

Client tables loaded into HDFS

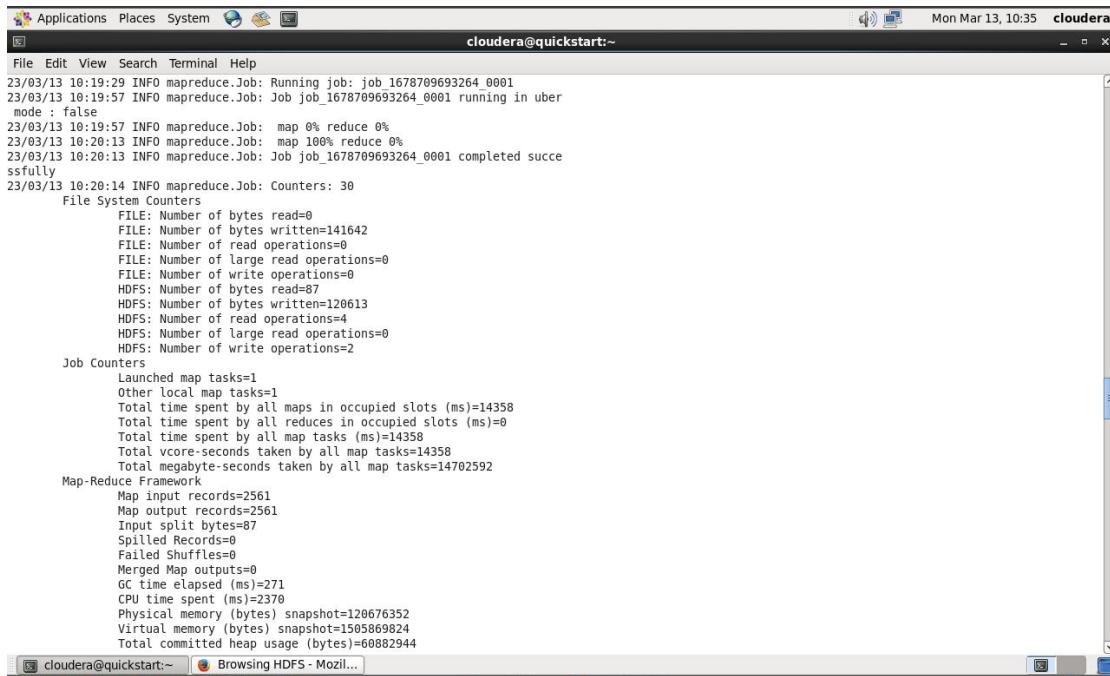
Browsing HDFS - Mozilla Firefox								
localhost:50070/explorer.html#/user/hive/warehouse								
Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started								
drwxrwxrwx	cloudera	supergroup	0 B	Mon Mar 13 10:20:25 -0700 2023	0	0 B	address	
drwxrwxrwx	cloudera	supergroup	0 B	Mon Mar 13 10:21:14 -0700 2023	0	0 B	claim	
drwxrwxrwx	cloudera	supergroup	0 B	Mon Mar 13 10:22:14 -0700 2023	0	0 B	contain	
drwxrwxrwx	cloudera	supergroup	0 B	Mon Mar 13 10:23:11 -0700 2023	0	0 B	disease	
drwxrwxrwx	cloudera	supergroup	0 B	Mon Mar 13 10:23:56 -0700 2023	0	0 B	insurancecompany	
drwxrwxrwx	cloudera	supergroup	0 B	Mon Mar 13 10:24:35 -0700 2023	0	0 B	insuranceplan	
drwxrwxrwx	cloudera	supergroup	0 B	Mon Mar 13 10:25:20 -0700 2023	0	0 B	keep	
drwxrwxrwx	cloudera	supergroup	0 B	Mon Mar 13 10:26:04 -0700 2023	0	0 B	medicine	
drwxrwxrwx	cloudera	supergroup	0 B	Mon Mar 13 10:26:44 -0700 2023	0	0 B	patient	
drwxrwxrwx	cloudera	supergroup	0 B	Mon Mar 13 10:27:29 -0700 2023	0	0 B	patient_details	
drwxrwxrwx	cloudera	supergroup	0 B	Mon Mar 13 10:28:08 -0700 2023	0	0 B	person	
drwxrwxrwx	cloudera	supergroup	0 B	Mon Mar 13 10:28:46 -0700 2023	0	0 B	pharmacy	
drwxrwxrwx	cloudera	supergroup	0 B	Mon Mar 13 10:29:28 -0700 2023	0	0 B	prescription	
drwxrwxrwx	cloudera	supergroup	0 B	Mon Mar 13 10:30:15 -0700 2023	0	0 B	treatment	

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

cloudera@quickstart:~ | Browsing HDFS - Mozilla Firefox

Import command

```
Applications Places System Terminal Help cloudera@quickstart:~ Mon Mar 13, 10:33 cloudera
File Edit View Search Terminal Help
Manager.java:52)
    at org.apache.sqoop.manager.CatalogQueryManager.listTables(CatalogQueryM
anager.java:102)
    ... 7 more
[cloudera@quickstart ~]$ sqoop import-all-tables --connect jdbc:mysql://localhos
t:3306/healthcare --username root --password --hive-import --m 1
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
23/03/13 10:18:47 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.8.0
Missing argument for option: password
[cloudera@quickstart ~]$ sqoop import-all-tables --connect jdbc:mysql://localhos
t:3306/healthcare --username root --password cloudera --hive-import --m 1
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
23/03/13 10:19:05 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.8.0
23/03/13 10:19:05 WARN tool.BaseSqoopTool: Setting your password on the command-
line is insecure. Consider using -P instead.
23/03/13 10:19:05 INFO tool.BaseSqoopTool: Using Hive-specific delimiters for ou-
put. You can override
23/03/13 10:19:05 INFO tool.BaseSqoopTool: delimiters with --fields-terminated-b
y, etc.
23/03/13 10:19:06 INFO manager.MySQLManager: Preparing to use a MySQL streaming
resultset.
23/03/13 10:19:07 INFO tool.CodeGenTool: Beginning code generation
23/03/13 10:19:07 INFO manager.SqlManager: Executing SQL statement: SELECT t.* F
ROM `address` AS t LIMIT 1
23/03/13 10:19:07 INFO manager.SqlManager: Executing SQL statement: SELECT t.* F
ROM `address` AS t LIMIT 1
23/03/13 10:19:07 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/ha
doop-mapreduce
Note: /tmp/sqoop-cloudera/compile/636c91746041007289edac7d1c28e60a/address.java
uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
23/03/13 10:19:14 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-clou
dera/compile/636c91746041007289edac7d1c28e60a/address.jar
23/03/13 10:19:14 WARN manager.MySQLManager: It looks like you are importing fro
m mysql.
23/03/13 10:19:14 WARN manager.MySQLManager: This transfer can be faster! Use th
e --direct
```

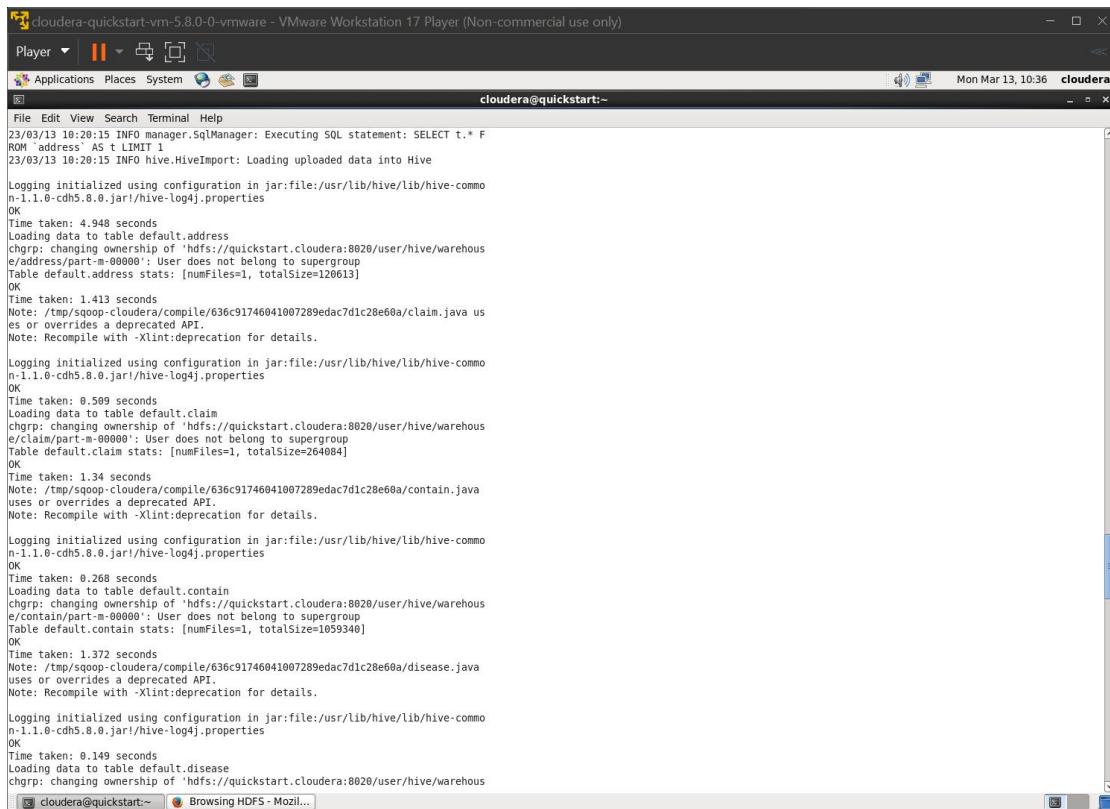


```

Applications Places System cloudera@quickstart:~ Mon Mar 13, 10:35 cloudera
File Edit View Search Terminal Help
23/03/13 10:19:29 INFO mapreduce.Job: Running job: job_1678709693264_0001
23/03/13 10:19:57 INFO mapreduce.Job: Job job_1678709693264_0001 running in uber
mode : false
23/03/13 10:19:57 INFO mapreduce.Job: map 0% reduce 0%
23/03/13 10:20:13 INFO mapreduce.Job: map 100% reduce 0%
23/03/13 10:20:13 INFO mapreduce.Job: Job job_1678709693264_0001 completed succe
ssfully
23/03/13 10:20:14 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=141642
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=87
    HDFS: Number of bytes written=120613
    HDFS: Number of read operations=4
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=1
    Other local map tasks=1
    Total time spent by all maps in occupied slots (ms)=14358
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=14358
    Total vcore-seconds taken by all map tasks=14358
    Total megabyte-seconds taken by all map tasks=14702592
  Map-Reduce Framework
    Map input records=2561
    Map output records=2561
    Input split bytes=87
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=271
    CPU time spent (ms)=2370
    Physical memory (bytes) snapshot=120676352
    Virtual memory (bytes) snapshot=1505869824
    Total committed heap usage (bytes)=60882944
cloudera@quickstart:~ | Browsing HDFS - Mozilla...

