1) Starting the mysql by giving username as root and password as cloudera.

mysql -u root -pcloudera

```
File  Edit  View  Search  Terminal  Help
[cloudera@quickstart ~]$ mysql -u root -pcloudera
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 546
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>
```

**show databases;**

```
File  Edit  View  Search  Terminal  Help
owners.

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

mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| cm                 |
| firehose           |
| hue                |
| metastore          |
| mysql              |
| nav                |
| navms              |
| oozie              |
| retail_db          |
| rman               |
| sentry             |
+--------------------+
12 rows in set (0.00 sec)

mysql>
```

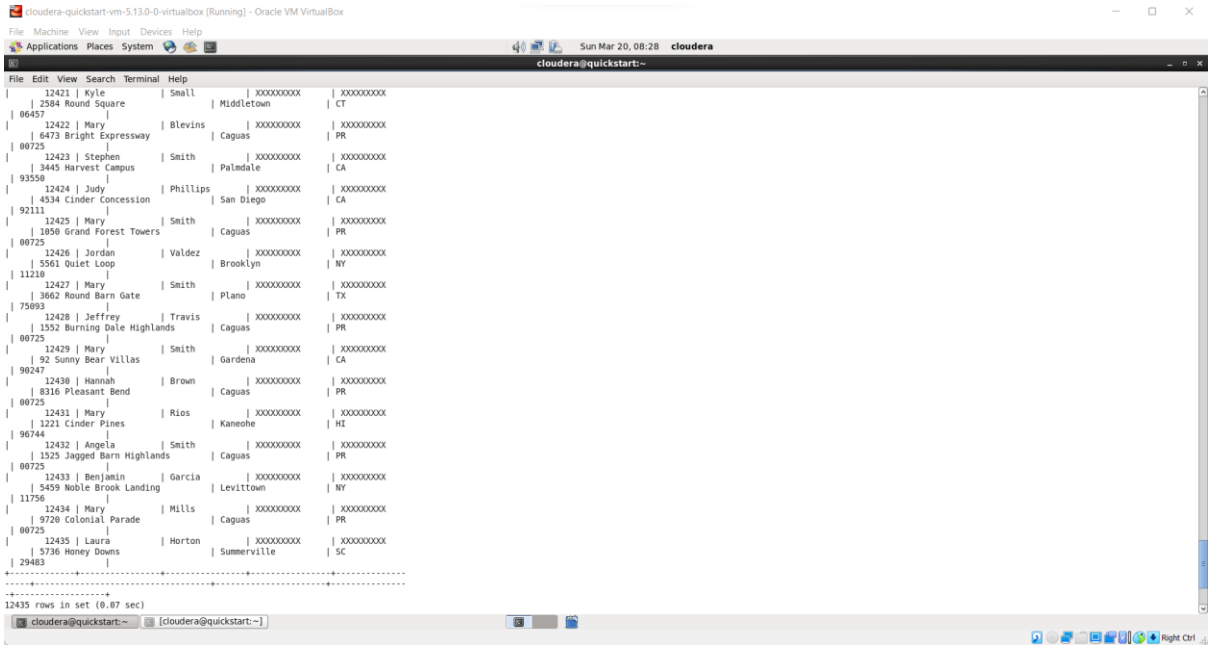
**Use retail_db;**

```
mysql> use retail_db;
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>
```

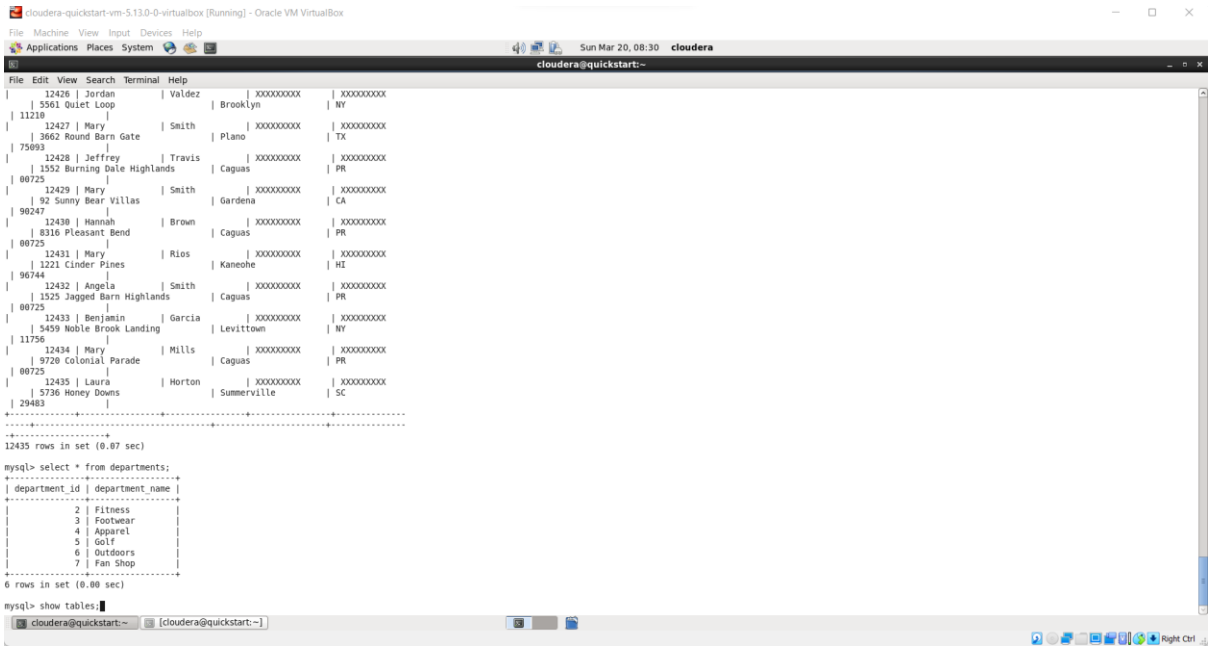
**Show tables;**

```
mysql> show tables;
+----------------------+
| Tables_in_retail_db  |
+----------------------+
| categories           |
| customers            |
| departments          |
| order_items          |
| orders               |
| products             |
+----------------------+
6 rows in set (0.00 sec)

mysql>
```

