

Assignment 11.4

Creating a table and loading dataset into the table

```
hive> create table estate(street string,city string,zip int,state string,beds int,baths int,sq_feet int,type string,
>
> sale_date string,price int,latitude string,longitude string)
>
> row format delimited
>
> fields terminated by ',';
OK
Time taken: 4.424 seconds
..
```

```
hive> load data local inpath '/home/acadgild/realestate.csv' into table estate;
Loading data to table default.estate
Table default.estate stats: [numFiles=1, totalSize=114088]
OK
Time taken: 6.931 seconds
..
```

1. Write a hive query to find the city wise list all the Condos which is not less than ten thousand.

```
hive> select * from estate where ((price>10000) and (type == 'Condo')) group by city;
Query ID = acadgild_20170528225050_883b9b60-ac35-44c3-a0b5-bdb3711fe183
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
```

Output

```
ANTELOPE
AUBURN
CAMERON PARK
CARMICHAEL
CITRUS HEIGHTS
ELK GROVE
FAIR OAKS
FOLSOM
GOLD RIVER
LINCOLN
RANCHO CORDOVA
ROSEVILLE
SACRAMENTO
Time taken: 171.058 seconds, Fetched: 13 row(s)
```

2. Write a hive query to find which is the cheapest Condo in CA. name the city,street and price for the Condo.

```
hive> select street,city,price from estate where type == 'Condo' order by price limit 1;
Query ID = acadgild_20170528230101_6eb2448e-7181-4454-a233-16e70fd46034
Total jobs = 1
Launching Job 1 out of 1
```

output

```
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 16.21 sec HDFS Read: 113399 HDFS Write: 33 SUCCESS
Total MapReduce CPU Time Spent: 16 seconds 210 msec
OK
5840 WALERGA RD SACRAMENTO 40000
```

3. Write a hive query to list top 5 residency details which lie in the budget of 60000-120000, an area more than 1450, sold after 17th may, min bedroom 3 and, min bathroom 2.

```
hive> select street,city,zip,state,beds,baths,sq feet,type,sale_date,price from estate where ((price>=60000 and price<=120000)
and (sq feet>1450) and (beds>2) and (baths>1)) order by price limit 5;
Query ID = acadgild_20170529195353_c90e7159-3947-40db-ae5a-cfcea63cc793
Total jobs = 1
Launching Job 1 out of 1
```

output

```
7342 GIGI PL SACRAMENTO 95828 CA 4 4 1995 Multi-Family Mon May 19 00:00:00 EDT 2008 120000
0
Time taken: 92.065 seconds, Fetched: 1 row(s)
hive>
```

4. Write a hive query to separate list of residential apartments with more than 2 beds. Also include columns in following order City,Baths,Sq_feet,Price,flat_type,Beds respectively

creating a separate table resident

```
hive>
> create table resident(City string,Baths int,sq_feet int,price int) partitioned by (flat_type string,Beds int);
OK
Time taken: 1.018 seconds
```

```
hive> set hive.exec.dynamic.partition.mode=nonstrict;
hive> insert into table resident partition(flat_type,Beds) select city,baths,sq_feet,price,type,beds from estate where (beds>2 and type='Residential');
Query ID = acadgild_20170529203535_e21315e8-c931-4edc-8548-4a8f7fbb59b4
```

output

```
acadgild@localhost:~
FAIR OAKS      5      2846      680000      Residential      5
SACRAMENTO    2      2484      699000      Residential      5
SACRAMENTO    2      1516      150000      Residential      5
SACRAMENTO    3      3076      240000      Residential      5
SACRAMENTO    3      2992      295000      Residential      5
LINCOLN 3      0      304000      Residential      5
ELK GROVE     3      3468      320000      Residential      5
ANTELOPE      3      2346      320000      Residential      5
ANTELOPE      3      2347      325000      Residential      5
RANCHO CORDOVA 3      3281      395100      Residential      5
SACRAMENTO    3      4246      452000      Residential      5
EL DORADO HILLS 3      0      487500      Residential      5
ROSEVILLE    3      0      520000      Residential      5
WILTON 2      3741      579093      Residential      5
GRANITE BAY   3      2896      676200      Residential      5
WILTON 3      3788      691659      Residential      5
GRANITE BAY   3      3670      760000      Residential      5
SACRAMENTO    2      1712      168000      Residential      5
ELK GROVE     2      2254      222750      Residential      5
SACRAMENTO    3      2258      232500      Residential      5
SACRAMENTO    3      2494      276000      Residential      5
SACRAMENTO    3      2280      279000      Residential      5
ELK GROVE     3      2592      303000      Residential      5
LINCOLN 3      2457      312000      Residential      5
ANTELOPE      3      3134      315000      Residential      5
ROCKLIN 3      2606      346375      Residential      5
ROSEVILLE    3      0      347225      Residential      5
ELK GROVE     3      2877      349000      Residential      5
ELK GROVE     3      2875      397000      Residential      5
ROSEVILLE    3      0      438000      Residential      5
FOLSOM 3      2787      441000      Residential      5
SACRAMENTO    3      2824      445000      Residential      5
EL DORADO HILLS 4      0      508000      Residential      5
SACRAMENTO    4      3863      598695      Residential      5
ROSEVILLE    3      0      600000      Residential      5
ROSEVILLE    3      3579      610000      Residential      5
EL DORADO HILLS 5      0      830000      Residential      6
ELK GROVE     3      2555      300000      Residential      6
Time taken: 1.004 seconds, Fetched: 715 row(s)
hive>
```