Hive Scripts and Result-Screenshots

1. Create Hive table and some basic query operations.

1. First, create the Hive table using the output txt file. (I first created a test file).

```
create external table test2 (weekName string,timePeriod int,address string,lon
double,lat double,avgVolume double) row format delimited fields terminated by
',' location '/user/xf569/CBDA/test2-output/part-r-00000';
```

2. Some trial commands

```
show tables;
```



```
0: jdbc:hive2://babar.es.its.nyu.edu:10000/> show tables;
INFO  : Compiling command(queryId=hive_20191110191818_88ab19e6-5420-4c98-ad60-8e5bc0f242cb): show tables
INFO  : Semantic Analysis Completed
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:tab_name, type:string, comment:from deserializer)], properties:null)
INFO  : Completed compiling command(queryId=hive_20191110191818_88ab19e6-5420-4c98-ad60-8e5bc0f242cb); Time taken: 0.063 seconds
INFO  : Executing command(queryId=hive_20191110191818_88ab19e6-5420-4c98-ad60-8e5bc0f242cb): show tables
INFO  : Starting task [Stage-0:DDL] in serial mode
INFO  : Completed executing command(queryId=hive_20191110191818_88ab19e6-5420-4c98-ad60-8e5bc0f242cb); Time taken: 0.135 seconds
INFO  : OK
+-----------+--+
| tab_name  |
+-----------+--+
| test2     |
| w1        |
| w10       |
| w11       |
| w3        |
| w30       |
+-----------+--+
6 rows selected (0.205 seconds)
```

```
describe test2;
```

```
+------------+-----------+----------+--+
|  col_name  | data_type | comment  |
+------------+-----------+----------+--+
| weekname   | string    |          |
| timeperiod | int       |          |
| address    | string    |          |
| lon        | double    |          |
| lat        | double    |          |
| avgvolume  | double    |          |
+------------+-----------+----------+--+
```

```
select * from test2;
```

```
+----------------+------------------+-----------------+-------------------+--------------------+------------------+--+
| test2.weekname | test2.timeperiod |  test2.address  |    test2.lon      |     test2.lat      | test2.avgvolume  |
+----------------+------------------+-----------------+-------------------+--------------------+------------------+--+
| 1              | 0                | 1 Ave & E 16 St | 40.73221853       | -73.98165557       | 5.5              |
| 1              | 3                | 8 Ave & W 31 St | 40.7505853470215  | -73.9946848154068  | 2.0              |
| 2              | 10               | E 3 St & 1 Ave  | 40.72467721       | -73.98783413       | 2.0              |
| 4              | 0                | 1 Ave & E 16 St | 40.73221853       | -73.98165557       | 5.0              |
| 4              | 3                | 8 Ave & W 31 St | 40.7505853470215  | -73.9946848154068  | 4.0              |
+----------------+------------------+-----------------+-------------------+--------------------+------------------+--+
5 rows selected (0.77 seconds)
```

```
select * from test2 where weekname = "2" and timeperiod = 10;
```

3. Create the final table for all data.

In Hive command environment:

```
create external table average1 (weekName string,timePeriod int,address
string,lon double,lat double,avgVolume double) row format delimited fields
terminated by ',' location '/user/xf569/CBDA/average-output/';
```



```
select * from average1 limit 20;
```


2. Hive analysis for our project

1. Track the Commuting Time: get insights of NYC office area.

Sample: Mean Value of Weekdays (Mon-Fri) From 7 AM-9 AM:

HiveQL:

```
FROM average1 SELECT address,lon,lat,(sum(avgvolume)/5.0) AS meanAvg WHERE
((weekname != "6" and weekname != "7") and (timeperiod = 7 or timeperiod = 8))
GROUP BY address,lon,lat SORT by meanAvg desc limit 10;
```

It might take a long time to run.

