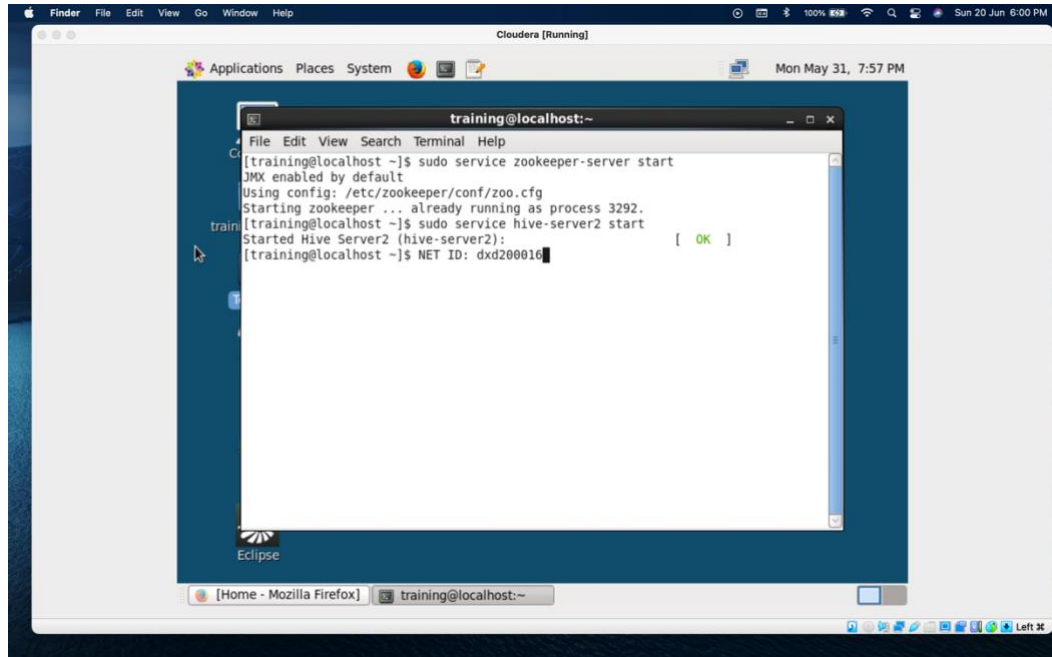# IMPALA

Divyansh Dahiya

LAB Chapter 06: IMPALA OR HIVE

Create and Populate Tables in Impala or Hive

In this lab I will define Impala/Hive tables to model and view data in HDFS.

1. First, I started the terminal and ran the commands to start the Hive server.
   Command: sudo service zookeeper-server start
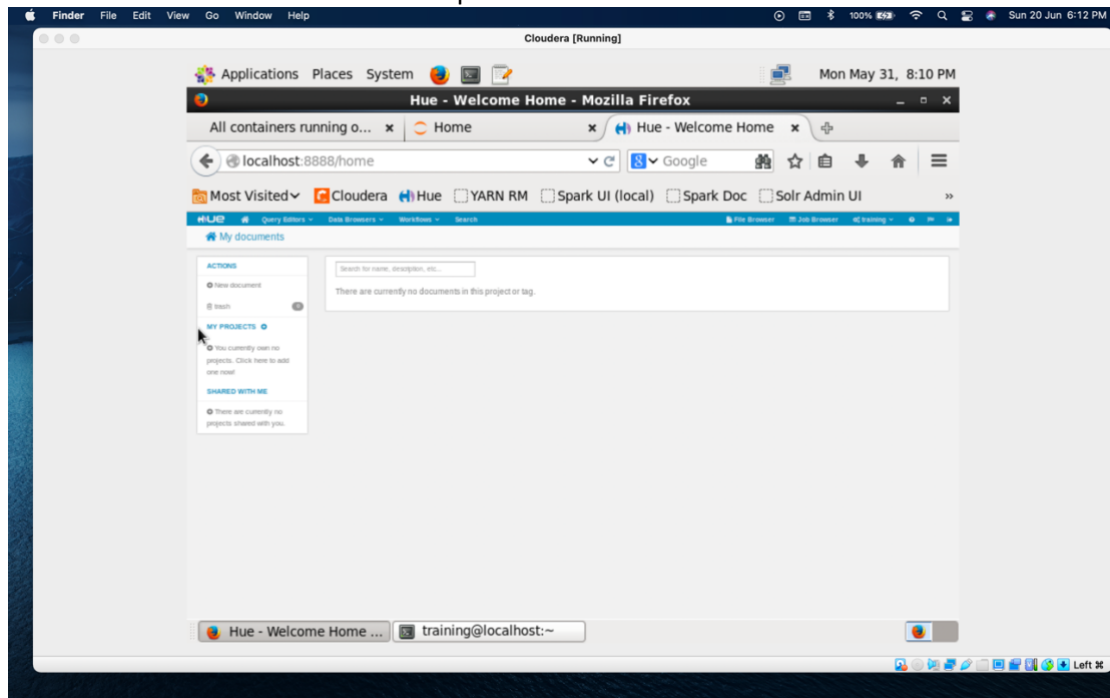   sudo service hive-server2 start