**Select * from customers;**

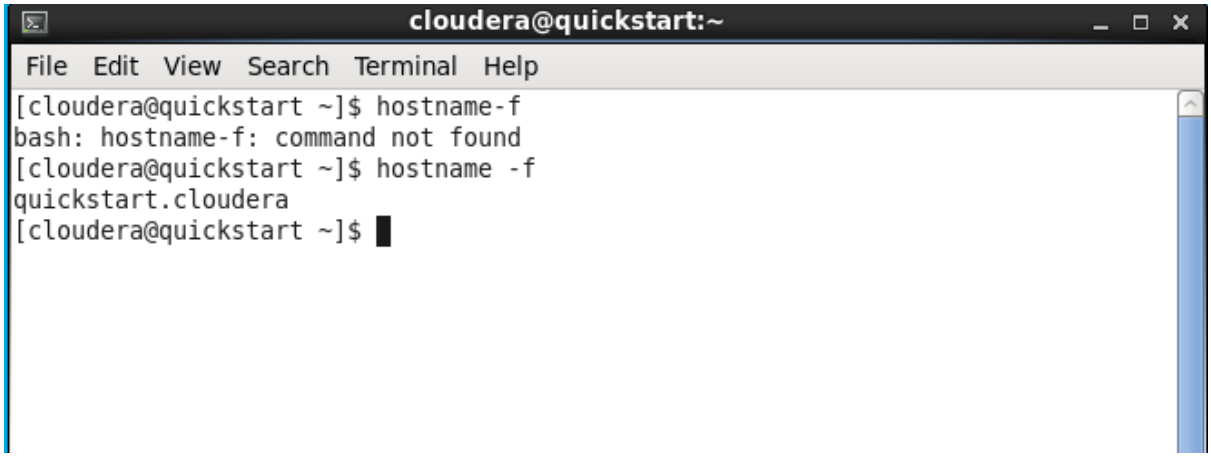



**Select * from departments;**




**Open the new terminal for running command for sqoop.**

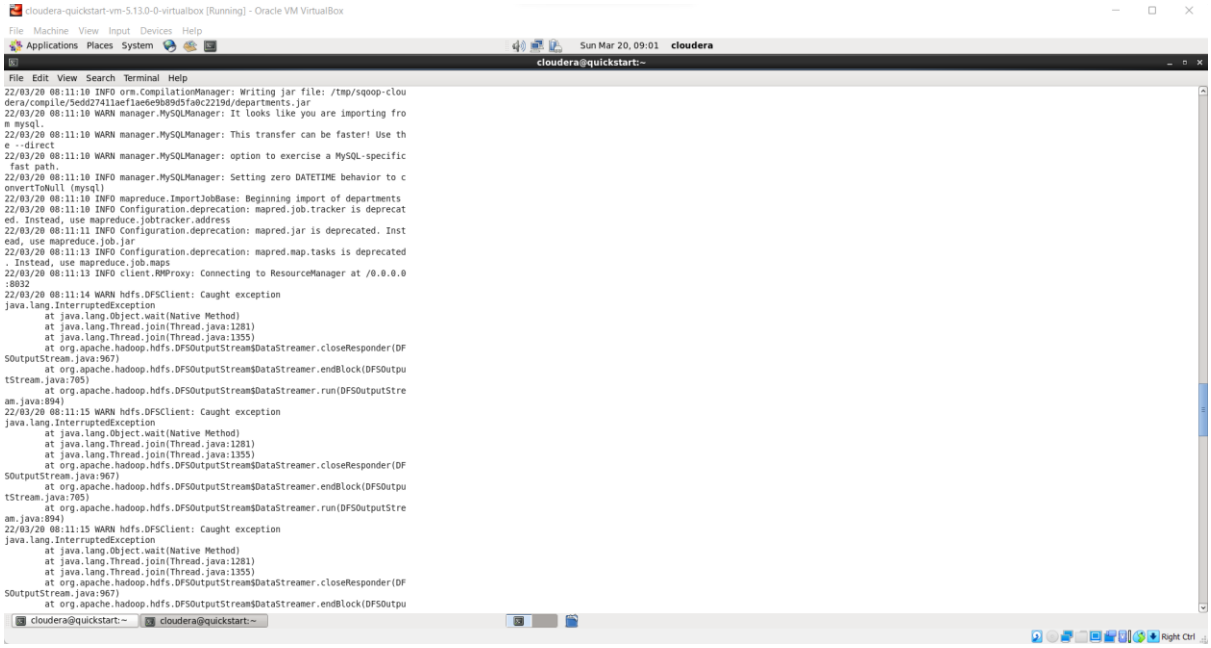
**Hostname-f**

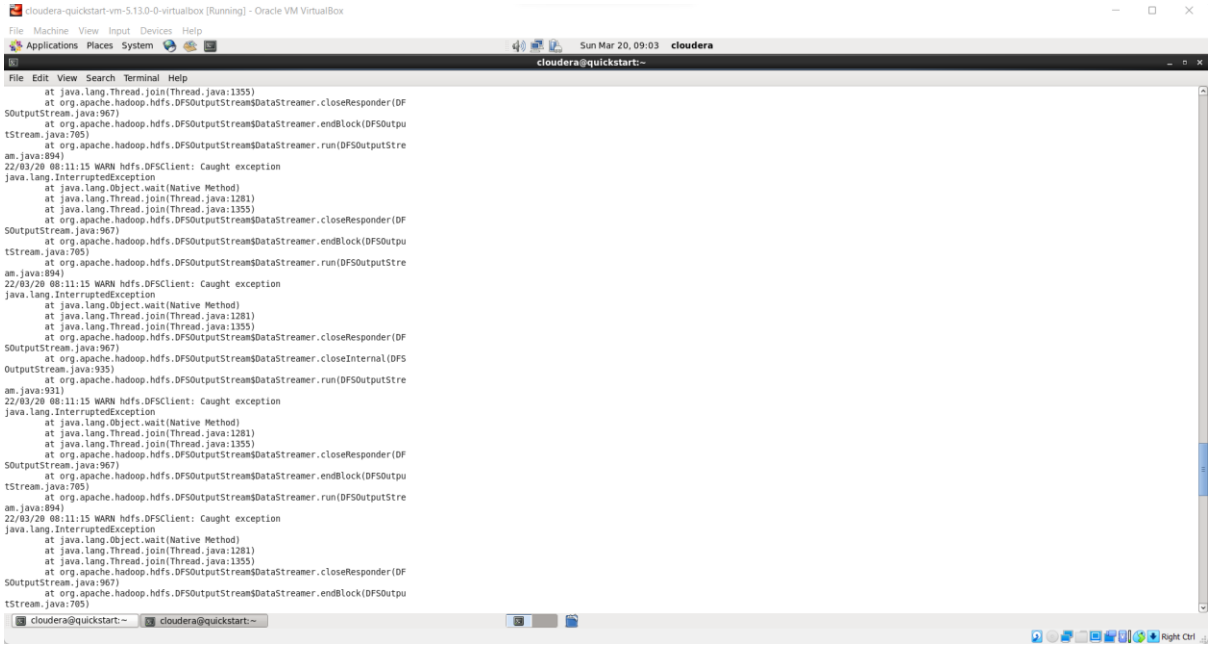


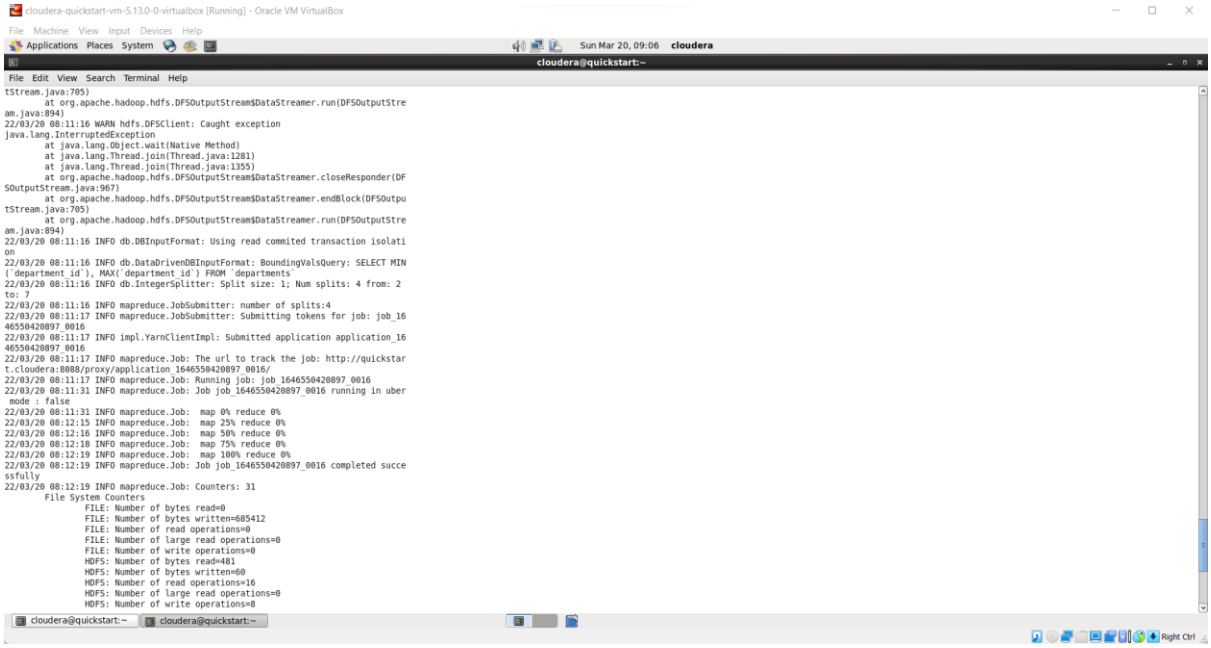

**sqoop list-databases --connect jdbc:mysql://quickstart:3306/ --password cloudera --username root;**

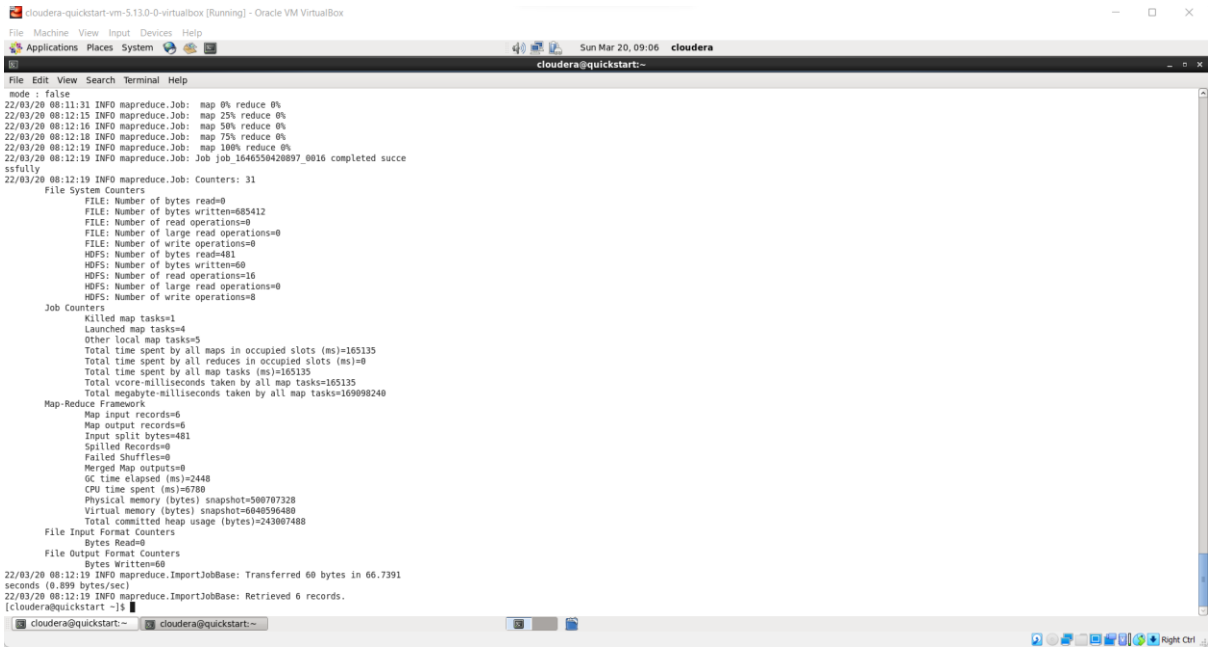
```
[cloudera@quickstart ~]$ sqoop list-tables --connect jdbc:mysql://quickstart:330
6/retail_db --password cloudera --username root;
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
22/03/08 10:24:17 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
22/03/08 10:24:17 WARN tool.BaseSqoopTool: Setting your password on the command-
line is insecure. Consider using -P instead.
22/03/08 10:24:19 INFO manager.MySQLManager: Preparing to use a MySQL streaming
resultset.
categories
customers
departments
order_items
orders
products
[cloudera@quickstart ~]$
```

sqoop list-tables --connect jdbc:mysql://quickstart:3306/retail_db --password cloudera --username root --table departments;


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 20, 09:01 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
22/03/20 08:11:10 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/5edd27411aef1ae6e9b89d5fa0c2219d/departments.jar
22/03/20 08:11:10 WARN manager.MySQLManager: It looks like you are importing from mysql.
22/03/20 08:11:10 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
22/03/20 08:11:10 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
22/03/20 08:11:10 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
22/03/20 08:11:10 INFO mapreduce.ImportJobBase: Beginning import of departments
22/03/20 08:11:10 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
22/03/20 08:11:11 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
22/03/20 08:11:13 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
22/03/20 08:11:13 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/03/20 08:11:14 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
cloudera@quickstart:~
cloudera@quickstart:~
Right Ctrl


