

HIVE PROJECT

STEP 1:

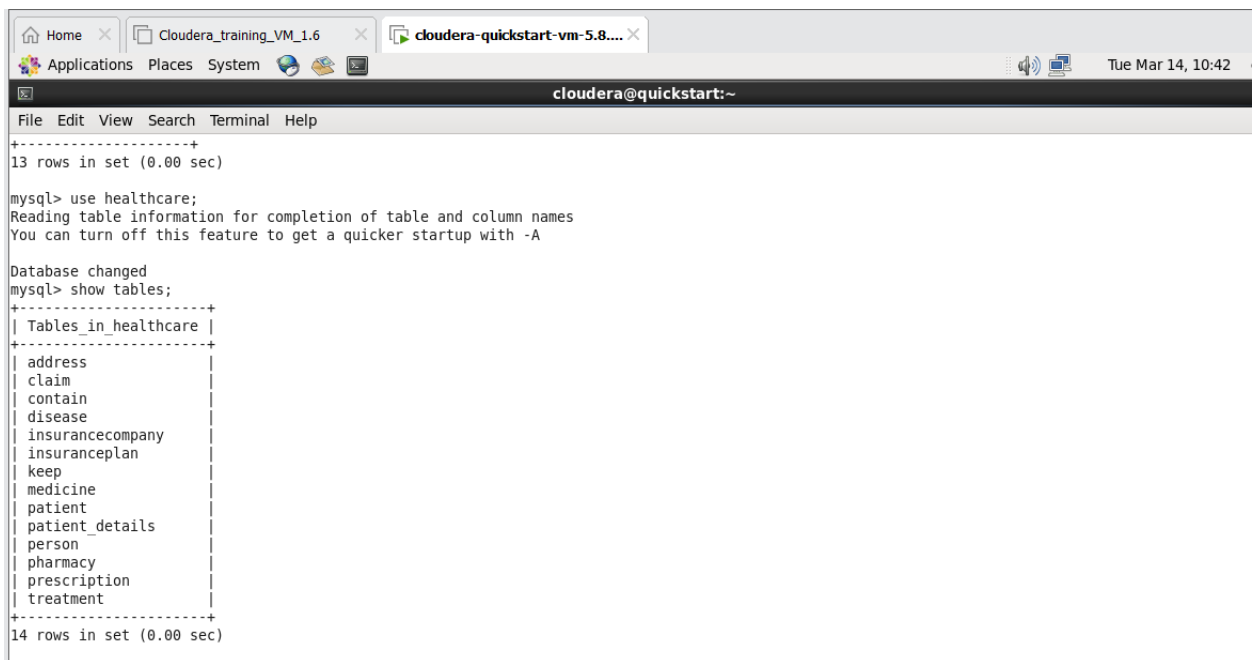
Import the SQL dump file in Cloudera MySQL Environment.

```
mysql -u root -p healthcare < healthcare_dump.sql
```

STEP 2:

Import all tables from MySQL to hive.

```
scoop import-all-tables --connect jdbc:mysql://localhost:3306/healthcare --username root  
--hive-import -m 1
```

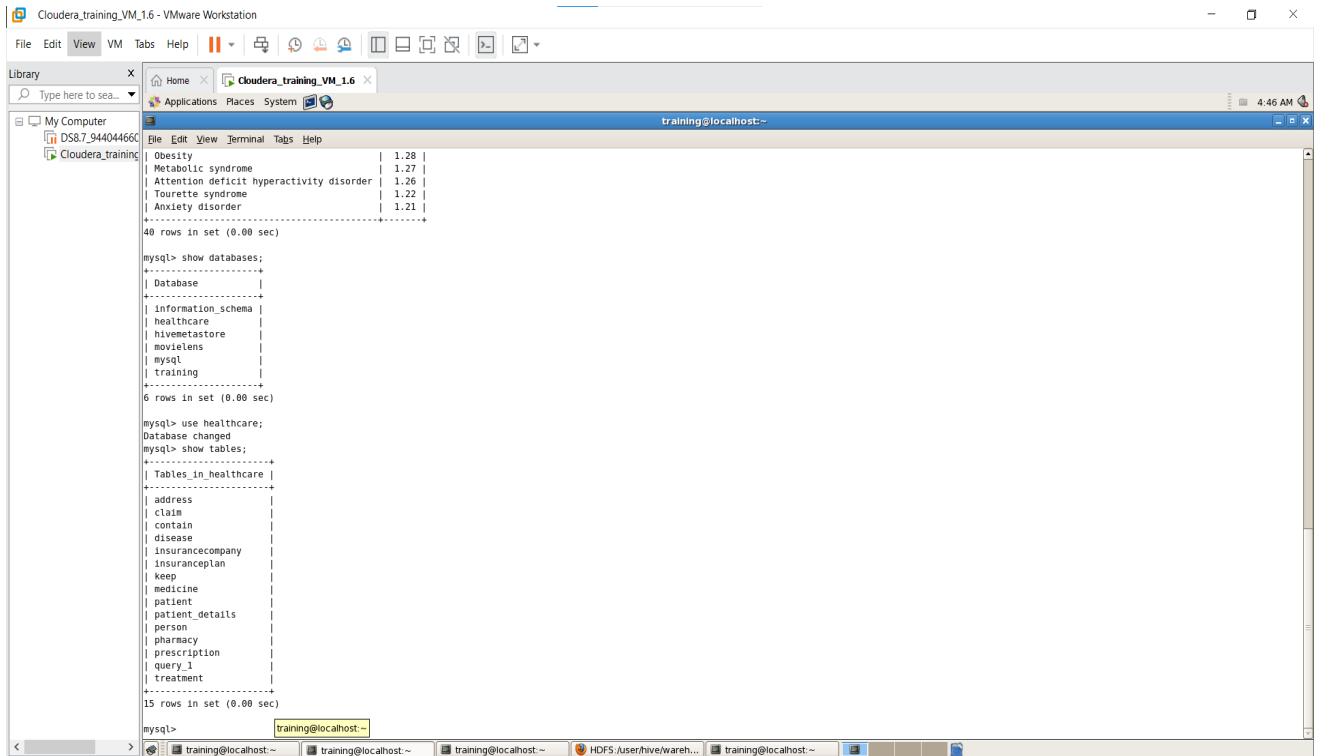


The screenshot shows a terminal window titled 'cloudera@quickstart:~'. The terminal output displays the results of a MySQL query. It starts with '13 rows in set (0.00 sec)', followed by 'mysql> use healthcare;' and a message about reading table information. Then, 'mysql> show tables;' is executed, resulting in a list of 14 tables: address, claim, contain, disease, insurancecompany, insuranceplan, keep, medicine, patient, patient_details, person, pharmacy, prescription, and treatment. The output ends with '14 rows in set (0.00 sec)'.

```
File Edit View Search Terminal Help
+-----+
13 rows in set (0.00 sec)

mysql> use healthcare;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> show tables;
+-----+
| Tables_in_healthcare |
+-----+
| address              |
| claim                |
| contain              |
| disease              |
| insurancecompany      |
| insuranceplan         |
| keep                  |
| medicine              |
| patient              |
| patient_details      |
| person                |
| pharmacy              |
| prescription          |
| treatment             |
+-----+
14 rows in set (0.00 sec)
```



