# IST769 Homework Submission #7

## Basic Information

Your Name: Thulasi Ram Ruppa Krishnan   
Your SUID: 460746269  
Your Email: truppakr@syr.edu  
Date Due: 05-14-2020   
Homework #: HW7

## Instructions

For each answer, please include your answer as text, and any screenshot(s) which demonstrate your answer was executed. Most importantly, make sure to include evidence your answer is correct. This will most likely be a screenshot. If you had issues, problems, or had to make assumptions include them in your answer.

# Exercises

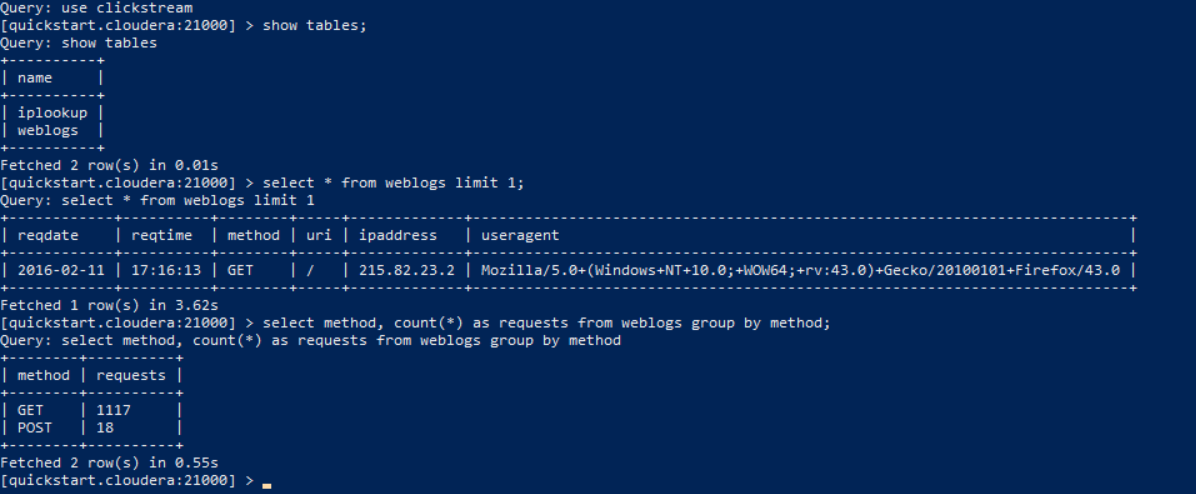
Complete each of the following exercises. If you are unsure how to accomplish the task, please consult the coursework videos where there are explanations and demos.

1. From Impala, use the two external tables **weblogs** created from **clickstream/logs\_noheader** and **iplookup** created from **clickstream/iplookup\_noheader** you created in the previous assignment to complete this question. Use the impala shell to answer the following questions, making sure to include the SELECT query you used to answer it.
   1. How many GET and POST requests are there in the weblogs?

**Answer:**

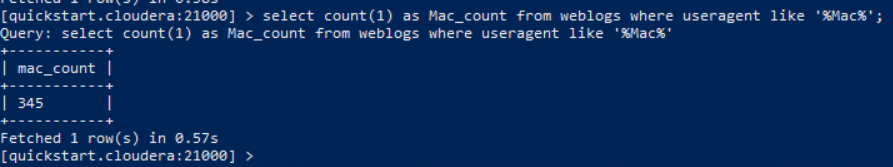
**1117 GET requests**

**18 POST requests**

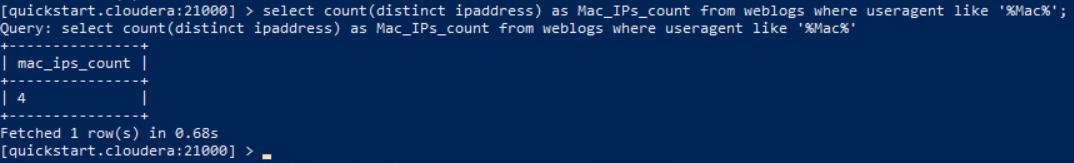


* 1. How many requests have Mac in the user agent?

