IST769 Homework Submission

Name: **Sathish Kumar Rajendiran**
SUID: **666555028**
Email: srajendi@syr.edu
Due Date: **08/24/2021**
Task: **Impala, HBase and HCatalog**
Homework #:**7**

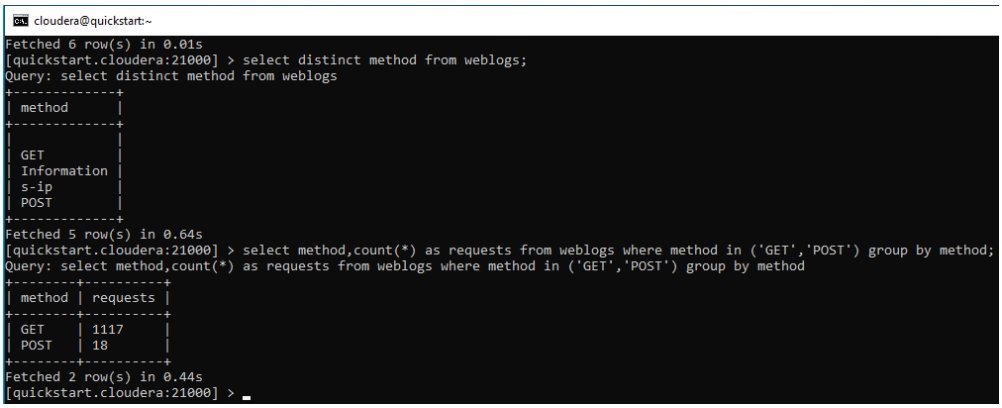
Exercise(s):

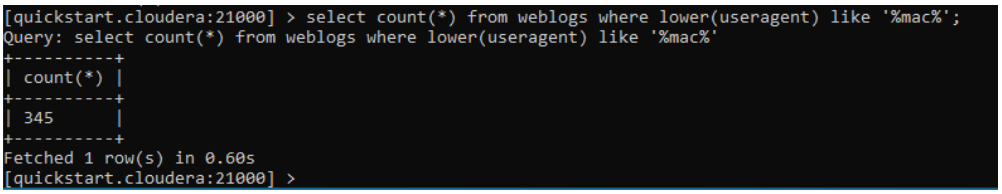
1. From Impala, use the two external tables weblogs created from clickstream/logs_noheader and iplookup created from clickstream/iplookup_noheader you created in the previous assignment to complete this question. Use the impala shell to answer the following questions, making sure to include the SELECT query you used to answer it.
    a. How many GET and POST requests are there in the weblogs?
    b. How many requests have Mac in the user agent?
    c. How many hosts (ip addresses) have Mac in the user agent?

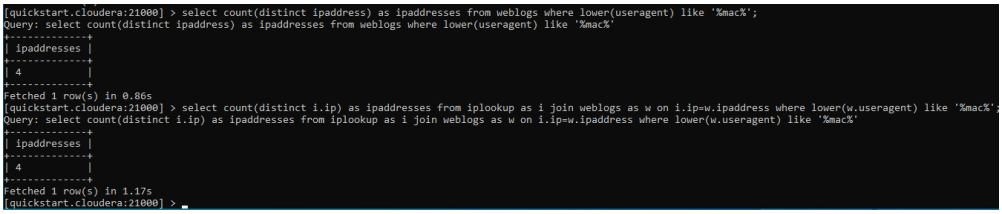
Solution:

```
-- Query to find number of GET and POST requests
select method, count(*) as requests
from weblogs
where method in ('GET','POST')
group by method;

--Query to find number of requests have MAC in User agent
select count(*)
from weblogs
where lower(useragent) like '%mac%';

--Query to find number of hosts have MAC in User agent
select count(distinct ipaddress) as ipaddresses
from weblogs
where lower(useragent) like '%mac%';

select count(distinct i.ip) as ipaddresses
from iplookup i join weblogs w on i.ip = w.ipaddress
where
where lower(w.useragent) like '%mac%';
```

Evidence:







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

Solution:

```
-- create a Hbase table as computers with column family info
t = create 'computers', 'info'

 -- verify the table existence
scan 'computers'
```

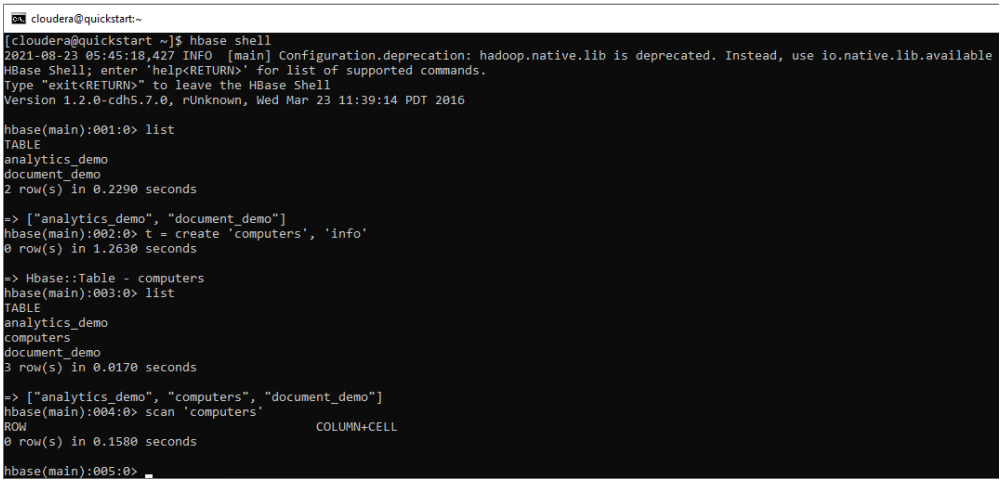
-- Add records to computers table

```
put ‘computers’, ‘1’, ‘info:Model’, ‘Dell’
put ‘computers’, ‘1’, ‘info:GB_RAM’, ‘16’
put ‘computers’, ‘1’, ‘info:TB_Disk, ‘1’

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:Model’, ‘HP
put ‘computers’, ‘3’, ‘info:GB_RAM’, ‘8’
put ‘computers’, ‘3’, ‘info:TB_Disk, ‘1’

put ‘computers’, ‘4’, ‘info:Model’, ‘Acer
put ‘computers’, ‘4’, ‘info:GB_RAM’, ‘16’
put ‘computers’, ‘4’, ‘info:TB_Disk, ‘2’
```

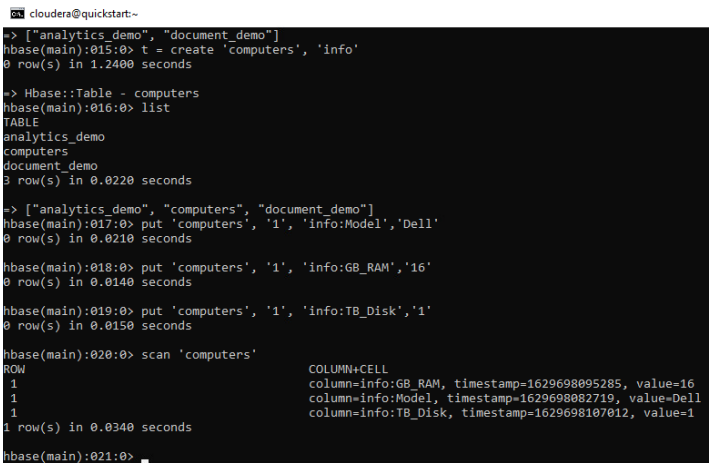
-- verify the data existence

```
scan ‘computers’
```

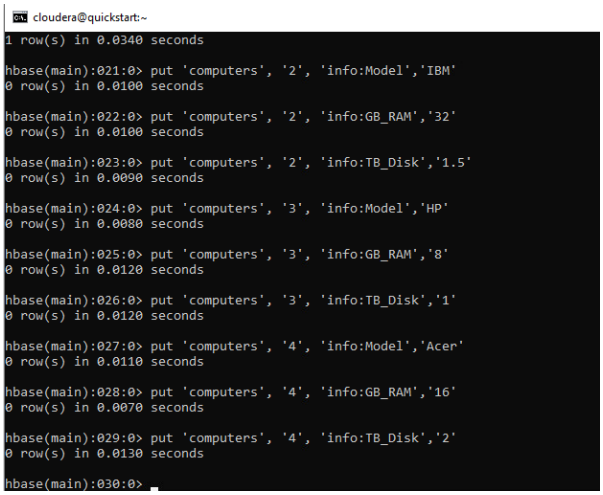
Evidence:

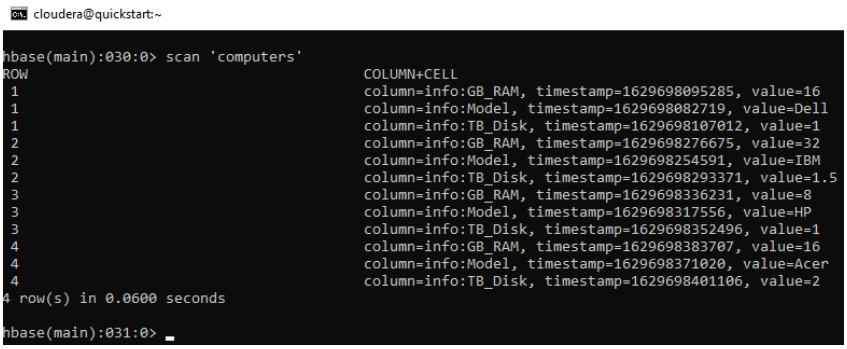


```
cloudera@quickstart:~
[cloudera@quickstart ~]$ hbase shell
2021-08-23 05:45:18,427 INFO  [main] Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.7.0, rUnknown, Wed Mar 23 11:39:14 PDT 2016

hbase(main):001:0> list
TABLE
analytics_demo
document_demo
2 row(s) in 0.2290 seconds

=> ["analytics_demo", "document_demo"]
hbase(main):002:0> t = create 'computers', 'info'
0 row(s) in 1.2630 seconds

=> Hbase::Table - computers
hbase(main):003:0> list
TABLE
analytics_demo
computers
document_demo
3 row(s) in 0.0170 seconds

=> ["analytics_demo", "computers", "document_demo"]
hbase(main):004:0> scan 'computers'
ROW                                   COLUMN+CELL
0 row(s) in 0.1580 seconds

hbase(main):005:0>
```



```
cloudera@quickstart:~
=> ["analytics_demo", "document_demo"]
hbase(main):015:0> t = create 'computers', 'info'
0 row(s) in 1.2400 seconds

=> Hbase::Table - computers
hbase(main):016:0> list
TABLE
analytics_demo
computers
document_demo
3 row(s) in 0.0220 seconds

=> ["analytics_demo", "computers", "document_demo"]
hbase(main):017:0> put 'computers', '1', 'info:Model','Dell'
0 row(s) in 0.0210 seconds

hbase(main):018:0> put 'computers', '1', 'info:GB_RAM','16'
0 row(s) in 0.0140 seconds

hbase(main):019:0> put 'computers', '1', 'info:TB_Disk','1'
0 row(s) in 0.0150 seconds

hbase(main):020:0> scan 'computers'
ROW                                   COLUMN+CELL
 1                                    column=info:GB_RAM, timestamp=1629698095285, value=16
 1                                    column=info:Model, timestamp=1629698082719, value=Dell
 1                                    column=info:TB_Disk, timestamp=1629698107012, value=1
1 row(s) in 0.0340 seconds

hbase(main):021:0>
```



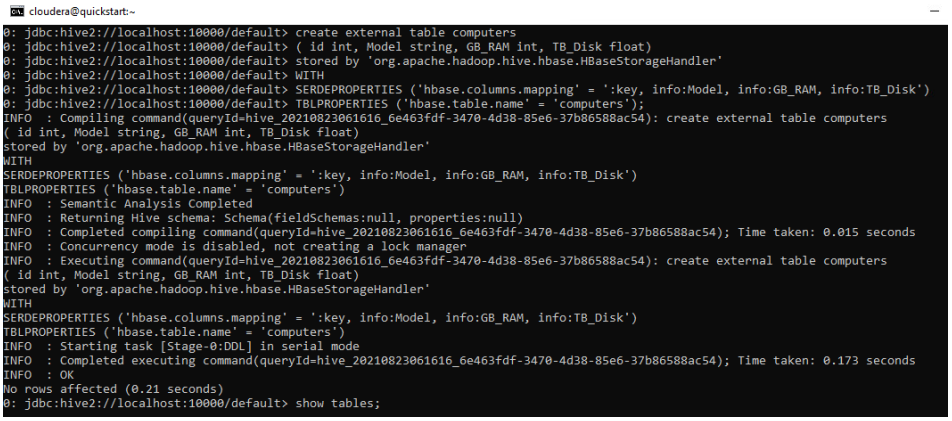


3. From the Hive shell, write an HQL statement to create an external Hive table from the HBase **computers** table. Then write a hive query to add up the total ram and disk across all computers. Your answer should include all HQL statements.