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 20, 09:03 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeInternal(DFSOutputStream.java:935)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:931)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
cloudera@quickstart:~
cloudera@quickstart:~
Right Ctrl


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 20, 09:06 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
tStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStre
am.java:894)
22/03/20 08:11:16 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DF
SOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutpu
tStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStre
am.java:894)
22/03/20 08:11:16 INFO db.DBInputFormat: Using read commited transaction isolati
on
22/03/20 08:11:16 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN
(`department_id`), MAX(`department_id`) FROM `departments`
22/03/20 08:11:16 INFO db.IntegerSplitter: Split size: 1; Num splits: 4 from: 2
to: 7
22/03/20 08:11:16 INFO mapreduce.JobSubmitter: number of splits:4
22/03/20 08:11:17 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_16
46550420897_0016
22/03/20 08:11:17 INFO impl.YarnClientImpl: Submitted application application_16
46550420897_0016
22/03/20 08:11:17 INFO mapreduce.Job: The url to track the job: http://quickstar
t.cloudera:8088/proxy/application_1646550420897_0016/
22/03/20 08:11:17 INFO mapreduce.Job: Running job: job_1646550420897_0016
22/03/20 08:11:31 INFO mapreduce.Job: Job job_1646550420897_0016 running in uber
mode : false
22/03/20 08:11:31 INFO mapreduce.Job: map 0% reduce 0%
22/03/20 08:12:15 INFO mapreduce.Job: map 25% reduce 0%
22/03/20 08:12:16 INFO mapreduce.Job: map 50% reduce 0%
22/03/20 08:12:18 INFO mapreduce.Job: map 75% reduce 0%
22/03/20 08:12:19 INFO mapreduce.Job: map 100% reduce 0%
22/03/20 08:12:19 INFO mapreduce.Job: Job job_1646550420897_0016 completed succe
ssfully
22/03/20 08:12:19 INFO mapreduce.Job: Counters: 31
File System Counters
FILE: Number of bytes read=0
FILE: Number of bytes written=685412
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=481
HDFS: Number of bytes written=60
HDFS: Number of read operations=16
HDFS: Number of large read operations=0
HDFS: Number of write operations=8
cloudera@quickstart:~
cloudera@quickstart:~
Right Ctrl


