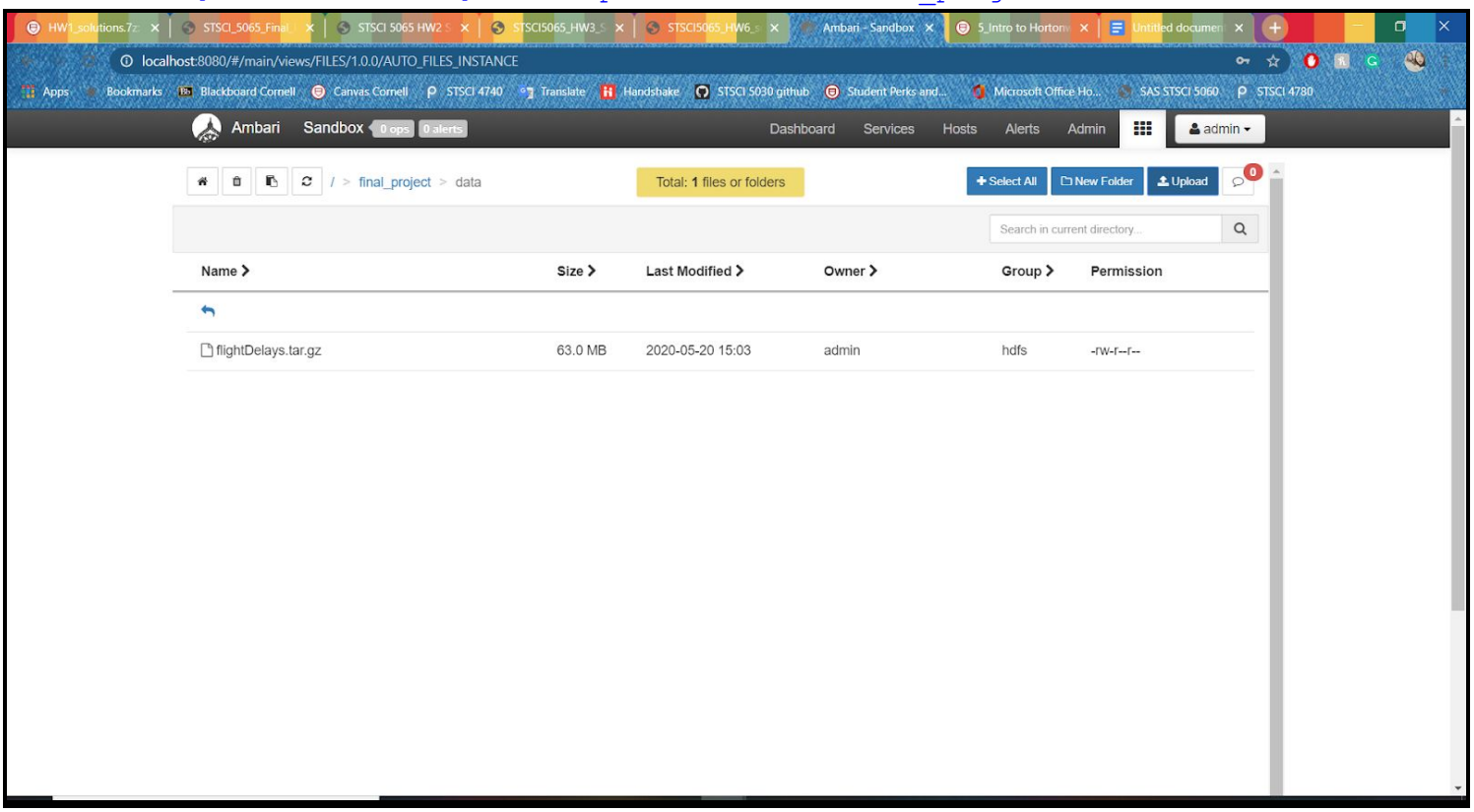
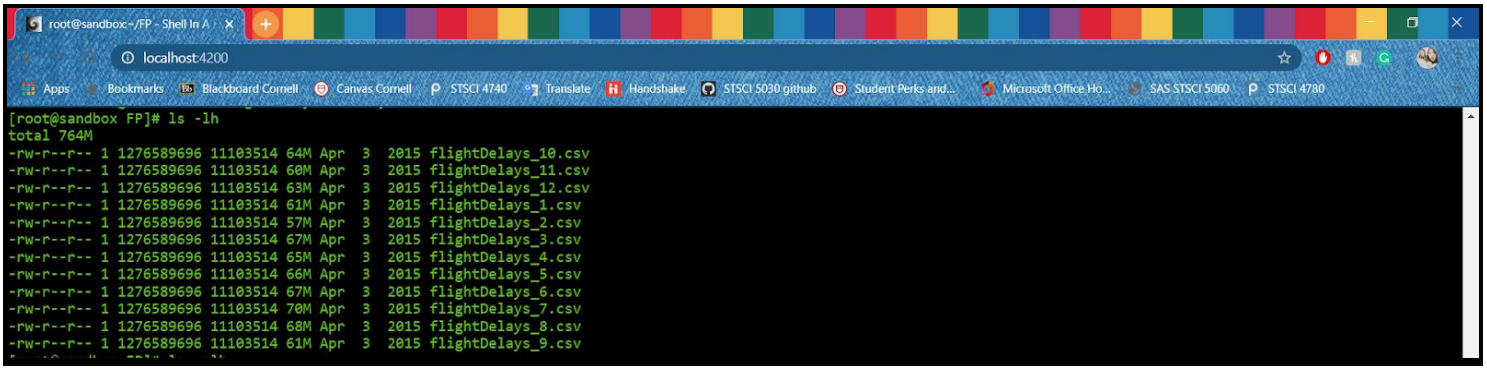
1.

```
[root@sandbox ~]# hadoop fs -mkdir /final_project
[root@sandbox ~]# hadoop fs -mkdir /final_project/data
```



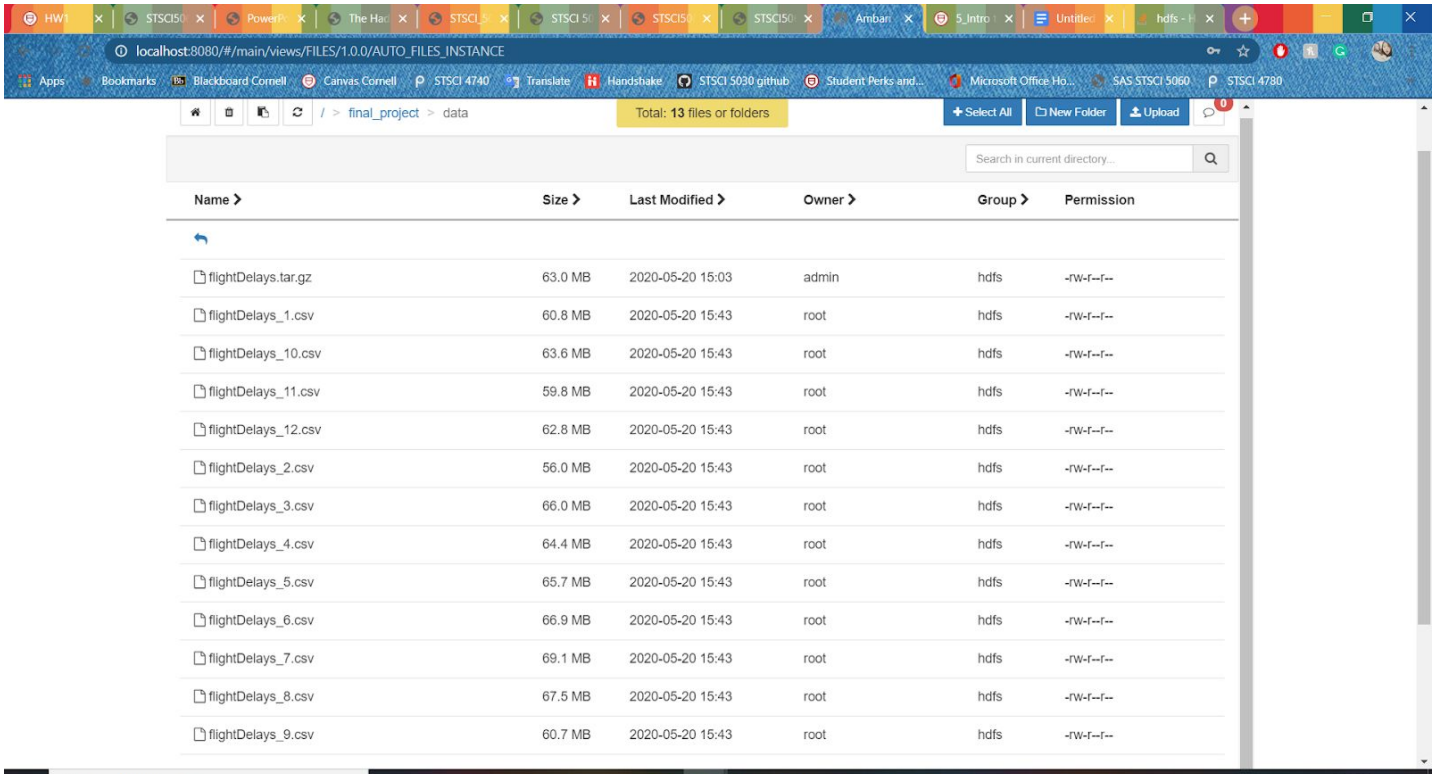
2.

```
[root@sandbox ~]# mkdir FP
[root@sandbox ~]# cd FP
[root@sandbox FP]# hadoop fs -copyToLocal
/final_project/data/flightDelays.tar.gz
[root@sandbox FP]# gzip -d flightDelays.tar.gz
[root@sandbox FP]# tar -xf flightDelays.tar
[root@sandbox FP]# ls -lh
```



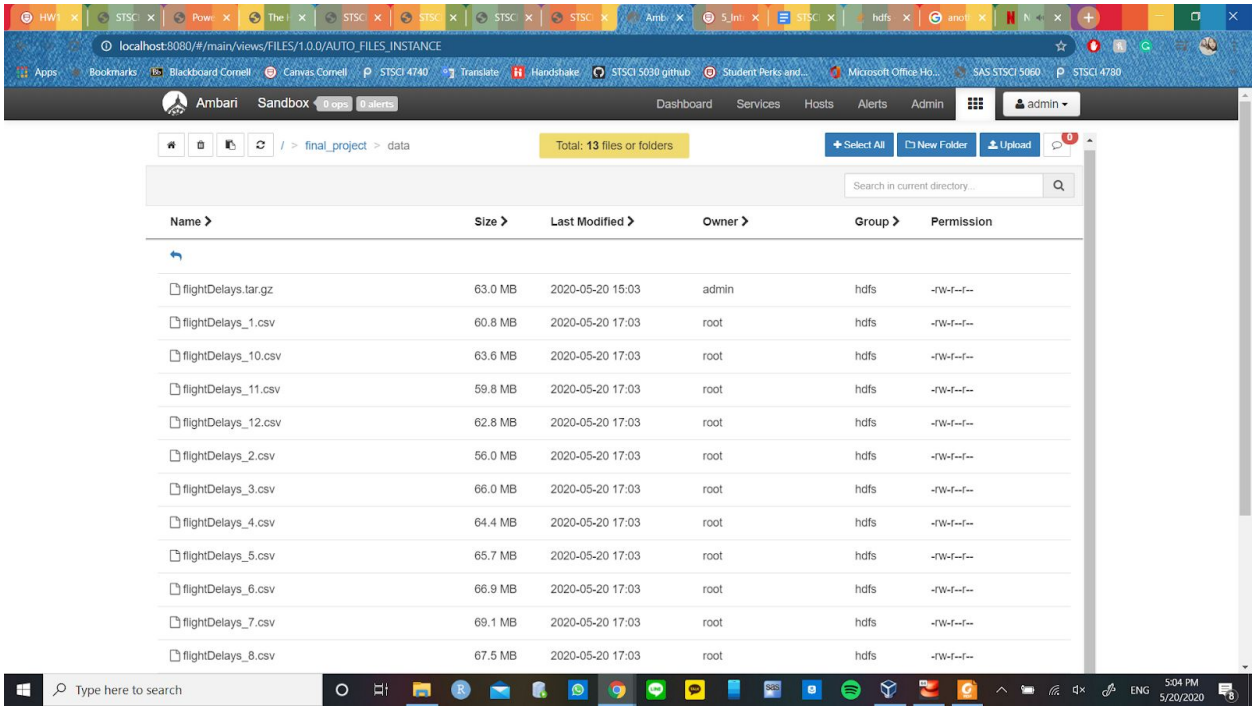
3.

```
[root@sandbox FP]# hadoop fs -copyFromLocal flightDelays_1.csv
flightDelays_2.csv flightDelays_3.csv flightDelays_4.csv
flightDelays_5.csv flightDelays_6.csv flightDelays_7.csv
flightDelays_8.csv flightDelays_9.csv flightDelays_10.csv
flightDelays_11.csv flightDelays_12.csv /final_project/data
```



The first time I used copyFromLocal from the FP folder only having to type out the file names to the HDFS directory.

```
[root@sandbox FP]# cd ~
[root@sandbox ~]# hadoop fs -copyFromLocal
~/FP/flightDelays_1.csv ~/FP/flightDelays_2.csv
~/FP/flightDelays_3.csv ~/FP/flightDelays_4.csv
~/FP/flightDelays_5.csv ~/FP/flightDelays_6.csv
~/FP/flightDelays_7.csv ~/FP/flightDelays_8.csv
~/FP/flightDelays_9.csv ~/FP/flightDelays_10.csv
~/FP/flightDelays_11.csv ~/FP/flightDelays_12.csv
/final_project/data
```

