IST769 Homework Submission

Name: **Sathish Kumar Rajendiran**
SUID: **666555028**
Email: srajendi@syr.edu
Due Date: **08/17/2021**
Task: **Hive and Pig**
Homework #:**6**

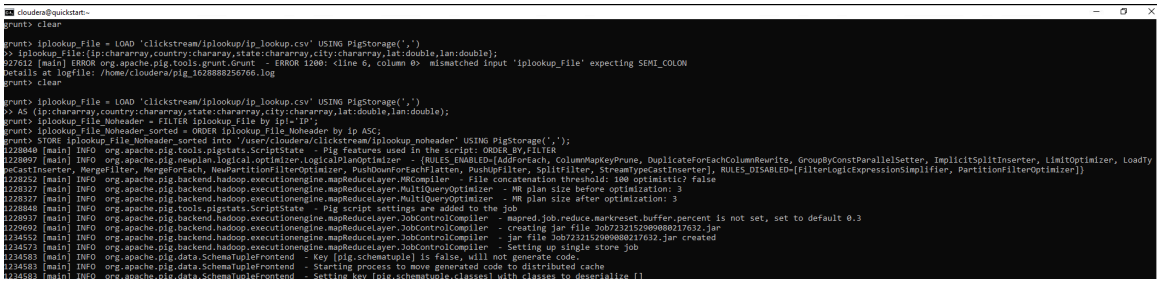
Exercise(s):

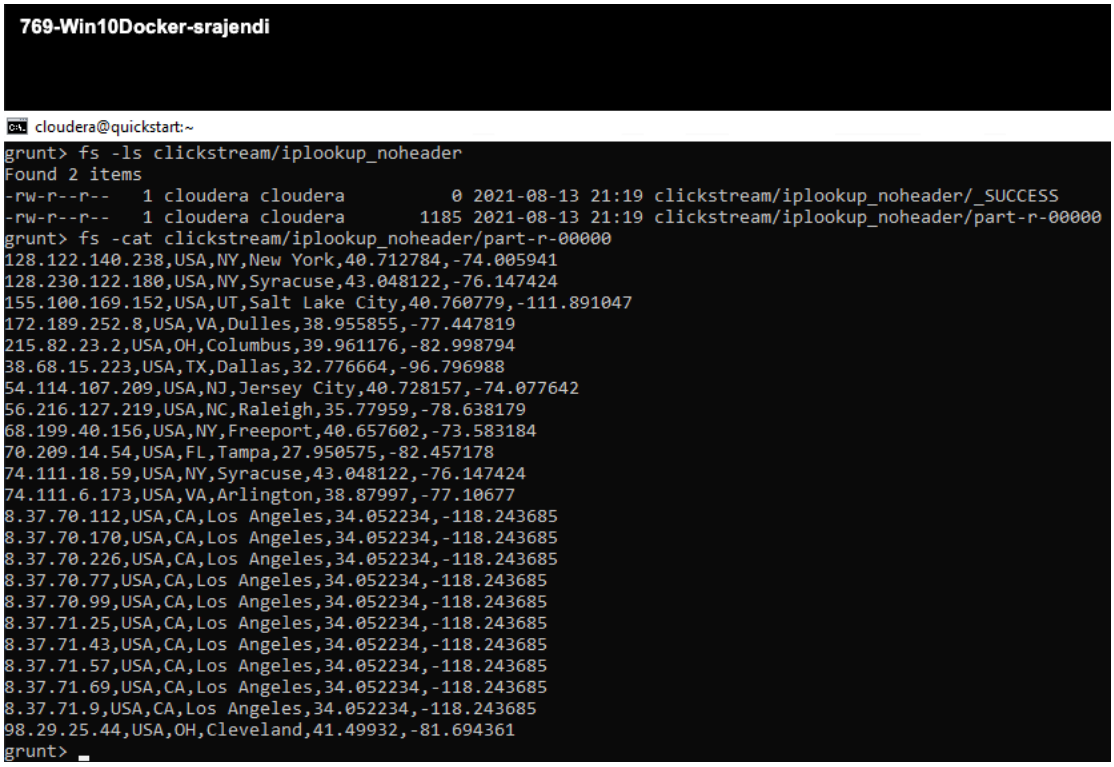
1. Load the comma-delimited HDFS dataset at clickstream/iplookup into a relation with an explicit schema. Use filter logic to remove the first row (which contains a header) then sort the output by IP and dump a comma-delimited data set to clickstream/iplookup_noheader. Record all your Pig commands required to complete your transformation.

Solution:

--list available folders
fs -ls clickstream/iplookup
--view the contents of the file
fs -cat clickstream/iplookup/ip_lookup.csv
--Load the file using PigStorage with explicit schema
iplookup_File = LOAD 'clickstream/iplookup/ip_lookup.csv' USING PigStorage(',') AS (ip:chararray,country:chararray,state:chararray,city:chararray,lat:double,lan:double);
--Remove the header row from the file
iplookup_File_Noheader = FILTER iplookup_File by ip! ='IP';
--Sort the file
iplookup_File_Noheader_sorted = ORDER iplookup_File_Noheader by ip ASC;
--Store the file into HDFS
STORE iplookup_File_Noheader_sorted into '/user/cloudera/clickstream/iplookup_noheader' USING PigStorage(',');
--view the contents of the file
fs -cat clickstream/iplookup_noheader/part-r-00000

Evidence:





2. Write Pig commands to produce a count of IP Addresses by state codes, sorted by the count with highest values first, like this:
   (CA, 10)
   (NY, 4)
   (VA, 2)
   Etc….
   Record all your Pig commands required to complete your transformation.

Solution:

--Group by State

iplookup_by_state = GROUP iplookup_File_Noheader_sorted by state;

--view the schema

describe iplookup_by_state;

--count number of ip's by state

iplookup_by_state_n_counts = FOREACH iplookup_by_state GENERATE group as state, COUNT(iplookup_File_Noheader_sorted.ip) as counts;

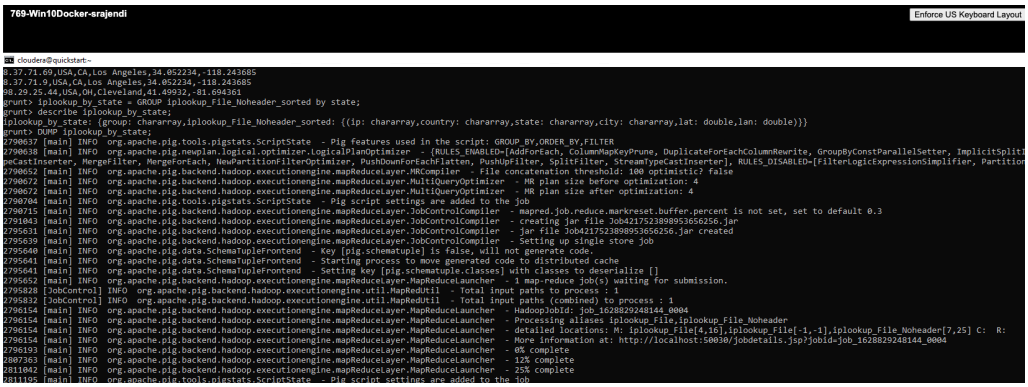
--sort number of ip's by state

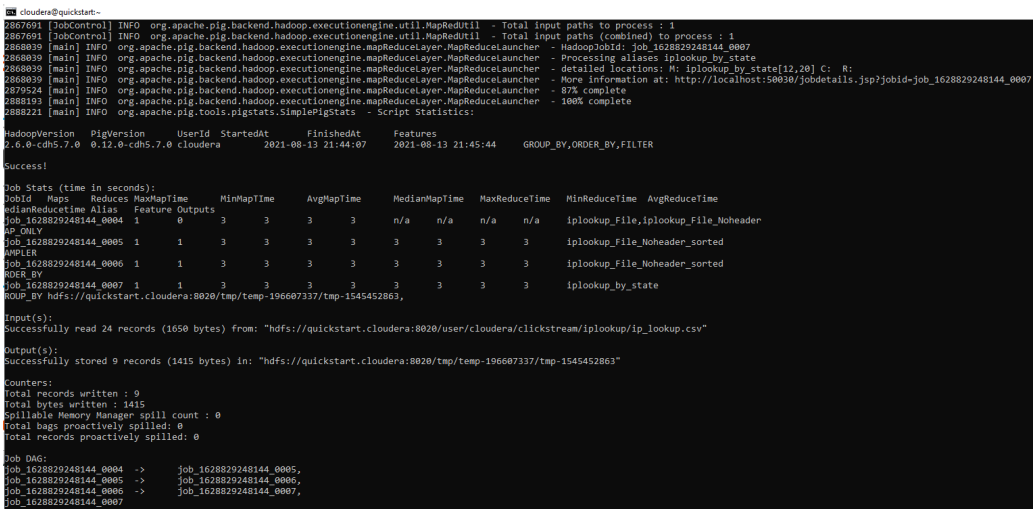
iplookup_by_state_n_counts_sorted = ORDER iplookup_by_state_n_counts DESC;

--view the output

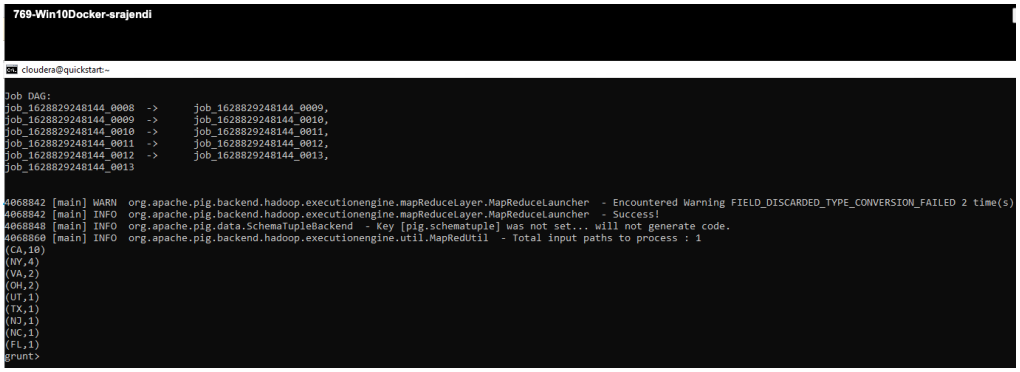
DUMP iplookup_by_state_n_counts_sorted;

Evidence:






```
grunt> describe iplookup_by_state;
iplookup_by_state: {group: chararray,iplookup_File_Noheader_sorted: {(ip: chararray,country: chararray,state: chararray,city: chararray,lat: double,lan: double)}}
grunt> iplookup_by_state_n_counts = FOREACH iplookup_by_state GENERATE group as state,COUNT(iplookup_File_Noheader_sorted.ip) as counts;
grunt> iplookup_by_state_n_counts_sorted = ORDER iplookup_by_state_n_counts by counts DESC;
grunt> DUMP iplookup_by_state_n_counts_sorted;
3899313 [main] INFO  org.apache.pig.tools.pigstats.ScriptState  - Pig features used in the script: GROUP_BY,ORDER_BY,FILTER
3899315 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer  - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstP
zer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[Fi
imizer]}
3899326 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler  - File concatenation threshold: 100 optimistic? false
3899344 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.CombinerOptimizer  - Choosing to move algebraic foreach to combiner
3899350 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer  - MR plan size before optimization: 6
3899351 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer  - MR plan size after optimization: 6
3899383 [main] INFO  org.apache.pig.tools.pigstats.ScriptState  - Pig script settings are added to the job
3899397 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
3899674 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - creating jar file Job7174910676869897026.jar
3904466 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - jar file Job7174910676869897026.jar created
3904480 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - Setting up single store job
3904481 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Key [pig.schematuple] is false, will not generate code.
3904481 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Starting process to move generated code to distributed cache
3904481 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Setting key [pig.schematuple.classes] with classes to deserialize []
3904504 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 1 map-reduce job(s) waiting for submission.
3904648 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths to process : 1
3904653 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths (combined) to process : 1
3905009 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - HadoopJobId: job_1628829248144_0008
3905009 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Processing aliases iplookup_File,iplookup_File_Noheader
3905009 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - detailed locations: M: iplookup_File[4,16],iplookup_File[-1,-1],iplookup_F
3905009 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_16288
3905050 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 0% complete
3916537 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 8% complete
3920213 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 16% complete
3920365 [main] INFO  org.apache.pig.tools.pigstats.ScriptState  - Pig script settings are added to the job
3920376 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
3920378 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - Reduce phase detected, estimating # of required reducers.
3920378 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - Using reducer estimator: org.apache.pig.backend.hadoop.executionengine.map
3920394 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator  - BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=1385
3920394 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - Setting Parallelism to 1
3920688 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - creating jar file Job7020432360865629210.jar
3925294 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - jar file Job7020432360865629210.jar created
3925302 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - Setting up single store job
3925304 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Key [pig.schematuple] is false, will not generate code.
3925304 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Starting process to move generated code to distributed cache
3925304 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Setting key [pig.schematuple.classes] with classes to deserialize []
3925327 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 1 map-reduce job(s) waiting for submission.
3925472 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths to process : 1
3925472 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths (combined) to process : 1
3925830 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - HadoopJobId: job_1628829248144_0009
3925830 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Processing aliases iplookup_File_Noheader_sorted
```




```
Job DAG:
job_1628829248144_0008  ->      job_1628829248144_0009,
job_1628829248144_0009  ->      job_1628829248144_0010,
job_1628829248144_0010  ->      job_1628829248144_0011,
job_1628829248144_0011  ->      job_1628829248144_0012,
job_1628829248144_0012  ->      job_1628829248144_0013,
job_1628829248144_0013


4068842 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 2 time(s)
4068842 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Success!
4068848 [main] INFO  org.apache.pig.data.SchemaTupleBackend  - Key [pig.schematuple] was not set... will not generate code.
4068860 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths to process : 1
(CA,10)
(NY,4)
(VA,2)
(OH,2)
(UT,1)
(TX,1)
(NJ,1)
(NC,1)
(FL,1)
grunt>
```

3. Use pig to load the web log files from **clickstream/logs** using the following schema: reqdate:chararray, reqtime:chararray, x1:int, method:chararray, uri:chararray, x2:int ,x3:int, x4:int ,ipaddress:chararray, useragent:chararray, filter any rows which begin with a "#" (these are header rows and should be removed, then writes out the reqdate, reqtime, method, uri, ipaddress and useragent columns to a tab-delimited data set in HDFS **clickstream/logs_noheader**. HINT: The data is space delimited.