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 20, 09:06 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
mode : false
22/03/20 08:11:31 INFO mapreduce.Job: map 0% reduce 0%
22/03/20 08:12:15 INFO mapreduce.Job: map 25% reduce 0%
22/03/20 08:12:16 INFO mapreduce.Job: map 50% reduce 0%
22/03/20 08:12:18 INFO mapreduce.Job: map 75% reduce 0%
22/03/20 08:12:19 INFO mapreduce.Job: map 100% reduce 0%
22/03/20 08:12:19 INFO mapreduce.Job: Job job_1646550420897_0016 completed succe
ssfully
22/03/20 08:12:19 INFO mapreduce.Job: Counters: 31
File System Counters
FILE: Number of bytes read=0
FILE: Number of bytes written=685412
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=481
HDFS: Number of bytes written=60
HDFS: Number of read operations=16
HDFS: Number of large read operations=0
HDFS: Number of write operations=8
Job Counters
Killed map tasks=1
Launched map tasks=4
Other local map tasks=5
Total time spent by all maps in occupied slots (ms)=165135
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=165135
Total vcore-milliseconds taken by all map tasks=165135
Total megabyte-milliseconds taken by all map tasks=169098240
Map-Reduce Framework
Map input records=6
Map output records=6
Input split bytes=481
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=2448
CPU time spent (ms)=6780
Physical memory (bytes) snapshot=500707328
Virtual memory (bytes) snapshot=6040596480
Total committed heap usage (bytes)=243007488
File Input Format Counters
Bytes Read=0
File Output Format Counters
Bytes Written=60
22/03/20 08:12:19 INFO mapreduce.ImportJobBase: Transferred 60 bytes in 66.7391
seconds (0.899 bytes/sec)
22/03/20 08:12:19 INFO mapreduce.ImportJobBase: Retrieved 6 records.
[cloudera@quickstart ~]$
cloudera@quickstart:~
cloudera@quickstart:~
Right Ctrl

**hadoop fs –ls**

```
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=60
22/03/20 08:12:19 INFO mapreduce.ImportJobBase: Transferred 60 bytes in 66.7391
seconds (0.899 bytes/sec)
22/03/20 08:12:19 INFO mapreduce.ImportJobBase: Retrieved 6 records.
[cloudera@quickstart ~]$ hadoop fs -ls
Found 2 items
drwxr-xr-x   - cloudera cloudera          0 2022-03-20 08:12 department1
drwxr-xr-x   - cloudera cloudera          0 2022-03-20 07:59 departments
[cloudera@quickstart ~]$
```

**hadoop fs –ls deparments;**

```
[cloudera@quickstart ~]$ hadoop fs -ls
Found 1 items
drwxr-xr-x   - cloudera cloudera          0 2022-03-20 07:59 departments
[cloudera@quickstart ~]$ hadoop fs -ls departments;
Found 5 items
-rw-r--r--   1 cloudera cloudera          0 2022-03-20 07:59 departments/_SUCCES
S
-rw-r--r--   1 cloudera cloudera         21 2022-03-20 07:59 departments/part-m-
00000
-rw-r--r--   1 cloudera cloudera         10 2022-03-20 07:59 departments/part-m-
00001
-rw-r--r--   1 cloudera cloudera          7 2022-03-20 07:59 departments/part-m-
00002
-rw-r--r--   1 cloudera cloudera         22 2022-03-20 07:59 departments/part-m-
00003
```

**hadoop fs -cat /user/cloudera/departments/part-m-00000**

**hadoop fs -cat /user/cloudera/departments/part-m-00001**

**hadoop fs -cat /user/cloudera/departments/part-m-00002**

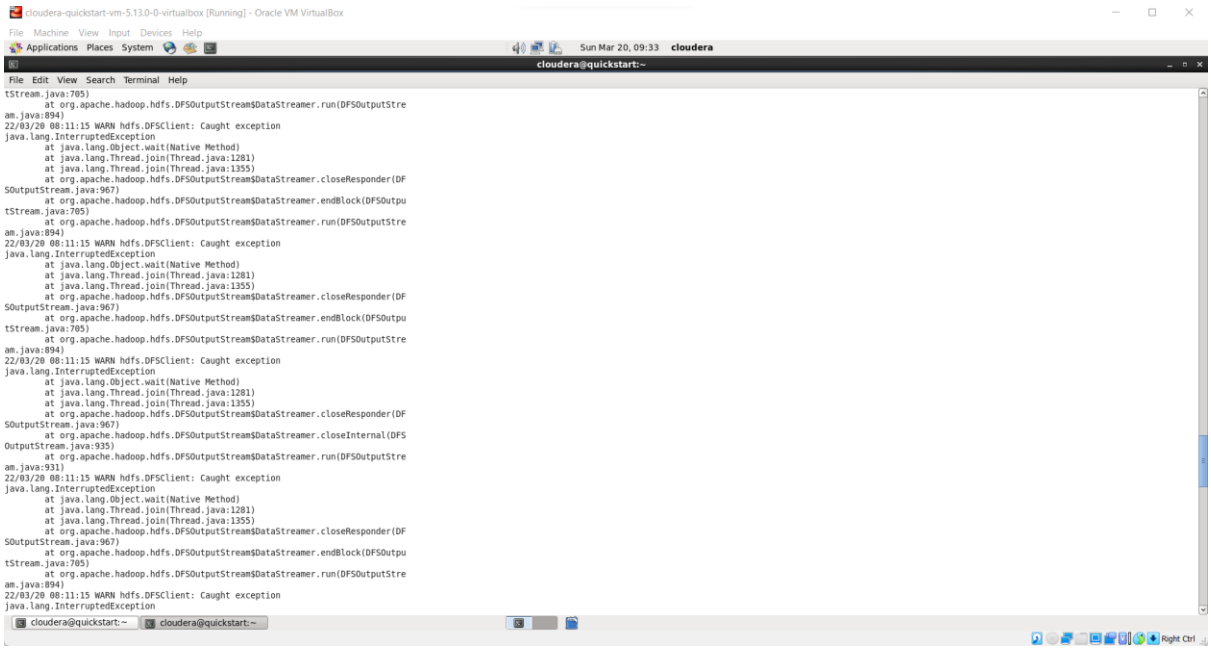
**hadoop fs -cat /user/cloudera/departments/part-m-00003**

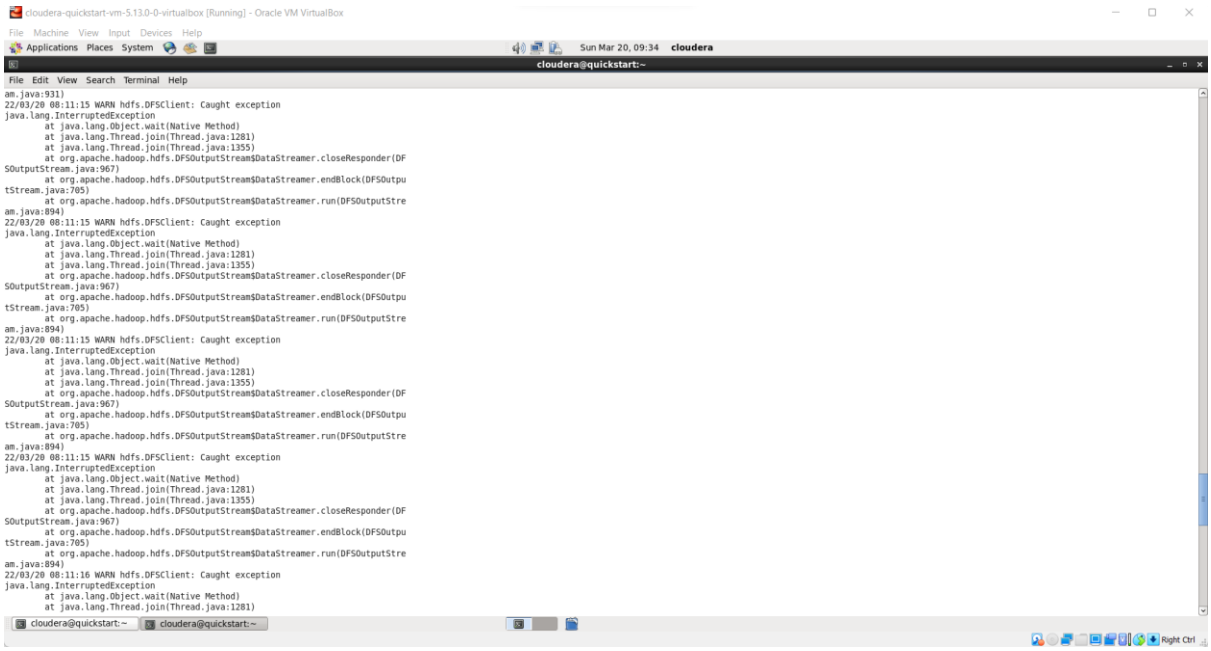
```
[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/departments/part-m-00000
2,Fitness
3,Footwear
[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/departments/part-m-00001
4,Apparel
[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/departments/part-m-00002
5,Golf
[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/departments/part-m-00006
cat: `/user/cloudera/departments/part-m-00006': No such file or directory
```

