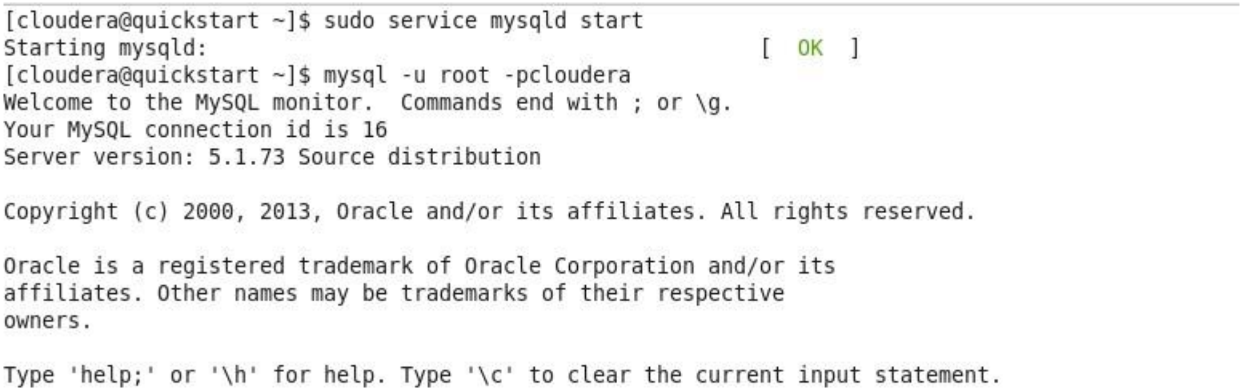
ICE-5

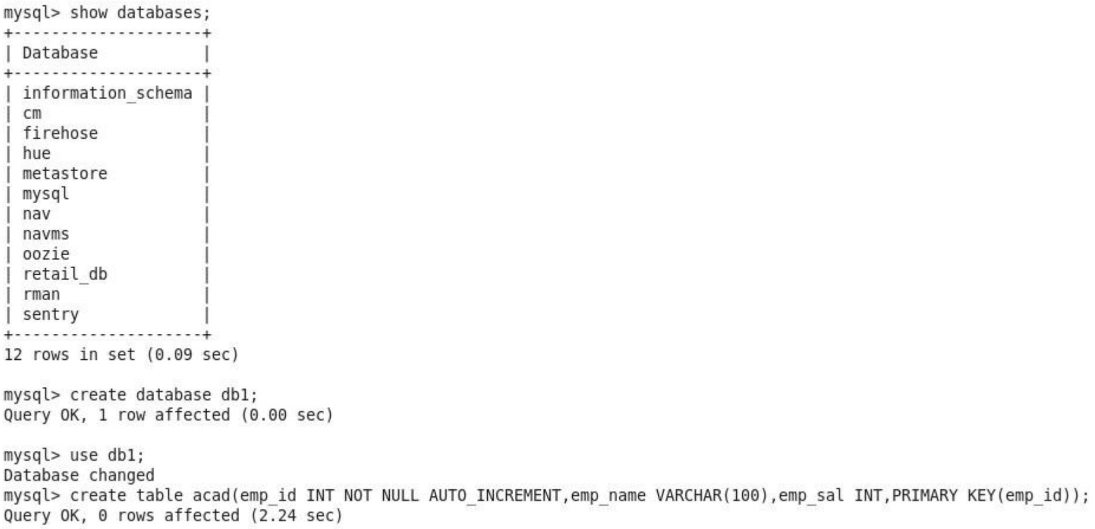
Step 1 : Starting and entering MySql



```
[cloudera@quickstart ~]$ sudo service mysqld start
Starting mysqld:                                           [  OK  ]
[cloudera@quickstart ~]$ mysql -u root -pcloudera
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 16
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.
```

- Firstly we begin the MySql service using sudo service command.
- Then we enter MySql shell using username and password.

