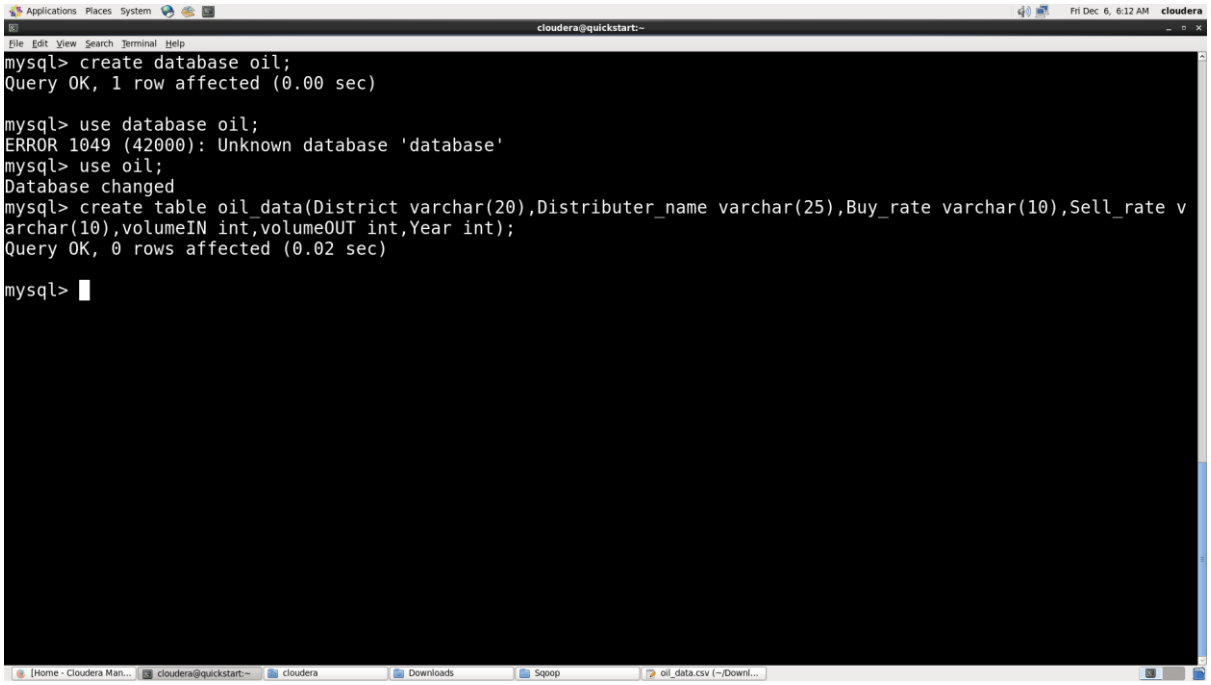
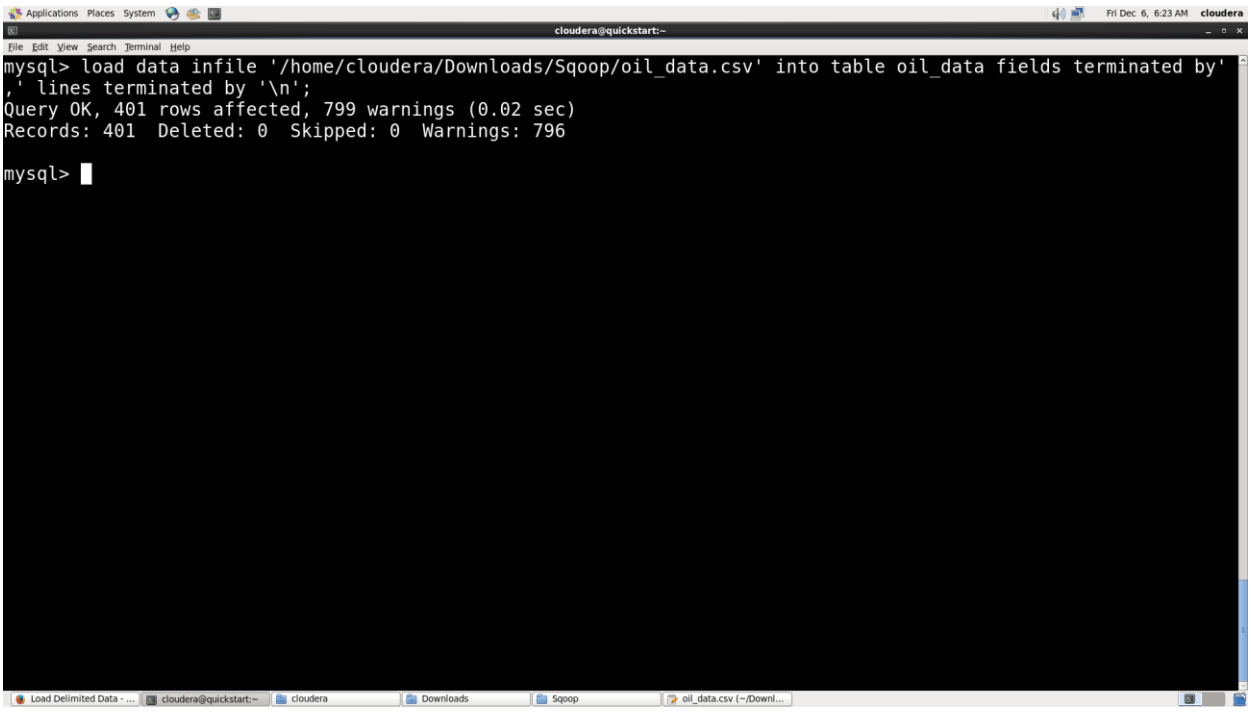
SQOOP

Name : Aniket Walse
Roll : 41

Lab 1.

