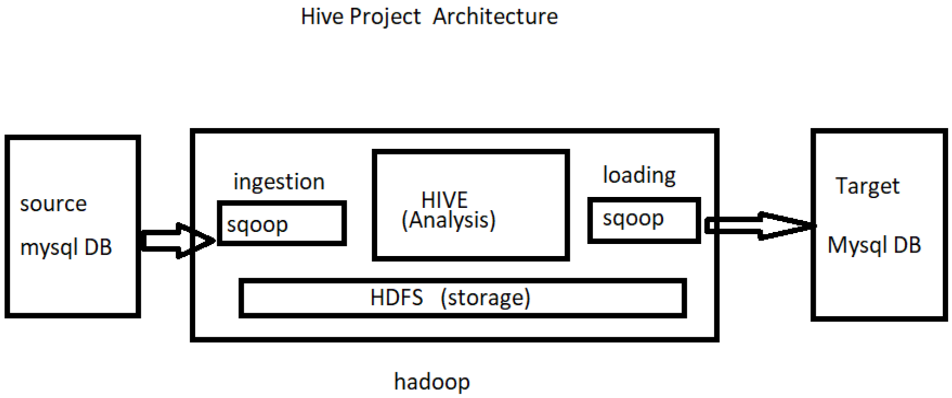
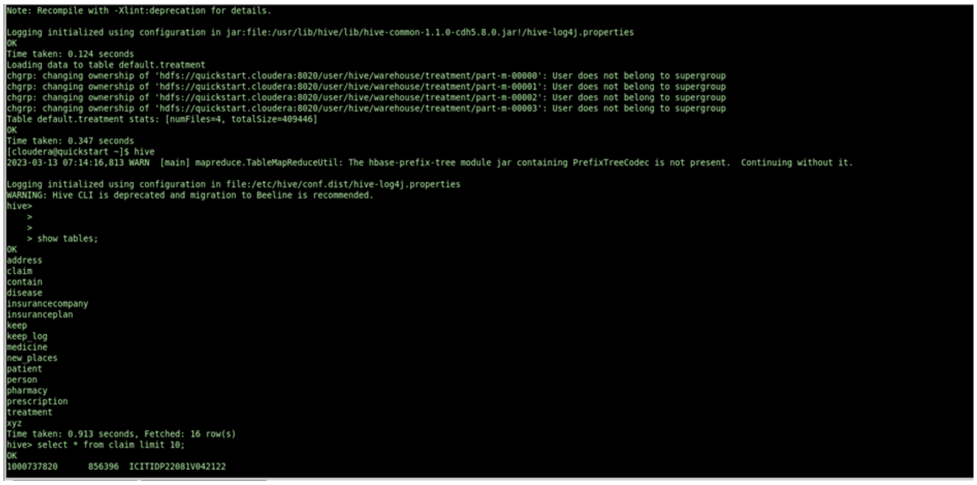
Hive HealthCare Analytics

Project Architecture:



Importing 13 tables from source database

sqoop import-all-tables --connect jdbc:mysql://localhost:3306/healthcare --username root --password cloudera --hive-import --m 1



```
Note: Recompile with -Xlint:deprecation for details.

Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-common-1.1.0-cdh5.8.0.jar!/hive-log4j.properties
OK
Time taken: 0.124 seconds
Loading data to table default.treatment
chgrp: changing ownership of 'hdfs://quickstart.cloudera:8020/user/hive/warehouse/treatment/part-m-00000': User does not belong to supergroup
chgrp: changing ownership of 'hdfs://quickstart.cloudera:8020/user/hive/warehouse/treatment/part-m-00001': User does not belong to supergroup
chgrp: changing ownership of 'hdfs://quickstart.cloudera:8020/user/hive/warehouse/treatment/part-m-00002': User does not belong to supergroup
chgrp: changing ownership of 'hdfs://quickstart.cloudera:8020/user/hive/warehouse/treatment/part-m-00003': User does not belong to supergroup
Table default.treatment stats: [numFiles=4, totalSize=409446]
OK
Time taken: 0.347 seconds
[cloudera@quickstart ~]$ hive
2023-03-13 07:14:16,813 WARN  [main] mapreduce.TableMapReduceUtil: The hbase-prefix-tree module jar containing PrefixTreeCodec is not present.  Continuing without it.

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive>
    >
    >
    > show tables;
OK
address
claim
contain
disease
insurancecompany
insuranceplan
keep
keep_log
medicine
new_places
patient
person
pharmacy
prescription
treatment
xyz
Time taken: 0.913 seconds, Fetched: 16 row(s)
hive> select * from claim limit 10;
OK
1000737820      856396  ICITIDP22081V042122
```

Analysing tables using Hive:

Step1: Creating hive external table to store analysis output
Step2: inserting Analysis result into hive external tables
Step3: creating table in source database

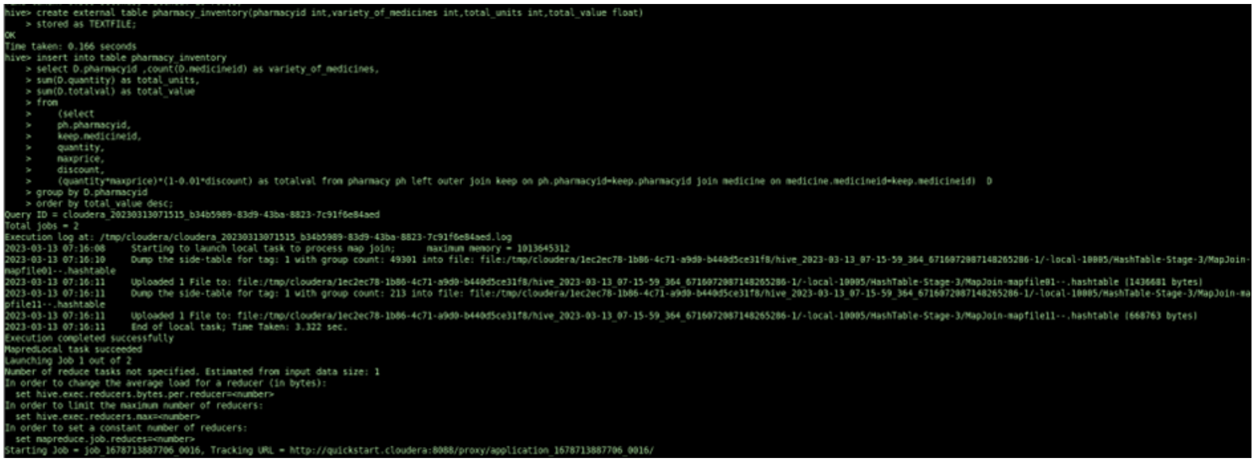
Step5: exporting the table data from hive external table to table in target Database

Step6: Checking the exported table in target database

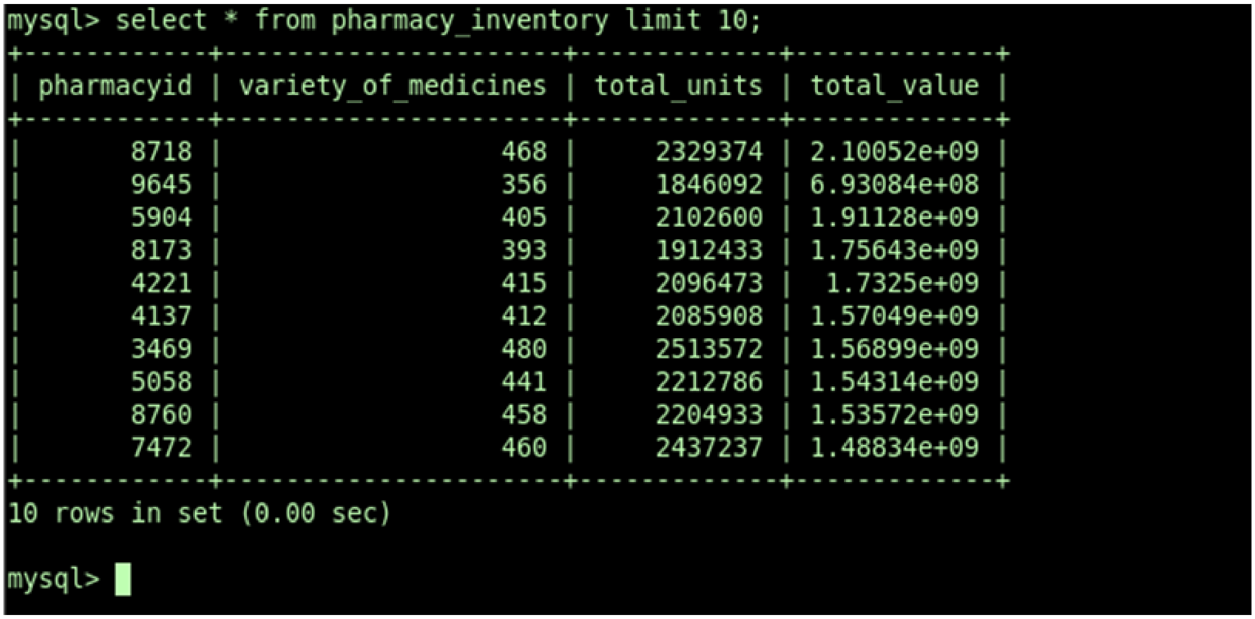
Problem statement1:

The Healthcare department wants a report about the inventory of pharmacies.
Generate a report on their behalf that shows how many units of medicine each pharmacy has in their inventory, the total maximum retail price of those medicines, and the total price of all the medicines after discount.
Note: discount field in keep signifies the percentage of discount on the maximum price.

```
   create external table ph_inv(pharmacyid int,variety_of_medicines int,total_units int,total_value float)
   ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
   LINES TERMINATED BY '\n';



   insert into table ph_inv
   select D.pharmacyid ,count(D.medicineid) as variety_of_medicines,
   sum(D.quantity) as total_units,
   sum(D.totalval) as total_value
   from
      (select
      ph.pharmacyid,
      keep.medicineid,
      quantity,
      maxprice,
      discount,
      (quantity*maxprice)*(1-0.01*discount) as totalval from pharmacy ph left outer join keep on
ph.pharmacyid=keep.pharmacyid join medicine on medicine.medicineid=keep.medicineid)  D
   group by D.pharmacyid
   order by total_value desc;


---in mysqldb
   create table pharmacy_inventory(pharmacyid int,variety_of_medicines int,total_units int,total_value float);

sqoop export --connect jdbc:mysql://localhost:3306/results --username root --password cloudera --table
pharmacy_inventory --export-dir /user/hive/warehouse/ph_inv --input-fields-terminated-by ','
```



```
hive> create external table pharmacy_inventory(pharmacyid int,variety_of_medicines int,total_units int,total_value float)
    > stored as TEXTFILE;
OK
Time taken: 0.166 seconds
hive> insert into table pharmacy_inventory
    > select D.pharmacyid ,count(D.medicineid) as variety_of_medicines,
    > sum(D.quantity) as total_units,
    > sum(D.totalval) as total_value
    > from
    >     (select
    >     ph.pharmacyid,
    >     keep.medicineid,
    >     quantity,
    >     maxprice,
    >     discount,
    >     (quantity*maxprice)*(1-0.01*discount) as totalval from pharmacy ph left outer join keep on ph.pharmacyid=keep.pharmacyid join medicine on medicine.medicineid=keep.medicineid)  D
    > group by D.pharmacyid
    > order by total_value desc;
Query ID = cloudera_20230313071515_b34b5989-83d9-43ba-8823-7c91f6e84aed
Total jobs = 2
Execution log at: /tmp/cloudera/cloudera_20230313071515_b34b5989-83d9-43ba-8823-7c91f6e84aed.log
2023-03-13 07:16:08     Starting to launch local task to process map join;      maximum memory = 1013645312
2023-03-13 07:16:10     Dump the side-table for tag: 1 with group count: 49301 into file: file:/tmp/cloudera/1ec2ec78-1b86-4c71-a9d0-b440d5ce31f8/hive_2023-03-13_07-15-59_364_6716072087148265286-1/-local-10005/HashTable-Stage-3/MapJoin-mapfile01--.hashtable
2023-03-13 07:16:11     Uploaded 1 File to: file:/tmp/cloudera/1ec2ec78-1b86-4c71-a9d0-b440d5ce31f8/hive_2023-03-13_07-15-59_364_6716072087148265286-1/-local-10005/HashTable-Stage-3/MapJoin-mapfile01--.hashtable (1436681 bytes)
2023-03-13 07:16:11     Dump the side-table for tag: 1 with group count: 213 into file: file:/tmp/cloudera/1ec2ec78-1b86-4c71-a9d0-b440d5ce31f8/hive_2023-03-13_07-15-59_364_6716072087148265286-1/-local-10005/HashTable-Stage-3/MapJoin-mapfile11--.hashtable
2023-03-13 07:16:11     Uploaded 1 File to: file:/tmp/cloudera/1ec2ec78-1b86-4c71-a9d0-b440d5ce31f8/hive_2023-03-13_07-15-59_364_6716072087148265286-1/-local-10005/HashTable-Stage-3/MapJoin-mapfile11--.hashtable (668763 bytes)
2023-03-13 07:16:11     End of local task; Time Taken: 3.322 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1678713887706_0016, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678713887706_0016/
```

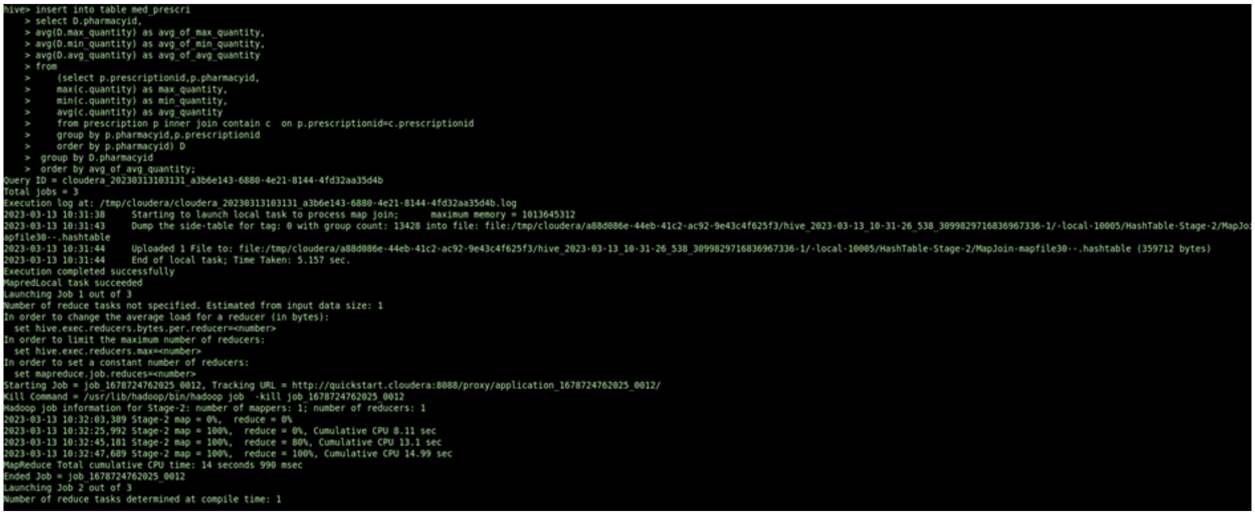


```
mysql> select * from pharmacy_inventory limit 10;
+------------+----------------------+-------------+-------------+
| pharmacyid | variety_of_medicines | total_units | total_value |
+------------+----------------------+-------------+-------------+
|       8718 |                  468 |     2329374 | 2.10052e+09 |
|       9645 |                  356 |     1846092 | 6.93084e+08 |
|       5904 |                  405 |     2102600 | 1.91128e+09 |
|       8173 |                  393 |     1912433 | 1.75643e+09 |
|       4221 |                  415 |     2096473 |  1.7325e+09 |
|       4137 |                  412 |     2085908 | 1.57049e+09 |
|       3469 |                  480 |     2513572 | 1.56899e+09 |
|       5058 |                  441 |     2212786 | 1.54314e+09 |
|       8760 |                  458 |     2204933 | 1.53572e+09 |
|       7472 |                  460 |     2437237 | 1.48834e+09 |
+------------+----------------------+-------------+-------------+
10 rows in set (0.00 sec)

mysql> 
```

