IST769 Impala, HBase and HCatalog

Your Name: Tajudeen Abdulazeez   
Your SUID: 69687-7373-0   
Your Email: toabdula@syr.edu   
Date Due:   
Homework #: 7

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?
   2. How many requests have Mac in the user agent?
   3. How many hosts (ip addresses) have Mac in the user agent?

1. From the HBase shell, include the commands required to complete the following.
   1. Create a table named **computers** with column family **info**.
   2. 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.