Solution:

--view the log files

fs -ls clickstream/logs

--view the content of log file

fs -cat clickstream/logs/u_ex160211.log

--LOAD the content of log files

clickstream_logs = LOAD 'clickstream/logs/*' USING PigStorage(' ') AS (reqdate:chararray,reqtime:chararray,X1:int,method:chararray,uri:chararray,X2:int,X3:int,X4:int,ipaddress:chararray,useragent:chararray);

--remove the header rows with "#"

clickstream_logs_clean = FILTER clickstream_logs by reqdate! =’#’;

--prepare the final file

clickstream_logs_final = FOREACH clickstream_logs_clean GENERATE reqdate, reqtime, method, uri, ipaddress, useragent;

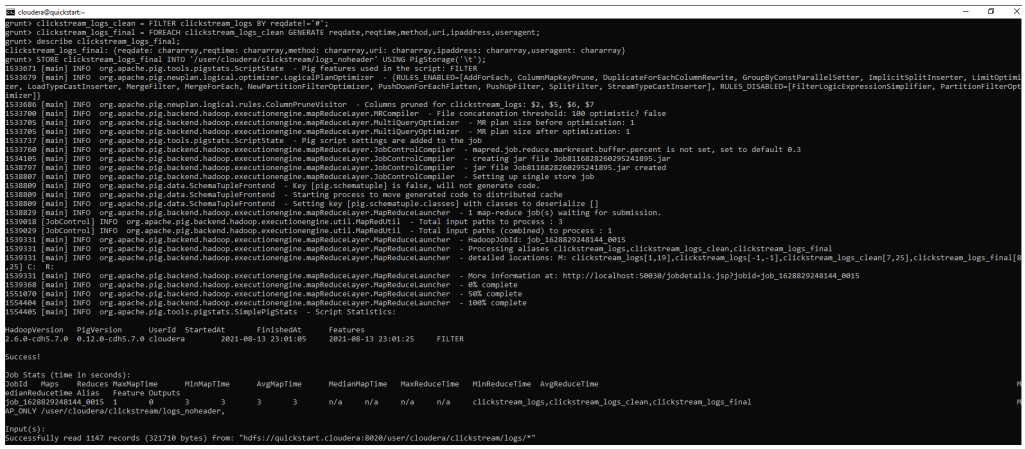
--store the file into HDFS

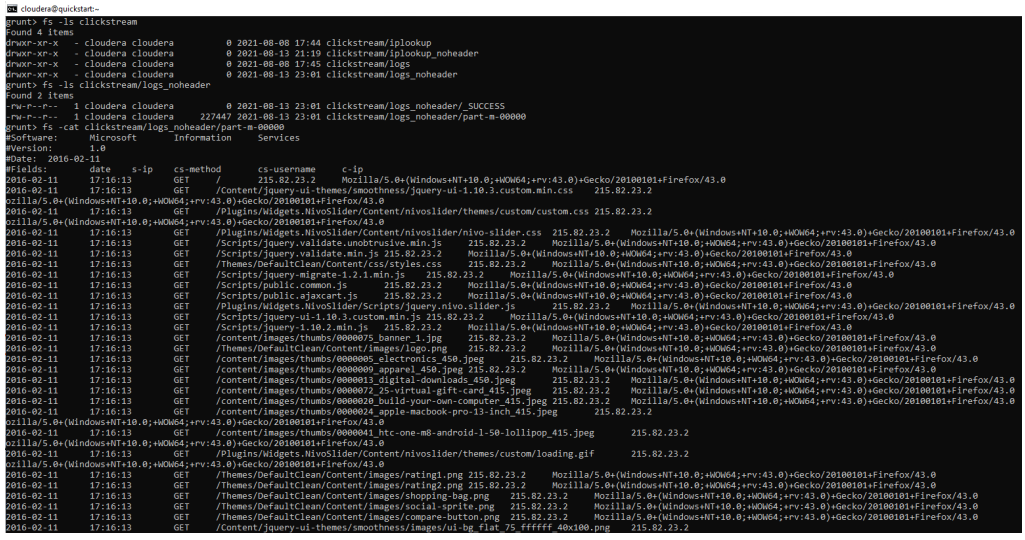
STORE clickstream_logs_final INTO ‘/user/cloudera/clickstream/logs_noheader’ USING PigStorage(‘\t’);

--view the file created

fs -cat clickstream/logs_noheader/part-m-00000

Evidence:





4. Use hive to create two external tables for the **clickstream/logs_noheader** and **clickstream/iplookup_noheader** files you created in the previous steps. These tables should be named **weblogs** and **iplookup** respectively and should be placed in the **clickstream** database. Be sure to record all HQL steps to complete the operations.

Solution:

--view the log files

hdfs dfs -ls clickstream/*header

--connect to Hive

beeline -u jdbc:hive2://localhost:10000/default -u cloudera -p cloudera

--create 'clickstream' database

CREATE database clickstream;

--view the log files on HDFS from Hive

!sh hdfs dfs -ls clickstream/*header

--create external table for 'iplookup'

create external table iplookup (ip string, country string, state string, city string, lat double, lan double) row format delimited fields terminated by ',' location '/user/cloudera/clickstream/iplookup_noheader/';

--create external table for 'weblogs'

create external table weblogs (reqdate date, reqtime timestamp, method string, uri string, ipaddress string, useragent string) row format delimited fields terminated by '\t' location '/user/cloudera/clickstream/logs_noheader/';

--view the hive tables schema

describe iplookup;

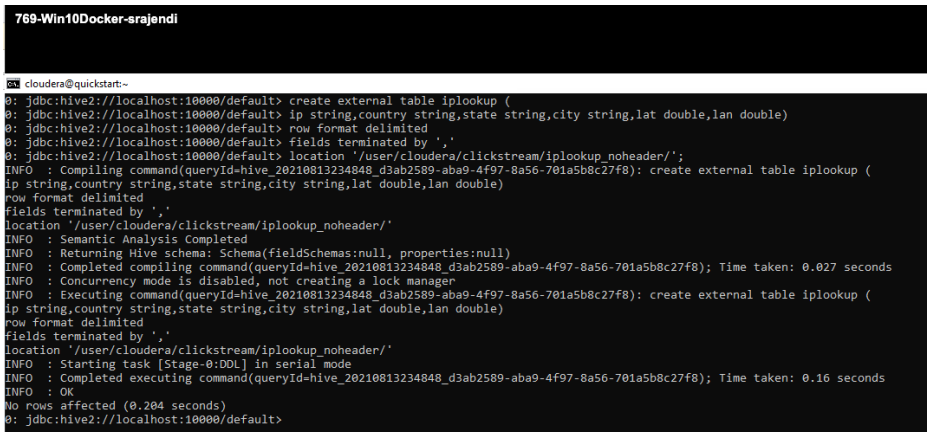
describe weblogs;

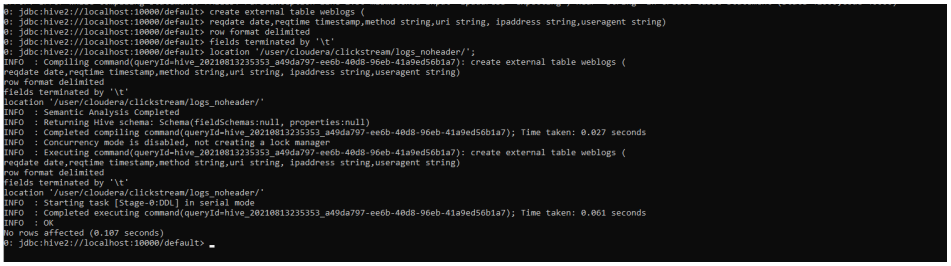
--view the data hive tables

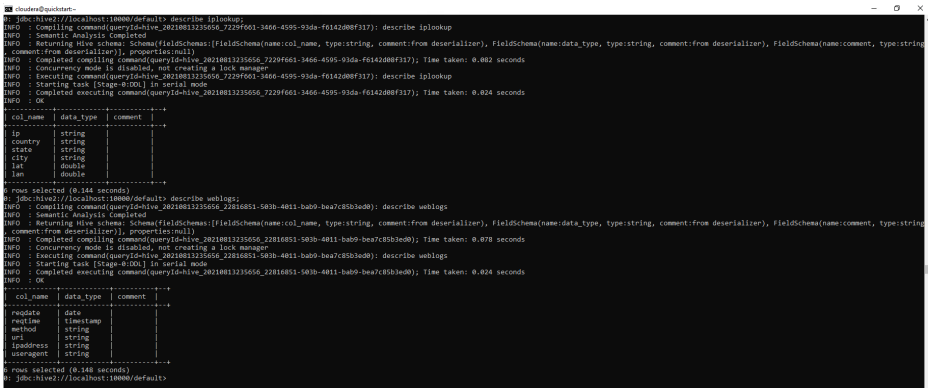
select * from iplookup limit 5;

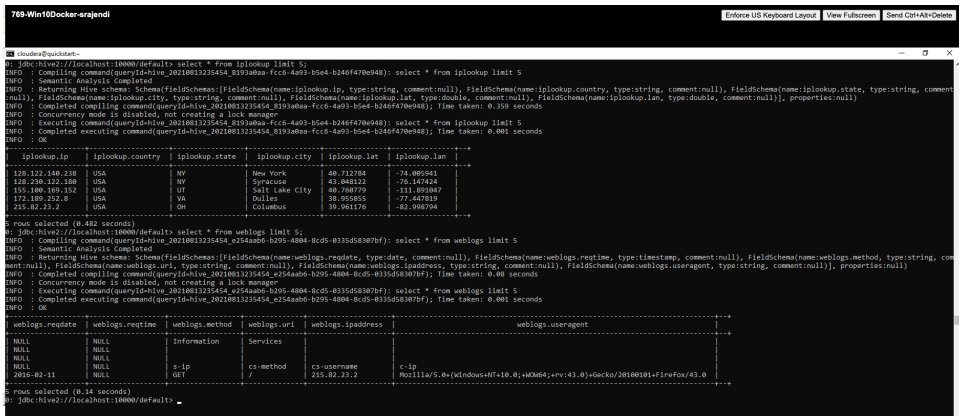
select * from weblogs limit 5;

Evidence:









5. Write an HQL query to display the name of the city and the number of HTTP requests from that city (NOTE: each row in the web logs is an HTTP request). Order the output so cities with the most requests are at the top. If you complete the query correctly, you should see Syracuse has 272-page requests and Los Angeles has 24.

Solution:

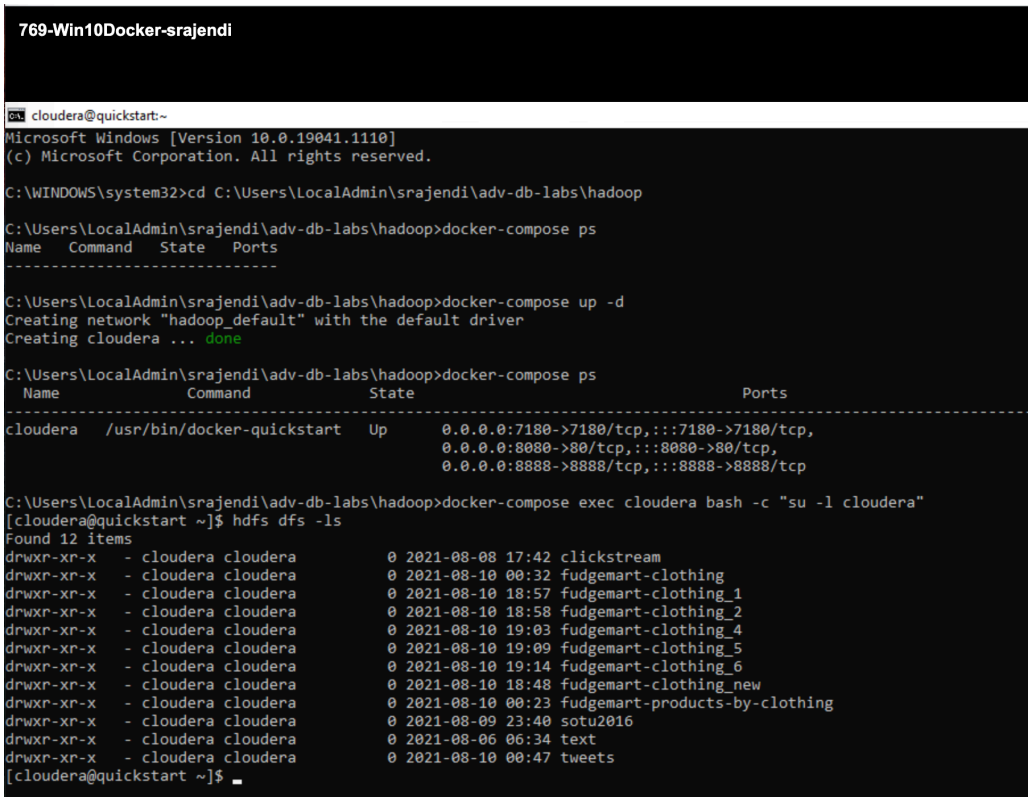
```
--HQL to select cities with most ip requests
select  i.city, count(w.method) as http_requests
from iplookup as i join weblogs as w
on i.ip = w.ipaddress
group by i.city
order by http_requests desc;
```

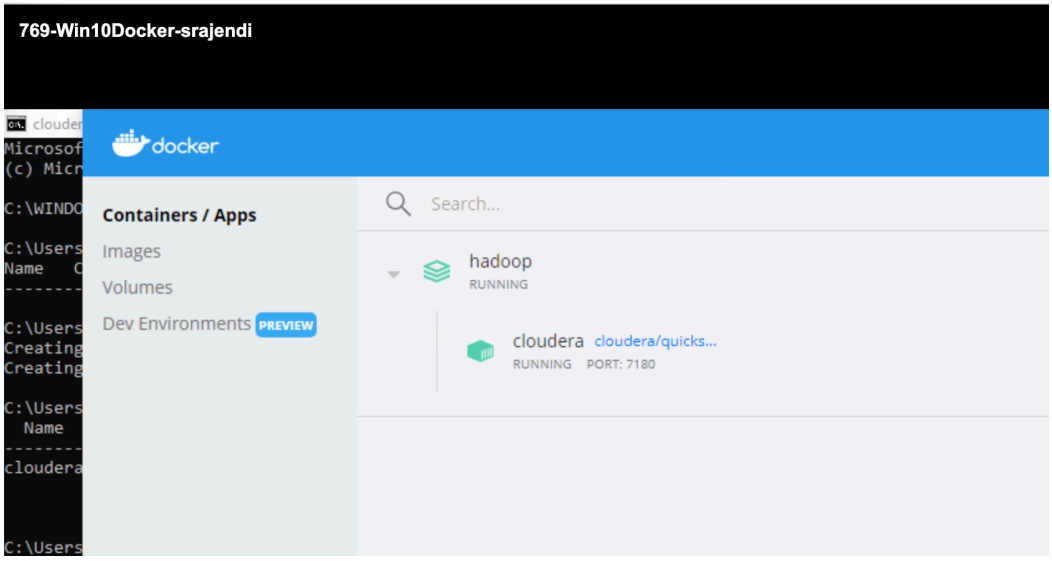
Evidence:

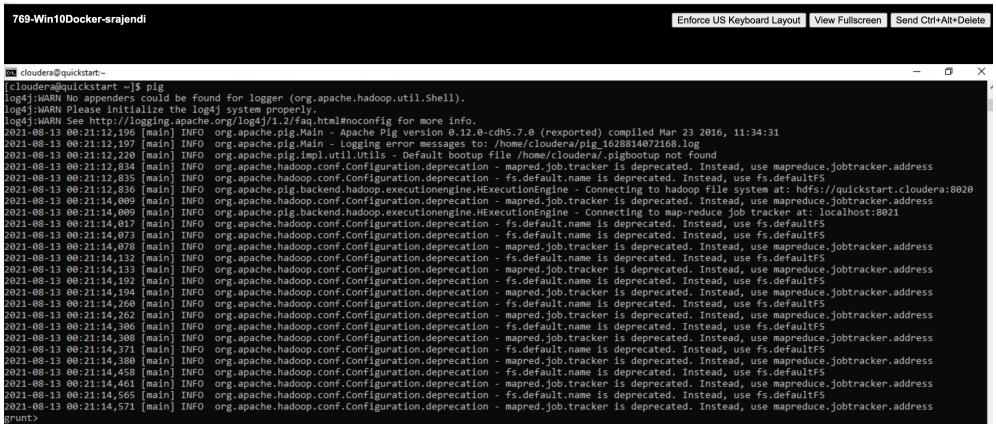
```
769-Win10Docker-srajendi

cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> select
0: jdbc:hive2://localhost:10000/default> i.city,count(w.method) as http_requests
0: jdbc:hive2://localhost:10000/default> from iplookup as i join weblogs as w
0: jdbc:hive2://localhost:10000/default> on i.ip = w.ipaddress
0: jdbc:hive2://localhost:10000/default> group by i.city
0: jdbc:hive2://localhost:10000/default> order by http_requests desc;
```

```
INFO  : MapReduce Jobs Launched:
INFO  : Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 3.62 sec   HDFS Read: 237563 HDFS Write: 456 SUCCESS
INFO  : Stage-Stage-3: Map: 1  Reduce: 1   Cumulative CPU: 2.45 sec   HDFS Read: 5093 HDFS Write: 164 SUCCESS
INFO  : Total MapReduce CPU Time Spent: 6 seconds 70 msec
INFO  : Completed executing command(queryId=hive_20210814000404_239f1460-9b8f-4cbc-a59e-337aa1ebd637); Time taken: 47.451 seconds
INFO  : OK
+----------------+----------------+--+
|     i.city     | http_requests  |
+----------------+----------------+--+
| Syracuse       | 272            |
| Columbus       | 152            |
| Dulles         | 91             |
| Jersey City    | 86             |
| Dallas         | 85             |
| Freeport       | 78             |
| Arlington      | 72             |
| New York       | 68             |
| Raleigh        | 54             |
| Tampa          | 54             |
| Salt Lake City | 51             |
| Cleveland      | 35             |
| Los Angeles    | 24             |
+----------------+----------------+--+
13 rows selected (47.911 seconds)
0: jdbc:hive2://localhost:10000/default>
```

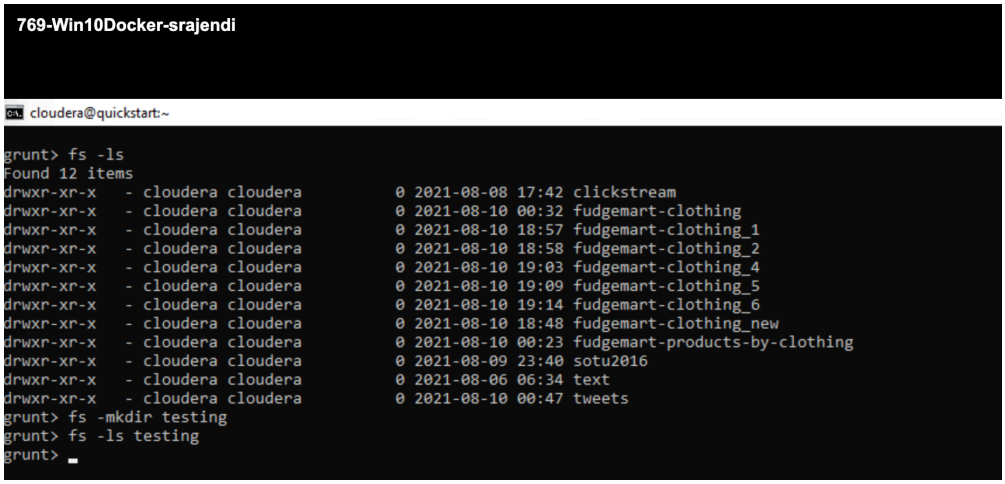
Appendix

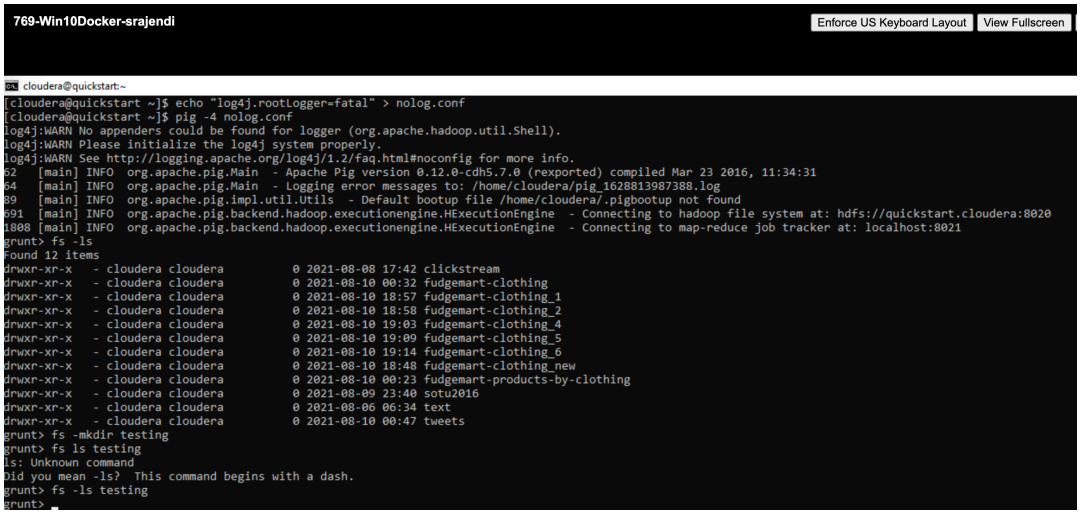

769-Win10Docker-srajendi
cloudera@quickstart:~
Microsoft Windows [Version 10.0.19041.1110]
(c) Microsoft Corporation. All rights reserved.
C:\WINDOWS\system32>cd C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop
C:\Users\LocalAdmin\srajendi\adv-db-labs\hadoop>docker-compose ps
Name Command State Ports
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




```
grunt> quit
[cloudera@quickstart ~]$ echo "log4j.rootLogger=fatal" > nolog.conf
[cloudera@quickstart ~]$ pig -4 nolog.conf
log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
64   [main] INFO  org.apache.pig.Main  - Apache Pig version 0.12.0-cdh5.7.0 (rexported) compiled Mar 23 2016, 11:34:31
66   [main] INFO  org.apache.pig.Main  - Logging error messages to: /home/cloudera/pig_1628814117980.log
89   [main] INFO  org.apache.pig.impl.util.Utils  - Default bootup file /home/cloudera/.pigbootup not found
686  [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine  - Connecting to hadoop file system at: hdfs://quickstart.cloudera:8020
1772 [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine  - Connecting to map-reduce job tracker at: localhost:8021
grunt>
```

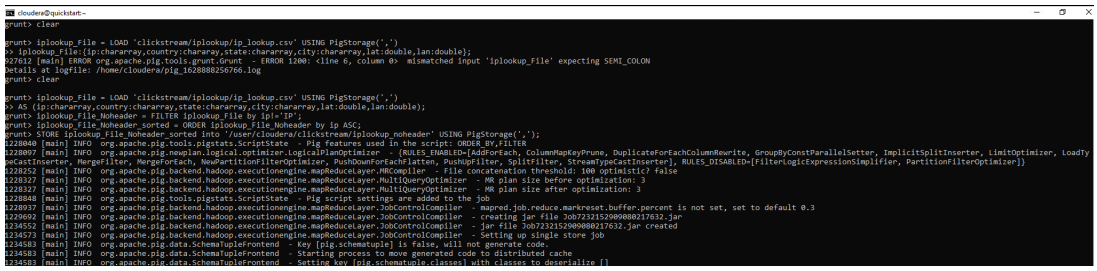




```
cloudera@quickstart:~
[cloudera@quickstart ~]$ clear
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 7 items
drwxr-xr-x   - cloudera cloudera          0 2021-08-08 17:42 clickstream
drwxr-xr-x   - cloudera cloudera          0 2021-08-10 00:32 fudgemart-clothing
drwxr-xr-x   - cloudera cloudera          0 2021-08-10 00:23 fudgemart-products-by-clothing
drwxr-xr-x   - cloudera cloudera          0 2021-08-09 23:40 sotu2016
drwxr-xr-x   - cloudera cloudera          0 2021-08-13 00:23 testing
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 ~]$ hdfs dfs -ls clickstream/iplookup
Found 1 items
-rw-r--r--   1 cloudera cloudera       1251 2021-08-08 17:44 clickstream/iplookup/ip_lookup.csv
[cloudera@quickstart ~]$ hdfs dfs -cat clickstream/iplookup/ip_lookup.csv
IP,Country,State,City,ApproxLat,ApproxLng
172.189.252.8,USA,VA,Dulles,38.955855,-77.447819
215.82.23.2,USA,OH,Columbus,39.961176,-82.998794
98.29.25.44,USA,OH,Cleveland,41.49932,-81.694361
68.199.40.156,USA,NY,Freeport,40.657602,-73.583184
155.100.169.152,USA,UT,Salt Lake City,40.760779,-111.891047
38.68.15.223,USA,TX,Dallas,32.776664,-96.796988
70.209.14.54,USA,FL,Tampa,27.950575,-82.457178
74.111.6.173,USA,VA,Arlington,38.87997,-77.10677
128.230.122.180,USA,NY,Syracuse,43.048122,-76.147424
```

769-Win10Docker-srajendi

```
cloudera@quickstart:~
[cloudera@quickstart ~]$ echo "log4j.rootLogger=fatal" > nolog.conf
[cloudera@quickstart ~]$ pig -4 nolog.conf
log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
85   [main] INFO  org.apache.pig.Main  - Apache Pig version 0.12.0-cdh5.7.0 (rexported) compiled Mar 23 2016, 11:34:31
87   [main] INFO  org.apache.pig.Main  - Logging error messages to: /home/cloudera/pig_1628888256766.log
112  [main] INFO  org.apache.pig.impl.util.Utils  - Default bootup file /home/cloudera/.pigbootup not found
672  [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine  - Connecting to hadoop file system at: hdfs://quickstart.cloudera:8020
1746 [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine  - Connecting to map-reduce job tracker at: localhost:8021
grunt> fs -ls
Found 7 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 00:23 fudgemart-products-by-clothing
drwxr-xr-x   - cloudera cloudera          0 2021-08-09 23:40 sotu2016
drwxr-xr-x   - cloudera cloudera          0 2021-08-13 00:23 testing
drwxr-xr-x   - cloudera cloudera          0 2021-08-06 06:34 text
drwxr-xr-x   - cloudera cloudera          0 2021-08-10 00:47 tweets
grunt> fs -ls clickstream
Found 2 items
drwxr-xr-x   - cloudera cloudera          0 2021-08-08 17:44 clickstream/iplookup
drwxr-xr-x   - cloudera cloudera          0 2021-08-08 17:45 clickstream/logs
grunt> fs -ls clickstream/iplookup
Found 1 items
-rw-r--r--   1 cloudera cloudera       1251 2021-08-08 17:44 clickstream/iplookup/ip_lookup.csv
grunt> fs -cat clickstream/iplookup/ip_lookup.csv
IP,Country,State,City,ApproxLat,ApproxLng
172.189.252.8,USA,VA,Dulles,38.955855,-77.447819
215.82.23.2,USA,OH,Columbus,39.961176,-82.998794
98.29.25.44,USA,OH,Cleveland,41.49932,-81.694361
68.199.40.156,USA,NY,Freeport,40.657602,-73.583184
155.100.169.152,USA,UT,Salt Lake City,40.760779,-111.891047
38.68.15.223,USA,TX,Dallas,32.776664,-96.796988
70.209.14.54,USA,FL,Tampa,27.950575,-82.457178
74.111.6.173,USA,VA,Arlington,38.87997,-77.10677
128.230.122.180,USA,NY,Syracuse,43.048122,-76.147424
128.122.140.238,USA,NY,New York,40.712784,-74.005941
56.216.127.219,USA,NC,Raleigh,35.77959,-78.638179
```



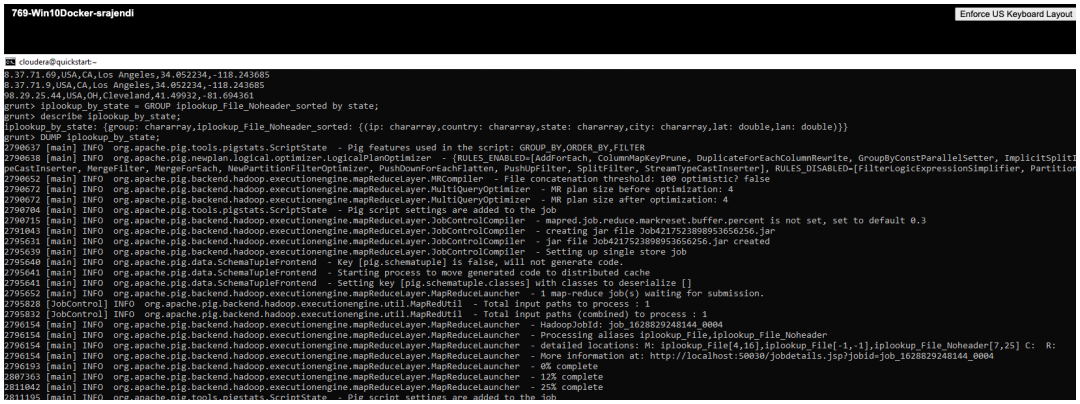
```
cloudera@quickstart:~
1288280 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 1 map-reduce job(s) waiting for submission.
1288442 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths to process : 1
1288445 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths (combined) to process : 1
1288787 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - HadoopJobId: job_1628829248144_0003
1288787 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Processing aliases iplookup_File_Noheader_sorted
1288787 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - detailed locations: M: iplookup_File_Noheader_sorted[8,32] C:  R:
1288787 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_1628829248144_0003
1300610 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 83% complete
1300811 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 100% complete
1300844 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats  - Script Statistics:

HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.6.0-cdh5.7.0  0.12.0-cdh5.7.0 cloudera        2021-08-13 21:18:05     2021-08-13 21:19:25     ORDER_BY,FILTER

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime  AvgReduceTime
edianReducetime Alias   Feature Outputs
job_1628829248144_0001  1       0       3       3       3       n/a     n/a     n/a     n/a     iplookup_File,iplookup_File_Noheader
AP_ONLY
job_1628829248144_0002  1       1       3       3       3       3       3       3       3       iplookup_File_Noheader_sorted
AMPLER
job_1628829248144_0003  1       1       4       4       4       4       3       3       3       3       iplookup_File_Noheader_sorted
RDER_BY /user/cloudera/clickstream/iplookup_noheader,

Input(s):
Successfully read 24 records (1650 bytes) from: "hdfs://quickstart.cloudera:8020/user/cloudera/clickstream/iplookup/ip_lookup.csv"

Output(s):
Successfully stored 23 records (1185 bytes) in: "/user/cloudera/clickstream/iplookup_noheader"
```