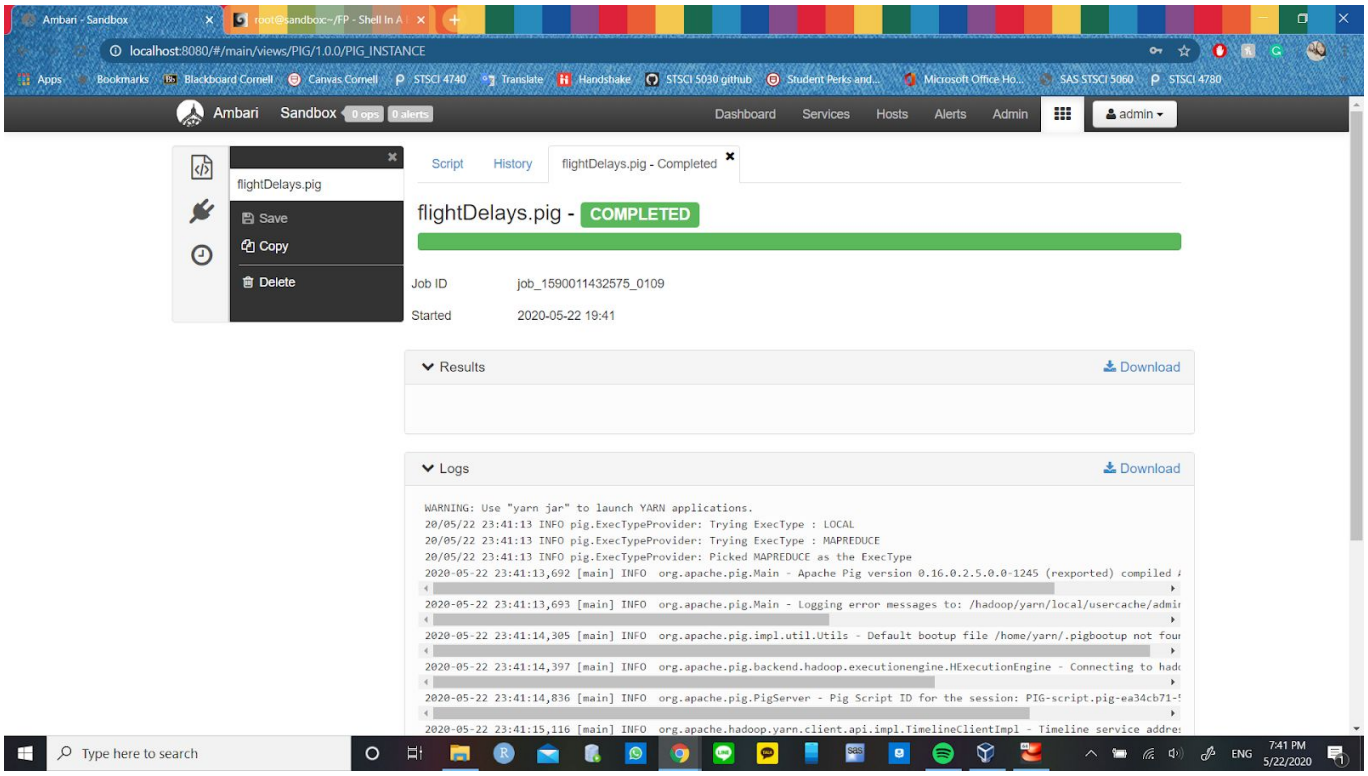


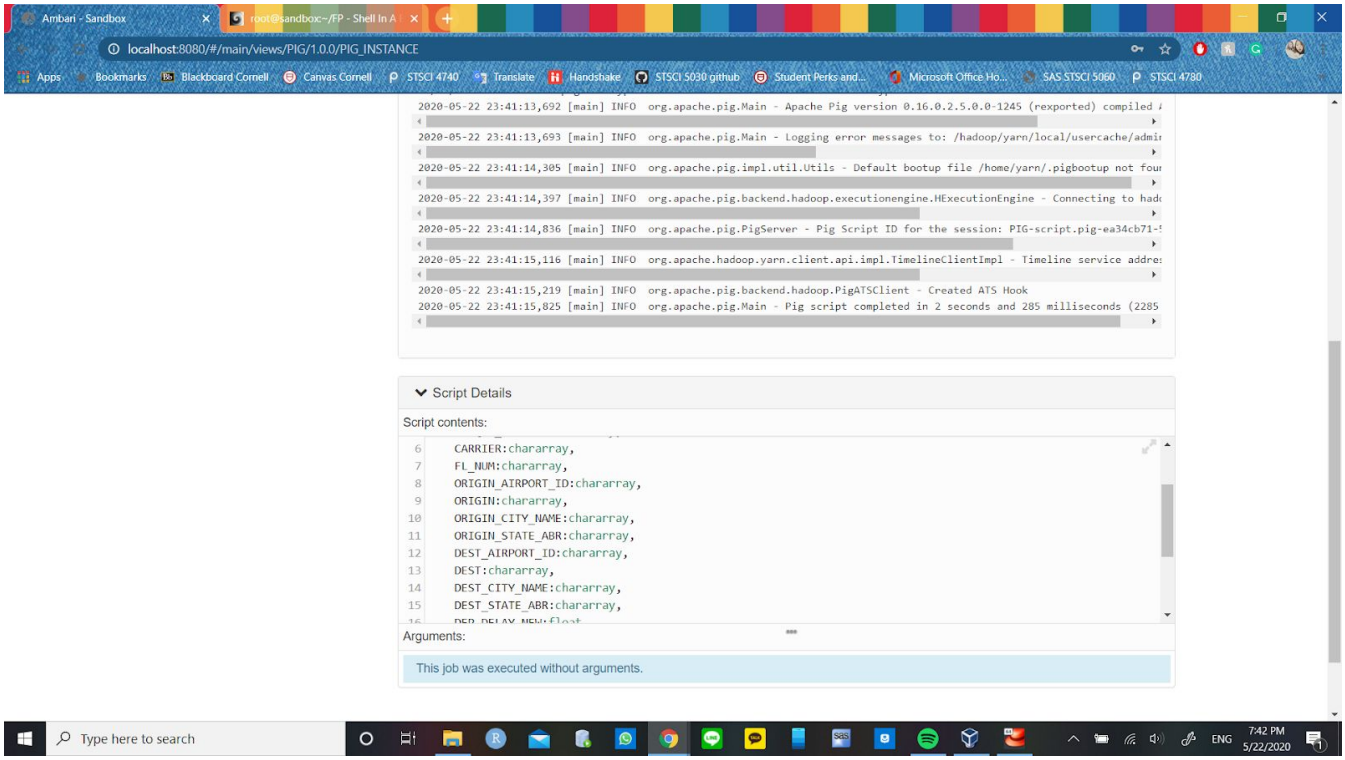
The second method I went back to the root and copied it from the root using the full address rather than from the FP folder.

4.
```
[root@sandbox ~]# hive
hive> CREATE DATABASE FPdb
    > LOCATION '/final_project';
hive> DESCRIBE DATABASE FPdb;
```

5.
```
flightDelays = LOAD '/final_project/data/flightDelays_*'
USING org.apache.pig.piggybank.storage.CSVExcelStorage()
as (YEAR:int,
FL_DATE:chararray,
UNIQUE_CARRIER:chararray,
CARRIER:chararray,
FL_NUM:chararray,
ORIGIN_AIRPORT_ID:chararray,
ORIGIN:chararray,
ORIGIN_CITY_NAME:chararray,
ORIGIN_STATE_ABR:chararray,
```

DEST_AIRPORT_ID:chararray,
DEST:chararray,
DEST_CITY_NAME:chararray,
DEST_STATE_ABR:chararray,
DEP_DELAY_NEW:float,
ARR_DELAY:float,
ARR_DELAY_NEW:float,
CARRIER_DELAY:float,
WEATHER_DELAY:float,
NAS_DELAY:float,
SECURITY_DELAY:float,
LATE_AIRCRAFT_DELAY:float);

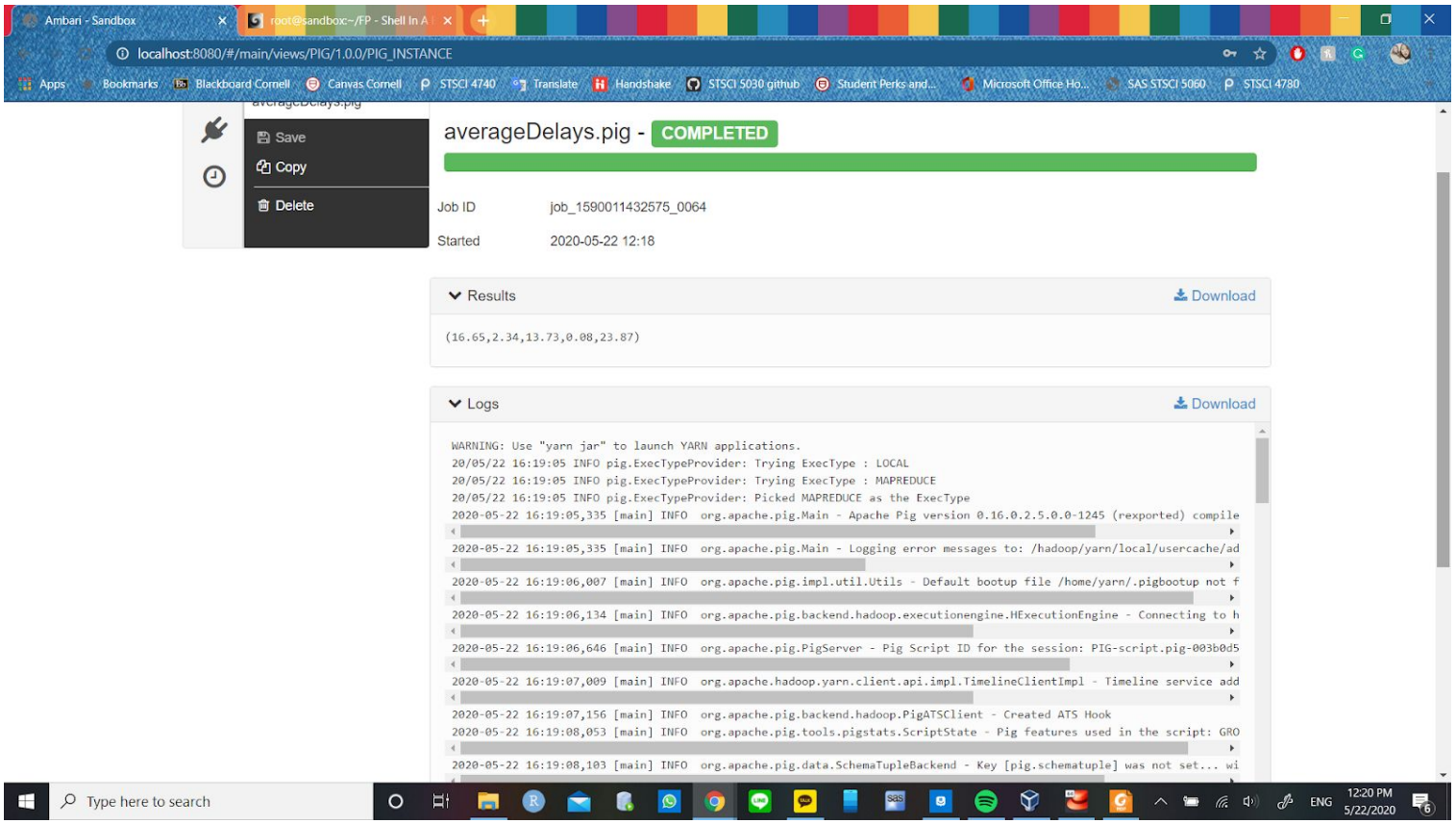


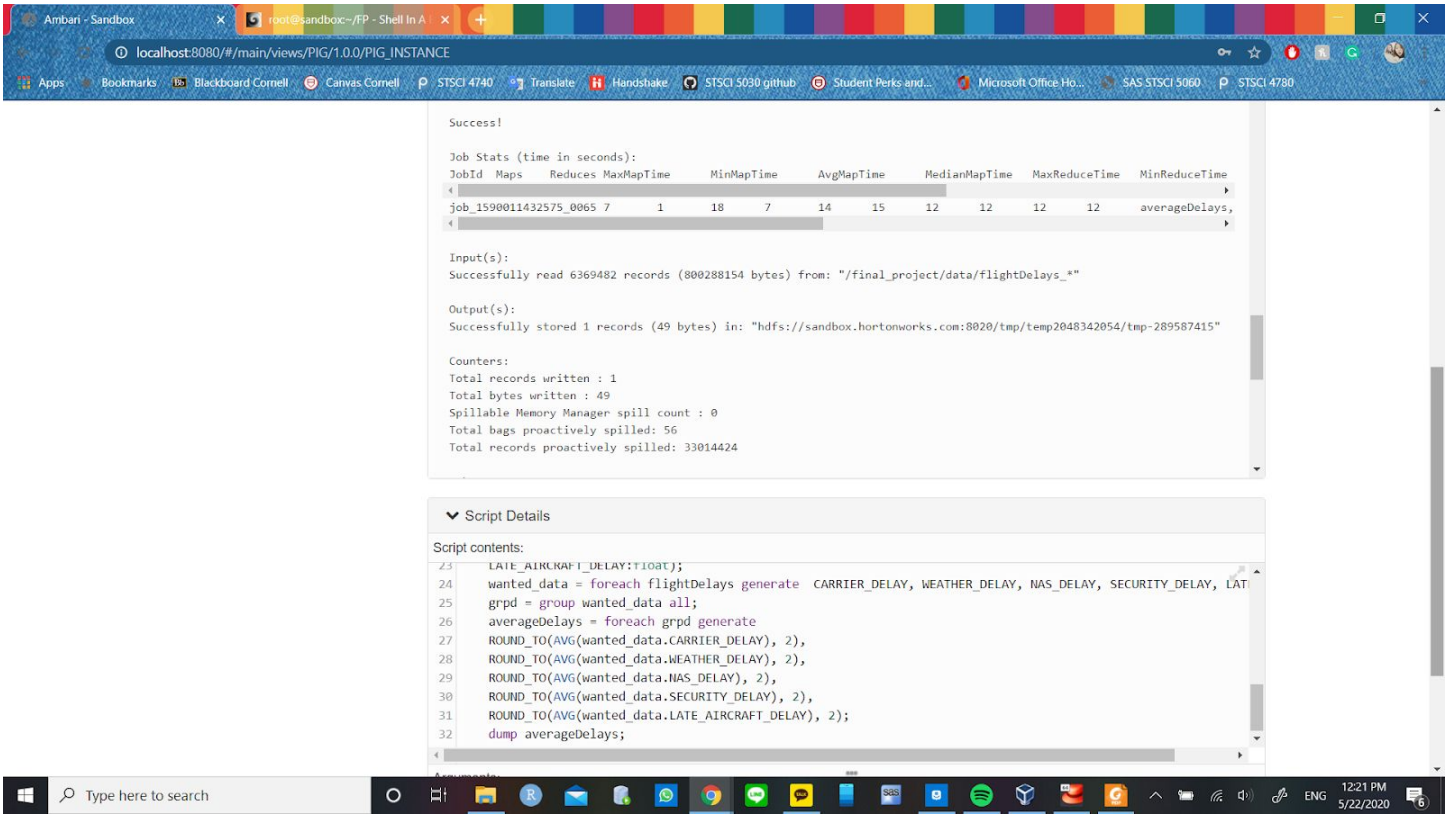


6.

```
flightDelays = LOAD '/final_project/data/flightDelays_*'
USING org.apache.pig.piggybank.storage.CSVExcelStorage()
as (YEAR:int,
FL_DATE:chararray,
UNIQUE_CARRIER:chararray,
CARRIER:chararray,
FL_NUM:chararray,
ORIGIN_AIRPORT_ID:chararray,
ORIGIN:chararray,
ORIGIN_CITY_NAME:chararray,
ORIGIN_STATE_ABR:chararray,
DEST_AIRPORT_ID:chararray,
DEST:chararray,
DEST_CITY_NAME:chararray,
DEST_STATE_ABR:chararray,
DEP_DELAY_NEW:float,
ARR_DELAY:float,
ARR_DELAY_NEW:float,
CARRIER_DELAY:float,
WEATHER_DELAY:float,
NAS_DELAY:float,
SECURITY_DELAY:float,
```

```
LATE_AIRCRAFT_DELAY:float);

wanted_data = foreach flightDelays generate  CARRIER_DELAY, WEATHER_DELAY,
NAS_DELAY, SECURITY_DELAY, LATE_AIRCRAFT_DELAY;
grpd = group wanted_data all;
averageDelays = foreach grpd generate
ROUND_TO(AVG(wanted_data.CARRIER_DELAY), 2),
ROUND_TO(AVG(wanted_data.WEATHER_DELAY), 2),
ROUND_TO(AVG(wanted_data.NAS_DELAY), 2),
ROUND_TO(AVG(wanted_data.SECURITY_DELAY), 2),
ROUND_TO(AVG(wanted_data.LATE_AIRCRAFT_DELAY), 2);
dump averageDelays;
```