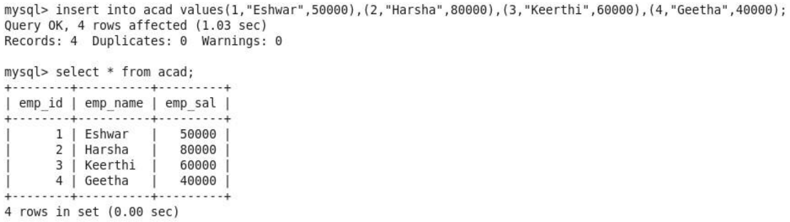
Step 2: Display data and creating database



```
mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| cm                 |
| firehose           |
| hue                |
| metastore          |
| mysql              |
| nav                |
| navms              |
| oozie              |
| retail_db          |
| rman               |
| sentry             |
+--------------------+
12 rows in set (0.09 sec)

mysql> create database db1;
Query OK, 1 row affected (0.00 sec)

mysql> use db1;
Database changed
mysql> create table acad(emp_id INT NOT NULL AUTO_INCREMENT,emp_name VARCHAR(100),emp_sal INT,PRIMARY KEY(emp_id));
Query OK, 0 rows affected (2.24 sec)
```

- Then we display the available databases using show command.
- Then we create database db1 using create command and then use it.
- Then we create table name acad with emp_id, emp_name and emp_sal as columns and emp_id being the primary key.

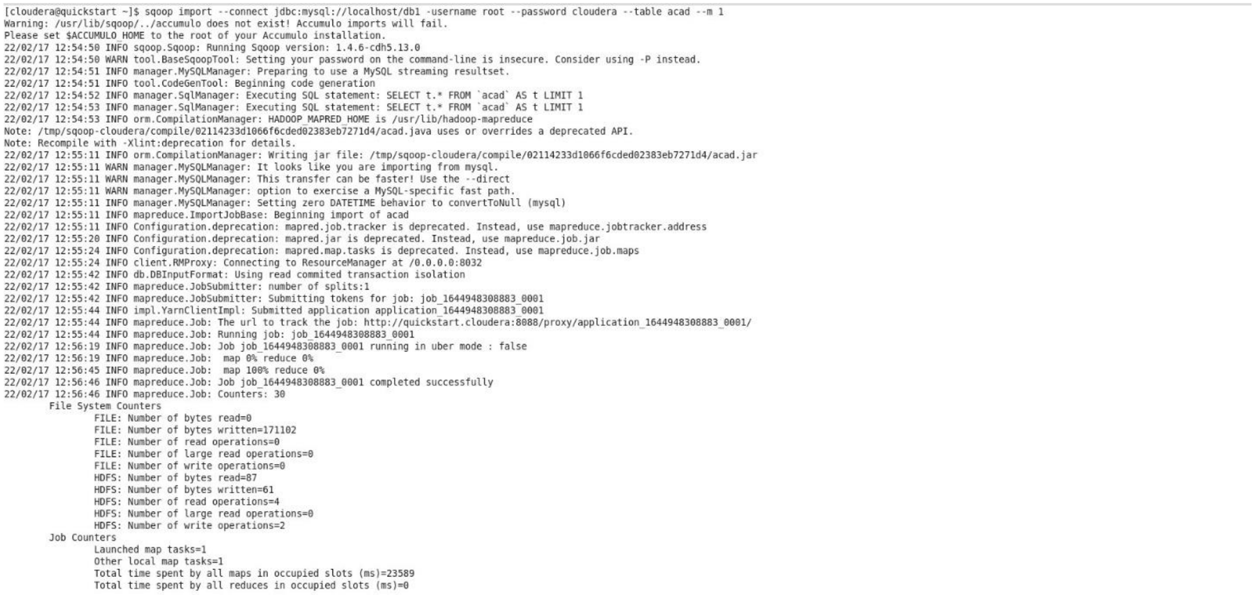
Step 3 : Create and insert values into table