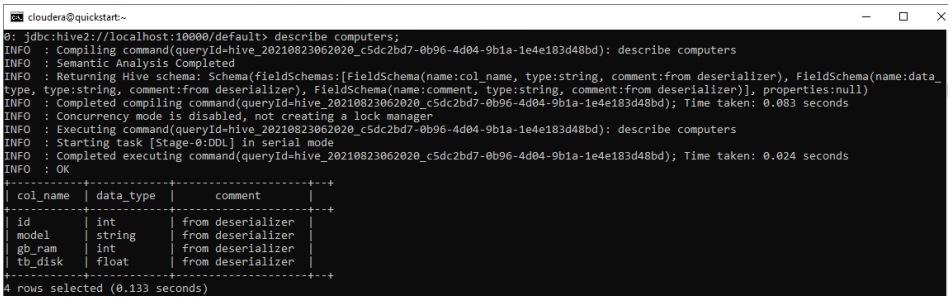
Solution:

```
-- create an external table in Hive from Hbase (computers)
create external table computers
 (id int, Model string, GB_RAM int, TB_Disk float)
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');

-- verify the data existence
describe computers;
select * from computers;

-- HQL statement to add total ram and disk across computers
select
  sum(GB_RAM) as Total_Memory
, sum(TB_Disk) as Total_Storage
from computers;
```

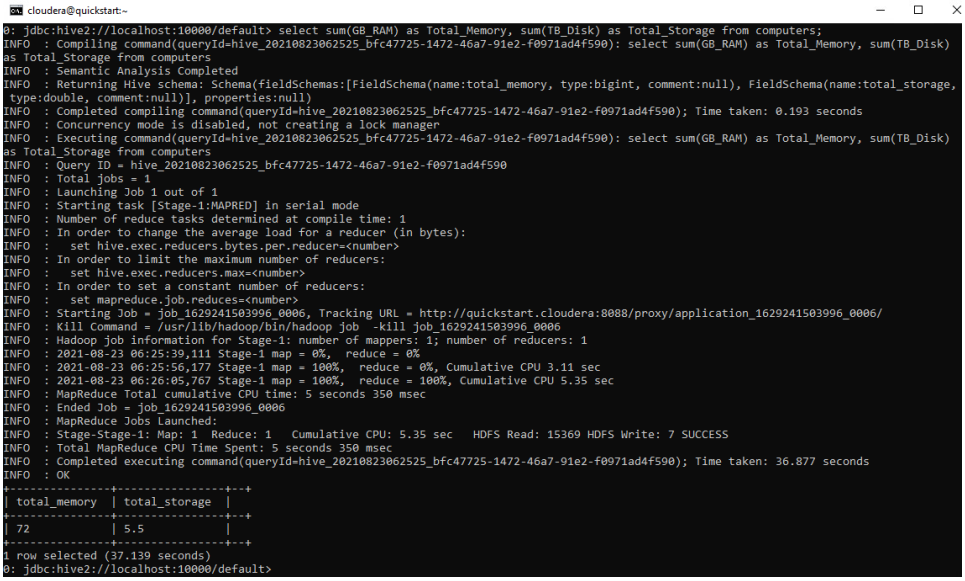
Evidence:



```
cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> create external table computers
0: jdbc:hive2://localhost:10000/default> ( id int, Model string, GB_RAM int, TB_Disk float)
0: jdbc:hive2://localhost:10000/default> stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
0: jdbc:hive2://localhost:10000/default> WITH
0: jdbc:hive2://localhost:10000/default> SERDEPROPERTIES ('hbase.columns.mapping' = ':key, info:Model, info:GB_RAM, info:TB_Disk')
0: jdbc:hive2://localhost:10000/default> TBLPROPERTIES ('hbase.table.name' = 'computers');
INFO  : Compiling command(queryId=hive_20210823061616_6e463fdf-3470-4d38-85e6-37b86588ac54): create external table computers
( id int, Model string, GB_RAM int, TB_Disk float)
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')
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20210823061616_6e463fdf-3470-4d38-85e6-37b86588ac54); Time taken: 0.015 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210823061616_6e463fdf-3470-4d38-85e6-37b86588ac54): create external table computers
( id int, Model string, GB_RAM int, TB_Disk float)
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')
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210823061616_6e463fdf-3470-4d38-85e6-37b86588ac54); Time taken: 0.173 seconds
INFO  : OK
No rows affected (0.21 seconds)
0: jdbc:hive2://localhost:10000/default> show tables;
```



```
cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> describe computers;
INFO  : Compiling command(queryId=hive_20210823062020_c5dc2bd7-0b96-4d04-9b1a-1e4e183d48bd): describe computers
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from deserializer), FieldSchema(name:data_
type, type:string, comment:from deserializer), FieldSchema(name:comment, type:string, comment:from deserializer)], properties:null)
INFO  : Completed compiling command(queryId=hive_20210823062020_c5dc2bd7-0b96-4d04-9b1a-1e4e183d48bd); Time taken: 0.083 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210823062020_c5dc2bd7-0b96-4d04-9b1a-1e4e183d48bd): describe computers
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210823062020_c5dc2bd7-0b96-4d04-9b1a-1e4e183d48bd); Time taken: 0.024 seconds
INFO  : OK
+-----------+------------+--------------------+--+
| col_name  | data_type  |      comment       |
+-----------+------------+--------------------+--+
| id        | int        | from deserializer  |
| model     | string     | from deserializer  |
| gb_ram    | int        | from deserializer  |
| tb_disk   | float      | from deserializer  |
+-----------+------------+--------------------+--+
4 rows selected (0.133 seconds)
```



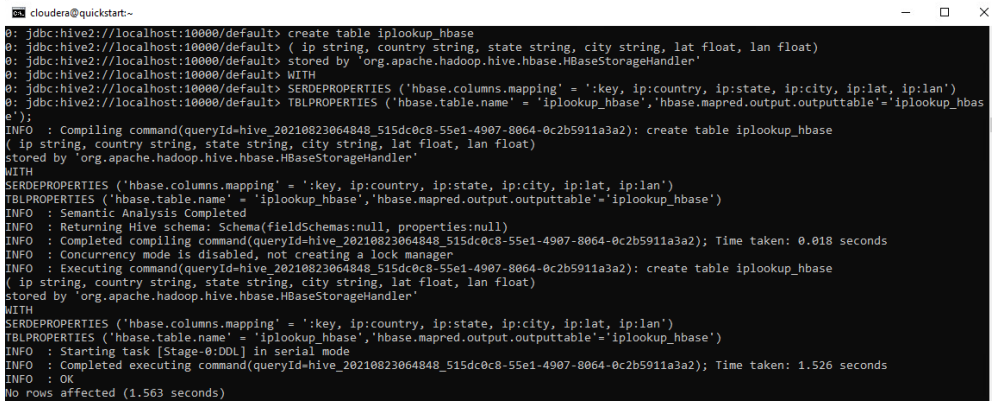
```
cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> select sum(GB_RAM) as Total_Memory, sum(TB_Disk) as Total_Storage from computers;
INFO  : Compiling command(queryId=hive_20210823062525_bfc47725-1472-46a7-91e2-f0971ad4f590): select sum(GB_RAM) as Total_Memory, sum(TB_Disk)
as Total_Storage from computers
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:total_memory, type:bigint, comment:null), FieldSchema(name:total_storage,
 type:double, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20210823062525_bfc47725-1472-46a7-91e2-f0971ad4f590); Time taken: 0.193 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210823062525_bfc47725-1472-46a7-91e2-f0971ad4f590): select sum(GB_RAM) as Total_Memory, sum(TB_Disk)
as Total_Storage from computers
INFO  : Query ID = hive_20210823062525_bfc47725-1472-46a7-91e2-f0971ad4f590
INFO  : Total jobs = 1
INFO  : Launching Job 1 out of 1
INFO  : Starting task [Stage-1:MAPRED] in serial mode
INFO  : Number of reduce tasks determined at compile time: 1
INFO  : In order to change the average load for a reducer (in bytes):
INFO  :   set hive.exec.reducers.bytes.per.reducer=<number>
INFO  : In order to limit the maximum number of reducers:
INFO  :   set hive.exec.reducers.max=<number>
INFO  : In order to set a constant number of reducers:
INFO  :   set mapreduce.job.reduces=<number>
INFO  : Starting Job = job_1629241503996_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1629241503996_0006/
INFO  : Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1629241503996_0006
INFO  : Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
INFO  : 2021-08-23 06:25:39,111 Stage-1 map = 0%,  reduce = 0%
INFO  : 2021-08-23 06:25:56,177 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.11 sec
INFO  : 2021-08-23 06:26:05,767 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.35 sec
INFO  : MapReduce Total cumulative CPU time: 5 seconds 350 msec
INFO  : Ended Job = job_1629241503996_0006
INFO  : MapReduce Jobs Launched:
INFO  : Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.35 sec   HDFS Read: 15369 HDFS Write: 7 SUCCESS
INFO  : Total MapReduce CPU Time Spent: 5 seconds 350 msec
INFO  : Completed executing command(queryId=hive_20210823062525_bfc47725-1472-46a7-91e2-f0971ad4f590); Time taken: 36.877 seconds
INFO  : OK
+--------------+---------------+--+
| total_memory  | total_storage  |
+--------------+---------------+--+
| 72           | 5.5           |
+--------------+---------------+--+
1 row selected (37.139 seconds)
0: jdbc:hive2://localhost:10000/default>
```

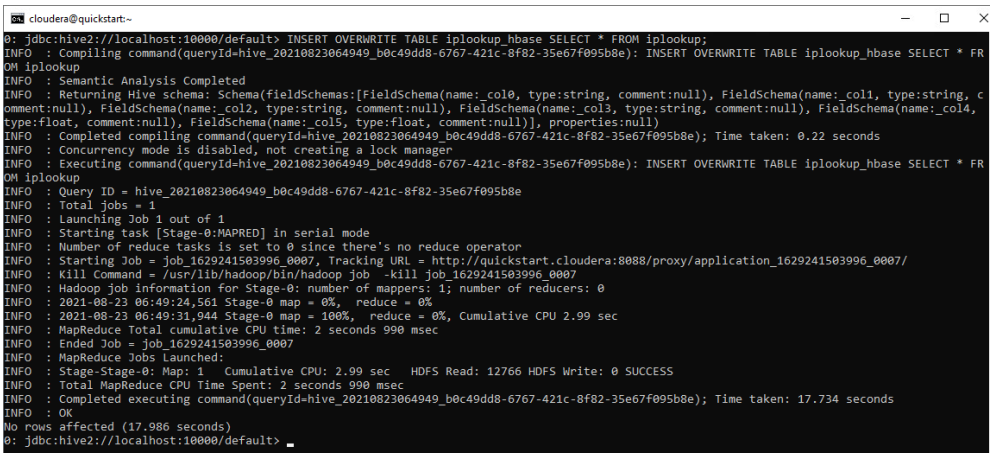
4. Use Hive to load the **iplookup** table you created from **clickstream/iplookup_noheader** into and HBase table, with IP address as key. Include the HQL Queries you wrote to make the table and load the data as the answer to your question.