2. Then I started the web browser and opened the Hue Web Interface.

3. Then, I click on the Query Editors, to view all the query editors.
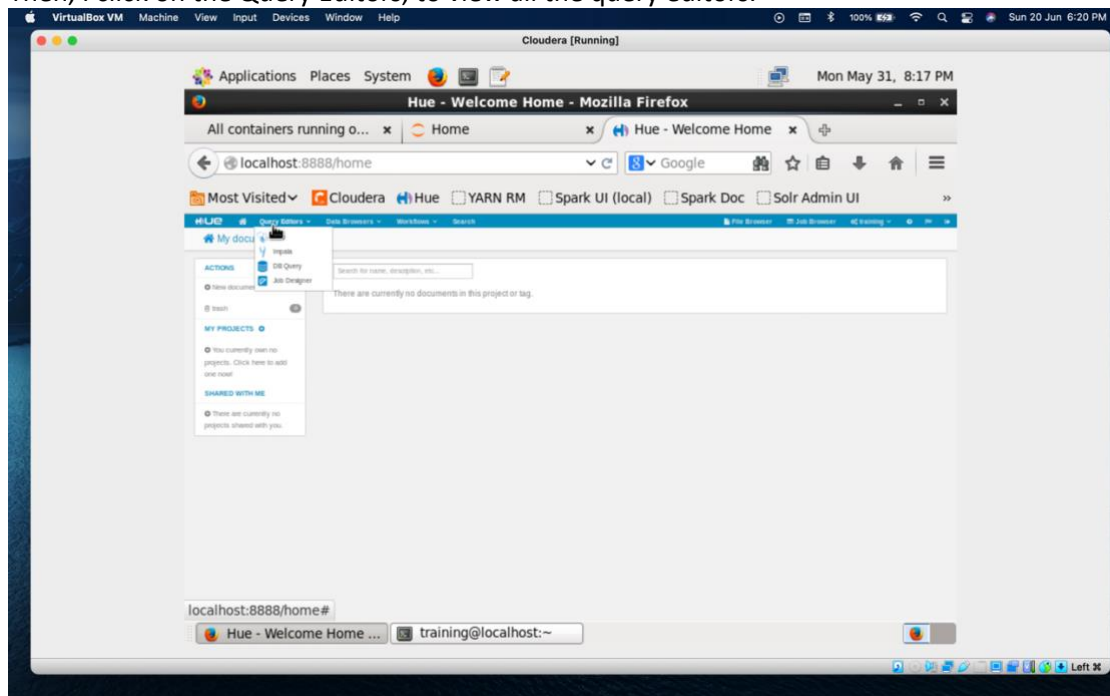
4. I selected the Impala Query editor and clicked on the New Query tab. Then wrote a SQL query to create a table and import the