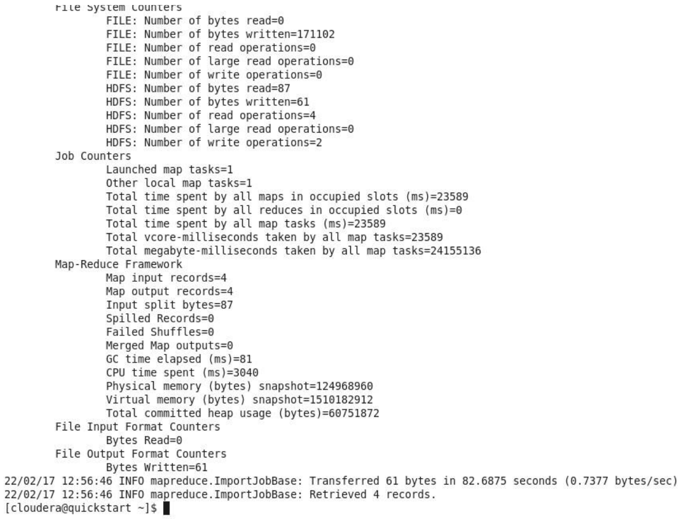
```
mysql> insert into acad values(1,"Eshwar",50000),(2,"Harsha",80000),(3,"Keerthi",60000),(4,"Geetha",40000);
Query OK, 4 rows affected (1.03 sec)
Records: 4  Duplicates: 0  Warnings: 0

mysql> select * from acad;
+--------+----------+---------+
| emp_id | emp_name | emp_sal |
+--------+----------+---------+
|      1 | Eshwar   |   50000 |
|      2 | Harsha   |   80000 |
|      3 | Keerthi  |   60000 |
|      4 | Geetha   |   40000 |
+--------+----------+---------+
4 rows in set (0.00 sec)
```

- Then we insert values into table acad using insert command,
- Then we display the table using select command.

Step 4 : Importing table from MySql to hadoop



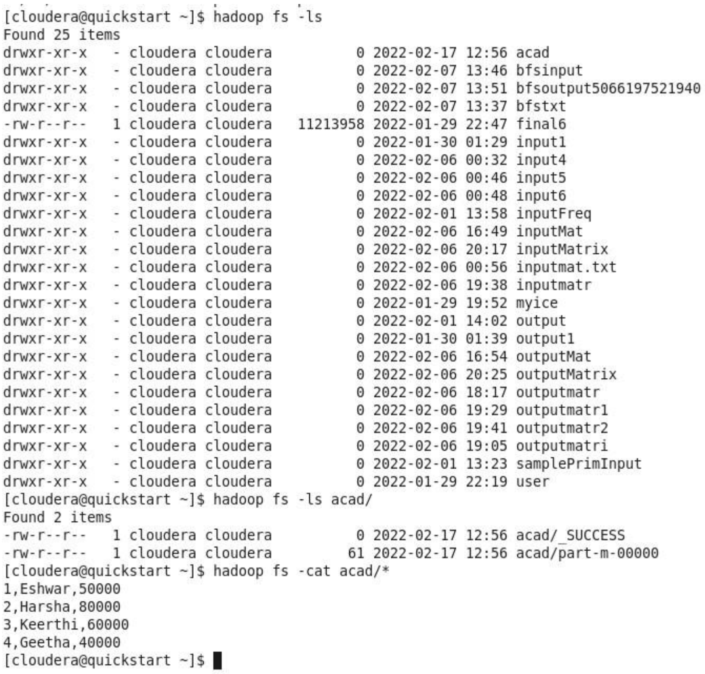
```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/db1 -username root --password cloudera --table acad --m 1
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
22/02/17 12:54:50 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
22/02/17 12:54:50 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
22/02/17 12:54:51 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
22/02/17 12:54:51 INFO tool.CodeGenTool: Beginning code generation
22/02/17 12:54:52 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `acad` AS t LIMIT 1
22/02/17 12:54:53 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `acad` AS t LIMIT 1
22/02/17 12:54:53 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/02114233d1066f6cded02383eb7271d4/acad.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
22/02/17 12:55:11 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/02114233d1066f6cded02383eb7271d4/acad.jar
22/02/17 12:55:11 WARN manager.MySQLManager: It looks like you are importing from mysql.
22/02/17 12:55:11 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
22/02/17 12:55:11 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
22/02/17 12:55:11 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
22/02/17 12:55:11 INFO mapreduce.ImportJobBase: Beginning import of acad
22/02/17 12:55:11 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
22/02/17 12:55:20 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
22/02/17 12:55:24 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
22/02/17 12:55:24 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/02/17 12:55:42 INFO db.DBInputFormat: Using read commited transaction isolation
22/02/17 12:55:42 INFO mapreduce.JobSubmitter: number of splits:1
22/02/17 12:55:42 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1644948308883_0001
22/02/17 12:55:44 INFO impl.YarnClientImpl: Submitted application application_1644948308883_0001
22/02/17 12:55:44 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1644948308883_0001/
22/02/17 12:55:44 INFO mapreduce.Job: Running job: job_1644948308883_0001
22/02/17 12:56:19 INFO mapreduce.Job: Job job_1644948308883_0001 running in uber mode : false
22/02/17 12:56:19 INFO mapreduce.Job:  map 0% reduce 0%
22/02/17 12:56:45 INFO mapreduce.Job:  map 100% reduce 0%
22/02/17 12:56:46 INFO mapreduce.Job: Job job_1644948308883_0001 completed successfully
22/02/17 12:56:46 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=171102
                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=61
                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)=23589
                Total time spent by all reduces in occupied slots (ms)=0
```



```
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=171102
                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=61
                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)=23589
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=23589
                Total vcore-milliseconds taken by all map tasks=23589
                Total megabyte-milliseconds taken by all map tasks=24155136
        Map-Reduce Framework
                Map input records=4
                Map output records=4
                Input split bytes=87
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=81
                CPU time spent (ms)=3040
                Physical memory (bytes) snapshot=124968960
                Virtual memory (bytes) snapshot=1510182912
                Total committed heap usage (bytes)=60751872
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=61
22/02/17 12:56:46 INFO mapreduce.ImportJobBase: Transferred 61 bytes in 82.6875 seconds (0.7377 bytes/sec)
22/02/17 12:56:46 INFO mapreduce.ImportJobBase: Retrieved 4 records.
[cloudera@quickstart ~]$
```