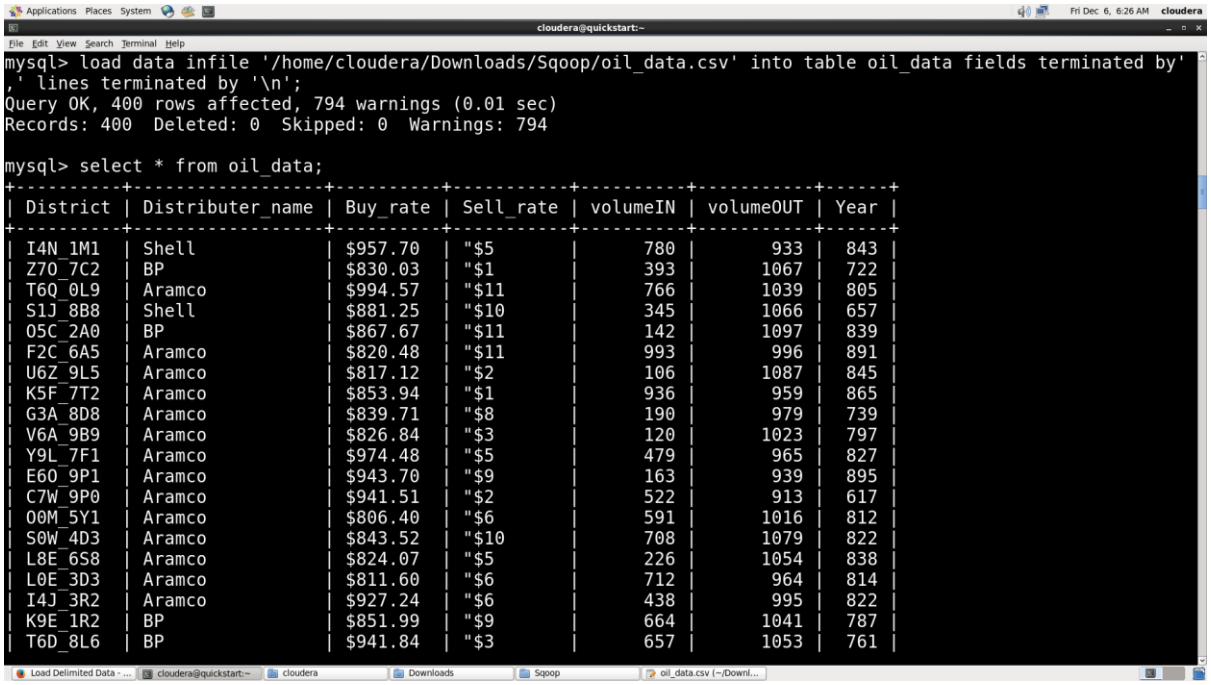
1. **Create database and table in Mysql according to oil dataset.**