#####

Problem Statement 1: Jimmy, from the healthcare department, has requested a report that shows how the number of treatments each age category of patients has gone through in the year 2022. The age category is as follows, Children (00-14 years), Youth (15-24 years), Adults (25-64 years), and Seniors (65 years and over). Assist Jimmy in generating the report.

Create External Table

```

create external table query1(counts int,category STRING)
row format delimited
fields terminated by ','
lines terminated by '\n'
LOCATION '/user/outputs/query1';

```

Insert Data Into External Table In Hive

```
INSERT OVERWRITE TABLE query1 SELECT COUNT(*), v1.category
FROM (
SELECT
CASE
WHEN YEAR(t1.date) - YEAR(dob) <= 14 THEN 'children'
WHEN YEAR(t1.date) - YEAR(dob) <= 24 THEN 'youth'
WHEN YEAR(t1.date) - YEAR(dob) <= 64 THEN 'adults'
ELSE 'senior citizen'
END AS category,
p.patientid AS patientid
FROM Patient p
INNER JOIN treatment t1 ON p.patientid = t1.patientid
WHERE YEAR(t1.date) = 2022
) AS v1
GROUP BY v1.category;
```

Browsing HDFS - Mozilla Firefox

localhost:50070/explorer.html#/user/outputs/query1

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/outputs/query1 Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rwxr-xr-x	cloudera	supergroup	53 B	Tue Mar 14 11:26:36 -0700 2023	1	128 MB	000000_0

Hadoop, 2016.

Firefox automatically sends some data to Mozilla so that we can improve your experience.

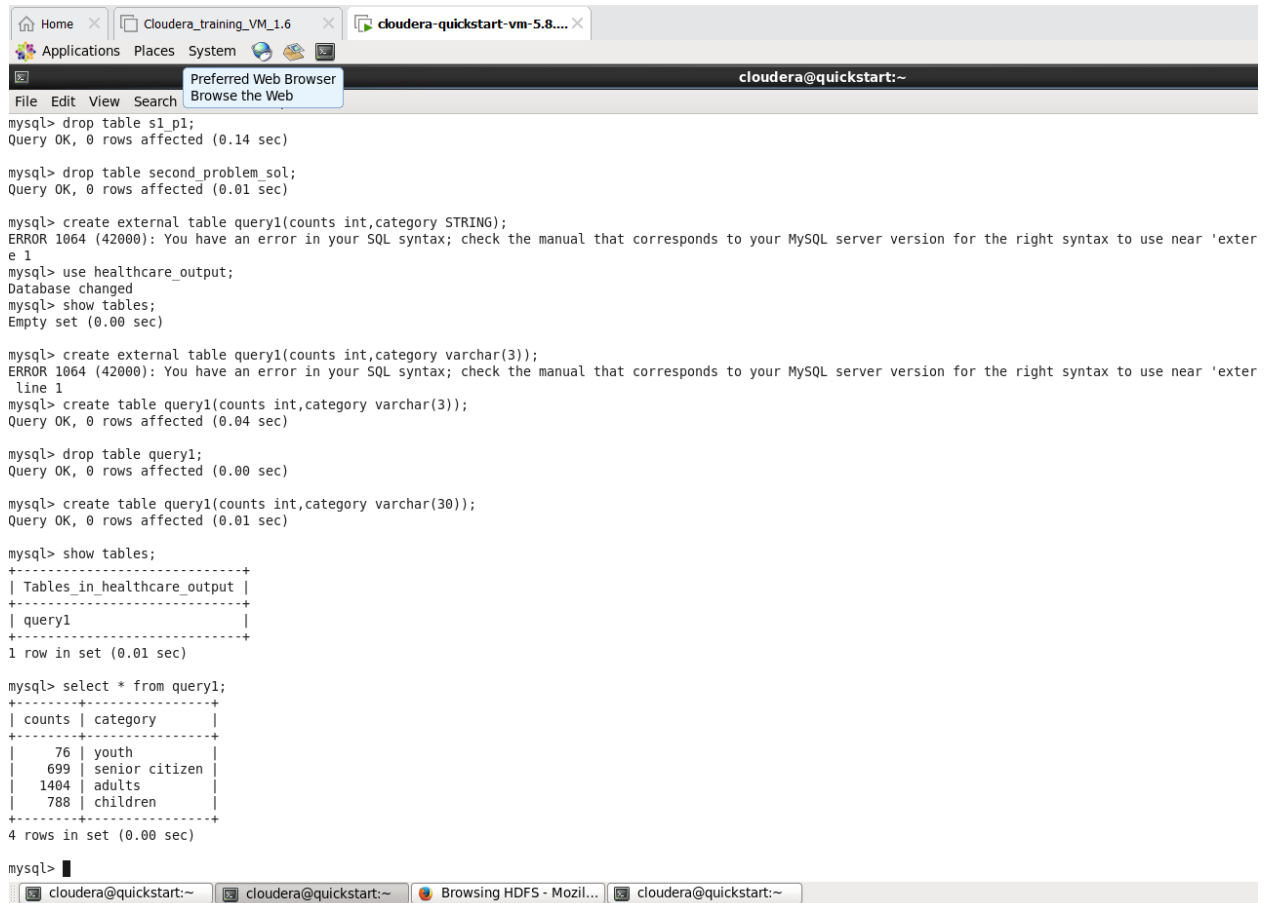
cloudera@quickstart:~ cloudera@quickstart:~ Browsing HDFS - Mozil... cloudera@quickstart:~