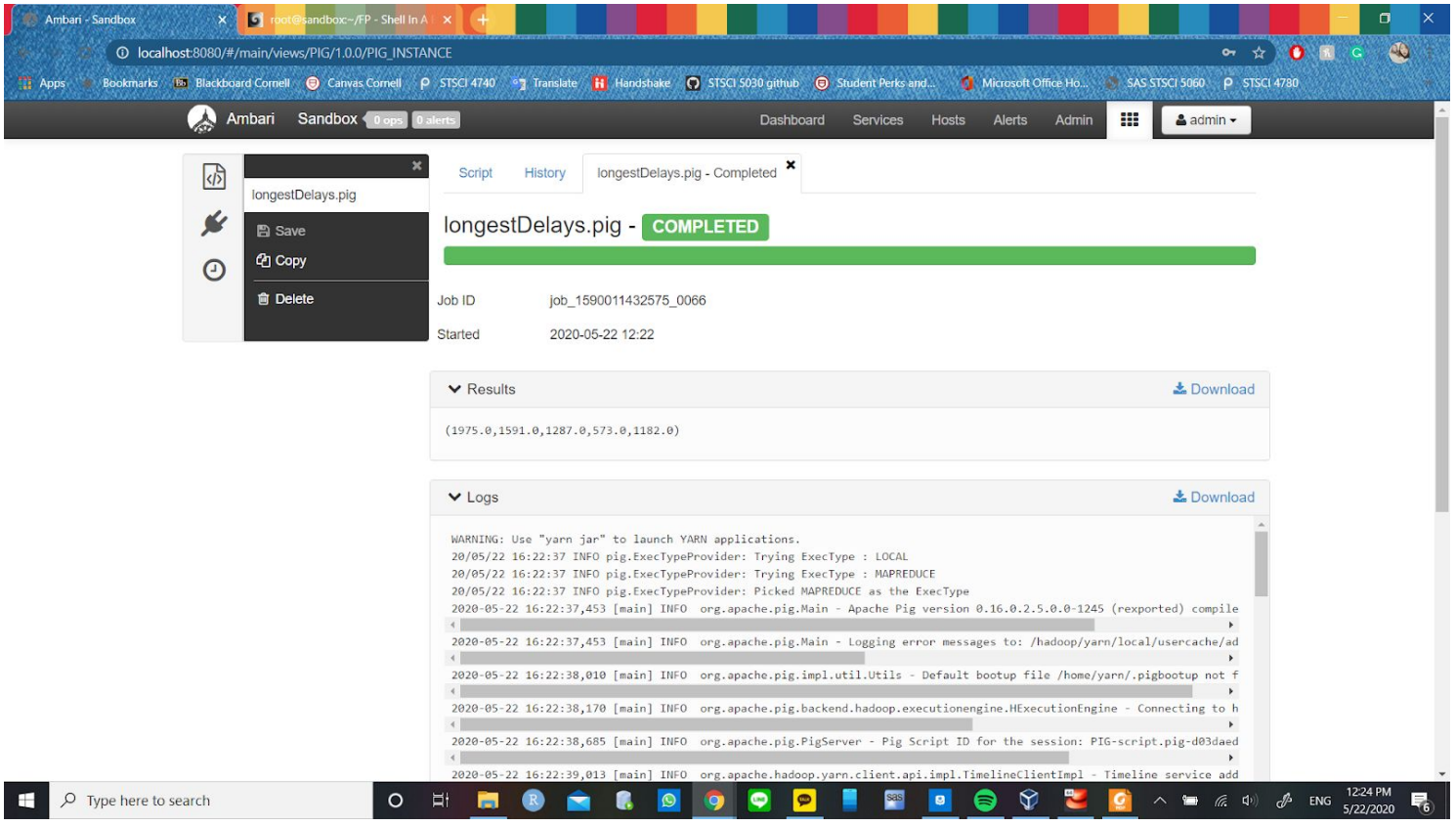


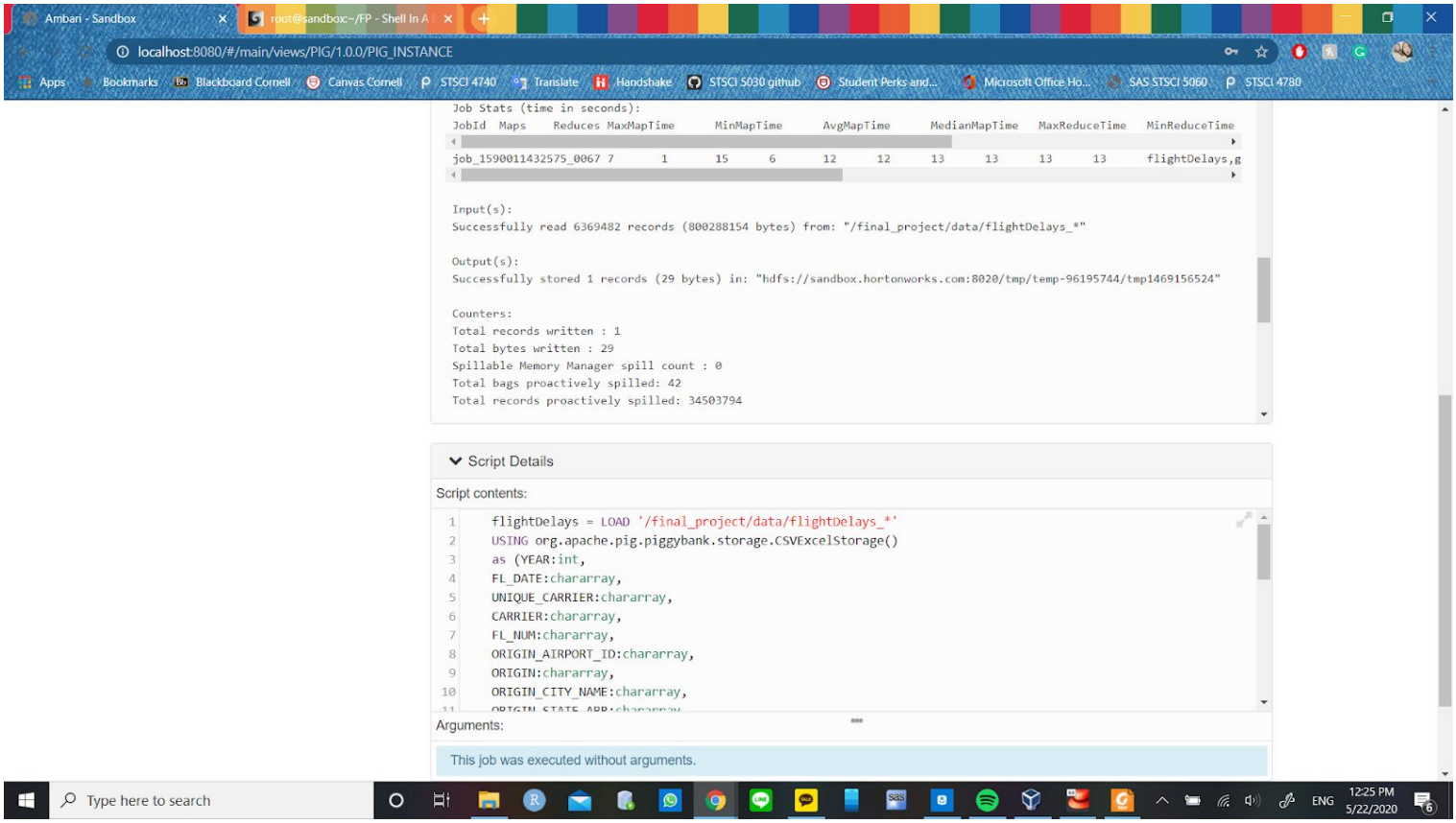


7.

```
flightDelays = LOAD '/final_project/data/flightDelays_*'
USING org.apache.pig.piggybank.storage.CSVExcelStorage()
as (YEAR:int,
FL_DATE:chararray,
UNIQUE_CARRIER:chararray,
CARRIER:chararray,
FL_NUM:chararray,
ORIGIN_AIRPORT_ID:chararray,
ORIGIN:chararray,
ORIGIN_CITY_NAME:chararray,
ORIGIN_STATE_ABR:chararray,
DEST_AIRPORT_ID:chararray,
DEST:chararray,
DEST_CITY_NAME:chararray,
DEST_STATE_ABR:chararray,
DEP_DELAY_NEW:float,
ARR_DELAY:float,
ARR_DELAY_NEW:float,
CARRIER_DELAY:float,
WEATHER_DELAY:float,
```

```
NAS_DELAY:float,
SECURITY_DELAY:float,
LATE_AIRCRAFT_DELAY:float);

wanted = foreach flightDelays generate CARRIER_DELAY, WEATHER_DELAY,
NAS_DELAY,
SECURITY_DELAY, LATE_AIRCRAFT_DELAY;
grp = group wanted all;
longDelays = foreach grp generate
MAX(wanted.CARRIER_DELAY),
MAX(wanted.WEATHER_DELAY),
MAX(wanted.NAS_DELAY),
MAX(wanted.SECURITY_DELAY),
MAX(wanted.LATE_AIRCRAFT_DELAY);
dump longDelays;
```





8.

flight_delays_udf.pig

```
flightDelays = LOAD '/final_project/data/flightDelays_*'
USING org.apache.pig.piggybank.storage.CSVExcelStorage()
as (YEAR:int,
FL_DATE:chararray,
UNIQUE_CARRIER:chararray,
CARRIER:chararray,
FL_NUM:chararray,
ORIGIN_AIRPORT_ID:chararray,
ORIGIN:chararray,
ORIGIN_CITY_NAME:chararray,
ORIGIN_STATE_ABR:chararray,
DEST_AIRPORT_ID:chararray,
DEST:chararray,
DEST_CITY_NAME:chararray,
DEST_STATE_ABR:chararray,
DEP_DELAY_NEW:float,
ARR_DELAY:float,
ARR_DELAY_NEW:float,
```

```
CARRIER_DELAY:float,
WEATHER_DELAY:float,
NAS_DELAY:float,
SECURITY_DELAY:float,
LATE_AIRCRAFT_DELAY:float);

REGISTER 'hdfs://sandbox.hortonworks.com:8020/final_project/flight_delay_udf.py'
using jython as myudf;
clean_c = filter flightDelays by CARRIER_DELAY == 1975.0;
max_c_delay = foreach clean_c generate myudf.get_max((($0,'YEAR'),($1,'FL_DATE'),
($2,'UNIQUE_CARRIER'),($3,'CARRIER'),($4,'FL_NUM'),($5,'ORIGIN_AIRPORT_ID'),
($6, 'ORIGIN'),
($7,'ORIGIN_CITY_NAME'),($8,'ORIGIN_STATE_ABR'),($9,'DEST_AIRPORT_ID'),
($10,'DEST'),($11,'DEST_CITY_NAME'),($12,'DEST_STATE_ABR'),($13,'DEP_DELAY
_NEW'),
($14,'ARR_DELAY'),($15,'ARR_DELAY_NEW'),($16,'CARRIER_DELAY'),($17,'WEATH
ER_DELAY'),
($18,'NAS_DELAY'),($19,'SECURITY_DELAY'),($20,'LATE_AIRCRAFT_DELAY')));
dump max_c_delay;
```

flight_delay_udf.py

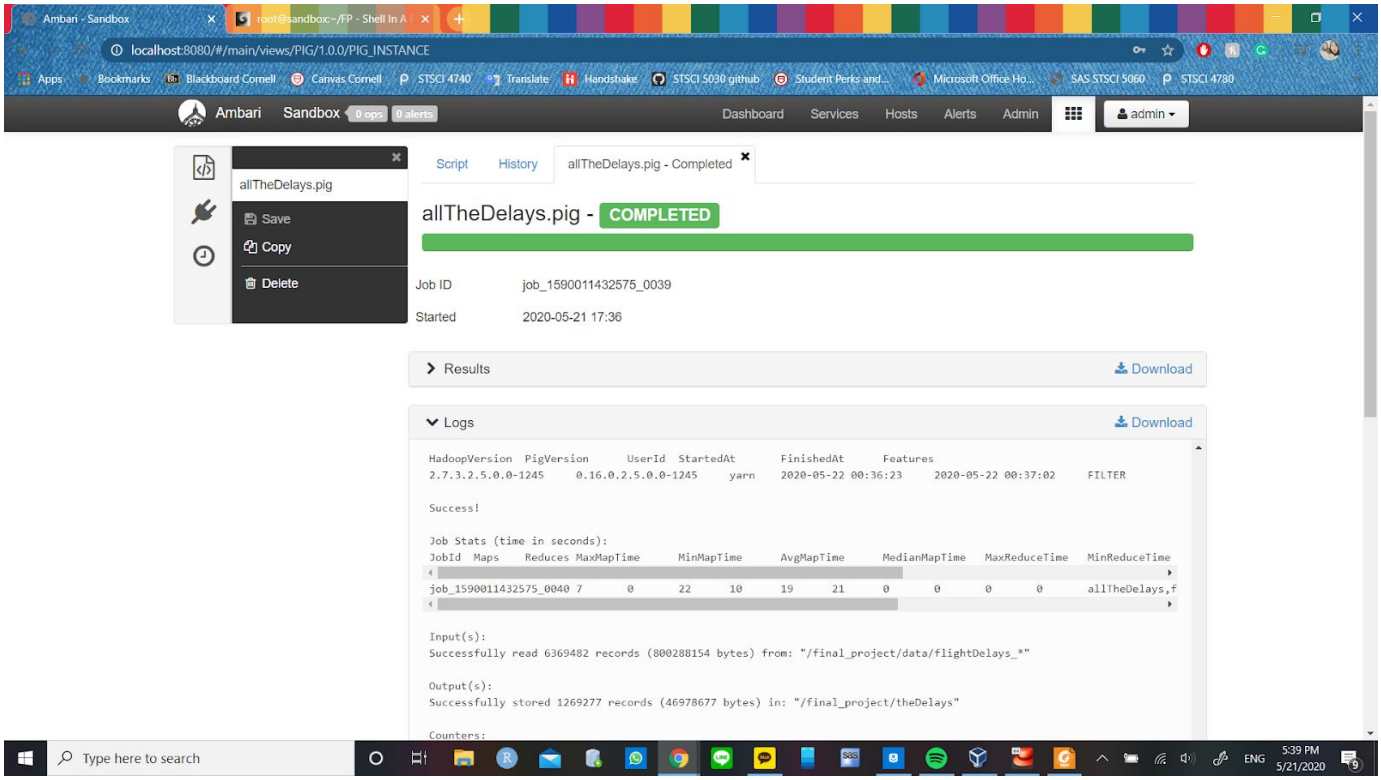
```
@outputSchema("dly:chararray")
def get_max(data):
   maxdelay = max(data[16][0])
   carrier = data[16][1]
   dly = 'The maximum $s is $.1f. The details of the delay are
as follows: \n \
%s:%d \n \
%s:%s \n \
%s:%s \n \
%s:%s \n \
%s:%s \n \
%s:%s \n \
%s:%s \n \
%s:%s \n \
%s:%s \n \
%s:%s \n \
%s:%s \n \
%s:%s \n \
%s:%.1f \n \
%s:%.1f \n \
%s:%.1f \n \
```

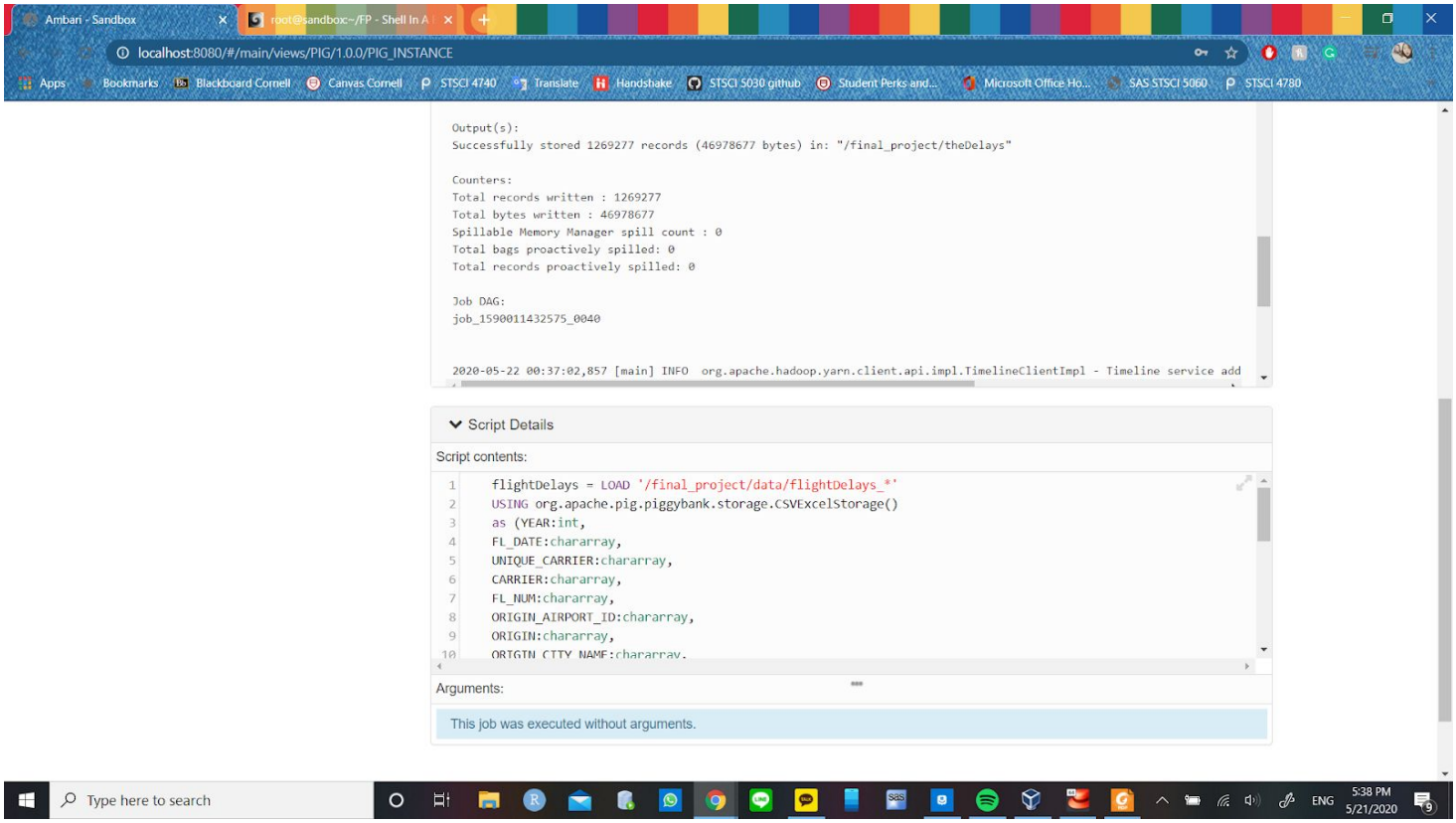
```
%s:%.1f \n \
%s:%.1f \n \
%s:%.1f \n \
%s:%.1f \n \
%s:%.1f \n' %(carrier,
maxdelay,data[0][1],data[0][0],data[1][1],data[1][0],data[2]
[1],data[2][0],data[3][1],data[3][0],data[4][1],data[4][0],data
[5][1],data[5][0],data[6][1],data[6][0],data[7][1],data[7][0],d
ata[8][1],data[8][0],data[9][1],data[9][0],data[10][1],data[10]
[0],data[11][1],data[11][0],data[12][1],data[12][0],data[13][1]
,data[13][0],data[14][1],data[14][0],data[15][1],data[15][0],da
ta[16][1],data[16][0],data[17][1],data[17][0],data[18][1],data[
18][0],data[19][1],data[19][0],data[20][1],data[20][0])
   return dly
```