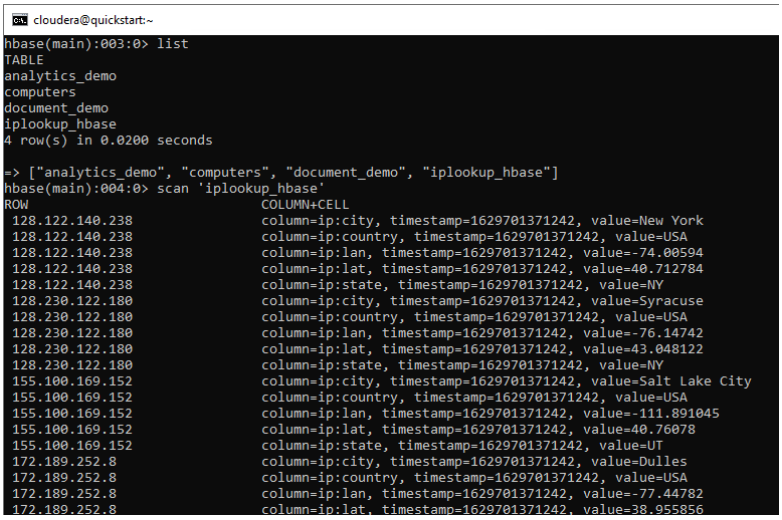
Solution:

```
-- create a table in Hbase from Hive (iplookup)
create table iplookup_hbase
 (ip string, country string, state string, city string, lat float, lan float)
stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
WITH
SERDEPROPERTIES ('hbase.columns.mapping'=':key, ip:country, ip:state, ip:city, ip:lat, ip:lan')
TBLPROPERTIES ('hbase.table.name'='iplookup_hbase', 'hbase.mapred.output.outputtable' =
'iplookup_hbase');

-- HQL to load the data
INSERT OVERWRITE TABLE iplookup_hbase SELECT * FROM iplookup;

-- Hbase shell to query the data
scan 'iplookup_hbase'
```

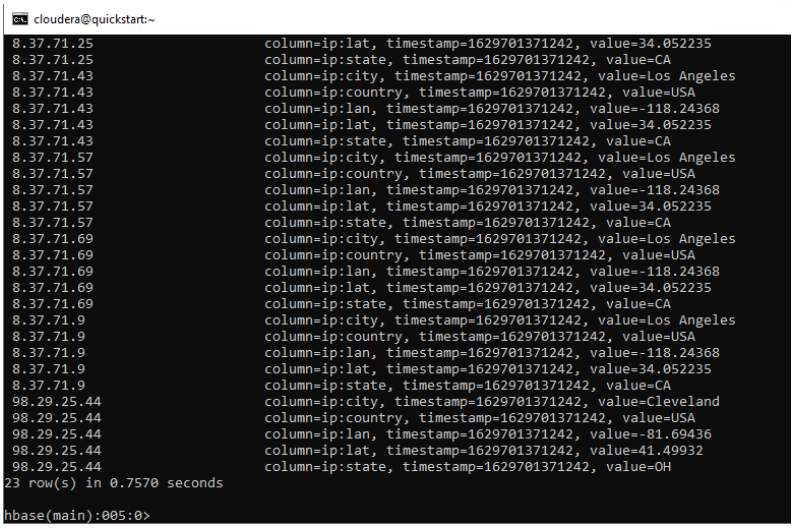
Evidence:







```
cloudera@quickstart:~
hbase(main):003:0> list
TABLE
analytics_demo
computers
document_demo
iplookup_hbase
4 row(s) in 0.0200 seconds

=> ["analytics_demo", "computers", "document_demo", "iplookup_hbase"]
hbase(main):004:0> scan 'iplookup_hbase'
ROW                                   COLUMN+CELL
 128.122.140.238                      column=ip:city, timestamp=1629701371242, value=New York
 128.122.140.238                      column=ip:country, timestamp=1629701371242, value=USA
 128.122.140.238                      column=ip:lan, timestamp=1629701371242, value=-74.00594
 128.122.140.238                      column=ip:lat, timestamp=1629701371242, value=40.712784
 128.122.140.238                      column=ip:state, timestamp=1629701371242, value=NY
 128.230.122.180                      column=ip:city, timestamp=1629701371242, value=Syracuse
 128.230.122.180                      column=ip:country, timestamp=1629701371242, value=USA
 128.230.122.180                      column=ip:lan, timestamp=1629701371242, value=-76.14742
 128.230.122.180                      column=ip:lat, timestamp=1629701371242, value=43.048122
 128.230.122.180                      column=ip:state, timestamp=1629701371242, value=NY
 155.100.169.152                      column=ip:city, timestamp=1629701371242, value=Salt Lake City
 155.100.169.152                      column=ip:country, timestamp=1629701371242, value=USA
 155.100.169.152                      column=ip:lan, timestamp=1629701371242, value=-111.891045
 155.100.169.152                      column=ip:lat, timestamp=1629701371242, value=40.76078
 155.100.169.152                      column=ip:state, timestamp=1629701371242, value=UT
 172.189.252.8                        column=ip:city, timestamp=1629701371242, value=Dulles
 172.189.252.8                        column=ip:country, timestamp=1629701371242, value=USA
 172.189.252.8                        column=ip:lan, timestamp=1629701371242, value=-77.44782
 172.189.252.8                        column=ip:lat, timestamp=1629701371242, value=38.955856
```



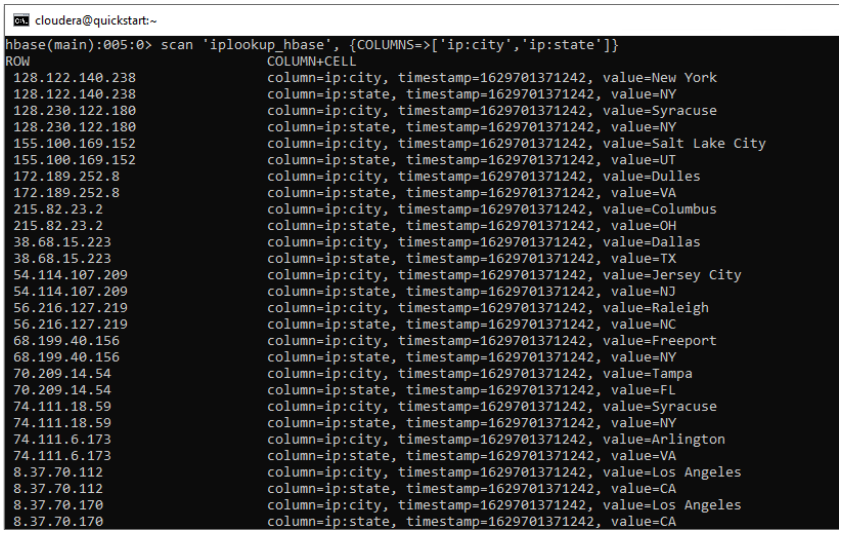
```
cloudera@quickstart:~
 8.37.71.25                           column=ip:lat, timestamp=1629701371242, value=34.052235
 8.37.71.25                           column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.43                           column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.43                           column=ip:country, timestamp=1629701371242, value=USA
 8.37.71.43                           column=ip:lan, timestamp=1629701371242, value=-118.24368
 8.37.71.43                           column=ip:lat, timestamp=1629701371242, value=34.052235
 8.37.71.43                           column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.57                           column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.57                           column=ip:country, timestamp=1629701371242, value=USA
 8.37.71.57                           column=ip:lan, timestamp=1629701371242, value=-118.24368
 8.37.71.57                           column=ip:lat, timestamp=1629701371242, value=34.052235
 8.37.71.57                           column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.69                           column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.69                           column=ip:country, timestamp=1629701371242, value=USA
 8.37.71.69                           column=ip:lan, timestamp=1629701371242, value=-118.24368
 8.37.71.69                           column=ip:lat, timestamp=1629701371242, value=34.052235
 8.37.71.69                           column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.9                            column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.9                            column=ip:country, timestamp=1629701371242, value=USA
 8.37.71.9                            column=ip:lan, timestamp=1629701371242, value=-118.24368
 8.37.71.9                            column=ip:lat, timestamp=1629701371242, value=34.052235
 8.37.71.9                            column=ip:state, timestamp=1629701371242, value=CA
 98.29.25.44                          column=ip:city, timestamp=1629701371242, value=Cleveland
 98.29.25.44                          column=ip:country, timestamp=1629701371242, value=USA
 98.29.25.44                          column=ip:lan, timestamp=1629701371242, value=-81.69436
 98.29.25.44                          column=ip:lat, timestamp=1629701371242, value=41.49932
 98.29.25.44                          column=ip:state, timestamp=1629701371242, value=OH
23 row(s) in 0.7570 seconds

hbase(main):005:0>
```

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

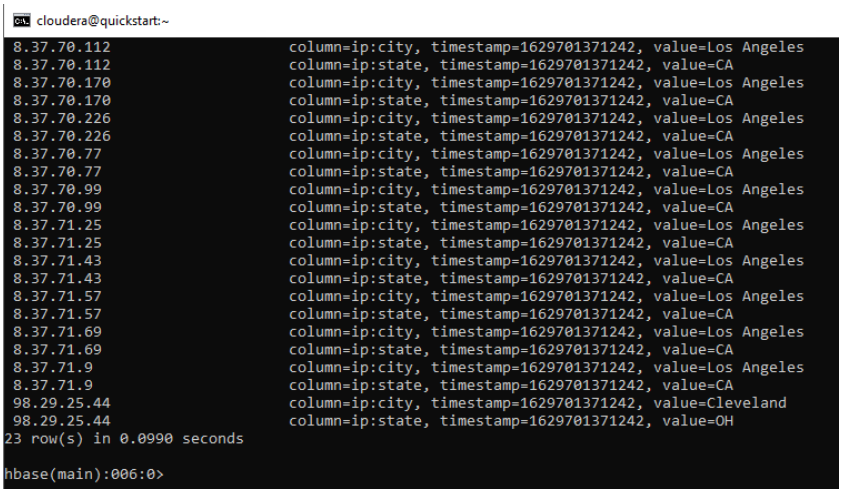
Solution:

-- Hbase shell to retrieve city and state info from iplookup table
scan 'iplookup_hbase', {COLUMNS => ['ip:city', 'ip:state']}

Evidence:

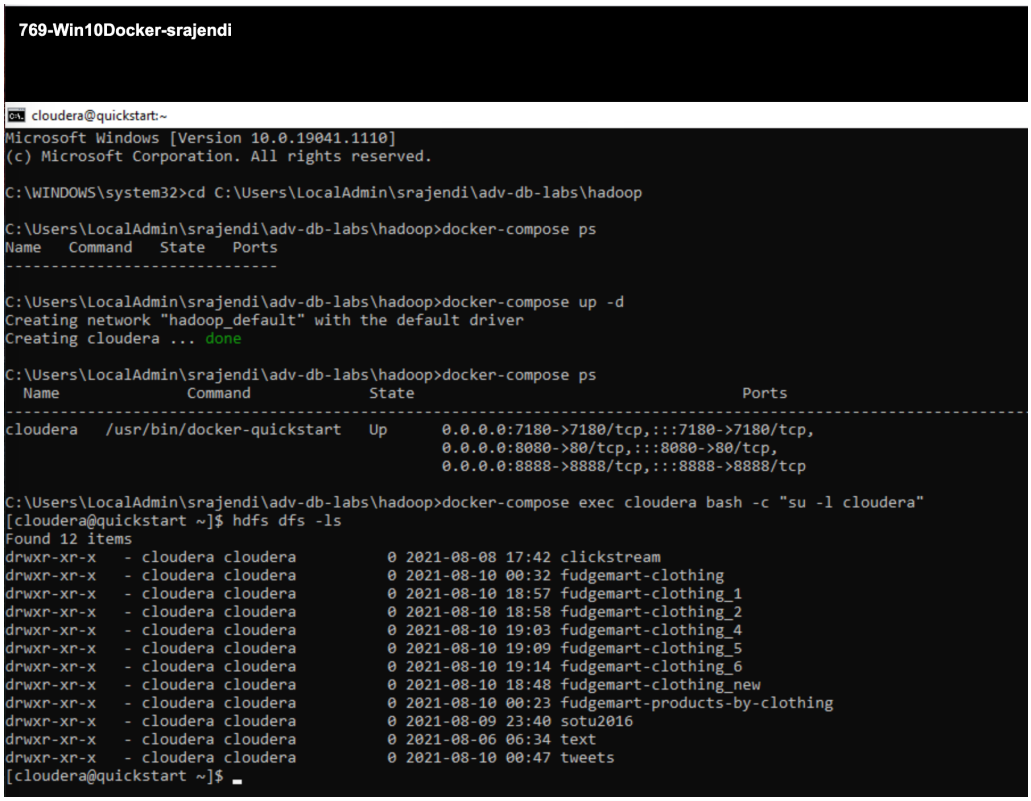


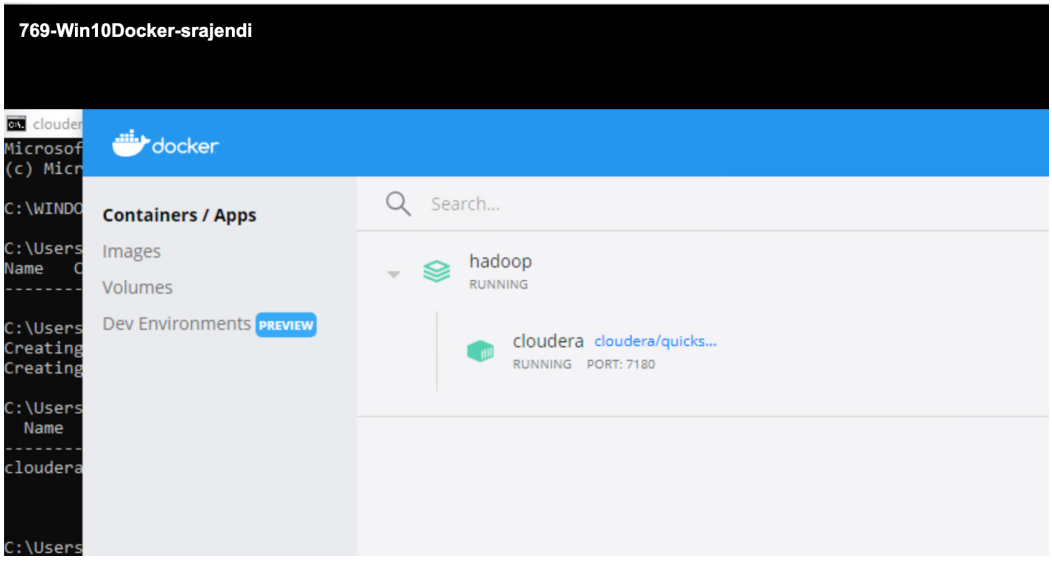
```
cloudera@quickstart:~
hbase(main):005:0> scan 'iplookup_hbase', {COLUMNS=>['ip:city','ip:state']}
ROW                                  COLUMN+CELL
 128.122.140.238                     column=ip:city, timestamp=1629701371242, value=New York
 128.122.140.238                     column=ip:state, timestamp=1629701371242, value=NY
 128.230.122.180                     column=ip:city, timestamp=1629701371242, value=Syracuse
 128.230.122.180                     column=ip:state, timestamp=1629701371242, value=NY
 155.100.169.152                     column=ip:city, timestamp=1629701371242, value=Salt Lake City
 155.100.169.152                     column=ip:state, timestamp=1629701371242, value=UT
 172.189.252.8                       column=ip:city, timestamp=1629701371242, value=Dulles
 172.189.252.8                       column=ip:state, timestamp=1629701371242, value=VA
 215.82.23.2                         column=ip:city, timestamp=1629701371242, value=Columbus
 215.82.23.2                         column=ip:state, timestamp=1629701371242, value=OH
 38.68.15.223                        column=ip:city, timestamp=1629701371242, value=Dallas
 38.68.15.223                        column=ip:state, timestamp=1629701371242, value=TX
 54.114.107.209                      column=ip:city, timestamp=1629701371242, value=Jersey City
 54.114.107.209                      column=ip:state, timestamp=1629701371242, value=NJ
 56.216.127.219                      column=ip:city, timestamp=1629701371242, value=Raleigh
 56.216.127.219                      column=ip:state, timestamp=1629701371242, value=NC
 68.199.40.156                       column=ip:city, timestamp=1629701371242, value=Freeport
 68.199.40.156                       column=ip:state, timestamp=1629701371242, value=NY
 70.209.14.54                        column=ip:city, timestamp=1629701371242, value=Tampa
 70.209.14.54                        column=ip:state, timestamp=1629701371242, value=FL
 74.111.18.59                        column=ip:city, timestamp=1629701371242, value=Syracuse
 74.111.18.59                        column=ip:state, timestamp=1629701371242, value=NY
 74.111.6.173                        column=ip:city, timestamp=1629701371242, value=Arlington
 74.111.6.173                        column=ip:state, timestamp=1629701371242, value=VA
 8.37.70.112                         column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.70.112                         column=ip:state, timestamp=1629701371242, value=CA
 8.37.70.170                         column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.70.170                         column=ip:state, timestamp=1629701371242, value=CA
```



```
cloudera@quickstart:~
 8.37.70.112                         column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.70.112                         column=ip:state, timestamp=1629701371242, value=CA
 8.37.70.170                         column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.70.170                         column=ip:state, timestamp=1629701371242, value=CA
 8.37.70.226                         column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.70.226                         column=ip:state, timestamp=1629701371242, value=CA
 8.37.70.77                          column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.70.77                          column=ip:state, timestamp=1629701371242, value=CA
 8.37.70.99                          column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.70.99                          column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.25                          column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.25                          column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.43                          column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.43                          column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.57                          column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.57                          column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.69                          column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.69                          column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.9                           column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.9                           column=ip:state, timestamp=1629701371242, value=CA
 98.29.25.44                         column=ip:city, timestamp=1629701371242, value=Cleveland
 98.29.25.44                         column=ip:state, timestamp=1629701371242, value=OH
23 row(s) in 0.0990 seconds

hbase(main):006:0>
```

Appendix


769-Win10Docker-srajendi
cloudera@quickstart:~
Microsoft Windows [Version 10.0.19041.1110]
(c) Microsoft Corporation. All rights reserved.
C:\WINDOWS\system32>cd C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose ps
Name Command State Ports
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose up -d
Creating network "hadoop_default" with the default driver
Creating cloudera ... done
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose ps
Name Command State Ports
cloudera /usr/bin/docker-quickstart Up 0.0.0.0:7180->7180/tcp,:::7180->7180/tcp,
0.0.0.0:8080->80/tcp,:::8080->80/tcp,
0.0.0.0:8888->8888/tcp,:::8888->8888/tcp
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose exec cloudera bash -c "su -l cloudera"
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 12 items
drwxr-xr-x - cloudera cloudera 0 2021-08-08 17:42 clickstream
drwxr-xr-x - cloudera cloudera 0 2021-08-10 00:32 fudgemart-clothing
drwxr-xr-x - cloudera cloudera 0 2021-08-10 18:57 fudgemart-clothing_1
drwxr-xr-x - cloudera cloudera 0 2021-08-10 18:58 fudgemart-clothing_2
drwxr-xr-x - cloudera cloudera 0 2021-08-10 19:03 fudgemart-clothing_4
drwxr-xr-x - cloudera cloudera 0 2021-08-10 19:09 fudgemart-clothing_5
drwxr-xr-x - cloudera cloudera 0 2021-08-10 19:14 fudgemart-clothing_6
drwxr-xr-x - cloudera cloudera 0 2021-08-10 18:48 fudgemart-clothing_new
drwxr-xr-x - cloudera cloudera 0 2021-08-10 00:23 fudgemart-products-by-clothing
drwxr-xr-x - cloudera cloudera 0 2021-08-09 23:40 sotu2016
drwxr-xr-x - cloudera cloudera 0 2021-08-06 06:34 text
drwxr-xr-x - cloudera cloudera 0 2021-08-10 00:47 tweets
[cloudera@quickstart ~]$


769-Win10Docker-srajendi
docker
Containers / Apps
Images
Volumes
Dev Environments
PREVIEW
Search...
hadoop
RUNNING
cloudera
cloudera/quicks...
RUNNING PORT: 7180

```
cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> clear
0: jdbc:hive2://localhost:10000/default> [cloudera@quickstart ~]$
[cloudera@quickstart ~]$ clear
[cloudera@quickstart ~]$ beeline -u jdbc:hive2://localhost:10000/default -u cloudera -p cloudera
2021-08-23 06:02:56,149 WARN  [main] mapreduce.TableMapReduceUtil: The hbase-prefix-tree module jar containing PrefixTreeCodec is not present.
  Continuing without it.
scan complete in 3ms
Connecting to jdbc:hive2://localhost:10000/default
Connected to: Apache Hive (version 1.1.0-cdh5.7.0)
Driver: Hive JDBC (version 1.1.0-cdh5.7.0)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 1.1.0-cdh5.7.0 by Apache Hive
0: jdbc:hive2://localhost:10000/default> show databases;
INFO  : Compiling command(queryId=hive_20210823060303_042f4588-b0ba-4ca2-9b38-63c6271b4255): show databases
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:database_name, type:string, comment:from deserializer)], properties:null)
INFO  : Completed compiling command(queryId=hive_20210823060303_042f4588-b0ba-4ca2-9b38-63c6271b4255); Time taken: 0.014 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210823060303_042f4588-b0ba-4ca2-9b38-63c6271b4255): show databases
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210823060303_042f4588-b0ba-4ca2-9b38-63c6271b4255); Time taken: 0.046 seconds
INFO  : OK
+----------------+--+
| database_name  |
+----------------+--+
| clickstream    |
| default        |
+----------------+--+
2 rows selected (0.216 seconds)
```