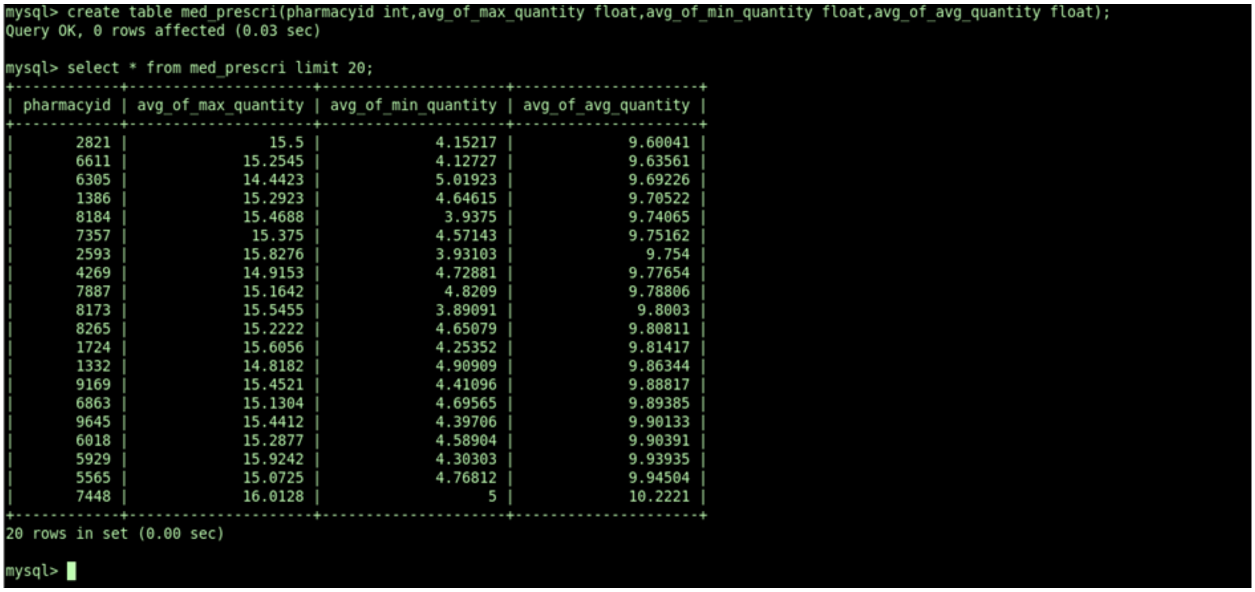
Problem statement 2:

**The healthcare department suspects that some pharmacies prescribe more medicines than others
in a single prescription, for them, generate a report that finds for each pharmacy the maximum,
minimum and average number of medicines prescribed in their prescriptions.**

```
create external table med_prescri(pharmacyid int,avg_of_max_quantity
float,avg_of_min_quantity float,avg_of_avg_quantity float)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
 LINES TERMINATED BY '\n';

    insert into table med_prescri
    select D.pharmacyid,
    avg(D.max_quantity) as avg_of_max_quantity,
    avg(D.min_quantity) as avg_of_min_quantity,
    avg(D.avg_quantity) as avg_of_avg_quantity
    from
        (select p.prescriptionid,p.pharmacyid,
        max(c.quantity) as max_quantity,
        min(c.quantity) as min_quantity,
        avg(c.quantity) as avg_quantity
        from prescription p inner join contain c  on p.prescriptionid=c.prescriptionid
        group by p.pharmacyid,p.prescriptionid
        order by p.pharmacyid) D
    group by D.pharmacyid
    order by avg_of_avg_quantity;

--in mysqldb
    create table med_prescri(pharmacyid int,avg_of_max_quantity
float,avg_of_min_quantity float,avg_of_avg_quantity float);

sqoop export --connect jdbc:mysql://localhost:3306/results --username root --password
cloudera --table med_prescri --export-dir /user/hive/warehouse/med_prescri
--input-fields-terminated-by ','
```



```
hive> insert into table med_prescri
    > select D.pharmacyid,
    > avg(D.max_quantity) as avg_of_max_quantity,
    > avg(D.min_quantity) as avg_of_min_quantity,
    > avg(D.avg_quantity) as avg_of_avg_quantity
    > from
    >     (select p.prescriptionid,p.pharmacyid,
    >     max(c.quantity) as max_quantity,
    >     min(c.quantity) as min_quantity,
    >     avg(c.quantity) as avg_quantity
    >     from prescription p inner join contain c  on p.prescriptionid=c.prescriptionid
    >     group by p.pharmacyid,p.prescriptionid
    >     order by p.pharmacyid) D
    >  group by D.pharmacyid
    >  order by avg_of_avg_quantity;
Query ID = cloudera_20230313103131_a3b6e143-6880-4e21-8144-4fd32aa35d4b
Total jobs = 3
Execution log at: /tmp/cloudera/cloudera_20230313103131_a3b6e143-6880-4e21-8144-4fd32aa35d4b.log
2023-03-13 10:31:38     Starting to launch local task to process map join;      maximum memory = 1013645312
2023-03-13 10:31:43     Dump the side-table for tag: 0 with group count: 13428 into file: file:/tmp/cloudera/a88d086e-44eb-41c2-ac92-9e43c4f625f3/hive_2023-03-13_10-31-26_538_3099829716836967336-1/-local-10005/HashTable-Stage-2/MapJo
apfile30--.hashtable
2023-03-13 10:31:44     Uploaded 1 File to: file:/tmp/cloudera/a88d086e-44eb-41c2-ac92-9e43c4f625f3/hive_2023-03-13_10-31-26_538_3099829716836967336-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile30--.hashtable (359712 bytes)
2023-03-13 10:31:44     End of local task; Time Taken: 5.157 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 3
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1678724762025_0012, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678724762025_0012/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1678724762025_0012
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-03-13 10:32:03,389 Stage-2 map = 0%,  reduce = 0%
2023-03-13 10:32:25,992 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 8.11 sec
2023-03-13 10:32:45,181 Stage-2 map = 100%,  reduce = 80%, Cumulative CPU 13.1 sec
2023-03-13 10:32:47,689 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 14.99 sec
MapReduce Total cumulative CPU time: 14 seconds 990 msec
Ended Job = job_1678724762025_0012
Launching Job 2 out of 3
Number of reduce tasks determined at compile time: 1
```



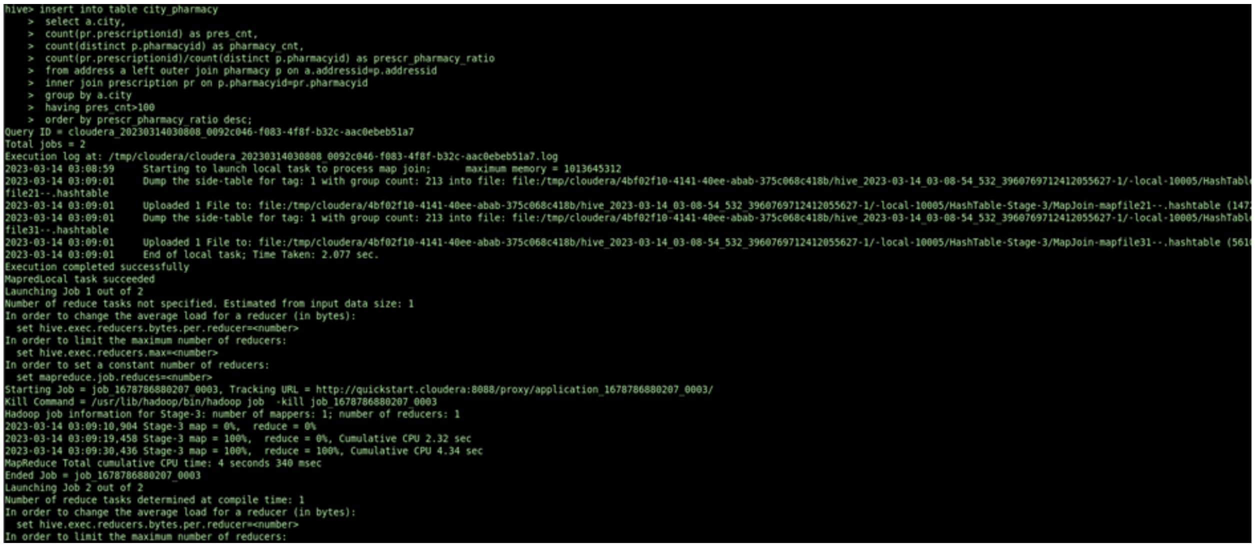
```
mysql> create table med_prescri(pharmacyid int,avg_of_max_quantity float,avg_of_min_quantity float,avg_of_avg_quantity float);
Query OK, 0 rows affected (0.03 sec)

mysql> select * from med_prescri limit 20;
+------------+---------------------+---------------------+---------------------+
| pharmacyid | avg_of_max_quantity | avg_of_min_quantity | avg_of_avg_quantity |
+------------+---------------------+---------------------+---------------------+
|       2821 |                15.5 |             4.15217 |             9.60041 |
|       6611 |             15.2545 |             4.12727 |             9.63561 |
|       6305 |             14.4423 |             5.01923 |             9.69226 |
|       1386 |             15.2923 |             4.64615 |             9.70522 |
|       8184 |             15.4688 |              3.9375 |             9.74065 |
|       7357 |              15.375 |             4.57143 |             9.75162 |
|       2593 |             15.8276 |             3.93103 |               9.754 |
|       4269 |             14.9153 |             4.72881 |             9.77654 |
|       7887 |             15.1642 |              4.8209 |             9.78806 |
|       8173 |             15.5455 |             3.89091 |              9.8003 |
|       8265 |             15.2222 |             4.65079 |             9.80811 |
|       1724 |             15.6056 |             4.25352 |             9.81417 |
|       1332 |             14.8182 |             4.90909 |             9.86344 |
|       9169 |             15.4521 |             4.41096 |             9.88817 |
|       6863 |             15.1304 |             4.69565 |             9.89385 |
|       9645 |             15.4412 |             4.39706 |             9.90133 |
|       6018 |             15.2877 |             4.58904 |             9.90391 |
|       5929 |             15.9242 |             4.30303 |             9.93935 |
|       5565 |             15.0725 |             4.76812 |             9.94504 |
|       7448 |             16.0128 |                   5 |             10.2221 |
+------------+---------------------+---------------------+---------------------+
20 rows in set (0.00 sec)

mysql> 
```

Problem Statement3:

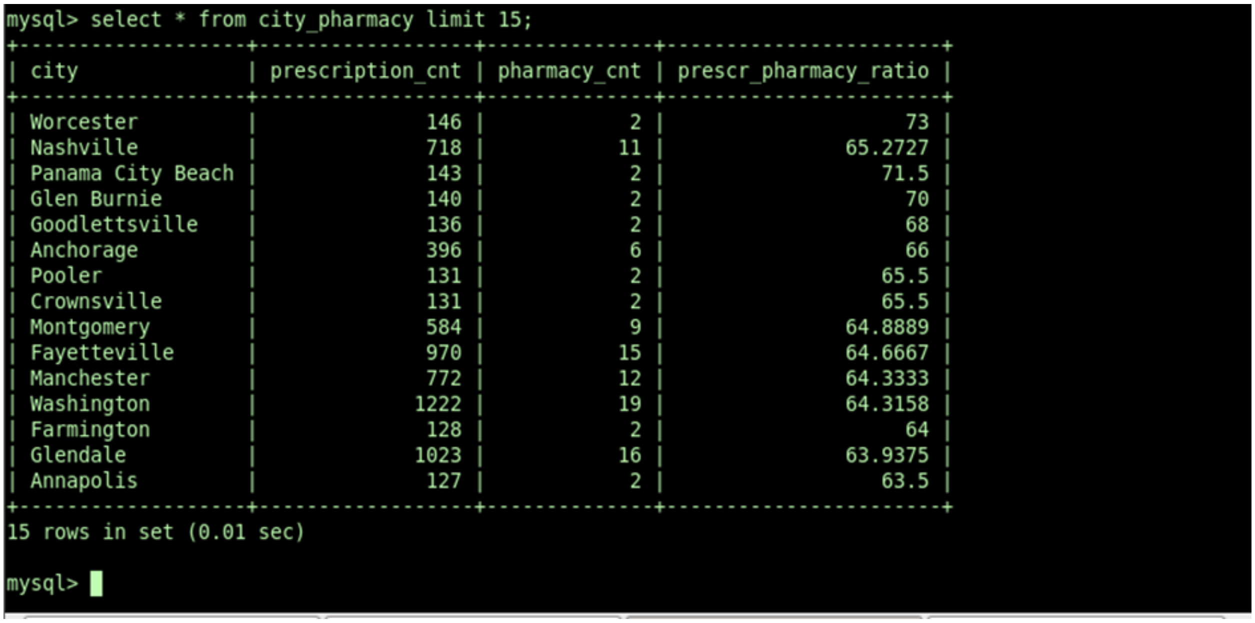
A company needs to set up 3 new pharmacies,
they have come up with an idea that the pharmacy can be set up in
cities where the pharmacy-to-prescription ratio is the lowest and the number of prescriptions
should exceed 100.
Assist the company to identify those cities where the pharmacy can be set up.

```
 create external table city_pharmacy(city string,prescription_cnt   int,pharmacy_cnt
int,prescr_pharmacy_ratio float)
    ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    LINES TERMINATED BY '\n';

    insert into table city_pharmacy
    select a.city,
    count(pr.prescriptionid) as pres_cnt,
    count(distinct p.pharmacyid) as pharmacy_cnt,
    count(pr.prescriptionid)/count(distinct p.pharmacyid) as prescr_pharmacy_ratio
    from address a left outer join pharmacy p on a.addressid=p.addressid
    inner join prescription pr on p.pharmacyid=pr.pharmacyid
    group by a.city
    having pres_cnt>100
    order by prescr_pharmacy_ratio desc;

--in mysqldb
    create table city_pharmacy(city varchar(20),prescription_cnt int,pharmacy_cnt
int,prescr_pharmacy_ratio float);
```