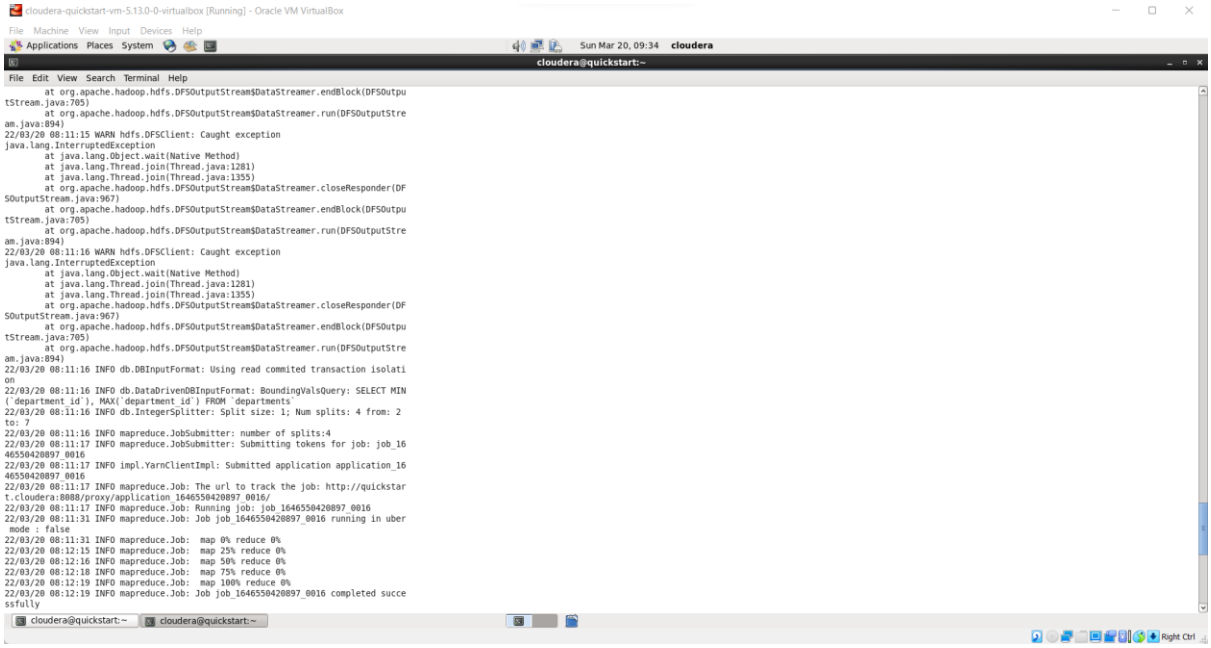
**hadoop fs -cat /user/cloudera/departments/part***

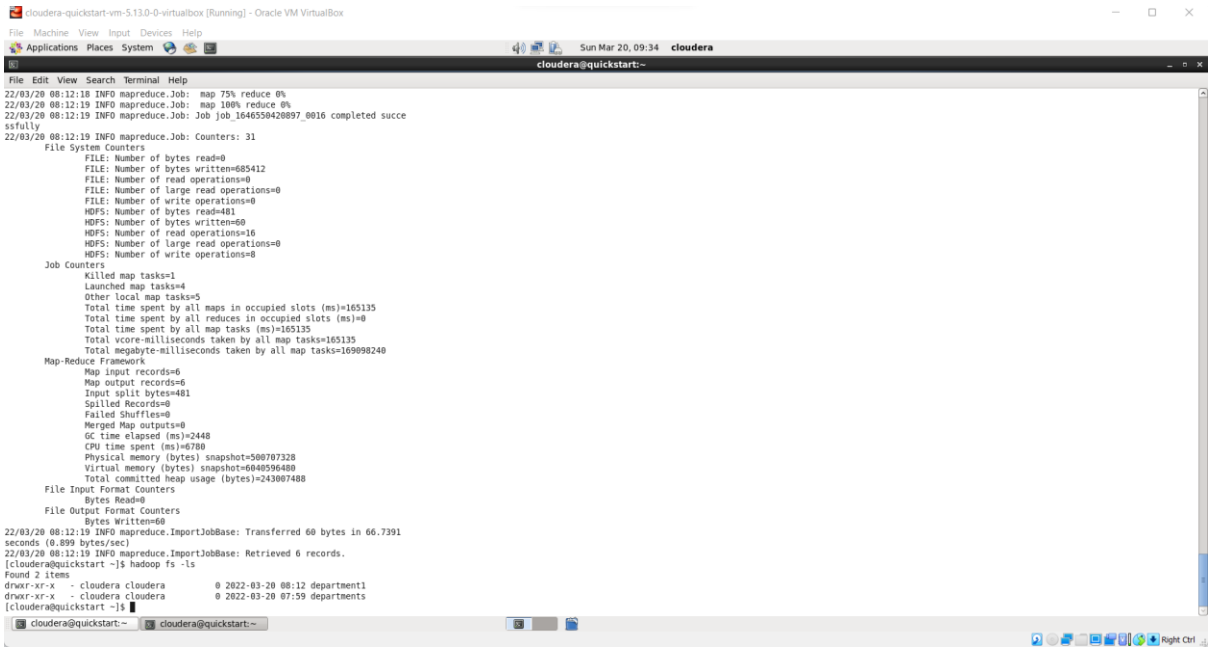
```
[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/departments/part*
2,Fitness
3,Footwear
4,Apparel
5,Golf
6,Outdoors
7,Fan Shop
```

**sqoop import --connect jdbc:mysql://quickstart:3306/retail_db --password cloudera --username root --table departments --target-dir /user/cloudera/department1**


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 20, 09:33 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
tStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStre
am.java:894)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DF
SOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutpu
tStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStre
am.java:894)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DF
SOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutpu
tStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStre
am.java:894)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DF
SOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeInternal(DFS
OutputStream.java:935)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStre
am.java:931)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DF
SOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutpu
tStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStre
am.java:894)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
cloudera@quickstart:~
cloudera@quickstart:~
Right Ctrl


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 20, 09:34 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
am.java:931)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DF
SOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutpu
tStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStre
am.java:894)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DF
SOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutpu
tStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStre
am.java:894)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DF
SOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutpu
tStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStre
am.java:894)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DF
SOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutpu
tStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStre
am.java:894)
22/03/20 08:11:16 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
cloudera@quickstart:~
cloudera@quickstart:~
Right Ctrl