```

IMPORTING TABLES



```

cloudera-quickstart-vm-5.8.0-0-vmware - VMware Workstation 17 Player (Non-commercial use only)
Player | || | ↻
Applications Places System cloudera@quickstart:~ Mon Mar 13, 10:36 cloudera
File Edit View Search Terminal Help
23/03/13 10:20:15 INFO manager.SqlManager: Executing SQL statement: SELECT t.* F
ROM `address` AS t LIMIT 1
23/03/13 10:20:15 INFO hive.HiveImport: Loading uploaded data into Hive
Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-commo
n-1.1.0-cdh5.8.0.jar!/hive-log4j.properties
OK
Time taken: 4.948 seconds
Loading data to table default.address
chgrp: changing ownership of `hdfs://quickstart.cloudera:8020/user/hive/warehouse
e/address/part-m-00000`: User does not belong to supergroup
Table default.address stats: [numFiles=1, totalSize=120613]
OK
Time taken: 1.413 seconds
Note: /tmp/sqoop-cloudera/compile/636c91746041007289edac7d1c28e60a/claim.java us
es or overrides a deprecated API
Note: Recompile with -Xlint:deprecation for details.
Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-commo
n-1.1.0-cdh5.8.0.jar!/hive-log4j.properties
OK
Time taken: 0.599 seconds
Loading data to table default.claim
chgrp: changing ownership of `hdfs://quickstart.cloudera:8020/user/hive/warehouse
e/claim/part-m-00000`: User does not belong to supergroup
Table default.claim stats: [numFiles=1, totalSize=264084]
OK
Time taken: 1.34 seconds
Note: /tmp/sqoop-cloudera/compile/636c91746041007289edac7d1c28e60a/contain.java
uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-commo
n-1.1.0-cdh5.8.0.jar!/hive-log4j.properties
OK
Time taken: 0.268 seconds
Loading data to table default.contain
chgrp: changing ownership of `hdfs://quickstart.cloudera:8020/user/hive/warehouse
e/contain/part-m-00000`: User does not belong to supergroup
Table default.contain stats: [numFiles=1, totalSize=1059340]
OK
Time taken: 1.372 seconds
Note: /tmp/sqoop-cloudera/compile/636c91746041007289edac7d1c28e60a/disease.java
uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-commo
n-1.1.0-cdh5.8.0.jar!/hive-log4j.properties
OK
Time taken: 0.149 seconds
Loading data to table default.disease
chgrp: changing ownership of `hdfs://quickstart.cloudera:8020/user/hive/warehouse

```

```
cloudera-quickstart-vm-5.8.0-0-vmware - VMware Workstation 17 Player (Non-commercial use only)
Player || □ ☰
Applications Places System 🌐 📁 Mon Mar 13, 10:37 cloudera
File Edit View Search Terminal Help
.cloudera@quickstart:~ .logging initialized using configuration in jar:/file:/usr/lib/hive/lib/hive-commo
1-1.1.0-cdh5.8.0.jar!/hive-log4j.properties
OK
Time taken: 0.164 seconds
>Loading data to table default.medicine
.chgrp: changing ownership of '/dfs:/quickstart.cloudera:8020/user/hive/warehouse
/medicine/part-m-00000': User does not belong to supergroup
Table default.medicine stats: [numFiles=1, totalSize=647977]
OK
Time taken: 0.593 seconds
Note: /tmp/sqoop-cloudera/compile/636c91746041007289edac7d1c28e60a/patient.java
uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.

.logging initialized using configuration in jar:/file:/usr/lib/hive/lib/hive-commo
1-1.1.0-cdh5.8.0.jar!/hive-log4j.properties
OK
Time taken: 0.121 seconds
>.loading data to table default.patient
.chgrp: changing ownership of '/dfs:/quickstart.cloudera:8020/user/hive/warehouse
/patient/part-m-00000': User does not belong to supergroup
Table default.patient stats: [numFiles=1, totalSize=1528]
OK
Time taken: 0.62 seconds
Note: /tmp/sqoop-cloudera/compile/636c91746041007289edac7d1c28e60a/patient_detail
.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.

.logging initialized using configuration in jar:/file:/usr/lib/hive/lib/hive-commo
1-1.1.0-cdh5.8.0.jar!/hive-log4j.properties
OK
Time taken: 0.159 seconds
>.loading data to table default.patient_details
.chgrp: changing ownership of '/dfs:/quickstart.cloudera:8020/user/hive/warehouse
/patient_details/part-m-00000': User does not belong to supergroup
Table default.patient_details stats: [numFiles=1, totalSize=108106]
OK
Time taken: 0.682 seconds
Note: /tmp/sqoop-cloudera/compile/636c91746041007289edac7d1c28e60a/person.java
uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.

.logging initialized using configuration in jar:/file:/usr/lib/hive/lib/hive-commo
1-1.1.0-cdh5.8.0.jar!/hive-log4j.properties
OK
Time taken: 0.164 seconds
>.loading data to table default.person
.chgrp: changing ownership of '/dfs:/quickstart.cloudera:8020/user/hive/warehouse
/person/part-m-00000': User does not belong to supergroup
Table default.person stats: [numFiles=1, totalSize=120069]
OK
Time taken: 0.544 seconds
| cloudera@quickstart:~ | Browsing HDFS - Mozilla...
Loading data to table default.person
.chgrp: changing ownership of '/dfs:/quickstart.cloudera:8020/user/hive/warehouse
/person/part-m-00000': User does not belong to supergroup
Table default.person stats: [numFiles=1, totalSize=120069]
OK
Time taken: 0.544 seconds
Note: /tmp/sqoop-cloudera/compile/636c91746041007289edac7d1c28e60a/pharmacy.java
uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.

Logging initialized using configuration in jar:/file:/usr/lib/hive/lib/hive-commo
n-1.1.0-cdh5.8.0.jar!/hive-log4j.properties
OK
Time taken: 0.106 seconds
>Loading data to table default.pharmacy
.chgrp: changing ownership of '/dfs:/quickstart.cloudera:8020/user/hive/warehouse
/pharmacy/part-m-00000': User does not belong to supergroup
Table default.pharmacy stats: [numFiles=1, totalSize=8095]
OK
Time taken: 0.592 seconds
Note: /tmp/sqoop-cloudera/compile/636c91746041007289edac7d1c28e60a/prescription.
java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.

Logging initialized using configuration in jar:/file:/usr/lib/hive/lib/hive-commo
n-1.1.0-cdh5.8.0.jar!/hive-log4j.properties
OK
Time taken: 0.159 seconds
>.loading data to table default.prescription
.chgrp: changing ownership of '/dfs:/quickstart.cloudera:8020/user/hive/warehouse
/prescription/part-m-00000': User does not belong to supergroup
Table default.prescription stats: [numFiles=1, totalSize=32227]
OK
Time taken: 0.955 seconds
Note: /tmp/sqoop-cloudera/compile/636c91746041007289edac7d1c28e60a/treatment.jav
a uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.

Logging initialized using configuration in jar:/file:/usr/lib/hive/lib/hive-commo
n-1.1.0-cdh5.8.0.jar!/hive-log4j.properties
OK
Time taken: 0.899 seconds
>Loading data to table default.treatment
.chgrp: changing ownership of '/dfs:/quickstart.cloudera:8020/user/hive/warehouse
/treatment/part-m-00000': User does not belong to supergroup
Table default.treatment stats: [numFiles=1, totalSize=409446]
OK
Time taken: 0.688 seconds
| cloudera@quickstart:~ | Browsing HDFS - Mozilla...
| cloudera@quickstart:~ | Browsing HDFS - Mozilla...
```

QUERY-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 view v_1 as select count(*),v1.category from
```

```

(select (case
when round(datediff(current_date(),dob)/365) <14 then 'child'
when round(datediff(current_date(),dob)/365) <24 then 'youth'
when round(datediff(current_date(),dob)/365) <64 then 'adult'
else 'senior citizen'
end) as Category
from patient p inner join treatment t1 on p.patientid = t1.patientid where year(t1.date)=2022) as v1
GROUP BY v1.Category;

```

```

File Edit View Search Terminal Help
        at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 7:5 cannot recognize input near "category" 'from' 'patient' in expression specification
hive> select count(*),v1.category from
  > (select (case
  > when round(datediff(current_date(),dob)/365) <14 then 'child'
  > when round(datediff(current_date(),dob)/365) <24 then 'youth'
  > when round(datediff(current_date(),dob)/365) <64 then 'adult'
  > else 'senior citizen'
  > end) as Category
  > from patient p inner join treatment t1 on p.patientid = t1.patientid where year(t1.date)=2022) as v1 GROUP BY v1.Category;
Query ID = cloudera_20230313110000_f9e9c7b1-c24f-47ff-952c-7d7f550c5611
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20230313110000_f9e9c7b1-c24f-47ff-952c-7d7f550c5611.log
2023-03-13 11:00:48 Starting to launch local task to process map join; maximum memory = 1013645312
2023-03-13 11:00:52 Dump the side-table for tag: 0 with group count: 1126 into file: file:/tmp/cloudera/f43d3868-c9fc-4403-8c4b-503424505e36/hive_2023-03-13_11-00-35_039_100256624
0271672201-1-local-10004/HashTable-Stage-2/MapJoin-mapfile10--hashtable
2023-03-13 11:00:52 Uploaded 1 File to: file:/tmp/cloudera/f43d3868-c9fc-4403-8c4b-503424505e36/hive_2023-03-13_11-00-35_039_1002566240271672201-1-local-10004/HashTable-Stage-2/M
apJoin-mapfile10--hashtable (37601 bytes)
2023-03-13 11:00:52 End of local task, Time Taken: 3.654 sec.
Execution of mapjoin succeeded
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_1678709693264_0017, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678709693264_0017/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1678709693264_0017
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-03-13 11:01:10,349 Stage-2 map = 0%, reduce = 0%
2023-03-13 11:01:30,415 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 6.79 sec
2023-03-13 11:01:48,165 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 10.31 sec
MapReduce Total cumulative CPU time: 10 seconds 310 msec
Ended Job : job_1678709693264_0017
MapReduce Jobs Launched:
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 10.31 sec HDFS Read: 423868 HDFS Write: 50 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 310 msec
0K
1299    adult
749    child
807    senior citizen
112    youth
Time taken: 75.419 seconds, Fetched: 4 row(s)
hive> 

```

QUERY 1 EXPORT