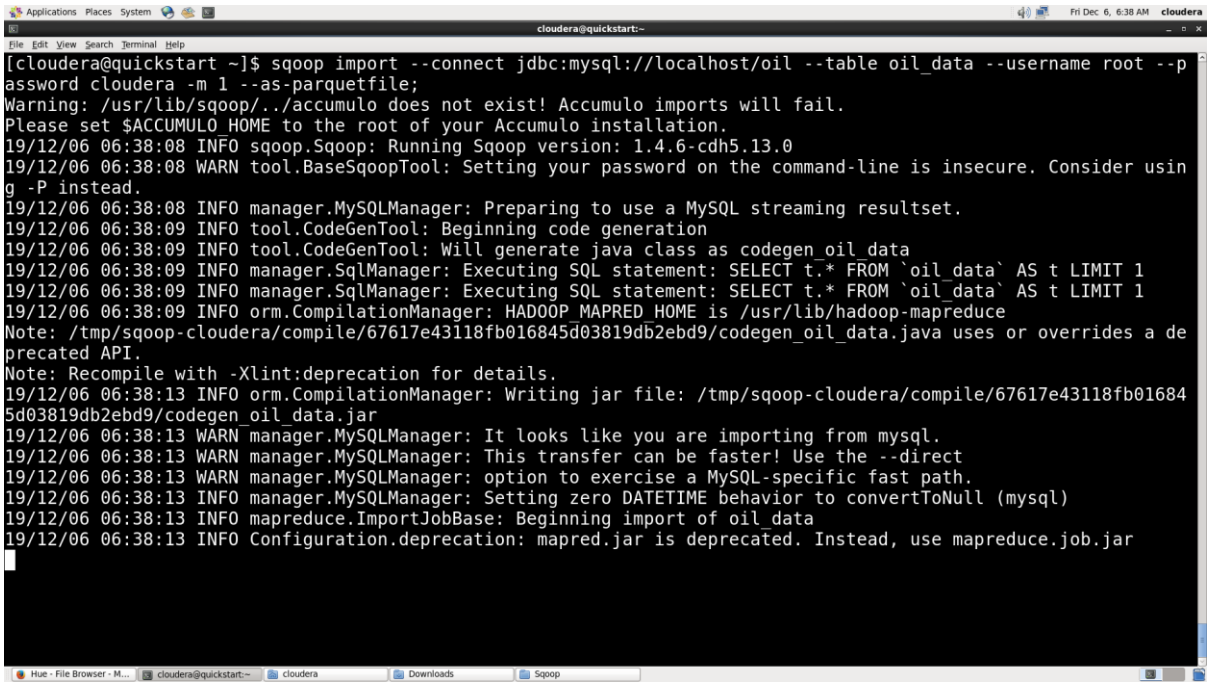


SQOOP

Name : Aniket Walse
Roll : 41



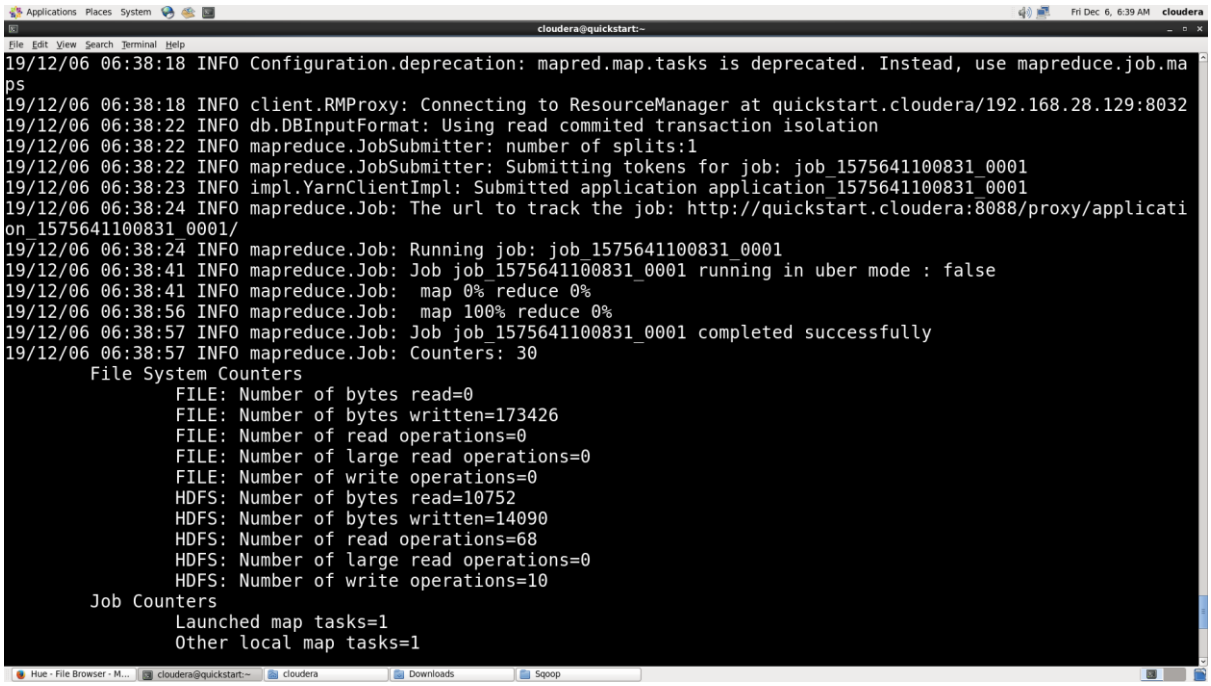
2. Sqoop import oil data into HDFS as parquet file.