sqoop export --connect jdbc:mysql://localhost:3306/results --username root --password cloudera --table city_pharmacy --export-dir /user/hive/warehouse/city_pharmacy --input-fields-terminated-by ','



```
hive> insert into table city_pharmacy
    >  select a.city,
    >  count(pr.prescriptionid) as pres_cnt,
    >  count(distinct p.pharmacyid) as pharmacy_cnt,
    >  count(pr.prescriptionid)/count(distinct p.pharmacyid) as prescr_pharmacy_ratio
    >  from address a left outer join pharmacy p on a.addressid=p.addressid
    >  inner join prescription pr on p.pharmacyid=pr.pharmacyid
    >  group by a.city
    >  having pres_cnt>100
    >  order by prescr_pharmacy_ratio desc;
Query ID = cloudera_20230314030808_0092c046-f083-4f8f-b32c-aac0ebeb51a7
Total jobs = 2
Execution log at: /tmp/cloudera/cloudera_20230314030808_0092c046-f083-4f8f-b32c-aac0ebeb51a7.log
2023-03-14 03:08:59     Starting to launch local task to process map join;      maximum memory = 1013645312
2023-03-14 03:09:01     Dump the side-table for tag: 1 with group count: 213 into file: file:/tmp/cloudera/4bf02f10-4141-40ee-abab-375c068c418b/hive_2023-03-14_03-08-54_532_3960769712412055627-1/-local-10005/HashTable
file21--.hashtable
2023-03-14 03:09:01     Uploaded 1 File to: file:/tmp/cloudera/4bf02f10-4141-40ee-abab-375c068c418b/hive_2023-03-14_03-08-54_532_3960769712412055627-1/-local-10005/HashTable-Stage-3/MapJoin-mapfile21--.hashtable (147
2023-03-14 03:09:01     Dump the side-table for tag: 1 with group count: 213 into file: file:/tmp/cloudera/4bf02f10-4141-40ee-abab-375c068c418b/hive_2023-03-14_03-08-54_532_3960769712412055627-1/-local-10005/HashTable
file31--.hashtable
2023-03-14 03:09:01     Uploaded 1 File to: file:/tmp/cloudera/4bf02f10-4141-40ee-abab-375c068c418b/hive_2023-03-14_03-08-54_532_3960769712412055627-1/-local-10005/HashTable-Stage-3/MapJoin-mapfile31--.hashtable (561
2023-03-14 03:09:01     End of local task; Time Taken: 2.077 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1678786880207_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678786880207_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1678786880207_0003
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2023-03-14 03:09:10,904 Stage-3 map = 0%,  reduce = 0%
2023-03-14 03:09:19,458 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 2.32 sec
2023-03-14 03:09:30,436 Stage-3 map = 100%,  reduce = 100%, Cumulative CPU 4.34 sec
MapReduce Total cumulative CPU time: 4 seconds 340 msec
Ended Job = job_1678786880207_0003
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
```



```
mysql> select * from city_pharmacy limit 15;
+-------------------+------------------+--------------+-----------------------+
| city              | prescription_cnt | pharmacy_cnt | prescr_pharmacy_ratio |
+-------------------+------------------+--------------+-----------------------+
| Worcester         |              146 |            2 |                    73 |
| Nashville         |              718 |           11 |               65.2727 |
| Panama City Beach |              143 |            2 |                  71.5 |
| Glen Burnie       |              140 |            2 |                    70 |
| Goodlettsville    |              136 |            2 |                    68 |
| Anchorage         |              396 |            6 |                    66 |
| Pooler            |              131 |            2 |                  65.5 |
| Crownsville       |              131 |            2 |                  65.5 |
| Montgomery        |              584 |            9 |               64.8889 |
| Fayetteville      |              970 |           15 |               64.6667 |
| Manchester        |              772 |           12 |               64.3333 |
| Washington        |             1222 |           19 |               64.3158 |
| Farmington        |              128 |            2 |                    64 |
| Glendale          |             1023 |           16 |               63.9375 |
| Annapolis         |              127 |            2 |                  63.5 |
+-------------------+------------------+--------------+-----------------------+
15 rows in set (0.01 sec)

mysql> █
```

Problem Statement 4:

**"HealthDirect" pharmacy finds it difficult to deal with the product type of medicine being displayed in numerical form, they want the product type in words.
Also, they want to filter the medicines based on tax criteria.
Display only the medicines of product categories 1, 2, and 3 for medicines that come under tax category I and medicines of product categories 4, 5, and 6 for medicines that come under tax category II.**

Write a SQL query to solve this problem.
ProductType numerical form and ProductType in words are given by
1 - Generic,
2 - Patent,
3 - Reference,
4 - Similar,
5 - New,
6 - Specific,
7 - Biological,
8 – Dinamized

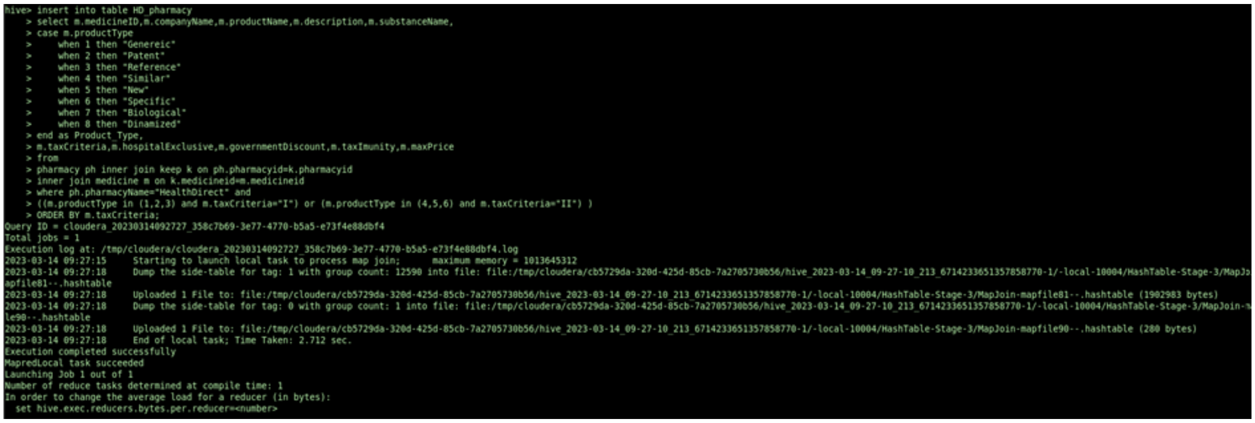
3 random rows and the column names of the Medicine table are given for reference.
Medicine (medicineID, companyName, productName, description, substanceName, productType, taxCriteria, hospitalExclusive, governmentDiscount, taxImunity, maxPrice)

```
   create external table HD_pharmacy(medicineID int,companyName string,productName
string,description string,substanceName string,Product_Type string,taxCriteria
string,hospitalExclusive string,governmentDiscount string,taxImunity string,maxPrice
float)
   ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
   LINES TERMINATED BY '\n';

   insert into table HD_pharmacy
   select
m.medicineID,m.companyName,m.productName,m.description,m.substanceName,
   case m.productType
      when 1 then "Genereic"
      when 2 then "Patent"
      when 3 then "Reference"
      when 4 then "Similar"
      when 5 then "New"
      when 6 then "Specific"
      when 7 then "Biological"
      when 8 then "Dinamized"
   end as Product_Type,
   m.taxCriteria,m.hospitalExclusive,m.governmentDiscount,m.taxImunity,m.maxPrice
   from
   pharmacy ph inner join keep k on ph.pharmacyid=k.pharmacyid
   inner join medicine m on k.medicineid=m.medicineid
   where ph.pharmacyName="HealthDirect" and
   ((m.productType in (1,2,3) and m.taxCriteria="I") or (m.productType in (4,5,6) and
m.taxCriteria="II") )
   ORDER BY m.taxCriteria;
```

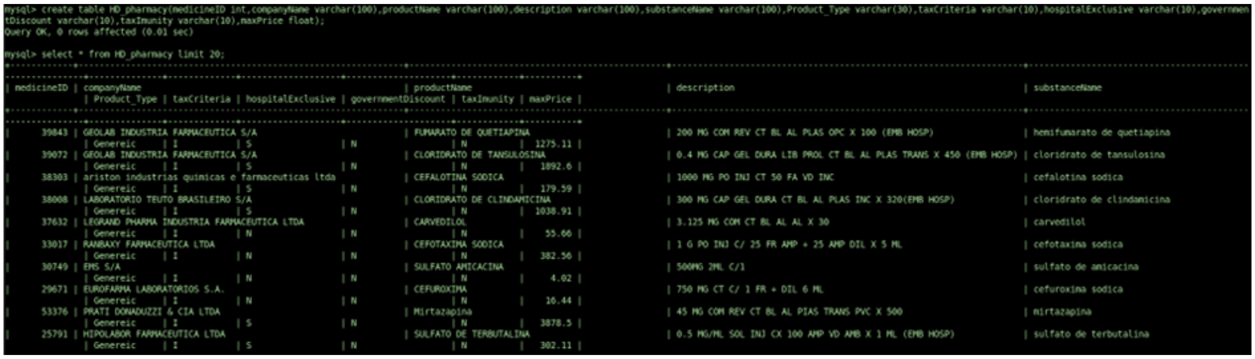
--in mysqldb

```
   create table HD_pharmacy(medicineID int,companyName varchar(100),productName varchar(100),description varchar(100),substanceName varchar(100),Product_Type varchar(30),taxCriteria varchar(10),hospitalExclusive varchar(10),governmentDiscount varchar(10),taxImunity varchar(10),maxPrice float);
```

```
sqoop export --connect jdbc:mysql://localhost:3306/results --username root --password cloudera --table HD_pharmacy --export-dir /user/hive/warehouse/hd_pharmacy
--input-fields-terminated-by ','
```



```
hive> insert into table HD_pharmacy
    > select m.medicineID,m.companyName,m.productName,m.description,m.substanceName,
    > case m.productType
    >     when 1 then "Genereic"
    >     when 2 then "Patent"
    >     when 3 then "Reference"
    >     when 4 then "Similar"
    >     when 5 then "New"
    >     when 6 then "Specific"
    >     when 7 then "Biological"
    >     when 8 then "Dinamized"
    > end as Product_Type,
    > m.taxCriteria,m.hospitalExclusive,m.governmentDiscount,m.taxImunity,m.maxPrice
    > from
    > pharmacy ph inner join keep k on ph.pharmacyid=k.pharmacyid
    > inner join medicine m on k.medicineid=m.medicineid
    > where ph.pharmacyName="HealthDirect" and
    > ((m.productType in (1,2,3) and m.taxCriteria="I") or (m.productType in (4,5,6) and m.taxCriteria="II") )
    > ORDER BY m.taxCriteria;
Query ID = cloudera_20230314092727_358c7b69-3e77-4770-b5a5-e73f4e88dbf4
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20230314092727_358c7b69-3e77-4770-b5a5-e73f4e88dbf4.log
2023-03-14 09:27:15     Starting to launch local task to process map join;      maximum memory = 1013645312
2023-03-14 09:27:18     Dump the side-table for tag: 1 with group count: 12590 into file: file:/tmp/cloudera/cb5729da-320d-425d-85cb-7a2705730b56/hive_2023-03-14_09-27-10_213_6714233651357858770-1/-local-10004/HashTable-Stage-3/MapJo
apfile81--.hashtable
2023-03-14 09:27:18     Uploaded 1 File to: file:/tmp/cloudera/cb5729da-320d-425d-85cb-7a2705730b56/hive_2023-03-14_09-27-10_213_6714233651357858770-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile81--.hashtable (1902983 bytes)
2023-03-14 09:27:18     Dump the side-table for tag: 0 with group count: 1 into file: file:/tmp/cloudera/cb5729da-320d-425d-85cb-7a2705730b56/hive_2023-03-14_09-27-10_213_6714233651357858770-1/-local-10004/HashTable-Stage-3/MapJoin-m
le90--.hashtable
2023-03-14 09:27:18     Uploaded 1 File to: file:/tmp/cloudera/cb5729da-320d-425d-85cb-7a2705730b56/hive_2023-03-14_09-27-10_213_6714233651357858770-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile90--.hashtable (280 bytes)
2023-03-14 09:27:18     End of local task; Time Taken: 2.712 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
```