Screenshot of the result:

2. Track the Friday night: get insights of Nightlife/Entertainment are in NYC.

Sample: Friday Night 8PM-9PM

```
select * from average1 where (weekname = "5" and timeperiod = 20) sort by
avgvolume desc limit 10;
```

```
INFO  : Starting Job = job_1569350662793_69694, Tracking URL = http://babar.es.its.nyu.edu:8088/proxy/application_1569350662793_69694/
INFO  : Kill Command = /opt/cloudera/parcels/CDH-5.15.2-1.cdh5.15.2.p0.3/lib/hadoop/bin/hadoop job  -kill job_1569350662793_69694
INFO  : Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
INFO  : 2019-11-30 10:58:52,509 Stage-1 map = 0%,  reduce = 0%
INFO  : 2019-11-30 10:58:58,722 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.73 sec
INFO  : 2019-11-30 10:59:05,897 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 7.37 sec
INFO  : MapReduce Total cumulative CPU time: 7 seconds 370 msec
INFO  : Ended Job = job_1569350662793_69694
INFO  : Launching Job 2 out of 2
INFO  : Starting task [Stage-2: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  : number of splits:1
INFO  : Submitting tokens for job: job_1569350662793_69695
INFO  : The url to track the job: http://babar.es.its.nyu.edu:8088/proxy/application_1569350662793_69695/
INFO  : Starting Job = job_1569350662793_69695, Tracking URL = http://babar.es.its.nyu.edu:8088/proxy/application_1569350662793_69695/
INFO  : Kill Command = /opt/cloudera/parcels/CDH-5.15.2-1.cdh5.15.2.p0.3/lib/hadoop/bin/hadoop job  -kill job_1569350662793_69695
INFO  : Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
INFO  : 2019-11-30 10:59:23,684 Stage-2 map = 0%,  reduce = 0%
INFO  : 2019-11-30 10:59:28,825 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.5 sec
INFO  : 2019-11-30 10:59:36,016 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 5.31 sec
INFO  : MapReduce Total cumulative CPU time: 5 seconds 310 msec
INFO  : Ended Job = job_1569350662793_69695
INFO  : MapReduce Jobs Launched:
INFO  : Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.37 sec   HDFS Read: 5492198 HDFS Write: 707 SUCCESS
INFO  : Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 5.31 sec   HDFS Read: 6634 HDFS Write: 567 SUCCESS
INFO  : Total MapReduce CPU Time Spent: 12 seconds 680 msec
INFO  : Completed executing command(queryId=hive_20191130105858_ac87076f-8bec-4f14-ac9a-93ec0d13f29e); Time taken: 61.528 seconds
INFO  : OK
+------------------+---------------------+-------------------------------+-------------------+----------------------+--------------------+--+
| average1.weekname | average1.timeperiod |        average1.address       |    average1.lon   |     average1.lat     | average1.avgvolume |
+------------------+---------------------+-------------------------------+-------------------+----------------------+--------------------+--+
| 5                | 20                  | E 7 St & Avenue A              | 40.72621788       | -73.98379855         | 40.0               |
| 5                | 20                  | E 39 St & 2 Ave                | 40.74780373       | -73.9734419          | 18.0               |
| 5                | 20                  | E 12 St & Avenue B             | 40.728048571362606 | -73.97881150245667  | 17.8               |
| 5                | 20                  | W 21 St & 6 Ave                | 40.74173969       | -73.99415556         | 17.0               |
| 5                | 20                  | Cooper Square & Astor Pl       | 40.72951496224949 | -73.99075269699097   | 16.8               |
| 5                | 20                  | Pershing Square North          | 40.751873         | -73.977706           | 15.6               |
| 5                | 20                  | LaGuardia Pl & W 3 St          | 40.72917025       | -73.99810231         | 15.5               |
| 5                | 20                  | Carmine St & 6 Ave             | 40.73038599       | -74.00214988         | 15.0               |
| 5                | 20                  | 1 Ave & E 16 St                | 40.73221853       | -73.98165557         | 14.6               |
| 5                | 20                  | Greenwich St & W Houston St    | 40.728846         | -74.008591           | 14.4               |
+------------------+---------------------+-------------------------------+-------------------+----------------------+--------------------+--+
10 rows selected (62.191 seconds)
0: jdbc:hive2://babar.es.its.nyu.edu:10000/>
```