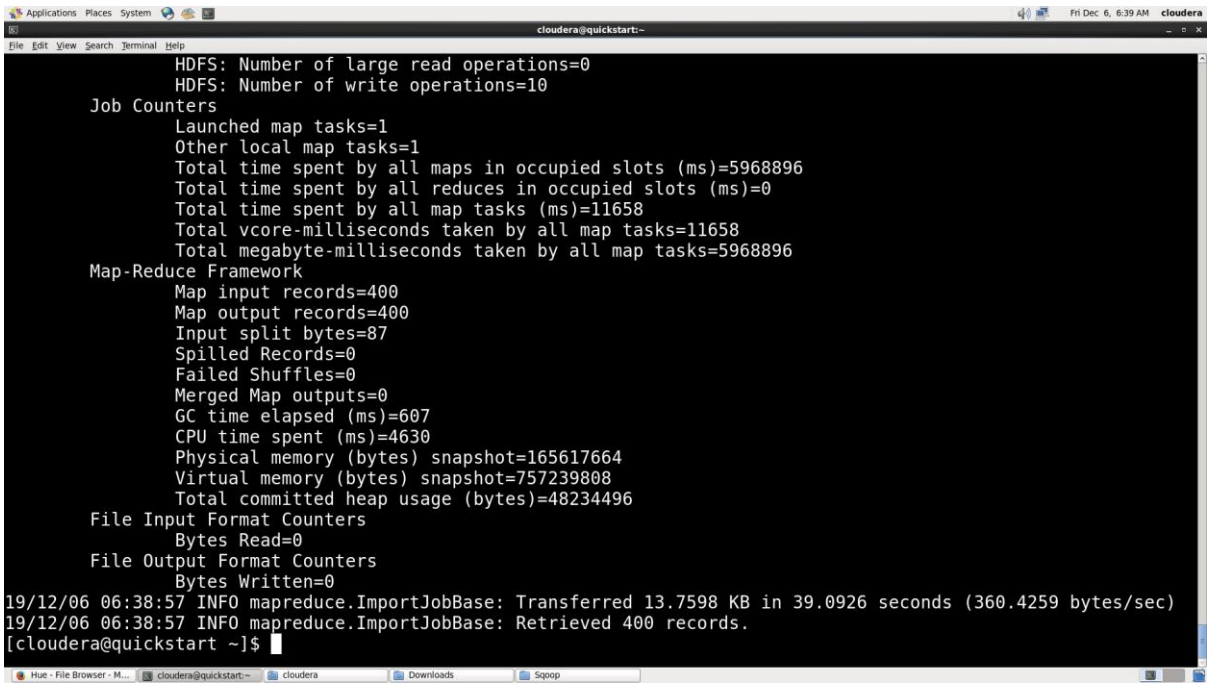
SQOOP

Name : Aniket Walse

Roll : 41



```
19/12/06 06:38:18 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.ma
ps
19/12/06 06:38:18 INFO client.RMProxy: Connecting to ResourceManager at quickstart.cloudera/192.168.28.129:8032
19/12/06 06:38:22 INFO db.DBInputFormat: Using read commited transaction isolation
19/12/06 06:38:22 INFO mapreduce.JobSubmitter: number of splits:1
19/12/06 06:38:22 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1575641100831_0001
19/12/06 06:38:23 INFO impl.YarnClientImpl: Submitted application application_1575641100831_0001
19/12/06 06:38:24 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/applicati
on_1575641100831_0001/
19/12/06 06:38:24 INFO mapreduce.Job: Running job: job_1575641100831_0001
19/12/06 06:38:41 INFO mapreduce.Job: Job job_1575641100831_0001 running in uber mode : false
19/12/06 06:38:41 INFO mapreduce.Job:  map 0% reduce 0%
19/12/06 06:38:56 INFO mapreduce.Job:  map 100% reduce 0%
19/12/06 06:38:57 INFO mapreduce.Job: Job job_1575641100831_0001 completed successfully
19/12/06 06:38:57 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=173426
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=10752
                HDFS: Number of bytes written=14090
                HDFS: Number of read operations=68
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=10
        Job Counters
                Launched map tasks=1
                Other local map tasks=1
```