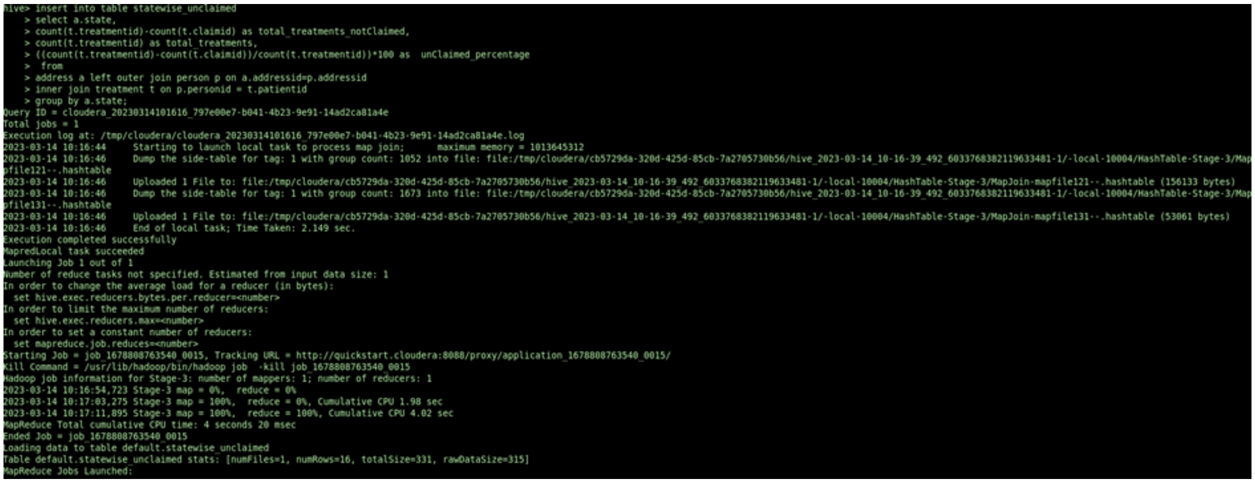
```
mysql> create table HD_pharmacy(medicineID int,companyName varchar(100),productName varchar(100),description varchar(100),substanceName varchar(100),Product_Type varchar(30),taxCriteria varchar(10),hospitalExclusive varchar(10),governmen
tDiscount varchar(10),taxImunity varchar(10),maxPrice float);
Query OK, 0 rows affected (0.01 sec)

mysql> select * from HD_pharmacy limit 20;
+------------+--------------------------------------------------------+------------------------------+--------------------------------------------------------------+-------------------------------+
-------------+-------------+-------------------+--------------------+------------+----------+
| medicineID | companyName                                            | productName                  | description                                                  | substanceName
             | Product_Type | taxCriteria | hospitalExclusive | governmentDiscount | taxImunity | maxPrice |
+------------+--------------------------------------------------------+------------------------------+--------------------------------------------------------------+-------------------------------+
-------------+-------------+-------------------+--------------------+------------+----------+
|      39843 | GEOLAB INDUSTRIA FARMACEUTICA S/A                      | FUMARATO DE QUETIAPINA       | 200 MG COM REV CT BL AL PLAS OPC X 100 (EMB HOSP)            | hemifumarato de quetiapina
             | Genereic     | I           | S                 | N                  | N          |  1275.11 |
|      39072 | GEOLAB INDUSTRIA FARMACEUTICA S/A                      | CLORIDRATO DE TANSULOSINA    | 0.4 MG CAP GEL DURA LIB PROL CT BL AL PLAS TRANS X 450 (EMB HOSP) | cloridrato de tansulosina
             | Genereic     | I           | S                 | N                  | N          |   1892.6 |
|      38303 | ariston industrias quimicas e farmaceuticas ltda       | CEFALOTINA SODICA            | 1000 MG PO INJ CT 50 FA VD INC                               | cefalotina sodica
             | Genereic     | I           | S                 | N                  | N          |   179.59 |
|      38008 | LABORATORIO TEUTO BRASILEIRO S/A                       | CLORIDRATO DE CLINDAMICINA   | 300 MG CAP GEL DURA CT BL AL PLAS INC X 320(EMB HOSP)        | cloridrato de clindamicina
             | Genereic     | I           | S                 | N                  | N          |  1038.91 |
|      37632 | LEGRAND PHARMA INDUSTRIA FARMACEUTICA LTDA             | CARVEDILOL                   | 3.125 MG COM CT BL AL AL X 30                                | carvedilol
             | Genereic     | I           | N                 | N                  | N          |    55.66 |
|      33017 | RANBAXY FARMACEUTICA LTDA                              | CEFOTAXIMA SODICA            | 1 G PO INJ C/ 25 FR AMP + 25 AMP DIL X 5 ML                  | cefotaxima sodica
             | Genereic     | I           | N                 | N                  | N          |   382.56 |
|      30749 | EMS S/A                                                | SULFATO AMICACINA            | 500MG 2ML C/1                                                | sulfato de amicacina
             | Genereic     | I           | N                 | N                  | N          |     4.02 |
|      29671 | EUROFARMA LABORATORIOS S.A.                            | CEFUROXIMA                   | 750 MG CT C/ 1 FR + DIL 6 ML                                 | cefuroxima sodica
             | Genereic     | I           | N                 | N                  | N          |    16.44 |
|      53376 | PRATI DONADUZZI & CIA LTDA                             | Mirtazapina                  | 45 MG COM REV CT BL AL PIAS TRANS PVC X 500                  | mirtazapina
             | Genereic     | I           | S                 | N                  | N          |   3878.5 |
|      25791 | HIPOLABOR FARMACEUTICA LTDA                            | SULFATO DE TERBUTALINA       | 0.5 MG/ML SOL INJ CX 100 AMP VD AMB X 1 ML (EMB HOSP)        | sulfato de terbutalina
             | Genereic     | I           | S                 | N                  | N          |   302.11 |
```

Problem Statement 5:

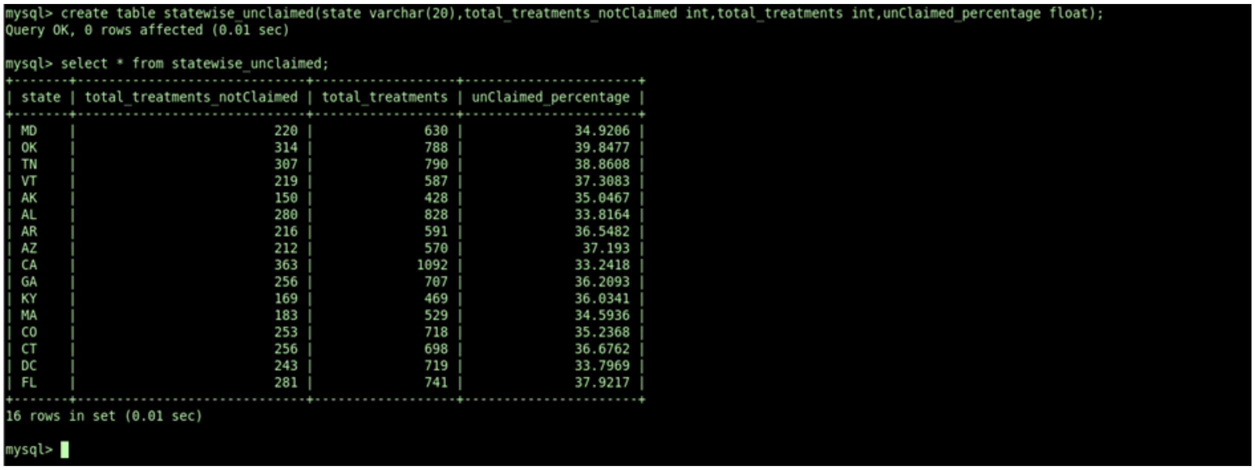
Sarah, from the healthcare department, has noticed many people do not claim insurance for
their treatment. She has requested a state-wise report of the percentage of treatments that
took place without claiming insurance. Assist Sarah by creating a report
as per her requirement.

```
   create external table statewise_unclaimed(state string,total_treatments_notClaimed int,total_treatments int,unClaimed_percentage float)
```

```
    ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    LINES TERMINATED BY '\n';

    insert into table statewise_unclaimed
    select a.state,
    count(t.treatmentid)-count(t.claimid) as total_treatments_notClaimed,
    count(t.treatmentid) as total_treatments,
    ((count(t.treatmentid)-count(t.claimid))/count(t.treatmentid))*100 as
unClaimed_percentage
    from
    address a left outer join person p on a.addressid=p.addressid
    inner join treatment t on p.personid = t.patientid
    group by a.state;

--in mysqldb
    create table statewise_unclaimed(state varchar(20),total_treatments_notClaimed
int,total_treatments int,unClaimed_percentage float);

sqoop export --connect jdbc:mysql://localhost:3306/results --username root --password
cloudera --table statewise_unclaimed --export-dir
/user/hive/warehouse/statewise_unclaimed --input-fields-terminated-by ','
```



```
hive> insert into table statewise_unclaimed
    > select a.state,
    > count(t.treatmentid)-count(t.claimid) as total_treatments_notClaimed,
    > count(t.treatmentid) as total_treatments,
    > ((count(t.treatmentid)-count(t.claimid))/count(t.treatmentid))*100 as  unClaimed_percentage
    >  from
    > address a left outer join person p on a.addressid=p.addressid
    > inner join treatment t on p.personid = t.patientid
    > group by a.state;
Query ID = cloudera_20230314101616_797e00e7-b041-4b23-9e91-14ad2ca81a4e
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20230314101616_797e00e7-b041-4b23-9e91-14ad2ca81a4e.log
2023-03-14 10:16:44     Starting to launch local task to process map join;      maximum memory = 1013645312
2023-03-14 10:16:46     Dump the side-table for tag: 1 with group count: 1052 into file: file:/tmp/cloudera/cb5729da-320d-425d-85cb-7a2705730b56/hive_2023-03-14_10-16-39_492_6033768382119633481-1/-local-10004/HashTable-Stage-3/Map
pfile121--.hashtable
2023-03-14 10:16:46     Uploaded 1 File to: file:/tmp/cloudera/cb5729da-320d-425d-85cb-7a2705730b56/hive_2023-03-14_10-16-39_492_6033768382119633481-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile121--.hashtable (156133 bytes)
2023-03-14 10:16:46     Dump the side-table for tag: 1 with group count: 1673 into file: file:/tmp/cloudera/cb5729da-320d-425d-85cb-7a2705730b56/hive_2023-03-14_10-16-39_492_6033768382119633481-1/-local-10004/HashTable-Stage-3/Map
pfile131--.hashtable
2023-03-14 10:16:46     Uploaded 1 File to: file:/tmp/cloudera/cb5729da-320d-425d-85cb-7a2705730b56/hive_2023-03-14_10-16-39_492_6033768382119633481-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile131--.hashtable (53061 bytes)
2023-03-14 10:16:46     End of local task; Time Taken: 2.149 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1678808763540_0015, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678808763540_0015/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1678808763540_0015
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2023-03-14 10:16:54,723 Stage-3 map = 0%,  reduce = 0%
2023-03-14 10:17:03,275 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.98 sec
2023-03-14 10:17:11,895 Stage-3 map = 100%,  reduce = 100%, Cumulative CPU 4.02 sec
MapReduce Total cumulative CPU time: 4 seconds 20 msec
Ended Job = job_1678808763540_0015
Loading data to table default.statewise_unclaimed
Table default.statewise_unclaimed stats: [numFiles=1, numRows=16, totalSize=331, rawDataSize=315]
MapReduce Jobs Launched:
```



```
mysql> create table statewise_unclaimed(state varchar(20),total_treatments_notClaimed int,total_treatments int,unClaimed_percentage float);
Query OK, 0 rows affected (0.01 sec)

mysql> select * from statewise_unclaimed;
+-------+-----------------------------+------------------+----------------------+
| state | total_treatments_notClaimed | total_treatments | unClaimed_percentage |
+-------+-----------------------------+------------------+----------------------+
| MD    |                         220 |              630 |              34.9206 |
| OK    |                         314 |              788 |              39.8477 |
| TN    |                         307 |              790 |              38.8608 |
| VT    |                         219 |              587 |              37.3083 |
| AK    |                         150 |              428 |              35.0467 |
| AL    |                         280 |              828 |              33.8164 |
| AR    |                         216 |              591 |              36.5482 |
| AZ    |                         212 |              570 |               37.193 |
| CA    |                         363 |             1092 |              33.2418 |
| GA    |                         256 |              707 |              36.2093 |
| KY    |                         169 |              469 |              36.0341 |
| MA    |                         183 |              529 |              34.5936 |
| CO    |                         253 |              718 |              35.2368 |
| CT    |                         256 |              698 |              36.6762 |
| DC    |                         243 |              719 |              33.7969 |
| FL    |                         281 |              741 |              37.9217 |
+-------+-----------------------------+------------------+----------------------+
16 rows in set (0.01 sec)

mysql> 
```

/*

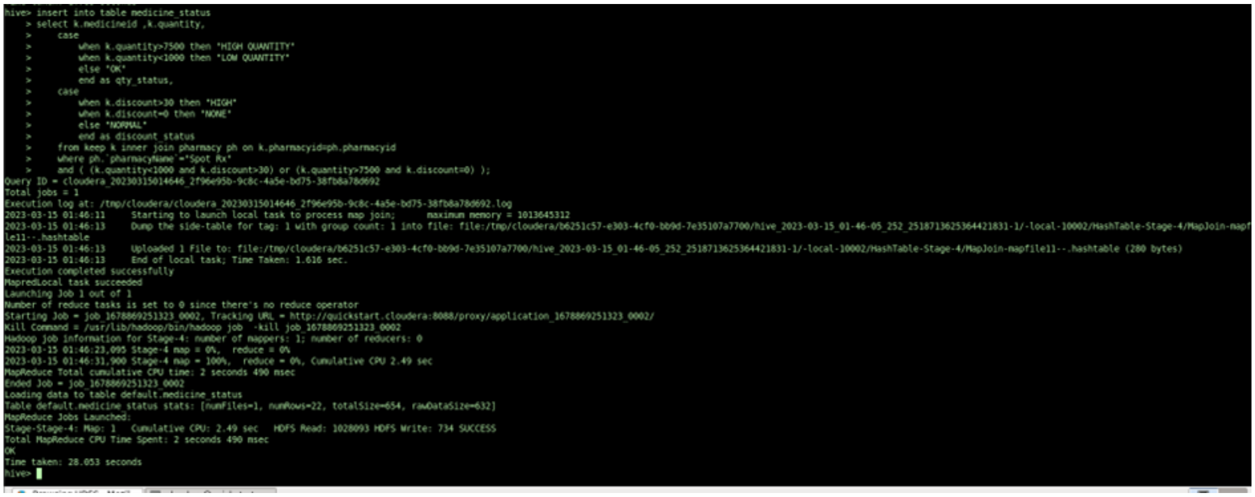
Problem Statement6:

In the Inventory of a pharmacy 'Spot Rx' the quantity of medicine is considered ‘HIGH QUANTITY’
when the quantity exceeds 7500
and ‘LOW QUANTITY’ when the quantity falls short of 1000. The discount is considered “HIGH”
if the discount rate on a product is 30% or higher, and the discount is considered “NONE”
when the discount rate on a product is 0%.
'Spot Rx' needs to find all the Low quantity products with high discounts and all the high-quantity
products with no discount so they can adjust the discount rate according to the demand.
Write a query for the pharmacy listing all the necessary details relevant to the given requirement.