- Then we import the table which is in the MySql to Hadoop using sqoop import command by entering username and password and specifying mapper as 1.

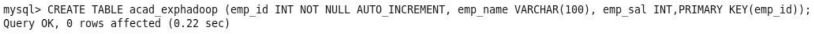
Step 5 : Display the files



```
[cloudera@quickstart ~]$ hadoop fs -ls
Found 25 items
drwxr-xr-x   - cloudera cloudera          0 2022-02-17 12:56 acad
drwxr-xr-x   - cloudera cloudera          0 2022-02-07 13:46 bfsinput
drwxr-xr-x   - cloudera cloudera          0 2022-02-07 13:51 bfsoutput5066197521940
drwxr-xr-x   - cloudera cloudera          0 2022-02-07 13:37 bfstxt
-rw-r--r--   1 cloudera cloudera   11213958 2022-01-29 22:47 final6
drwxr-xr-x   - cloudera cloudera          0 2022-01-30 01:29 input1
drwxr-xr-x   - cloudera cloudera          0 2022-02-06 00:32 input4
drwxr-xr-x   - cloudera cloudera          0 2022-02-06 00:46 input5
drwxr-xr-x   - cloudera cloudera          0 2022-02-06 00:48 input6
drwxr-xr-x   - cloudera cloudera          0 2022-02-01 13:58 inputFreq
drwxr-xr-x   - cloudera cloudera          0 2022-02-06 16:49 inputMat
drwxr-xr-x   - cloudera cloudera          0 2022-02-06 20:17 inputMatrix
drwxr-xr-x   - cloudera cloudera          0 2022-02-06 00:56 inputmat.txt
drwxr-xr-x   - cloudera cloudera          0 2022-02-06 19:38 inputmatr
drwxr-xr-x   - cloudera cloudera          0 2022-01-29 19:52 myice
drwxr-xr-x   - cloudera cloudera          0 2022-02-01 14:02 output
drwxr-xr-x   - cloudera cloudera          0 2022-01-30 01:39 output1
drwxr-xr-x   - cloudera cloudera          0 2022-02-06 16:54 outputMat
drwxr-xr-x   - cloudera cloudera          0 2022-02-06 20:25 outputMatrix
drwxr-xr-x   - cloudera cloudera          0 2022-02-06 18:17 outputmatr
drwxr-xr-x   - cloudera cloudera          0 2022-02-06 19:29 outputmatr1
drwxr-xr-x   - cloudera cloudera          0 2022-02-06 19:41 outputmatr2
drwxr-xr-x   - cloudera cloudera          0 2022-02-06 19:05 outputmatri
drwxr-xr-x   - cloudera cloudera          0 2022-02-01 13:23 samplePrimInput
drwxr-xr-x   - cloudera cloudera          0 2022-01-29 22:19 user
[cloudera@quickstart ~]$ hadoop fs -ls acad/
Found 2 items
-rw-r--r--   1 cloudera cloudera          0 2022-02-17 12:56 acad/_SUCCESS
-rw-r--r--   1 cloudera cloudera         61 2022-02-17 12:56 acad/part-m-00000
[cloudera@quickstart ~]$ hadoop fs -cat acad/*
1,Eshwar,50000
2,Harsha,80000
3,Keerthi,60000
4,Geetha,40000
[cloudera@quickstart ~]$ █
```

- Then we check the files using -ls command to find the imported file.