```

File Edit View Search Terminal Help
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=566096
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=781
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=0
Job Counters
  Launched map tasks=4
  Data-local map tasks=4
  Total time spent by all maps in occupied slots (ms)=213223
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=213223
  Total vcore-seconds taken by all map tasks=213223
  Total megabyte-seconds taken by all map tasks=218340352
Map-Reduce Framework
  Map input records=4
  Map output records=4
  Input split bytes=631
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=4001
  CPU time spent (ms)=7010
  Physical memory (bytes) snapshot=420995072
  Virtual memory (bytes) snapshot=6015369216
  Total committed heap usage (bytes)=243531776
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=0
23/03/14 06:32:06 INFO mapreduce.ExportJobBase: Transferred 781 bytes in 82.2841 seconds (9.4915 bytes/sec)
23/03/14 06:32:06 INFO mapreduce.ExportJobBase: Exported 4 records.
[cloudera@quickstart ~]$ 

```

Client DB Table

```
Welcome to the MySQL monitor.  Commands may be typed by you.
Your MySQL connection id is 130
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use client;
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> select * from solution1;
+-----+-----+
| count | category |
+-----+-----+
| 112  | youth   |
| 1299 | adult   |
| 749  | child   |
| 807  | senior citizen |
+-----+-----+
4 rows in set (0.00 sec)

mysql>
```

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

QUERY 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 view v_2 as 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 ;
```

EXTERNAL TABLE:

```
create external table opt2(diseasename varchar(50), count_female varchar(10), count_male varchar(10), ratio double)
row format delimited fields terminated by ',' lines terminated by '\n' location '/user/hive/output2';
```

OVERWRITE:

```
Hive> insert overwrite table opt8 select * from v_2 ;
```

EXPORT:

```
[cloudera@quickstart ~]$ sqoop export --connect jdbc:mysql://localhost:3306/client --username root --P --table
solution8 --export-dir /user/hive/output2/000000_0 --input-fields-terminated-by ','
```

OUTPUT:

```
File Edit View Search Terminal Help
cloudera@quickstart:~$ File Edit View Search Terminal Help
cloudera@quickstart:~$ MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Reduce: 1 Cumulative CPU: 9.57 sec HDFS Read: 425217 HDFS Write: 1942 SUCCESS
Stage-Stage-4: Map: 1 Reduce: 1 Cumulative CPU: 4.96 sec HDFS Read: 6928 HDFS Write: 1724 SUCCESS
Total Mapreduce CPU time Spent: 14 seconds 534 msec
OK
Alzheimer's disease          173      95      1.8210526215789475
Amyotrophic lateral sclerosis 165      106      1.5566037735849056
Anorexia nervosa             177      96      1.84375
Anxiety disorder              153      126      1.2142857142857142
Asthma                      144      101      1.4257425742574257
Atherosclerosis               174      112      1.5535714285714286
Attention deficit hyperactivity disorder 158      125      1.264
Autism                      156      94      1.6595744688051063
Autoimmune diseases          165      102      1.61764740588235294
Bipolar disorder              166      114      1.456145350877193
Cancer                      191      103      1.854368932038835
Chronic fatigue syndrome     158      107      1.4766355140186015
Chronic obstructive pulmonary disease 152      97      1.5670103092783585
Coronary heart disease        149      97      1.5360824742268042
Crohn's disease              182      102      1.7843137254981962
Dementia                    162      99      1.8
Depression                  170      82      2.073170731707317
Diabetes mellitus type 1     174      93      1.8709677419354838
Diabetes mellitus type 2     178      99      1.7979797979797978
Dilated cardiomyopathy       191      110      1.7363636363636363
Epilepsy                     153      96      1.59375
Fibromyalgia                178      109      1.3629032258864515
Hirschsprung's syndrome     169      124      1.4324324324324325
Irritable bowel syndrome     184      184      1.7692307692307692
Low back pain                159      111      1.4324324324324325
Lupus                       158      88      1.7954545454545454
Metabolic syndrome           161      127      1.26771653543367088
Multiple sclerosis            173      88      1.965909090909090998
Myocardial infarction        190      107      1.7757089345794392
Obesity                     157      123      1.2764227642276422
Obsessive?compulsive disorder 175      118      1.5909090909090908
Panic disorder               158      116      1.4363636363636363
Parkinson's disease           145      94      1.5425531914893618
Psoriasis                   157      93      1.6881728430187527
Rheumatoid arthritis         156      115      1.3885369734513274
Sciatica                     96      117      1.7708333333333333
Schizophrenia                180      117      1.6239310931624
Stroke                      183      120      1.6339285714285714
Thromboangiitis obliterans   175      96      1.82291666666666667
Tourette syndrome              153      125      1.224
Vasculitis                   175      121      1.4462889917355373
Time taken: 129.812 seconds, Fetched: 40 row(s)
hive> 
```

QUERY2 export:

```
sqoop export --connect jdbc:mysql://localhost:3306/client --username root --P --table solution2 --export-dir /user/hive/output2/000000_0 --input-fields-terminated-by ''
```

```

Applications Places System cloudera@quickstart:~ Tue Mar 14, 07:21 cloudera
File Edit View Search Terminal Help Change account settings and status
FILE: Number of bytes read=0
FILE: Number of bytes written=566188
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=5170
HDFS: Number of bytes written=0
HDFS: Number of read operations=19
HDFS: Number of large read operations=0
HDFS: Number of write operations=0
Job Counters
Launched map tasks=4
Data-local map tasks=4
Total time spent by all maps in occupied slots (ms)=215259
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=215259
Total vcore-seconds taken by all map tasks=215259
Total megabyte-seconds taken by all map tasks=220425216
Map-Reduce Framework
Map input records=40
Map output records=40
Input split bytes=631
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=3936
CPU time spent (ms)=6900
Physical memory (bytes) snapshot=451825664
Virtual memory (bytes) snapshot=6015369216
Total committed heap usage (bytes)=243531776
File Input Format Counters
Bytes Read=0
File Output Format Counters
Bytes Written=0
23/03/14 07:19:40 INFO mapreduce.ExportJobBase: Transferred 5.0488 KB in 85.4636 seconds (60.4936 bytes/sec)
23/03/14 07:19:40 INFO mapreduce.ExportJobBase: Exported 40 records.
[cloudera@quickstart ~]$ 

```

Query2 client db:

```

Applications Places System cloudera@quickstart:~ Tue Mar 14, 07:22 cloudera
File Edit View Search Terminal Help
mysql> create table solution2(diseasename varchar(100), count_m varchar(10), count_f varchar(10), ratio double);
Query OK, 0 rows affected (0.01 sec)

mysql> select * from solution2;
+-----+-----+-----+-----+
| diseasename | count_m | count_f | ratio |
+-----+-----+-----+-----+
| Alzheimer's disease | 173 | 95 | 1.82105263157899 |
| Amyotrophic lateral sclerosis | 165 | 106 | 1.55666377358491 |
| 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 | 1.65957446880511 |
| Autoimmune diseases | 165 | 102 | 1.61764705882353 |
| Bipolar disorder | 166 | 114 | 1.45614035087719 |
| Panic disorder | 158 | 110 | 1.43636363636364 |
| Parkinson's disease | 145 | 94 | 1.54255319148938 |
| 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 | 1.8229166666666667 |
| Tourette syndrome | 153 | 125 | 1.224 |
| Vasculitis | 175 | 121 | 1.44628099173554 |
| Cancer | 191 | 103 | 1.85436893203884 |
| Chronic fatigue syndrome | 158 | 107 | 1.47663551401869 |
| Chronic obstructive pulmonary disease | 152 | 97 | 1.56701036927833 |
| Coronary heart disease | 149 | 97 | 1.5368824742268 |
| Crohn's disease | 182 | 102 | 1.7843137254944 |
| Dementia | 162 | 90 | 1.8 |
| Depression | 170 | 82 | 2.07317073170732 |
| Diabetes mellitus type 1 | 174 | 93 | 1.87096774193548 |
| Diabetes mellitus type 2 | 178 | 99 | 1.7979797979798 |
| Dilated cardiomyopathy | 191 | 110 | 1.73636363636363 |
| Epilepsy | 153 | 96 | 1.39575 |
| Guillain?Barré syndrome | 169 | 124 | 1.36200322586645 |
| Irritable bowel syndrome | 184 | 104 | 1.76923076923077 |
| Low back pain | 159 | 111 | 1.43243243243243 |
| Lupus | 158 | 88 | 1.79545454545455 |
| Metabolic syndrome | 161 | 127 | 1.26711653543307 |
| Multiple sclerosis | 173 | 88 | 1.9659090909090909 |
| Myocardial infarction | 190 | 107 | 1.77578093457944 |
| Obesity | 157 | 123 | 1.27642276422764 |
| Obsessive?compulsive disorder | 175 | 110 | 1.5909090909090909 |
+-----+
40 rows in set (0.00 sec)

mysql> 

```

QUERY 3:

It is suspected by healthcare research department that the substance "ranitidine" might be causing some side effects. Find the top 3 companies using the substance in their medicine so that they can be informed about it.

```
create view v_3 as select * from  
(select dense_rank() over(partition by companyname order by quantity desc) as  
denseno,quantity,m.companyname as cname  
from keep join medicine m on keep.medicineid=m.medicineid where substancename='ranitidina' )k  
where k.denseno=1 limit 3;
```

```
Applications Places System cloudera@quickstart:~  
File Edit View Search Terminal Help  
File org.apache.hadoop.util.RunJar.main\(\[RunJar.java:136\]\)  
FAILED: ParseException line 216:27 cannot recognize input near ''densoeno'' ',' 'quantity' in expression specification  
hive> select * from \(select dense\_rank\(\) over\(partition by companyname order by quantity desc\) as denso, quantity, m.companyname as cname  
> from keep join medicine m on keep.medicined=m.medicined where substituencename='ranitidina' \)  
where k.denso <= 10 limit 3;  
FAILED: SemanticException Hivesql specified without GROUP BY  
hive> select \* from \\(select dense\\_rank\\(\\) over\\(partition by companyname order by quantity desc\\) as denso, quantity, m.companyname as cname  
> from keep join medicine m on keep.medicined=m.medicined where substituencename='ranitidina' \\)  
where k.denso <= 10 limit 3;  
Query ID: 2023-03-14\\\_042424\\\_7049dd0\\\_816d-4849-aas1-b54643fd6ab5f  
Total jobs = 0  
Execution Log: /tmp/cloudera/\\\_cloudera\\\_2023-03-14\\\_042424\\\_7049dd0\\\_816d-4849-aas1-b54643fd6ab5f.log  
Driver log: /tmp/cloudera/\\\_cloudera\\\_2023-03-14\\\_042424\\\_7049dd0\\\_816d-4849-aas1-b54643fd6ab5f.log  
Maximum memory = 1013645312  
2023-03-14 08:24:21.19 Starting the map-side table for tag: 8 with group count: 33261 into file: /tmp/cloudera/\\_efc948fb-f6cd-4fc-a964-7ef6f5bf5ab9/hive\\_2023-03-14\\_02-42-14\\_983\\_7702275694545613532-1-local-10004/HesTable-Stage-2/M  
apJoin-mapfile0-->hashable \\(97731 bytes\\)  
2023-03-14 08:24:21.19 End of local task. Time Taken: 1.308 sec.  
MapredLocal task succeeded  
Launching Job 1 out of 1  
Job successfully completed  
Launching Job 1 out of 1  
Job successfully completed  
In order to change the average load for a reducer \\(in bytes\\):  
set hive.exec.reducers.bytes.per.reducer=<number>  
To set the number of reducers:  
set hive.exec.reducers.max=<number>  
To set a constant number of reducers:  
set mapred.reduce.tasks=<number>  
Starting Job = job\\_1678799693264\\_0020, Tracking URL = http://quickstart.cloudera:8888/proxy/application\\\_1678799693264\\\_0020  
KILL Command : /usr/lib/hadoop/bin/hadoop job -kill job\\\_1678799693264\\\_0020  
2023-03-14 08:24:21.19 Stage-2 map = 0B, reduce = 0B  
2023-03-14 08:24:21.19 Stage-2 map = 0B, reduce = 0B, Cumulative CPU 1.46 sec  
2023-03-14 08:24:21.19 Stage-2 map = 100B, reduce = 100B, Cumulative CPU 2.55 sec  
MapReduce Total cumulative CPU time: 2 seconds 558 msec  
MapReduce Job = job\\_1678799693264\\_0020  
MapReduce Job Launching  
Stage-2: Map 1: Reduce 1: Cumulative CPU: 2.55 sec HDFS Read: 6494215 HDFS Write: 113 SUCCESS  
Total MapReduce Time Spent: 2 seconds 558 msec  
hive>  
1 6011 BIDAFARMA FARMACEUTICA LTDA  
1 9803 BIDAFARMA FARMACEUTICA LTDA  
1 9536 HIPOLIBRAN FARMACEUTICA LTDA  
Time taken: 251.83 seconds, Fetched: 3 rows\\(s\\)  
hive>
```

Query 3 Sqoop export

```
sqoop export --connect jdbc:mysql://localhost:3306/client --username root --P --table solution3 --export-dir /user/hive/output3/000000_0 --input-fields-terminated-by ''
```

```
File Edit View Search Terminal Help
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=566148
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=945
    HDFS: Number of bytes written=0
    HDFS: Number of read operations=19
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=0
  Job Counters
    Launched map tasks=4
    Data-local map tasks=4
    Total time spent by all maps in occupied slots (ms)=281467
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=281467
    Total vcore-seconds taken by all map tasks=281467
    Total megabyte-seconds taken by all map tasks=288222208
  Map-Reduce Framework
    Map input records=3
    Map output records=3
    Input split bytes=631
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=5243
    CPU time spent (ms)=6610
    Physical memory (bytes) snapshot=461275136
    Virtual memory (bytes) snapshot=6015205376
    Total committed heap usage (bytes)=243531776
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=0
23/03/14 07:41:00 INFO mapreduce.ExportJobBase: Transferred 945 bytes in 125.0225 seconds (7.5586 bytes/sec)
23/03/14 07:41:00 INFO mapreduce.ExportJobBase: Exported 3 records.
[cloudera@quickstart ~]$
```

Clientdb solution3 :

```
mysql> create table solution3(denseno int,quantity int,pharmacy varchar(100));
Query OK, 0 rows affected (0.01 sec)

mysql> select * from solution3;
+-----+-----+-----+
| denseno | quantity | pharmacy |
+-----+-----+-----+
|      1 |     9530 | HIPOLABOR FARMACEUTICA LTDA |
|      1 |     6011 | BIOFARMA FARMACEUTICA LTDA |
|      1 |     9686 | CIFARMA CIENTIFICA FARMACEUTICA LTDA |
+-----+-----+-----+
3 rows in set (0.00 sec)

mysql>
```

QUERY 4:

Jackson has requested a detailed pharmacy report that shows each pharmacy name, and how many prescriptions they have prescribed for each disease in the year 2022, along with this Jackson also needs to view how many prescriptions were prescribed by each pharmacy, and the total number prescriptions were prescribed for each disease.

Assist Jackson to create this report.

```
create view v_123 as
select pharmacynname,diseasename,count(*) as total_prescriptions
from pharmacy join prescription on pharmacy.pharmacyid=prescription.pharmacyid
join treatment on treatment.treatmentid=prescription.treatmentid
join disease on disease.diseaseid=treatment.diseaseid
group by disease.diseasename,pharmacy.pharmacynname
with rollup;

create view v_234 as
select pharmacynname,diseasename,count(prescriptionid) as no_of_prescriptions_2022
from pharmacy join prescription on pharmacy.pharmacyid=prescription.pharmacyid
join treatment on treatment.treatmentid=prescription.treatmentid
join disease on disease.diseaseid=treatment.diseaseid
where year(treatment.date)=2022
group by disease.diseasename,pharmacy.pharmacynname
with rollup;

create view v_4 as select
v1.pharmacynname,v1.diseasename,v2.no_of_prescriptions_2022,v1.total_prescriptions from
v_123 v1 left join v_234 v2 on v1.pharmacynname=v2.pharmacynname and
v1.diseasename=v2.diseasename;
```

```

Applications Places System cloudera@quickstart:~ Tue Mar 14, 03:25 cloudera
File Edit View Search Terminal Help
RX Express Vasculitis NULL 2
RefillWise Vasculitis NULL 2
Rejuvia Drugs Vasculitis 1 2
Reliable Pharmacy Vasculitis NULL 2
Reliable Rexall Vasculitis NULL 2
Revco Discount Drugs Vasculitis 2 4
Right Drugs Vasculitis 1 2
Rite Aid Vasculitis NULL 3
Rocky's Drug Vasculitis NULL 2
Roosevelt Clinic Vasculitis NULL 1
RxToMe Vasculitis 3 4
Rxtra Vasculitis NULL 1
Sand Point Pharmacy Vasculitis NULL 1
Sav-On Vasculitis NULL 1
ScriptSave Vasculitis NULL 3
ScriptSite Specialty Vasculitis 2 2
Sharp Specialty Pharmacy Vasculitis NULL 1
Simple Meds Vasculitis NULL 1
Smart Pharmacy Vasculitis 2 3
Southwest Pharmacy Vasculitis 2 2
Spot Rx Vasculitis NULL 2
Sunwest Vasculitis NULL 1
The Chemist Vasculitis NULL 1
The Downtown Dispensary Vasculitis NULL 1
The Rx Advocates Vasculitis NULL 1
Thrifty Way Pharmacy Vasculitis NULL 1
Tru Script Vasculitis NULL 1
True Pill Vasculitis 1 3
Union Center Pharmacy Vasculitis 2 4
University Pharmacy Vasculitis NULL 1
Wellcare Vasculitis 1 3
Wellfresh Vasculitis 1 2
Wellman's Pharmacy Vasculitis 1 1
Wellness Craft Vasculitis NULL 2
Wellness RX Vasculitis NULL 1
Wellocity Vasculitis NULL 2
Wellpharma Vasculitis 1 2
Wellport Drugs Vasculitis 1 2
Welltrack Vasculitis NULL 1
Wellwise Vasculitis 1 1
White Pigeon Pharmacy Vasculitis 2 3
Time taken: 66.439 seconds, Fetched: 6815 row(s)
hive> create view v_2 as select v1.pharmacyname,v1.diseasename,v2.no_of_prescriptions_2022,v1.total_prescriptions from
> v_123 v1 left join v_234 v2 on v1.pharmacyname=v2.pharmacyname and v1.diseasename=v2.diseasename;
OK
Time taken: 0.234 seconds
hive> 