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 20, 09:34 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/03/20 08:11:15 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/03/20 08:11:16 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/03/20 08:11:16 INFO db.DBInputFormat: Using read commited transaction isolation
22/03/20 08:11:16 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN(`department_id`), MAX(`department_id`) FROM `departments`
22/03/20 08:11:16 INFO db.IntegerSplitter: Split size: 1; Num splits: 4 from: 2 to: 7
22/03/20 08:11:16 INFO mapreduce.JobSubmitter: number of splits:4
22/03/20 08:11:17 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1646550420897_0016
22/03/20 08:11:17 INFO impl.YarnClientImpl: Submitted application application_1646550420897_0016
22/03/20 08:11:17 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1646550420897_0016/
22/03/20 08:11:17 INFO mapreduce.Job: Running job: job_1646550420897_0016
22/03/20 08:11:31 INFO mapreduce.Job: Job job_1646550420897_0016 running in uber mode : false
22/03/20 08:11:31 INFO mapreduce.Job: map 0% reduce 0%
22/03/20 08:12:15 INFO mapreduce.Job: map 25% reduce 0%
22/03/20 08:12:16 INFO mapreduce.Job: map 50% reduce 0%
22/03/20 08:12:18 INFO mapreduce.Job: map 75% reduce 0%
22/03/20 08:12:19 INFO mapreduce.Job: map 100% reduce 0%
22/03/20 08:12:19 INFO mapreduce.Job: Job job_1646550420897_0016 completed successfully
cloudera@quickstart:~
cloudera@quickstart:~
Right Ctrl


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Mar 20, 09:34 cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
22/03/20 08:12:18 INFO mapreduce.Job: map 75% reduce 0%
22/03/20 08:12:19 INFO mapreduce.Job: map 100% reduce 0%
22/03/20 08:12:19 INFO mapreduce.Job: Job job_1646550420897_0016 completed successfully
22/03/20 08:12:19 INFO mapreduce.Job: Counters: 31
File System Counters
FILE: Number of bytes read=0
FILE: Number of bytes written=685412
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=481
HDFS: Number of bytes written=60
HDFS: Number of read operations=16
HDFS: Number of large read operations=0
HDFS: Number of write operations=8
Job Counters
Killed map tasks=1
Launched map tasks=4
Other local map tasks=5
Total time spent by all maps in occupied slots (ms)=165135
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=165135
Total vcore-milliseconds taken by all map tasks=165135
Total megabyte-milliseconds taken by all map tasks=169098240
Map-Reduce Framework
Map input records=6
Map output records=6
Input split bytes=481
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=2448
CPU time spent (ms)=6780
Physical memory (bytes) snapshot=500707328
Virtual memory (bytes) snapshot=6040596480
Total committed heap usage (bytes)=243007488
File Input Format Counters
Bytes Read=0
File Output Format Counters
Bytes Written=60
22/03/20 08:12:19 INFO mapreduce.ImportJobBase: Transferred 60 bytes in 66.7391 seconds (0.899 bytes/sec)
22/03/20 08:12:19 INFO mapreduce.ImportJobBase: Retrieved 6 records.
[cloudera@quickstart ~]$ hadoop fs -ls
Found 2 items
drwxr-xr-x - cloudera cloudera 0 2022-03-20 08:12 department1
drwxr-xr-x - cloudera cloudera 0 2022-03-20 07:59 departments
[cloudera@quickstart ~]$
cloudera@quickstart:~
cloudera@quickstart:~
Right Ctrl

**As we can see 6 records are retrived successfully.**

**hadoop fs -ls**

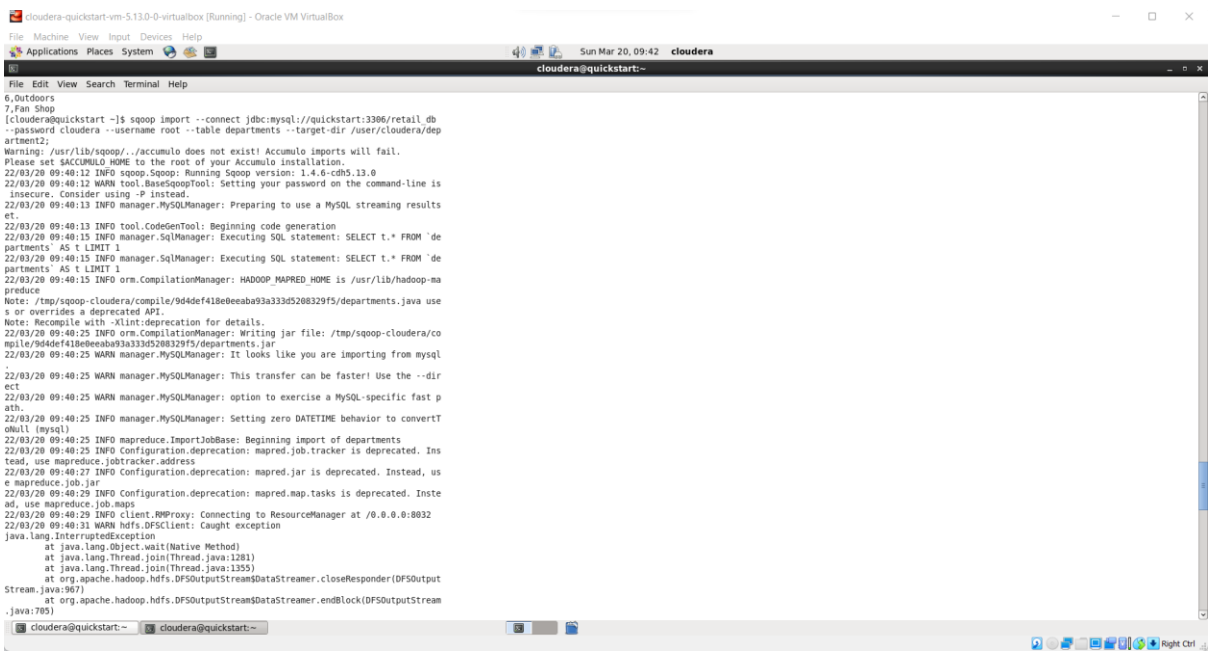
```
[cloudera@quickstart ~]$ hadoop fs -ls
Found 2 items
drwxr-xr-x   - cloudera cloudera          0 2022-03-20 08:12 department1
drwxr-xr-x   - cloudera cloudera          0 2022-03-20 07:59 departments
[cloudera@quickstart ~]$
```

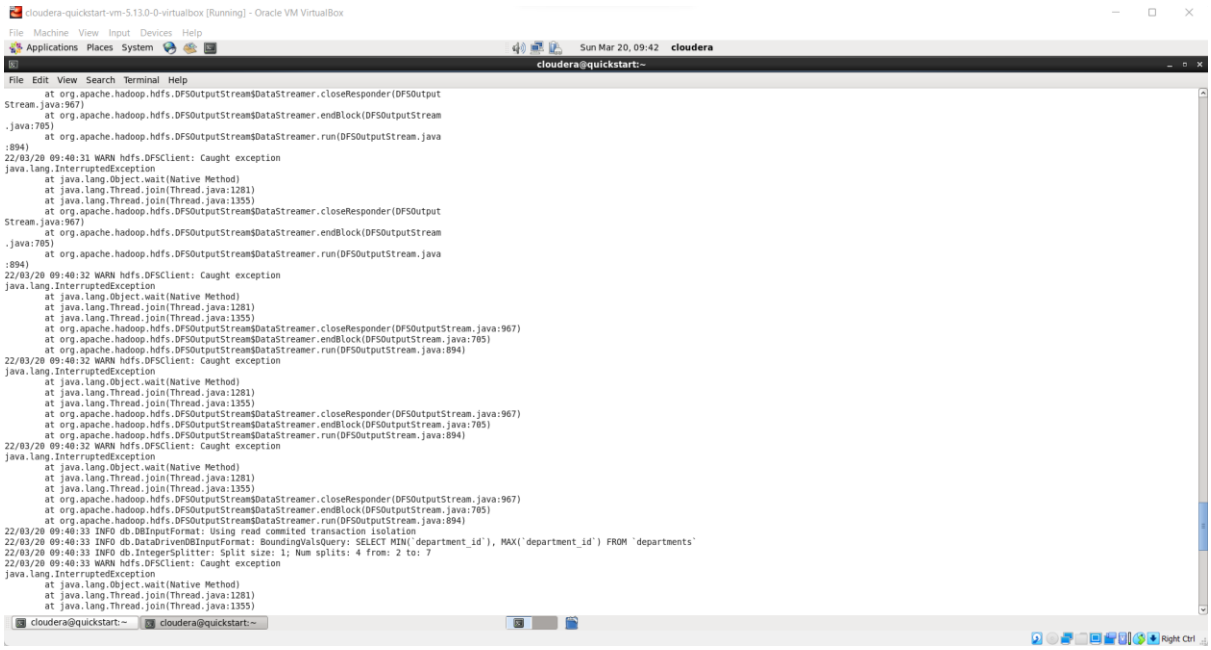
cloudera@quickstart:~ cloudera@quickstart:~