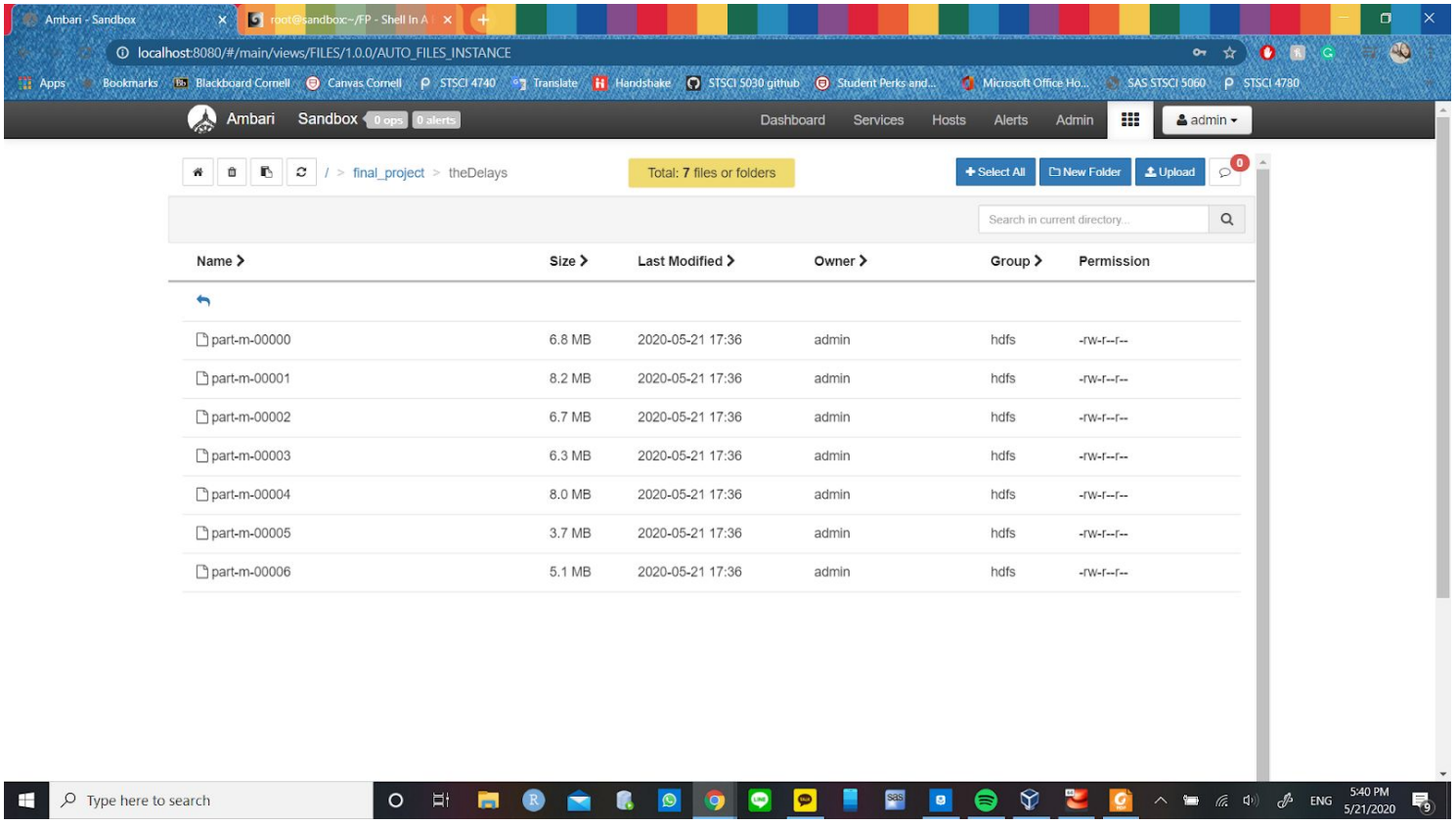
9.

```
flightDelays = LOAD '/final_project/data/flightDelays_*'
USING org.apache.pig.piggybank.storage.CSVExcelStorage()
as (YEAR:int,
FL_DATE:chararray,
UNIQUE_CARRIER:chararray,
CARRIER:chararray,
FL_NUM:chararray,
ORIGIN_AIRPORT_ID:chararray,
ORIGIN:chararray,
ORIGIN_CITY_NAME:chararray,
ORIGIN_STATE_ABR:chararray,
DEST_AIRPORT_ID:chararray,
DEST:chararray,
DEST_CITY_NAME:chararray,
DEST_STATE_ABR:chararray,
DEP_DELAY_NEW:float,
ARR_DELAY:float,
ARR_DELAY_NEW:float,
CARRIER_DELAY:float,
WEATHER_DELAY:float,
NAS_DELAY:float,
SECURITY_DELAY:float,
LATE_AIRCRAFT_DELAY:float);
allTheDelays = foreach flightDelays generate  FL_DATE, FL_NUM, CARRIER_DELAY,
WEATHER_DELAY, NAS_DELAY, SECURITY_DELAY, LATE_AIRCRAFT_DELAY;
```

```
theDelays = filter allTheDelays by NOT (CARRIER_DELAY is null or
WEATHER_DELAY is null or NAS_DELAY is null or SECURITY_DELAY is null or
LATE_AIRCRAFT_DELAY is null);
STORE theDelays INTO '/final_project/theDelays';
```

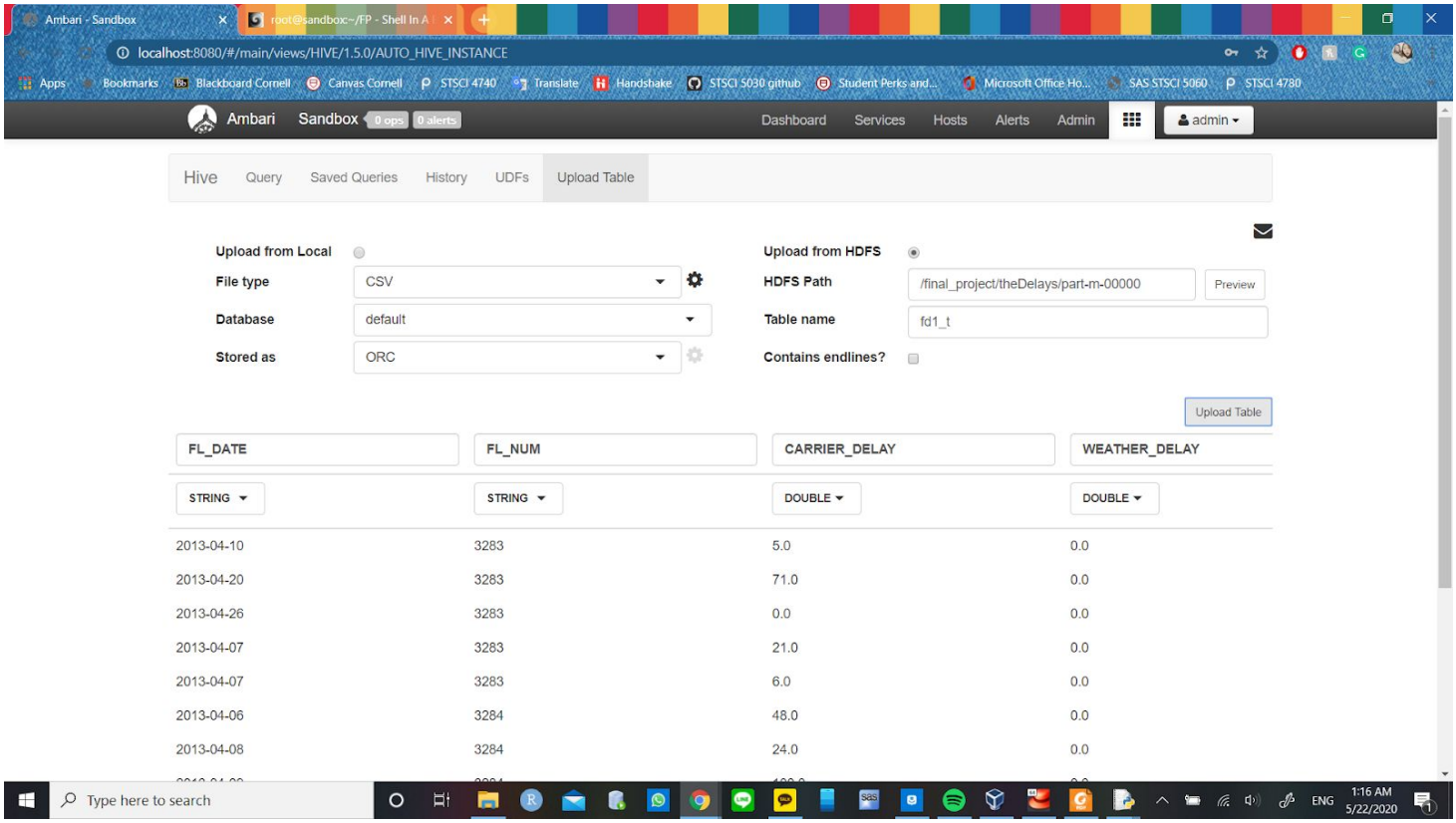


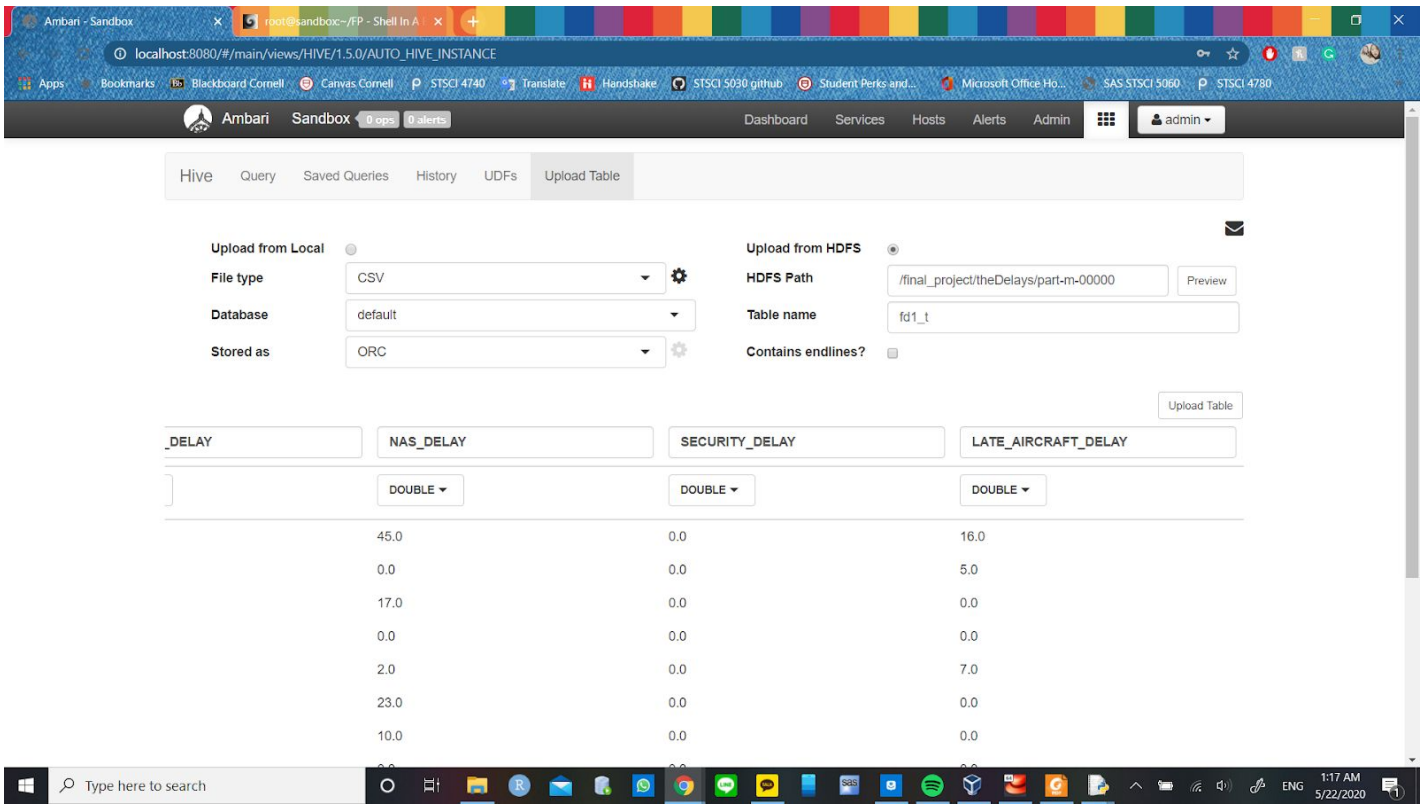

Ambari - Sandbox
root@sandbox:~/FP - Shell In A
localhost:8080/#/main/views/PIG/1.0.0/PIG_INSTANCE
Apps
Bookmarks
Blackboard Cornell
Canvas Cornell
STSCI 4740
Translate
Handshake
STSCI 5030 github
Student Perks and...
Microsoft Office Ho...
SAS STSCI 5060
STSCI 4780
Output(s):
Successfully stored 1269277 records (46978677 bytes) in: "/final_project/theDelays"
Counters:
Total records written : 1269277
Total bytes written : 46978677
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_1590011432575_0040
2020-05-22 00:37:02,857 [main] INFO org.apache.hadoop.yarn.client.api.impl.TimelineClientImpl - Timeline service add
Script Details
Script contents:
1 flightDelays = LOAD '/final_project/data/flightDelays_*'
2 USING org.apache.pig.piggybank.storage.CSVExcelStorage()
3 as (YEAR:int,
4 FL_DATE:chararray,
5 UNIQUE_CARRIER:chararray,
6 CARRIER:chararray,
7 FL_NUM:chararray,
8 ORIGIN_AIRPORT_ID:chararray,
9 ORIGIN:chararray,
10 ORIGIN_CITY_NAME:chararray,
Arguments:
This job was executed without arguments.
Type here to search
ENG
5:38 PM
5/21/2020


Ambari - Sandbox
root@sandbox:~/FP - Shell In A
localhost:8080/#/main/views/FILES/1.0.0/AUTO_FILES_INSTANCE
Apps
Bookmarks
Blackboard Cornell
Canvas Cornell
STSCI 4740
Translate
Handshake
STSCI 5030 github
Student Perks and...
Microsoft Office Ho...
SAS STSCI 5060
STSCI 4780
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
admin
/ > final_project > theDelays
Total: 7 files or folders
+ Select All
New Folder
Upload
Search in current directory...
Name
Size
Last Modified
Owner
Group
Permission
part-m-00000 6.8 MB 2020-05-21 17:36 admin hdfs -rw-r--r--
part-m-00001 8.2 MB 2020-05-21 17:36 admin hdfs -rw-r--r--
part-m-00002 6.7 MB 2020-05-21 17:36 admin hdfs -rw-r--r--
part-m-00003 6.3 MB 2020-05-21 17:36 admin hdfs -rw-r--r--
part-m-00004 8.0 MB 2020-05-21 17:36 admin hdfs -rw-r--r--
part-m-00005 3.7 MB 2020-05-21 17:36 admin hdfs -rw-r--r--
part-m-00006 5.1 MB 2020-05-21 17:36 admin hdfs -rw-r--r--
Type here to search
ENG
5:40 PM
5/21/2020