Create Output Table in Client DB

```
create table query1(counts int,category varchar(30));
```

Move Data to Client DB using Sqoop Export

```
sqoop export --connect jdbc:mysql://localhost:3306/healthcare_output  
--username root --password cludera --table query1  
--export-dir /user/outputs/query1/000000_0  
--input-fields-terminated-by ',';
```



```
mysql> drop table s1_p1;  
Query OK, 0 rows affected (0.14 sec)  
  
mysql> drop table second_problem_sol;  
Query OK, 0 rows affected (0.01 sec)  
  
mysql> create external table query1(counts int,category STRING);  
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near 'exter  
e 1  
mysql> use healthcare_output;  
Database changed  
mysql> show tables;  
Empty set (0.00 sec)  
  
mysql> create external table query1(counts int,category varchar(3));  
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near 'exter  
line 1  
mysql> create table query1(counts int,category varchar(3));  
Query OK, 0 rows affected (0.04 sec)  
  
mysql> drop table query1;  
Query OK, 0 rows affected (0.00 sec)  
  
mysql> create table query1(counts int,category varchar(30));  
Query OK, 0 rows affected (0.01 sec)  
  
mysql> show tables;  
+-----+  
| Tables_in_healthcare_output |  
+-----+  
| query1 |  
+-----+  
1 row in set (0.01 sec)  
  
mysql> select * from query1;  
+-----+  
| counts | category |  
+-----+  
| 76 | youth |  
| 699 | senior citizen |  
| 1404 | adults |  
| 788 | children |  
+-----+  
4 rows in set (0.00 sec)  
  
mysql>
```

#####

Problem Statement 2: Jimmy, from the healthcare department, wants to know which disease is infecting people of which gender more often.

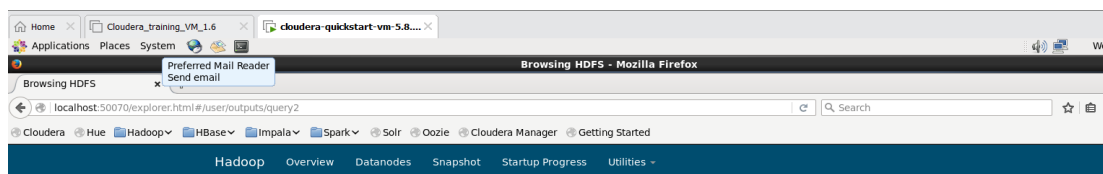
Assist Jimmy with this purpose by generating a report that shows for each disease the male-to-female ratio. Sort the data in a way that is helpful for Jimmy.

Create External Table

```
create external table query2 (diseasename varchar(50), malecount int, femalecount int,
malefemale double)
row format delimited
fields terminated by ','
lines terminated by '\n'
location '/user/outputs/query2';
```

Insert Data Into External Table In Hive

```
INSERT OVERWRITE TABLE query2 select diseasename,COUNT(IF(gender = 'male', 1, null))
count_male,
COUNT(IF(gender = 'female', 1, NULL)) count_female,
COUNT(IF(gender = 'male', 1, NULL))/COUNT(IF(gender = 'female', 1, NULL)) as ratio from
disease join treatment on treatment.diseaseid=disease.diseaseid
join patient on patient.patientid=treatment.patientid
join person on patient.patientid=person.personid
group by diseasename
order by diseasename ;
```



Browse Directory

/user/outputs/query2							Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rwxr-xr-x	cloudera	supergroup	1.68 KB	Wed Mar 15 02:14:56 -0700 2023	1	128 MB	000000_0

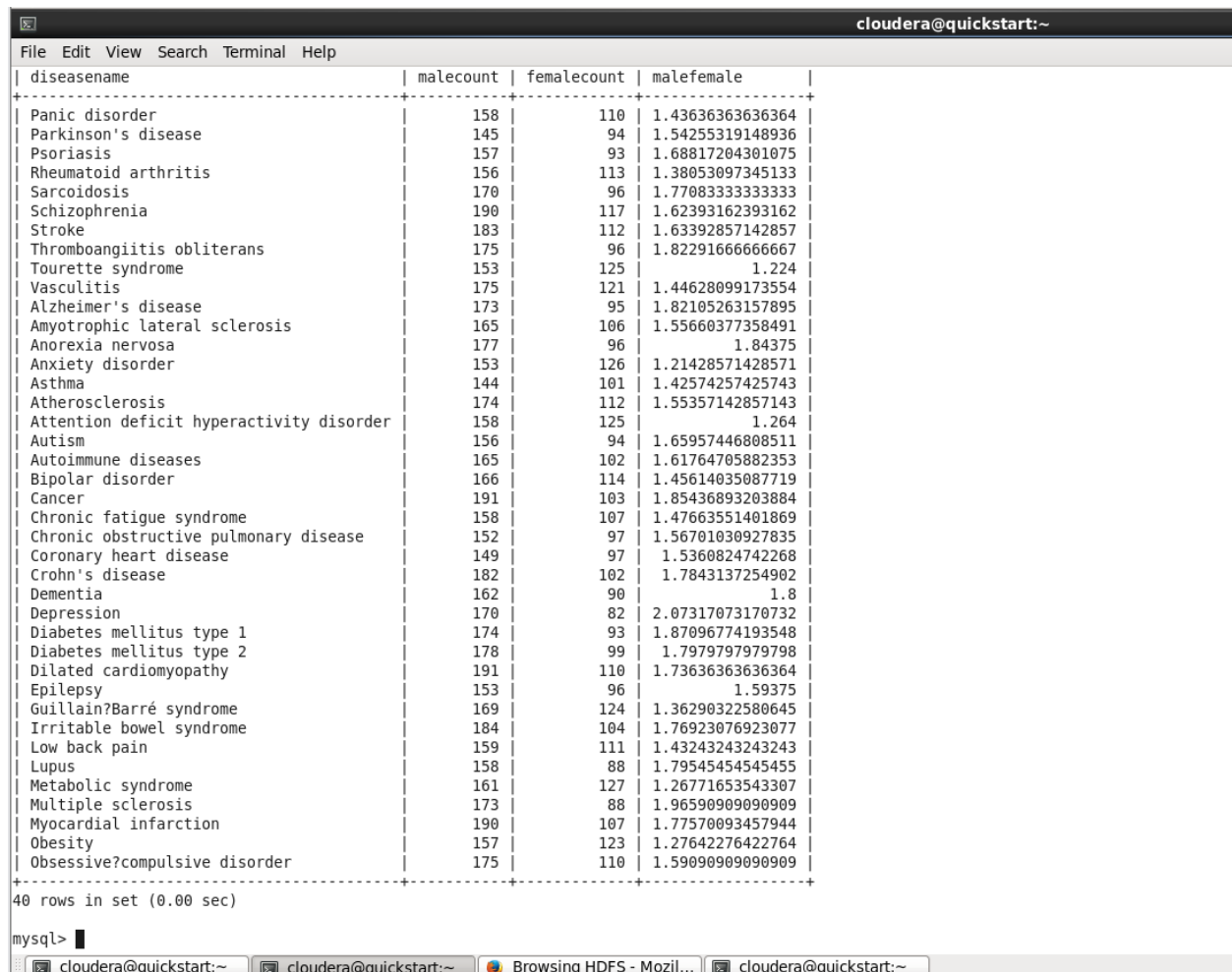
Hadoop, 2016.