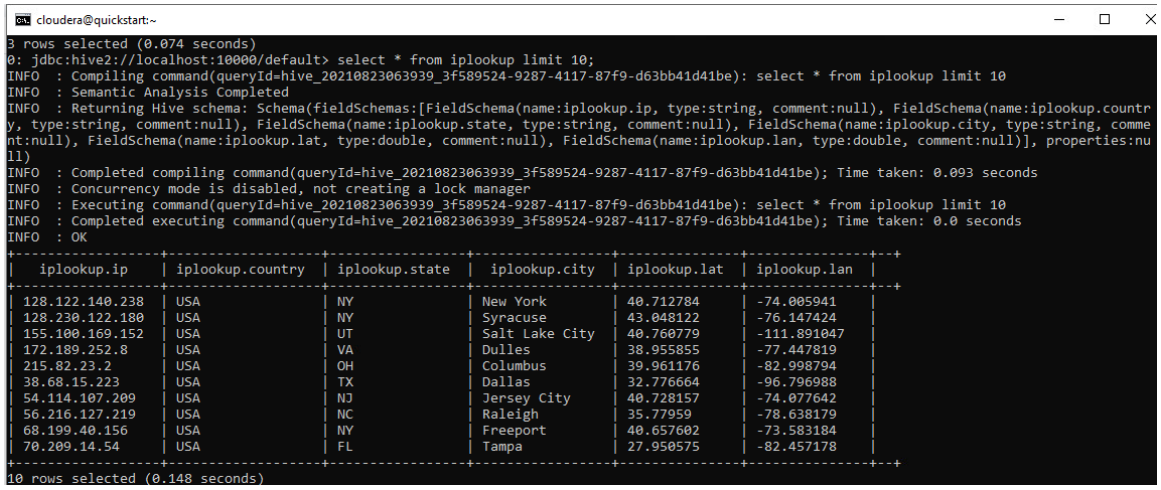
```
cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> use clickstream;
INFO  : Compiling command(queryId=hive_20210823060303_5e1a7c70-474f-447a-b52e-7dcccdf7bd5f): use clickstream
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20210823060303_5e1a7c70-474f-447a-b52e-7dcccdf7bd5f); Time taken: 0.027 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210823060303_5e1a7c70-474f-447a-b52e-7dcccdf7bd5f): use clickstream
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210823060303_5e1a7c70-474f-447a-b52e-7dcccdf7bd5f); Time taken: 0.015 seconds
INFO  : OK
No rows affected (0.066 seconds)
0: jdbc:hive2://localhost:10000/default> show tables;
INFO  : Compiling command(queryId=hive_20210823060404_62882c5d-1704-4e78-b31b-4ddcdc64724e): show tables
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:tab_name, type:string, comment:from deserializer)], properties:null)
INFO  : Completed compiling command(queryId=hive_20210823060404_62882c5d-1704-4e78-b31b-4ddcdc64724e); Time taken: 0.012 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210823060404_62882c5d-1704-4e78-b31b-4ddcdc64724e): show tables
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210823060404_62882c5d-1704-4e78-b31b-4ddcdc64724e); Time taken: 0.014 seconds
INFO  : OK
+-----------+--+
| tab_name  |
+-----------+--+
| iplookup  |
| weblogs   |
+-----------+--+
2 rows selected (0.057 seconds)
0: jdbc:hive2://localhost:10000/default>
```

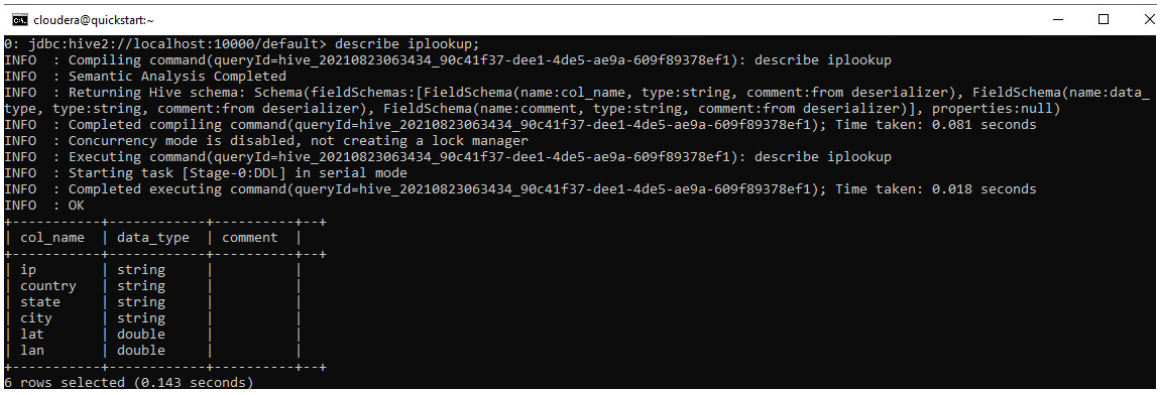
```
cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> create external table computers
0: jdbc:hive2://localhost:10000/default> ( id int, Model string, GB_RAM int, TB_Disk float)
0: jdbc:hive2://localhost:10000/default> stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
0: jdbc:hive2://localhost:10000/default> WITH
0: jdbc:hive2://localhost:10000/default> SERDEPROPERTIES ('hbase.columns.mapping' = ':key, info:Model, info:GB_RAM, info:TB_Disk')
0: jdbc:hive2://localhost:10000/default> TBLPROPERTIES ('hbase.table.name' = 'computers');
INFO  : Compiling command(queryId=hive_20210823061616_6e463fdf-3470-4d38-85e6-37b86588ac54): create external table computers
( id int, Model string, GB_RAM int, TB_Disk float)
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')
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20210823061616_6e463fdf-3470-4d38-85e6-37b86588ac54); Time taken: 0.015 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210823061616_6e463fdf-3470-4d38-85e6-37b86588ac54): create external table computers
( id int, Model string, GB_RAM int, TB_Disk float)
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')
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210823061616_6e463fdf-3470-4d38-85e6-37b86588ac54); Time taken: 0.173 seconds
INFO  : OK
No rows affected (0.21 seconds)
0: jdbc:hive2://localhost:10000/default> show tables;
```

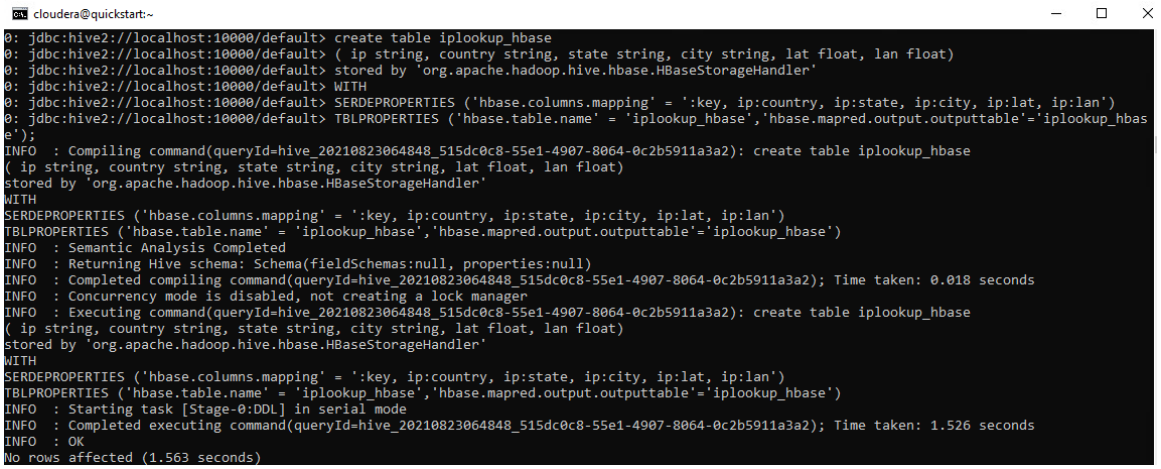
```
cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> show tables;
INFO  : Compiling command(queryId=hive_20210823061919_18b44205-f3ef-4a61-b1ee-ff0077cfcf20): show tables
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:tab_name, type:string, comment:from deserializer)], properties:null)
INFO  : Completed compiling command(queryId=hive_20210823061919_18b44205-f3ef-4a61-b1ee-ff0077cfcf20); Time taken: 0.008 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210823061919_18b44205-f3ef-4a61-b1ee-ff0077cfcf20): show tables
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210823061919_18b44205-f3ef-4a61-b1ee-ff0077cfcf20); Time taken: 0.012 seconds
INFO  : OK
+------------+--+
|  tab_name  |
+------------+--+
| computers  |
| iplookup   |
| weblogs    |
+------------+--+
3 rows selected (0.05 seconds)
```

```
cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> select * from computers;
INFO  : Compiling command(queryId=hive_20210823062020_45ec7b36-01e0-4af3-9b87-43f469599a59): select * from computers
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:computers.id, type:int, comment:null), FieldSchema(name:computers.model,
type:string, comment:null), FieldSchema(name:computers.gb_ram, type:int, comment:null), FieldSchema(name:computers.tb_disk, type:float, commen
t:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20210823062020_45ec7b36-01e0-4af3-9b87-43f469599a59); Time taken: 0.183 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210823062020_45ec7b36-01e0-4af3-9b87-43f469599a59): select * from computers
INFO  : Completed executing command(queryId=hive_20210823062020_45ec7b36-01e0-4af3-9b87-43f469599a59); Time taken: 0.001 seconds
INFO  : OK
+---------------+-----------------+------------------+--------------------+--+
| computers.id  | computers.model | computers.gb_ram | computers.tb_disk  |
+---------------+-----------------+------------------+--------------------+--+
| 1             | Dell            | 16               | 1.0                |
| 2             | IBM             | 32               | 1.5                |
| 3             | HP              | 8                | 1.0                |
| 4             | Acer            | 16               | 2.0                |
+---------------+-----------------+------------------+--------------------+--+
4 rows selected (0.417 seconds)
0: jdbc:hive2://localhost:10000/default>
```

```
cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> select sum(GB_RAM) as Total_Memory, sum(TB_Disk) as Total_Storage from computers;
INFO  : Compiling command(queryId=hive_20210823062525_bfc47725-1472-46a7-91e2-f0971ad4f590): select sum(GB_RAM) as Total_Memory, sum(TB_Disk)
as Total_Storage from computers
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:total_memory, type:bigint, comment:null), FieldSchema(name:total_storage,
 type:double, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20210823062525_bfc47725-1472-46a7-91e2-f0971ad4f590); Time taken: 0.193 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210823062525_bfc47725-1472-46a7-91e2-f0971ad4f590): select sum(GB_RAM) as Total_Memory, sum(TB_Disk)
as Total_Storage from computers
INFO  : Query ID = hive_20210823062525_bfc47725-1472-46a7-91e2-f0971ad4f590
INFO  : Total jobs = 1
INFO  : Launching Job 1 out of 1
INFO  : Starting task [Stage-1:MAPRED] in serial mode
INFO  : Number of reduce tasks determined at compile time: 1
INFO  : In order to change the average load for a reducer (in bytes):
INFO  :   set hive.exec.reducers.bytes.per.reducer=<number>
INFO  : In order to limit the maximum number of reducers:
INFO  :   set hive.exec.reducers.max=<number>
INFO  : In order to set a constant number of reducers:
INFO  :   set mapreduce.job.reduces=<number>
INFO  : Starting Job = job_1629241503996_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1629241503996_0006/
INFO  : Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1629241503996_0006
INFO  : Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
INFO  : 2021-08-23 06:25:39,111 Stage-1 map = 0%,  reduce = 0%
INFO  : 2021-08-23 06:25:56,177 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.11 sec
INFO  : 2021-08-23 06:26:05,767 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.35 sec
INFO  : MapReduce Total cumulative CPU time: 5 seconds 350 msec
INFO  : Ended Job = job_1629241503996_0006
INFO  : MapReduce Jobs Launched:
INFO  : Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.35 sec   HDFS Read: 15369 HDFS Write: 7 SUCCESS
INFO  : Total MapReduce CPU Time Spent: 5 seconds 350 msec
INFO  : Completed executing command(queryId=hive_20210823062525_bfc47725-1472-46a7-91e2-f0971ad4f590); Time taken: 36.877 seconds
INFO  : OK
+---------------+----------------+--+
| total_memory  | total_storage  |
+---------------+----------------+--+
| 72            | 5.5            |
+---------------+----------------+--+
1 row selected (37.139 seconds)
0: jdbc:hive2://localhost:10000/default>
```