webpage data, that was imported in the previous lab.w
   Command: CREATE EXTERNAL TABLE webpage (page_id SMALLINT, name STRING, assoc_files STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LOCATION '/loudacre/webpage'
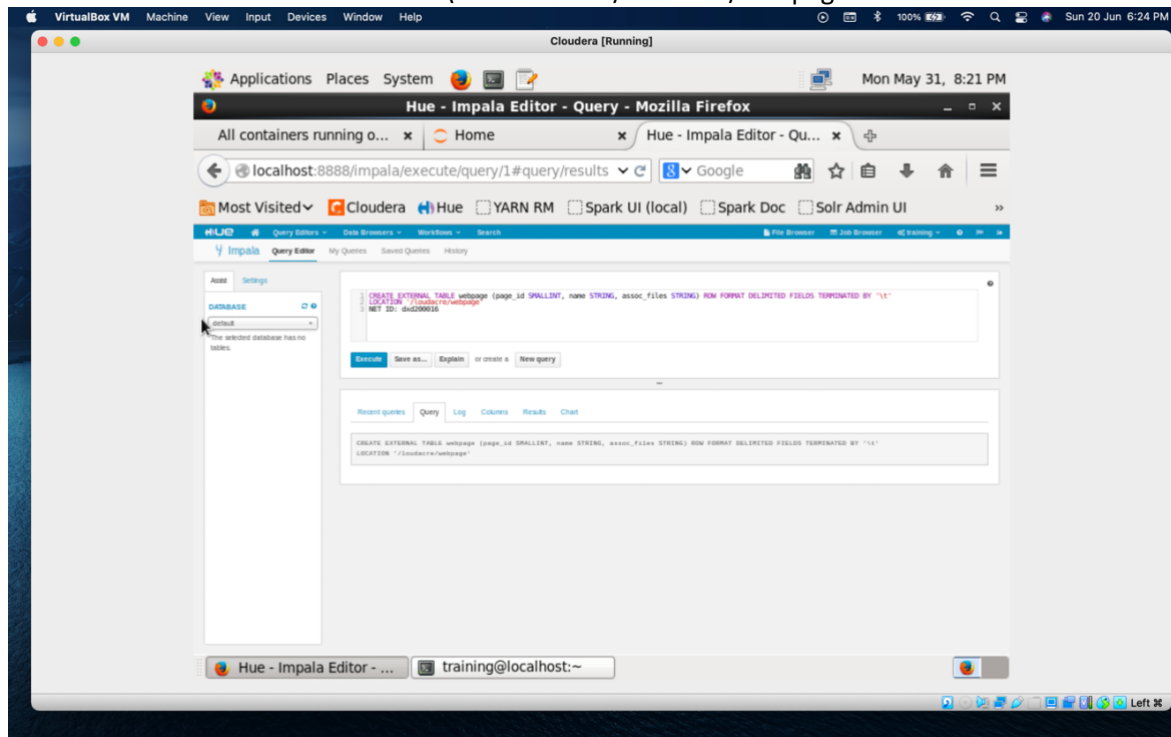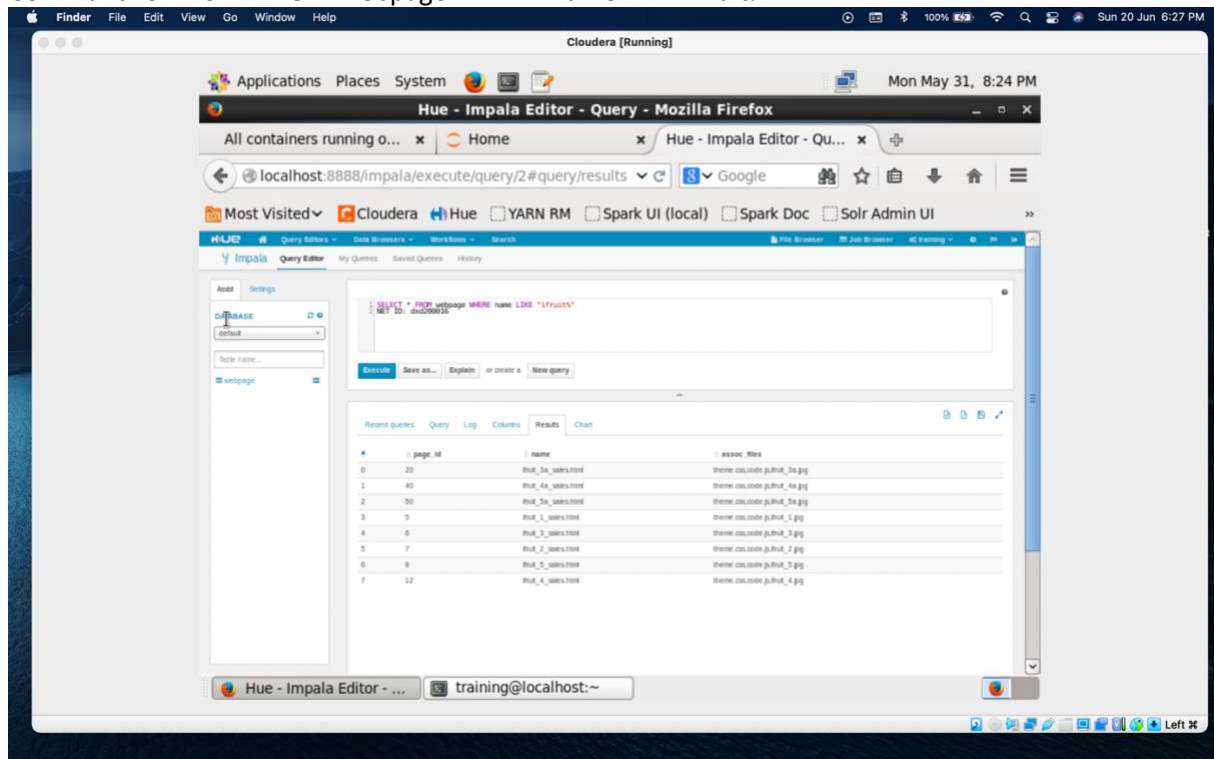