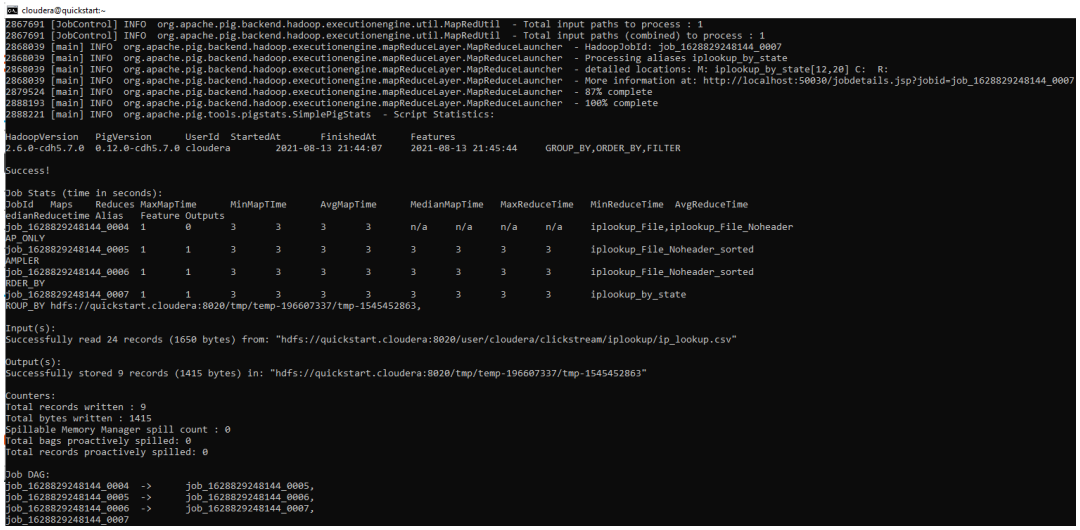
```
769-Win10Docker-srajendi
```

```
cloudera@quickstart:~
Successfully read 24 records (1650 bytes) from: "hdfs://quickstart.cloudera:8020/user/cloudera/clickstream/iplookup/ip_lookup.csv"

Output(s):
Successfully stored 23 records (1185 bytes) in: "/user/cloudera/clickstream/iplookup_noheader"

Counters:
Total records written : 23
Total bytes written : 1185
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1628829248144_0001  ->      job_1628829248144_0002,
job_1628829248144_0002  ->      job_1628829248144_0003,
job_1628829248144_0003


1317161 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 2 time(s).
1317161 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Success!
grunt> fs -ls clickstream
Found 3 items
drwxr-xr-x   - cloudera cloudera          0 2021-08-08 17:44 clickstream/iplookup
drwxr-xr-x   - cloudera cloudera          0 2021-08-13 21:19 clickstream/iplookup_noheader
drwxr-xr-x   - cloudera cloudera          0 2021-08-08 17:45 clickstream/logs
```

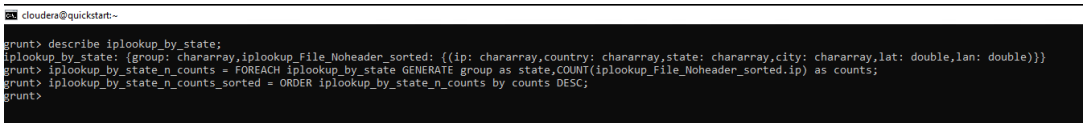
```
769-Win10Docker-srajendi
```

```
cloudera@quickstart:~
grunt> fs -ls clickstream/iplookup_noheader
Found 2 items
-rw-r--r--   1 cloudera cloudera          0 2021-08-13 21:19 clickstream/iplookup_noheader/_SUCCESS
-rw-r--r--   1 cloudera cloudera       1185 2021-08-13 21:19 clickstream/iplookup_noheader/part-r-00000
grunt> fs -cat clickstream/iplookup_noheader/part-r-00000
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
74.111.18.59,USA,NY,Syracuse,43.048122,-76.147424
74.111.6.173,USA,VA,Arlington,38.87997,-77.10677
8.37.70.112,USA,CA,Los Angeles,34.052234,-118.243685
8.37.70.170,USA,CA,Los Angeles,34.052234,-118.243685
8.37.70.226,USA,CA,Los Angeles,34.052234,-118.243685
8.37.70.77,USA,CA,Los Angeles,34.052234,-118.243685
8.37.70.99,USA,CA,Los Angeles,34.052234,-118.243685
8.37.71.25,USA,CA,Los Angeles,34.052234,-118.243685
8.37.71.43,USA,CA,Los Angeles,34.052234,-118.243685
8.37.71.57,USA,CA,Los Angeles,34.052234,-118.243685
8.37.71.69,USA,CA,Los Angeles,34.052234,-118.243685
8.37.71.9,USA,CA,Los Angeles,34.052234,-118.243685
98.29.25.44,USA,OH,Cleveland,41.49932,-81.694361
grunt>
```

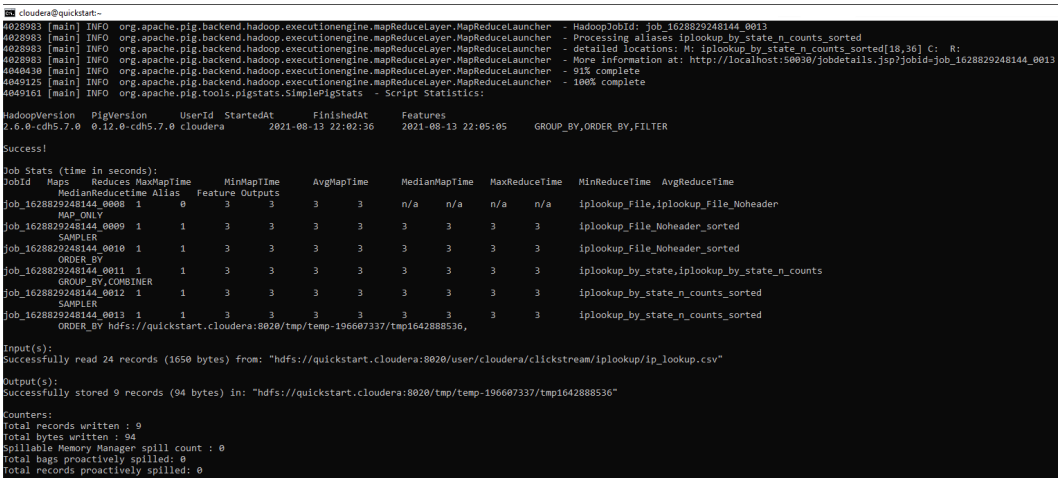

769-Win10Docker-srajendi
Enforce US Keyboard Layout
cloudera@quickstart:~
8.37.71.69,USA,CA,Los Angeles,34.052234,-118.243685
8.37.71.9,USA,CA,Los Angeles,34.052234,-118.243685
98.29.25.44,USA,OH,Cleveland,41.49932,-81.694361
grunt> iplookup_by_state = GROUP iplookup_File_Noheader_sorted by state;
grunt> describe iplookup_by_state;
iplookup_by_state: {group: chararray,iplookup_File_Noheader_sorted: {(ip: chararray,country: chararray,state: chararray,city: chararray,lat: double,lan: double)}}
grunt> DUMP iplookup_by_state;
2790637 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY,ORDER_BY,FILTER
2790638 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitI
nsCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, Partitior
2790652 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2790672 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 4
2790672 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 4
2790704 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2790715 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2791043 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job4217523898953656256.jar
2795631 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job4217523898953656256.jar created
2795639 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store job
2795640 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2795641 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cache
2795641 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Setting key [pig.schematuple.classes] with classes to deserialize []
2795652 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2795828 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2795832 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2796154 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1628829248144_0004
2796154 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases iplookup_File,iplookup_File_Noheader
2796154 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: iplookup_File[4,16],iplookup_File[-1,-1],iplookup_File_Noheader[7,25] C: R:
2796154 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_1628829248144_0004
2796193 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 0% complete
2807363 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 12% complete
2811042 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 25% complete
2811195 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job


cloudera@quickstart:~
2867691 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2867691 [JobControl] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
2868039 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - HadoopJobId: job_1628829248144_0007
2868039 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Processing aliases iplookup_by_state
2868039 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - detailed locations: M: iplookup_by_state[12,20] C: R:
2868039 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_1628829248144_0007
2879524 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 87% complete
2888193 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2888221 [main] INFO org.apache.pig.tools.pigstats.SimplePigStats - Script Statistics:
HadoopVersion PigVersion UserId StartedAt FinishedAt Features
2.6.0-cdh5.7.0 0.12.0-cdh5.7.0 cloudera 2021-08-13 21:44:07 2021-08-13 21:45:44 GROUP_BY,ORDER_BY,FILTER
Success!
Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime
edianReducetime Alias Feature Outputs
job_1628829248144_0004 1 0 3 3 3 3 n/a n/a n/a n/a iplookup_File,iplookup_File_Noheader
AP_ONLY
job_1628829248144_0005 1 1 3 3 3 3 3 3 3 3 iplookup_File_Noheader_sorted
AMPLER
job_1628829248144_0006 1 1 3 3 3 3 3 3 3 3 iplookup_File_Noheader_sorted
RDER_BY
job_1628829248144_0007 1 1 3 3 3 3 3 3 3 3 iplookup_by_state
ROUP_BY hdfs://quickstart.cloudera:8020/tmp/temp-196607337/tmp-1545452863,
Input(s):
Successfully read 24 records (1650 bytes) from: "hdfs://quickstart.cloudera:8020/user/cloudera/clickstream/iplookup/ip_lookup.csv"
Output(s):
Successfully stored 9 records (1415 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-196607337/tmp-1545452863"
Counters:
Total records written : 9
Total bytes written : 1415
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_1628829248144_0004 -> job_1628829248144_0005,
job_1628829248144_0005 -> job_1628829248144_0006,
job_1628829248144_0006 -> job_1628829248144_0007,
job_1628829248144_0007

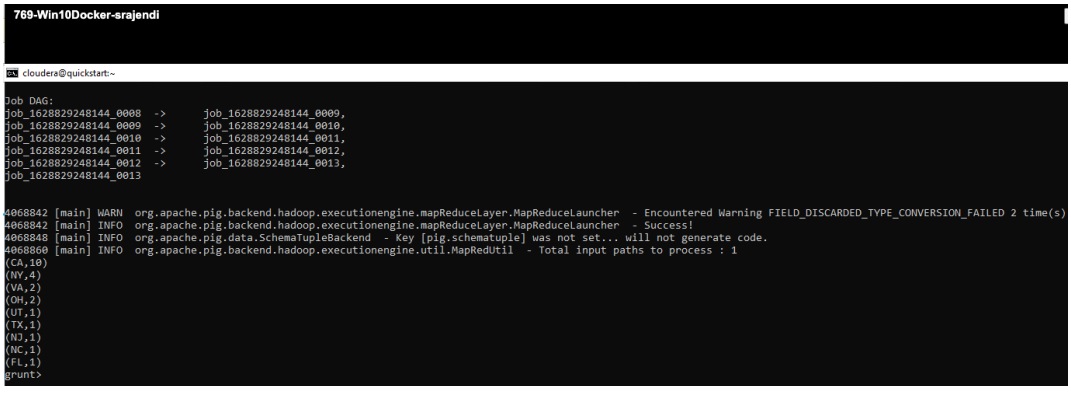

769-Win10Docker-srajendi
Enforce US Keyboard Layout
View Fullscreen
Send Ctrl+Alt+Delete
cloudera@quickstart:~
2898258 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 2 time(s).
2898259 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2898264 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2898278 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
{CA,{(8.37.71.9,USA,CA,Los Angeles,34.052234,-118.243685),(8.37.71.69,USA,CA,Los Angeles,34.052234,-118.243685),(8.37.71.57,USA,CA,Los Angeles,34.052234,-118.243685),(8.37.71.43,USA,CA,Los Angeles,34.052234,-118.243685),(8.37.71.25,USA,CA,Los Angeles,34.052234,-118.243685),(8.37.70.99,USA,CA,Los Angeles,34.052234,-118.243685),(8.37.70.77,USA,CA,Los Angeles,34.052234,-118.243685),(8.37.70.226,USA,CA,Los Angeles,34.052234,-118.243685),(8.37.70.170,USA,CA,Los Angeles,34.052234,-118.243685),(8.37.70.112,USA,CA,Los Angeles,34.052234,-118.243685)}}
{FL,{(70.209.14.54,USA,FL,Tampa,27.950575,-82.457178)}}
{NC,{(56.216.127.219,USA,NC,Raleigh,35.77959,-78.638179)}}
{NJ,{(64.114.107.209,USA,NJ,Jersey City,40.728157,-74.077642)}}
{NY,{(74.111.18.59,USA,NY,Syracuse,43.048122,-76.147424),(68.199.40.156,USA,NY,Freeport,40.657602,-73.583184),(128.122.140.238,USA,NY,New York,40.712784,-74.005941),(128.230.122.180,USA,NY,Syracuse,43.048122,-76.147424)}}
{OH,{(98.29.25.44,USA,OH,Cleveland,41.49932,-81.694361),(215.82.23.2,USA,OH,Columbus,39.961176,-82.998794)}}
{TX,{(38.68.15.223,USA,TX,Dallas,32.776664,-96.796988)}}
{UT,{(155.100.169.152,USA,UT,Salt Lake City,40.760779,-111.891047)}}
{VA,{(172.189.252.8,USA,VA,Dulles,38.955855,-77.447819),(74.111.6.173,USA,VA,Arlington,38.87997,-77.10677)}}
grunt>


cloudera@quickstart:~
grunt> describe iplookup_by_state;
iplookup_by_state: {group: chararray,iplookup_File_Noheader_sorted: {(ip: chararray,country: chararray,state: chararray,city: chararray,lat: double,lan: double)}}
grunt> iplookup_by_state_n_counts = FOREACH iplookup_by_state GENERATE group as state,COUNT(iplookup_File_Noheader_sorted.ip) as counts;
grunt> iplookup_by_state_n_counts_sorted = ORDER iplookup_by_state_n_counts by counts DESC;
grunt>


```
grunt> describe iplookup_by_state;
iplookup_by_state: {group: chararray,iplookup_File_Noheader_sorted: {(ip: chararray,country: chararray,state: chararray,city: chararray,lat: double,lan: double)}}
grunt> iplookup_by_state_n_counts = FOREACH iplookup_by_state GENERATE group as state,COUNT(iplookup_File_Noheader_sorted.ip) as counts;
grunt> iplookup_by_state_n_counts_sorted = ORDER iplookup_by_state_n_counts by counts DESC;
grunt> DUMP iplookup_by_state_n_counts_sorted;
3899313 [main] INFO  org.apache.pig.tools.pigstats.ScriptState  - Pig features used in the script: GROUP_BY,ORDER_BY,FILTER
3899315 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer  - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstP
zer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[Fi
imizer]}
3899326 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler  - File concatenation threshold: 100 optimistic? false
3899344 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.CombinerOptimizer  - Choosing to move algebraic foreach to combiner
3899350 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer  - MR plan size before optimization: 6
3899351 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer  - MR plan size after optimization: 6
3899383 [main] INFO  org.apache.pig.tools.pigstats.ScriptState  - Pig script settings are added to the job
3899397 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
3899674 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - creating jar file Job7174910676869897026.jar
3904466 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - jar file Job7174910676869897026.jar created
3904480 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - Setting up single store job
3904481 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Key [pig.schematuple] is false, will not generate code.
3904481 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Starting process to move generated code to distributed cache
3904481 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Setting key [pig.schematuple.classes] with classes to deserialize []
3904504 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 1 map-reduce job(s) waiting for submission.
3904648 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths to process : 1
3904653 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths (combined) to process : 1
3905009 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - HadoopJobId: job_1628829248144_0008
3905009 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Processing aliases iplookup_File,iplookup_File_Noheader
3905009 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - detailed locations: M: iplookup_File[4,16],iplookup_File[-1,-1],iplookup_F
3905009 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_16288
3905050 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 0% complete
3916537 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 8% complete
3920213 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 16% complete
3920365 [main] INFO  org.apache.pig.tools.pigstats.ScriptState  - Pig script settings are added to the job
3920376 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
3920378 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - Reduce phase detected, estimating # of required reducers.
3920378 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - Using reducer estimator: org.apache.pig.backend.hadoop.executionengine.ma
3920394 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.InputSizeReducerEstimator  - BytesPerReducer=1000000000 maxReducers=999 totalInputFileSize=1385
3920394 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - Setting Parallelism to 1
3920688 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - creating jar file Job7020432360865629210.jar
3925294 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - jar file Job7020432360865629210.jar created
3925302 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - Setting up single store job
3925304 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Key [pig.schematuple] is false, will not generate code.
3925304 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Starting process to move generated code to distributed cache
3925304 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Setting key [pig.schematuple.classes] with classes to deserialize []
3925327 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 1 map-reduce job(s) waiting for submission.
3925472 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths to process : 1
3925472 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths (combined) to process : 1
3925830 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - HadoopJobId: job_1628829248144_0009
3925830 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Processing aliases iplookup_File_Noheader_sorted
```




```
4028983 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - HadoopJobId: job_1628829248144_0013
4028983 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Processing aliases iplookup_by_state_n_counts_sorted
4028983 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - detailed locations: M: iplookup_by_state_n_counts_sorted[18,36] C:  R:
4028983 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_1628829248144_0013
4040430 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 91% complete
4049125 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 100% complete
4049161 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats  - Script Statistics:

HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.6.0-cdh5.7.0  0.12.0-cdh5.7.0 cloudera        2021-08-13 22:02:36     2021-08-13 22:05:05     GROUP_BY,ORDER_BY,FILTER

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime  AvgReduceTime
        MedianReducetime Alias  Feature Outputs
job_1628829248144_0008  1       0       3       3       3       n/a     n/a     n/a     n/a     iplookup_File,iplookup_File_Noheader
        MAP_ONLY
job_1628829248144_0009  1       1       3       3       3       3       3       3       3       iplookup_File_Noheader_sorted
        SAMPLER
job_1628829248144_0010  1       1       3       3       3       3       3       3       3       iplookup_File_Noheader_sorted
        ORDER_BY
job_1628829248144_0011  1       1       3       3       3       3       3       3       3       iplookup_by_state,iplookup_by_state_n_counts
        GROUP_BY,COMBINER
job_1628829248144_0012  1       1       3       3       3       3       3       3       3       iplookup_by_state_n_counts_sorted
        SAMPLER
job_1628829248144_0013  1       1       3       3       3       3       3       3       3       iplookup_by_state_n_counts_sorted
        ORDER_BY hdfs://quickstart.cloudera:8020/tmp/temp-196607337/tmp1642888536,

Input(s):
Successfully read 24 records (1650 bytes) from: "hdfs://quickstart.cloudera:8020/user/cloudera/clickstream/iplookup/ip_lookup.csv"

Output(s):
Successfully stored 9 records (94 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp-196607337/tmp1642888536"

Counters:
Total records written : 9
Total bytes written : 94
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
```

