the database.

Line 4: This line specifies the table in the database we want to import.

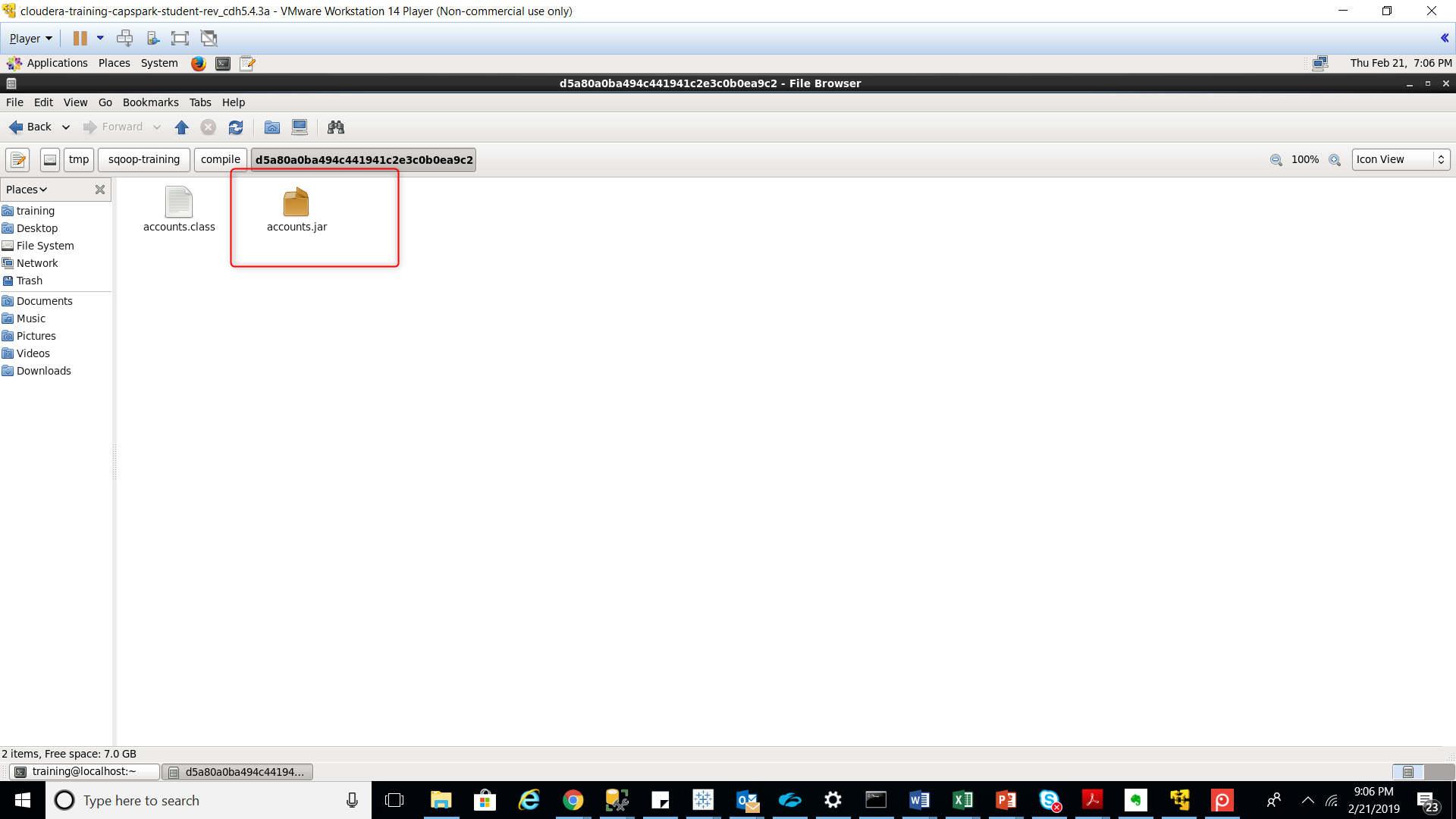
Line 5: This line specifies the location in HDFS that we would like to import the database table