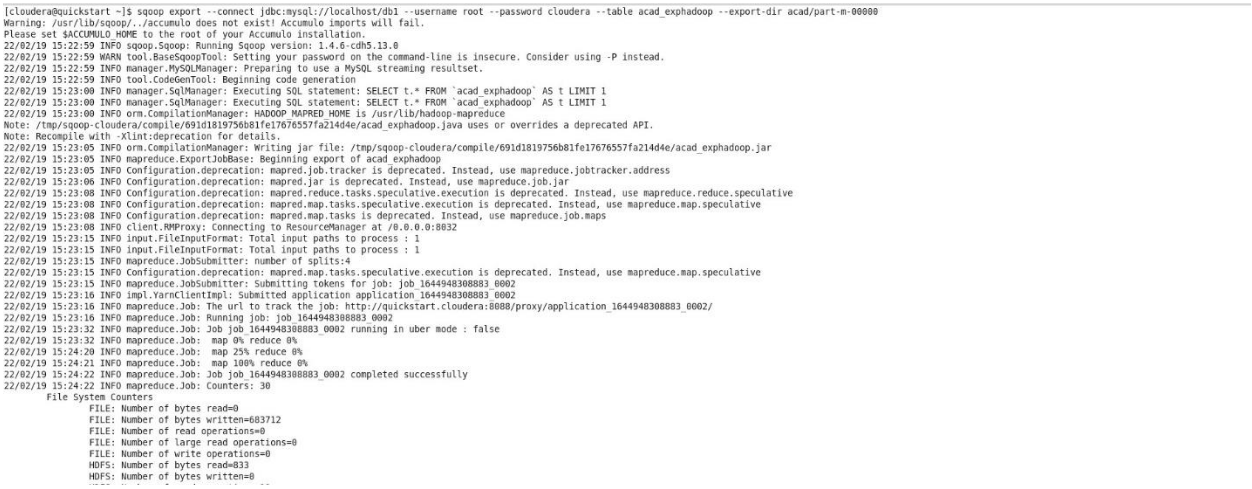
Step 6 : Create table in MySql



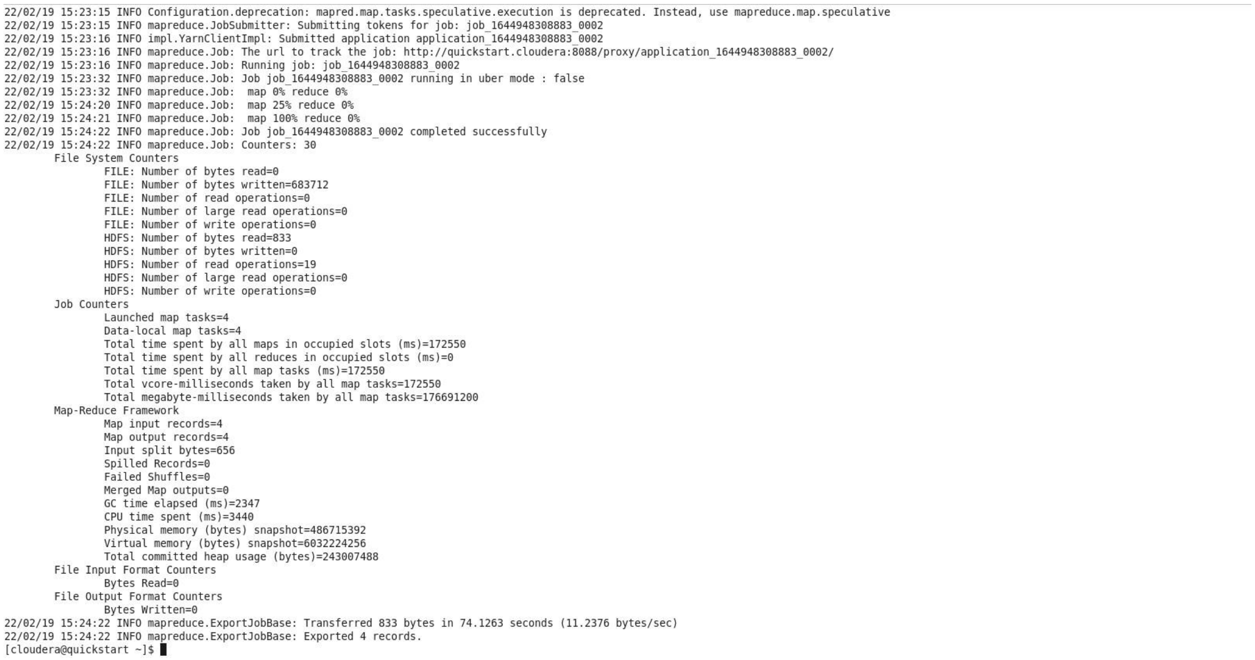
```
mysql> CREATE TABLE acad_exphadoop (emp_id INT NOT NULL AUTO_INCREMENT, emp_name VARCHAR(100), emp_sal INT,PRIMARY KEY(emp_id));
Query OK, 0 rows affected (0.22 sec)
```