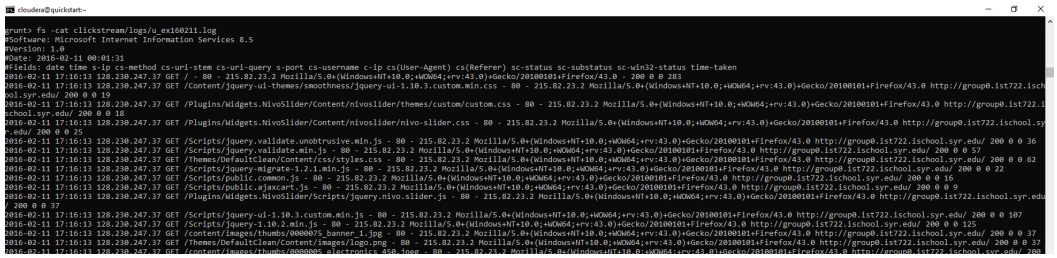



```
Job DAG:
job_1628829248144_0008  ->      job_1628829248144_0009,
job_1628829248144_0009  ->      job_1628829248144_0010,
job_1628829248144_0010  ->      job_1628829248144_0011,
job_1628829248144_0011  ->      job_1628829248144_0012,
job_1628829248144_0012  ->      job_1628829248144_0013,
job_1628829248144_0013


4068842 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 2 time(s)
4068842 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Success!
4068848 [main] INFO  org.apache.pig.data.SchemaTupleBackend  - Key [pig.schematuple] was not set... will not generate code.
4068860 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths to process : 1
(CA,10)
(NY,4)
(VA,2)
(OH,2)
(UT,1)
(TX,1)
(NJ,1)
(NC,1)
(FL,1)
grunt>
```

769-Win10Docker-srajendi

cloudera@quickstart:~

```
[cloudera@quickstart ~]$ fs -ls
-bash: fs: command not found
[cloudera@quickstart ~]$ pig -4 nolog.conf
log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
97   [main] INFO  org.apache.pig.Main  - Apache Pig version 0.12.0-cdh5.7.0 (rexported) compiled Mar 23 2016, 11:34:31
98   [main] INFO  org.apache.pig.Main  - Logging error messages to: /home/cloudera/pig_1628894131487.log
133  [main] INFO  org.apache.pig.impl.util.Utils  - Default bootup file /home/cloudera/.pigbootup not found
760  [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine  - Connecting to hadoop file system at: hdfs://quickstart.cloudera:8020
1859 [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine  - Connecting to map-reduce job tracker at: localhost:8021
grunt> fs -ls
Found 7 items
drwxr-xr-x   - cloudera cloudera          0 2021-08-13 21:19 clickstream
drwxr-xr-x   - cloudera cloudera          0 2021-08-10 00:32 fudgemart-clothing
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 2021-08-13 00:23 testing
drwxr-xr-x   - cloudera cloudera          0 2021-08-06 06:34 text
drwxr-xr-x   - cloudera cloudera          0 2021-08-10 00:47 tweets
grunt> fs -ls clickstream
Found 3 items
drwxr-xr-x   - cloudera cloudera          0 2021-08-08 17:44 clickstream/iplookup
drwxr-xr-x   - cloudera cloudera          0 2021-08-13 21:19 clickstream/iplookup_noheader
drwxr-xr-x   - cloudera cloudera          0 2021-08-08 17:45 clickstream/logs
grunt> fs -ls clickstream/logs
Found 3 items
-rw-r--r--   1 cloudera cloudera     137233 2021-08-08 17:45 clickstream/logs/u_ex160211.log
-rw-r--r--   1 cloudera cloudera      78658 2021-08-08 17:45 clickstream/logs/u_ex160212.log
-rw-r--r--   1 cloudera cloudera     105235 2021-08-08 17:45 clickstream/logs/u_ex160213.log
grunt>
```



769-Win10Docker-srajendi

cloudera@quickstart:~

```
grunt> clickstream_logs = LOAD 'clickstream/logs/*' USING PigStorage(' ')
>> AS (reqdate:chararray,reqtime:chararray,X1:int,method:chararray,uri:chararray,X2:int,X3:int,X4:int,ipaddress:chararray,useragent:chararray);
grunt> describe clickstream_logs;
clickstream_logs: {reqdate: chararray,reqtime: chararray,X1: int,method: chararray,uri: chararray,X2: int,X3: int,X4: int,ipaddress: chararray,useragent: chararray}
grunt> DUMP clickstream_logs;
```

cloudera@quickstart:~

```
grunt> clickstream_logs = LOAD 'clickstream/logs/*' USING PigStorage(' ')
>> AS (reqdate:chararray,reqtime:chararray,X1:int,method:chararray,uri:chararray,X2:int,X3:int,X4:int,ipaddress:chararray,useragent:chararray);
grunt> describe clickstream_logs;
clickstream_logs: {reqdate: chararray,reqtime: chararray,X1: int,method: chararray,uri: chararray,X2: int,X3: int,X4: int,ipaddress: chararray,useragent: chararray}
grunt> DUMP clickstream_logs;
648244 [main] INFO  org.apache.pig.tools.pigstats.ScriptState  - Pig features used in the script: UNKNOWN
648303 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer  - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter,
er, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpres
mizer]}
648455 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler  - File concatenation threshold: 100 optimistic? false
648488 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer  - MR plan size before optimization: 1
648488 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer  - MR plan size after optimization: 1
648854 [main] INFO  org.apache.pig.tools.pigstats.ScriptState  - Pig script settings are added to the job
648908 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
649409 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - creating jar file Job5334611041575104.jar
654269 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - jar file Job5334611041575104.jar created
654292 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - Setting up single store job
654303 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Key [pig.schematuple] is false, will not generate code.
654303 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Starting process to move generated code to distributed cache
654303 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Setting key [pig.schematuple.classes] with classes to deserialize []
654351 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 1 map-reduce job(s) waiting for submission.
654788 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths to process : 3
654819 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths (combined) to process : 1
655395 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - HadoopJobId: job_1628829248144_0014
655395 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Processing aliases clickstream_logs
655395 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - detailed locations: M: clickstream_logs[1,19],clickstream_logs[-1,-1] C:  R:
655395 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_1628829248144_0014
655449 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 0% complete
666617 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 50% complete
671099 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 100% complete
671101 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats  - Script Statistics:

HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.6.0-cdh5.7.0  0.12.0-cdh5.7.0 cloudera        2021-08-13 22:46:20     2021-08-13 22:46:42     UNKNOWN

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime  AvgReduceTime
edianReducetime Alias   Feature Outputs
job_1628829248144_0014  1       0       3       3       3       n/a     n/a     n/a     n/a     clickstream_logs
AP_ONLY hdfs://quickstart.cloudera:8020/tmp/temp1444246572/tmp677452236,

Input(s):
Successfully read 1147 records (321710 bytes) from: "hdfs://quickstart.cloudera:8020/user/cloudera/clickstream/logs/*"

Output(s):
Successfully stored 1147 records (252612 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp1444246572/tmp677452236"
```

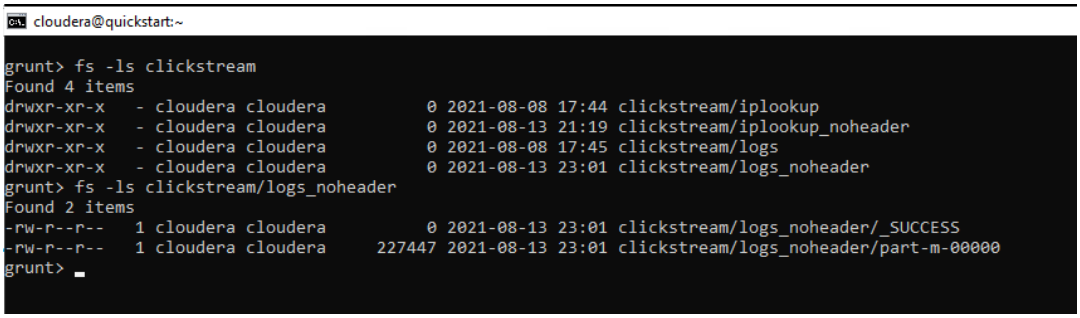
```
cloudera@quickstart:~
Successfully stored 1147 records (252612 bytes) in: "hdfs://quickstart.cloudera:8020/tmp/temp1444246572/tmp677452236"

Counters:
Total records written : 1147
Total bytes written : 252612
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1628829248144_0014


671197 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Encountered Warning ACCESSING_NON_EXISTENT_FIELD 57 time(s).
671197 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 3423 time(s).
671197 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Success!
671202 [main] INFO  org.apache.pig.data.SchemaTupleBackend  - Key [pig.schematuple] was not set... will not generate code.
671220 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths to process : 1
(#Software:,Microsoft,,Information,Services,8,,,,)
(#Version:,1.0,,,,,,,,)
(#Date:,2016-02-11,,,,,,,,)
(#Fields:,date,,s-ip,cs-method,,,,cs-username,c-ip)
(2016-02-11,17:16:13,,GET,/,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/Content/jquery-ui-themes/smoothness/jquery-ui-1.10.3.custom.min.css,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/Plugins/Widgets.NivoSlider/Content/nivoslider/themes/custom/custom.css,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/Plugins/Widgets.NivoSlider/Content/nivoslider/nivo-slider.css,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/Scripts/jquery.validate.unobtrusive.min.js,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/Scripts/jquery.validate.min.js,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/Themes/DefaultClean/Content/css/styles.css,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/Scripts/jquery-migrate-1.2.1.min.js,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/Scripts/public.common.js,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/Scripts/public.ajaxcart.js,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/Plugins/Widgets.NivoSlider/Scripts/jquery.nivo.slider.js,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/Scripts/jquery-ui-1.10.3.custom.min.js,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/Scripts/jquery-1.10.2.min.js,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/content/images/thumbs/0000075_banner_1.jpg,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/Themes/DefaultClean/Content/images/logo.png,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/content/images/thumbs/0000005_electronics_450.jpeg,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/content/images/thumbs/0000009_apparel_450.jpeg,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/content/images/thumbs/0000013_digital-downloads_450.jpeg,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/content/images/thumbs/0000072_25-virtual-gift-card_415.jpeg,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/content/images/thumbs/0000020_build-your-own-computer_415.jpeg,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/content/images/thumbs/0000024_apple-macbook-pro-13-inch_415.jpeg,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/content/images/thumbs/0000041_htc-one-m8-android-l-50-lollipop_415.jpeg,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/Plugins/Widgets.NivoSlider/Content/nivoslider/themes/custom/loading.gif,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
(2016-02-11,17:16:13,,GET,/Themes/DefaultClean/Content/images/ratings.png,,80,,215.82.23.2,Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0)
```