```

```

Applications Places System cloudera@quickstart:~ Tue Mar 14, 04:45 cloudera
File Edit View Search Terminal Help
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1678709693264_0038
Hadoop job information for Stage-16: number of mappers: 1; number of reducers: 0
2023-03-14 04:32:23,881 Stage-16 map = 0%, reduce = 0%
2023-03-14 04:32:40,269 Stage-16 map = 100%, reduce = 0%, Cumulative CPU 3.1 sec
MapReduce Total cumulative CPU time: 3 seconds 100 msec
Ended Job = job_1678709693264_0038
MapReduce Jobs Launched:
Stage-Stage-4: Map: 1 Reduce: 1 Cumulative CPU: 12.15 sec HDFS Read: 339250 HDFS Write: 351789 SUCCESS
Stage-Stage-12: Map: 1 Reduce: 1 Cumulative CPU: 10.61 sec HDFS Read: 340118 HDFS Write: 156207 SUCCESS
Stage-Stage-16: Map: 1 Cumulative CPU: 3.1 sec HDFS Read: 5558 HDFS Write: 15 SUCCESS
Total MapReduce CPU Time Spent: 25 seconds 860 msec
OK
NULL NULL NULL 13428
Time taken: 231.779 seconds, Fetched: 1 row(s)
hive> create external table opt4(pharmacyname varchar(30),diseasename varchar(30),no_of_prescription int,total_prescription int) comment 'pharmacy details' row form
at delimited fields terminated by '\t' lines terminated by '\n' location '/user/hive/output4';
OK
Time taken: 0.217 seconds
hive> insert overwrite table output4 select * from v_4;
FAILED: SemanticException [Error 10001]: Line 1:23 Table not found 'output4'
hive> insert overwrite table output4 select * from v_4;
FAILED: SemanticException [Error 10001]: Line 1:23 Table not found 'output4'
hive> insert overwrite table opt4 select * from v_4;
Query ID = cloudera_20230314044343_0efbef32-8c66-49e3-a187-22c0866f0f2c
Total jobs = 4
Execution log at: /tmp/cloudera/cloudera_20230314044343_0efbef32-8c66-49e3-a187-22c0866f0f2c.log
2023-03-14 04:43:45 Starting to launch local task to process map join; maximum memory = 1013645312
2023-03-14 04:43:49 Dump the side-table for tag: 1 with group count: 40 into file: file:/tmp/cloudera/9ac20b85-fb06-4ebc-b86d-4f48e6ef39bd/hive_2023-03-14_04-43-30_814_7790535282660878174-1/-local-10012/HashTable-Stage-4/MapJoin-mapfile81--.hashtable
2023-03-14 04:43:49 Uploaded 1 File to: file:/tmp/cloudera/9ac20b85-fb06-4ebc-b86d-4f48e6ef39bd/hive_2023-03-14_04-43-30_814_7790535282660878174-1/-local-10012/HashTable-Stage-4/MapJoin-mapfile81--.hashtable(1754 bytes)
2023-03-14 04:43:49 Dump the side-table for tag: 1 with group count: 10885 into file: file:/tmp/cloudera/9ac20b85-fb06-4ebc-b86d-4f48e6ef39bd/hive_2023-03-14_04-43-30_814_7790535282660878174-1/-local-10012/HashTable-Stage-4/MapJoin-mapfile81--.hashtable
2023-03-14 04:43:50 Uploaded 1 File to: file:/tmp/cloudera/9ac20b85-fb06-4ebc-b86d-4f48e6ef39bd/hive_2023-03-14_04-43-30_814_7790535282660878174-1/-local-10012/HashTable-Stage-4/MapJoin-mapfile91--.hashtable(25183 bytes)
2023-03-14 04:43:50 Dump the side-table for tag: 0 with group count: 213 into file: file:/tmp/cloudera/9ac20b85-fb06-4ebc-b86d-4f48e6ef39bd/hive_2023-03-14_04-43-30_814_7790535282660878174-1/-local-10012/HashTable-Stage-4/MapJoin-mapfile100--.hashtable
2023-03-14 04:43:50 Uploaded 1 File to: file:/tmp/cloudera/9ac20b85-fb06-4ebc-b86d-4f48e6ef39bd/hive_2023-03-14_04-43-30_814_7790535282660878174-1/-local-10012/HashTable-Stage-4/MapJoin-mapfile100--.hashtable(7967 bytes)
2023-03-14 04:43:50 End of local task; Time Taken: 5.064 sec.
Execution completed successfully
MapredLocal task succeeded
Execution log at: /tmp/cloudera/cloudera_20230314044343_0efbef32-8c66-49e3-a187-22c0866f0f2c.log
2023-03-14 04:44:00 Starting to launch local task to process map join; maximum memory = 1013645312
2023-03-14 04:44:14 Dump the side-table for tag: 1 with group count: 40 into file: file:/tmp/cloudera/9ac20b85-fb06-4ebc-b86d-4f48e6ef39bd/hive_2023-03-14_04-43-30_814_7790535282660878174-1/-local-10014/HashTable-Stage-13/MapJoin-mapfile111--.hashtable
2023-03-14 04:44:14 Uploaded 1 File to: file:/tmp/cloudera/9ac20b85-fb06-4ebc-b86d-4f48e6ef39bd/hive_2023-03-14_04-43-30_814_7790535282660878174-1/-local-10014/HashTable-Stage-13/MapJoin-mapfile111--.hashtable(1754 bytes)
2023-03-14 04:44:14 Dump the side-table for tag: 1 with group count: 2967 into file: file:/tmp/cloudera/9ac20b85-fb06-4ebc-b86d-4f48e6ef39bd/hive_2023-03-14_04-43-30_814_7790535282660878174-1/-local-10014/HashTable-Stage-13/MapJoin-mapfile121--.hashtable
2023-03-14 04:44:15 Uploaded 1 File to: file:/tmp/cloudera/9ac20b85-fb06-4ebc-b86d-4f48e6ef39bd/hive_2023-03-14_04-43-30_814_7790535282660878174-1/-local-10014/HashTable-Stage-13/MapJoin-mapfile121--.hashtable(68834 bytes)

```

Query4 export:

```

Applications Places System Tue Mar 14, 08:06 cludera
File Edit View Search Terminal Help Change account settings and status
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=566324
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=269680
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=16
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=0
Job Counters
  Launched map tasks=4
  Data-local map tasks=4
  Total time spent by all maps in occupied slots (ms)=42035
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=42035
  Total vcore-seconds taken by all map tasks=42035
  Total megabyte-seconds taken by all map tasks=43043840
Map-Reduce Framework
  Map input records=6815
  Map output records=6815
  Input split bytes=556
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=562
  CPU time spent (ms)=2810
  Physical memory (bytes) snapshot=507924480
  Virtual memory (bytes) snapshot=6019416064
  Total committed heap usage (bytes)=243531776
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=0
23/03/14 08:05:56 INFO mapreduce.ExportJobBase: Transferred 263.3594 KB in 23.7566 seconds (11.0857 KB/sec)
23/03/14 08:05:56 INFO mapreduce.ExportJobBase: Exported 6815 records.
[cludera@quickstart ~]$ 

```

Clientdb Solution 4

```

mysql> create table solution4(pharmaname varchar(100), prescription varchar(100
),prescription_count_2022 varchar(10), total_prescriptions bigint);
Query OK, 0 rows affected (0.00 sec)

mysql> select * from solution4 limit 10;
+-----+-----+-----+
| pharmaname      | prescription      | prescription_count_2022 | total_p
descriptions |
+-----+-----+-----+
| \N           | \N           | \N           |
| 13428 |          | Alzheimer's disease | \N
| 326 |          | Alzheimer's disease | \N
| Absolute Care | Alzheimer's disease | \N
| 1 |          | Alzheimer's disease | \N
| Acculife Drug Stores | Alzheimer's disease | 2
| 5 |          | Alzheimer's disease | \N
| Adams Drugs | Alzheimer's disease | 1
| 2 |          | Alzheimer's disease | \N
| Alliance Community | Alzheimer's disease | \N
| 2 |          | Alzheimer's disease | \N
| Ally Scripts | Alzheimer's disease | \N
| 1 |          | Alzheimer's disease | \N
| Apotheco | Alzheimer's disease | \N
| 2 |          | Alzheimer's disease | \N
| Banks Apothecary | Alzheimer's disease | \N
| 1 |          | Alzheimer's disease | \N
| Bartell Drugs | Alzheimer's disease | 1
| 1 |          | Alzheimer's disease | \N
+-----+-----+-----+
10 rows in set (0.00 sec)

mysql> select count(*) from solution4;
+-----+
| count(*) |
+-----+
| 6815 |
+-----+
1 row in set (0.00 sec)

mysql> 

```

QUERY 5:

The healthcare department wants to know which category of patients is being affected the most by each disease.

Assist the department in creating a report regarding this.

Provided the healthcare department has categorized the patients into the following category.