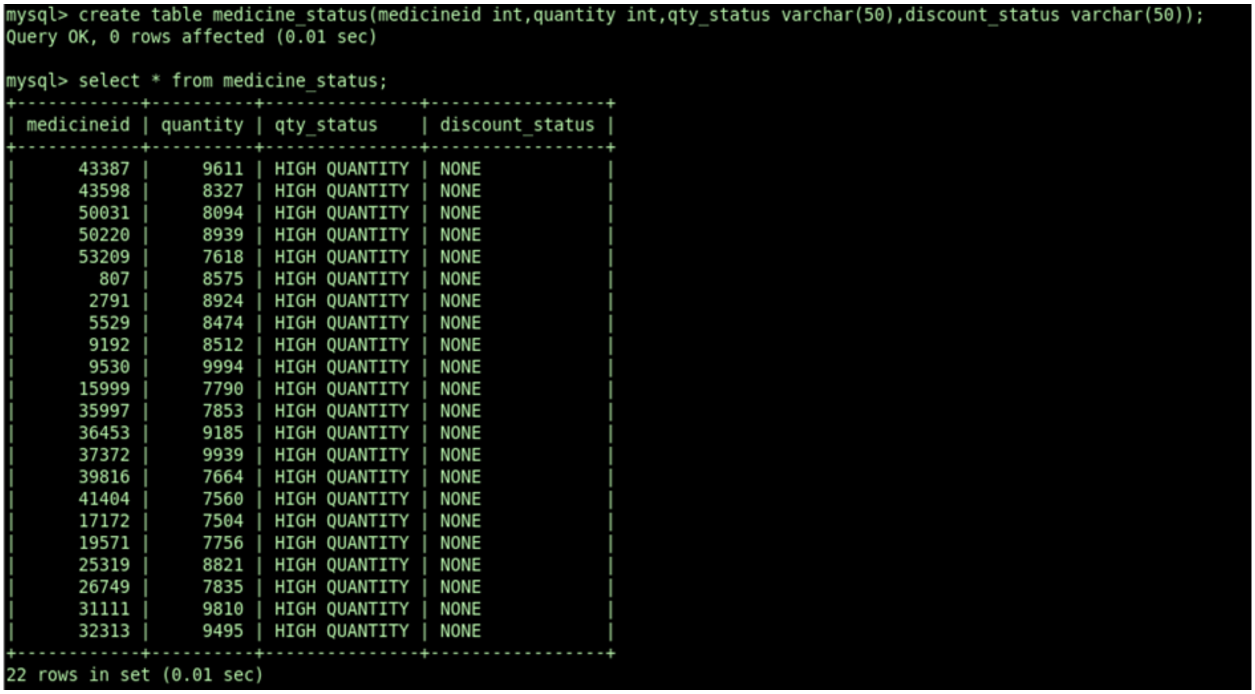
Hint: Inventory is reflected in the Keep table.

```
   create external table medicine_status(medicineid int,quantity int,qty_status string,discount_status string)
   ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
   LINES TERMINATED BY '\n';

   insert into table medicine_status
   select k.medicineid ,k.quantity,
      case
         when k.quantity>7500 then "HIGH QUANTITY"
         when k.quantity<1000 then "LOW QUANTITY"
         else "OK"
```

```
          end as qty_status,
        case
          when k.discount>30 then "HIGH"
          when k.discount=0 then "NONE"
          else "NORMAL"
          end as discount_status
        from keep k inner join pharmacy ph on k.pharmacyid=ph.pharmacyid
        where ph.`pharmacyName`="Spot Rx"
        and ( (k.quantity<1000 and k.discount>30) or (k.quantity>7500 and k.discount=0) );

--in mysqldb
    create table medicine_status(medicineid int,quantity int,qty_status
varchar(50),discount_status varchar(50));

sqoop export --connect jdbc:mysql://localhost:3306/results --username root --password
cloudera --table medicine_status --export-dir /user/hive/warehouse/medicine_status
--input-fields-terminated-by ','
```



```
hive> insert into table medicine_status
    > select k.medicineid ,k.quantity,
    >     case
    >         when k.quantity>7500 then "HIGH QUANTITY"
    >         when k.quantity<1000 then "LOW QUANTITY"
    >         else "OK"
    >         end as qty_status,
    >     case
    >         when k.discount>30 then "HIGH"
    >         when k.discount=0 then "NONE"
    >         else "NORMAL"
    >         end as discount_status
    >     from keep k inner join pharmacy ph on k.pharmacyid=ph.pharmacyid
    >     where ph.`pharmacyName`="Spot Rx"
    >     and ( (k.quantity<1000 and k.discount>30) or (k.quantity>7500 and k.discount=0) );
Query ID = cloudera_20230315014646_2f96e95b-9c8c-4a5e-bd75-38fb8a78d692
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20230315014646_2f96e95b-9c8c-4a5e-bd75-38fb8a78d692.log
2023-03-15 01:46:11     Starting to launch local task to process map join;      maximum memory = 1013645312
2023-03-15 01:46:13     Dump the side-table for tag: 1 with group count: 1 into file: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_01-46-05_252_2518713625364421831-1/-local-10002/HashTable-Stage-4/MapJoin-mapf
ile11--.hashtable
2023-03-15 01:46:13     Uploaded 1 File to: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_01-46-05_252_2518713625364421831-1/-local-10002/HashTable-Stage-4/MapJoin-mapfile11--.hashtable (280 bytes)
2023-03-15 01:46:13     End of local task; Time Taken: 1.616 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1678869251323_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678869251323_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1678869251323_0002
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 0
2023-03-15 01:46:23,095 Stage-4 map = 0%,  reduce = 0%
2023-03-15 01:46:31,900 Stage-4 map = 100%,  reduce = 0%, Cumulative CPU 2.49 sec
MapReduce Total cumulative CPU time: 2 seconds 490 msec
Ended Job = job_1678869251323_0002
Loading data to table default.medicine_status
Table default.medicine_status stats: [numFiles=1, numRows=22, totalSize=654, rawDataSize=632]
MapReduce Jobs Launched:
Stage-Stage-4: Map: 1   Cumulative CPU: 2.49 sec   HDFS Read: 1028093 HDFS Write: 734 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 490 msec
OK
Time taken: 28.053 seconds
hive>
```



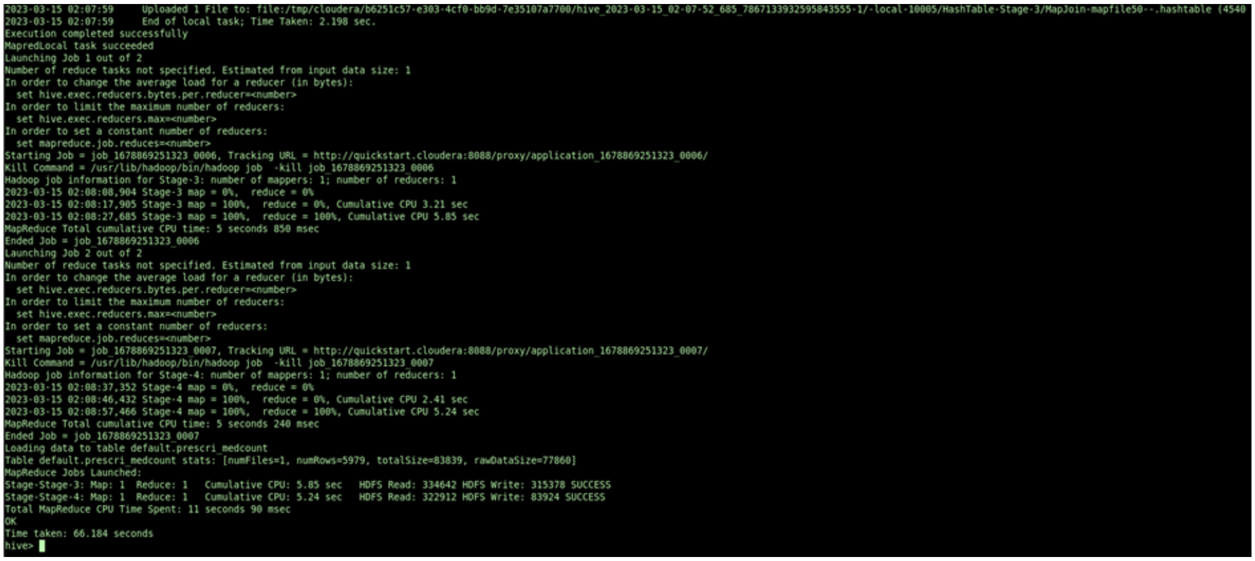
```
mysql> create table medicine_status(medicineid int,quantity int,qty_status varchar(50),discount_status varchar(50));
Query OK, 0 rows affected (0.01 sec)

mysql> select * from medicine_status;
+------------+----------+---------------+-----------------+
| medicineid | quantity | qty_status    | discount_status |
+------------+----------+---------------+-----------------+
|      43387 |     9611 | HIGH QUANTITY | NONE            |
|      43598 |     8327 | HIGH QUANTITY | NONE            |
|      50031 |     8094 | HIGH QUANTITY | NONE            |
|      50220 |     8939 | HIGH QUANTITY | NONE            |
|      53209 |     7618 | HIGH QUANTITY | NONE            |
|        807 |     8575 | HIGH QUANTITY | NONE            |
|       2791 |     8924 | HIGH QUANTITY | NONE            |
|       5529 |     8474 | HIGH QUANTITY | NONE            |
|       9192 |     8512 | HIGH QUANTITY | NONE            |
|       9530 |     9994 | HIGH QUANTITY | NONE            |
|      15999 |     7790 | HIGH QUANTITY | NONE            |
|      35997 |     7853 | HIGH QUANTITY | NONE            |
|      36453 |     9185 | HIGH QUANTITY | NONE            |
|      37372 |     9939 | HIGH QUANTITY | NONE            |
|      39816 |     7664 | HIGH QUANTITY | NONE            |
|      41404 |     7560 | HIGH QUANTITY | NONE            |
|      17172 |     7504 | HIGH QUANTITY | NONE            |
|      19571 |     7756 | HIGH QUANTITY | NONE            |
|      25319 |     8821 | HIGH QUANTITY | NONE            |
|      26749 |     7835 | HIGH QUANTITY | NONE            |
|      31111 |     9810 | HIGH QUANTITY | NONE            |
|      32313 |     9495 | HIGH QUANTITY | NONE            |
+------------+----------+---------------+-----------------+
22 rows in set (0.01 sec)
```

problem statement7:

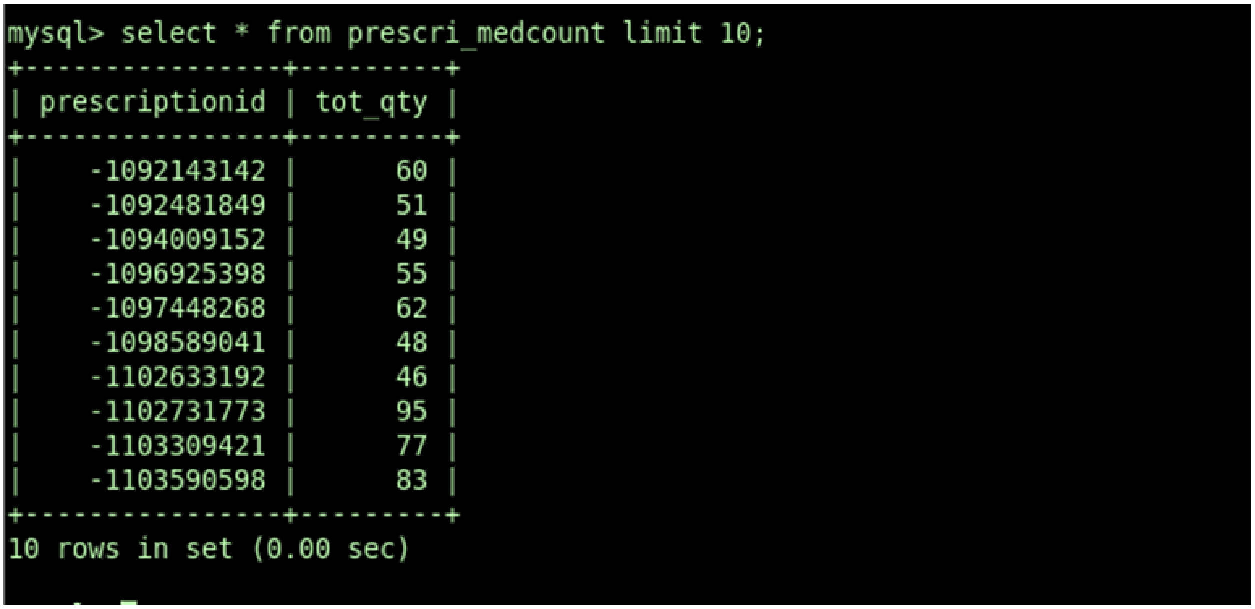
The total quantity of medicine in a prescription is the sum of the quantity of all the medicines in the prescription.
Select the prescriptions for which the total quantity of medicine exceeds
the avg of the total quantity of medicines for all the prescriptions.

```
    create external table prescri_medcount(prescriptionid int,tot_qty int)
    ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    LINES TERMINATED BY '\n';

    insert into table prescri_medcount
    select prescriptionid,tot_qty from
        (select pr.prescriptionid,sum(c.quantity) as tot_qty,
        avg(sum(c.quantity)) over() as avg_qty
        from
        pharmacy ph inner join Prescription pr on ph.pharmacyid=pr.pharmacyid
        inner join contain c on c.prescriptionid=pr.prescriptionid
        group by pr.prescriptionid) D
    where tot_qty > avg_qty;

--in mysqldb:
    create table prescri_medcount(prescriptionid int,tot_qty int);
```

sqoop export --connect jdbc:mysql://localhost:3306/results --username root --password cloudera --table prescri_medcount --export-dir /user/hive/warehouse/prescri_medcount --input-fields-terminated-by ','



```
2023-03-15 02:07:59     Uploaded 1 File to: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_02-07-52_685_7867133932595843555-1/-local-10005/HashTable-Stage-3/MapJoin-mapfile50--.hashtable (4540
2023-03-15 02:07:59     End of local task; Time Taken: 2.198 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1678869251323_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678869251323_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1678869251323_0006
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2023-03-15 02:08:08,904 Stage-3 map = 0%,  reduce = 0%
2023-03-15 02:08:17,905 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 3.21 sec
2023-03-15 02:08:27,685 Stage-3 map = 100%,  reduce = 100%, Cumulative CPU 5.85 sec
MapReduce Total cumulative CPU time: 5 seconds 850 msec
Ended Job = job_1678869251323_0006
Launching Job 2 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1678869251323_0007, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678869251323_0007/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1678869251323_0007
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 1
2023-03-15 02:08:37,352 Stage-4 map = 0%,  reduce = 0%
2023-03-15 02:08:46,432 Stage-4 map = 100%,  reduce = 0%, Cumulative CPU 2.41 sec
2023-03-15 02:08:57,466 Stage-4 map = 100%,  reduce = 100%, Cumulative CPU 5.24 sec
MapReduce Total cumulative CPU time: 5 seconds 240 msec
Ended Job = job_1678869251323_0007
Loading data to table default.prescri_medcount
Table default.prescri_medcount stats: [numFiles=1, numRows=5979, totalSize=83839, rawDataSize=77860]
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1  Reduce: 1   Cumulative CPU: 5.85 sec   HDFS Read: 334642 HDFS Write: 315378 SUCCESS
Stage-Stage-4: Map: 1  Reduce: 1   Cumulative CPU: 5.24 sec   HDFS Read: 322912 HDFS Write: 83924 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 90 msec
OK
Time taken: 66.184 seconds
hive> 
```



```
mysql> select * from prescri_medcount limit 10;
+----------------+---------+
| prescriptionid | tot_qty |
+----------------+---------+
|    -1092143142 |      60 |
|    -1092481849 |      51 |
|    -1094009152 |      49 |
|    -1096925398 |      55 |
|    -1097448268 |      62 |
|    -1098589041 |      48 |
|    -1102633192 |      46 |
|    -1102731773 |      95 |
|    -1103309421 |      77 |
|    -1103590598 |      83 |
+----------------+---------+
10 rows in set (0.00 sec)
```

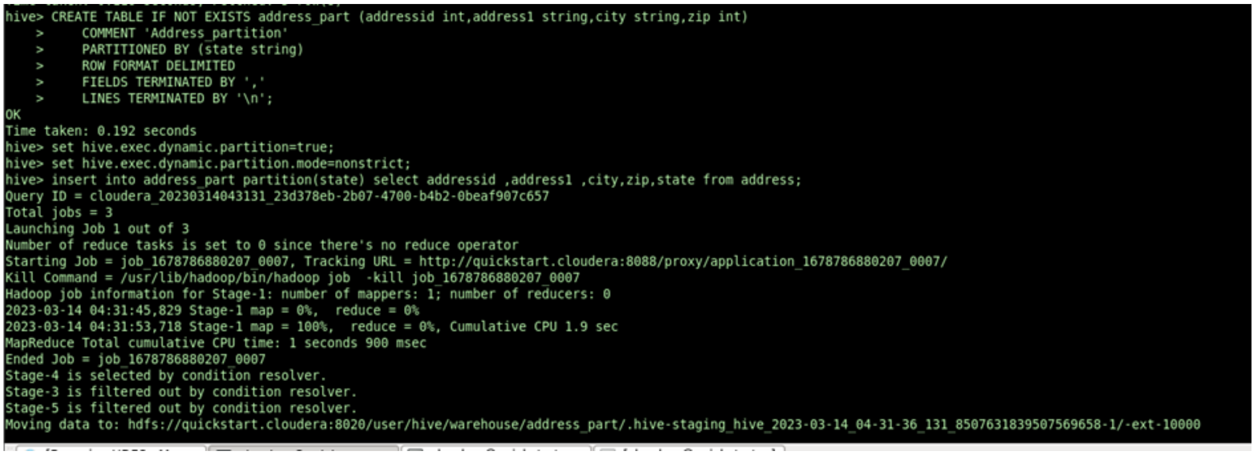
Problem Statement8:

The State of Alabama (AL) is trying to manage its healthcare resources more efficiently. For each city in their state, they need to identify the disease for which the maximum number of patients have gone for treatment. Assist the state for this purpose.