5. Now, on the left side menu, refreshing the Database tab, we can see that webpage table is created.
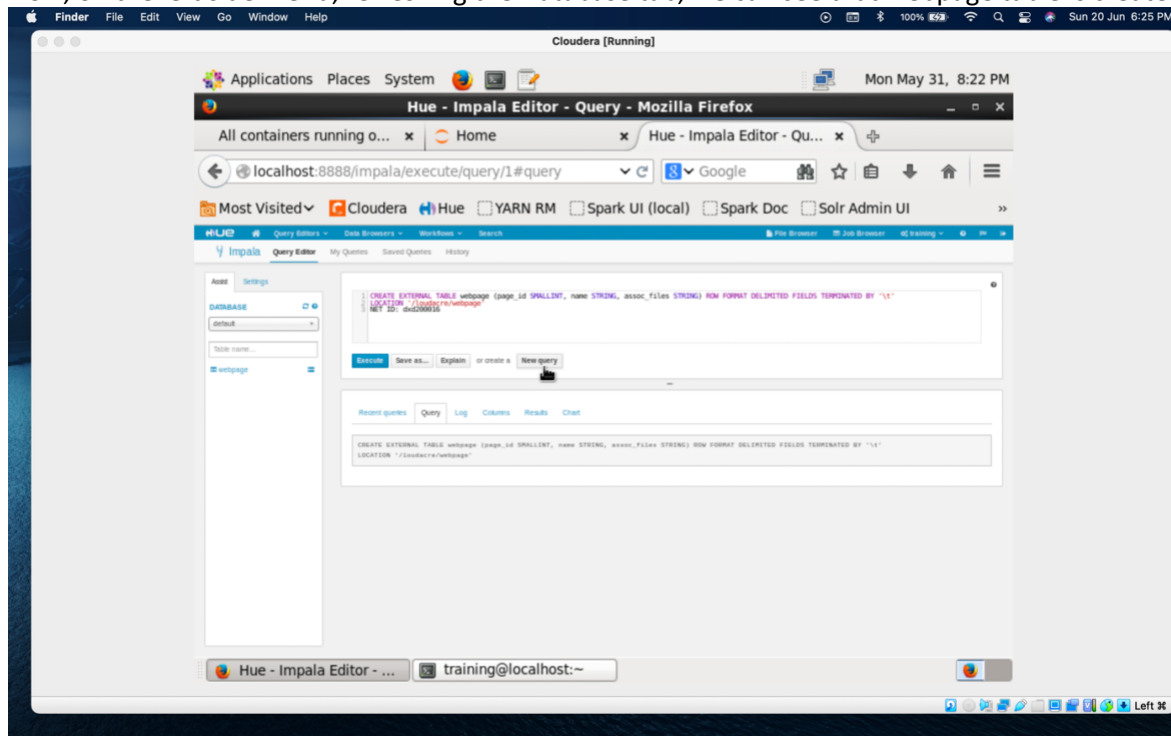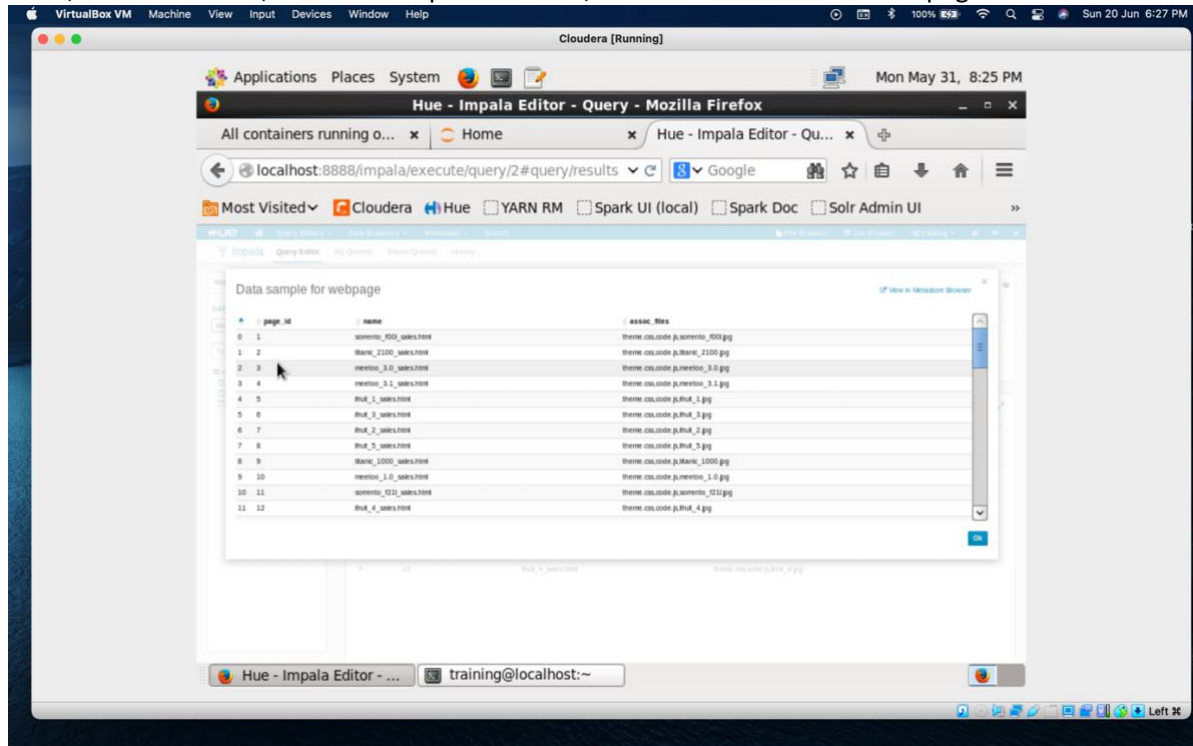
6. Now, I clicked on the webpage table to see the column definitions.
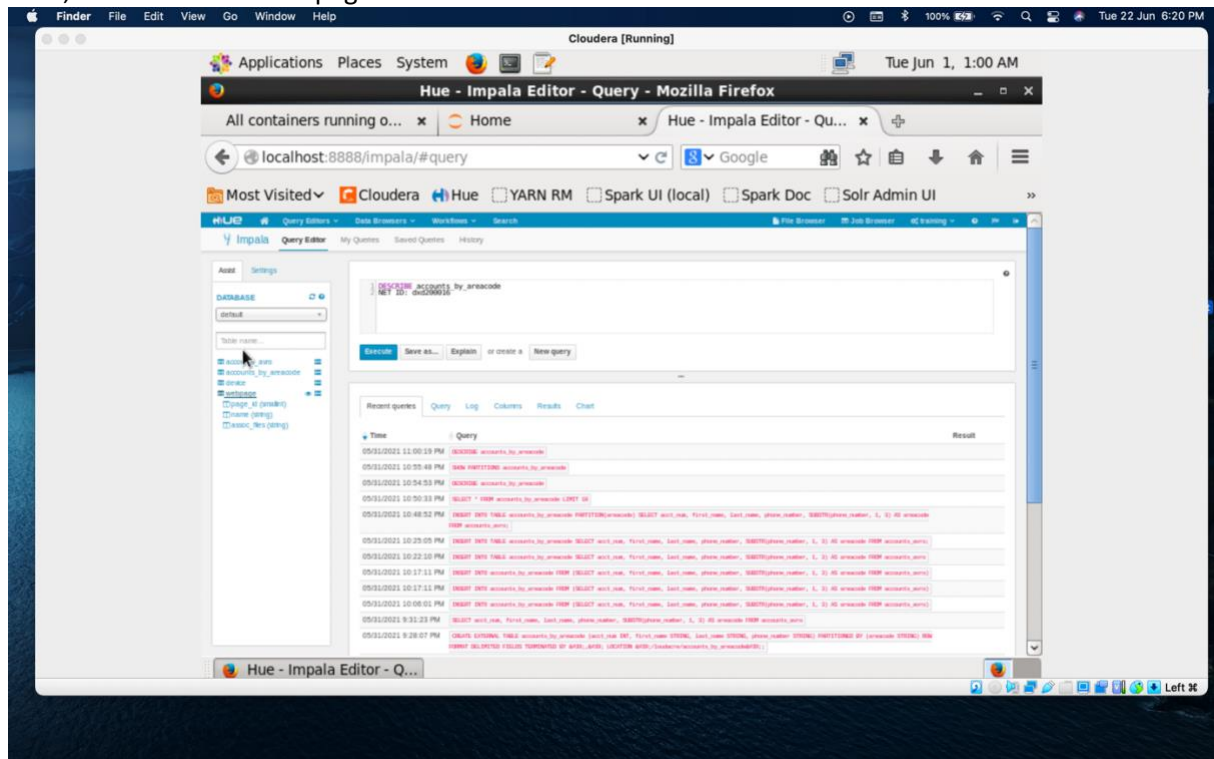
7. Then, I ran the query to view the data where name is ifruit.
   Command: SELECT * FROM webpage WHERE name LIKE "ifruit%"

8. Then, I clicked the on, Preview Sample Data icon, to view the data in the webpage table.

9. Now, I opened the terminal window and write the command to import the device table directly into the Hive Metastore.
Command: sqoop import \ --connect jdbc:mysql://localhost/loudacre \ --username training --password training \ --fields-terminated-by '\t' \ --table device \ --hive-import