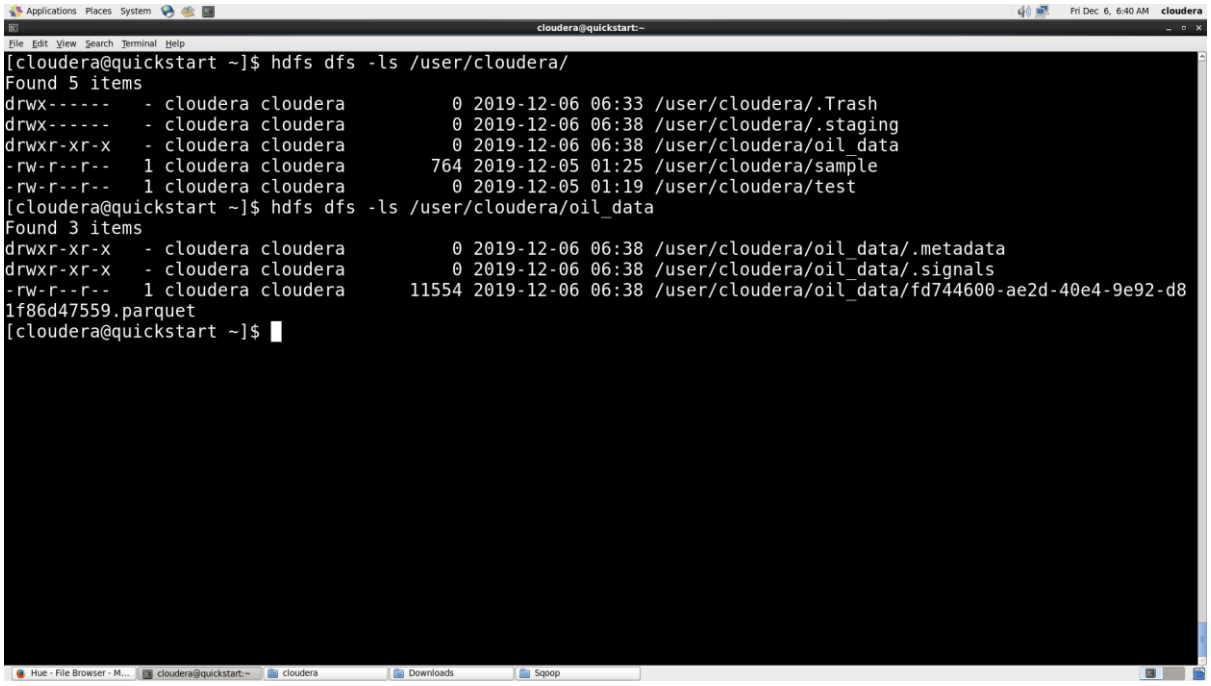


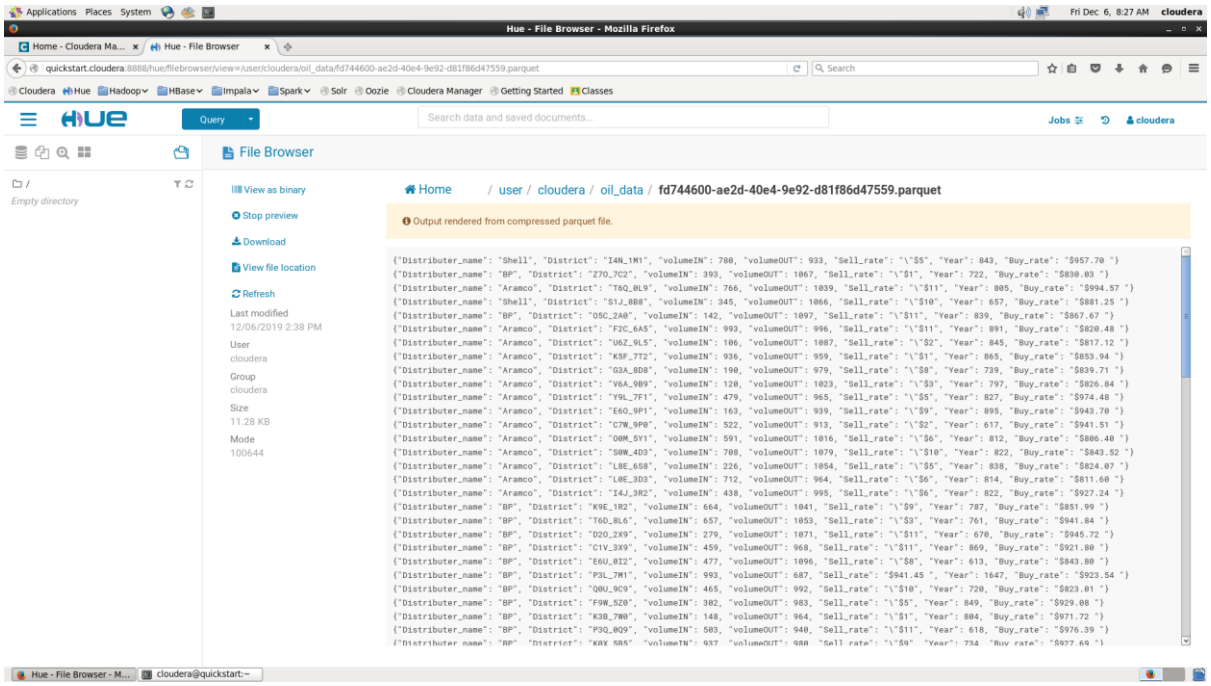
```
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=10
        Job Counters
                Launched map tasks=1
                Other local map tasks=1
                Total time spent by all maps in occupied slots (ms)=5968896
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=11658
                Total vcore-milliseconds taken by all map tasks=11658
                Total megabyte-milliseconds taken by all map tasks=5968896
        Map-Reduce Framework
                Map input records=400
                Map output records=400
                Input split bytes=87
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=607
                CPU time spent (ms)=4630
                Physical memory (bytes) snapshot=165617664
                Virtual memory (bytes) snapshot=757239808
                Total committed heap usage (bytes)=48234496
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
19/12/06 06:38:57 INFO mapreduce.ImportJobBase: Transferred 13.7598 KB in 39.0926 seconds (360.4259 bytes/sec)
19/12/06 06:38:57 INFO mapreduce.ImportJobBase: Retrieved 400 records.
[cloudera@quickstart ~]$
```