Note: The state of Alabama is represented as AL in Address Table.

----------address table partition---------

```
   CREATE TABLE IF NOT EXISTS address_part (addressid int,address1 string,city
string,zip int)
       COMMENT 'Address_partition'
       PARTITIONED BY (state string)
       ROW FORMAT DELIMITED
       FIELDS TERMINATED BY ','
       LINES TERMINATED BY '\n';

   insert into address_part partition(state) select addressid ,address1 ,city,zip,state from
address;
```



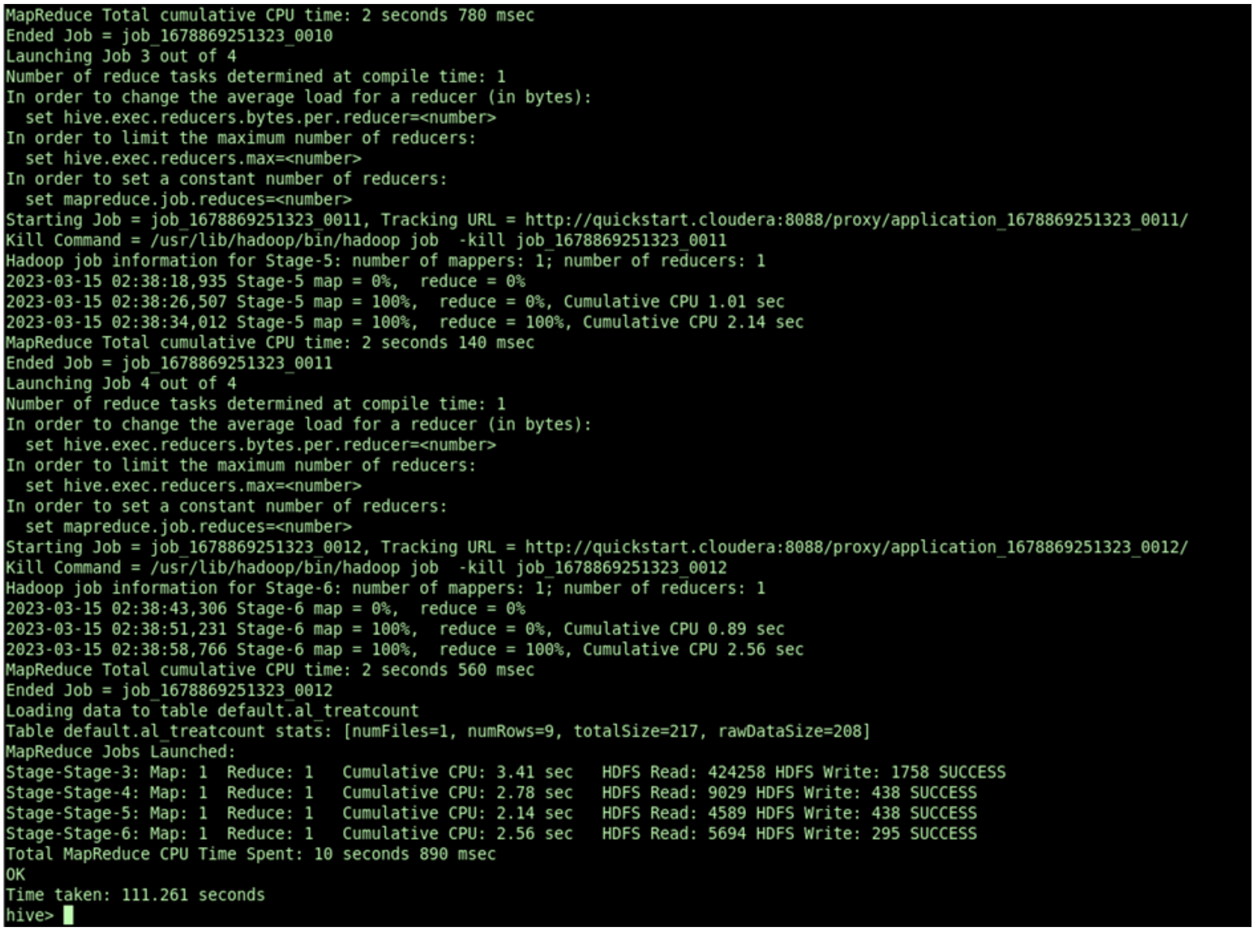
```
hive> CREATE TABLE IF NOT EXISTS address_part (addressid int,address1 string,city string,zip int)
    >     COMMENT 'Address_partition'
    >     PARTITIONED BY (state string)
    >     ROW FORMAT DELIMITED
    >     FIELDS TERMINATED BY ','
    >     LINES TERMINATED BY '\n';
OK
Time taken: 0.192 seconds
hive> set hive.exec.dynamic.partition=true;
hive> set hive.exec.dynamic.partition.mode=nonstrict;
hive> insert into address_part partition(state) select addressid ,address1 ,city,zip,state from address;
Query ID = cloudera_20230314043131_23d378eb-2b07-4700-b4b2-0beaf907c657
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1678786880207_0007, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678786880207_0007/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1678786880207_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-03-14 04:31:45,829 Stage-1 map = 0%,  reduce = 0%
2023-03-14 04:31:53,718 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.9 sec
MapReduce Total cumulative CPU time: 1 seconds 900 msec
Ended Job = job_1678786880207_0007
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/address_part/.hive-staging_hive_2023-03-14_04-31-36_131_8507631839507569658-1/-ext-10000
```

```
   create external table AL_treatcount(city string,diseaseid int,treat_cnt int)
   ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
   LINES TERMINATED BY '\n';

   insert into table AL_treatcount
   select city,diseaseid,treat_cnt
   from
      (select a.city,t.diseaseid,count( t.treatmentid) as treat_cnt,
      dense_rank() over(partition by a.city order by count( t.treatmentid) desc) as drnk
      from treatment t inner join person p on t.patientid=p.personid
      inner join address_part a on p.addressid=a.addressid
      where a.state="AL"
      group by a.city,t.diseaseid
      order by a.city asc ) D
```

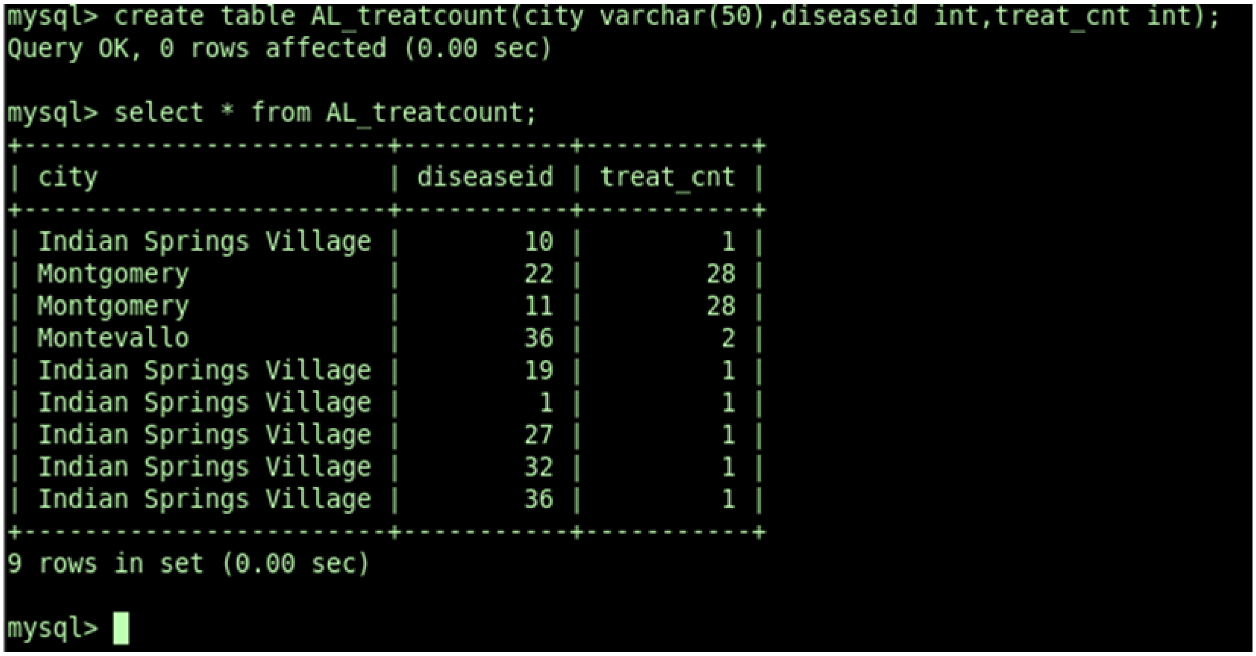
```
    where drnk=1
    order by treat_cnt desc;
```

```
--in mysqldb:
    create table AL_treatcount(city varchar(50),diseaseid int,treat_cnt int);
```

```
sqoop export --connect jdbc:mysql://localhost:3306/results --username root --password
cloudera --table AL_treatcount --export-dir /user/hive/warehouse/al_treatcount
--input-fields-terminated-by ','
```



```
MapReduce Total cumulative CPU time: 2 seconds 780 msec
Ended Job = job_1678869251323_0010
Launching Job 3 out of 4
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1678869251323_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678869251323_0011/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1678869251323_0011
Hadoop job information for Stage-5: number of mappers: 1; number of reducers: 1
2023-03-15 02:38:18,935 Stage-5 map = 0%,  reduce = 0%
2023-03-15 02:38:26,507 Stage-5 map = 100%,  reduce = 0%, Cumulative CPU 1.01 sec
2023-03-15 02:38:34,012 Stage-5 map = 100%,  reduce = 100%, Cumulative CPU 2.14 sec
MapReduce Total cumulative CPU time: 2 seconds 140 msec
Ended Job = job_1678869251323_0011
Launching Job 4 out of 4
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1678869251323_0012, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678869251323_0012/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1678869251323_0012
Hadoop job information for Stage-6: number of mappers: 1; number of reducers: 1
2023-03-15 02:38:43,306 Stage-6 map = 0%,  reduce = 0%
2023-03-15 02:38:51,231 Stage-6 map = 100%,  reduce = 0%, Cumulative CPU 0.89 sec
2023-03-15 02:38:58,766 Stage-6 map = 100%,  reduce = 100%, Cumulative CPU 2.56 sec
MapReduce Total cumulative CPU time: 2 seconds 560 msec
Ended Job = job_1678869251323_0012
Loading data to table default.al_treatcount
Table default.al_treatcount stats: [numFiles=1, numRows=9, totalSize=217, rawDataSize=208]
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1  Reduce: 1   Cumulative CPU: 3.41 sec   HDFS Read: 424258 HDFS Write: 1758 SUCCESS
Stage-Stage-4: Map: 1  Reduce: 1   Cumulative CPU: 2.78 sec   HDFS Read: 9029 HDFS Write: 438 SUCCESS
Stage-Stage-5: Map: 1  Reduce: 1   Cumulative CPU: 2.14 sec   HDFS Read: 4589 HDFS Write: 438 SUCCESS
Stage-Stage-6: Map: 1  Reduce: 1   Cumulative CPU: 2.56 sec   HDFS Read: 5694 HDFS Write: 295 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 890 msec
OK
Time taken: 111.261 seconds
hive>
```



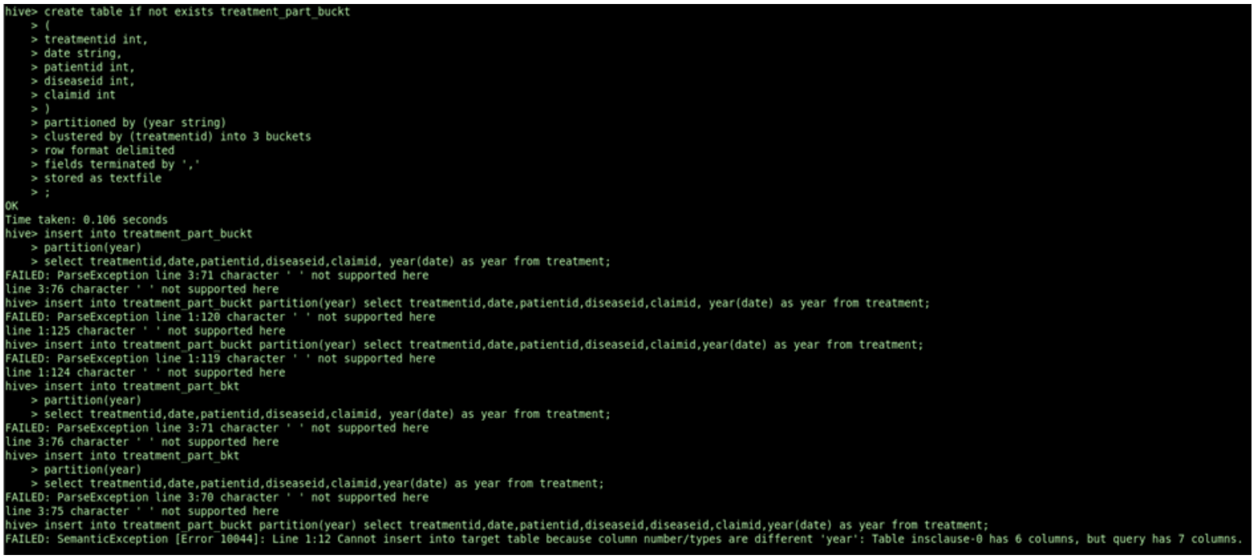
```
mysql> create table AL_treatcount(city varchar(50),diseaseid int,treat_cnt int);
Query OK, 0 rows affected (0.00 sec)

mysql> select * from AL_treatcount;
+------------------------+-----------+-----------+
| city                   | diseaseid | treat_cnt |
+------------------------+-----------+-----------+
| Indian Springs Village |        10 |         1 |
| Montgomery             |        22 |        28 |
| Montgomery             |        11 |        28 |
| Montevallo             |        36 |         2 |
| Indian Springs Village |        19 |         1 |
| Indian Springs Village |         1 |         1 |
| Indian Springs Village |        27 |         1 |
| Indian Springs Village |        32 |         1 |
| Indian Springs Village |        36 |         1 |
+------------------------+-----------+-----------+
9 rows in set (0.00 sec)

mysql> █
```

Problem Statement9:

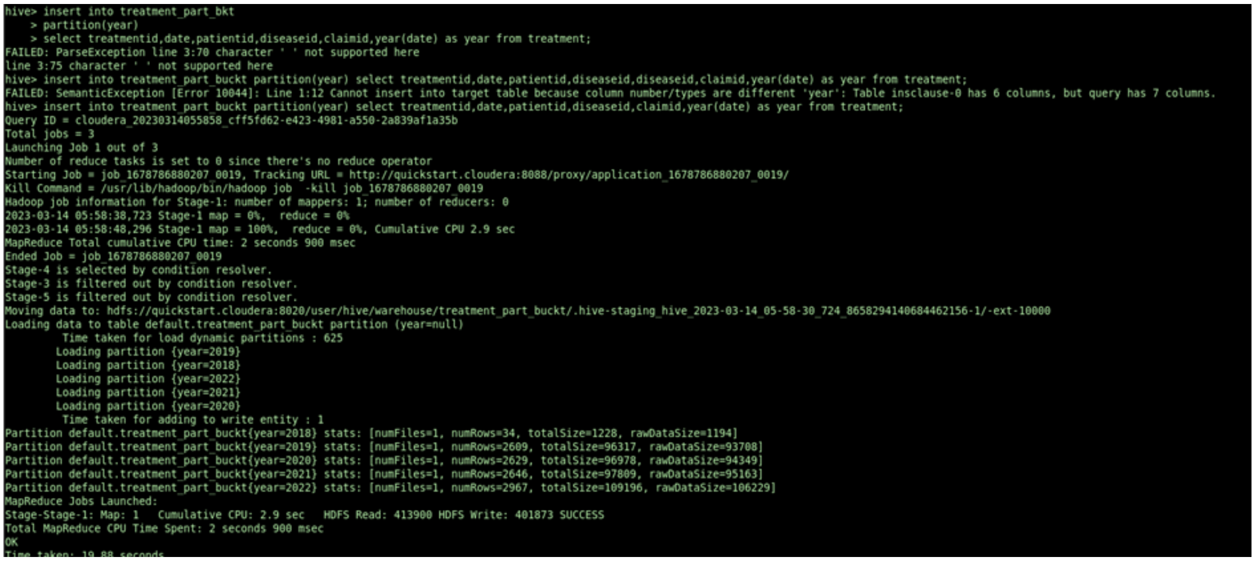
**The healthcare department wants a pharmacy report on the percentage of hospital-exclusive
medicine prescribed in the year 2022.
Assist the healthcare department to view for each pharmacy,
the pharmacy id, pharmacy name, total quantity of medicine prescribed in 2022,
total quantity of hospital-exclusive medicine prescribed by the pharmacy in 2022,
and the percentage of hospital-exclusive medicine to the total medicine prescribed in 2022.
Order the result in descending order of the percentage found.**

```
-----------partition & buckets on treatment----------------

create table if not exists treatment_part_buckt
(
treatmentid int,
date string,
patientid int,
diseaseid int,
claimid int
)
partitioned by (year string)
clustered by (treatmentid) into 3 buckets
row format delimited
```

```
fields terminated by ','
stored as textfile;

insert into treatment_part_bkt
partition(year)
select treatmentid,date,patientid,diseaseid,claimid,year(date) as year from treatment;

insert into treatment_part_buckt partition(year) select
treatmentid,date,patientid,diseaseid,claimid,year(date) as year from treatment;
```



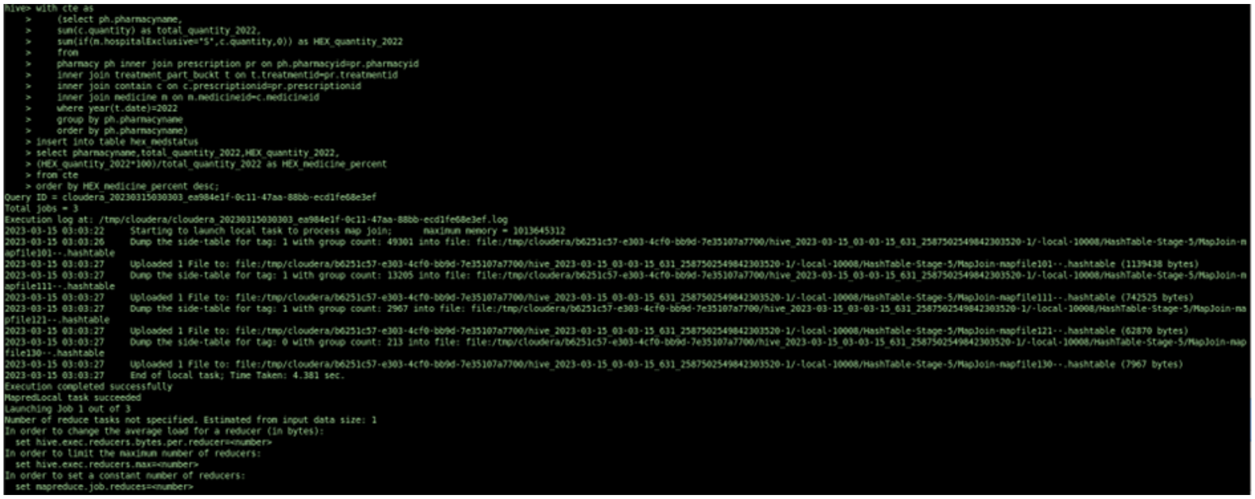
```
hive> create table if not exists treatment_part_buckt
    > (
    > treatmentid int,
    > date string,
    > patientid int,
    > diseaseid int,
    > claimid int
    > )
    > partitioned by (year string)
    > clustered by (treatmentid) into 3 buckets
    > row format delimited
    > fields terminated by ','
    > stored as textfile
    > ;
OK
Time taken: 0.106 seconds
hive> insert into treatment_part_buckt
    > partition(year)
    > select treatmentid,date,patientid,diseaseid,claimid, year(date) as year from treatment;
FAILED: ParseException line 3:71 character ' ' not supported here
line 3:76 character ' ' not supported here
hive> insert into treatment_part_buckt partition(year) select treatmentid,date,patientid,diseaseid,claimid, year(date) as year from treatment;
FAILED: ParseException line 1:120 character ' ' not supported here
line 1:125 character ' ' not supported here
hive> insert into treatment_part_buckt partition(year) select treatmentid,date,patientid,diseaseid,claimid,year(date) as year from treatment;
FAILED: ParseException line 1:119 character ' ' not supported here
line 1:124 character ' ' not supported here
hive> insert into treatment_part_bkt
    > partition(year)
    > select treatmentid,date,patientid,diseaseid,claimid, year(date) as year from treatment;
FAILED: ParseException line 3:71 character ' ' not supported here
line 3:76 character ' ' not supported here
hive> insert into treatment_part_bkt
    > partition(year)
    > select treatmentid,date,patientid,diseaseid,claimid,year(date) as year from treatment;
FAILED: ParseException line 3:70 character ' ' not supported here
line 3:75 character ' ' not supported here
hive> insert into treatment_part_buckt partition(year) select treatmentid,date,patientid,diseaseid,diseaseid,claimid,year(date) as year from treatment;
FAILED: SemanticException [Error 10044]: Line 1:12 Cannot insert into target table because column number/types are different 'year': Table insclause-0 has 6 columns, but query has 7 columns.
```



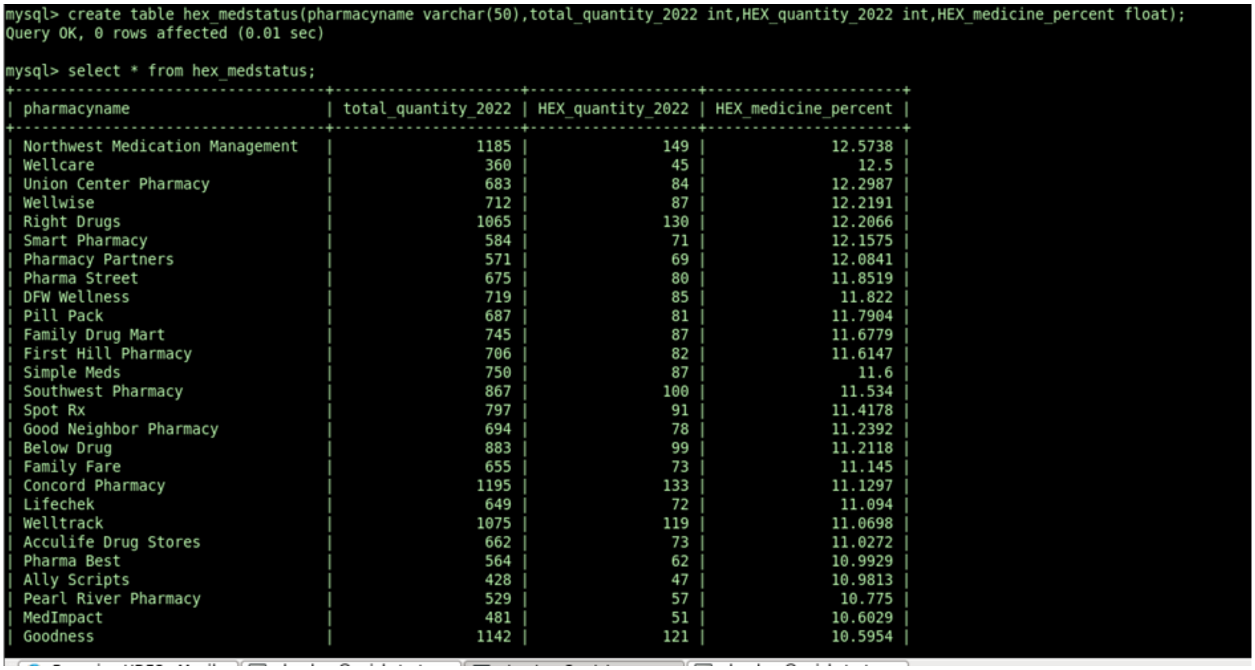
```
hive> insert into treatment_part_bkt
    > partition(year)
    > select treatmentid,date,patientid,diseaseid,claimid,year(date) as year from treatment;
FAILED: ParseException line 3:70 character ' ' not supported here
line 3:75 character ' ' not supported here
hive> insert into treatment_part_buckt partition(year) select treatmentid,date,patientid,diseaseid,diseaseid,claimid,year(date) as year from treatment;
FAILED: SemanticException [Error 10044]: Line 1:12 Cannot insert into target table because column number/types are different 'year': Table insclause-0 has 6 columns, but query has 7 columns.
hive> insert into treatment_part_buckt partition(year) select treatmentid,date,patientid,diseaseid,claimid,year(date) as year from treatment;
Query ID = cloudera_20230314055858_cff5fd62-e423-4981-a550-2a839af1a35b
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1678786880207_0019, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678786880207_0019/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1678786880207_0019
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-03-14 05:58:38,723 Stage-1 map = 0%,  reduce = 0%
2023-03-14 05:58:48,296 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.9 sec
MapReduce Total cumulative CPU time: 2 seconds 900 msec
Ended Job = job_1678786880207_0019
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/treatment_part_buckt/.hive-staging_hive_2023-03-14_05-58-30_724_8658294140684462156-1/-ext-10000
Loading data to table default.treatment_part_buckt partition (year=null)
         Time taken for load dynamic partitions : 625
        Loading partition {year=2019}
        Loading partition {year=2018}
        Loading partition {year=2022}
        Loading partition {year=2021}
        Loading partition {year=2020}
         Time taken for adding to write entity : 1
Partition default.treatment_part_buckt{year=2018} stats: [numFiles=1, numRows=34, totalSize=1228, rawDataSize=1194]
Partition default.treatment_part_buckt{year=2019} stats: [numFiles=1, numRows=2609, totalSize=96317, rawDataSize=93708]
Partition default.treatment_part_buckt{year=2020} stats: [numFiles=1, numRows=2629, totalSize=96978, rawDataSize=94349]
Partition default.treatment_part_buckt{year=2021} stats: [numFiles=1, numRows=2646, totalSize=97809, rawDataSize=95163]
Partition default.treatment_part_buckt{year=2022} stats: [numFiles=1, numRows=2967, totalSize=109196, rawDataSize=106229]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 2.9 sec   HDFS Read: 413900 HDFS Write: 401873 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 900 msec
OK
Time taken: 19.88 seconds
```

```
---------------------------------------------------------

create external table hex_medstatus(pharmacyname string,total_quantity_2022
int,HEX_quantity_2022 int,HEX_medicine_percent float)
```

```
   ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
   LINES TERMINATED BY '\n';


   with cte as
      (select ph.pharmacyname,
      sum(c.quantity) as total_quantity_2022,
      sum(if(m.hospitalExclusive="S",c.quantity,0)) as HEX_quantity_2022
      from
      pharmacy ph inner join prescription pr on ph.pharmacyid=pr.pharmacyid
      inner join treatment_part_buckt t on t.treatmentid=pr.treatmentid
      inner join contain c on c.prescriptionid=pr.prescriptionid
      inner join medicine m on m.medicineid=c.medicineid
      where year(t.date)=2022
      group by ph.pharmacyname
      order by ph.pharmacyname)
   insert into table hex_medstatus
   select pharmacyname,total_quantity_2022,HEX_quantity_2022,
   (HEX_quantity_2022*100)/total_quantity_2022 as HEX_medicine_percent
   from cte
   order by HEX_medicine_percent desc;

--in mysqldb

   create table hex_medstatus(pharmacyname varchar(50),total_quantity_2022
int,HEX_quantity_2022 int,HEX_medicine_percent float);

sqoop export --connect jdbc:mysql://localhost:3306/results --username root --password
cloudera --table hex_medstatus --export-dir /user/hive/warehouse/hex_medstatus
--input-fields-terminated-by ','
```



```
hive> with cte as
    >     (select ph.pharmacyname,
    >     sum(c.quantity) as total_quantity_2022,
    >     sum(if(m.hospitalExclusive="S",c.quantity,0)) as HEX_quantity_2022
    >     from
    >     pharmacy ph inner join prescription pr on ph.pharmacyid=pr.pharmacyid
    >     inner join treatment_part_buckt t on t.treatmentid=pr.treatmentid
    >     inner join contain c on c.prescriptionid=pr.prescriptionid
    >     inner join medicine m on m.medicineid=c.medicineid
    >     where year(t.date)=2022
    >     group by ph.pharmacyname
    >     order by ph.pharmacyname)
    > insert into table hex_medstatus
    > select pharmacyname,total_quantity_2022,HEX_quantity_2022,
    > (HEX_quantity_2022*100)/total_quantity_2022 as HEX_medicine_percent
    > from cte
    > order by HEX_medicine_percent desc;
Query ID = cloudera_20230315030303_ea984e1f-0c11-47aa-88bb-ecd1fe68e3ef
Total jobs = 3
Execution log at: /tmp/cloudera/cloudera_20230315030303_ea984e1f-0c11-47aa-88bb-ecd1fe68e3ef.log
2023-03-15 03:03:22     Starting to launch local task to process map join;      maximum memory = 1013645312
2023-03-15 03:03:26     Dump the side-table for tag: 1 with group count: 49301 into file: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-03-15_631_2587502549842303520-1/-local-10008/HashTable-Stage-5/MapJoin-m
apfile101--.hashtable
2023-03-15 03:03:27     Uploaded 1 File to: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-03-15_631_2587502549842303520-1/-local-10008/HashTable-Stage-5/MapJoin-mapfile101--.hashtable (1139438 bytes)
2023-03-15 03:03:27     Dump the side-table for tag: 1 with group count: 13205 into file: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-03-15_631_2587502549842303520-1/-local-10008/HashTable-Stage-5/MapJoin-m
apfile111--.hashtable
2023-03-15 03:03:27     Uploaded 1 File to: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-03-15_631_2587502549842303520-1/-local-10008/HashTable-Stage-5/MapJoin-mapfile111--.hashtable (742525 bytes)
2023-03-15 03:03:27     Dump the side-table for tag: 1 with group count: 2967 into file: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-03-15_631_2587502549842303520-1/-local-10008/HashTable-Stage-5/MapJoin-ma
pfile121--.hashtable
2023-03-15 03:03:27     Uploaded 1 File to: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-03-15_631_2587502549842303520-1/-local-10008/HashTable-Stage-5/MapJoin-mapfile121--.hashtable (62870 bytes)
2023-03-15 03:03:27     Dump the side-table for tag: 0 with group count: 213 into file: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-03-15_631_2587502549842303520-1/-local-10008/HashTable-Stage-5/MapJoin-map
file130--.hashtable
2023-03-15 03:03:27     Uploaded 1 File to: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-03-15_631_2587502549842303520-1/-local-10008/HashTable-Stage-5/MapJoin-mapfile130--.hashtable (7967 bytes)
2023-03-15 03:03:27     End of local task; Time Taken: 4.381 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 3
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
```



```
mysql> create table hex_medstatus(pharmacyname varchar(50),total_quantity_2022 int,HEX_quantity_2022 int,HEX_medicine_percent float);
Query OK, 0 rows affected (0.01 sec)

mysql> select * from hex_medstatus;
+---------------------------------+---------------------+-------------------+----------------------+
| pharmacyname                    | total_quantity_2022 | HEX_quantity_2022 | HEX_medicine_percent |
+---------------------------------+---------------------+-------------------+----------------------+
| Northwest Medication Management |                1185 |               149 |              12.5738 |
| Wellcare                        |                 360 |                45 |                 12.5 |
| Union Center Pharmacy           |                 683 |                84 |              12.2987 |
| Wellwise                        |                 712 |                87 |              12.2191 |
| Right Drugs                     |                1065 |               130 |              12.2066 |
| Smart Pharmacy                  |                 584 |                71 |              12.1575 |
| Pharmacy Partners               |                 571 |                69 |              12.0841 |
| Pharma Street                   |                 675 |                80 |              11.8519 |
| DFW Wellness                    |                 719 |                85 |               11.822 |
| Pill Pack                       |                 687 |                81 |              11.7904 |
| Family Drug Mart                |                 745 |                87 |              11.6779 |
| First Hill Pharmacy             |                 706 |                82 |              11.6147 |
| Simple Meds                     |                 750 |                87 |                 11.6 |
| Southwest Pharmacy              |                 867 |               100 |               11.534 |
| Spot Rx                         |                 797 |                91 |              11.4178 |
| Good Neighbor Pharmacy          |                 694 |                78 |              11.2392 |
| Below Drug                      |                 883 |                99 |              11.2118 |
| Family Fare                     |                 655 |                73 |               11.145 |
| Concord Pharmacy                |                1195 |               133 |              11.1297 |
| Lifechek                        |                 649 |                72 |               11.094 |
| Welltrack                       |                1075 |               119 |              11.0698 |
| Acculife Drug Stores            |                 662 |                73 |              11.0272 |
| Pharma Best                     |                 564 |                62 |              10.9929 |
| Ally Scripts                    |                 428 |                47 |              10.9813 |
| Pearl River Pharmacy            |                 529 |                57 |               10.775 |
| MedImpact                       |                 481 |                51 |              10.6029 |
| Goodness                        |                1142 |               121 |              10.5954 |
```