- Then we created table in MySql to which the Hadoop table will be exported.

Step 7 : Exporting table to MySql

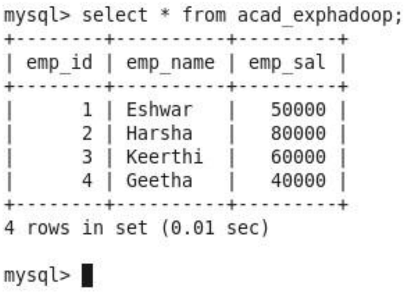


```
[cloudera@quickstart ~]$ sqoop export --connect jdbc:mysql://localhost/db1 --username root --password cloudera --table acad_exphadoop --export-dir acad/part-m-00000
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
22/02/19 15:22:59 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
22/02/19 15:22:59 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
22/02/19 15:22:59 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
22/02/19 15:22:59 INFO tool.CodeGenTool: Beginning code generation
22/02/19 15:23:00 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `acad_exphadoop` AS t LIMIT 1
22/02/19 15:23:00 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `acad_exphadoop` AS t LIMIT 1
22/02/19 15:23:00 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/691d1819756b81fe17676557fa214d4e/acad_exphadoop.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
22/02/19 15:23:05 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/691d1819756b81fe17676557fa214d4e/acad_exphadoop.jar
22/02/19 15:23:05 INFO mapreduce.ExportJobBase: Beginning export of acad_exphadoop
22/02/19 15:23:05 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
22/02/19 15:23:06 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
22/02/19 15:23:08 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
22/02/19 15:23:08 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
22/02/19 15:23:08 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
22/02/19 15:23:08 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/02/19 15:23:15 INFO input.FileInputFormat: Total input paths to process : 1
22/02/19 15:23:15 INFO input.FileInputFormat: Total input paths to process : 1
22/02/19 15:23:15 INFO mapreduce.JobSubmitter: number of splits:4
22/02/19 15:23:15 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
22/02/19 15:23:15 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1644948308883_0002
22/02/19 15:23:16 INFO impl.YarnClientImpl: Submitted application application_1644948308883_0002
22/02/19 15:23:16 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1644948308883_0002/
22/02/19 15:23:16 INFO mapreduce.Job: Running job: job_1644948308883_0002
22/02/19 15:23:32 INFO mapreduce.Job: Job job_1644948308883_0002 running in uber mode : false
22/02/19 15:23:32 INFO mapreduce.Job:  map 0% reduce 0%
22/02/19 15:24:20 INFO mapreduce.Job:  map 25% reduce 0%
22/02/19 15:24:21 INFO mapreduce.Job:  map 100% reduce 0%
22/02/19 15:24:22 INFO mapreduce.Job: Job job_1644948308883_0002 completed successfully
22/02/19 15:24:22 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=683712
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=833
                HDFS: Number of bytes written=0
```



```
22/02/19 15:23:15 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
22/02/19 15:23:15 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1644948308883_0002
22/02/19 15:23:16 INFO impl.YarnClientImpl: Submitted application application_1644948308883_0002
22/02/19 15:23:16 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1644948308883_0002/
22/02/19 15:23:16 INFO mapreduce.Job: Running job: job_1644948308883_0002
22/02/19 15:23:32 INFO mapreduce.Job: Job job_1644948308883_0002 running in uber mode : false
22/02/19 15:23:32 INFO mapreduce.Job:  map 0% reduce 0%
22/02/19 15:24:20 INFO mapreduce.Job:  map 25% reduce 0%
22/02/19 15:24:21 INFO mapreduce.Job:  map 100% reduce 0%
22/02/19 15:24:22 INFO mapreduce.Job: Job job_1644948308883_0002 completed successfully
22/02/19 15:24:22 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=683712
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=833
                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)=172550
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=172550
                Total vcore-milliseconds taken by all map tasks=172550
                Total megabyte-milliseconds taken by all map tasks=176691200
        Map-Reduce Framework
                Map input records=4
                Map output records=4
                Input split bytes=656
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=2347
                CPU time spent (ms)=3440
                Physical memory (bytes) snapshot=486715392
                Virtual memory (bytes) snapshot=6032224256
                Total committed heap usage (bytes)=243007488
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
22/02/19 15:24:22 INFO mapreduce.ExportJobBase: Transferred 833 bytes in 74.1263 seconds (11.2376 bytes/sec)
22/02/19 15:24:22 INFO mapreduce.ExportJobBase: Exported 4 records.
[cloudera@quickstart ~]$ █
```

