# IST769 Homework Submission Template

## Basic Information

Your Name: Wanyue XiaoYour

SUID: 720633297

Your Email: xwanyue@syr.edu

Date Due: March 26th, 2020

Homework #: HW07

## 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.

## Your Answers:

1.

1. How many GET and POST requests are there in the weblogs?

**Select count(method) from weblogs where method = ‘GET’;**

**-- 1117**

**Select count(method) from weblogs where method = ‘POST’;**

**-- 18**

A screenshot of a cell phone

Description automatically generated

1. How many requests have Mac in the user agent? -- 345

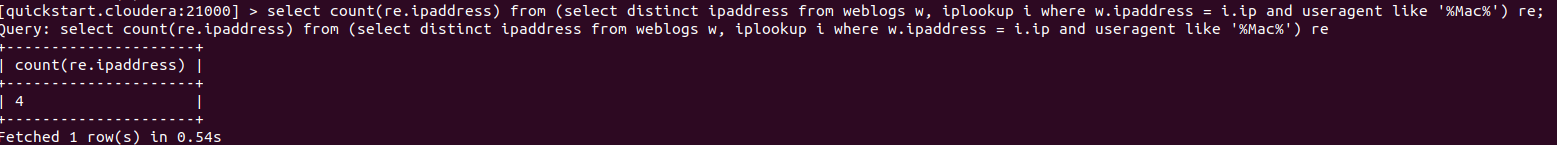
**Select count(useragent) from weblogs where useragent like ‘%Mac%’;**

A picture containing food, drawing

Description automatically generated

1. How many hosts (ip addresses) have Mac in the user agent? – 4 distinct ip address

**Select count(re.ipaddress) from (select distinct ipaddress from weblogs w, iplookup I where w.ipaddress = i.ip and useragent like ‘%Mac%’) re; -- 4computer**



2.

* 1. Create a table named **computers** with column family **info**.

**Creat ‘computers’, ’info’**

A picture containing meter

Description automatically generated

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

**Put ‘computers’, 1, ‘info:ComputerID’,’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:ComputerID’,’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:ComputerID’,’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:ComputerID’,’4’**

**Put ‘computers’, 4, ‘info:Model’,’Acer’**

**Put ‘computers’, 4, ‘info:GB\_Ram’,’16’**

**Put ‘computers’, 4, ‘info:TB\_Disk’,’2’**

A screenshot of a cell phone

Description automatically generated

Finally, check the result by using scan command:

**Scan ‘computers’**

A picture containing holding

Description automatically generated

3.

Create the table:

# Since the original hbase has two columns, which are key and computerID, show the same information. Here I will put the key to the computerID variable that is inside the computers table:

**create external table computers (ComputerID int, Model string, GB\_Ram string, TB\_Disk string) stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'**

**WITH SERDEPROPERTIES("hbase.columns.mapping" = "key, info:Model, info:GB\_Ram, info:TB\_Disk") TBLPROPERTIES("hbase.table.name"="computers");**

**Show tables;**

**Select \* from computers;**

A screenshot of a cell phone

Description automatically generated

**Select sum(gb\_ram), sum(tb\_disk) from computers;**

**# Here the \_c0 should be the sum result of gb\_ram while \_c1 should be the sum result of tb\_disk:**

A picture containing food

Description automatically generated

4.

# create a table inside hive:

**create table iplookup\_hbase (ip string, country string, state string, city string, approxlat decimal(10,6), approxlng decimal(10,6)) STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'**

**WITH SERDEPROPERTIES ("hbase.columns.mapping"=":key, info:country, info:state, info:city, info: approxlat, info: approxlng ") TBLPROPERTIES("hbase.table.name"="iplookup");**

**insert into iplookup\_hbase select ip, country, state, city, approxlat, approxlng from iplookup;**

**Select \* from iplookup\_hbase;**

A circuit board

Description automatically generated

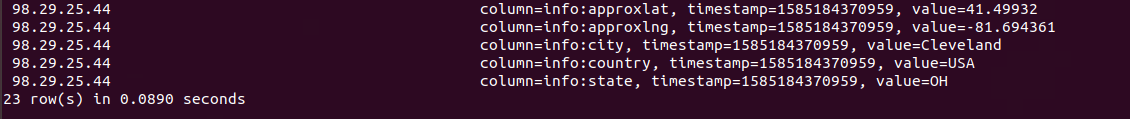
Back to hbase:

**list**

**Scan ‘iplookup’**

A close up of text on a black background

Description automatically generated

Result：  


5.

**Scan “iplookup”, {COLUMNS => [‘info:state’, ‘info:city’]}**

A close up of a logo

Description automatically generated

## Student Reflection:

To achieve the highest grade on the assignment you must be as descriptive and personal as possible with your reflection. Ask yourself the following questions.

## Student Reflection:

1. What new information have I learned from completing this assignment?

Hadoop, hive, impala, hbase, and their relationships

1. What were my challenges / roadblocks I encountered on the way to completing it?

Operation of Linux System

1. To be better prepared to attempt this assignment I should \_\_\_\_\_\_\_ ?

Review

1. If I could make this assignment better, I would \_\_\_\_\_?

Google

1. Rate your comfort level with completing this assignment:

1.

1 ==> I can do this on my own and explain how to do it.

2 ==> I can do this on my own without any help.

3 ==> I can do this with help or guidance from others. If you choose this level, please list those who helped you.

4 ==> I don't understand this at all yet and need extra help. If you choose this, please try to articulate that which you do not understand.