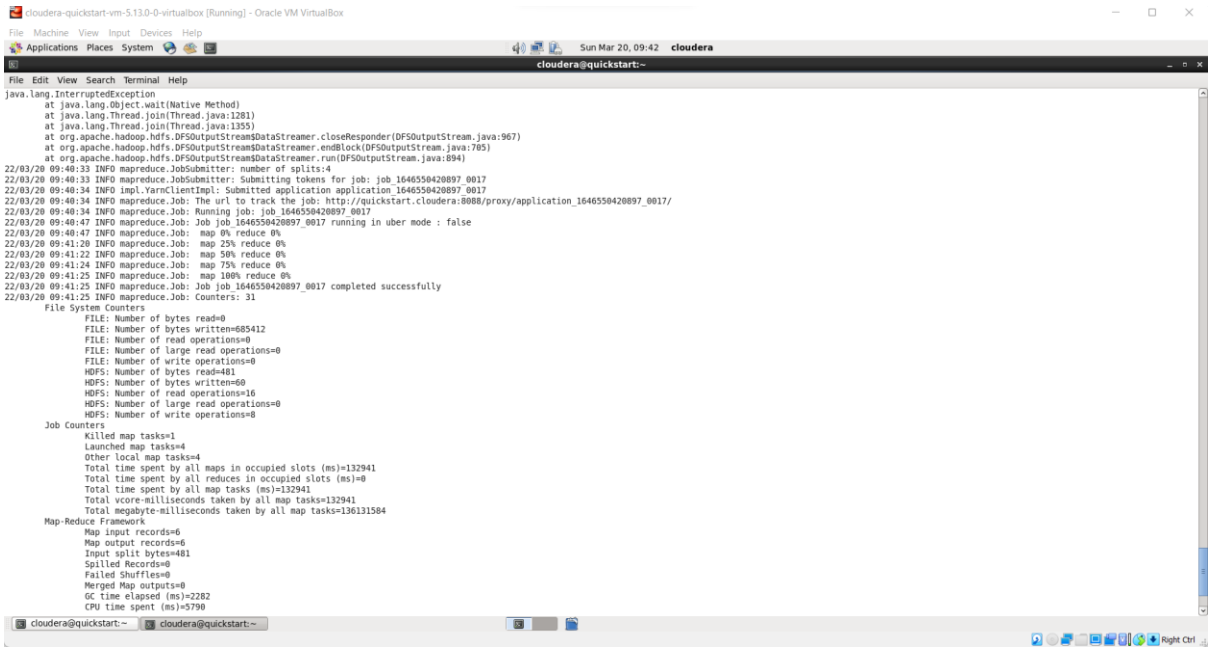
**hadoop fs -cat /user/cloudera/department1/part***

```
[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/department1/part*
2,Fitness
3,Footwear
4,Apparel
5,Golf
6,Outdoors
7,Fan Shop
[cloudera@quickstart ~]$
```

**sqoop import --connect jdbc:mysql://quickstart:3306/retail_db --password cloudera --username root --table departments --where "department_id>4" --target-dir/user/cloudera/department2;**






```
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=2282
                CPU time spent (ms)=5790
                Physical memory (bytes) snapshot=483442688
                Virtual memory (bytes) snapshot=6040580096
                Total committed heap usage (bytes)=243007488
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=60
22/03/20 09:41:26 INFO mapreduce.ImportJobBase: Transferred 60 bytes in 56.3145 seconds (1.0654 bytes/sec)
22/03/20 09:41:26 INFO mapreduce.ImportJobBase: Retrieved 6 records.
```

**As we can see 3 records are retrieved successfully**.

hadoop fs -ls /user/cloudera/department2

```
-ls/user/cloudera/department2: Unknown command
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/department2
Found 5 items
-rw-r--r--   1 cloudera cloudera          0 2022-03-20 09:41 /user/cloudera/department2
/_SUCCESS
-rw-r--r--   1 cloudera cloudera         21 2022-03-20 09:41 /user/cloudera/department2
/part-m-00000
-rw-r--r--   1 cloudera cloudera         10 2022-03-20 09:41 /user/cloudera/department2
/part-m-00001
-rw-r--r--   1 cloudera cloudera          7 2022-03-20 09:41 /user/cloudera/department2
/part-m-00002
-rw-r--r--   1 cloudera cloudera         22 2022-03-20 09:41 /user/cloudera/department2
/part-m-00003
[cloudera@quickstart ~]$
```

**hadoop fs -cat /user/cloudera/department2/part***

```
[cloudera@quickstart ~]$ hadoop fs -cat /user/cloudera/department2/part*
5,Golf
6,Outdoors
7,Fan Shop
[cloudera@quickstart ~]$
```

Now we will see the Export command. So what the export tool does is it will export the data from our hdfs to the RDBMS. So for that we need to have some table in mysql with some records so for that we will now move to mysql.

create table dpt(department_id int not null auto_increment, department_name varchar(50) not null, primary key(department_id));

```
mysql> create table dpt(department_id int not null auto_increment,department_name varchar(50) not null,primary key(department_id));
Query OK, 0 rows affected (0.05 sec)

mysql>
```

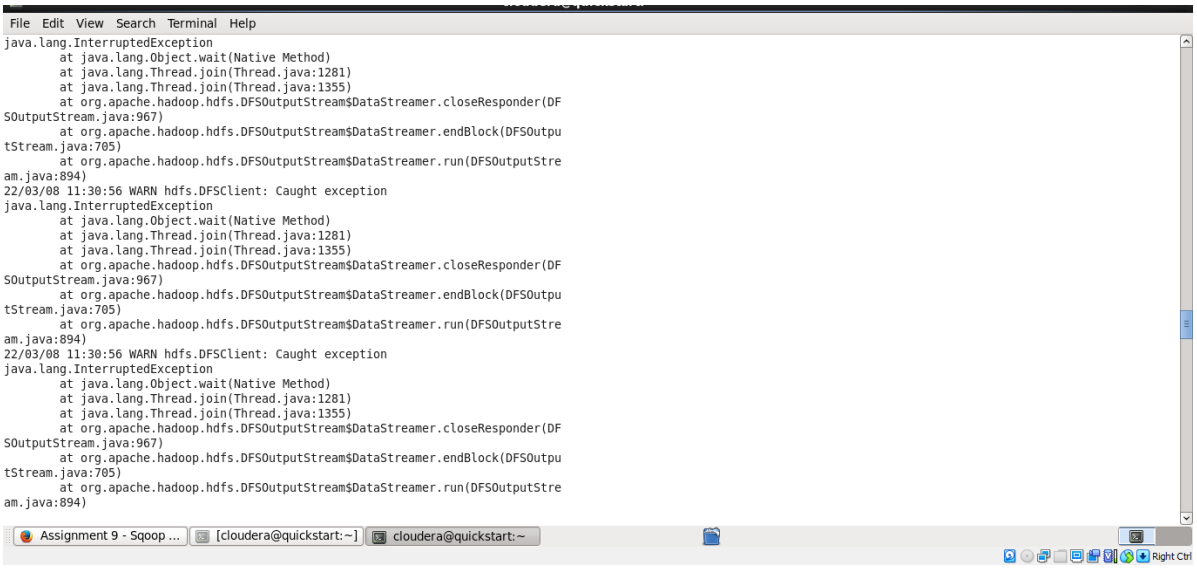
cloudera@quickstart:~ cloudera@quickstart:~