Create Output Table in Client DB

create table query2(diseasename varchar(50), malecount int, femalecount int, malefemale double);

Move Data to Client DB using Sqoop Export

```
sqoop export --connect jdbc:mysql://localhost:3306/healthcare_output
--username root --password cloudera --table query2
--export-dir /user/outputs/query2/000000_0
--input-fields-terminated-by ',';
```



diseasename	malecount	femalecount	malefemale
Panic disorder	158	110	1.43636363636364
Parkinson's disease	145	94	1.54255319148936
Psoriasis	157	93	1.68817204301075
Rheumatoid arthritis	156	113	1.38053097345133
Sarcoidosis	170	96	1.77083333333333
Schizophrenia	190	117	1.62393162393162
Stroke	183	112	1.63392857142857
Thromboangiitis obliterans	175	96	1.82291666666667
Tourette syndrome	153	125	1.224
Vasculitis	175	121	1.44628099173554
Alzheimer's disease	173	95	1.82105263157895
Amyotrophic lateral sclerosis	165	106	1.55660377358491
Anorexia nervosa	177	96	1.84375
Anxiety disorder	153	126	1.21428571428571
Asthma	144	101	1.42574257425743
Atherosclerosis	174	112	1.55357142857143
Attention deficit hyperactivity disorder	158	125	1.264
Autism	156	94	1.65957446808511
Autoimmune diseases	165	102	1.61764705882353
Bipolar disorder	166	114	1.45614035087719
Cancer	191	103	1.85436893203884
Chronic fatigue syndrome	158	107	1.47663551401869
Chronic obstructive pulmonary disease	152	97	1.56701030927835
Coronary heart disease	149	97	1.5360824742268
Crohn's disease	182	102	1.7843137254902
Dementia	162	90	1.8
Depression	170	82	2.07317073170732
Diabetes mellitus type 1	174	93	1.87096774193548
Diabetes mellitus type 2	178	99	1.79797979797979
Dilated cardiomyopathy	191	110	1.73636363636364
Epilepsy	153	96	1.59375
Guillain-Barré syndrome	169	124	1.36290322580645
Irritable bowel syndrome	184	104	1.76923076923077
Low back pain	159	111	1.43243243243243
Lupus	158	88	1.79545454545455
Metabolic syndrome	161	127	1.26771653543307
Multiple sclerosis	173	88	1.96590909090909
Myocardial infarction	190	107	1.77570093457944
Obesity	157	123	1.27642276422764
Obsessive-compulsive disorder	175	110	1.59090909090909

40 rows in set (0.00 sec)

mysql>

#####

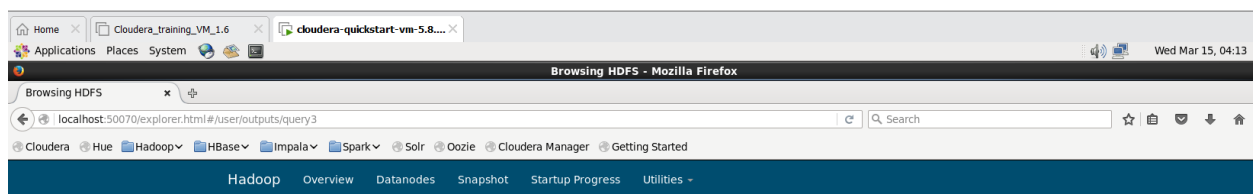
Problem Statement 3: Jacob, from insurance management, has noticed that insurance claims are not made for all the treatments. He also wants to figure out if the gender of the patient has any impact on the insurance claim. Assist Jacob in this situation by generating a report that finds for each gender the number of treatments, number of claims, and treatment-to-claim ratio. And notice if there is a significant difference between the treatment-to-claim ratio of male and female patients.

Create External Table

```
create external table query3
(gender string, count_claims int, count_treatments int, ration double)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ';'
LINES TERMINATED BY '\n'
LOCATION '/user/outputs/query3';
```

Insert Data Into External Table In Hive

```
WITH cte_table2 AS (
  SELECT pe.`gender` AS Gender, c.`claimID` AS Claims, t.`treatmentID` AS treatments
  FROM `claim` c
  JOIN `treatment` t ON c.`claimID` = t.`claimID`
  JOIN `patient` p ON p.`patientID` = t.`patientID`
  JOIN `person` pe ON pe.`personID` = p.`patientID`
)
INSERT OVERWRITE table query3
SELECT Gender, COUNT(Claims) AS `Total Number of Claims`,
       COUNT(treatments) AS `Total Number of treatments`,
       COUNT(Claims) / COUNT(treatments) AS Ratio
FROM cte_table2
GROUP BY Gender;
```



Browse Directory

/user/outputs/query3								Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
-rwxr-xr-x	cloudera	supergroup	40 B	Wed Mar 15 04:09:21 -0700 2023	1	128 MB	000000_0	

Hadoop, 2016.

Create Output Table in Client DB

```
CREATE TABLE query3(  
  gender varchar(10),  
  count_claims int,  
  count_treatments int,  
  ratio double  
);
```

*EXPORT_COMMAND-:

```
sqoop export --connect jdbc:mysql://localhost:3306/healthcare_output  
--username root --password cloudera --table query3  
--export-dir /user/outputs/query3/000000_0  
--input-fields-terminated-by ';;'
```

Query OK, 0 rows affected (0.01 sec)

mysql> CREATE TABLE query3(
-> gender varchar(10),
-> count_claims int,
-> count_treatments int,
-> ratio double
->);

Query OK, 0 rows affected (0.01 sec)

mysql> select * from query3;

gender	count_claims	count_treatments	ratio
female	2676	2676	1
male	4287	4287	1

2 rows in set (0.00 sec)

mysql> █

cloudera@quickstart:~ cloudera@quickstart:~ Browsing HDFS - Mozil...