3. Track the Sunday afternoon: get insights of where do New Yorkers spend the weekend afternoon.

Sample: Sunday 2PM-3PM

```sql
select * from average1 where (weekname = "7" and timeperiod = 14) sort by
avgvolume desc limit 10;
```

```
+------------------+---------------------+------------------------------------+--------------------+----------------------+--------------------+--+
| average1.weekname | average1.timeperiod |          average1.address          |    average1.lon    |     average1.lat     | average1.avgvolume |
+------------------+---------------------+------------------------------------+--------------------+----------------------+--------------------+--+
| 7                | 14                  | Broadway & E 14 St                 | 40.73454567        | -73.99074142         | 28.5               |
| 7                | 14                  | E 17 St & Broadway                 | 40.73704984        | -73.99009296         | 28.5               |
| 7                | 14                  | Columbus Ave & W 72 St             | 40.7770575         | -73.97898475         | 24.5               |
| 7                | 14                  | Broadway & W 58 St                 | 40.76695317        | -73.98169333         | 21.75              |
| 7                | 14                  | W 21 St & 6 Ave                    | 40.74173969        | -73.99415556         | 20.75              |
| 7                | 14                  | Cleveland Pl & Spring St           | 40.722103786686034 | -73.99724900722504  | 20.0               |
| 7                | 14                  | University Pl & E 14 St            | 40.73492695        | -73.99200509         | 19.5               |
| 7                | 14                  | Grand Army Plaza & Central Park S  | 40.7643971         | -73.97371465         | 18.75              |
| 7                | 14                  | W 22 St & 10 Ave                   | 40.74691959        | -74.00451887         | 18.25              |
| 7                | 14                  | Central Park S & 6 Ave             | 40.76590936        | -73.97634151         | 17.666666666666668 |
+------------------+---------------------+------------------------------------+--------------------+----------------------+--------------------+--+
10 rows selected (57.247 seconds)
0: jdbc:hive2://babar.es.its.nyu.edu:10000/>
```

4. Track the volume changing trends on a stable location on Monday (choose the place we find has most records on commuting time in step 1)

```sql
select * from average1 where (weekname = "1" and address = "Pershing Square
North") order by cast(timeperiod as int) asc;
```

```
INFO  : Ended Job = job_1569350662793_69716
INFO  : MapReduce Jobs Launched:
INFO  : Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.9 sec   HDFS Read: 5492935 HDFS Write: 1213 SUCCESS
INFO  : Total MapReduce CPU Time Spent: 7 seconds 900 msec
INFO  : Completed executing command(queryId=hive_20191130111717_14e33fbf-efa5-4cbe-9f28-2accc1be4b58); Time taken: 32.448 seconds
INFO  : OK
+--------------------+--------------------+------------------------+----------------+----------------+----------------------+--+
| average1.weekname  | average1.timeperiod |    average1.address    | average1.lon   | average1.lat   | average1.avgvolume   |
+--------------------+--------------------+------------------------+----------------+----------------+----------------------+--+
| 1                  | 0                  | Pershing Square North  | 40.751873      | -73.977706     | 1.25                 |
| 1                  | 1                  | Pershing Square North  | 40.751873      | -73.977706     | 2.3333333333333335   |
| 1                  | 3                  | Pershing Square North  | 40.751873      | -73.977706     | 1.0                  |
| 1                  | 4                  | Pershing Square North  | 40.751873      | -73.977706     | 1.0                  |
| 1                  | 5                  | Pershing Square North  | 40.751873      | -73.977706     | 3.0                  |
| 1                  | 6                  | Pershing Square North  | 40.751873      | -73.977706     | 3.5                  |
| 1                  | 7                  | Pershing Square North  | 40.751873      | -73.977706     | 2.0                  |
| 1                  | 8                  | Pershing Square North  | 40.751873      | -73.977706     | 9.25                 |
| 1                  | 9                  | Pershing Square North  | 40.751873      | -73.977706     | 9.25                 |
| 1                  | 10                 | Pershing Square North  | 40.751873      | -73.977706     | 13.25                |
| 1                  | 11                 | Pershing Square North  | 40.751873      | -73.977706     | 14.5                 |
| 1                  | 12                 | Pershing Square North  | 40.751873      | -73.977706     | 9.5                  |
| 1                  | 13                 | Pershing Square North  | 40.751873      | -73.977706     | 9.5                  |
| 1                  | 14                 | Pershing Square North  | 40.751873      | -73.977706     | 8.5                  |
| 1                  | 15                 | Pershing Square North  | 40.751873      | -73.977706     | 12.5                 |
| 1                  | 16                 | Pershing Square North  | 40.751873      | -73.977706     | 9.5                  |
| 1                  | 17                 | Pershing Square North  | 40.751873      | -73.977706     | 9.75                 |
| 1                  | 18                 | Pershing Square North  | 40.751873      | -73.977706     | 7.75                 |
| 1                  | 19                 | Pershing Square North  | 40.751873      | -73.977706     | 5.5                  |
| 1                  | 20                 | Pershing Square North  | 40.751873      | -73.977706     | 2.75                 |
| 1                  | 21                 | Pershing Square North  | 40.751873      | -73.977706     | 3.5                  |
| 1                  | 22                 | Pershing Square North  | 40.751873      | -73.977706     | 2.25                 |
| 1                  | 23                 | Pershing Square North  | 40.751873      | -73.977706     | 3.0                  |
+--------------------+--------------------+------------------------+----------------+----------------+----------------------+--+
23 rows selected (32.929 seconds)
0: jdbc:hive2://babar.es.its.nyu.edu:10000/>
```