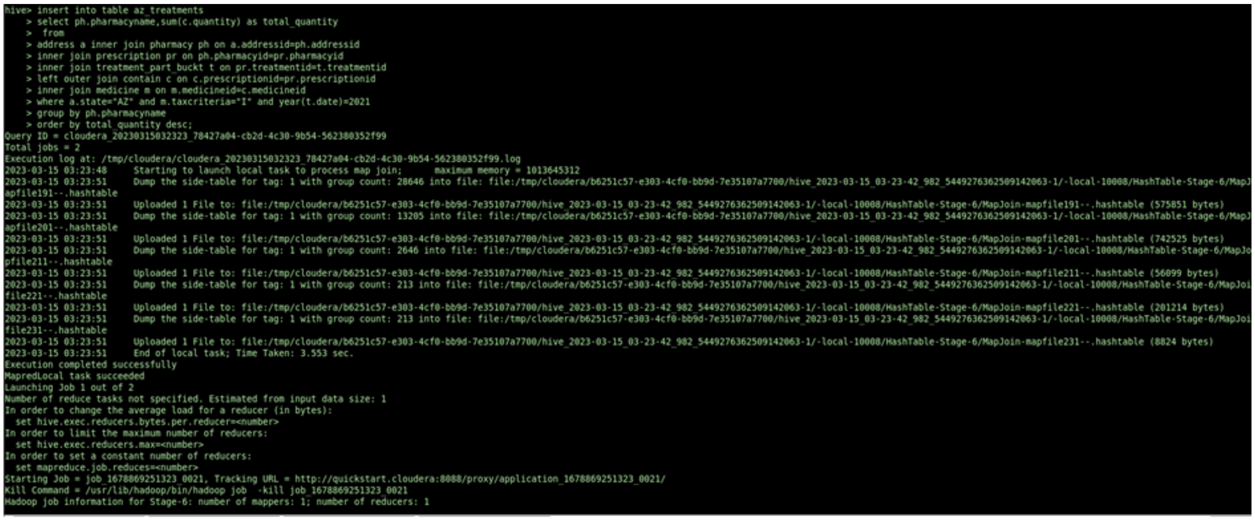
Problem Statement10:

Jhonny, from the finance department of Arizona(AZ), has requested a report that lists the total quantity of medicine each pharmacy in his state has prescribed that falls under Tax criteria I for treatments that took place in 2021. Assist Jhonny in generating the report.

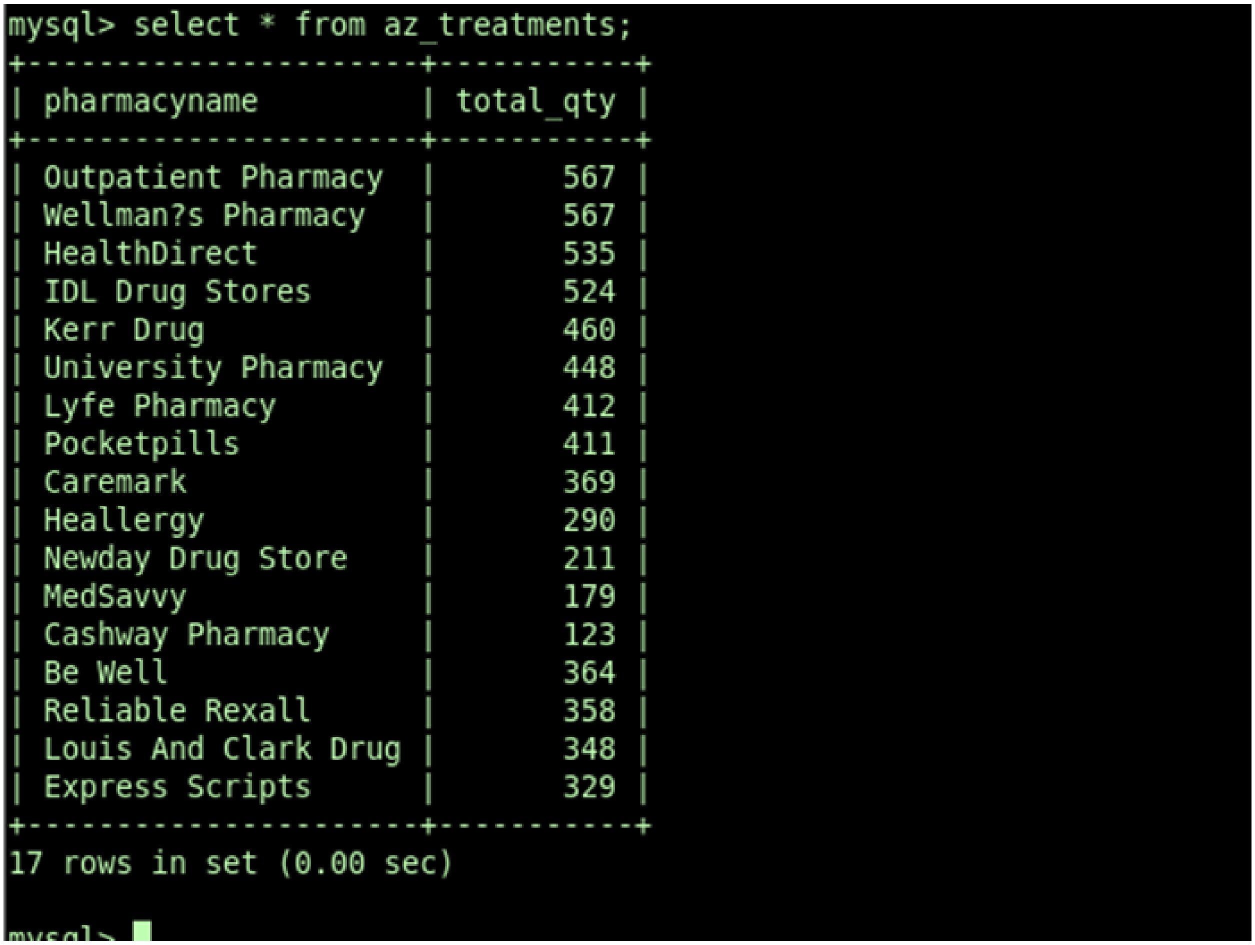
```
create external table az_treatments(pharmacyname string,total_qty int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n';

insert into table az_treatments
select ph.pharmacyname,sum(c.quantity) as total_quantity
from
address a inner join pharmacy ph on a.addressid=ph.addressid
inner join prescription pr on ph.pharmacyid=pr.pharmacyid
inner join treatment_part_buckt t on pr.treatmentid=t.treatmentid
left outer join contain c on c.prescriptionid=pr.prescriptionid
inner join medicine m on m.medicineid=c.medicineid
where a.state="AZ" and m.taxcriteria="I" and year(t.date)=2021
group by ph.pharmacyname
order by total_quantity desc;
```

```
--in mysqldb:
    create table az_treatments(pharmacyname varchar(50),total_qty int);
```

```
sqoop export --connect jdbc:mysql://localhost:3306/results --username root --password
cloudera --table az_treatments --export-dir /user/hive/warehouse/az_treatments
--input-fields-terminated-by ','
```



```
hive> insert into table az_treatments
    > select ph.pharmacyname,sum(c.quantity) as total_quantity
    >  from
    > address a inner join pharmacy ph on a.addressid=ph.addressid
    > inner join prescription pr on ph.pharmacyid=pr.pharmacyid
    > inner join treatment_part_buckt t on pr.treatmentid=t.treatmentid
    > left outer join contain c on c.prescriptionid=pr.prescriptionid
    > inner join medicine m on m.medicineid=c.medicineid
    > where a.state="AZ" and m.taxcriteria="I" and year(t.date)=2021
    > group by ph.pharmacyname
    > order by total_quantity desc;
Query ID = cloudera_20230315032323_78427a04-cb2d-4c30-9b54-562380352f99
Total jobs = 2
Execution log at: /tmp/cloudera/cloudera_20230315032323_78427a04-cb2d-4c30-9b54-562380352f99.log
2023-03-15 03:23:48     Starting to launch local task to process map join;      maximum memory = 1013645312
2023-03-15 03:23:51     Dump the side-table for tag: 1 with group count: 28646 into file: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-23-42_982_5449276362509142063-1/-local-10008/HashTable-Stage-6/MapJ
apfile191--.hashtable
2023-03-15 03:23:51     Uploaded 1 File to: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-23-42_982_5449276362509142063-1/-local-10008/HashTable-Stage-6/MapJoin-mapfile191--.hashtable (575851 bytes)
2023-03-15 03:23:51     Dump the side-table for tag: 1 with group count: 13205 into file: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-23-42_982_5449276362509142063-1/-local-10008/HashTable-Stage-6/MapJ
apfile201--.hashtable
2023-03-15 03:23:51     Uploaded 1 File to: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-23-42_982_5449276362509142063-1/-local-10008/HashTable-Stage-6/MapJoin-mapfile201--.hashtable (742525 bytes)
2023-03-15 03:23:51     Dump the side-table for tag: 1 with group count: 2646 into file: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-23-42_982_5449276362509142063-1/-local-10008/HashTable-Stage-6/MapJo
pfile211--.hashtable
2023-03-15 03:23:51     Uploaded 1 File to: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-23-42_982_5449276362509142063-1/-local-10008/HashTable-Stage-6/MapJoin-mapfile211--.hashtable (56099 bytes)
2023-03-15 03:23:51     Dump the side-table for tag: 1 with group count: 213 into file: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-23-42_982_5449276362509142063-1/-local-10008/HashTable-Stage-6/MapJoi
file221--.hashtable
2023-03-15 03:23:51     Uploaded 1 File to: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-23-42_982_5449276362509142063-1/-local-10008/HashTable-Stage-6/MapJoin-mapfile221--.hashtable (201214 bytes)
2023-03-15 03:23:51     Dump the side-table for tag: 1 with group count: 213 into file: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-23-42_982_5449276362509142063-1/-local-10008/HashTable-Stage-6/MapJoi
file231--.hashtable
2023-03-15 03:23:51     Uploaded 1 File to: file:/tmp/cloudera/b6251c57-e303-4cf0-bb9d-7e35107a7700/hive_2023-03-15_03-23-42_982_5449276362509142063-1/-local-10008/HashTable-Stage-6/MapJoin-mapfile231--.hashtable (8824 bytes)
2023-03-15 03:23:51     End of local task; Time Taken: 3.553 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1678869251323_0021, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678869251323_0021/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1678869251323_0021
Hadoop job information for Stage-6: number of mappers: 1; number of reducers: 1
```



```
mysql> select * from az_treatments;
+----------------------+-----------+
| pharmacyname         | total_qty |
+----------------------+-----------+
| Outpatient Pharmacy  |       567 |
| Wellman?s Pharmacy   |       567 |
| HealthDirect         |       535 |
| IDL Drug Stores      |       524 |
| Kerr Drug            |       460 |
| University Pharmacy  |       448 |
| Lyfe Pharmacy        |       412 |
| Pocketpills          |       411 |
| Caremark             |       369 |
| Heallergy            |       290 |
| Newday Drug Store    |       211 |
| MedSavvy             |       179 |
| Cashway Pharmacy     |       123 |
| Be Well              |       364 |
| Reliable Rexall      |       358 |
| Louis And Clark Drug |       348 |
| Express Scripts      |       329 |
+----------------------+-----------+
17 rows in set (0.00 sec)

mysql> 
```