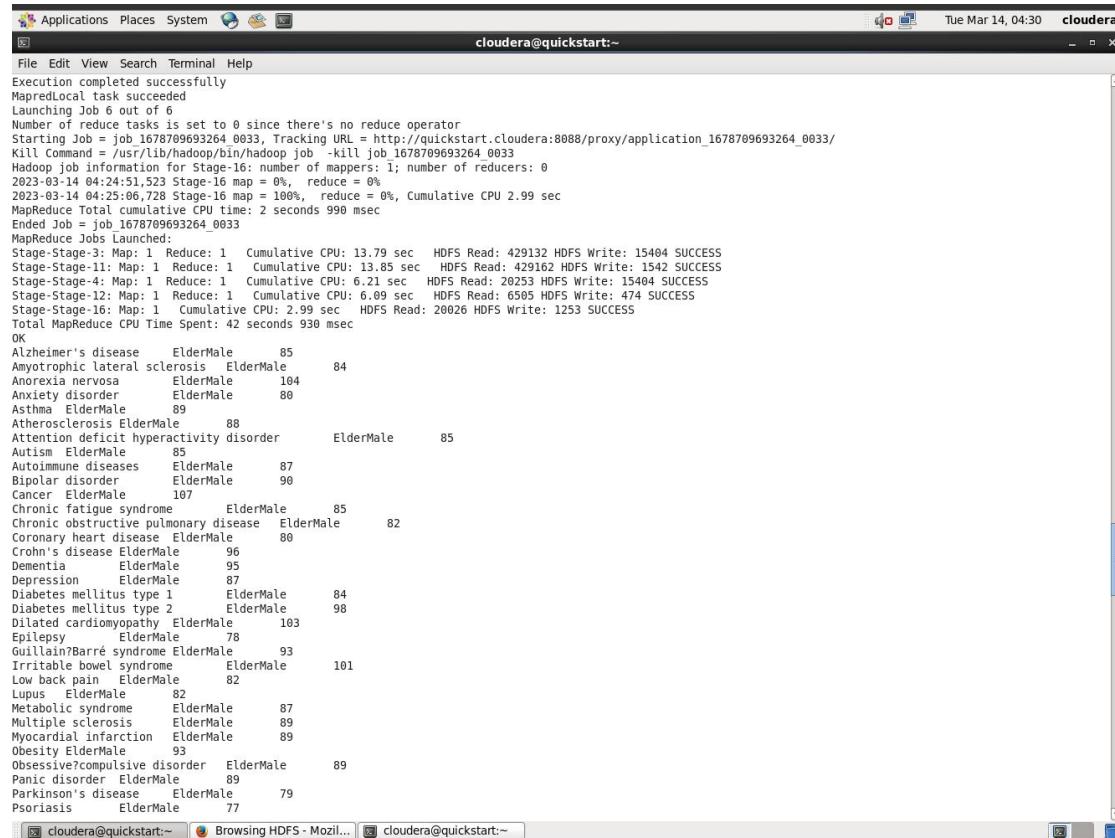
YoungMale: Born on or after 1st Jan 2005 and gender male.
 YoungFemale: Born on or after 1st Jan 2005 and gender female.
 AdultMale: Born before 1st Jan 2005 but on or after 1st Jan 1985 and gender male.
 AdultFemale: Born before 1st Jan 2005 but on or after 1st Jan 1985 and gender female.
 MidAgeMale: Born before 1st Jan 1985 but on or after 1st Jan 1970 and gender male.
 MidAgeFemale: Born before 1st Jan 1985 but on or after 1st Jan 1970 and gender female.
 ElderMale: Born before 1st Jan 1970, and gender male.
 ElderFemale: Born before 1st Jan 1970, and gender female.

```

create view v_category as
with cte_3 as (
select diseasename ,
case when p.dob>='2005-01-01' and pe.gender='male' then "YoungMale" when p.dob>='2005-01-01'
and pe.gender='female' then "YoungFemale" when p.dob<'2005-01-01' and p.dob>= '1985-01-01'and
pe.gender='male' then "AdultMale" when p.dob<'2005-01-01' and p.dob>= '1985-01-01'and
pe.gender='female' then "AdultFemale" when p.dob<'1985-01-01' and p.dob>= '1970-01-01'and
pe.gender='male' then "MidAgeMale" when p.dob<'1985-01-01' and p.dob>= '1970-01-01'and
pe.gender='female' then "MidAgeFemale" when p.dob<'1970-01-01'and pe.gender='male' then
"ElderMale" when p.dob<'1970-01-01'and pe.gender='female' then "ElderFemale" end as category
from person pe join patient p on pe.personid=p.patientid
join treatment t on p.patientid=t.patientid
join disease d on d.diseaseid=t.diseaseid)
select diseasename, count(*) over(partition by diseasename,category ) as count,category
from cte_3 ;
  
```

```

create view v_5 as select diseasename,category,count(*)from v_category group by
diseasename,category
having count(*)>= any (select max(count) from v_category group by diseasename);
  
```



The screenshot shows a terminal window titled 'cloudera@quickstart:~'. The window displays the execution of a MapReduce job and the resulting output of the view v_5.

Execution completed successfully

MapredLocal task succeeded

Launching Job 6 out of 6

Number of reduce tasks is set to 0 since there's no reduce operator

Starting Job = job_1678709693264_0033, Tracking URL: http://quickstart.cloudera:8088/proxy/application_1678709693264_0033/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1678709693264_0033

Hadoop job information for Stage-16: number of mappers: 1; number of reducers: 0

2023-03-14 04:24:51,523 Stage-16 map = 0%, reduce = 0%

2023-03-14 04:25:06,728 Stage-16 map = 100%, reduce = 0%, Cumulative CPU 2.99 sec

MapReduce Total cumulative CPU time: 2 seconds 990 msec

Ended Job = job_1678709693264_0033

MapReduce Jobs Launched:

- Stage-Stage-3: Map: 1 Reduce: 1 Cumulative CPU: 13.79 sec HDFS Read: 429132 HDFS Write: 15404 SUCCESS
- Stage-Stage-11: Map: 1 Reduce: 1 Cumulative CPU: 13.85 sec HDFS Read: 429162 HDFS Write: 1542 SUCCESS
- Stage-Stage-4: Map: 1 Reduce: 1 Cumulative CPU: 6.21 sec HDFS Read: 20253 HDFS Write: 15404 SUCCESS
- Stage-Stage-12: Map: 1 Reduce: 1 Cumulative CPU: 6.09 sec HDFS Read: 6505 HDFS Write: 474 SUCCESS
- Stage-Stage-16: Map: 1 Cumulative CPU: 2.99 sec HDFS Read: 20026 HDFS Write: 1253 SUCCESS

Total MapReduce CPU Time Spent: 42 seconds 930 msec

OK

Disease	Category	Count
Alzheimer's disease	ElderMale	85
Amyotrophic lateral sclerosis	ElderMale	84
Anorexia nervosa	ElderMale	104
Anxiety disorder	ElderMale	80
Asthma	ElderMale	89
Atherosclerosis	ElderMale	88
Attention deficit hyperactivity disorder	ElderMale	85
Autism	ElderMale	85
Autoimmune diseases	ElderMale	87
Bipolar disorder	ElderMale	90
Cancer	ElderMale	107
Chronic fatigue syndrome	ElderMale	85
Chronic obstructive pulmonary disease	ElderMale	82
Coronary heart disease	ElderMale	80
Crohn's disease	ElderMale	96
Dementia	ElderMale	95
Depression	ElderMale	87
Diabetes mellitus type 1	ElderMale	84
Diabetes mellitus type 2	ElderMale	98
Dilated cardiomyopathy	ElderMale	103
Epilepsy	ElderMale	78
Guillain?Barré syndrome	ElderMale	93
Irritable bowel syndrome	ElderMale	101
Low back pain	ElderMale	82
Lupus	ElderMale	82
Metabolic syndrome	ElderMale	87
Multiple sclerosis	ElderMale	89
Myocardial infarction	ElderMale	89
Obesity	ElderMale	93
Obsessive?compulsive disorder	ElderMale	89
Panic disorder	ElderMale	89
Parkinson's disease	ElderMale	79
Psoriasis	ElderMale	77

Query5 export:

```
sqoop export --connect jdbc:mysql://localhost:3306/client --username root --P --table solution5 --export-dir /user/hive/output5/000000_0 --input-fields-terminated-by ''
```

The screenshot shows a terminal window titled "cloudera@quickstart:~". The window displays the command "sqoop export" followed by its options and the resulting system counter output. The output includes details about file and system counters, job counters, and the Map-Reduce framework. It also shows the transfer of data and the number of records exported.

```
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=566152
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=3936
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=19
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=0

Job Counters
  Launched map tasks=4
  Data-local map tasks=4
  Total time spent by all maps in occupied slots (ms)=215063
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=215063
  Total vcore-seconds taken by all map tasks=215063
  Total megabyte-seconds taken by all map tasks=220224512

Map-Reduce Framework
  Map input records=40
  Map output records=40
  Input split bytes=631
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=3415
  CPU time spent (ms)=6750
  Physical memory (bytes) snapshot=502734848
  Virtual memory (bytes) snapshot=6015508480
  Total committed heap usage (bytes)=243531776

File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=0

23/03/14 07:52:06 INFO mapreduce.ExportJobBase: Transferred 3.8438 KB in 84.7906 seconds (46.4202 bytes/sec)
23/03/14 07:52:06 INFO mapreduce.ExportJobBase: Exported 40 records.
[cloudera@quickstart ~]$
```

Client db solution5:

```

Applications Places System  cloudera@quickstart:~ Tue Mar 14, 07:52
File Edit View Search Terminal Help

mysql> create table solution5(diseasename varchar(100), category varchar(100), count bigint);
Query OK, 0 rows affected (0.03 sec)

mysql> select * from solution5;
+-----+-----+-----+
| diseasename | category | count |
+-----+-----+-----+
| Panic disorder | ElderMale | 89 |
| Parkinson's disease | ElderMale | 79 |
| Psoriasis | ElderMale | 77 |
| Rheumatoid arthritis | ElderMale | 84 |
| Sarcoidosis | ElderMale | 92 |
| Schizophrenia | ElderMale | 100 |
| Stroke | ElderMale | 89 |
| Thromboangiitis obliterans | ElderMale | 88 |
| Tourette syndrome | ElderMale | 85 |
| Vasculitis | ElderMale | 88 |
| Alzheimer's disease | ElderMale | 85 |
| Amyotrophic lateral sclerosis | ElderMale | 84 |
| Anorexia nervosa | ElderMale | 104 |
| Anxiety disorder | ElderMale | 88 |
| Asthma | ElderMale | 89 |
| Atherosclerosis | ElderMale | 88 |
| Attention deficit hyperactivity disorder | ElderMale | 85 |
| Autism | ElderMale | 85 |
| Autoimmune diseases | ElderMale | 87 |
| Bipolar disorder | ElderMale | 90 |
| Cancer | ElderMale | 107 |
| Chronic fatigue syndrome | ElderMale | 85 |
| Chronic obstructive pulmonary disease | ElderMale | 82 |
| Coronary heart disease | ElderMale | 88 |
| Crohn's disease | ElderMale | 96 |
| Dementia | ElderMale | 95 |
| Depression | ElderMale | 87 |
| Diabetes mellitus type 1 | ElderMale | 84 |
| Diabetes mellitus type 2 | ElderMale | 98 |
| Dilated cardiomyopathy | ElderMale | 103 |
| Epilepsy | ElderMale | 78 |
| Guillain-Barré syndrome | ElderMale | 93 |
| Irritable bowel syndrome | ElderMale | 101 |
| Low back pain | ElderMale | 82 |
| Lupus | ElderMale | 82 |
| Metabolic syndrome | ElderMale | 87 |
| Multiple sclerosis | ElderMale | 89 |
| Myocardial infarction | ElderMale | 89 |
| Obesity | ElderMale | 93 |
| Obsessive-compulsive disorder | ElderMale | 89 |
+-----+-----+-----+
40 rows in set (0.00 sec)

mysql>