#####

Problem Statement 4: 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.

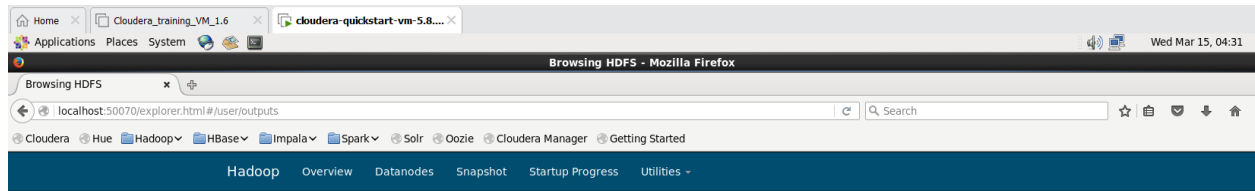
Note: discount field in keep signifies the percentage of discount on the maximum price.

Create External Table

```
create external table query4
(pharmacyName String, count_medicines int, total_price double, total_discounted_price double)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
LOCATION '/user/outputs/query4';
```

Insert Data Into External Table In Hive

```
with cte_table3 as (
  select `pharmacyName` as `Pharmacy Name`,
    count(m.`medicineID`) as `Total number of Medicines`,
    sum(m.`maxPrice`) as `Total Retail Price`,
    sum(m.`maxPrice` - (k.`discount` * 0.01)) as `Total Price of Medicines after discount`
  from pharmacy p
  join `keep` k on p.`pharmacyID` = k.`pharmacyID`
  join `medicine` m on m.`medicineID` = k.`medicineID`
  where p.`pharmacyID` = k.`pharmacyID`
  group by pharmacyName
)
INSERT OVERWRITE table query4
SELECT * FROM cte_table3;
```



Browse Directory

/user/outputs							Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	cloudera	supergroup	0 B	Tue Mar 14 11:26:39 -0700 2023	0	0 B	query1
drwxr-xr-x	cloudera	supergroup	0 B	Wed Mar 15 02:14:58 -0700 2023	0	0 B	query2
drwxr-xr-x	cloudera	supergroup	0 B	Wed Mar 15 04:09:24 -0700 2023	0	0 B	query3
drwxr-xr-x	cloudera	supergroup	0 B	Wed Mar 15 04:28:42 -0700 2023	0	0 B	query4

Hadoop, 2016.

Create Output Table in Client DB

```
CREATE TABLE query4(  
  pharmacyName Varchar(50),  
  count_medicines int,  
  total_price double,  
  total_discounted_price double  
);
```

*EXPORT_COMMAND-:

```
sqoop export \  
--connect jdbc:mysql://localhost:3306/healthcare_output \  
--username root \  
--password cloudera \  
--table query4 \  
--export-dir /user/outputs/query4/000000_0;
```

Home

Cloudera_training_VM_1.6

cloudera-quickstart-vm-5.8...

Applications

Places

System

cloudera@quickstart:~

File Edit View Search Terminal Help

Mediserv	136	135231.2	135210.8
Medisuite	449	181675.39	181606.59
Medlife	280	153321.98	153280.28
Midtown Express	193	42356.09	42325.59
MobiMeds	47	16448.69	16442.29
Modern Health	104	49718.78	49703.98
Neighborcare	405	379390.71	379329.31
Neighbors	438	210247.13	210182.33
New Era	141	70394.89	70373.19
Newday Drug Store	109	94521.37	94505.57
North East Pharmacy	216	62218.88	62184.68
Northport Pharmacy	463	169414.04	169349.94
Northwest Medication Management	28	5113.37	5109.97
Northwest Pharmacy	227	119308.49	119273.79
Omnicare	200	81862.69	81833.89
OptumRx	232	138356	138320.3
Outpatient Pharmacy	443	112925.02	112852.22
Pacific Medical	248	67936.98	67900.68
Pavilions Pharmacy	196	76314.62	76285.62
Pearl River Pharmacy	52	13221.87	13214.77
PersonalRX	237	68081.2	68045.1
Pharma Best	288	80614.59	80573.49
Pharma First	178	120551.8	120523.8
Pharma Street	421	245060.61	244994.41
PharmaMed	206	79438.01	79407.61
Pharmaca Integrative	233	90915.8	90880
Pharmacy Alliance	223	115386.49	115350.89
Pharmacy Express	58	12242.48	12233.48
Pharmacy Partners	369	159310.53	159257.73
Pharmanic	60	36700.36	36689.96
Pill Pack	383	99028.21	98973.01
Planet Health	125	53034.27	53017.47
Pocketpills	404	164646.4	164586.5
Postmeds	443	152959.4	152890.3
Powerhouse Pharmacy	81	39716.69	39706.19
PrecisionMed	91	50756.04	50741.64
Premier Long Term Care Pharm	404	155790.48	155727.08
Prescription Hope	372	121052.14	120998.94
Prescription Lifeline	359	143656.47	143599.57
Pride Pharmacy	475	269456.14	269384.54
Priority Pharmacy	230	72815.38	72782.5800000001
Protowell	121	46118.67	46101.47

213 rows in set (0.00 sec)

mysql>

cloudera@quickstart:~

cloudera@quickstart:~

Browsing HDFS - Mozil...

cloudera@quickstart:~

#####

Problem Statement 5: An Insurance company wants a state wise report of the treatments to claim ratio between 1st April 2021 and 31st March 2022 (days both included). Assist them to create such a report.