Line 6: This line tells Sqoop to represent null values as /N

Output of the command as below



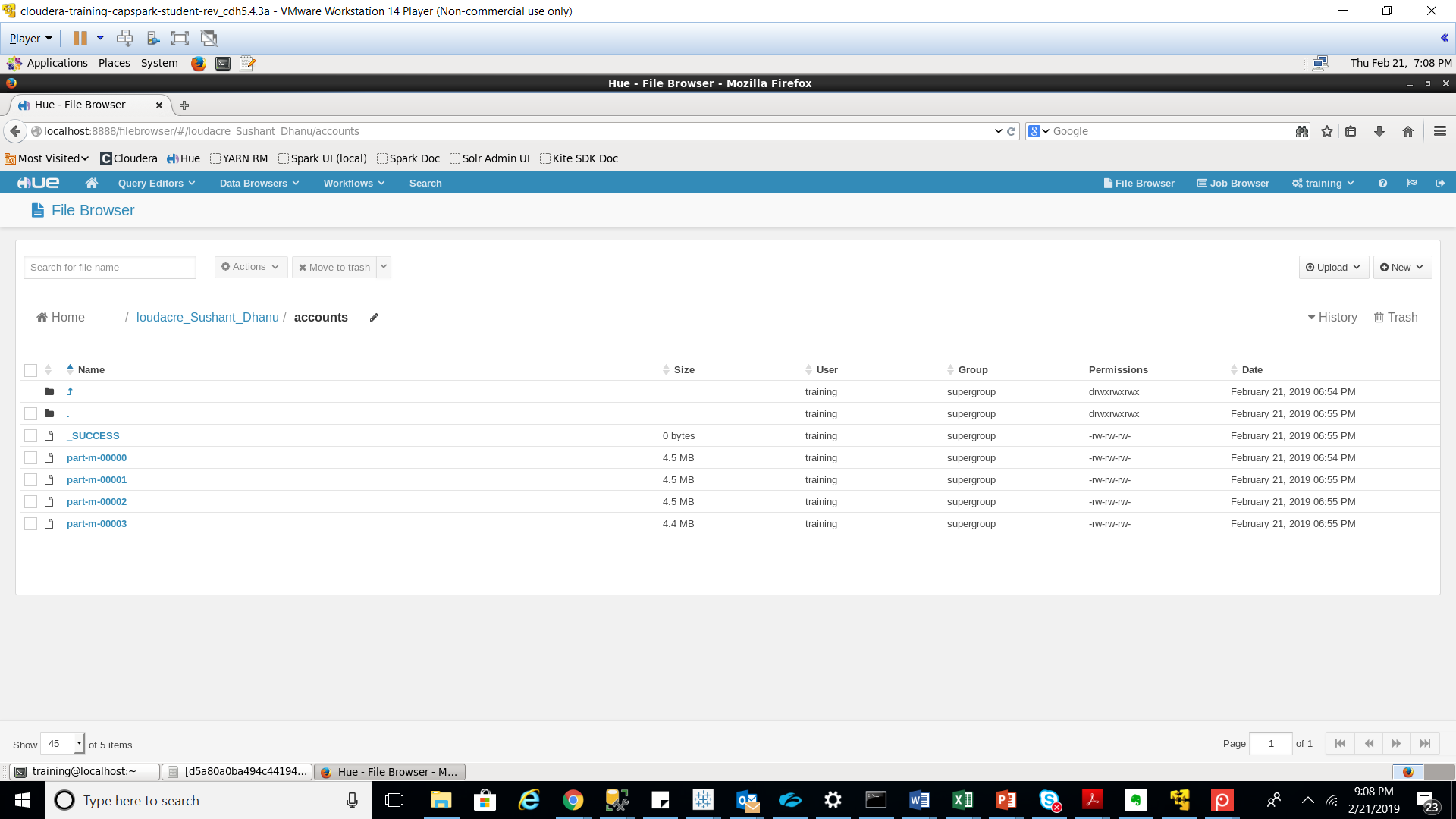
Sqoop generates a java source file for each table being imported, it also uses that file during the import process and in subsequent MapReduce processing



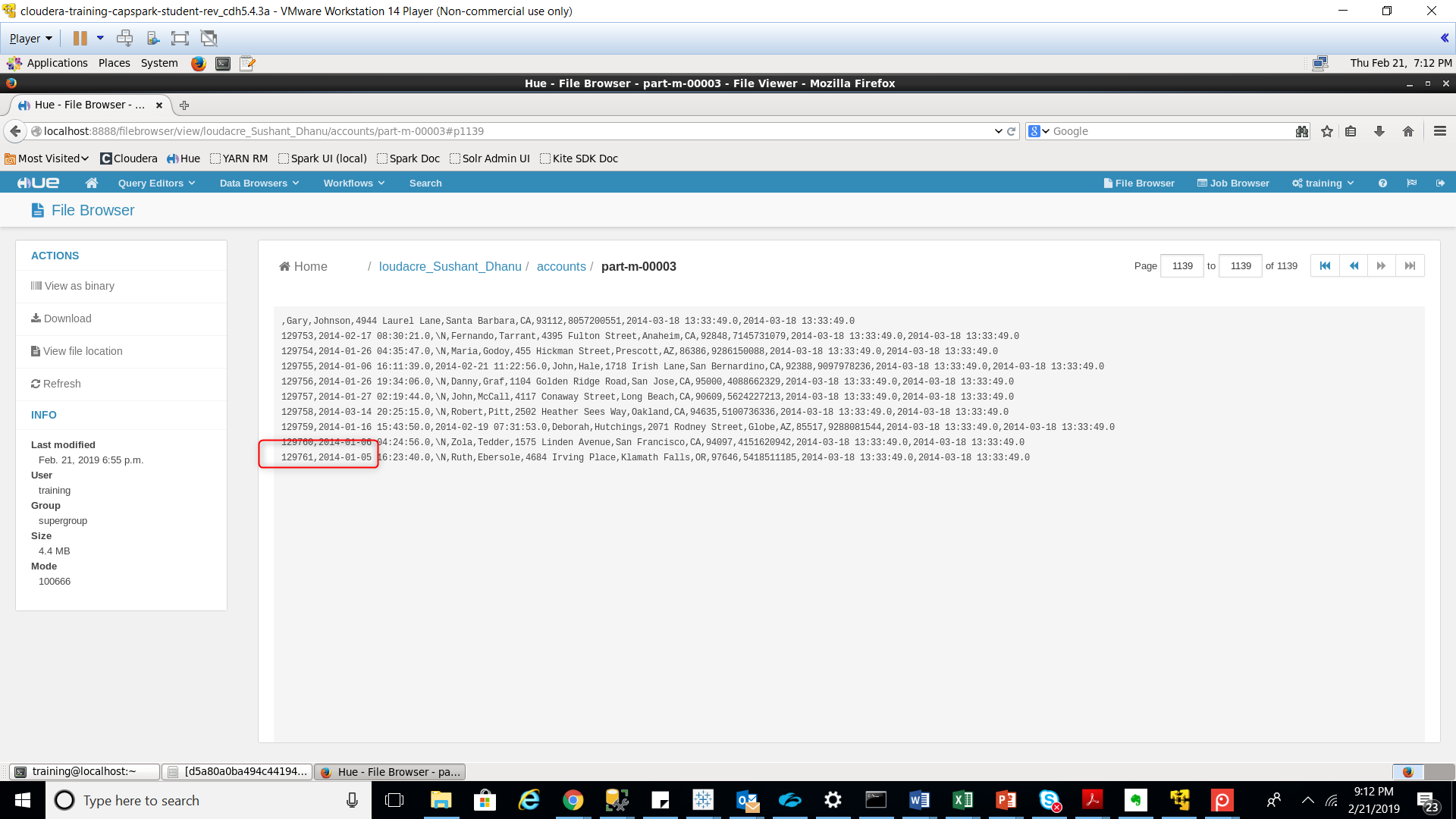
Using the file browser in HUE, navigate to the following path in HDFS:

/loudacre\_Sushant\_Dhanu/accounts

We see that 4 files are created because 4 mappers are used based on the size of the data.



Now open the last file and navigate to the last record of that file. Take a note of this Account # 129761.



Execute the following command in the terminal

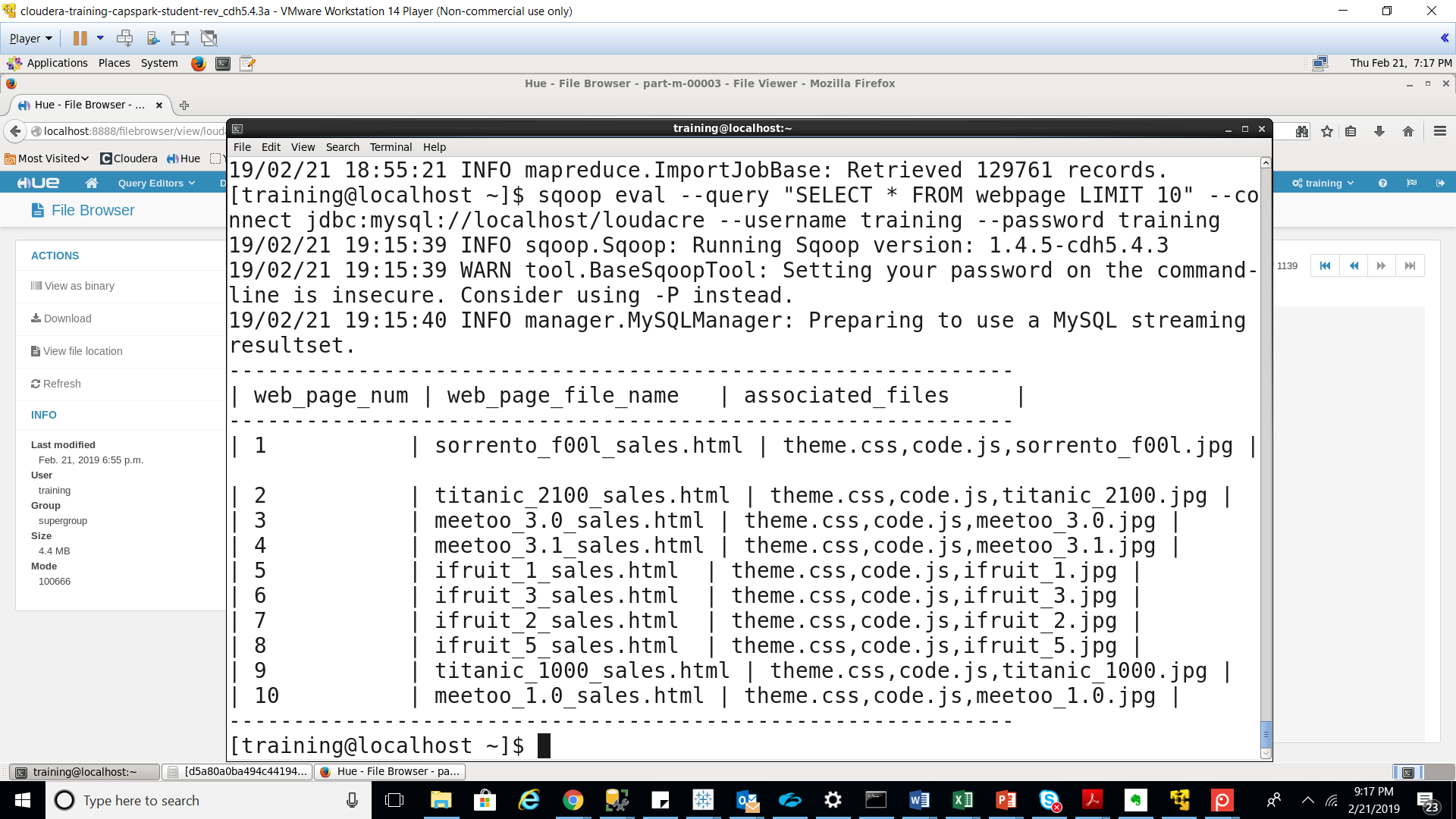
Sqoop eval \

--query “SELECT \* FROM webpage LIMIT 10” \

--connect jdbc:mysql://localhost/loudacre \

--username training –password training

The Sqoop eval tool evaluates a SQL statement and displays the results in the terminal.