SQOOP

Name : Aniket Walse
Roll : 41

3. Verify parquet file is created.



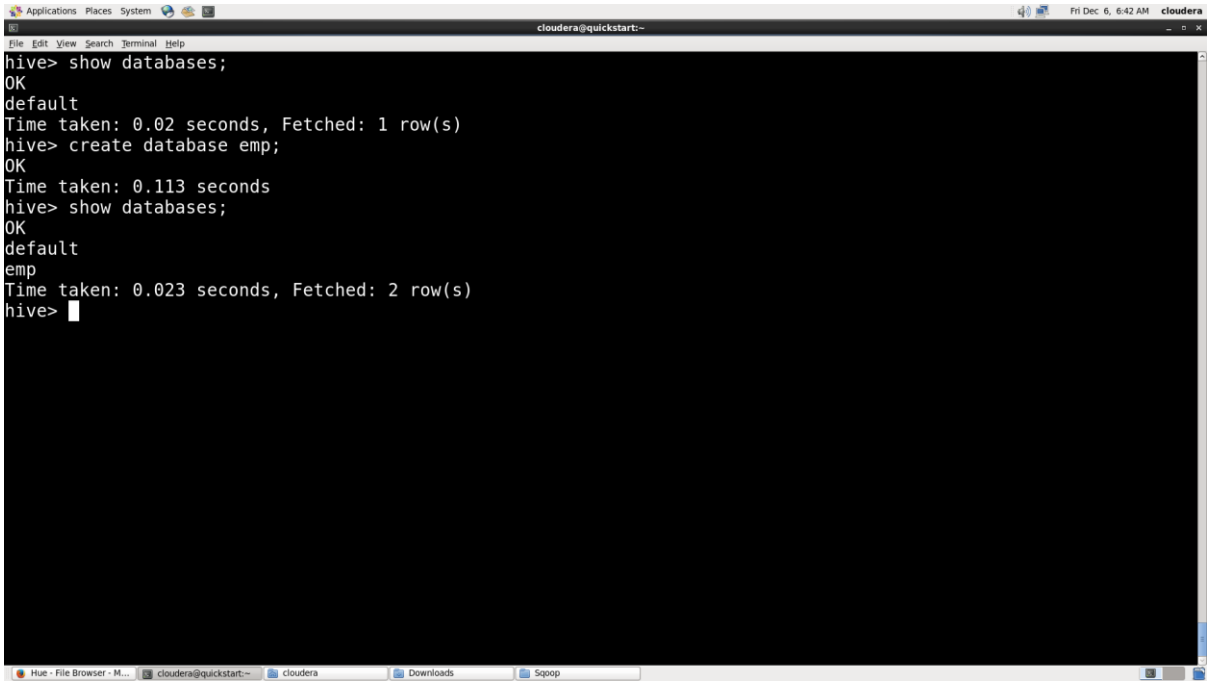


SQOOP

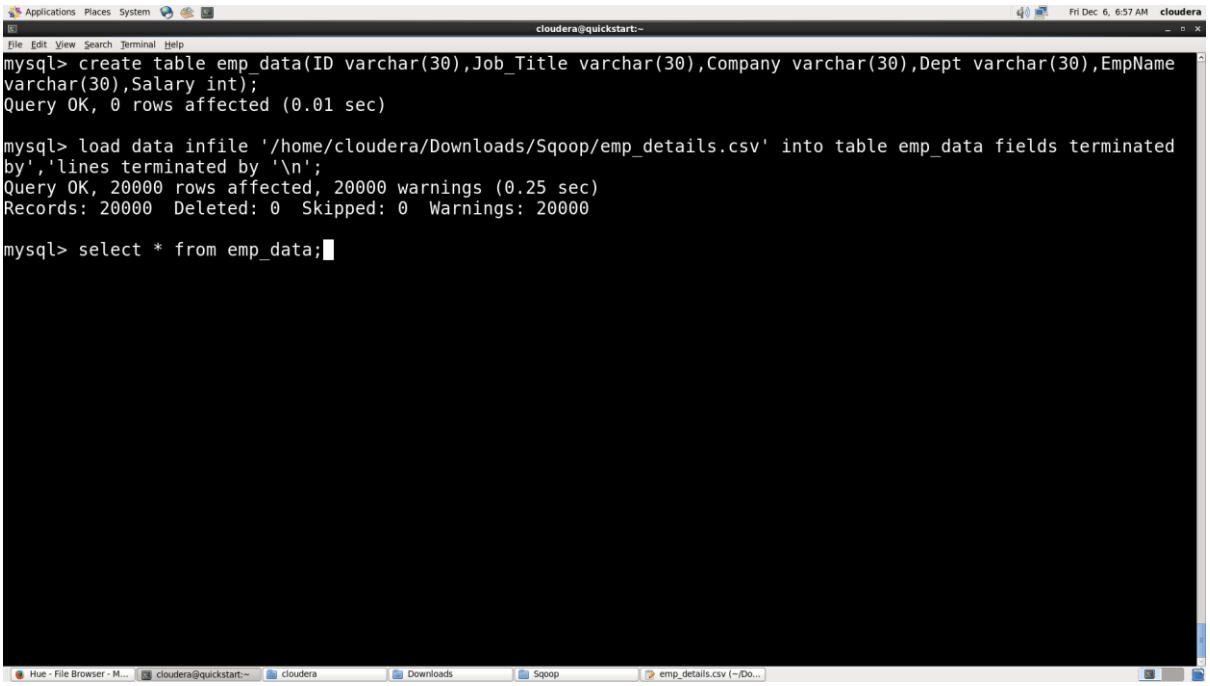
Name : Aniket Walse
Roll : 41

Lab 2.