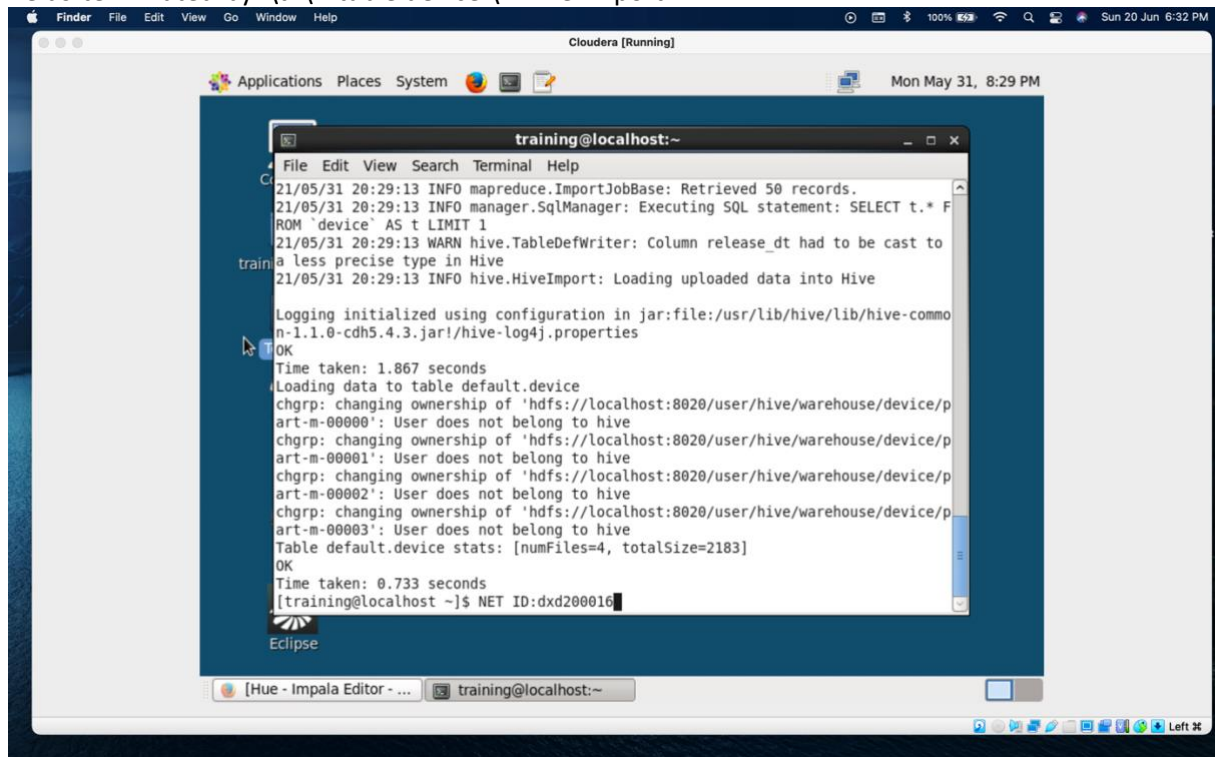
10. Then, using the Hue Web Interface, I opened the Hive Warehouse location to view imported copies of the data.
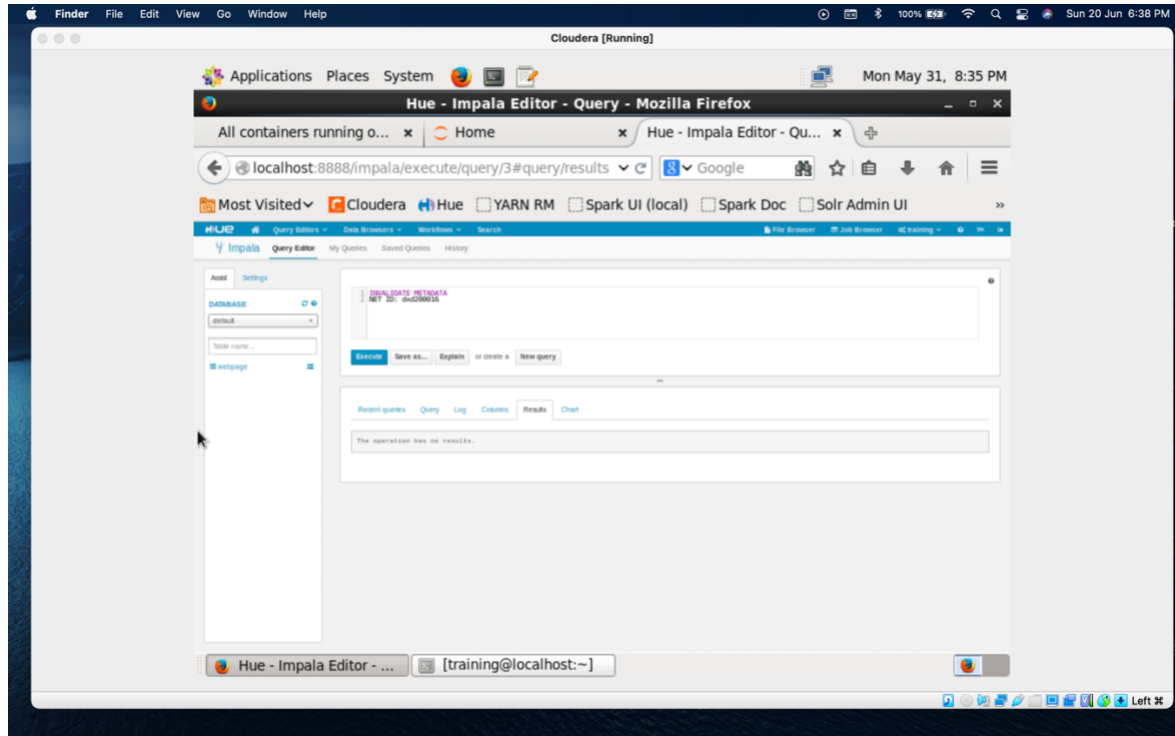   Location: /user/hive/warehouse/device