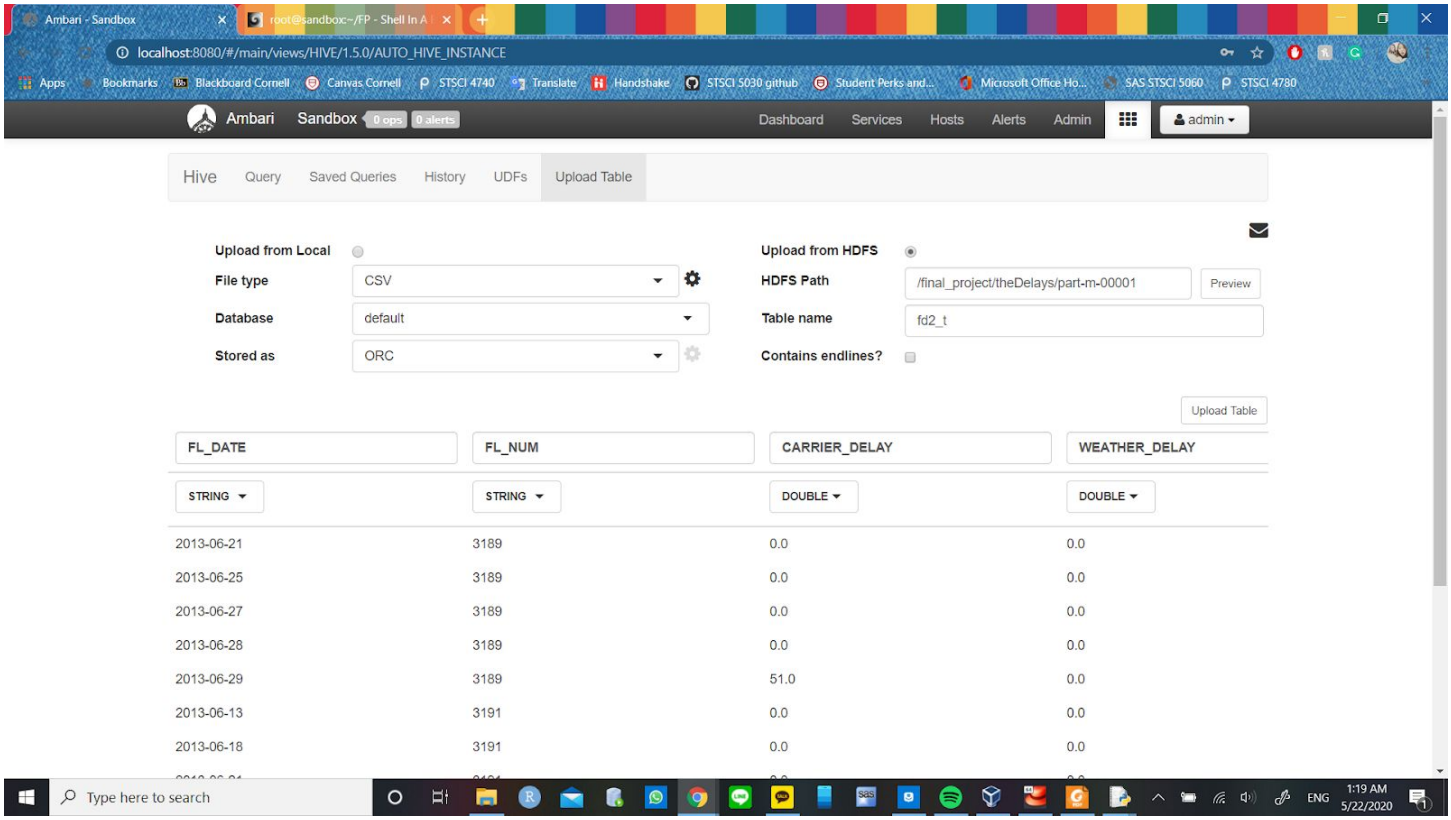
10.

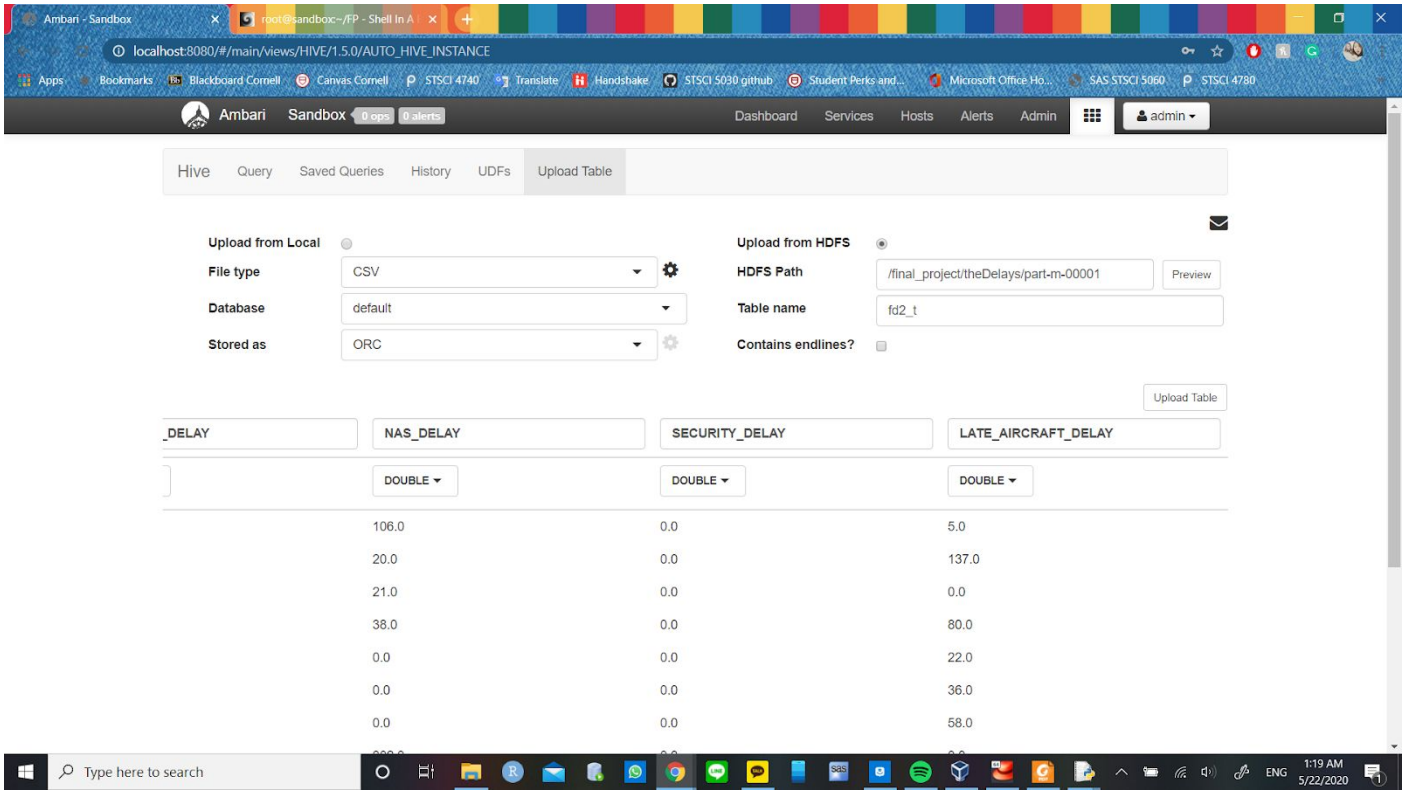
a. In default database fd1_t:





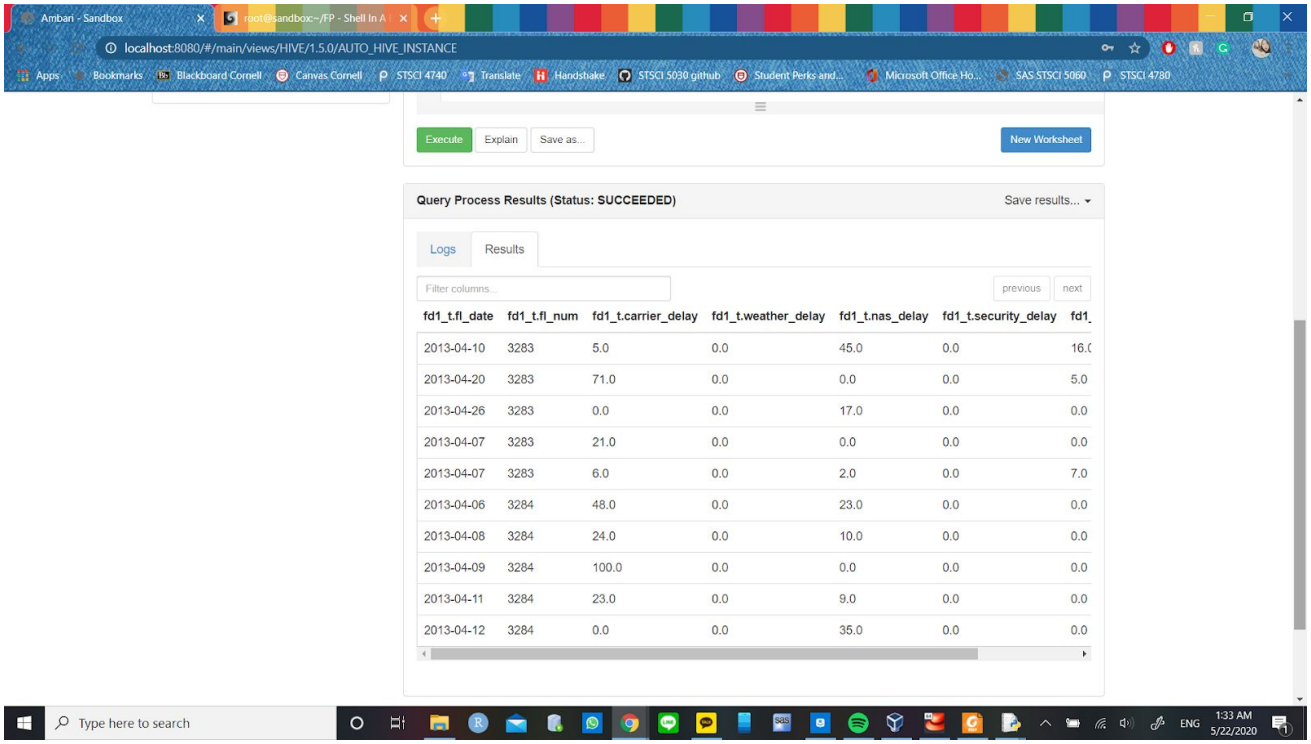
fd2_t:

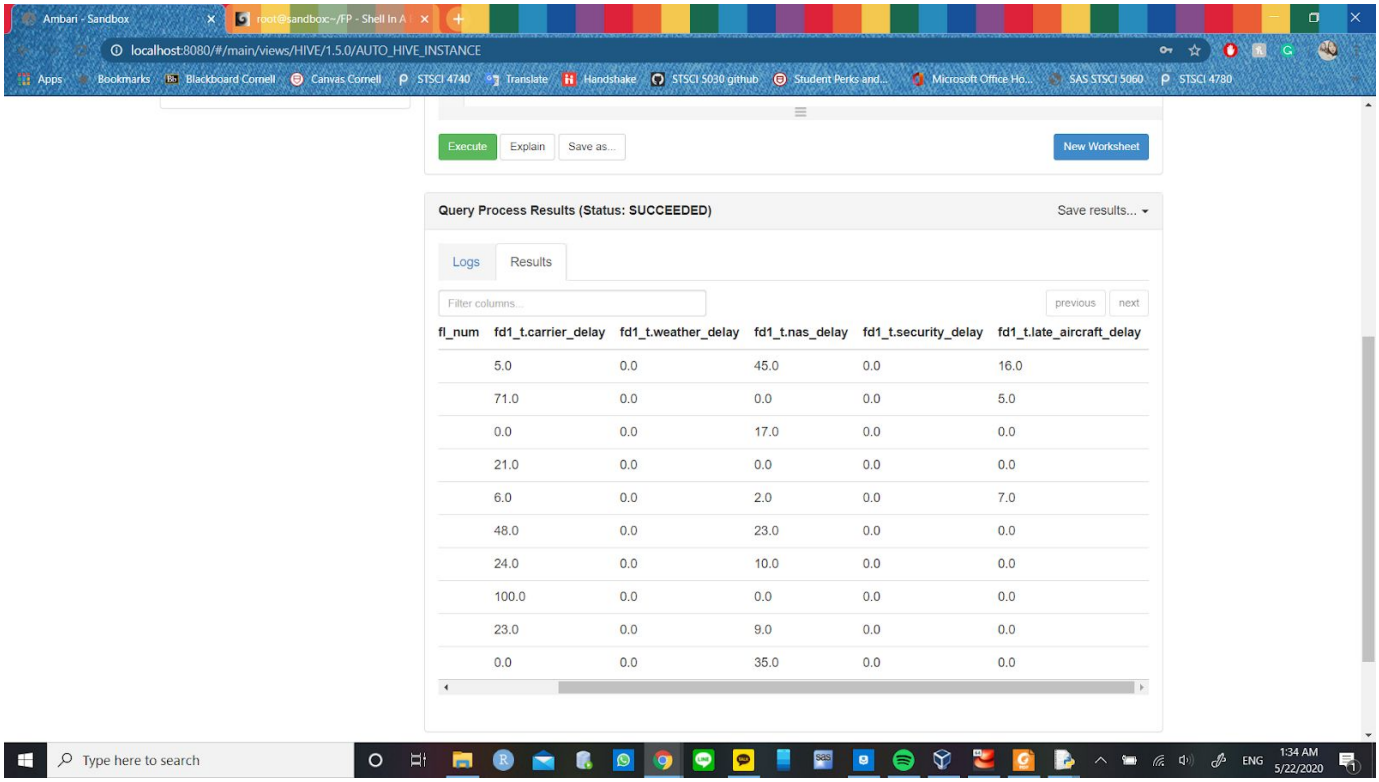

Ambari - Sandbox
localhost:8080/#/main/views/HIVE/1.5.0/AUTO_HIVE_INSTANCE
Ambari Sandbox 0 ops 0 alerts
Dashboard Services Hosts Alerts Admin admin
Hive Query Saved Queries History UDFs Upload Table
Upload from Local
File type CSV
Database default
Stored as ORC
Upload from HDFS
HDFS Path /final_project/theDelays/part-m-00001
Preview
Table name fd2_t
Contains endlines?
Upload Table
FL_DATE FL_NUM CARRIER_DELAY WEATHER_DELAY
STRING STRING DOUBLE DOUBLE
2013-06-21 3189 0.0 0.0
2013-06-25 3189 0.0 0.0
2013-06-27 3189 0.0 0.0
2013-06-28 3189 0.0 0.0
2013-06-29 3189 51.0 0.0
2013-06-13 3191 0.0 0.0
2013-06-18 3191 0.0 0.0
Type here to search
ENG 1:19 AM 5/22/2020


Ambari - Sandbox
localhost:8080/#/main/views/HIVE/1.5.0/AUTO_HIVE_INSTANCE
Ambari Sandbox 0 ops 0 alerts
Dashboard Services Hosts Alerts Admin admin
Hive Query Saved Queries History UDFs Upload Table
Upload from Local
File type CSV
Database default
Stored as ORC
Upload from HDFS
HDFS Path /final_project/theDelays/part-m-00001
Preview
Table name fd2_t
Contains endlines?
Upload Table
_DELAY NAS_DELAY SECURITY_DELAY LATE_AIRCRAFT_DELAY
DOUBLE DOUBLE DOUBLE
106.0 0.0 5.0
20.0 0.0 137.0
21.0 0.0 0.0
38.0 0.0 80.0
0.0 0.0 22.0
0.0 0.0 36.0
0.0 0.0 58.0
Type here to search
ENG 1:19 AM 5/22/2020