1. **Create a database in hive for storing emp details.**

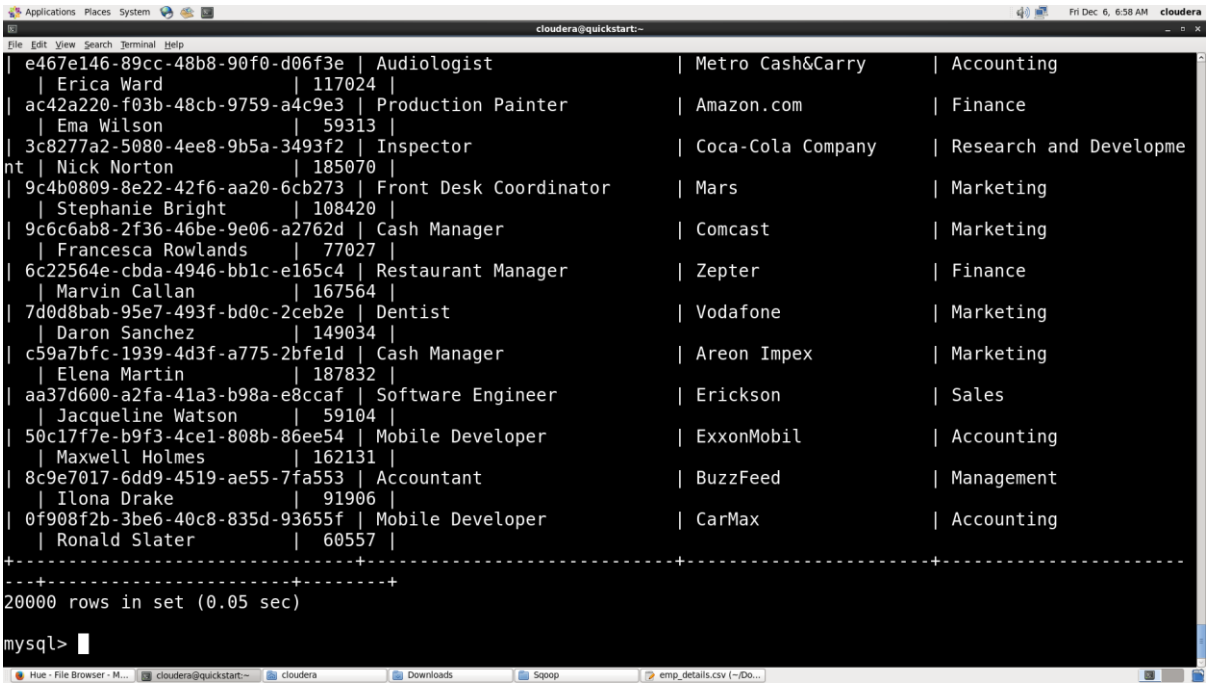


2. **Create a mysql database and table for storing emp_details dataset.**

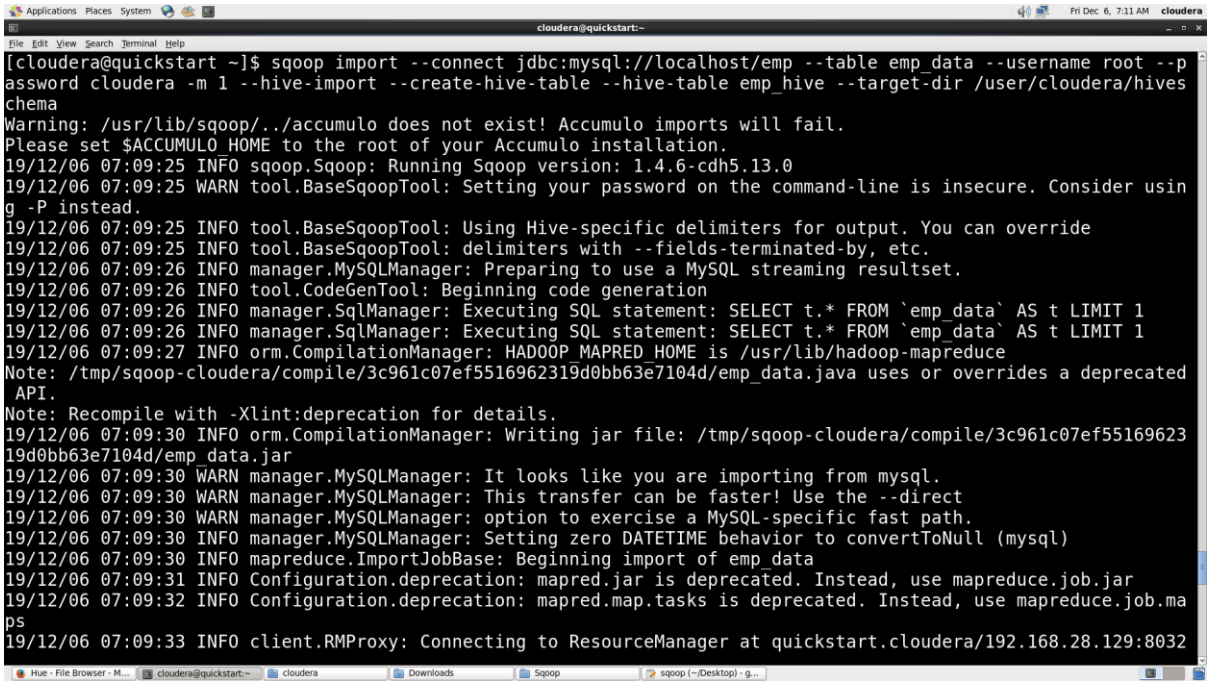


SQOOP

Name : Aniket Walse
Roll : 41

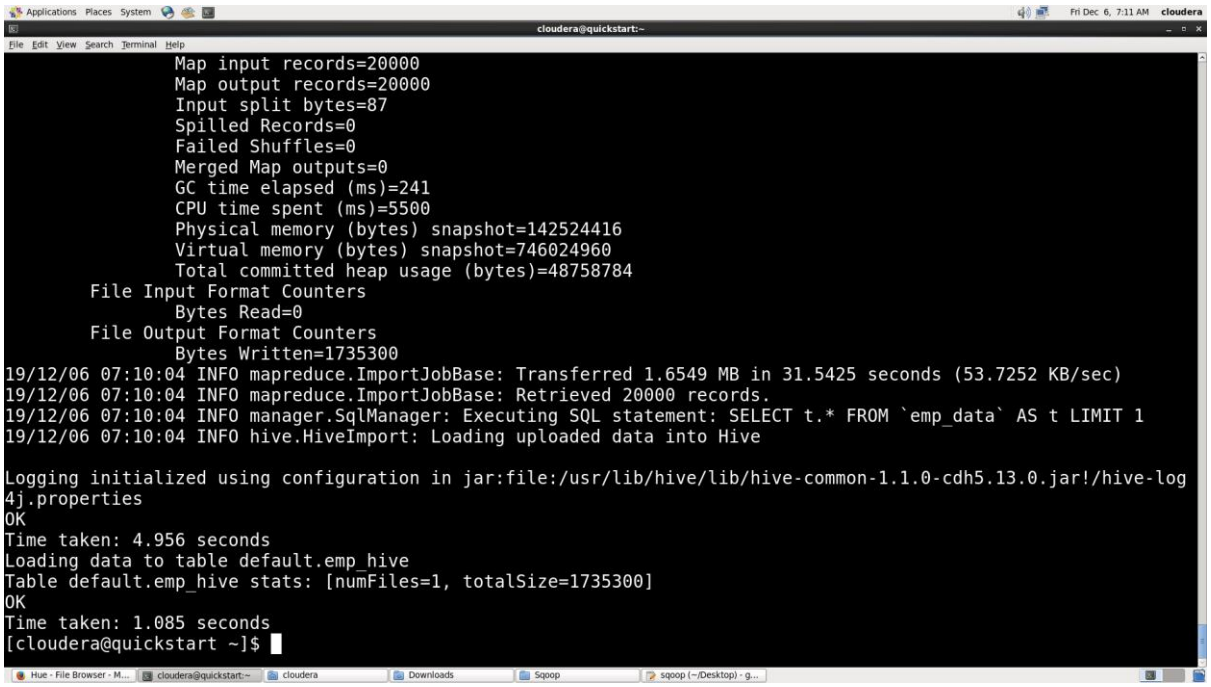