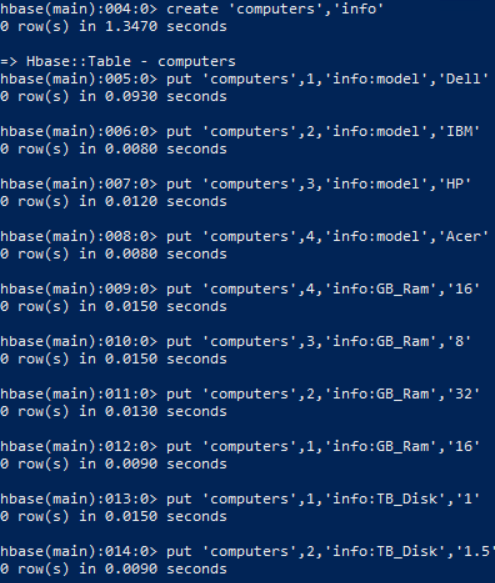
**Answer: 345**



* 1. How many hosts (ip addresses) have Mac in the user agent?  
     **Answer: 4**

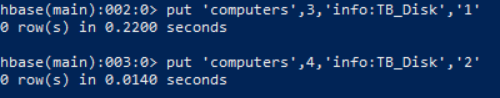
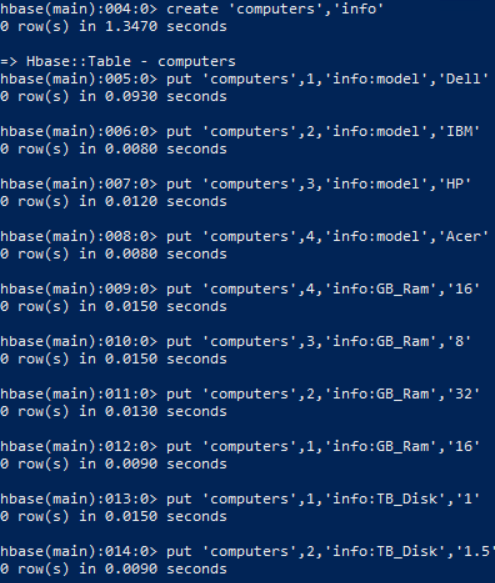


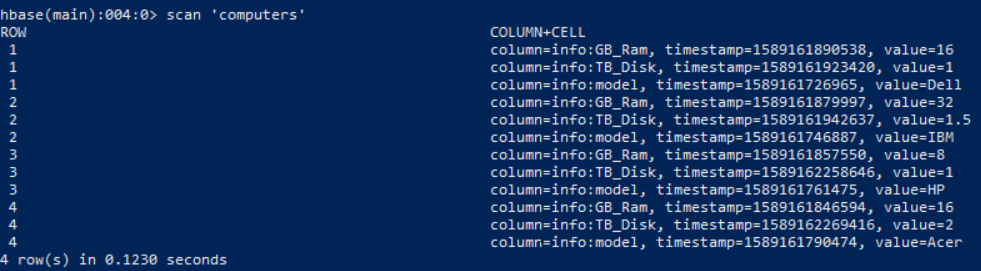
1. From the HBase shell, include the commands required to complete the following.
   1. Create a table named **computers** with column family **info**.



* 1. Issue HBase commands to write the following data to the table in the column family:

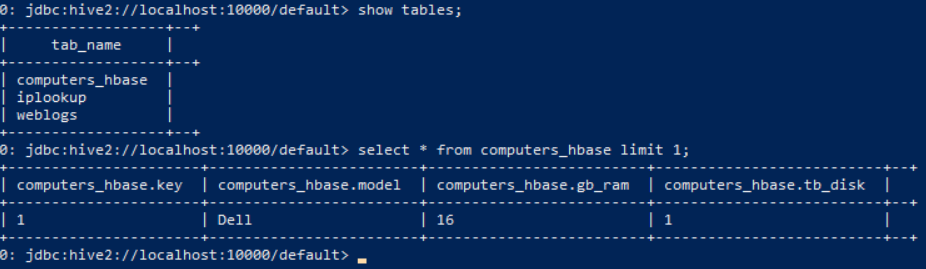
|  |  |  |  |
| --- | --- | --- | --- |
| Computer ID | Model | GB\_Ram | TB\_Disk |
| 1 | Dell | 16 | 1 |
| 2 | IBM | 32 | 1.5 |
| 3 | HP | 8 | 1 |
| 4 | Acer | 16 | 2 |