Create External Table

```
create external table query5 (state varchar(10), treat_count int, claim_count int, ratio double)
row format delimited
fields terminated by ','
lines terminated by '\n'
location '/user/outputs/query5';
SELECT address.state, COUNT(treatment.treatmentID) AS treat_count,
COUNT(claim.claimID) AS claim_count,
COUNT(treatment.treatmentID) / COUNT(claim.claimID) AS ratio
FROM address
INNER JOIN person ON address.addressID = person.addressID
INNER JOIN patient ON person.personID = patient.patientID
INNER JOIN treatment ON patient.patientID = treatment.patientID
LEFT JOIN claim ON treatment.claimID = claim.claimID
WHERE treatment.date BETWEEN '2021-04-01' AND '2022-03-31'
```

```
cloudera@quickstart:~$
File Edit View Search Terminal Help
(49400 bytes)
2023-03-15 10:30:35 Dump the side-table for tag: 1 with group count: 1673 into file: file:/tmp/cloudera/237fd4e4-6587-4fd0-9190-
Table-Stage-4/MapJoin-mapfile21-..hashtable
2023-03-15 10:30:35 Uploaded 1 File to: file:/tmp/cloudera/237fd4e4-6587-4fd0-9190-383a051b116a/hive_2023-03-15_10-30-14_651_70:
(53061 bytes)
2023-03-15 10:30:35 End of local task; Time Taken: 2.879 sec.
Execution completed successfully
MapReduce 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.reducers=<number>
Starting Job = job_1678811368319_0034, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678811368319_0034/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1678811368319_0034
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 1
2023-03-15 10:30:49,770 Stage-4 map = 0%, reduce = 0%
2023-03-15 10:31:10,441 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 9.14 sec
2023-03-15 10:31:20,071 Stage-4 map = 100%, reduce = 100%, Cumulative CPU 10.8 sec
MapReduce Total cumulative CPU time: 10 seconds 800 msec
Ended Job = job_1678811368319_0034
MapReduce Jobs Launched:
Stage-Stage-4: Map: 1 Reduce: 1 Cumulative CPU: 10.8 sec HDFS Read: 138448 HDFS Write: 466 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 800 msec
JK
K 98 67 1.462686567164179
L 213 130 1.6384615384615384
M 141 92 1.5326086956521738
N 135 82 1.646341463414634
O 267 182 1.467032967032967
P 182 114 1.5964912280701755
Q 196 135 1.451851851851852
R 167 110 1.518181818181818
S 192 114 1.6842105263157894
T 195 127 1.5354330708661417
U 128 87 1.471264367816092
V 142 96 1.4791666666666667
W 167 110 1.518181818181818
X 207 123 1.6829268292682926
Y 208 123 1.6910569105691058
Z 131 89 1.4719101123595506
Time taken: 66.616 seconds, Fetched: 16 row(s)
hive>
```

CREATE PARTITION TABLE

```
create table query5_p1 (addressid int , address1 string, city string, zip int) partitioned by (state string);
```

```
insert into query5_p1 partition(state) select addressid,address1,city,zip,state from address;
```

```
SELECT query5_p1.state, COUNT(treatment.treatmentID) AS treat_count,
COUNT(claim.claimID) AS claim_count,
COUNT(treatment.treatmentID) / COUNT(claim.claimID) AS ratio
FROM query5_p1
INNER JOIN person ON query5_p1.addressID = person.addressID
INNER JOIN patient ON person.personID = patient.patientID
INNER JOIN treatment ON patient.patientID = treatment.patientID
LEFT JOIN claim ON treatment.claimID = claim.claimID
WHERE treatment.date BETWEEN '2021-04-01' AND '2022-03-31'
GROUP BY query5_p1.state;
```

COMPARING NORMAL EXECUTION WITH PARTITION TABLE

```
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_1678811368319_0036, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678811368319_0036/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1678811368319_0036
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 1
2023-03-15 10:40:24,259 Stage-4 map = 0%, reduce = 0%
2023-03-15 10:40:33,858 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 3.29 sec
2023-03-15 10:40:42,392 Stage-4 map = 100%, reduce = 100%, Cumulative CPU 4.92 sec
MapReduce Total cumulative CPU time: 4 seconds 920 msec
Ended Job = job_1678811368319_0036
MapReduce Jobs Launched:
Stage-Stage-4: Map: 1 Reduce: 1 Cumulative CPU: 4.92 sec HDFS Read: 142225 HDFS Write: 466 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 920 msec
OK
AK      98      67      1.462686567164179
AL     213     130      1.6384615384615384
AR     141     92      1.5326086956521738
AZ     135     82      1.646341463414634
CA     267    182      1.467032967032967
CO     182    114      1.5964912280701755
CT     196    135      1.451851851851852
DC     167    110      1.518181818181818
FL     192    114      1.6842105263157894
GA     195    127      1.5354330708661417
KY     128     87      1.471264367816092
MA     142     96      1.4791666666666667
MD     167    110      1.518181818181818
OK     207    123      1.6829268292682926
TN     208    123      1.6910569105691058
VT     131     89      1.4719101123595506
Time taken: 36.825 seconds, Fetched: 16 row(s)
hive>
```

cloudera@quickstart:~ cloudera@quickstart:~ Browsing HDFS - Mozil... cloudera@quickstart:~

Use the mouse pointer inside a page's Ctrl-G

```
create table query5 (state varchar(10), treat_count int, claim_count int, ratio double)
```

```
sqoop export --connect jdbc:mysql://localhost:3306/healthcare_output --username root
--password cloudera --table query5 --export-dir /user/outputs/query5/000000_0
--input-fields-terminated-by ' ';
```

/user/hive/warehouse/query5_p1

Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxrwxrwx	cloudera	supergroup	0 B	Wed Mar 15 10:38:12 -0700 2023	0	0 B	state=AK
drwxrwxrwx	cloudera	supergroup	0 B	Wed Mar 15 10:38:12 -0700 2023	0	0 B	state=AL
drwxrwxrwx	cloudera	supergroup	0 B	Wed Mar 15 10:38:13 -0700 2023	0	0 B	state=AR
drwxrwxrwx	cloudera	supergroup	0 B	Wed Mar 15 10:38:12 -0700 2023	0	0 B	state=AZ
drwxrwxrwx	cloudera	supergroup	0 B	Wed Mar 15 10:38:13 -0700 2023	0	0 B	state=CA
drwxrwxrwx	cloudera	supergroup	0 B	Wed Mar 15 10:38:13 -0700 2023	0	0 B	state=CO
drwxrwxrwx	cloudera	supergroup	0 B	Wed Mar 15 10:38:13 -0700 2023	0	0 B	state=CT
drwxrwxrwx	cloudera	supergroup	0 B	Wed Mar 15 10:38:13 -0700 2023	0	0 B	state=DC
drwxrwxrwx	cloudera	supergroup	0 B	Wed Mar 15 10:38:13 -0700 2023	0	0 B	state=FL
drwxrwxrwx	cloudera	supergroup	0 B	Wed Mar 15 10:38:12 -0700 2023	0	0 B	state=GA
drwxrwxrwx	cloudera	supergroup	0 B	Wed Mar 15 10:38:13 -0700 2023	0	0 B	state=KY
drwxrwxrwx	cloudera	supergroup	0 B	Wed Mar 15 10:38:12 -0700 2023	0	0 B	state=MA
drwxrwxrwx	cloudera	supergroup	0 B	Wed Mar 15 10:38:13 -0700 2023	0	0 B	state=MD
drwxrwxrwx	cloudera	supergroup	0 B	Wed Mar 15 10:38:13 -0700 2023	0	0 B	state=OK
drwxrwxrwx	cloudera	supergroup	0 B	Wed Mar 15 10:38:13 -0700 2023	0	0 B	state=TN

some data to Mozilla so that we can improve your experience.