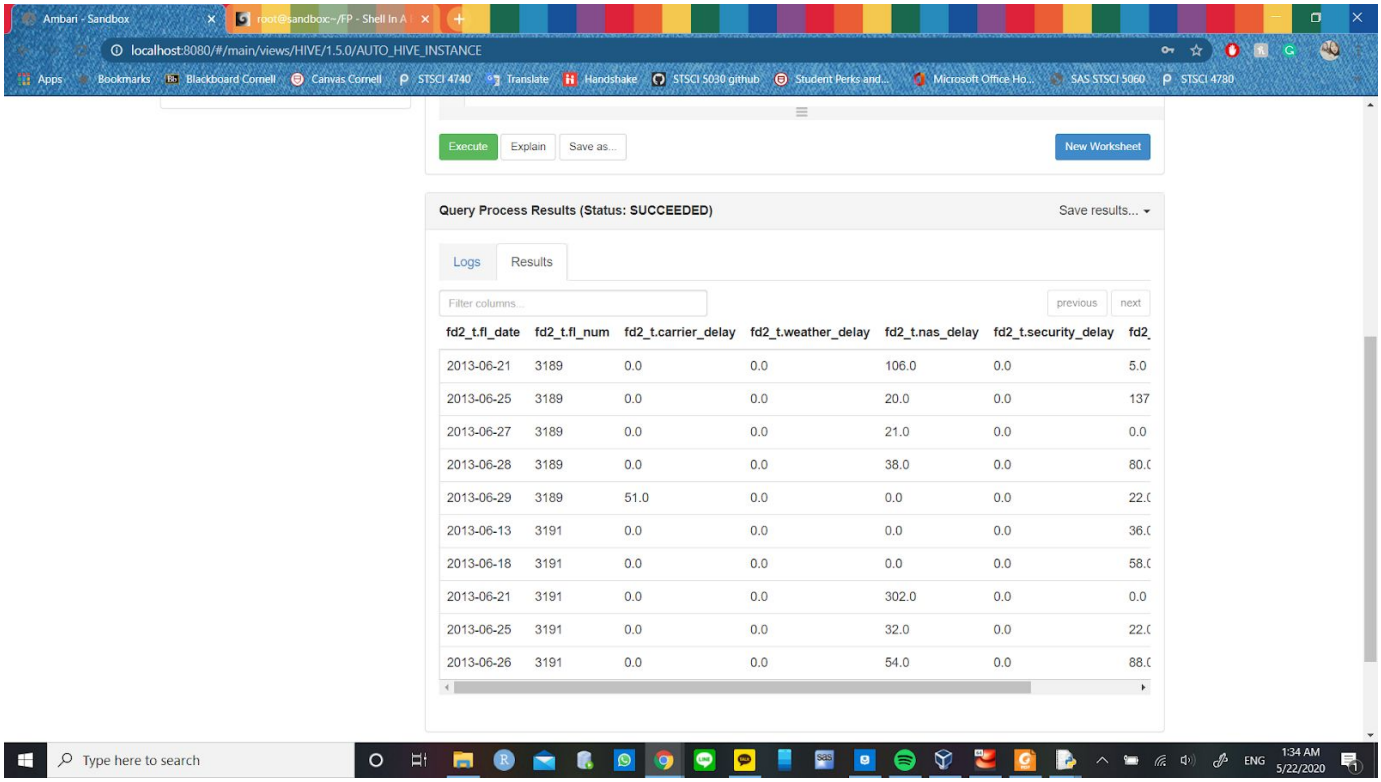
In FPdb database:

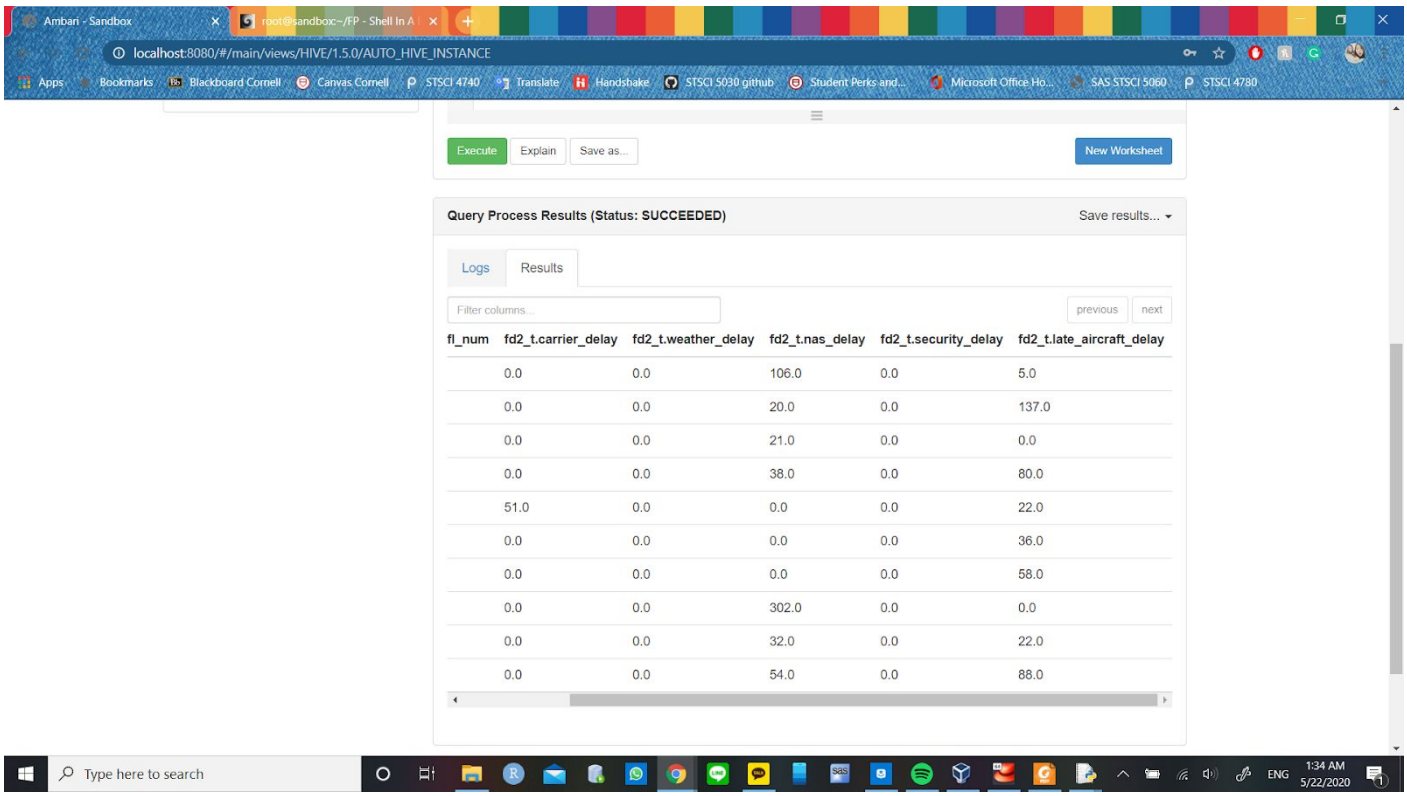
SELECT * FROM fd1_t LIMIT 10;





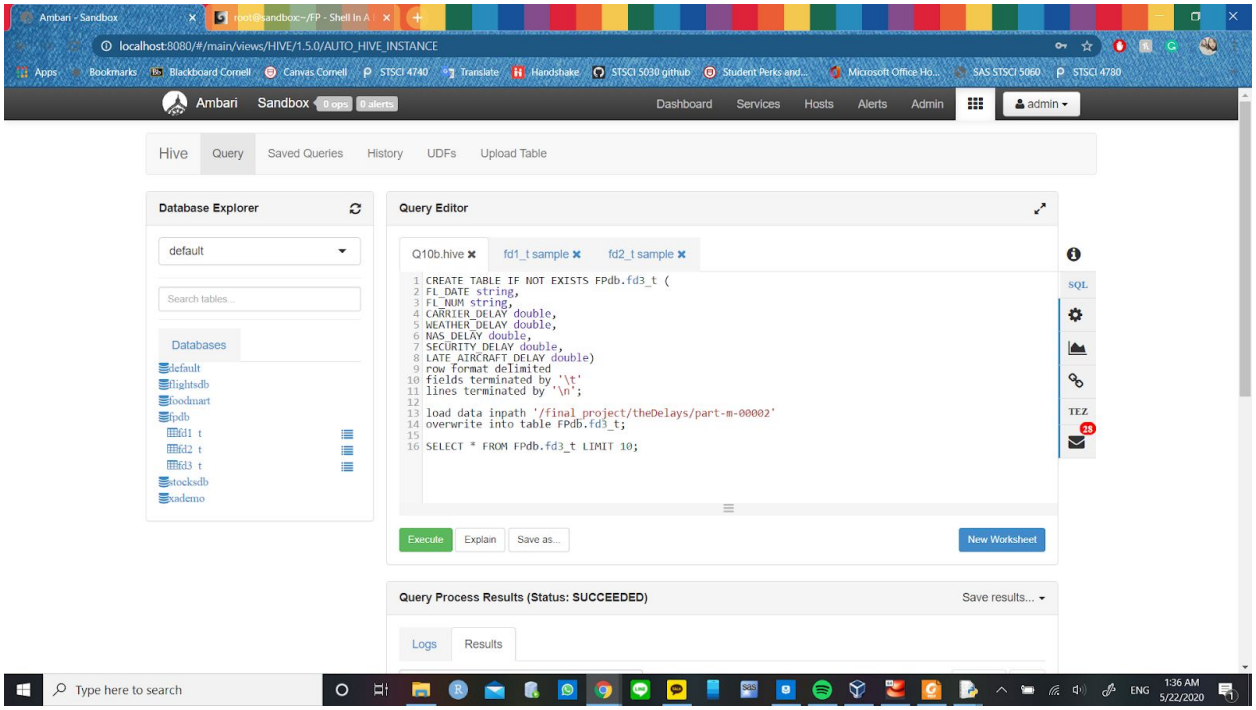
SELECT * FROM fd2_t LIMIT 10;

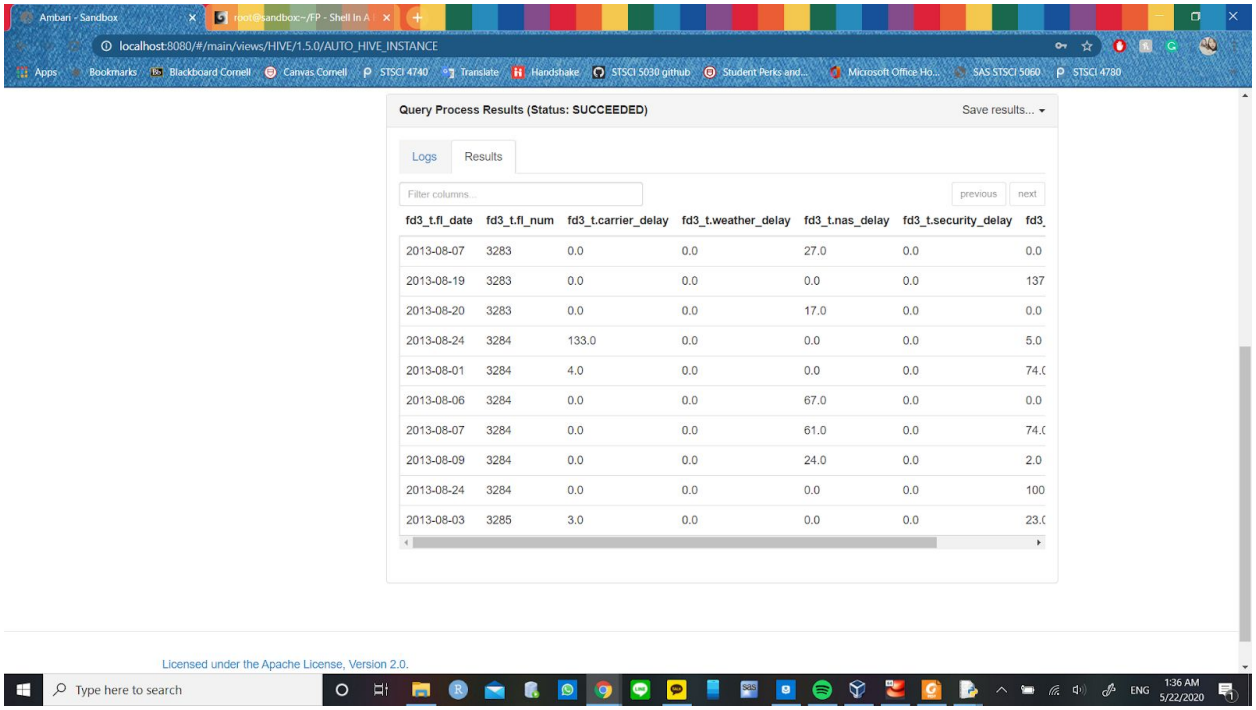


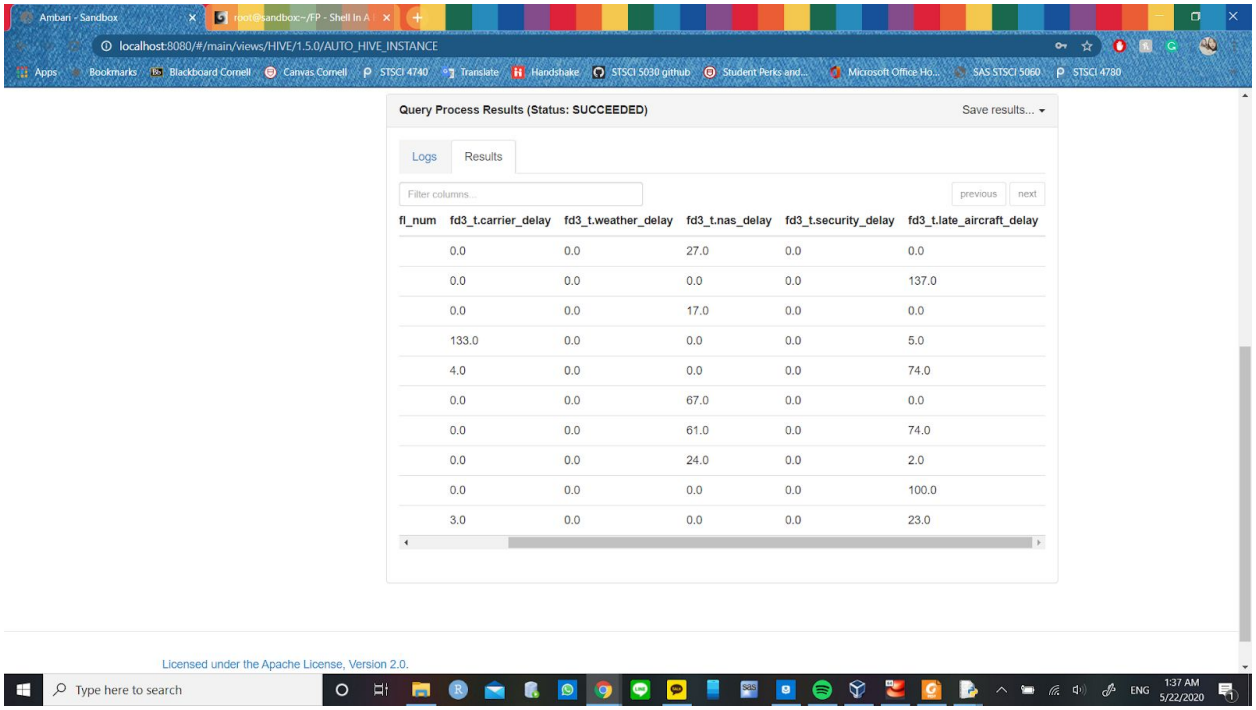


b.

```
CREATE TABLE IF NOT EXISTS FPdb.fd3_t (
FL_DATE string,
FL_NUM string,
CARRIER_DELAY double,
WEATHER_DELAY double,
NAS_DELAY double,
SECURITY_DELAY double,
LATE_AIRCRAFT_DELAY double)
row format delimited
fields terminated by '\t'
lines terminated by '\n';

load data inpath '/final_project/theDelays/part-m-00002'
overwrite into table FPdb.fd3_t;

SELECT * FROM FPdb.fd3_t LIMIT 10;
```