3. Sqoop import emp details data into hive table.

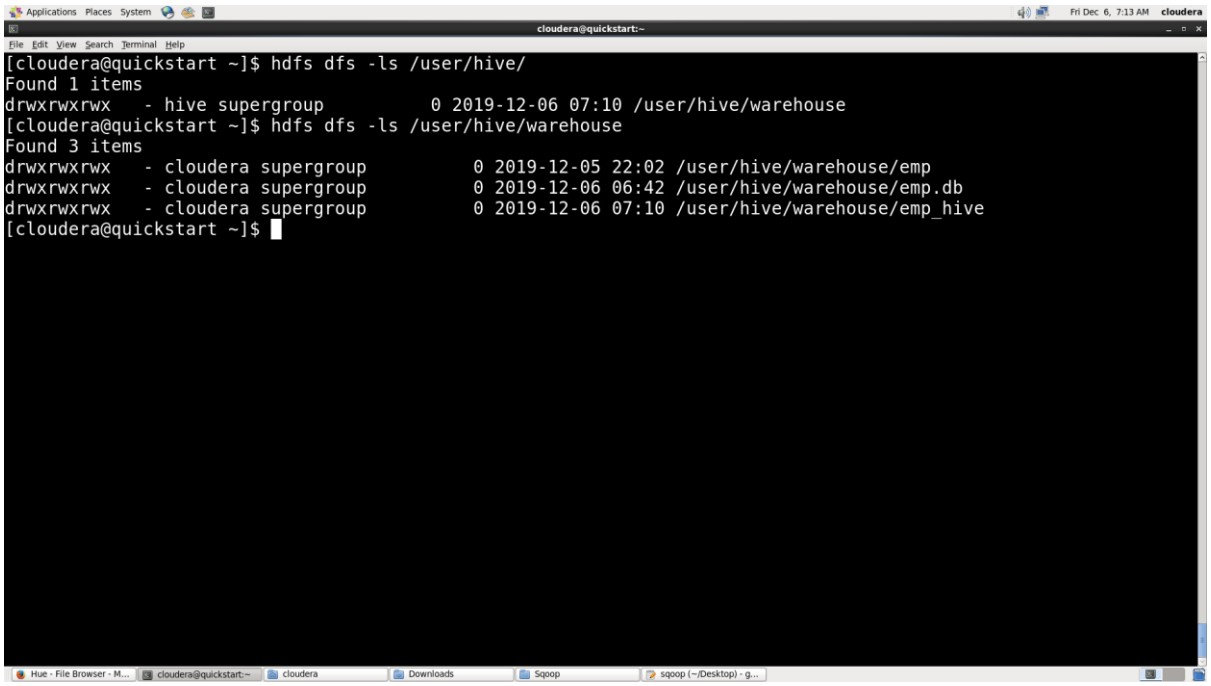


SQOOP

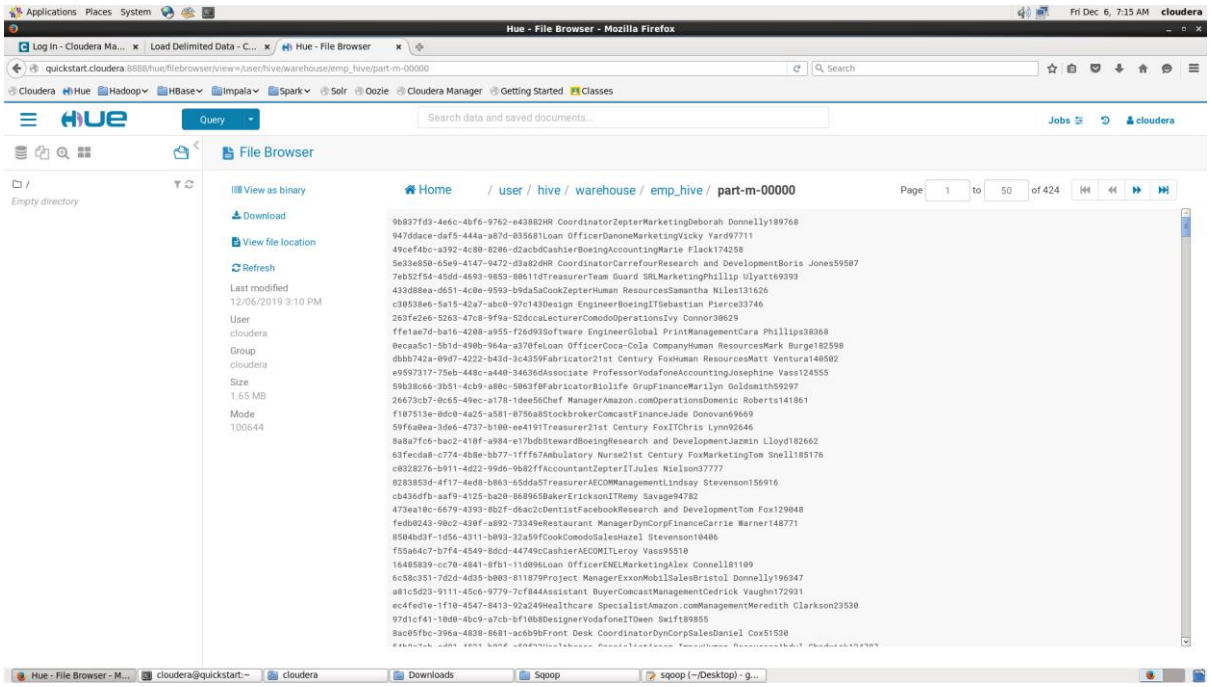
Name : Aniket Walse
Roll : 41



4. Verify if hive table is created and data has been loaded.



SQOOP

Name : Aniket Walse

Roll : 41