```

QUERY 6:

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.

```

select a.state,count(*)
from treatment t left join claim c on t.claimid=c.claimid
left join patient p on t.patientid=p.patientid
left join person pe on p.patientid=pe.personid
left join address_partitioned a on pe.addressid=a.addressid
where t.claimid IS NULL
group by a.state;

```

Before Partitioning

File Edit View Search Terminal Help

```
7/HashTable-Stage-4/MapJoin-mapfile162--.hashtable
2023-03-15 03:00:05      Uploaded 1 File to: file:/tmp/cloudera/b2c68120-3ee9-49f1-b4f3-9cd474831956/hive_2023-03-15_02-59-47_140_725838844012247361-1/-local-10007/HashTable-Stage-4/MapJoin-mapfile162--.hash
table (7026 bytes)
2023-03-15 03:00:05      Dump the side-table for tag: 1 with group count: 0 into file: file:/tmp/cloudera/b2c68120-3ee9-49f1-b4f3-9cd474831956/hive_2023-03-15_02-59-47_140_725838844012247361-1/-local-10007/H
ashTable-Stage-4/MapJoin-mapfile171--.hashtable
2023-03-15 03:00:05      Uploaded 1 File to: file:/tmp/cloudera/b2c68120-3ee9-49f1-b4f3-9cd474831956/hive_2023-03-15_02-59-47_140_725838844012247361-1/-local-10007/H
ashTable-Stage-4/MapJoin-mapfile171--.hashtable
2023-03-15 03:00:05      End of local task; Time Taken: 5.887 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_1678709693264_0150, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678709693264_0150/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1678709693264_0150
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 1
2023-03-15 03:00:25,067 Stage-4 map = 0%, reduce = 0%
2023-03-15 03:00:44,951 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 5.91 sec
2023-03-15 03:01:03,542 Stage-4 map = 100%, reduce = 100%, Cumulative CPU 9.22 sec
MapReduce Total cumulative CPU time: 9 seconds 220 msec
Ended Job = job_1678709693264_0150
MapReduce Jobs Launched:
Stage-Stage-4: Map: 1 Reduce: 1   Cumulative CPU: 9.22 sec   HDFS Read: 425484 HDFS Write: 112 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 220 msec
OK
AK     150
AL     280
AR     216
AZ     212
CA     363
CO     253
CT     256
DC     243
FL     281
GA     256
KY     169
MA     183
MD     220
OK     314
TN     307
VT     219
Time taken: 78.615 seconds, Fetched: 16 row(s)
hive>
```

After partitioning:

File Edit View Search Terminal Help

```
table (260 bytes)
2023-03-14 09:02:00      Dump the side-table for tag: 1 with group count: 0 into file: file:/tmp/cloudera/b2c68120-3ee9-49f1-b4f3-9cd474831956/hive_2023-03-14_21-01-32_875_2866428592903610360-1/-local-10007/H
ashTable-Stage-4/MapJoin-mapfile21--.hashtable
2023-03-14 09:02:00      Uploaded 1 File to: file:/tmp/cloudera/b2c68120-3ee9-49f1-b4f3-9cd474831956/hive_2023-03-14_21-01-32_875_2866428592903610360-1/-local-10007/HashTable-Stage-4/MapJoin-mapfile21--.hash
table (260 bytes)
2023-03-14 09:02:00      End of local task; Time Taken: 3.341 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_1678709693264_0107, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1678709693264_0107/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1678709693264_0107
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 1
2023-03-14 21:02:27,278 Stage-4 map = 0%, reduce = 0%
2023-03-14 21:02:45,668 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 3.69 sec
2023-03-14 21:03:02,404 Stage-4 map = 100%, reduce = 100%, Cumulative CPU 6.33 sec
MapReduce Total cumulative CPU time: 6 seconds 330 msec
Ended Job = job_1678709693264_0107
MapReduce Jobs Launched:
Stage-Stage-4: Map: 1 Reduce: 1   Cumulative CPU: 6.33 sec   HDFS Read: 15109 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 330 msec
OK
Time taken: 90.984 seconds
hive>
```

```
Applications Places System cloudera@quickstart:~ Wed Mar 15, 03:02 cloudera
File Edit View Search Terminal Help
cloudera/b2c68120-3ee9-49f1-b4f3-9cd474831956/hive_2023-03-15_02-57-44_763_6988705894246858029-1/-local-100
97/HashTable-Stage-4/MapJoin-mapfile132--.hashtable
2023-03-15 02:58:03     Uploaded 1 File to: file:/tmp/cloudera/b2c68120-3ee9-49f1-b4f3-9cd474831956/hiv
e_2023-03-15_02-57-44_763_6988705894246858029-1/-local-10007/HashTable-Stage-4/MapJoin-mapfile132--.has
htable (70226 bytes)
2023-03-15 02:58:03     Dump the side-table for tag: 1 with group count: 0 into file: file:/tmp/clouder
a/b2c68120-3ee9-49f1-b4f3-9cd474831956/hive_2023-03-15_02-57-44_763_6988705894246858029-1/-local-10007/
HashTable-Stage-4/MapJoin-mapfile141--.hashtable
2023-03-15 02:58:03     Uploaded 1 File to: file:/tmp/cloudera/b2c68120-3ee9-49f1-b4f3-9cd474831956/hiv
e_2023-03-15_02-57-44_763_6988705894246858029-1/-local-10007/HashTable-Stage-4/MapJoin-mapfile141--.has
htable (260 bytes)
2023-03-15 02:58:03     End of local task; Time Taken: 5.888 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_1678709693264_0149, Tracking URL = http://quickstart.cloudera:8088/proxy/application_
1678709693264_0149/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1678709693264_0149
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 1
2023-03-15 02:58:22,066 Stage-4 map = 0%,  reduce = 0%
2023-03-15 02:58:41,827 Stage-4 map = 100%,  reduce = 0%, Cumulative CPU 5.93 sec
2023-03-15 02:59:00,323 Stage-4 map = 100%,  reduce = 100%, Cumulative CPU 9.39 sec
MapReduce Total cumulative CPU time: 9 seconds 390 msec
Ended Job = job_1678709693264_0149
MapReduce Jobs Launched:
Stage-Stage-4: Map: 1 Reduce: 1 Cumulative CPU: 9.39 sec HDFS Read: 430058 HDFS Write: 112 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 390 msec
OK
4K      150
4L      280
4R      216
4Z      212
CA      363
CO      253
CT      256
DC      243
FL      281
GA      256
KY      169
MA      183
MD      220
OK      314
TN      307
VT      219
Time taken: 77.769 seconds, Fetched: 16 row(s)
```

Query 6 export:

```
File Edit View Search Terminal Help
23/03/15 03:32:49 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce
.map.speculative
23/03/15 03:32:50 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1678709693264_0156
23/03/15 03:32:51 INFO impl.YarnClientImpl: Submitted application application_1678709693264_0156
23/03/15 03:32:51 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1678709693264_0156/
23/03/15 03:32:51 INFO mapreduce.Job: Running job: job_1678709693264_0156
23/03/15 03:33:09 INFO mapreduce.Job: Job job_1678709693264_0156 running in uber mode : false
23/03/15 03:33:09 INFO mapreduce.Job: map 0% reduce 0%
23/03/15 03:34:10 INFO mapreduce.Job: map 50% reduce 0%
23/03/15 03:34:12 INFO mapreduce.Job: map 75% reduce 0%
23/03/15 03:34:13 INFO mapreduce.Job: map 100% reduce 0%
23/03/15 03:34:13 INFO mapreduce.Job: Job job_1678709693264_0156 completed successfully
23/03/15 03:34:13 INFO mapreduce.Job: Counters
File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=566044
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=828
    HDFS: Number of bytes written=0
    HDFS: Number of read operations=16
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=0
Job Counters
    Launched map tasks=4
    Data-local map tasks=4
    Total time spent by all maps in occupied slots (ms)=229647
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=229647
    Total vcore-seconds taken by all map tasks=229647
    Total megabyte-seconds taken by all map tasks=235158528
Map-Reduce Framework
    Map input records=16
    Map output records=16
    Input split bytes=536
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=3797
    CPU time spent (ms)=7020
    Physical memory (bytes) snapshot=457289728
    Virtual memory (bytes) snapshot=6015803392
    Total committed heap usage (bytes)=243531776
File Input Format Counters
    Bytes Read=0
File Output Format Counters
    Bytes Written=0
23/03/15 03:34:13 INFO mapreduce.ExportJobBase: Transferred 828 bytes in 89.0916 seconds (9.2938 bytes/sec)
23/03/15 03:34:13 INFO mapreduce.ExportJobBase: Exported 16 records.
[cloudera@quickstart ~] $ sudo mysql -u root -p
cloudera@quickstart:~ cloudera@quickstart:~ Browsing HDFS - Mozilla... Close Firefox
```

```

File Edit View Search Terminal Help
Physical memory (bytes) snapshot=457289728
Virtual memory (bytes) snapshot=6015803392
Total committed heap usage (bytes)=243531776
File Input Format Counters
Bytes Read=0
File Output Format Counters
Bytes Written=0
23/03/15 03:34:13 INFO mapreduce.ExportJobBase: Transferred 828 bytes in 89.0916 seconds (9.2938 bytes/sec)
23/03/15 03:34:13 INFO mapreduce.ExportJobBase: Exported 16 records.
[cloudera@quickstart ~]$ sudo mysql -u root -p
Enter password:
Welcome to the MySQL monitor. Commands end with ; or \g.
Your MySQL connection id is 198
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use client;
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> select * from solution6;
+-----+-----+
| state | count |
+-----+-----+
| AK   | 150  |
| AL   | 280  |
| AR   | 216  |
| AZ   | 212  |
| CA   | 363  |
| CO   | 253  |
| CT   | 256  |
| DC   | 243  |
| FL   | 281  |
| GA   | 256  |
| KY   | 169  |
| MA   | 183  |
| MD   | 220  |
| OK   | 314  |
| TN   | 307  |
| VT   | 219  |
+-----+-----+
16 rows in set (0.01 sec)

mysql> ■