5. Track the volume changing trends on a stable location on the mean of all weekdays

```
FROM average1 SELECT address,lon,lat,timeperiod,(sum(avgvolume)/5.0) AS meanAvg
WHERE ((weekname != "6" and weekname != "7") and address = "Pershing Square
North") GROUP BY address,lon,lat,timeperiod ORDER by cast(timeperiod as int)
asc
```

```
INFO  : 2019-11-30 11:39:31,519 Stage-2 map = 0%,   reduce = 0%
INFO  : 2019-11-30 11:39:37,671 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.56 sec
INFO  : 2019-11-30 11:39:43,830 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 4.68 sec
INFO  : MapReduce Total cumulative CPU time: 4 seconds 680 msec
INFO  : Ended Job = job_1569350662793_69747
INFO  : MapReduce Jobs Launched:
INFO  : Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 7.39 sec   HDFS Read: 5494399 HDFS Write: 1632 SUCCESS
INFO  : Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 4.68 sec   HDFS Read: 7332 HDFS Write: 1359 SUCCESS
INFO  : Total MapReduce CPU Time Spent: 12 seconds 70 msec
INFO  : Completed executing command(queryId=hive_20191130113838_de0403a0-bf63-497b-9739-7f7f71c64344); Time taken: 62.512 seconds
INFO  : OK
+------------------------+------------+-------------+------------+----------------------+--+
|        address         |    lon     |     lat     | timeperiod |       meanavg        |
+------------------------+------------+-------------+------------+----------------------+--+
| Pershing Square North  | 40.751873  | -73.977706  | 0          | 2.3833333333333333   |
| Pershing Square North  | 40.751873  | -73.977706  | 1          | 1.8833333333333335   |
| Pershing Square North  | 40.751873  | -73.977706  | 2          | 0.6                  |
| Pershing Square North  | 40.751873  | -73.977706  | 3          | 0.2                  |
| Pershing Square North  | 40.751873  | -73.977706  | 4          | 1.1733333333333333   |
| Pershing Square North  | 40.751873  | -73.977706  | 5          | 4.77                 |
| Pershing Square North  | 40.751873  | -73.977706  | 6          | 15.63                |
| Pershing Square North  | 40.751873  | -73.977706  | 7          | 41.71                |
| Pershing Square North  | 40.751873  | -73.977706  | 8          | 79.62                |
| Pershing Square North  | 40.751873  | -73.977706  | 9          | 74.37                |
| Pershing Square North  | 40.751873  | -73.977706  | 10         | 24.23                |
| Pershing Square North  | 40.751873  | -73.977706  | 11         | 15.37                |
| Pershing Square North  | 40.751873  | -73.977706  | 12         | 15.09                |
| Pershing Square North  | 40.751873  | -73.977706  | 13         | 17.310000000000002   |
| Pershing Square North  | 40.751873  | -73.977706  | 14         | 16.169999999999998   |
| Pershing Square North  | 40.751873  | -73.977706  | 15         | 22.169999999999998   |
| Pershing Square North  | 40.751873  | -73.977706  | 16         | 30.669999999999998   |
| Pershing Square North  | 40.751873  | -73.977706  | 17         | 64.79                |
| Pershing Square North  | 40.751873  | -73.977706  | 18         | 60.510000000000005   |
| Pershing Square North  | 40.751873  | -73.977706  | 19         | 29.419999999999998   |
| Pershing Square North  | 40.751873  | -73.977706  | 20         | 15.45                |
| Pershing Square North  | 40.751873  | -73.977706  | 21         | 8.370000000000001    |
| Pershing Square North  | 40.751873  | -73.977706  | 22         | 6.1                  |
| Pershing Square North  | 40.751873  | -73.977706  | 23         | 4.17                 |
+------------------------+------------+-------------+------------+----------------------+--+
24 rows selected (63.507 seconds)
0: jdbc:hive2://babar.es.its.nyu.edu:10000/>
```