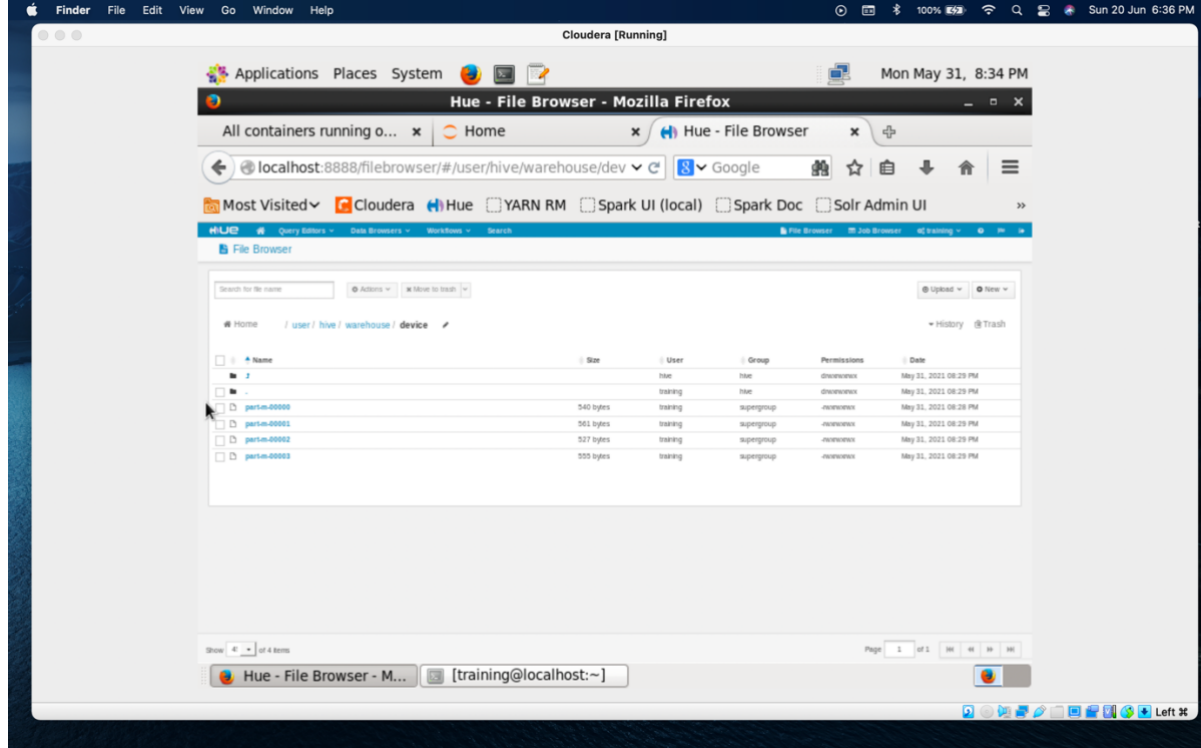
11. The contents of one of the file part-m-00000 from device table can be seen in the below screenshot.
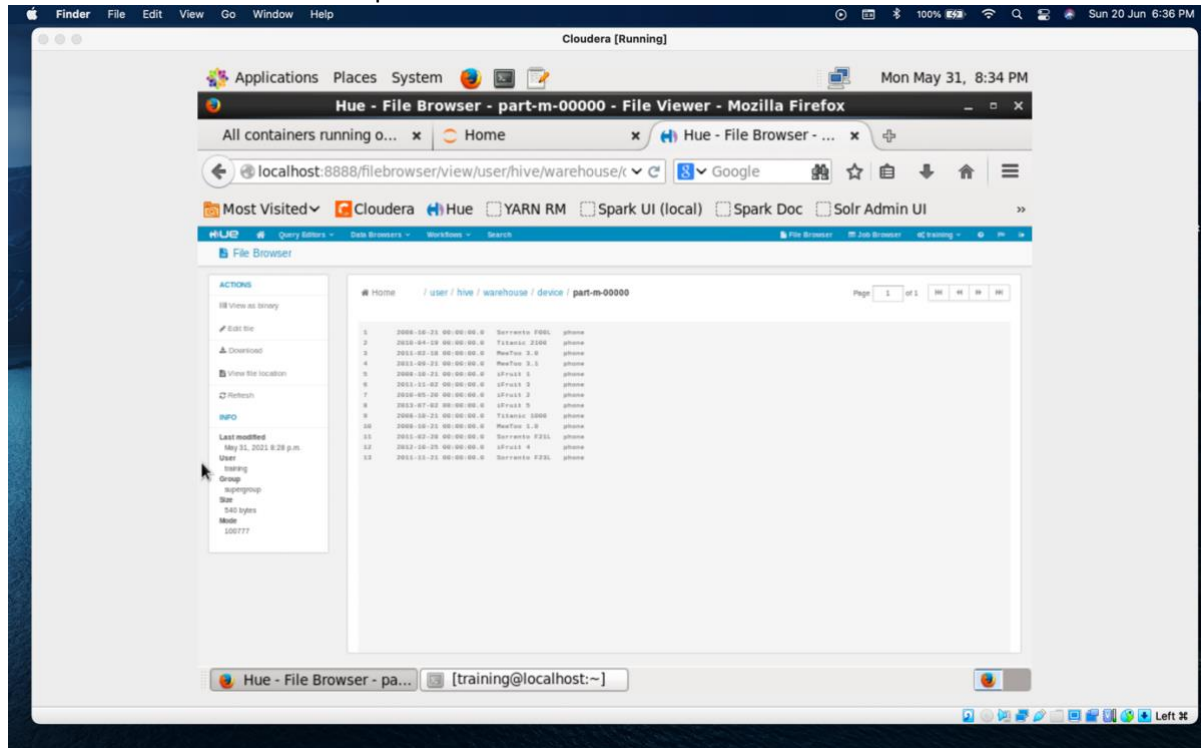
12. Then to refresh the Impala metadata cache I ran the following command.
    Command: INVALIDATE METADATA

13. At last, I rant the command to view the first 10 records from the device table.
    Command: SELECT * FROM device LIMIT 10;