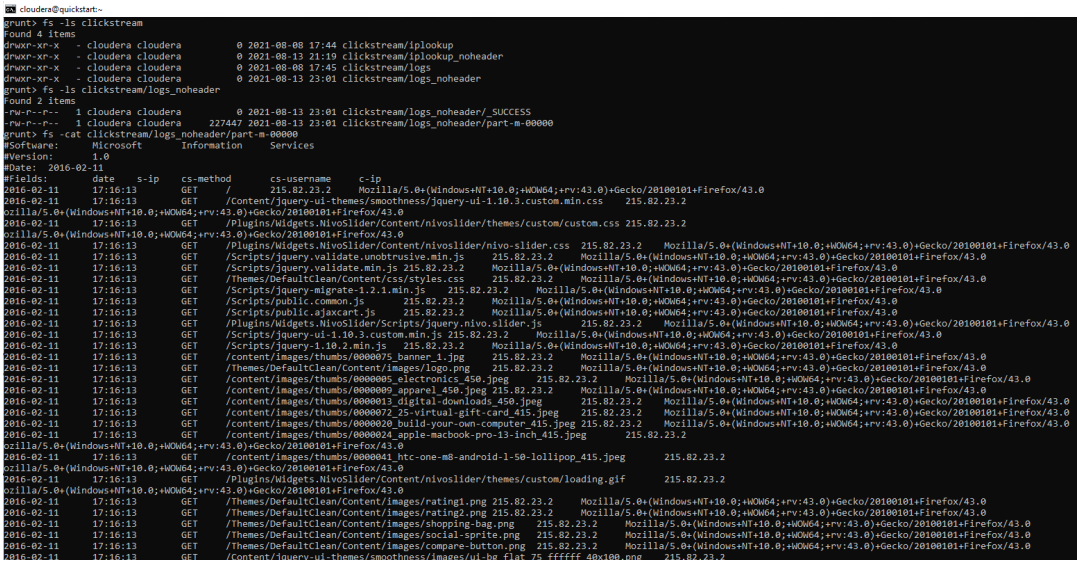
769-Win10Docker-srajendi

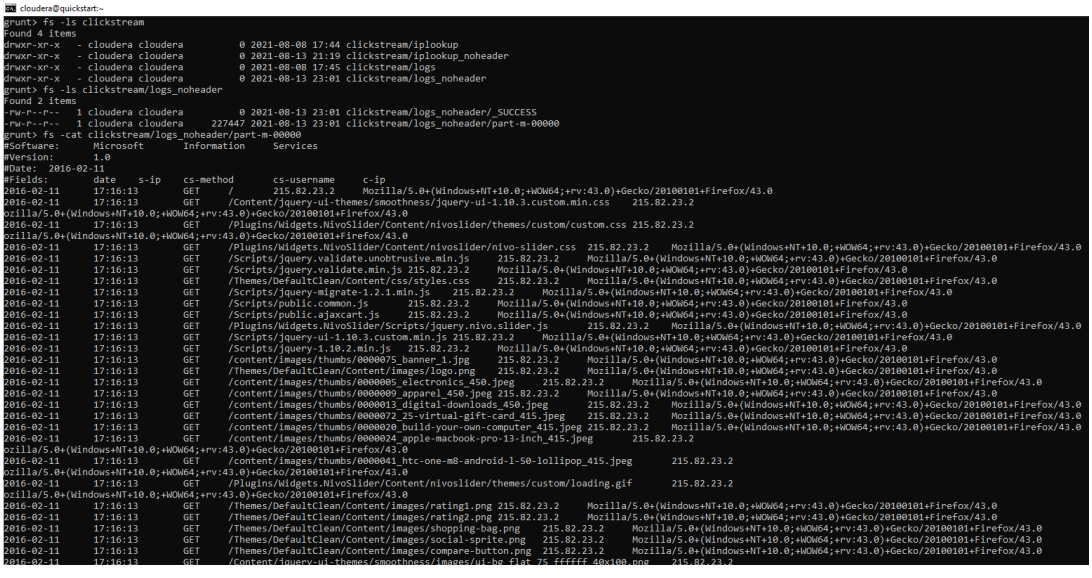
```
cloudera@quickstart:~

grunt> clickstream_logs_clean = FILTER clickstream_logs BY reqdate!='#';
grunt> clickstream_logs_final = FOREACH clickstream_logs_clean GENERATE reqdate,reqtime,method,uri,ipaddress,useragent;
grunt> describe clickstream_logs_final;
clickstream_logs_final: {reqdate: chararray,reqtime: chararray,method: chararray,uri: chararray,ipaddress: chararray,useragent: chararray}
grunt>
```

```
cloudera@quickstart:~
grunt> clickstream_logs_clean = FILTER clickstream_logs BY reqdate!='#';
grunt> clickstream_logs_final = FOREACH clickstream_logs_clean GENERATE reqdate,reqtime,method,uri,ipaddress,useragent;
grunt> describe clickstream_logs_final;
clickstream_logs_final: {reqdate: chararray,reqtime: chararray,method: chararray,uri: chararray,ipaddress: chararray,useragent: chararray}
grunt> STORE clickstream_logs_final INTO '/user/cloudera/clickstream/logs_noheader' USING PigStorage('\t');
1533671 [main] INFO  org.apache.pig.tools.pigstats.ScriptState  - Pig features used in the script: FILTER
1533676 [main] INFO  org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer  - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, DuplicateForEachColumnRewrite, GroupByConstParallelSetter, ImplicitSplitInserter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NewPartitionFilterOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter], RULES_DISABLED=[FilterLogicExpressionSimplifier, PartitionFilterOptimizer]}
1533680 [main] INFO  org.apache.pig.newplan.logical.rules.ColumnPruneVisitor  - Columns pruned for clickstream_logs: $2, $5, $6, $7
1533700 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler  - File concatenation threshold: 100 optimistic? false
1533705 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer  - MR plan size before optimization: 1
1533705 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer  - MR plan size after optimization: 1
1533737 [main] INFO  org.apache.pig.tools.pigstats.ScriptState  - Pig script settings are added to the job
1533760 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
1534105 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - creating jar file Job8116822602952241895.jar
1538797 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - jar file Job8116822602952241895.jar created
1538807 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler  - Setting up single store job
1538809 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Key [pig.schematuple] is false, will not generate code.
1538809 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Starting process to move generated code to distributed cache
1538809 [main] INFO  org.apache.pig.data.SchemaTupleFrontend  - Setting key [pig.schematuple.classes] with classes to deserialize []
1538829 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 1 map-reduce job(s) waiting for submission.
1539018 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths to process : 3
1539029 [JobControl] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths (combined) to process : 1
1539331 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - HadoopJobId: job_1628829248144_0015
1539331 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Processing aliases clickstream_logs,clickstream_logs_clean,clickstream_logs_final
1539331 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - detailed locations: M: clickstream_logs[1,10],clickstream_logs[-1,-1],clickstream_logs_clean[7,25],clickstream_logs_final[8,25] C:  R:
1539331 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - More information at: http://localhost:50030/jobdetails.jsp?jobid=job_1628829248144_0015
1539368 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 0% complete
1551070 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 50% complete
1554404 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - 100% complete
1554405 [main] INFO  org.apache.pig.tools.pigstats.SimplePigStats  - Script Statistics:

HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
2.6.0-cdh5.7.0  0.12.0-cdh5.7.0 cloudera        2021-08-13 23:01:05     2021-08-13 23:01:25     FILTER

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime                                                                   M
edianReduceTime Alias   Feature Outputs
job_1628829248144_0015  1       0       3       3       3       3       n/a     n/a     n/a     n/a     clickstream_logs,clickstream_logs_clean,clickstream_logs_final                   M
AP_ONLY /user/cloudera/clickstream/logs_noheader,

Input(s):
Successfully read 1147 records (321710 bytes) from: "hdfs://quickstart.cloudera:8020/user/cloudera/clickstream/logs/*"
```

```
cloudera@quickstart:~

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime
edianReduceTime Alias   Feature Outputs
job_1628829248144_0015  1       0       3       3       3       3       n/a     n/a     n/a     n/a     clickstream_logs,clickstream_logs_clean,clickstream_logs_final
AP_ONLY /user/cloudera/clickstream/logs_noheader,

Input(s):
Successfully read 1147 records (321710 bytes) from: "hdfs://quickstart.cloudera:8020/user/cloudera/clickstream/logs/*"

Output(s):
Successfully stored 1147 records (227447 bytes) in: "/user/cloudera/clickstream/logs_noheader"

Counters:
Total records written : 1147
Total bytes written : 227447
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1628829248144_0015


1554514 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher  - Success!
grunt>
```


cloudera@quickstart:~
grunt> fs -ls clickstream
Found 4 items
drwxr-xr-x   - cloudera cloudera   0 2021-08-08 17:44 clickstream/iplookup
drwxr-xr-x   - cloudera cloudera   0 2021-08-13 21:19 clickstream/iplookup_noheader
drwxr-xr-x   - cloudera cloudera   0 2021-08-08 17:45 clickstream/logs
drwxr-xr-x   - cloudera cloudera   0 2021-08-13 23:01 clickstream/logs_noheader
grunt> fs -ls clickstream/logs_noheader
Found 2 items
-rw-r--r--   1 cloudera cloudera   0 2021-08-13 23:01 clickstream/logs_noheader/_SUCCESS
-rw-r--r--   1 cloudera cloudera   227447 2021-08-13 23:01 clickstream/logs_noheader/part-m-00000
grunt>


cloudera@quickstart:~
grunt> fs -ls clickstream
Found 4 items
drwxr-xr-x   - cloudera cloudera   0 2021-08-08 17:44 clickstream/iplookup
drwxr-xr-x   - cloudera cloudera   0 2021-08-13 21:19 clickstream/iplookup_noheader
drwxr-xr-x   - cloudera cloudera   0 2021-08-08 17:45 clickstream/logs
drwxr-xr-x   - cloudera cloudera   0 2021-08-13 23:01 clickstream/logs_noheader
grunt> fs -ls clickstream/logs_noheader
Found 2 items
-rw-r--r--   1 cloudera cloudera   0 2021-08-13 23:01 clickstream/logs_noheader/_SUCCESS
-rw-r--r--   1 cloudera cloudera   227447 2021-08-13 23:01 clickstream/logs_noheader/part-m-00000
grunt> fs -cat clickstream/logs_noheader/part-m-00000
#Software: Microsoft   Information   Services
#Version: 1.0
#Date: 2016-02-11
#Fields: date   s-ip   cs-method   cs-username   c-ip
2016-02-11   17:16:13   GET   /   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Content/jquery-ui-themes/smoothness/jquery-ui-1.10.3.custom.min.css   215.82.23.2
ozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Plugins/Widgets.NivoSlider/Content/nivoslider/themes/custom/custom.css 215.82.23.2
ozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Plugins/Widgets.NivoSlider/Content/nivoslider/nivo-slider.css 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Scripts/jquery.validate.unobtrusive.min.js   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Scripts/jquery.validate.min.js 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Themes/DefaultClean/Content/css/styles.css   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Scripts/jquery-migrate-1.2.1.min.js   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Scripts/public.common.js   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Scripts/public.ajaxcart.js   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Plugins/Widgets.NivoSlider/Scripts/jquery.nivo.slider.js   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Scripts/jquery-ui-1.10.3.custom.min.js 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Scripts/jquery-1.10.2.min.js   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000075_banner_1.jpg   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Themes/DefaultClean/Content/images/logo.png   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000005_electronics_450.jpeg   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000009_apparel_450.jpeg 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000013_digital-downloads_450.jpeg   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000072_25-virtual-gift-card_415.jpeg   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000020_build-your-own-computer_415.jpeg 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000024_apple-macbook-pro-13-inch_415.jpeg   215.82.23.2
ozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000041_htc-one-m8-android-l-50-lollipop_415.jpeg   215.82.23.2
ozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Plugins/Widgets.NivoSlider/Content/nivoslider/themes/custom/loading.gif   215.82.23.2
ozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Themes/DefaultClean/Content/images/rating1.png 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Themes/DefaultClean/Content/images/rating2.png 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Themes/DefaultClean/Content/images/shopping-bag.png   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Themes/DefaultClean/Content/images/social-sprite.png   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Themes/DefaultClean/Content/images/compare-button.png 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Content/jquery-ui-themes/smoothness/images/ui-bg_flat_75_ffffff_40x100.png   215.82.23.2


cloudera@quickstart:~
grunt> fs -ls clickstream
Found 4 items
drwxr-xr-x   - cloudera cloudera   0 2021-08-08 17:44 clickstream/iplookup
drwxr-xr-x   - cloudera cloudera   0 2021-08-13 21:19 clickstream/iplookup_noheader
drwxr-xr-x   - cloudera cloudera   0 2021-08-08 17:45 clickstream/logs
drwxr-xr-x   - cloudera cloudera   0 2021-08-13 23:01 clickstream/logs_noheader
grunt> fs -ls clickstream/logs_noheader
Found 2 items
-rw-r--r--   1 cloudera cloudera   0 2021-08-13 23:01 clickstream/logs_noheader/_SUCCESS
-rw-r--r--   1 cloudera cloudera   227447 2021-08-13 23:01 clickstream/logs_noheader/part-m-00000
grunt> fs -cat clickstream/logs_noheader/part-m-00000
#Software: Microsoft   Information   Services
#Version: 1.0
#Date: 2016-02-11
#Fields: date   s-ip   cs-method   cs-username   c-ip
2016-02-11   17:16:13   GET   /   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Content/jquery-ui-themes/smoothness/jquery-ui-1.10.3.custom.min.css   215.82.23.2
ozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Plugins/Widgets.NivoSlider/Content/nivoslider/themes/custom/custom.css 215.82.23.2
ozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Plugins/Widgets.NivoSlider/Content/nivoslider/nivo-slider.css 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Scripts/jquery.validate.unobtrusive.min.js   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Scripts/jquery.validate.min.js 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Themes/DefaultClean/Content/css/styles.css   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Scripts/jquery-migrate-1.2.1.min.js   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Scripts/public.common.js   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Scripts/public.ajaxcart.js   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Plugins/Widgets.NivoSlider/Scripts/jquery.nivo.slider.js   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Scripts/jquery-ui-1.10.3.custom.min.js 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Scripts/jquery-1.10.2.min.js   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000075_banner_1.jpg   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Themes/DefaultClean/Content/images/logo.png   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000005_electronics_450.jpeg   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000009_apparel_450.jpeg 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000013_digital-downloads_450.jpeg   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000072_25-virtual-gift-card_415.jpeg   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000020_build-your-own-computer_415.jpeg 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000024_apple-macbook-pro-13-inch_415.jpeg   215.82.23.2
ozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /content/images/thumbs/0000041_htc-one-m8-android-l-50-lollipop_415.jpeg   215.82.23.2
ozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Plugins/Widgets.NivoSlider/Content/nivoslider/themes/custom/loading.gif   215.82.23.2
ozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Themes/DefaultClean/Content/images/rating1.png 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Themes/DefaultClean/Content/images/rating2.png 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Themes/DefaultClean/Content/images/shopping-bag.png   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Themes/DefaultClean/Content/images/social-sprite.png   215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Themes/DefaultClean/Content/images/compare-button.png 215.82.23.2   Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
2016-02-11   17:16:13   GET   /Content/jquery-ui-themes/smoothness/images/ui-bg_flat_75_ffffff_40x100.png   215.82.23.2

769-Win10Docker-srajendi

cloudera@quickstart:~

```
grunt> quit
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 7 items
drwxr-xr-x   - cloudera cloudera          0 2021-08-13 23:01 clickstream
drwxr-xr-x   - cloudera cloudera          0 2021-08-10 00:32 fudgemart-clothing
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 2021-08-13 00:23 testing
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 ~]$ hdfs dfs -ls clickstream
Found 4 items
drwxr-xr-x   - cloudera cloudera          0 2021-08-08 17:44 clickstream/iplookup
drwxr-xr-x   - cloudera cloudera          0 2021-08-13 21:19 clickstream/iplookup_noheader
drwxr-xr-x   - cloudera cloudera          0 2021-08-08 17:45 clickstream/logs
drwxr-xr-x   - cloudera cloudera          0 2021-08-13 23:01 clickstream/logs_noheader
[cloudera@quickstart ~]$ hdfs dfs -ls clickstream/*header
Found 2 items
-rw-r--r--   1 cloudera cloudera          0 2021-08-13 21:19 clickstream/iplookup_noheader/_SUCCESS
-rw-r--r--   1 cloudera cloudera       1185 2021-08-13 21:19 clickstream/iplookup_noheader/part-r-00000
Found 2 items
-rw-r--r--   1 cloudera cloudera          0 2021-08-13 23:01 clickstream/logs_noheader/_SUCCESS
-rw-r--r--   1 cloudera cloudera     227447 2021-08-13 23:01 clickstream/logs_noheader/part-m-00000
[cloudera@quickstart ~]$
```

769-Win10Docker-srajendi

cloudera@quickstart:~

```
[cloudera@quickstart ~]$ beeline -u jdbc:hive2://localhost:10000/default -u cloudera -p cloudera
2021-08-13 23:15:01,472 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_20210813231515_4b33b72a-dc97-4428-974e-a04502dffc72): 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_20210813231515_4b33b72a-dc97-4428-974e-a04502dffc72); Time taken: 1.298 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210813231515_4b33b72a-dc97-4428-974e-a04502dffc72): show databases
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210813231515_4b33b72a-dc97-4428-974e-a04502dffc72); Time taken: 0.419 seconds
INFO  : OK
+----------------+--+
| database_name  |
+----------------+--+
| clickstream    |
| default        |
+----------------+--+
2 rows selected (2.162 seconds)
```

cloudera@quickstart:~

```
[cloudera@quickstart ~]$ beeline -u jdbc:hive2://localhost:10000/default -u cloudera -p cloudera
2021-08-13 23:26:29,123 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_20210813232626_92d6ab46-8595-48f2-96c5-ce0c5b27efd0): 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_20210813232626_92d6ab46-8595-48f2-96c5-ce0c5b27efd0); Time taken: 0.031 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210813232626_92d6ab46-8595-48f2-96c5-ce0c5b27efd0): show databases
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210813232626_92d6ab46-8595-48f2-96c5-ce0c5b27efd0); Time taken: 0.049 seconds
INFO  : OK
+----------------+--+
| database_name  |
+----------------+--+
| clickstream    |
| default        |
+----------------+--+
2 rows selected (0.224 seconds)
```