Execute the following command

Sqoop import \

--connect jdbc:mysql://localhost/loudacre \

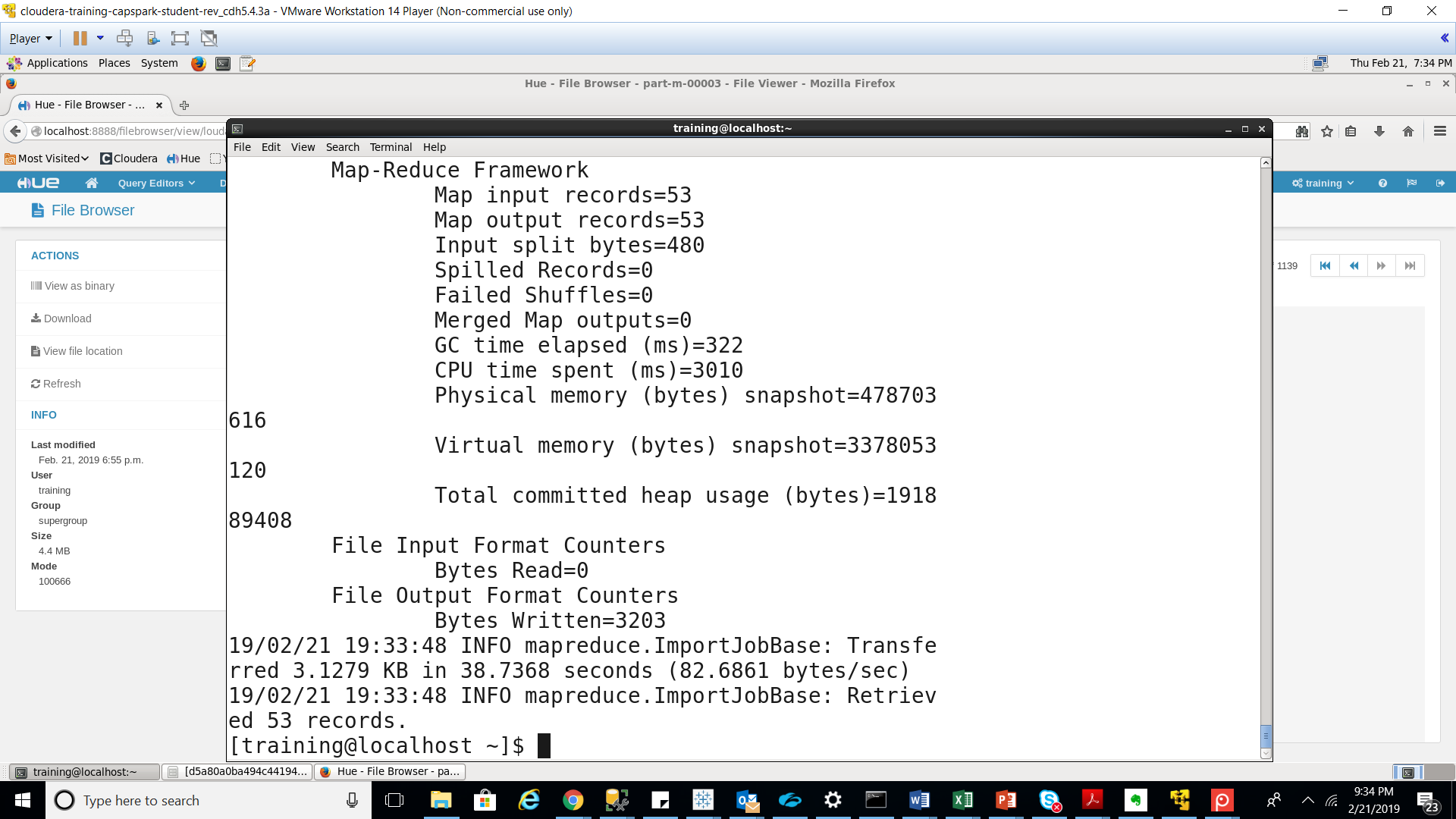
--username training –password training \

--table webpage \

--traget-dir /loudacre\_Sushant\_Dhanu/webpae \

--fields-terminated-by “\t”

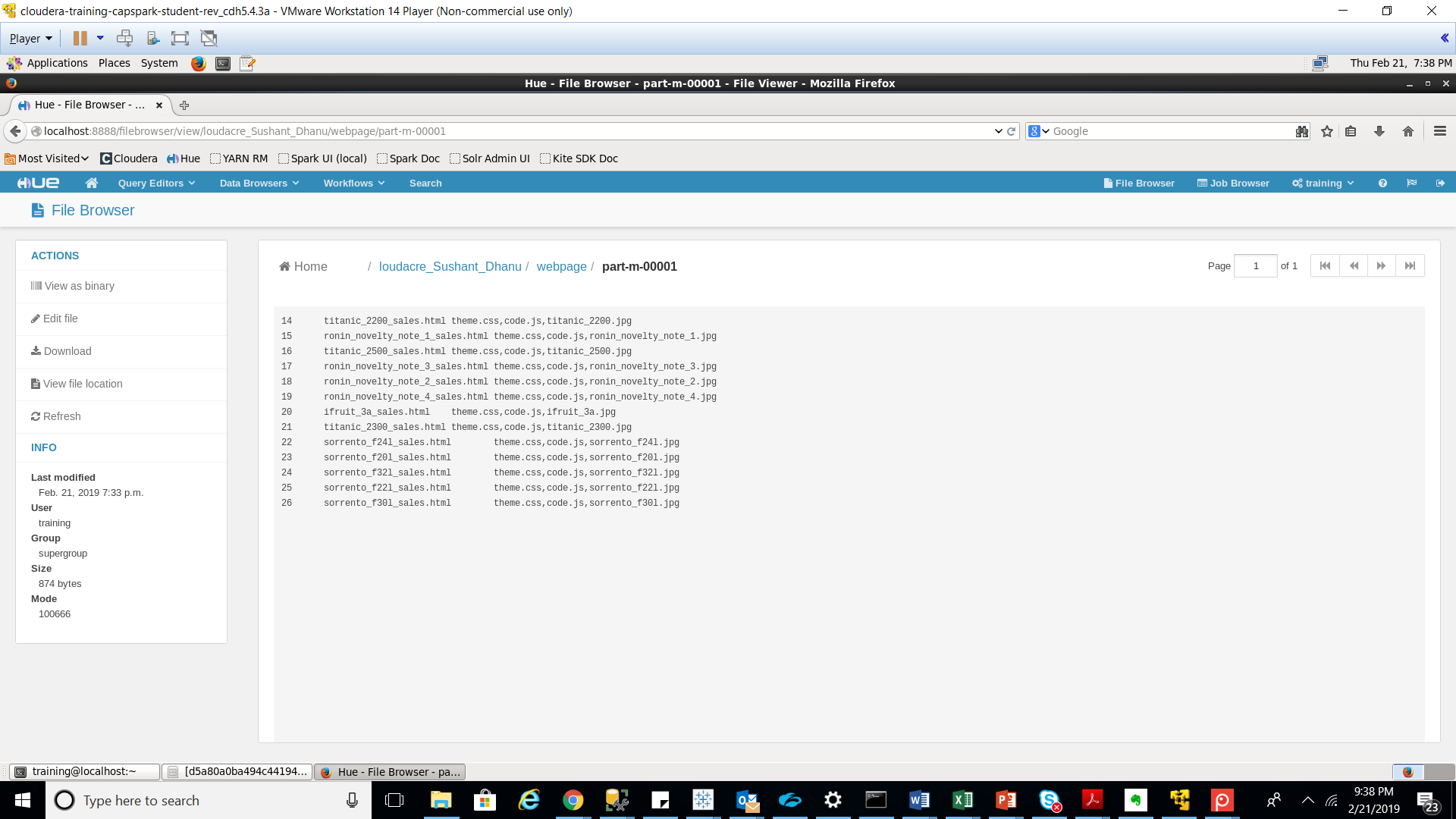
We are importing the webpage table into HDFS as a tab delimited file instead of a comma delimited file



