# IST769 Homework 7 Submission

## Basic Information

Your Name: Srihari Busam   
Your SUID: sbusam  
Your Email: sbusam@syr.edu  
Date Due: 11/18/2021  
Homework #: 7

## QUESTIONS:

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?
2. 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.
2. 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.
3. From the HBase shell, write an HBase query to retrieve the city and state columns for all rows in the **iplookup** table.

## ANSWERS:

### **ANSWER 1:**

|  |
| --- |
| CODE |
| --Query1  SELECT method, count(\*) as method\_count FROM weblogs GROUP BY method;  --Query2  SELECT COUNT(\*) as mac\_count FROM weblogs WHERE useragent like '%Mac%';  --Query 3  -- I am assuming that this question wants the valid ip addresses from both weblogs and iplookup tables  SELECT COUNT(DISTINCT(ip.ip)) AS unique\_ip\_mac  FROM weblogs wl JOIN iplookup ip  ON (wl.ipaddress = ip.ip)  WHERE wl.useragent LIKE '%Mac%'; |

**Screenshot #Query1**

A screenshot of a computer

Description automatically generated

**Screenshot #Query2**

A screenshot of a computer

Description automatically generated

**Screenshot #Query3**

A screenshot of a computer

Description automatically generated

### **ANSWER 2:**

|  |
| --- |
| CODE |
| create 'computers', 'info'  put 'computers', 1, 'info:computer\_id',1  put 'computers', 1, 'info:model','Dell'  put 'computers', 1, 'info:gb\_ram',16  put 'computers', 1, 'info:tb\_disk',1  put 'computers', 2, 'info:computer\_id',2  put 'computers', 2, 'info:model','IBM'  put 'computers', 2, 'info:gb\_ram',32  put 'computers', 2, 'info:tb\_disk',1.5  put 'computers', 3, 'info:computer\_id',3  put 'computers', 3, 'info:model','HP'  put 'computers', 3, 'info:gb\_ram',8  put 'computers', 3, 'info:tb\_disk',1  put 'computers', 4, 'info:computer\_id',4  put 'computers', 4, 'info:model','Acer'  put 'computers', 4, 'info:gb\_ram',16  put 'computers', 4, 'info:tb\_disk',2 |

**Screen shot to show computers table from hbase**Text

Description automatically generated

### **ANSWER 3:**

|  |
| --- |
| CODE |
| CREATE EXTERNAL TABLE computers\_hbase (key int, computer\_id int, model string, gb\_ram int, tb\_disk float)  stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'  WITH SERDEPROPERTIES("hbase.columns.mapping" = "info:computer\_id,info:model,info:gb\_ram,info:tb\_disk")  TBLPROPERTIES("hbase.table.name"="computers");  select sum(gb\_ram) as total\_ram from computers\_hbase; |

**Screenshot showing table through hive**Text

Description automatically generated

**Screenshot Query though hive**

Text

Description automatically generated

### **ANSWER 4:**

|  |
| --- |
| CODE |
| create table iplookup\_hbase (ip string, country string, state string, city string, approx\_lat double, approx\_lng double)  STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'  WITH SERDEPROPERTIES("hbase.columns.mapping" = ":key, ipl:country, ipl:state, ipl:city, ipl:approx\_lat, ipl:approx\_lng")  TBLPROPERTIES("hbase.table.name"="iplookup");  INSERT INTO iplookup\_hbase select ip, country, state, city, approx\_lat, approx\_lng from iplookup; |

|  |
| --- |
| CODE on HBASE to validate. Query to get cities from the imported data |
| scan 'iplookup', { COLUMNS => ['ipl:city']} |

**Screenshot of the city query data from HBASE**

Text

Description automatically generated

### **ANSWER 5:**

|  |
| --- |
| CODE: Query to get city and state from HBASE |
| scan 'iplookup', { COLUMNS => ['ipl:city', 'ipl:state']} |

**Screenshot to get city and state for the imported data from HBASE**

Text

Description automatically generated