769-Win10Docker-srajendi

cloudera@quickstart:~

```
0: jdbc:hive2://localhost:10000/default> DROP DATABASE IF EXISTS clickstream CASCADE;
INFO  : Compiling command(queryId=hive_20210813232727_3d960416-98b1-460b-92c1-f86f496fcbe8): DROP DATABASE IF EXISTS clickstream CASCADE
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20210813232727_3d960416-98b1-460b-92c1-f86f496fcbe8); Time taken: 0.665 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210813232727_3d960416-98b1-460b-92c1-f86f496fcbe8): DROP DATABASE IF EXISTS clickstream CASCADE
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210813232727_3d960416-98b1-460b-92c1-f86f496fcbe8); Time taken: 7.948 seconds
INFO  : OK
No rows affected (8.656 seconds)
0: jdbc:hive2://localhost:10000/default> show databases;
INFO  : Compiling command(queryId=hive_20210813232727_8c103cad-6b1e-4dbf-8a82-66746a8cf1b2): 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_20210813232727_8c103cad-6b1e-4dbf-8a82-66746a8cf1b2); Time taken: 0.007 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210813232727_8c103cad-6b1e-4dbf-8a82-66746a8cf1b2): show databases
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210813232727_8c103cad-6b1e-4dbf-8a82-66746a8cf1b2); Time taken: 0.019 seconds
INFO  : OK
+----------------+--+
| database_name  |
+----------------+--+
| default        |
+----------------+--+
1 row selected (0.061 seconds)
0: jdbc:hive2://localhost:10000/default>
```

769-Win10Docker-srajendi

cloudera@quickstart:~

```
0: jdbc:hive2://localhost:10000/default> CREATE DATABASE clickstream;
INFO  : Compiling command(queryId=hive_20210813232828_04815a86-f0fa-49f4-8d31-214de74d3e6d): CREATE DATABASE clickstream
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20210813232828_04815a86-f0fa-49f4-8d31-214de74d3e6d); Time taken: 0.007 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210813232828_04815a86-f0fa-49f4-8d31-214de74d3e6d): CREATE DATABASE clickstream
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210813232828_04815a86-f0fa-49f4-8d31-214de74d3e6d); Time taken: 0.212 seconds
INFO  : OK
No rows affected (0.243 seconds)
0: jdbc:hive2://localhost:10000/default> show databases;
INFO  : Compiling command(queryId=hive_20210813232828_666a979d-f78a-4c96-ab33-f216dac5f4c4): 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_20210813232828_666a979d-f78a-4c96-ab33-f216dac5f4c4); Time taken: 0.008 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210813232828_666a979d-f78a-4c96-ab33-f216dac5f4c4): show databases
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210813232828_666a979d-f78a-4c96-ab33-f216dac5f4c4); Time taken: 0.006 seconds
INFO  : OK
+----------------+--+
| database_name  |
+----------------+--+
| clickstream    |
| default        |
+----------------+--+
2 rows selected (0.055 seconds)
0: jdbc:hive2://localhost:10000/default> show tables;
INFO  : Compiling command(queryId=hive_20210813232929_506dd7cd-f3eb-49be-9578-7d2262b007c2): show tablesream
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_20210813232929_506dd7cd-f3eb-49be-9578-7d2262b007c2); Time taken: 0.007 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210813232929_506dd7cd-f3eb-49be-9578-7d2262b007c2): show tablesream
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210813232929_506dd7cd-f3eb-49be-9578-7d2262b007c2); Time taken: 0.012 seconds
INFO  : OK
+-----------+--+ (0.059 seconds)
| tab_name  |://localhost:10000/default> swo
+-----------+--+
+-----------+--+
No rows selected (0.055 seconds)
0: jdbc:hive2://localhost:10000/default>
```

769-Win10Docker-srajendi

cloudera@quickstart:~

```
0: jdbc:hive2://localhost:10000/default> show tables;
INFO  : Compiling command(queryId=hive_20210813232929_506dd7cd-f3eb-49be-9578-7d2262b007c2): show tablesream
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_20210813232929_506dd7cd-f3eb-49be-9578-7d2262b007c2); Time taken: 0.007 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210813232929_506dd7cd-f3eb-49be-9578-7d2262b007c2): show tablesream
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210813232929_506dd7cd-f3eb-49be-9578-7d2262b007c2); Time taken: 0.012 seconds
INFO  : OK
+-----------+--+ (0.059 seconds)
| tab_name  |://localhost:10000/default> swo
+-----------+--+
+-----------+--+
No rows selected (0.055 seconds)
0: jdbc:hive2://localhost:10000/default> !sh hdfs dfs -ls clickstream/*header
Found 2 items
-rw-r--r--   1 cloudera cloudera          0 2021-08-13 21:19 clickstream/iplookup_noheader/_SUCCESS
-rw-r--r--   1 cloudera cloudera       1185 2021-08-13 21:19 clickstream/iplookup_noheader/part-r-00000
Found 2 items
-rw-r--r--   1 cloudera cloudera          0 2021-08-13 23:01 clickstream/logs_noheader/_SUCCESS
-rw-r--r--   1 cloudera cloudera     227447 2021-08-13 23:01 clickstream/logs_noheader/part-m-00000
0: jdbc:hive2://localhost:10000/default>
```

769-Win10Docker-srajendi

cloudera@quickstart:~

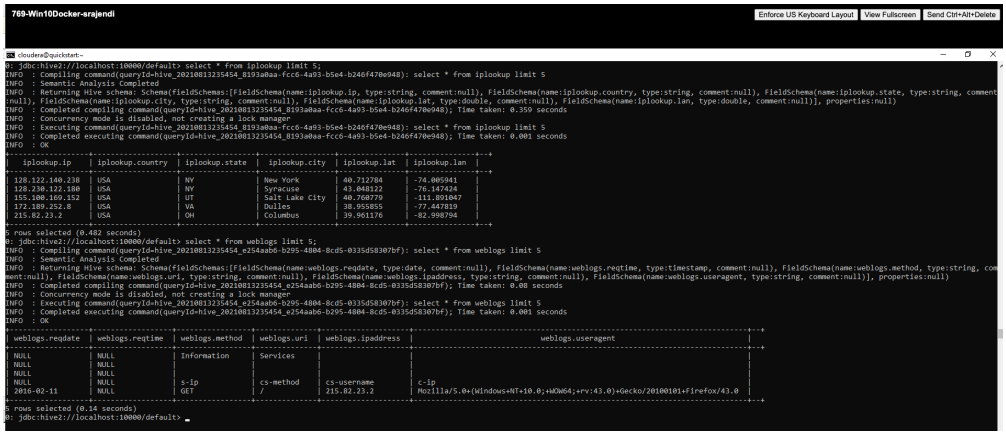
```
0: jdbc:hive2://localhost:10000/default> create external table iplookup (
0: jdbc:hive2://localhost:10000/default> ip string,country string,state string,city string,lat double,lan double)
0: jdbc:hive2://localhost:10000/default> row format delimited
0: jdbc:hive2://localhost:10000/default> fields terminated by ','
0: jdbc:hive2://localhost:10000/default> location '/user/cloudera/clickstream/iplookup_noheader/';
INFO  : Compiling command(queryId=hive_20210813234848_d3ab2589-aba9-4f97-8a56-701a5b8c27f8): create external table iplookup (
ip string,country string,state string,city string,lat double,lan double)
row format delimited
fields terminated by ','
location '/user/cloudera/clickstream/iplookup_noheader/'
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20210813234848_d3ab2589-aba9-4f97-8a56-701a5b8c27f8); Time taken: 0.027 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210813234848_d3ab2589-aba9-4f97-8a56-701a5b8c27f8): create external table iplookup (
ip string,country string,state string,city string,lat double,lan double)
row format delimited
fields terminated by ','
location '/user/cloudera/clickstream/iplookup_noheader/'
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210813234848_d3ab2589-aba9-4f97-8a56-701a5b8c27f8); Time taken: 0.16 seconds
INFO  : OK
No rows affected (0.204 seconds)
0: jdbc:hive2://localhost:10000/default>
```

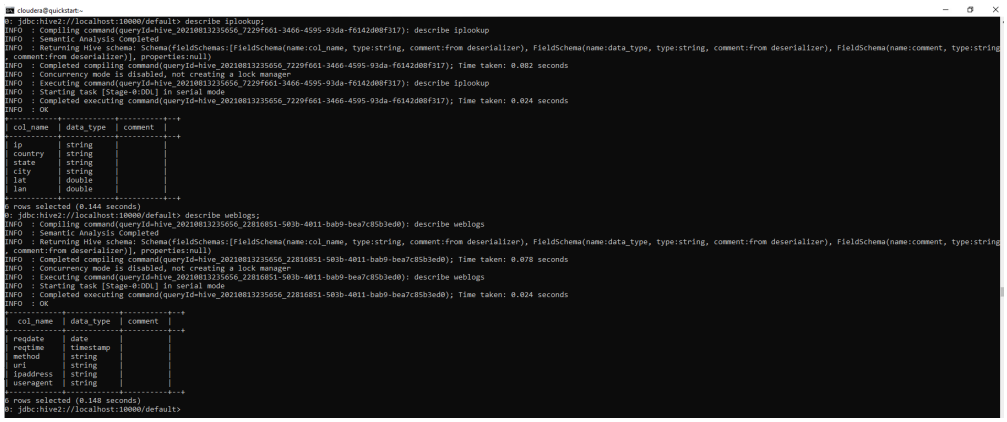
769-Win10Docker-srajendi

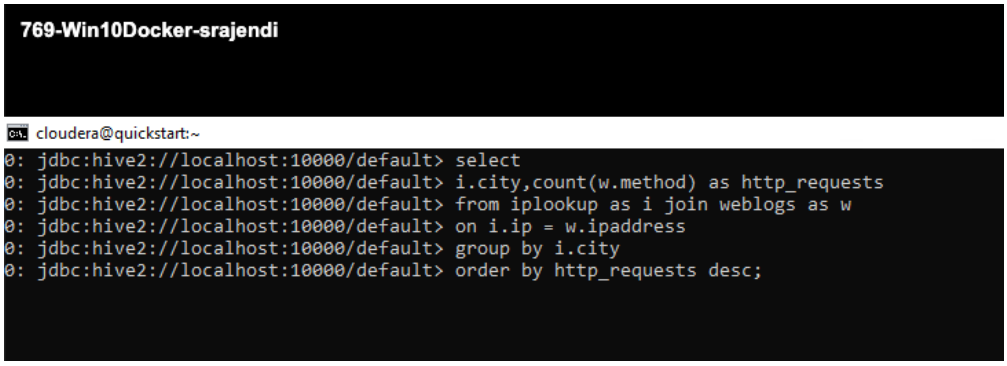
Enforce US Keyboard

cloudera@quickstart:~

```
0: jdbc:hive2://localhost:10000/default> create external table weblogs (
0: jdbc:hive2://localhost:10000/default> reqdate date,reqtime timestamp,method string,uri string ipaddress string,useragent string)
0: jdbc:hive2://localhost:10000/default> row format delimited
0: jdbc:hive2://localhost:10000/default> fields terminated by '\t'
0: jdbc:hive2://localhost:10000/default> location '/user/cloudera/clickstream/logs_noheader/';
Error: Error while compiling statement: FAILED: ParseException line 2:56 mismatched input 'ipaddress' expecting ) near 'string' in create table statement (state=42000,code=40000)
0: jdbc:hive2://localhost:10000/default> create external table weblogs (
0: jdbc:hive2://localhost:10000/default> reqdate date,reqtime timestamp,method string,uri string, ipaddress string,useragent string)
0: jdbc:hive2://localhost:10000/default> row format delimited
0: jdbc:hive2://localhost:10000/default> fields terminated by '\t'
0: jdbc:hive2://localhost:10000/default> location '/user/cloudera/clickstream/logs_noheader/';
INFO  : Compiling command(queryId=hive_20210813235353_a49da797-ee6b-40d8-96eb-41a9ed56b1a7): create external table weblogs (
reqdate date,reqtime timestamp,method string,uri string, ipaddress string,useragent string)
row format delimited
fields terminated by '\t'
location '/user/cloudera/clickstream/logs_noheader/'
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO  : Completed compiling command(queryId=hive_20210813235353_a49da797-ee6b-40d8-96eb-41a9ed56b1a7); Time taken: 0.027 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210813235353_a49da797-ee6b-40d8-96eb-41a9ed56b1a7): create external table weblogs (
reqdate date,reqtime timestamp,method string,uri string, ipaddress string,useragent string)
row format delimited
fields terminated by '\t'
location '/user/cloudera/clickstream/logs_noheader/'
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20210813235353_a49da797-ee6b-40d8-96eb-41a9ed56b1a7); Time taken: 0.061 seconds
INFO  : OK
No rows affected (0.107 seconds)
0: jdbc:hive2://localhost:10000/default>
```


769-Win10Docker-srajendi
Enforce US Keyboard Layout
View Fullscreen
Send Ctrl+Alt+Delete
cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> select * from iplookup limit 5;
INFO : Compiling command(queryId=hive_20210813235454_8193a0aa-fcc6-4a93-b5e4-b246f470e948): select * from iplookup limit 5
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_20210813235454_8193a0aa-fcc6-4a93-b5e4-b246f470e948); Time taken: 0.359 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20210813235454_8193a0aa-fcc6-4a93-b5e4-b246f470e948): select * from iplookup limit 5
INFO : Completed executing command(queryId=hive_20210813235454_8193a0aa-fcc6-4a93-b5e4-b246f470e948); Time taken: 0.001 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
5 rows selected (0.482 seconds)
0: jdbc:hive2://localhost:10000/default> select * from weblogs limit 5;
INFO : Compiling command(queryId=hive_20210813235454_e254aab6-b295-4804-8cd5-0335d58307bf): select * from weblogs limit 5
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:weblogs.reqdate, type:date, comment:null), FieldSchema(name:weblogs.reqtime, type:timestamp, comment:null), FieldSchema(name:weblogs.method, type:string, comment:null), FieldSchema(name:weblogs.uri, type:string, comment:null), FieldSchema(name:weblogs.ipaddress, type:string, comment:null), FieldSchema(name:weblogs.useragent, type:string, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20210813235454_e254aab6-b295-4804-8cd5-0335d58307bf); Time taken: 0.08 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20210813235454_e254aab6-b295-4804-8cd5-0335d58307bf): select * from weblogs limit 5
INFO : Completed executing command(queryId=hive_20210813235454_e254aab6-b295-4804-8cd5-0335d58307bf); Time taken: 0.001 seconds
INFO : OK
weblogs.reqdate	weblogs.reqtime	weblogs.method	weblogs.uri	weblogs.ipaddress	weblogs.useragent
NULL	NULL	Information	Services		
NULL	NULL				
NULL	NULL				
NULL	NULL	s-ip	cs-method	cs-username	c-ip
2016-02-11	NULL	GET	/	215.82.23.2	Mozilla/5.0+(Windows+NT+10.0;+WOW64;+rv:43.0)+Gecko/20100101+Firefox/43.0
5 rows selected (0.14 seconds)
0: jdbc:hive2://localhost:10000/default>