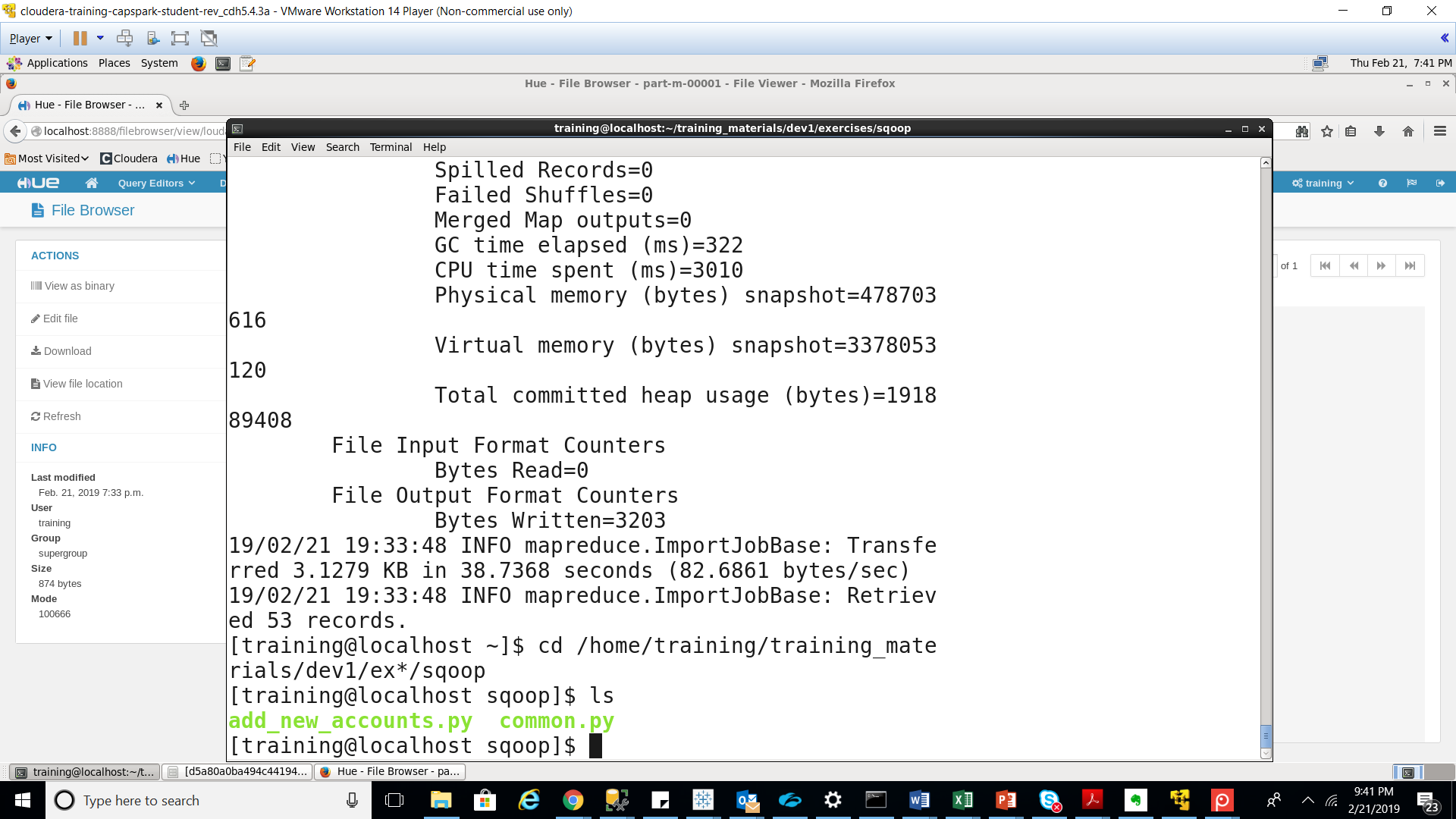
Using the file browser in HUE, navigate to the following path in HDFS:

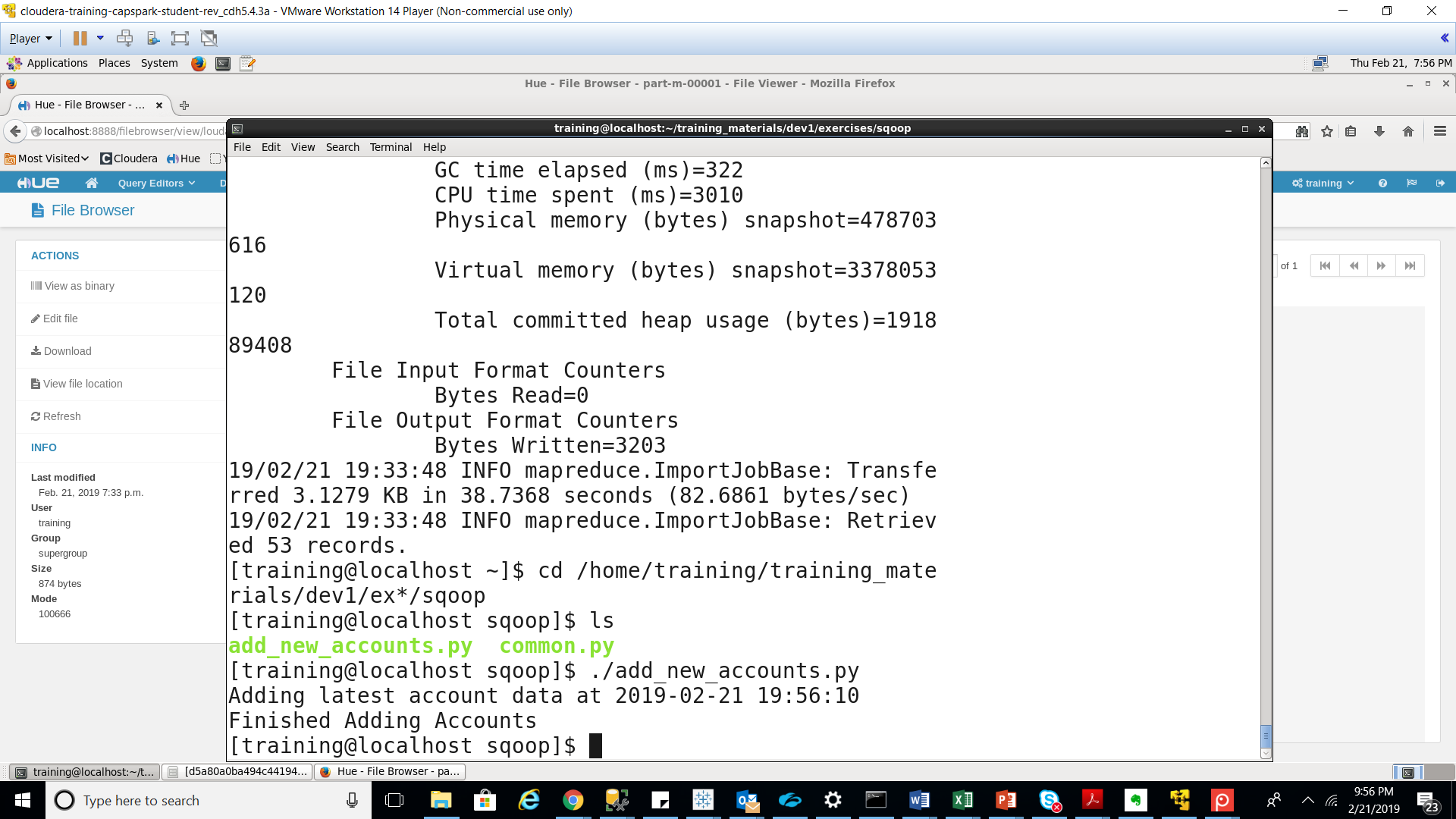
/loudacre\_Sushant\_Dhanu/webpage

We are confirming that the webpage that was imported into HDFS as a tab delimited file



We are runnning a python script that will add records to the accounts table in the loudacre database.