Browsing HDFS - Mozil...

cloudera@quickstart:~

cloudera@quickstart:~

Choose

```
mysql> select * from query5;
```

state	treat_count	claim_count	ratio
OK	207	123	1.68292682926829
TN	208	123	1.69105691056911
VT	131	89	1.47191011235955
AK	98	67	1.46268656716418
AL	213	130	1.63846153846154
AR	141	92	1.53260869565217
AZ	135	82	1.64634146341463
CA	267	182	1.46703296703297
CO	182	114	1.59649122807018
CT	196	135	1.45185185185185
DC	167	110	1.51818181818182
FL	192	114	1.68421052631579
GA	195	127	1.53543307086614
KY	128	87	1.47126436781609
MA	142	96	1.47916666666667
MD	167	110	1.51818181818182

```
16 rows in set (0.01 sec)

mysql>
```

#####

Problem Statement 6: 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

```
create external table query6
(city String, pharmacy_prescription_ratio double)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
LOCATION '/user/outputs/query6';
```

Insert Data Into External Table In Hive

```
INSERT INTO query6
SELECT a.city,
       ROUND(COUNT(DISTINCT pha.pharmacyID) / COUNT(DISTINCT pre.prescriptionID),4)
AS pharmacy_prescription_ratio
FROM pharmacy pha
JOIN address a ON a.addressID = pha.addressID
JOIN prescription pre ON pha.pharmacyID = pre.pharmacyID
GROUP BY a.city
HAVING COUNT(pre.prescriptionID) > 100
```

ORDER BY pharmacy_prescription_ratio;

Browse Directory

/user/outputs/query6 Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rwxr-xr-x	cloudera	supergroup	415 B	Thu Mar 16 12:27:57 -0700 2023	1	128 MB	000000_0

Hadoop, 2016.

Create Output Table in Client DB

```
CREATE TABLE query6(  
city Varchar(50),  
pharmacy_prescription_ratio double  
);
```

EXPORT COMMAND

```
sqoop export --connect jdbc:mysql://localhost:3306/healthcare_output --username root  
--password cloudera --table query6 --export-dir /user/outputs/query6/000000_0  
--input-fields-terminated-by ',';
```



```
mysql> select * from query6;
```

city	pharmacy_prescription_ratio
Union City	0.0162
Panama City	0.0165
Oklahoma City	0.0166
Savannah	0.0166
Castro Valley	0.0174
Worcester	0.0137
Panama City Beach	0.014
Glen Burnie	0.0143
Goodlettsville	0.0147
Anchorage	0.0152
Crownsville	0.0153
Pooler	0.0153
Nashville	0.0153
Montgomery	0.0154
Washington	0.0155
Manchester	0.0155
Fayetteville	0.0155
Glendale	0.0156
Farmington	0.0156
Annapolis	0.0157
Louisville	0.0158
Edmond	0.0158
Arvada	0.0161

```
23 rows in set (0.03 sec)

mysql>
```

#####

Problem Statement 7: 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

```
create external table query7
(pharmacyName String, max_count_medicines int, min_count_medicines int,
avg_count_medicines double)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
LOCATION '/user/outputs/query7';
```

Insert Data Into External Table In Hive

```
INSERT OVERWRITE Table query7
select pharmacyName, max(c.quantity) as `maximum number of medicines`,
min(c.quantity) as `minimum number of medicines`,
round(avg(c.quantity),0) as `average number of medicines`
from pharmacy p
join prescription pr on p.`pharmacyID` = pr.`pharmacyID`
```

```
join contain c on pr.`prescriptionID` = c.`prescriptionID`  
group by pharmacyName;
```

[Hadoop](#) [Overview](#) [Datanodes](#) [Snapshot](#) [Startup Progress](#) [Utilities](#) ▾

Browse Directory

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	cloudera	supergroup	0 B	Thu Mar 16 12:51:27 -0700 2023	0	0 B	.hive-staging_hive_2023-03-16_12-51-01_803_784951287687143882-1

Hadoop, 2016.

Create Output Table in Client DB

```
CREATE TABLE query7(  
  pharmacyName Varchar(50),  
  max_count_medicines int, min_count_medicines int, avg_count_medicines double  
);
```

EXPORT COMMAND

```
sqoop export --connect jdbc:mysql://localhost:3306/healthcare_output --username root  
--password cloudera --table query7 --export-dir /user/outputs/query6/0000000_0  
--input-fields-terminated-by ',';
```

pharmacyName	count	total_price	total_discounted_price
Mediserv	20	1	10
Medisuite	20	1	10
Medlife	20	1	10
Midtown Express	20	1	10
MobiMeds	20	1	10
Modern Health	20	1	11
Neighborcare	20	1	10
Neighbors	20	1	10
New Era	20	1	11
Newday Drug Store	20	1	11
North East Pharmacy	20	1	11
Northport Pharmacy	20	1	10
Northwest Medication Management	20	1	11
Northwest Pharmacy	20	1	11
Omnicare	20	1	10
OptumRx	20	1	11
Outpatient Pharmacy	20	1	10
Pacific Medical	20	1	11
Pavilions Pharmacy	20	1	10
Pearl River Pharmacy	20	1	11
PersonalRX	20	1	10
Pharma Best	20	1	11
Pharma First	20	1	11
Pharma Street	20	1	11
PharmaMed	20	1	11
Pharmaca Integrative	20	1	11
Pharmacy Alliance	20	1	10
Pharmacy Express	20	1	11
Pharmacy Partners	20	1	10
Pharmanic	20	1	11
Pill Pack	20	1	11
Planet Health	20	1	10
Pocketpills	20	1	11
Postmeds	20	1	11
Powerhouse Pharmacy	20	1	11
PrecisionMed	20	1	10
Premier Long Term Care Pharm	20	1	11
Prescription Hope	20	1	11
Prescription Lifeline	20	1	10
Pride Pharmacy	20	1	10
Priority Pharmacy	20	1	11
Protowell	20	1	11