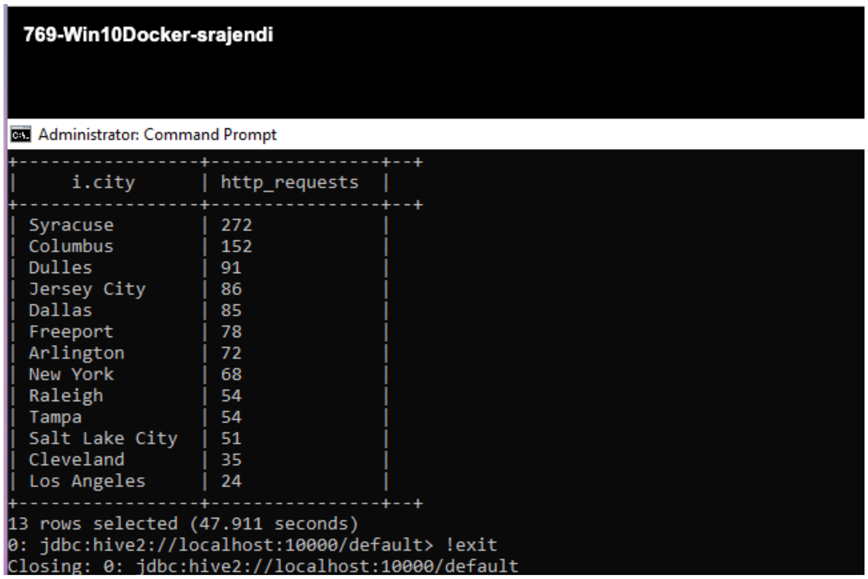

cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> describe iplookup;
INFO : Compiling command(queryId=hive_20210813235656_7229f661-3466-4595-93da-f6142d08f317): describe iplookup
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, 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_20210813235656_7229f661-3466-4595-93da-f6142d08f317); Time taken: 0.082 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20210813235656_7229f661-3466-4595-93da-f6142d08f317): describe iplookup
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20210813235656_7229f661-3466-4595-93da-f6142d08f317); Time taken: 0.024 seconds
INFO : OK
col_name	data_type	comment
ip	string	
country	string	
state	string	
city	string	
lat	double	
lan	double	
6 rows selected (0.144 seconds)
0: jdbc:hive2://localhost:10000/default> describe weblogs;
INFO : Compiling command(queryId=hive_20210813235656_22816851-503b-4011-bab9-bea7c85b3ed0): describe weblogs
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_20210813235656_22816851-503b-4011-bab9-bea7c85b3ed0); Time taken: 0.078 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20210813235656_22816851-503b-4011-bab9-bea7c85b3ed0): describe weblogs
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20210813235656_22816851-503b-4011-bab9-bea7c85b3ed0); Time taken: 0.024 seconds
INFO : OK
col_name	data_type	comment
reqdate	date	
reqtime	timestamp	
method	string	
uri	string	
ipaddress	string	
useragent	string	
6 rows selected (0.148 seconds)
0: jdbc:hive2://localhost:10000/default>


769-Win10Docker-srajendi
cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> select
0: jdbc:hive2://localhost:10000/default> i.city,count(w.method) as http_requests
0: jdbc:hive2://localhost:10000/default> from iplookup as i join weblogs as w
0: jdbc:hive2://localhost:10000/default> on i.ip = w.ipaddress
0: jdbc:hive2://localhost:10000/default> group by i.city
0: jdbc:hive2://localhost:10000/default> order by http_requests desc;

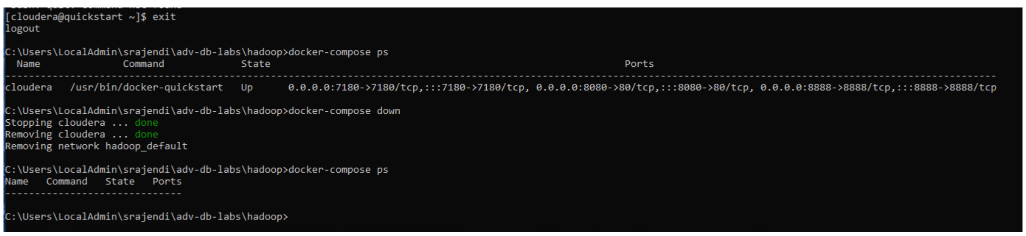
```
cloudera@quickstart:~
0: jdbc:hive2://localhost:10000/default> select
0: jdbc:hive2://localhost:10000/default> i.city,count(w.method) as http_requests
0: jdbc:hive2://localhost:10000/default> from iplookup as i join weblogs as w
0: jdbc:hive2://localhost:10000/default> on i.ip = w.ipaddress
0: jdbc:hive2://localhost:10000/default> group by i.city
0: jdbc:hive2://localhost:10000/default> order by http_requests desc;
INFO  : Compiling command(queryId=hive_20210814000404_239f1460-9b8f-4cbc-a59e-337aa1ebd637): select
i.city,count(w.method) as http_requests
from iplookup as i join weblogs as w
on i.ip = w.ipaddress
group by i.city
order by http_requests desc
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:i.city, type:string, comment:null), FieldSchema(name:http_requests, type:bigint, comment:null)], properties:null)
INFO  : Completed compiling command(queryId=hive_20210814000404_239f1460-9b8f-4cbc-a59e-337aa1ebd637); Time taken: 0.394 seconds
INFO  : Concurrency mode is disabled, not creating a lock manager
INFO  : Executing command(queryId=hive_20210814000404_239f1460-9b8f-4cbc-a59e-337aa1ebd637): select
i.city,count(w.method) as http_requests
from iplookup as i join weblogs as w
on i.ip = w.ipaddress
group by i.city
order by http_requests desc
INFO  : Query ID = hive_20210814000404_239f1460-9b8f-4cbc-a59e-337aa1ebd637
INFO  : Total jobs = 2
INFO  : Starting task [Stage-6:MAPREDLOCAL] in serial mode
INFO  : Execution completed successfully
INFO  : MapredLocal task succeeded
INFO  : Launching Job 1 out of 2
INFO  : Starting task [Stage-2:MAPRED] in serial mode
INFO  : Number of reduce tasks not specified. Estimated from input data size: 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_1628829248144_0016, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1628829248144_0016/
INFO  : Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1628829248144_0016
INFO  : Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
INFO  : 2021-08-14 00:04:56,675 Stage-2 map = 0%,  reduce = 0%
INFO  : 2021-08-14 00:05:03,089 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.15 sec
```

769-Win10Docker-srajendi

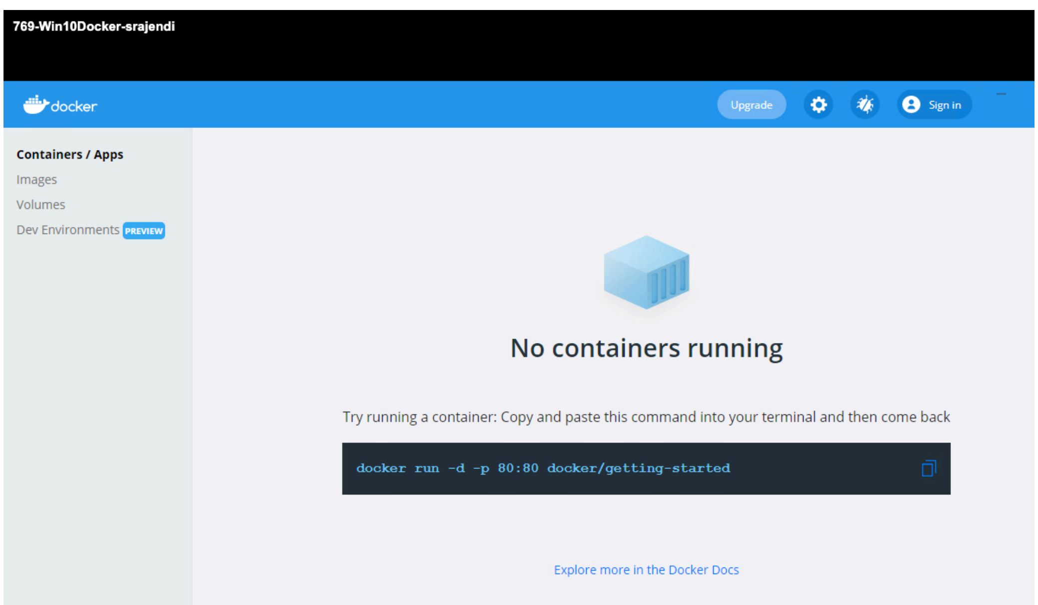
```
cloudera@quickstart:~
INFO  : Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1628829248144_0016
INFO  : Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
INFO  : 2021-08-14 00:04:56,675 Stage-2 map = 0%,  reduce = 0%
INFO  : 2021-08-14 00:05:03,089 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.15 sec
INFO  : 2021-08-14 00:05:11,509 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 3.62 sec
INFO  : MapReduce Total cumulative CPU time: 3 seconds 620 msec
INFO  : Ended Job = job_1628829248144_0016
INFO  : Launching Job 2 out of 2
INFO  : Starting task [Stage-3: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_1628829248144_0017, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1628829248144_0017/
INFO  : Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1628829248144_0017
INFO  : Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
INFO  : 2021-08-14 00:05:18,582 Stage-3 map = 0%,  reduce = 0%
INFO  : 2021-08-14 00:05:23,856 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 0.95 sec
INFO  : 2021-08-14 00:05:30,175 Stage-3 map = 100%,  reduce = 100%, Cumulative CPU 2.45 sec
INFO  : MapReduce Total cumulative CPU time: 2 seconds 450 msec
INFO  : Ended Job = job_1628829248144_0017
INFO  : MapReduce Jobs Launched:
INFO  : Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 3.62 sec   HDFS Read: 237563 HDFS Write: 456 SUCCESS
INFO  : Stage-Stage-3: Map: 1  Reduce: 1   Cumulative CPU: 2.45 sec   HDFS Read: 5093 HDFS Write: 164 SUCCESS
INFO  : Total MapReduce CPU Time Spent: 6 seconds 70 msec
INFO  : Completed executing command(queryId=hive_20210814000404_239f1460-9b8f-4cbc-a59e-337aa1ebd637); Time taken: 47.451 seconds
INFO  : OK
+-----------------+----------------+--+
|     i.city      | http_requests  |
+-----------------+----------------+--+
| Syracuse        | 272            |
| Columbus        | 152            |
| Dulles          | 91             |
| Jersey City     | 86             |
| Dallas          | 85             |
| Freeport        | 78             |
| Arlington       | 72             |
| New York        | 68             |
| Raleigh         | 54             |
| Tampa           | 54             |
| Salt Lake City  | 51             |
| Cleveland       | 35             |
| Los Angeles     | 24             |
+-----------------+----------------+--+
13 rows selected (47.911 seconds)
0: jdbc:hive2://localhost:10000/default>
```

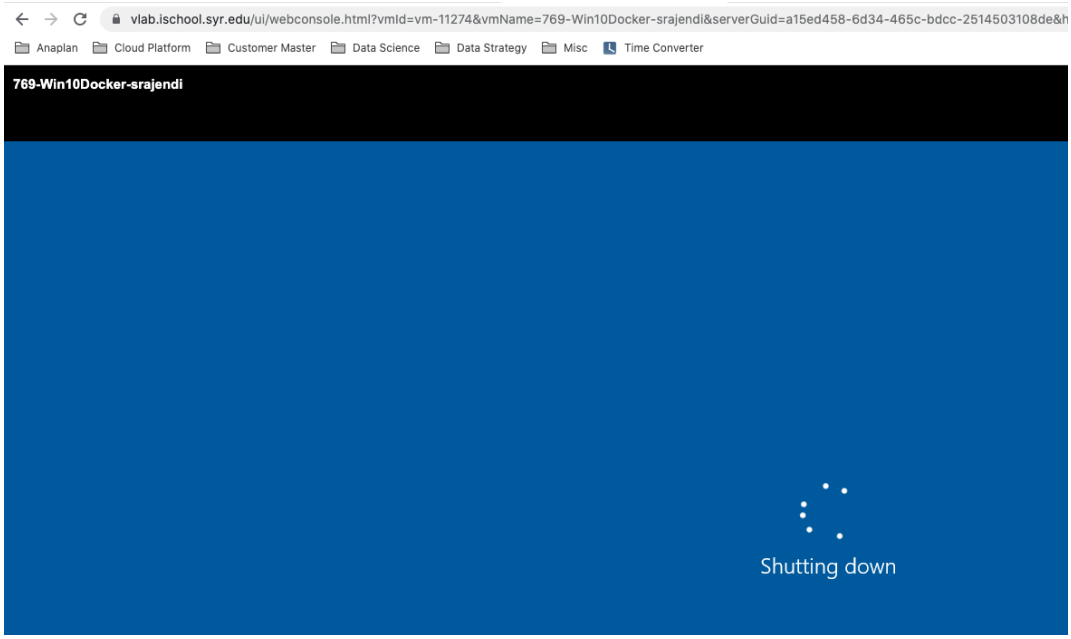


```
769-Win10Docker-srajendi

Administrator: Command Prompt
+------------------+-----------------+--+
|      i.city      | http_requests  |
+------------------+-----------------+--+
| Syracuse         | 272             |
| Columbus         | 152             |
| Dulles           | 91              |
| Jersey City      | 86              |
| Dallas           | 85              |
| Freeport         | 78              |
| Arlington        | 72              |
| New York         | 68              |
| Raleigh          | 54              |
| Tampa            | 54              |
| Salt Lake City   | 51              |
| Cleveland        | 35              |
| Los Angeles      | 24              |
+------------------+-----------------+--+
13 rows selected (47.911 seconds)
0: jdbc:hive2://localhost:10000/default> !exit
Closing: 0: jdbc:hive2://localhost:10000/default
```

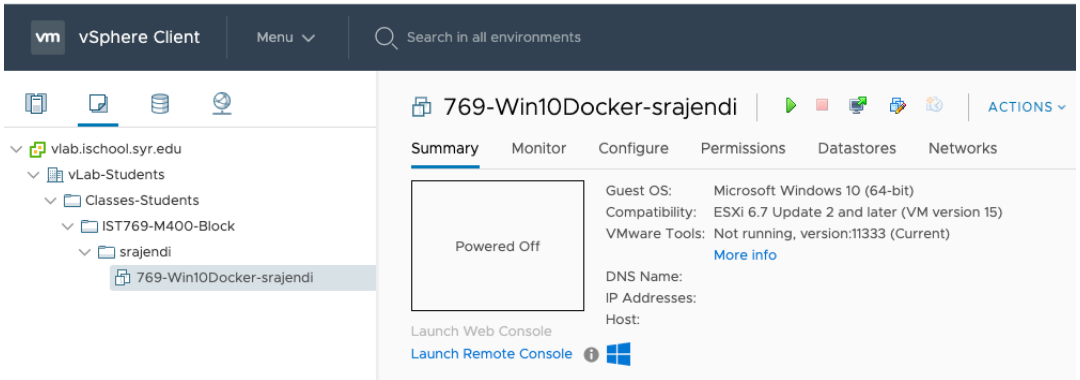


```
[cloudera@quickstart ~]$ exit
logout

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>
```




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.


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