Ambari - Sandbox
root@sandbox:~/FP - Shell In A
localhost:8080/#/main/views/HIVE/1.5.0/AUTO_HIVE_INSTANCE
Ambari
Sandbox
Dashboard
Services
Hosts
Alerts
Admin
admin
Hive
Query
Saved Queries
History
UDFs
Upload Table
Database Explorer
default
Search tables...
Databases
default
flightsdb
foodmart
fpdb
fd1_t
fd2_t
fd3_t
stocksdb
xademo
Query Editor
Q10b.hive
fd1_t sample
fd2_t sample
CREATE TABLE IF NOT EXISTS FPdb.fd3_t (
FL_DATE string,
FL_NUM string,
CARRIER_DELAY double,
WEATHER_DELAY double,
NAS_DELAY double,
SECURITY_DELAY double,
LATE_AIRCRAFT_DELAY double)
row format delimited
fields terminated by '\t'
lines terminated by '\n';
load data inpath '/final_project/theDelays/part-m-00002'
overwrite into table FPdb.fd3_t;
SELECT * FROM FPdb.fd3_t LIMIT 10;
SQL
TEZ
Execute
Explain
Save as...
New Worksheet
Query Process Results (Status: SUCCEEDED)
Save results...
Logs
Results
Type here to search
ENG
1:36 AM
5/22/2020


Ambari - Sandbox
localhost:8080/#/main/views/HIVE/1.5.0/AUTO_HIVE_INSTANCE
Query Process Results (Status: SUCCEEDED)
Save results...
Logs
Results
Filter columns...
previous
next
fd3_t.fl_date fd3_t.fl_num fd3_t.carrier_delay fd3_t.weather_delay fd3_t.nas_delay fd3_t.security_delay fd3_
2013-08-07 3283 0.0 0.0 27.0 0.0 0.0
2013-08-19 3283 0.0 0.0 0.0 0.0 137
2013-08-20 3283 0.0 0.0 17.0 0.0 0.0
2013-08-24 3284 133.0 0.0 0.0 0.0 5.0
2013-08-01 3284 4.0 0.0 0.0 0.0 74.(
2013-08-06 3284 0.0 0.0 67.0 0.0 0.0
2013-08-07 3284 0.0 0.0 61.0 0.0 74.(
2013-08-09 3284 0.0 0.0 24.0 0.0 2.0
2013-08-24 3284 0.0 0.0 0.0 0.0 100
2013-08-03 3285 3.0 0.0 0.0 0.0 23.(
Licensed under the Apache License, Version 2.0.
Type here to search
ENG
1:36 AM
5/22/2020

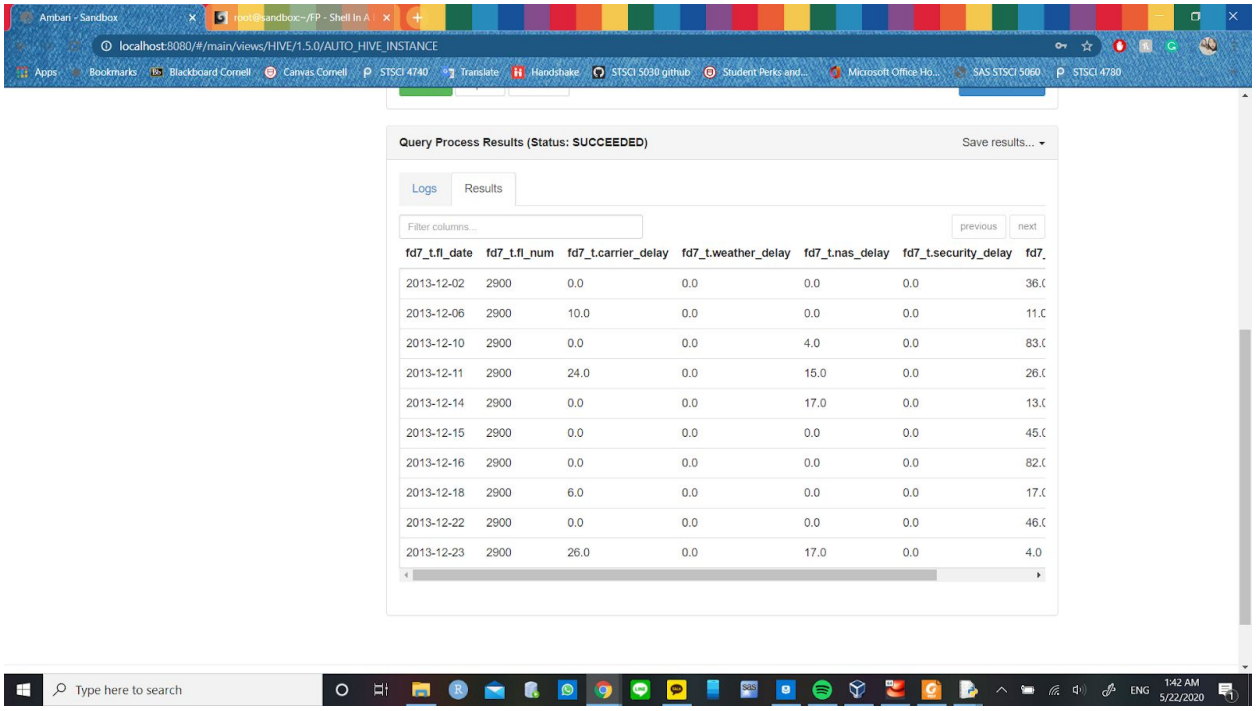


c.

```
create table FPdb.fd4_t
LIKE FPdb.fd3_t;

load data inpath '/final_project/theDelays/part-m-00003'
overwrite into table FPdb.fd4_t;


create table FPdb.fd5_t
LIKE FPdb.fd3_t;

load data inpath '/final_project/theDelays/part-m-00004'
overwrite into table FPdb.fd5_t;


create table FPdb.fd6_t
LIKE FPdb.fd3_t;

load data inpath '/final_project/theDelays/part-m-00005'
overwrite into table FPdb.fd6_t;


create table FPdb.fd7_t
LIKE FPdb.fd3_t;
```

```
load data inpath '/final_project/theDelays/part-m-00006'
overwrite into table FPdb.fd7_t;
```

```
SELECT * FROM FPdb.fd7_t LIMIT 10;
```

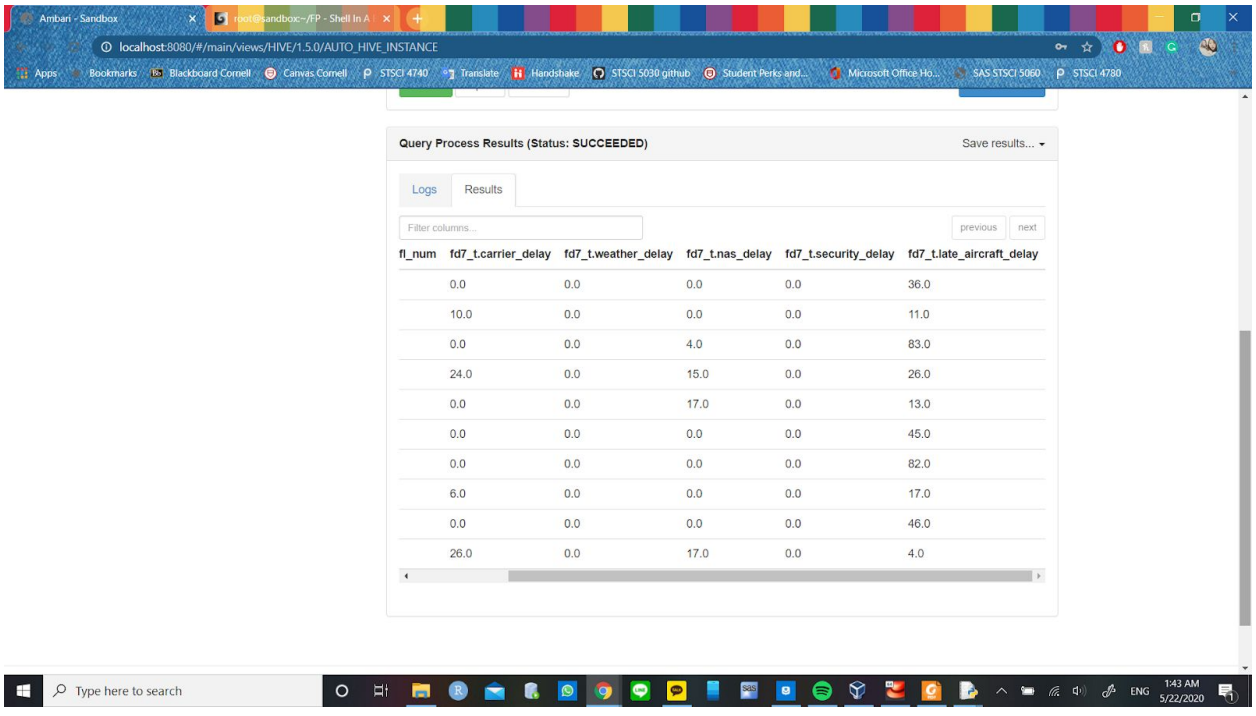


Query Process Results (Status: SUCCEEDED) — Save results...

Logs | Results

Filter columns... previous next

fd7_t.fl_date	fd7_t.fl_num	fd7_t.carrier_delay	fd7_t.weather_delay	fd7_t.nas_delay	fd7_t.security_delay	fd7_
2013-12-02	2900	0.0	0.0	0.0	0.0	36.0
2013-12-06	2900	10.0	0.0	0.0	0.0	11.0
2013-12-10	2900	0.0	0.0	4.0	0.0	83.0
2013-12-11	2900	24.0	0.0	15.0	0.0	26.0
2013-12-14	2900	0.0	0.0	17.0	0.0	13.0
2013-12-15	2900	0.0	0.0	0.0	0.0	45.0
2013-12-16	2900	0.0	0.0	0.0	0.0	82.0
2013-12-18	2900	6.0	0.0	0.0	0.0	17.0
2013-12-22	2900	0.0	0.0	0.0	0.0	46.0
2013-12-23	2900	26.0	0.0	17.0	0.0	4.0



Query Process Results (Status: SUCCEEDED) — Save results...

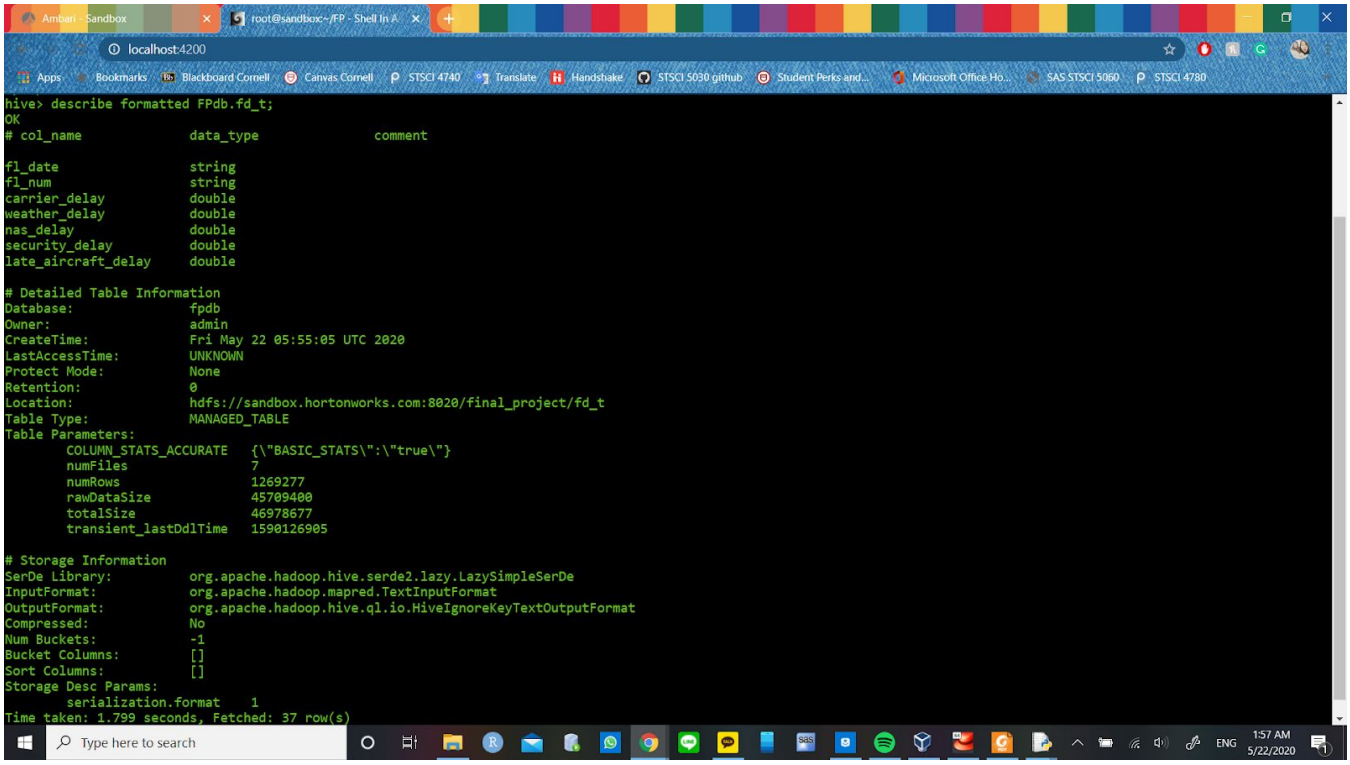
Logs | Results

Filter columns... previous next

fl_num	fd7_t.carrier_delay	fd7_t.weather_delay	fd7_t.nas_delay	fd7_t.security_delay	fd7_t.late_aircraft_delay
	0.0	0.0	0.0	0.0	36.0
	10.0	0.0	0.0	0.0	11.0
	0.0	0.0	4.0	0.0	83.0
	24.0	0.0	15.0	0.0	26.0
	0.0	0.0	17.0	0.0	13.0
	0.0	0.0	0.0	0.0	45.0
	0.0	0.0	0.0	0.0	82.0
	6.0	0.0	0.0	0.0	17.0
	0.0	0.0	0.0	0.0	46.0
	26.0	0.0	17.0	0.0	4.0

11.

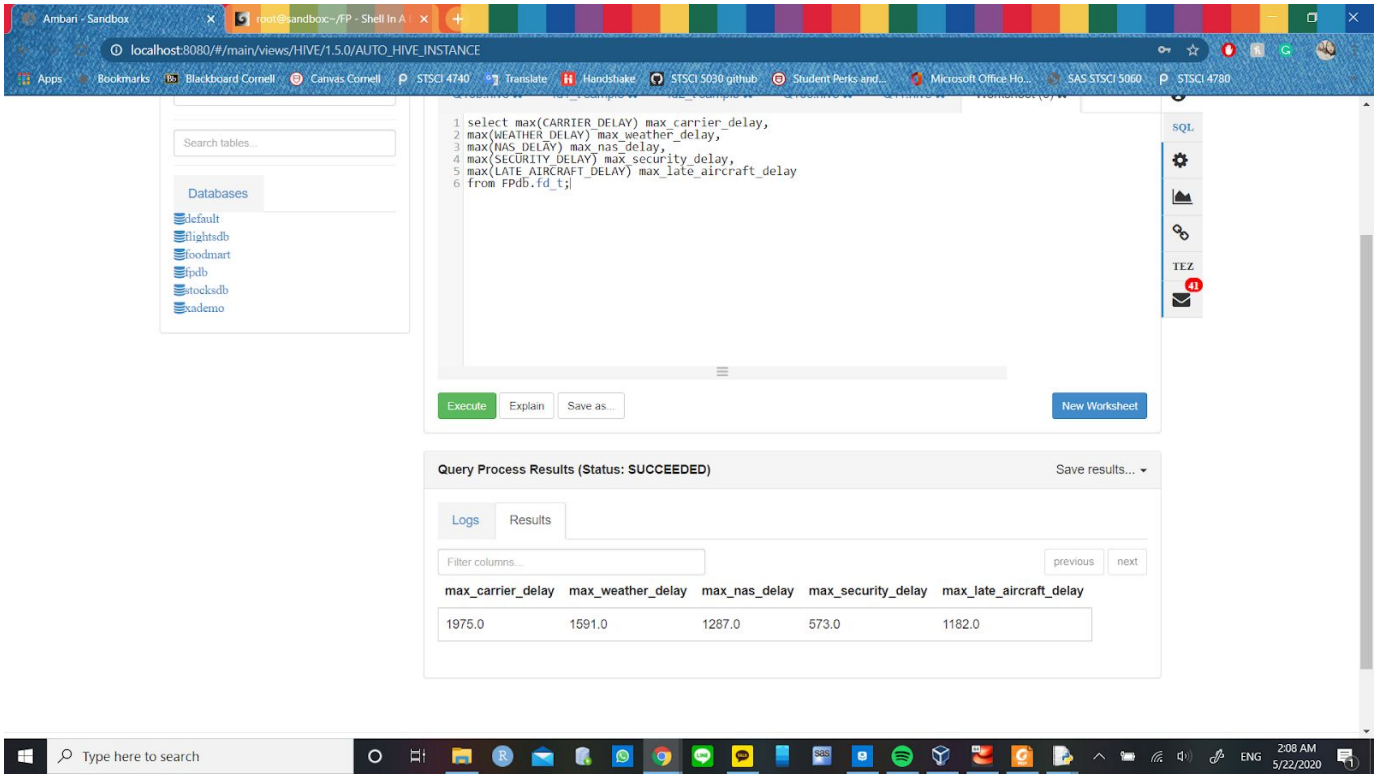
```
CREATE TABLE IF NOT EXISTS FPdb.fd_t AS
SELECT * FROM FPdb.fd1_t
UNION ALL
SELECT * FROM FPdb.fd2_t
UNION ALL
SELECT * FROM FPdb.fd3_t
UNION ALL
SELECT * FROM FPdb.fd4_t
UNION ALL
SELECT * FROM FPdb.fd5_t
UNION ALL
SELECT * FROM FPdb.fd6_t
UNION ALL
SELECT * FROM FPdb.fd7_t;
```