cloudera@quickstart:~
3 rows selected (0.074 seconds)
0: jdbc:hive2://localhost:10000/default> select * from iplookup limit 10;
INFO : Compiling command(queryId=hive_20210823063939_3f589524-9287-4117-87f9-d63bb41d41be): select * from iplookup limit 10
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:iplookup.ip, type:string, comment:null), FieldSchema(name:iplookup.country, type:string, comment:null), FieldSchema(name:iplookup.state, type:string, comment:null), FieldSchema(name:iplookup.city, type:string, comment:null), FieldSchema(name:iplookup.lat, type:double, comment:null), FieldSchema(name:iplookup.lan, type:double, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20210823063939_3f589524-9287-4117-87f9-d63bb41d41be); Time taken: 0.093 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20210823063939_3f589524-9287-4117-87f9-d63bb41d41be): select * from iplookup limit 10
INFO : Completed executing command(queryId=hive_20210823063939_3f589524-9287-4117-87f9-d63bb41d41be); Time taken: 0.0 seconds
INFO : OK
iplookup.ip	iplookup.country	iplookup.state	iplookup.city	iplookup.lat	iplookup.lan
128.122.140.238	USA	NY	New York	40.712784	-74.005941
128.230.122.180	USA	NY	Syracuse	43.048122	-76.147424
155.100.169.152	USA	UT	Salt Lake City	40.760779	-111.891047
172.189.252.8	USA	VA	Dulles	38.955855	-77.447819
215.82.23.2	USA	OH	Columbus	39.961176	-82.998794
38.68.15.223	USA	TX	Dallas	32.776664	-96.796988
54.114.107.209	USA	NJ	Jersey City	40.728157	-74.077642
56.216.127.219	USA	NC	Raleigh	35.77959	-78.638179
68.199.40.156	USA	NY	Freeport	40.657602	-73.583184
70.209.14.54	USA	FL	Tampa	27.950575	-82.457178
10 rows selected (0.148 seconds)

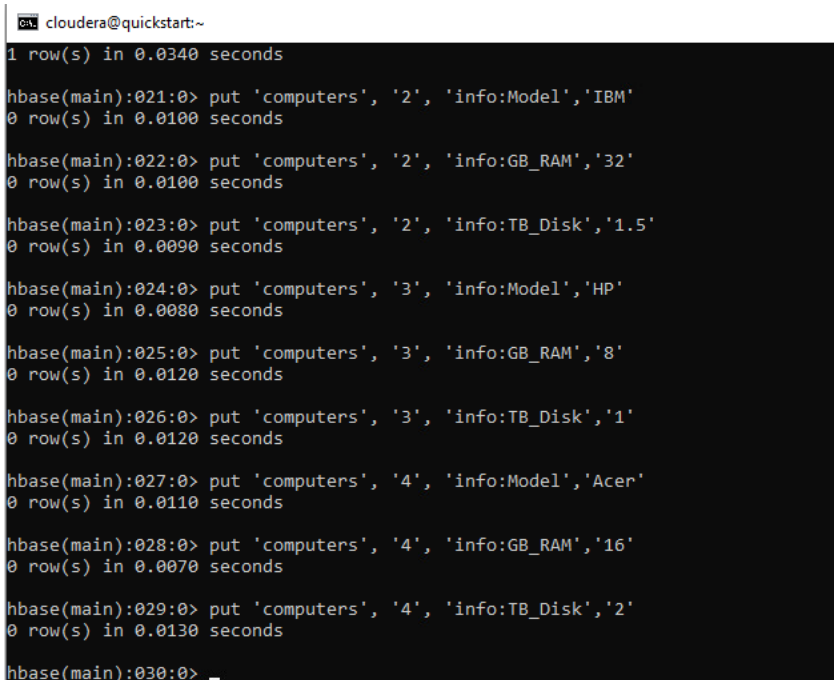

cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> describe iplookup;
INFO : Compiling command(queryId=hive_20210823063434_90c41f37-dee1-4de5-ae9a-609f89378ef1): describe iplookup
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from deserializer), FieldSchema(name:data_type, type:string, comment:from deserializer), FieldSchema(name:comment, type:string, comment:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=hive_20210823063434_90c41f37-dee1-4de5-ae9a-609f89378ef1); Time taken: 0.081 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20210823063434_90c41f37-dee1-4de5-ae9a-609f89378ef1): describe iplookup
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20210823063434_90c41f37-dee1-4de5-ae9a-609f89378ef1); Time taken: 0.018 seconds
INFO : OK
col_name	data_type	comment
ip	string	
country	string	
state	string	
city	string	
lat	double	
lan	double	
6 rows selected (0.143 seconds)

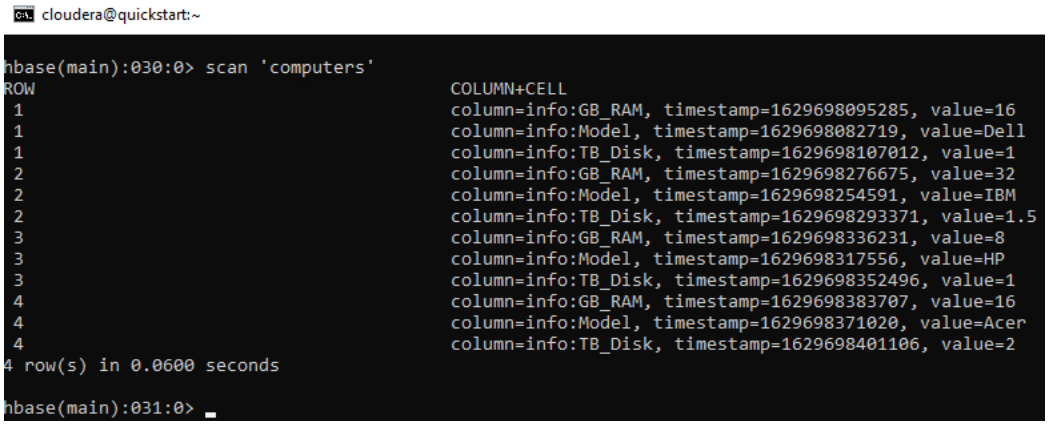

cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> create table iplookup_hbase
0: jdbc:hive2://localhost:10000/default> (ip string, country string, state string, city string, lat float, lan float)
0: jdbc:hive2://localhost:10000/default> stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
0: jdbc:hive2://localhost:10000/default> WITH
0: jdbc:hive2://localhost:10000/default> SERDEPROPERTIES ('hbase.columns.mapping' = ':key, ip:country, ip:state, ip:city, ip:lat, ip:lan')
0: jdbc:hive2://localhost:10000/default> TBLPROPERTIES ('hbase.table.name' = 'iplookup_hbase','hbase.mapred.output.outputtable'='iplookup_hbase');
INFO : Compiling command(queryId=hive_20210823064848_515dc0c8-55e1-4907-8064-0c2b5911a3a2): create table iplookup_hbase
(ip string, country string, state string, city string, lat float, lan float)
stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
WITH
SERDEPROPERTIES ('hbase.columns.mapping' = ':key, ip:country, ip:state, ip:city, ip:lat, ip:lan')
TBLPROPERTIES ('hbase.table.name' = 'iplookup_hbase','hbase.mapred.output.outputtable'='iplookup_hbase')
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=hive_20210823064848_515dc0c8-55e1-4907-8064-0c2b5911a3a2); Time taken: 0.018 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20210823064848_515dc0c8-55e1-4907-8064-0c2b5911a3a2): create table iplookup_hbase
(ip string, country string, state string, city string, lat float, lan float)
stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
WITH
SERDEPROPERTIES ('hbase.columns.mapping' = ':key, ip:country, ip:state, ip:city, ip:lat, ip:lan')
TBLPROPERTIES ('hbase.table.name' = 'iplookup_hbase','hbase.mapred.output.outputtable'='iplookup_hbase')
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20210823064848_515dc0c8-55e1-4907-8064-0c2b5911a3a2); Time taken: 1.526 seconds
INFO : OK
No rows affected (1.563 seconds)

```
cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> INSERT OVERWRITE TABLE iplookup_hbase SELECT * FROM iplookup;
INFO  : Compiling command(queryId=hive_20210823064949_b0c49dd8-6767-421c-8f82-35e67f095b8e): INSERT OVERWRITE TABLE iplookup_hbase SELECT * FR
OM iplookup
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:_col0, type:string, comment:null), FieldSchema(name:_col1, type:string, c
omment:null), FieldSchema(name:_col2, type:string, comment:null), FieldSchema(name:_col3, type:string, comment:null), FieldSchema(name:_col4,
type:float, comment:null), FieldSchema(name:_col5, type:float, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20210823064949_b0c49dd8-6767-421c-8f82-35e67f095b8e); Time taken: 0.22 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210823064949_b0c49dd8-6767-421c-8f82-35e67f095b8e): INSERT OVERWRITE TABLE iplookup_hbase SELECT * FR
OM iplookup
INFO  : Query ID = hive_20210823064949_b0c49dd8-6767-421c-8f82-35e67f095b8e
INFO  : Total jobs = 1
INFO  : Launching Job 1 out of 1
INFO  : Starting task [Stage-0:MAPRED] in serial mode
INFO  : Number of reduce tasks is set to 0 since there's no reduce operator
INFO  : Starting Job = job_1629241503996_0007, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1629241503996_0007/
INFO  : Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1629241503996_0007
INFO  : Hadoop job information for Stage-0: number of mappers: 1; number of reducers: 0
INFO  : 2021-08-23 06:49:24,561 Stage-0 map = 0%,  reduce = 0%
INFO  : 2021-08-23 06:49:31,944 Stage-0 map = 100%,  reduce = 0%, Cumulative CPU 2.99 sec
INFO  : MapReduce Total cumulative CPU time: 2 seconds 990 msec
INFO  : Ended Job = job_1629241503996_0007
INFO  : MapReduce Jobs Launched:
INFO  : Stage-Stage-0: Map: 1   Cumulative CPU: 2.99 sec   HDFS Read: 12766 HDFS Write: 0 SUCCESS
INFO  : Total MapReduce CPU Time Spent: 2 seconds 990 msec
INFO  : Completed executing command(queryId=hive_20210823064949_b0c49dd8-6767-421c-8f82-35e67f095b8e); Time taken: 17.734 seconds
INFO  : OK
No rows affected (17.986 seconds)
0: jdbc:hive2://localhost:10000/default>
```

```
cloudera@quickstart:~
[cloudera@quickstart ~]$ hbase shell
2021-08-23 05:45:18,427 INFO  [main] Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.7.0, rUnknown, Wed Mar 23 11:39:14 PDT 2016

hbase(main):001:0> list
TABLE
analytics_demo
document_demo
2 row(s) in 0.2290 seconds

=> ["analytics_demo", "document_demo"]
hbase(main):002:0> t = create 'computers', 'info'
0 row(s) in 1.2630 seconds

=> Hbase::Table - computers
hbase(main):003:0> list
TABLE
analytics_demo
computers
document_demo
3 row(s) in 0.0170 seconds

=> ["analytics_demo", "computers", "document_demo"]
hbase(main):004:0> scan 'computers'
ROW                                   COLUMN+CELL
0 row(s) in 0.1580 seconds

hbase(main):005:0>
```

```
cloudera@quickstart:~
=> ["analytics_demo", "document_demo"]
hbase(main):015:0> t = create 'computers', 'info'
0 row(s) in 1.2400 seconds

=> Hbase::Table - computers
hbase(main):016:0> list
TABLE
analytics_demo
computers
document_demo
3 row(s) in 0.0220 seconds

=> ["analytics_demo", "computers", "document_demo"]
hbase(main):017:0> put 'computers', '1', 'info:Model','Dell'
0 row(s) in 0.0210 seconds

hbase(main):018:0> put 'computers', '1', 'info:GB_RAM','16'
0 row(s) in 0.0140 seconds

hbase(main):019:0> put 'computers', '1', 'info:TB_Disk','1'
0 row(s) in 0.0150 seconds

hbase(main):020:0> scan 'computers'
ROW                                   COLUMN+CELL
 1                                    column=info:GB_RAM, timestamp=1629698095285, value=16
 1                                    column=info:Model, timestamp=1629698082719, value=Dell
 1                                    column=info:TB_Disk, timestamp=1629698107012, value=1
1 row(s) in 0.0340 seconds

hbase(main):021:0>
```