213 rows in set (0.01 sec)

mysql>

#####

Problem Statement 8: 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

create external table query8

(pharmacyName String, count_medicines int, total_price double, total_discounted_price double)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

```
LOCATION '/user/outputs/query8';
```

Insert Data Into External Table In Hive

```
with cte_table3 as (  
  select `pharmacyName` as `Pharmacy Name`,  
         count(m.`medicineID`) as `Total number of Medicines`,  
         sum(m.`maxPrice`) as `Total Retail Price`,  
         sum(m.`maxPrice` - (k.`discount` * 0.01)) as `Total Price of Medicines after discount`  
  from pharmacy p  
 join `keep` k on p.`pharmacyID` = k.`pharmacyID`  
 join `medicine` m on m.`medicineID` = k.`medicineID`  
 where p.`pharmacyID` = k.`pharmacyID`  
 group by pharmacyName  
)  
INSERT OVERWRITE table query8  
SELECT * FROM cte_table3;
```

Hadoop

Overview

Datanodes

Snapshot

Startup Progress

Utilities

Browse Directory

/user/outputs/query8

Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rwxr-xr-x	cloudera	supergroup	10.75 KB	Thu Mar 16 13:03:35 -0700 2023	1	128 MB	000000_0

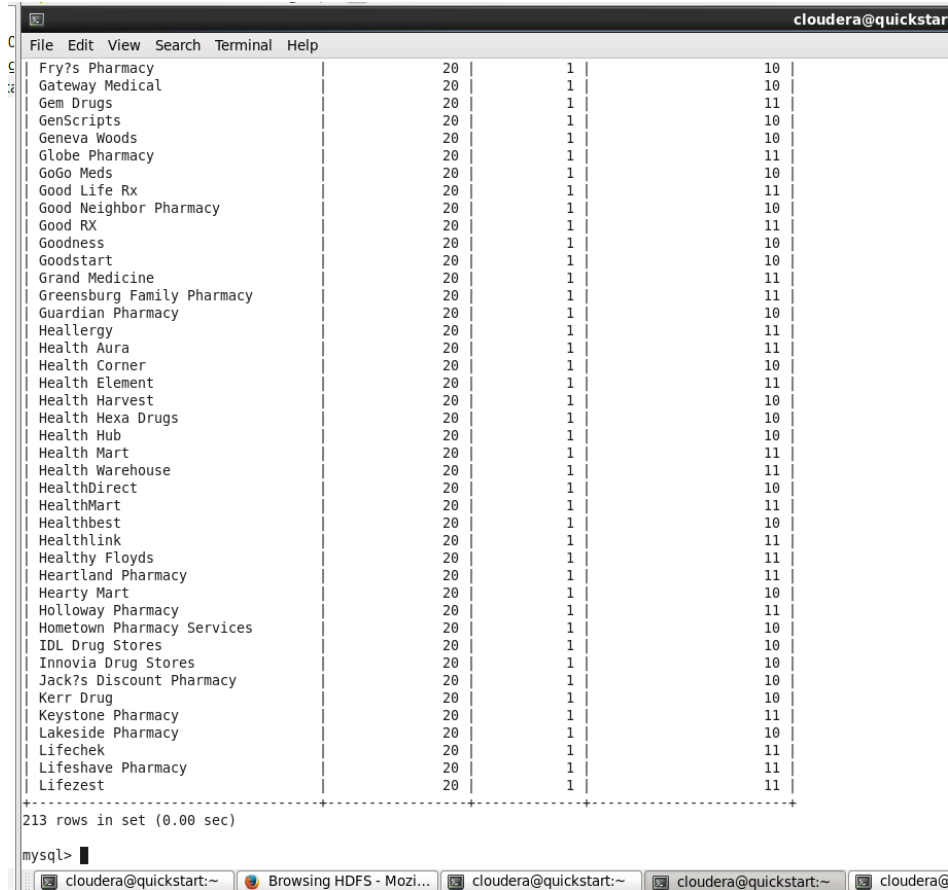
Hadoop, 2016.

Create Output Table in Client DB

```
CREATE TABLE query8(  
  pharmacyName Varchar(50),  
  count_medicines int,  
  total_price double,  
  total_discounted_price double  
);
```

EXPORT COMMAND

```
sqoop export --connect jdbc:mysql://localhost:3306/healthcare_output --username root
--password cloudera --table query8 --export-dir /user/outputs/query6/000000_0
--input-fields-terminated-by ',';
```



Fry?s Pharmacy	20	1	10
Gateway Medical	20	1	10
Gem Drugs	20	1	11
GenScripts	20	1	10
Geneva Woods	20	1	10
Globe Pharmacy	20	1	11
GoGo Meds	20	1	10
Good Life Rx	20	1	11
Good Neighbor Pharmacy	20	1	10
Good RX	20	1	11
Goodness	20	1	10
Goodstart	20	1	10
Grand Medicine	20	1	11
Greensburg Family Pharmacy	20	1	11
Guardian Pharmacy	20	1	10
Heallergy	20	1	11
Health Aura	20	1	11
Health Corner	20	1	10
Health Element	20	1	11
Health Harvest	20	1	10
Health Hexa Drugs	20	1	10
Health Hub	20	1	10
Health Mart	20	1	11
Health Warehouse	20	1	11
HealthDirect	20	1	10
HealthMart	20	1	11
Healthbest	20	1	10
Healthlink	20	1	11
Healthy Floyds	20	1	11
Heartland Pharmacy	20	1	11
Hearty Mart	20	1	10
Holloway Pharmacy	20	1	11
Hometown Pharmacy Services	20	1	10
IDL Drug Stores	20	1	10
Innovia Drug Stores	20	1	10
Jack?s Discount Pharmacy	20	1	10
Kerr Drug	20	1	10
Keystone Pharmacy	20	1	11
Lakeside Pharmacy	20	1	10
Lifechek	20	1	11
Lifeshave Pharmacy	20	1	11
Lifezest	20	1	11

213 rows in set (0.00 sec)

mysql>

#####