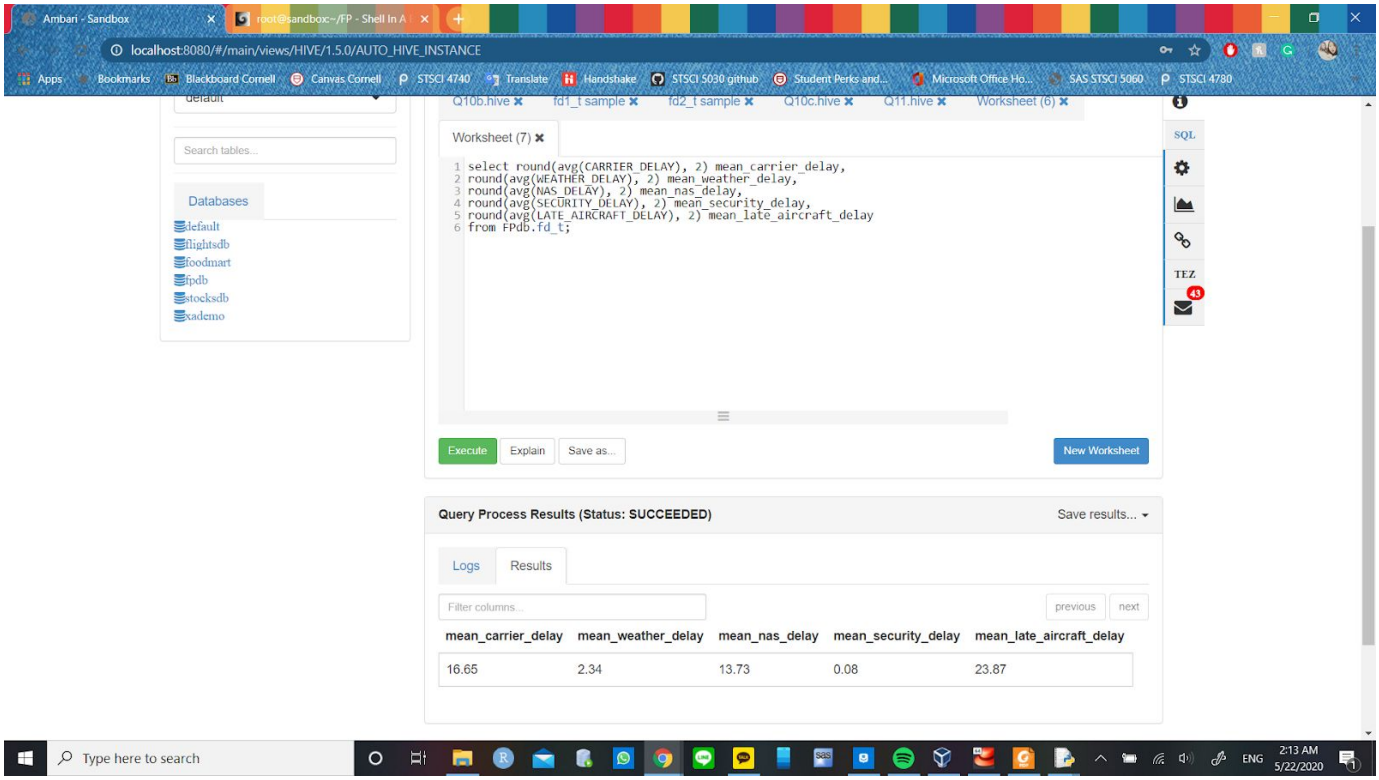
12.

```
select max(CARRIER_DELAY) max_carrier_delay,
max(WEATHER_DELAY) max_weather_delay,
max(NAS_DELAY) max_nas_delay,
max(SECURITY_DELAY) max_security_delay,
```

```
max(LATE_AIRCRAFT_DELAY) max_late_aircraft_delay
from FPdb.fd_t;
```




```
select round(avg(CARRIER_DELAY), 2) mean_carrier_delay,
round(avg(WEATHER_DELAY), 2) mean_weather_delay,
round(avg(NAS_DELAY), 2) mean_nas_delay,
round(avg(SECURITY_DELAY), 2) mean_security_delay,
round(avg(LATE_AIRCRAFT_DELAY), 2) mean_late_aircraft_delay
from FPdb.fd_t;
```



13.

```
hive> use FPdb;

hive> create view averageDelays_v as
    > select round(avg(CARRIER_DELAY), 2) mean_carrier_delay,
    > round(avg(WEATHER_DELAY), 2) mean_weather_delay,
    > round(avg(NAS_DELAY), 2) mean_nas_delay,
    > round(avg(SECURITY_DELAY), 2) mean_security_delay,
    > round(avg(LATE_AIRCRAFT_DELAY), 2)
mean_late_aircraft_delay
    > from fd_t;
```

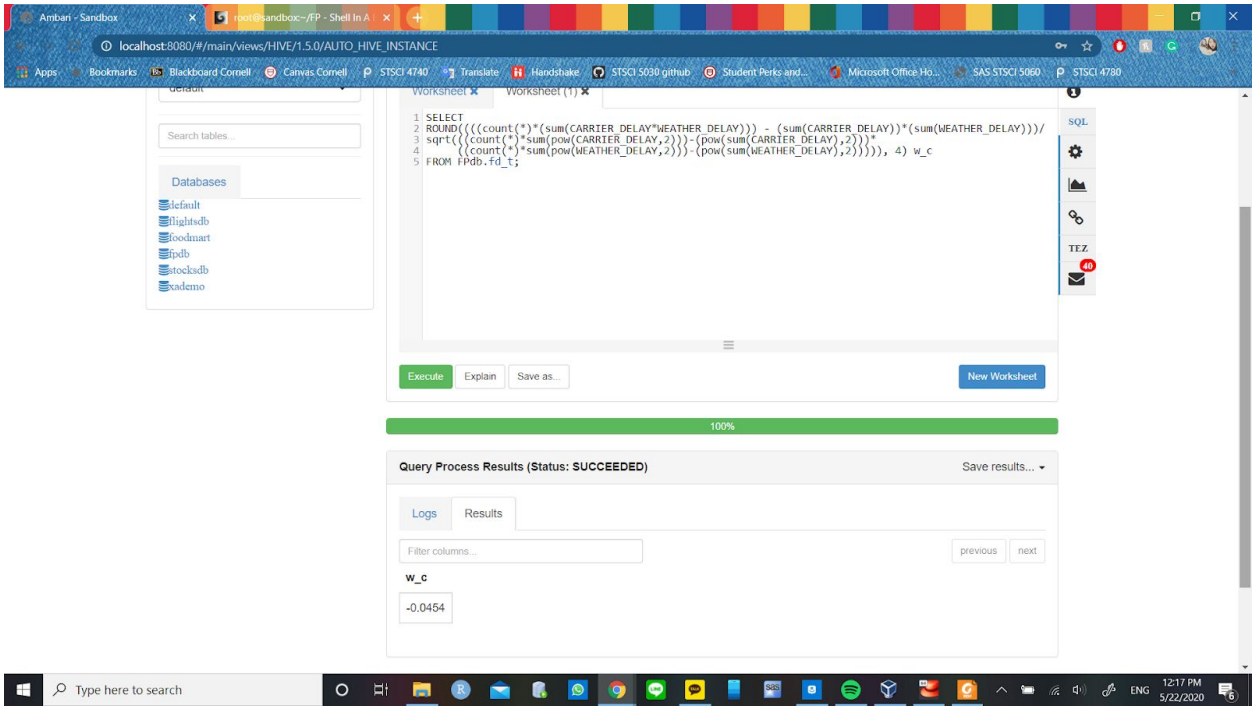
FindMaxAverageDelayType.py

```
#!/usr/bin/python

import sys

def
findMaxAverageDelayType(carrier,weather,nas,security,late_aircr
aft):
```

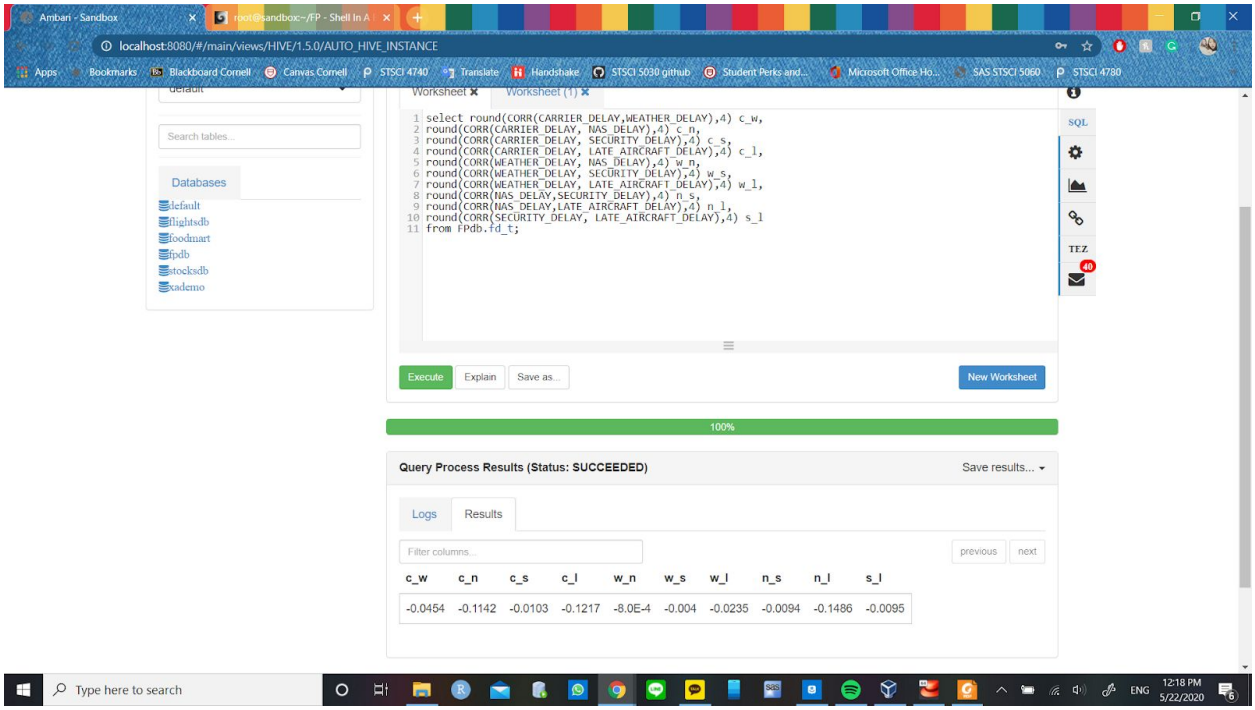
```
 delayTypes = {0:'carrier_delay', 1:'weather_delay',
2:'nas_delay', 3:'security_de
lay', 4:'late_aircraft_delay'}
 delays = [carrier,weather,nas,security,late_aircraft]
 maxAvgValue = max(delays)
 return delayTypes[delays.index(maxAvgValue)], maxAvgValue

for line in sys.STDIN:
 print 'The delay category with the longest average delays is
%s; the average delay
time is %.2f'
%(findMaxAverageDelayType(float(carrier),float(weather),float(n
as),\
float(security),float(late_aircraft))[0],
findMaxAverageDelayType(float(carrier),float(weather),float(nas
),\float(security),float(late_aircraft))[1])

hive> ADD IN /FP/FindMaxAverageDelayType.py;
hive> SELECT
TRANSFORM(mean_carrier_delay,mean_weather_delay,mean_nas_delay,
mean_se
curity_delay, mean_late_aircraft_delay)
   > USING 'python /FP/FindMaxAverageDelayType.py'
   > AS maxAvgDelayType
   > FROM FPdb.averageDelays_v;
```

14.

```
SELECT
ROUND((((count(*)*(sum(CARRIER_DELAY*WEATHER_DELAY))) -
(sum(CARRIER_DELAY))*(sum(WEATHER_DELAY)))/
sqrt(((count(*)*sum(pow(CARRIER_DELAY,2)))-(pow(sum(CARRIER_DELAY),2)))*
((count(*)*sum(pow(WEATHER_DELAY,2)))-(pow(sum(WEATHER_DELAY),2))))), 4)
w_c
FROM FPdb.fd_t;
```




```
select round(CORR(CARRIER_DELAY,WEATHER_DELAY),4) c_w,
round(CORR(CARRIER_DELAY, NAS_DELAY),4) c_n,
round(CORR(CARRIER_DELAY, SECURITY_DELAY),4) c_s,
round(CORR(CARRIER_DELAY, LATE_AIRCRAFT_DELAY),4) c_l,
round(CORR(WEATHER_DELAY, NAS_DELAY),4) w_n,
round(CORR(WEATHER_DELAY, SECURITY_DELAY),4) w_s,
round(CORR(WEATHER_DELAY, LATE_AIRCRAFT_DELAY),4) w_l,
round(CORR(NAS_DELAY,SECURITY_DELAY),4) n_s,
round(CORR(NAS_DELAY,LATE_AIRCRAFT_DELAY),4) n_l,
round(CORR(SECURITY_DELAY, LATE_AIRCRAFT_DELAY),4) s_l
from FPdb.fd_t;
```


Ambari - Sandbox
root@sandbox:~/FP - Shell In A
localhost:8080/#/main/views/HIVE/1.5.0/AUTO_HIVE_INSTANCE
Apps
Bookmarks
Blackboard Cornell
Canvas Cornell
STSCI 4740
Translate
Handshake
STSCI 5030 github
Student Perks and...
Microsoft Office Ho...
SAS STSCI 5060
STSCI 4780
default
Search tables...
Databases
default
flightsdb
foodmart
fpdb
stocksdb
xademo
Worksheet
Worksheet (1)
1 select round(CORR(CARRIER_DELAY,WEATHER_DELAY),4) c_w,
2 round(CORR(CARRIER_DELAY, NAS_DELAY),4) c_n,
3 round(CORR(CARRIER_DELAY, SECURITY_DELAY),4) c_s,
4 round(CORR(CARRIER_DELAY, LATE_AIRCRAFT_DELAY),4) c_l,
5 round(CORR(WEATHER_DELAY, NAS_DELAY),4) w_n,
6 round(CORR(WEATHER_DELAY, SECURITY_DELAY),4) w_s,
7 round(CORR(WEATHER_DELAY, LATE_AIRCRAFT_DELAY),4) w_l,
8 round(CORR(NAS_DELAY,SECURITY_DELAY),4) n_s,
9 round(CORR(NAS_DELAY,LATE_AIRCRAFT_DELAY),4) n_l,
10 round(CORR(SECURITY_DELAY, LATE_AIRCRAFT_DELAY),4) s_l
11 from FPdb.fd_t;
SQL
TEZ
Execute
Explain
Save as...
New Worksheet
100%
Query Process Results (Status: SUCCEEDED)
Save results...
Logs
Results
Filter columns...
previous
next
c_w c_n c_s c_l w_n w_s w_l n_s n_l s_l
-0.0454 -0.1142 -0.0103 -0.1217 -8.0E-4 -0.004 -0.0235 -0.0094 -0.1486 -0.0095
Type here to search
ENG
12:18 PM
5/22/2020