cloudera@quickstart:~
1 row(s) in 0.0340 seconds
hbase(main):021:0> put 'computers', '2', 'info:Model','IBM'
0 row(s) in 0.0100 seconds
hbase(main):022:0> put 'computers', '2', 'info:GB_RAM','32'
0 row(s) in 0.0100 seconds
hbase(main):023:0> put 'computers', '2', 'info:TB_Disk','1.5'
0 row(s) in 0.0090 seconds
hbase(main):024:0> put 'computers', '3', 'info:Model','HP'
0 row(s) in 0.0080 seconds
hbase(main):025:0> put 'computers', '3', 'info:GB_RAM','8'
0 row(s) in 0.0120 seconds
hbase(main):026:0> put 'computers', '3', 'info:TB_Disk','1'
0 row(s) in 0.0120 seconds
hbase(main):027:0> put 'computers', '4', 'info:Model','Acer'
0 row(s) in 0.0110 seconds
hbase(main):028:0> put 'computers', '4', 'info:GB_RAM','16'
0 row(s) in 0.0070 seconds
hbase(main):029:0> put 'computers', '4', 'info:TB_Disk','2'
0 row(s) in 0.0130 seconds
hbase(main):030:0>

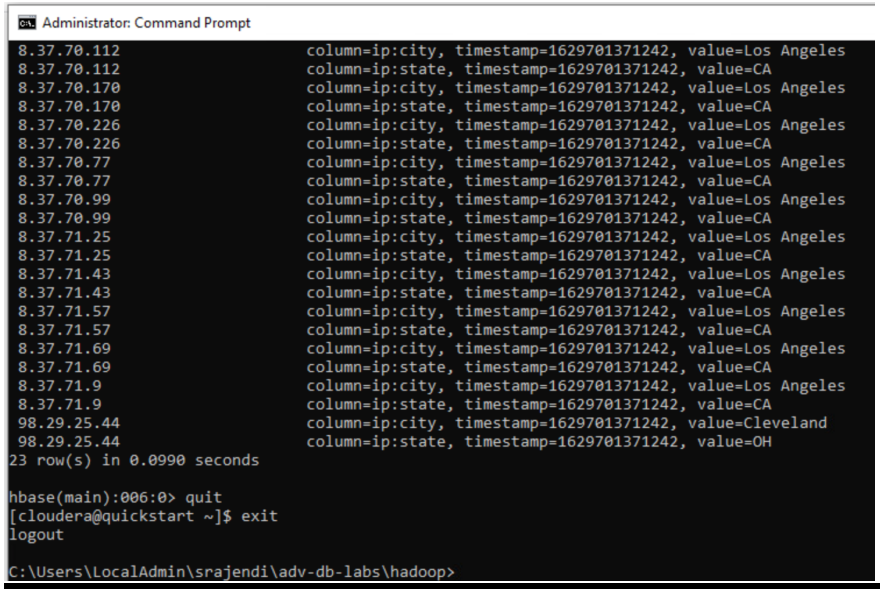

cloudera@quickstart:~
hbase(main):030:0> scan 'computers'
ROW COLUMN+CELL
1 column=info:GB_RAM, timestamp=1629698095285, value=16
1 column=info:Model, timestamp=1629698082719, value=Dell
1 column=info:TB_Disk, timestamp=1629698107012, value=1
2 column=info:GB_RAM, timestamp=1629698276675, value=32
2 column=info:Model, timestamp=1629698254591, value=IBM
2 column=info:TB_Disk, timestamp=1629698293371, value=1.5
3 column=info:GB_RAM, timestamp=1629698336231, value=8
3 column=info:Model, timestamp=1629698317556, value=HP
3 column=info:TB_Disk, timestamp=1629698352496, value=1
4 column=info:GB_RAM, timestamp=1629698383707, value=16
4 column=info:Model, timestamp=1629698371020, value=Acer
4 column=info:TB_Disk, timestamp=1629698401106, value=2
4 row(s) in 0.0600 seconds
hbase(main):031:0>

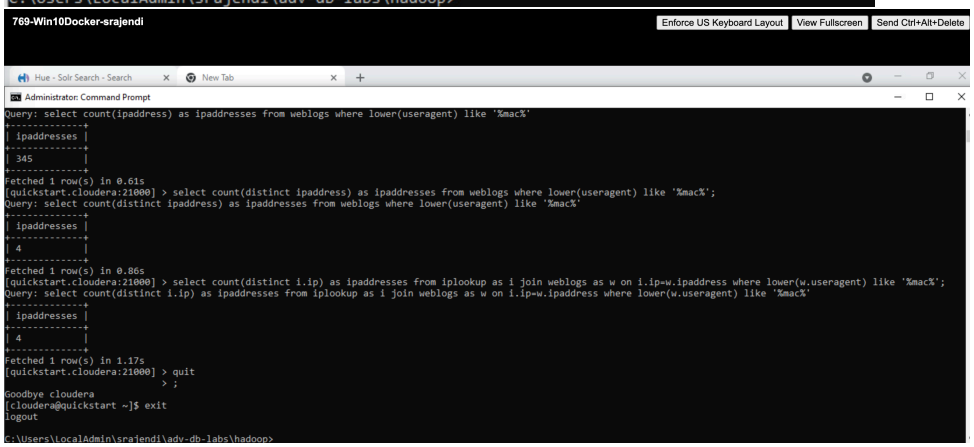
```
cloudera@quickstart:~
hbase(main):003:0> list
TABLE
analytics_demo
computers
document_demo
iplookup_hbase
4 row(s) in 0.0200 seconds

=> ["analytics_demo", "computers", "document_demo", "iplookup_hbase"]
hbase(main):004:0> scan 'iplookup_hbase'
ROW                             COLUMN+CELL
 128.122.140.238                column=ip:city, timestamp=1629701371242, value=New York
 128.122.140.238                column=ip:country, timestamp=1629701371242, value=USA
 128.122.140.238                column=ip:lan, timestamp=1629701371242, value=-74.00594
 128.122.140.238                column=ip:lat, timestamp=1629701371242, value=40.712784
 128.122.140.238                column=ip:state, timestamp=1629701371242, value=NY
 128.230.122.180                column=ip:city, timestamp=1629701371242, value=Syracuse
 128.230.122.180                column=ip:country, timestamp=1629701371242, value=USA
 128.230.122.180                column=ip:lan, timestamp=1629701371242, value=-76.14742
 128.230.122.180                column=ip:lat, timestamp=1629701371242, value=43.048122
 128.230.122.180                column=ip:state, timestamp=1629701371242, value=NY
 155.100.169.152                column=ip:city, timestamp=1629701371242, value=Salt Lake City
 155.100.169.152                column=ip:country, timestamp=1629701371242, value=USA
 155.100.169.152                column=ip:lan, timestamp=1629701371242, value=-111.891045
 155.100.169.152                column=ip:lat, timestamp=1629701371242, value=40.76078
 155.100.169.152                column=ip:state, timestamp=1629701371242, value=UT
 172.189.252.8                  column=ip:city, timestamp=1629701371242, value=Dulles
 172.189.252.8                  column=ip:country, timestamp=1629701371242, value=USA
 172.189.252.8                  column=ip:lan, timestamp=1629701371242, value=-77.44782
 172.189.252.8                  column=ip:lat, timestamp=1629701371242, value=38.955856
```

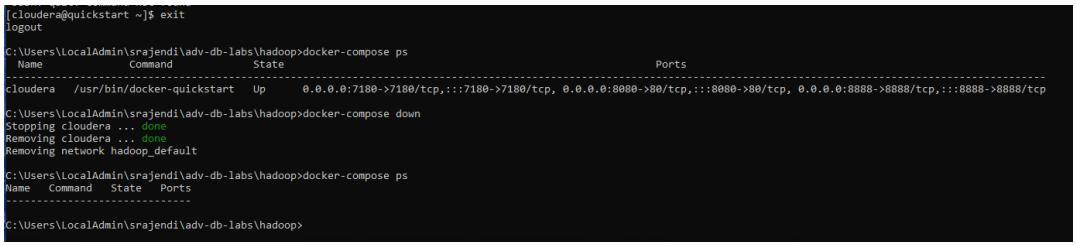
```
cloudera@quickstart:~
 8.37.71.25                     column=ip:lat, timestamp=1629701371242, value=34.052235
 8.37.71.25                     column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.43                     column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.43                     column=ip:country, timestamp=1629701371242, value=USA
 8.37.71.43                     column=ip:lan, timestamp=1629701371242, value=-118.24368
 8.37.71.43                     column=ip:lat, timestamp=1629701371242, value=34.052235
 8.37.71.43                     column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.57                     column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.57                     column=ip:country, timestamp=1629701371242, value=USA
 8.37.71.57                     column=ip:lan, timestamp=1629701371242, value=-118.24368
 8.37.71.57                     column=ip:lat, timestamp=1629701371242, value=34.052235
 8.37.71.57                     column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.69                     column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.69                     column=ip:country, timestamp=1629701371242, value=USA
 8.37.71.69                     column=ip:lan, timestamp=1629701371242, value=-118.24368
 8.37.71.69                     column=ip:lat, timestamp=1629701371242, value=34.052235
 8.37.71.69                     column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.9                      column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.9                      column=ip:country, timestamp=1629701371242, value=USA
 8.37.71.9                      column=ip:lan, timestamp=1629701371242, value=-118.24368
 8.37.71.9                      column=ip:lat, timestamp=1629701371242, value=34.052235
 8.37.71.9                      column=ip:state, timestamp=1629701371242, value=CA
 98.29.25.44                    column=ip:city, timestamp=1629701371242, value=Cleveland
 98.29.25.44                    column=ip:country, timestamp=1629701371242, value=USA
 98.29.25.44                    column=ip:lan, timestamp=1629701371242, value=-81.69436
 98.29.25.44                    column=ip:lat, timestamp=1629701371242, value=41.49932
 98.29.25.44                    column=ip:state, timestamp=1629701371242, value=OH
23 row(s) in 0.7570 seconds

hbase(main):005:0>
```