Sqoop import \

--connect jdbc:mysql://localhost/loudacre \

--username training –password training \

--table accounts \

--increamental append \

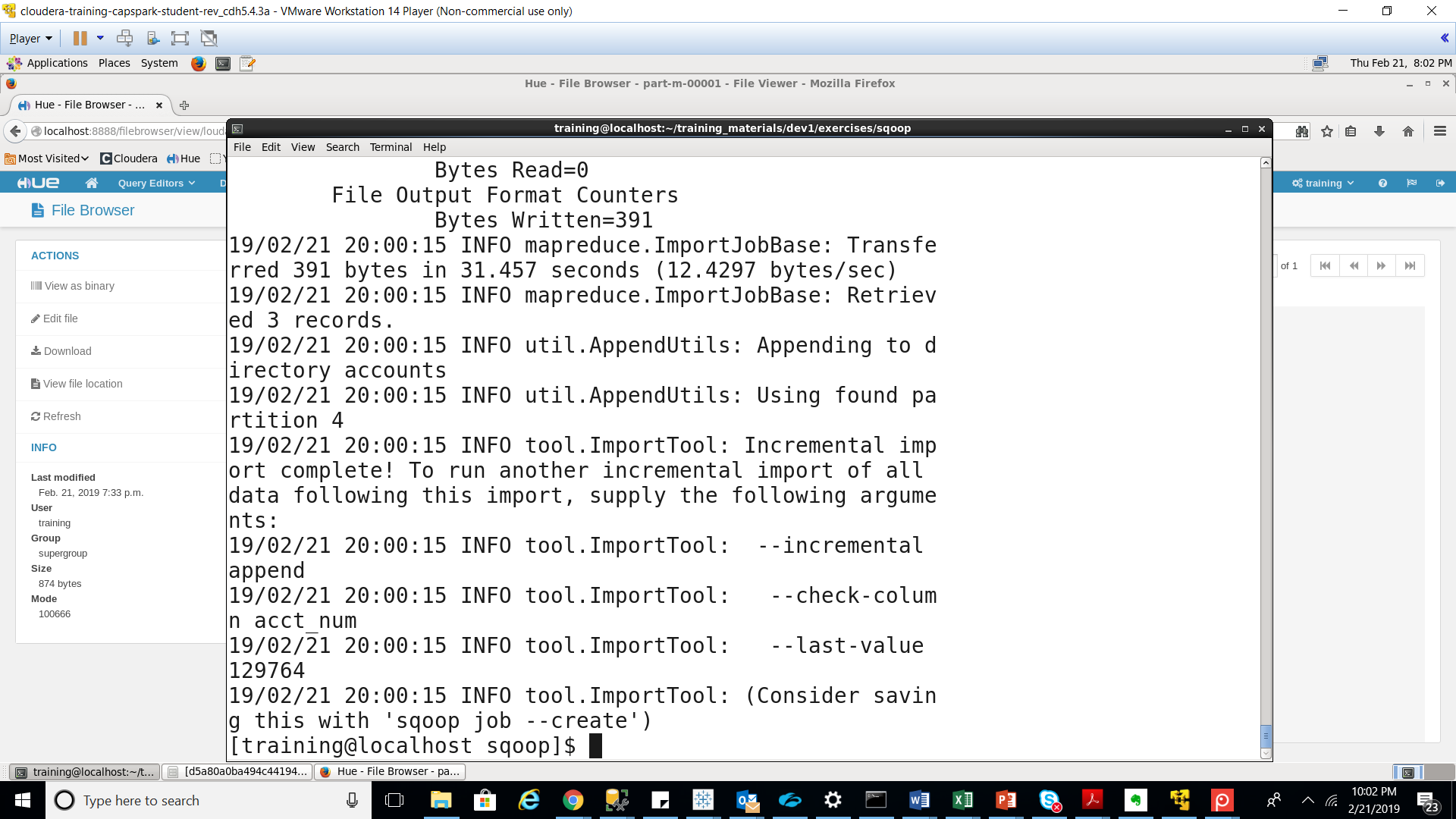
--target-dir /loudacre\_Sushant\_Dhanu/accounts \