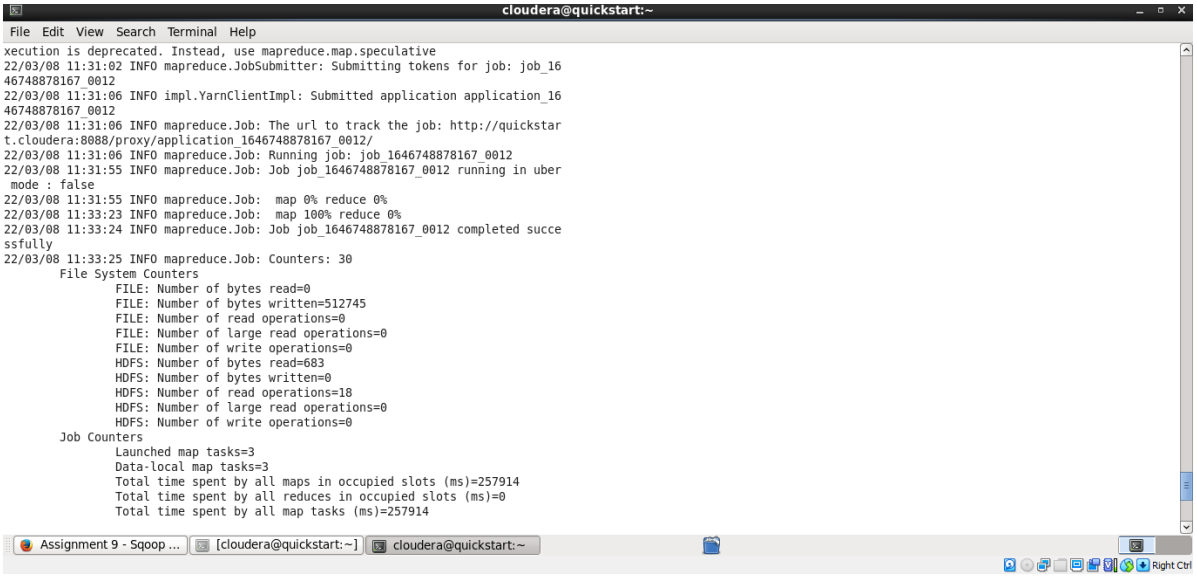
**Select * from dpt;**

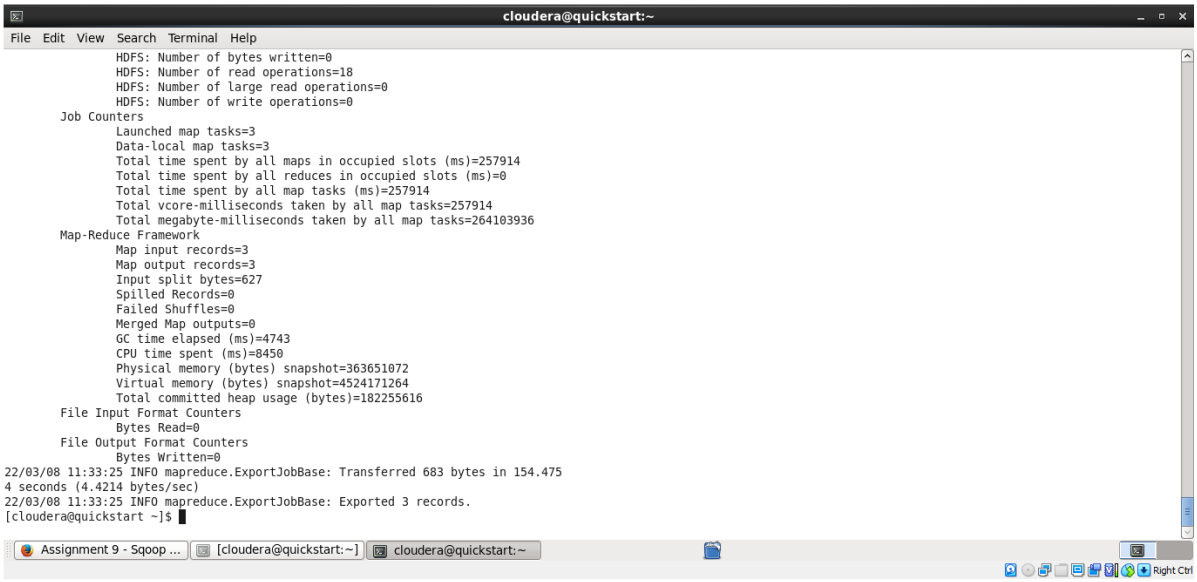
```
mysql> select * from dpt;
Empty set (0.01 sec)

mysql>
```

Now we will exporting the data from the hdfs to dpt table of mysql. Now we will move to the sqoop terminal.

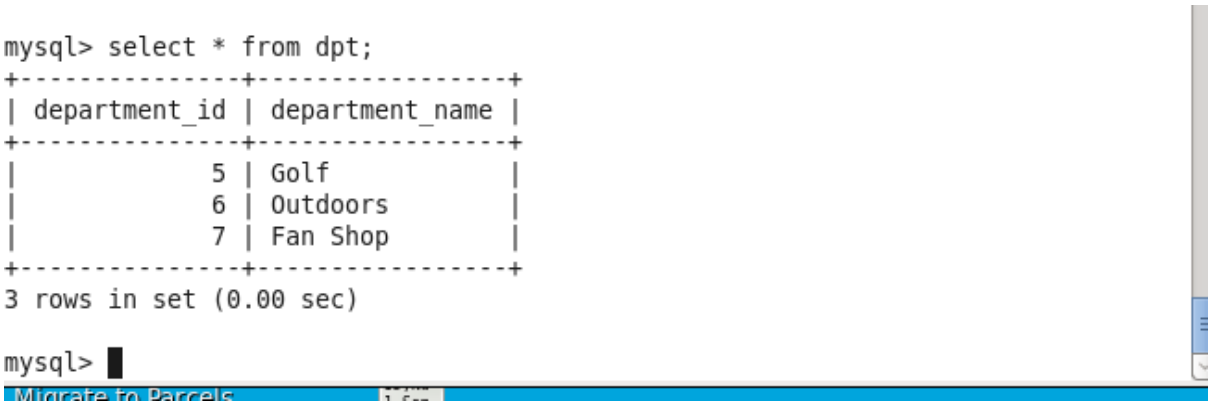
sqoop export --connect jdbc:mysql://quickstart:3306/retail_db --password cloudera --username root --table dpt --export-dir /user/cloudera/department2;






**Select * from dpt;**




As we can see these 3 records which are present in department2 table are successfully exported inside the dpt table of mysql.