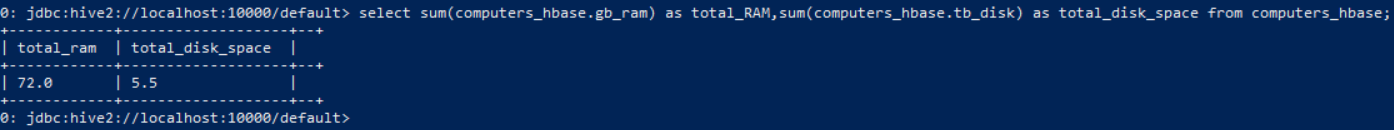




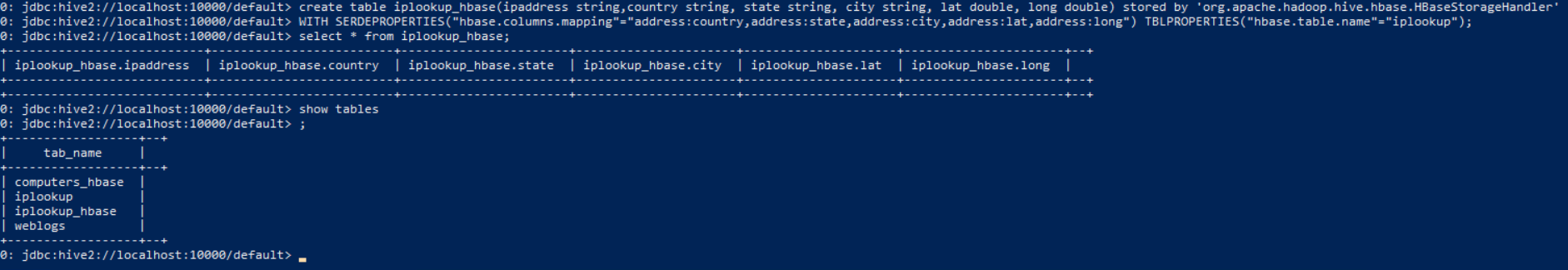
1. From the Hive shell, write an HQL statement to create an external Hive table from the HBase **computers** table. Then write a hive query to add up the total ram and disk across all computers. Your answer should include all HQL statements.

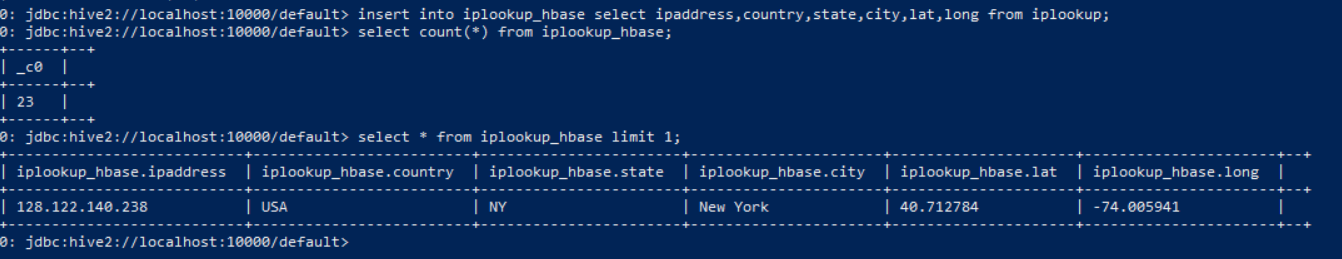






1. Use Hive to load the **iplookup** table you created from **clickstream/iplookup\_noheader** into and HBase table, with IP address as key. Include the HQL Queries you wrote to make the table and load the data as the answer to your question.





1. From the HBase shell, write an HBase query to retrieve the city and state columns for all rows in the **iplookup** table.



# Tear-Down / Troubleshooting Common Issues

**IMPORTANT NOTE:** Hadoop is a complex environment with many interworking parts. Since there are many services running in the container, it is suggested that you bring down the container completely whenever:

1. You are finished with the lab, or
2. Whenever your computer goes to sleep. When that happens, the running Hadoop container is going to lose connectivity among its services.

When you are **finished with the homework or stopping work for an extended period of time** you should bring down the environment like this:

1. From the terminal window where you typed docker-compose up -d type in the following:  
   docker-compose down

If your **Hadoop container seems to have broken services**. Tear down the container and bring it back up again:

1. From the terminal window where you typed docker-compose up -d type in the following:  
   docker-compose down and then docker-compose up -d