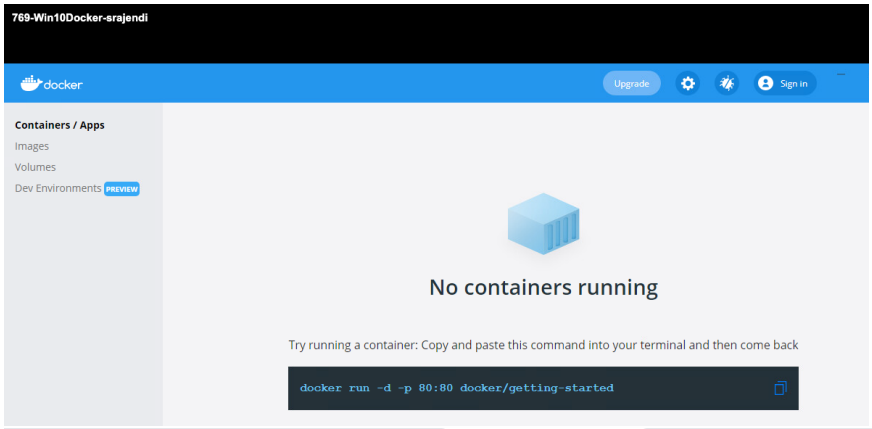
```
cloudera@quickstart:~
hbase(main):005:0> scan 'iplookup_hbase', {COLUMNS=>['ip:city','ip:state']}
ROW                           COLUMN+CELL
 128.122.140.238              column=ip:city, timestamp=1629701371242, value=New York
 128.122.140.238              column=ip:state, timestamp=1629701371242, value=NY
 128.230.122.180              column=ip:city, timestamp=1629701371242, value=Syracuse
 128.230.122.180              column=ip:state, timestamp=1629701371242, value=NY
 155.100.169.152              column=ip:city, timestamp=1629701371242, value=Salt Lake City
 155.100.169.152              column=ip:state, timestamp=1629701371242, value=UT
 172.189.252.8                column=ip:city, timestamp=1629701371242, value=Dulles
 172.189.252.8                column=ip:state, timestamp=1629701371242, value=VA
 215.82.23.2                  column=ip:city, timestamp=1629701371242, value=Columbus
 215.82.23.2                  column=ip:state, timestamp=1629701371242, value=OH
 38.68.15.223                 column=ip:city, timestamp=1629701371242, value=Dallas
 38.68.15.223                 column=ip:state, timestamp=1629701371242, value=TX
 54.114.107.209               column=ip:city, timestamp=1629701371242, value=Jersey City
 54.114.107.209               column=ip:state, timestamp=1629701371242, value=NJ
 56.216.127.219               column=ip:city, timestamp=1629701371242, value=Raleigh
 56.216.127.219               column=ip:state, timestamp=1629701371242, value=NC
 68.199.40.156                column=ip:city, timestamp=1629701371242, value=Freeport
 68.199.40.156                column=ip:state, timestamp=1629701371242, value=NY
 70.209.14.54                 column=ip:city, timestamp=1629701371242, value=Tampa
 70.209.14.54                 column=ip:state, timestamp=1629701371242, value=FL
 74.111.18.59                 column=ip:city, timestamp=1629701371242, value=Syracuse
 74.111.18.59                 column=ip:state, timestamp=1629701371242, value=NY
 74.111.6.173                 column=ip:city, timestamp=1629701371242, value=Arlington
 74.111.6.173                 column=ip:state, timestamp=1629701371242, value=VA
 8.37.70.112                  column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.70.112                  column=ip:state, timestamp=1629701371242, value=CA
 8.37.70.170                  column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.70.170                  column=ip:state, timestamp=1629701371242, value=CA
```

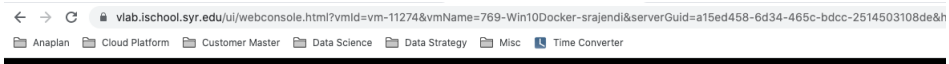
```
cloudera@quickstart:~
 8.37.70.112                  column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.70.112                  column=ip:state, timestamp=1629701371242, value=CA
 8.37.70.170                  column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.70.170                  column=ip:state, timestamp=1629701371242, value=CA
 8.37.70.226                  column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.70.226                  column=ip:state, timestamp=1629701371242, value=CA
 8.37.70.77                   column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.70.77                   column=ip:state, timestamp=1629701371242, value=CA
 8.37.70.99                   column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.70.99                   column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.25                   column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.25                   column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.43                   column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.43                   column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.57                   column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.57                   column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.69                   column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.69                   column=ip:state, timestamp=1629701371242, value=CA
 8.37.71.9                    column=ip:city, timestamp=1629701371242, value=Los Angeles
 8.37.71.9                    column=ip:state, timestamp=1629701371242, value=CA
 98.29.25.44                  column=ip:city, timestamp=1629701371242, value=Cleveland
 98.29.25.44                  column=ip:state, timestamp=1629701371242, value=OH
23 row(s) in 0.0990 seconds

hbase(main):006:0>
```


Administrator: Command Prompt
8.37.70.112 column=ip:city, timestamp=1629701371242, value=Los Angeles
8.37.70.112 column=ip:state, timestamp=1629701371242, value=CA
8.37.70.170 column=ip:city, timestamp=1629701371242, value=Los Angeles
8.37.70.170 column=ip:state, timestamp=1629701371242, value=CA
8.37.70.226 column=ip:city, timestamp=1629701371242, value=Los Angeles
8.37.70.226 column=ip:state, timestamp=1629701371242, value=CA
8.37.70.77 column=ip:city, timestamp=1629701371242, value=Los Angeles
8.37.70.77 column=ip:state, timestamp=1629701371242, value=CA
8.37.70.99 column=ip:city, timestamp=1629701371242, value=Los Angeles
8.37.70.99 column=ip:state, timestamp=1629701371242, value=CA
8.37.71.25 column=ip:city, timestamp=1629701371242, value=Los Angeles
8.37.71.25 column=ip:state, timestamp=1629701371242, value=CA
8.37.71.43 column=ip:city, timestamp=1629701371242, value=Los Angeles
8.37.71.43 column=ip:state, timestamp=1629701371242, value=CA
8.37.71.57 column=ip:city, timestamp=1629701371242, value=Los Angeles
8.37.71.57 column=ip:state, timestamp=1629701371242, value=CA
8.37.71.69 column=ip:city, timestamp=1629701371242, value=Los Angeles
8.37.71.69 column=ip:state, timestamp=1629701371242, value=CA
8.37.71.9 column=ip:city, timestamp=1629701371242, value=Los Angeles
8.37.71.9 column=ip:state, timestamp=1629701371242, value=CA
98.29.25.44 column=ip:city, timestamp=1629701371242, value=Cleveland
98.29.25.44 column=ip:state, timestamp=1629701371242, value=OH
23 row(s) in 0.0990 seconds
hbase(main):006:0> quit
[cloudera@quickstart ~]$ exit
logout
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>

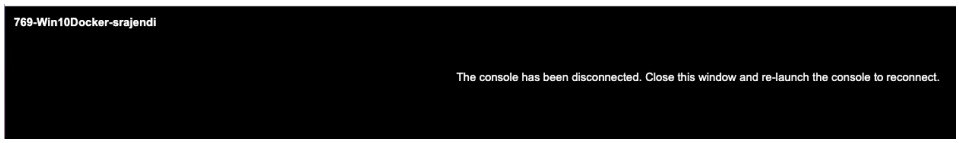

769-Win10Docker-srajendi
Enforce US Keyboard Layout
View Fullscreen
Send Ctrl+Alt+Delete
Hue - Solr Search - Search
New Tab
Administrator: Command Prompt
Query: select count(ipaddress) as ipaddresses from weblogs where lower(useragent) like '%mac%'
| ipaddresses |
| 345 |
Fetched 1 row(s) in 0.61s
[quickstart.cloudera:21000] > select count(distinct ipaddress) as ipaddresses from weblogs where lower(useragent) like '%mac%';
Query: select count(distinct ipaddress) as ipaddresses from weblogs where lower(useragent) like '%mac%'
| ipaddresses |
| 4 |
Fetched 1 row(s) in 0.86s
[quickstart.cloudera:21000] > select count(distinct i.ip) as ipaddresses from iplookup as i join weblogs as w on i.ip=w.ipaddress where lower(w.useragent) like '%mac%';
Query: select count(distinct i.ip) as ipaddresses from iplookup as i join weblogs as w on i.ip=w.ipaddress where lower(w.useragent) like '%mac%'
| ipaddresses |
| 4 |
Fetched 1 row(s) in 1.17s
[quickstart.cloudera:21000] > quit
> ;
Goodbye cloudera
[cloudera@quickstart ~]$ exit
logout
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>

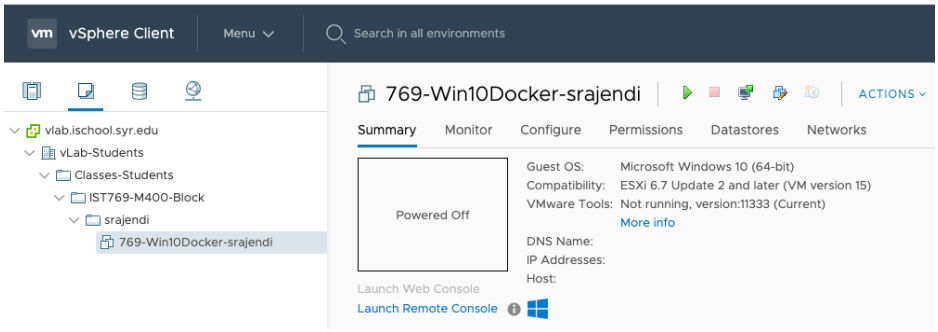

[cloudera@quickstart ~]$ exit
logout
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose ps
Name Command State Ports
cloudera /usr/bin/docker-quickstart Up 0.0.0.0:7180->7180/tcp,:::7180->7180/tcp, 0.0.0.0:8080->80/tcp,:::8080->80/tcp, 0.0.0.0:8888->8888/tcp,:::8888->8888/tcp
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose down
Stopping cloudera ... done
Removing cloudera ... done
Removing network hadoop_default
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose ps
Name Command State Ports
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>


769-Win10Docker-srajendi
docker
Upgrade
Sign in
Containers / Apps
Images
Volumes
Dev Environments
PREVIEW
No containers running
Try running a container: Copy and paste this command into your terminal and then come back
docker run -d -p 80:80 docker/getting-started


vlab.ischool.syr.edu/ui/webconsole.html?vmId=vm-11274&vmName=769-Win10Docker-srajendi&serverGuid=a15ed458-6d34-465c-bdcc-2514503108de&h
Anaplan
Cloud Platform
Customer Master
Data Science
Data Strategy
Misc
Time Converter


769-Win10Docker-srajendi
Shutting down


769-Win10Docker-srajendi
The console has been disconnected. Close this window and re-launch the console to reconnect.


vm
vSphere Client
Menu
Search in all environments
vlab.ischool.syr.edu
vLab-Students
Classes-Students
IST769-M400-Block
srajendi
769-Win10Docker-srajendi
769-Win10Docker-srajendi
ACTIONS
Summary
Monitor
Configure
Permissions
Datastores
Networks
Powered Off
Guest OS: Microsoft Windows 10 (64-bit)
Compatibility: ESXi 6.7 Update 2 and later (VM version 15)
VMware Tools: Not running, version:11333 (Current)
More info
DNS Name:
IP Addresses:
Host:
Launch Web Console
Launch Remote Console