```

QUERY 7:

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 view v_7 as SELECT address_partitioned.state, COUNT(treatment.treatmentID) AS treat_count,
COUNT(claim.claimID) AS claim_count,
COUNT(treatment.treatmentID) / COUNT(claim.claimID) AS ratio
FROM address_partitioned
INNER JOIN person ON address_partitioned.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 address_partitioned.state;

```

Query 7 export:

```
Applications Places System 
File Edit View Search Terminal Help
File org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:789)
23/03/15 04:21:40 INFO mapreduce.JobSubmitter: number of splits:4
23/03/15 04:21:40 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
23/03/15 04:21:40 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1678709693264_0159
23/03/15 04:21:42 INFO impl.YarnClientImpl: Submitted application application_1678709693264_0159
23/03/15 04:21:42 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1678709693264_0159/
23/03/15 04:21:42 INFO mapreduce.Job: Job: Running job; job_1678709693264_0159
23/03/15 04:22:01 INFO mapreduce.Job: Job job_1678709693264_0159 running in uber mode : false
23/03/15 04:22:01 INFO mapreduce.Job: map 0% reduce 0%
23/03/15 04:22:57 INFO mapreduce.Job: map 50% reduce 0%
23/03/15 04:22:58 INFO mapreduce.Job: map 75% reduce 0%
23/03/15 04:22:59 INFO mapreduce.Job: map 100% reduce 0%
23/03/15 04:22:59 INFO mapreduce.Job: Job job_1678709693264_0159 completed successfully
23/03/15 04:22:59 INFO mapreduce.Job: Counters: 30
File System Counters
FILE: Number of bytes read=0
FILE: Number of bytes written=566196
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=1873
HDFS: Number of bytes written=0
HDFS: Number of read operations=19
HDFS: Number of large read operations=0
HDFS: Number of write operations=0
Job Counters
Launched map tasks=4
Data-local map tasks=4
Total time spent by all maps in occupied slots (ms)=214486
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=214486
Total vcore-seconds taken by all map tasks=214486
Total megabyte-seconds taken by all map tasks=219633664
Map-Reduce Framework
Map input records=16
Map output records=16
Input split bytes=631
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=3763
CPU time spent (ms)=6810
Physical memory (bytes) snapshot=459182080
Virtual memory (bytes) snapshot=6015827968
Total committed heap usage (bytes)=243531776
File Input Format Counters
Bytes Read=0
File Output Format Counters
Bytes Written=0
23/03/15 04:22:59 INFO mapreduce.ExportJobBase: Transferred 1.8291 KB in 83.7931 seconds (22.3527 bytes/sec)
23/03/15 04:22:59 INFO mapreduce.ExportJobBase: Exported 16 records.
[cloudera@quickstart ~]$
```

ClientDB solution7

```

mysql> use client;
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> select * from solution7;
+-----+-----+-----+-----+
| state | treat_count | claim_count | ratio |
+-----+-----+-----+-----+
| AK   |    98 |     67 | 1.46268656716418 |
| AL   |   213 |    130 | 1.63846153846154 |
| AR   |   141 |     92 | 1.5326869565217 |
| AZ   |   135 |     82 | 1.64634146341463 |
| CA   |   267 |    182 | 1.46703296703297 |
| OK   |   207 |    123 | 1.68292682926829 |
| TN   |   208 |    123 | 1.69105691056911 |
| VT   |   131 |     89 | 1.47191011235955 |
| 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 (0.00 sec)

mysql> █
[cloudera@quickstart:~] [cloudera@quickstart:~] [Browsing HDFS - Mozilla...]

```

QUERY 8:

```

create view q8 as SELECT pharmacyname,count(v1.medicineID),sum(v1.maxprice),sum(v1.discount_price) FROM
pharmacy INNER JOIN (SELECT k.pharmacyid,m.medicineID,m.maxprice,m.maxprice-(m.maxprice*(k.discount/100)) AS
discount_price FROM keep k INNER JOIN medicine m on m.medicineid=k.medicineid ) AS v1 ON
pharmacy.pharmacyid = v1.pharmacyid GROUP BY pharmacyname;

```

EXTERNAL TABLE:

```

create external table opt8(pharmacyname varchar(50),count int,sum_max float,sum_discount float) row format
delimited fields terminated by ',' lines terminated by '\n' location '/user/hive/output8'; OVERWRITE: Hive> insert
overwrite table opt8 select * from q8;

```

EXPORT:

```
[cloudera@quickstart ~]$ sqoop export --connect jdbc:mysql://localhost:3306/client --username root --P --table
solution8 --export-dir /user/hive/output8/000000_0 --input-fields-terminated-by ','
```

OUTPUT:

```

File Edit View Search Terminal Help
Simple Meds      119      61676.97      49584.935999999994
Smart Pharmacy    256      127257.2000000006      105543.31400000003
Southside Family Pharmacy  177      125551.50000000003      111140.788
Southwest Pharmacy  461      208976.48000000007      189040.648
Spot Rx 455     170213.61999999994      144761.668
Sunwest 70       21048.04      19219.43400000005
Sure Save        331      216128.8299999996      180142.759
The Chemist       85       12250.04      10041.68000000002
The Compounding Pharmacy  258      75204.67000000003      60301.86699999998
The Downtown Dispensary 20      2194.270000000004      1875.972000000002
The Pill Club     219      95385.3100000001      80828.51900000001
The Rx Advocates  369      128145.8899999997      106118.84600000002
Thrifty Way Pharmacy  119      27634.16999999999      22690.524
Trinity Pharmacy   104      35381.71      30131.41699999998
Tru Script        149      37241.2099999999      31971.06699999995
True Pill         490      240604.0399999986      195707.70400000003
Union Center Pharmacy 114      44018.2999999996      37536.82500000004
University Pharmacy 486      148628.3800000006      123578.171
Vista Pharmacy    320      164827.39      143476.6770000003
Vitasource        130      29284.4999999996      24454.091
Wellcare          295      109663.1200000013      90347.1320000001
Wellfresh          229      68059.2500000001      59326.3850000001
Wellman's Pharmacy 252      146250.06      117875.4219999999
Wellness Craft    469      234017.4699999994      205501.8779999985
Wellness RX       438      149251.0599999994      123646.0899999994
Wellocity          352      143719.9000000008      122105.37400000005
Wellpharma        260      95757.02      79300.84
Wellport Drugs    406      362389.7399999976      297014.9469999997
Welltrack          301      157715.2899999998      129926.207
Wellwise           139      41686.2399999999      36278.576
White Pigeon Pharmacy 191      81827.0100000002      65362.819
Zenrx            319      262604.8799999999      198715.14700000003
Time taken: 41.058 seconds, Fetched: 213 row(s)
hive> █

```

SQOOP EXPORT query8:

```
Applications Places System cloudera@quickstart:~  
File Edit View Search Terminal Help  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=99722  
Total vcore-seconds taken by all map tasks=99722  
Total megabyte-seconds taken by all map tasks=102115328  
Map-Reduce Framework  
    Map input records=213  
    Map output records=213  
    Input split bytes=556  
    Spilled Records=0  
    Failed Shuffles=0  
    Merged Map outputs=0  
    GC time elapsed (ms)=824  
    CPU time spent (ms)=3220  
    Physical memory (bytes) snapshot=423038976  
    Virtual memory (bytes) snapshot=6015205376  
    Total committed heap usage (bytes)=243531776  
File Input Format Counters  
    Bytes Read=0  
File Output Format Counters  
    Bytes Written=0  
23/03/17 05:07:49 INFO mapreduce.ExportJobBase: Transferred 18.3428 KB in 44.164  
5 seconds (425.2966 bytes/sec)  
23/03/17 05:07:49 INFO mapreduce.ExportJobBase: Exported 213 records.  
[cloudera@quickstart ~]$
```

Client DB solution8 :

```
Applications Places System cloudera@quickstart:~ Fri Mar 17, 5:08 AM  
File Edit View Search Terminal Help  
mysql> use client;  
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> select * from solution8;  
+-----+-----+-----+-----+  
| pharmacyname | count | sum_max | sum_discount |  
+-----+-----+-----+-----+  
| Absolute Care | 460 | 244531.8 | 221840.6 |  
| Acculife Drug Stores | 441 | 126163.52 | 104723.445 |  
| Adams Drugs | 375 | 235377.12 | 202660.36 |  
| Alliance Community | 136 | 97313.96 | 82905.76 |  
| Ally Scripts | 205 | 54419.88 | 46148.84 |  
| Apotheco | 454 | 207948 | 174585.45 |  
| Assured Rx | 266 | 147459.34 | 131109.1 |  
| Atlas Drugs | 41 | 11520.18 | 9505.121 |  
| Banks Apothecary | 208 | 137390.56 | 117488.055 |  
| Bartell Drugs | 144 | 141589.52 | 134924.31 |  
| Be Well | 221 | 73483.27 | 64780.188 |  
| Below Drug | 415 | 248536.55 | 231600.9 |  
| Better Life | 200 | 105992.73 | 89069.64 |  
| Better You | 251 | 146893.27 | 120894.945 |  
| Beyond Health | 331 | 154187.23 | 134258.45 |  
| Bio Scrip | 441 | 296968.7 | 241029.86 |  
| Bioplus Specialty | 368 | 145346.92 | 122033.93 |  
| Blink Health | 473 | 169583.45 | 140611.17 |  
| Brennen Drugs Co | 263 | 93118.48 | 79607.09 |  
| Cardinal Health | 313 | 113821.08 | 96577.23 |  
| Care Pharmacy | 468 | 212823.12 | 185698.98 |  
| CareFirst | 131 | 72796.94 | 60257.465 |  
| Caremark | 391 | 226842.33 | 203845.73 |  
| Carepoint | 428 | 134650.34 | 115062.24 |  
| Cash Saver Pharmacy | 305 | 101332.34 | 86456.516 |  
| Cashway Pharmacy | 400 | 262923.3 | 226297.47 |  
| Cedra Pharmacy | 353 | 133574.94 | 119078.18 |  
| Central Drug Store | 281 | 186039.12 | 152678.36 |  
| Central Rx | 573.32 | 53517.145 |  
+-----+  
cloudera@quickstart:~
```