--null-non-string ‘\\N’ \

--check-column acct\_num \

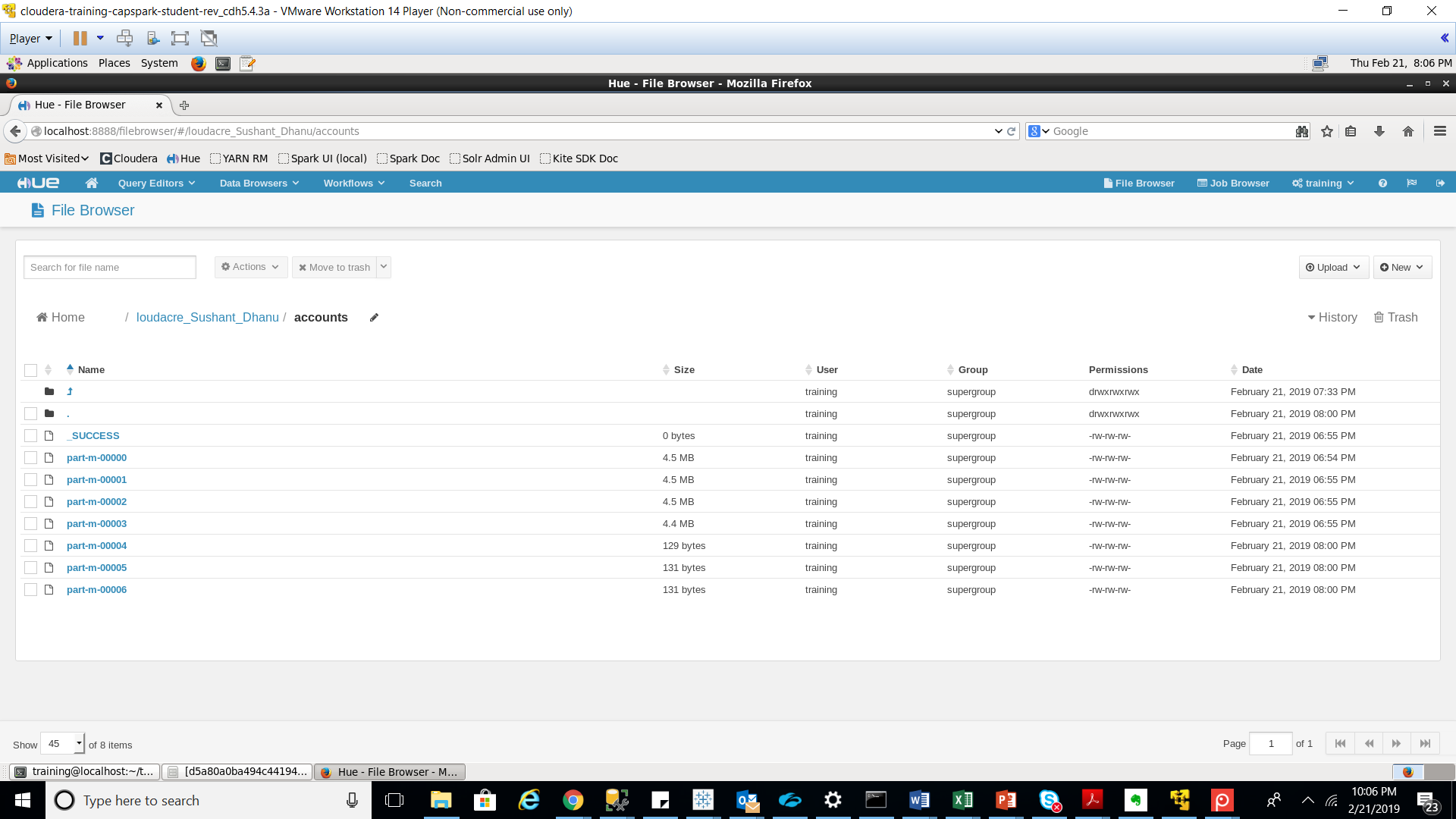
--last-value 129761

We are performing an incremental append based on the acct\_num column. We are importing those records whose acct\_num is greater than the largest\_acct\_num we noted



We are using HUE to confirm whether or not Sqoop performed our incremental import.

Yes, the incremental export was successful as we can see 3 new files have been added to the folder due to the incremental update



Execute the following command

Sqoop import \

--connect jdbc:mysql://localhost/loudacre \

--username training –password training \

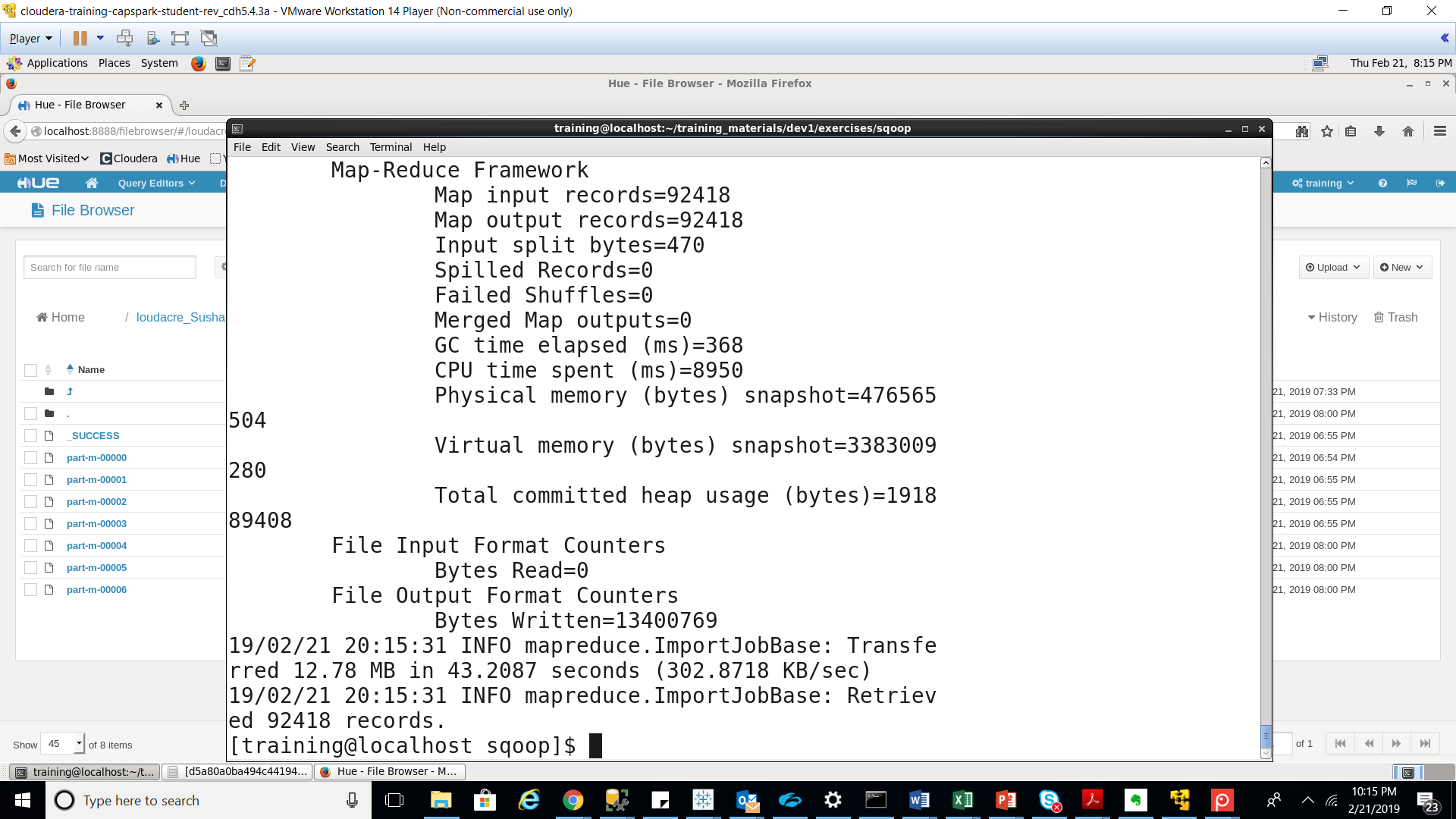
--table accounts \

--where “state=’CA’” \

--target-dir /loudacre\_Sushant/accounts \

--null-non-string ‘\\N’ \

We are using the ‘where” clause in the Sqoop import statement to import only those records in the accounts table that contain state = CA.



We are using HUE to confirm whether or not Sqoop performed our partial import.