- Then we export the table from Hadoop using sqoop export command.
- In this command we specify the path of Hadoop file and username and password to access the localhost and we specify the path to MySql table to where the file will be exported.

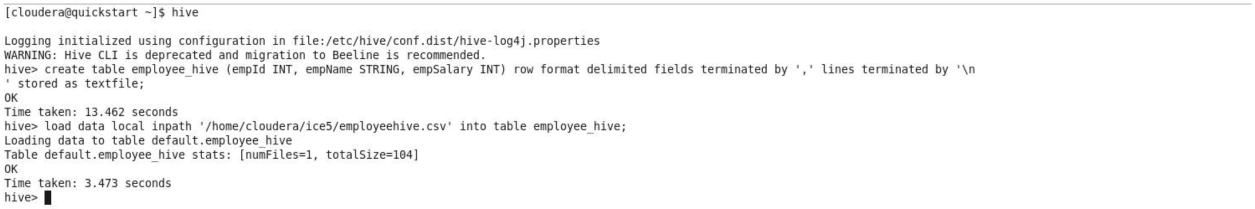


```
mysql> select * from acad_exphadoop;
+--------+----------+---------+
| emp_id | emp_name | emp_sal |
+--------+----------+---------+
|      1 | Eshwar   |   50000 |
|      2 | Harsha   |   80000 |
|      3 | Keerthi  |   60000 |
|      4 | Geetha   |   40000 |
+--------+----------+---------+
4 rows in set (0.01 sec)

mysql> █
```

- Then we display the table using select command.

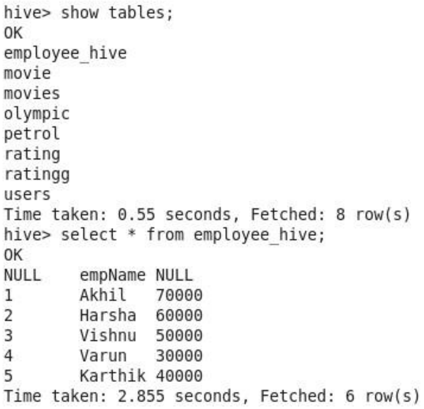
Creating Hive tables using HQL

Step 1 : Create table in Hive



```
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> create table employee_hive (empId INT, empName STRING, empSalary INT) row format delimited fields terminated by ',' lines terminated by '\n
' stored as textfile;
OK
Time taken: 13.462 seconds
hive> load data local inpath '/home/cloudera/ice5/employeehive.csv' into table employee_hive;
Loading data to table default.employee_hive
Table default.employee_hive stats: [numFiles=1, totalSize=104]
OK
Time taken: 3.473 seconds
hive> █
```

- Firstly we enter into hive using hive command.
- Then we create table with emp_id, emp_Name and emp_Salary as columns and we also specify the datatypes of columns.
- Then we load the data employee.csv into the table using load data command by specifying the path.

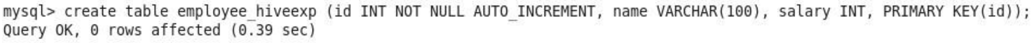


```
hive> show tables;
OK
employee_hive
movie
movies
olympic
petrol
rating
ratingg
users
Time taken: 0.55 seconds, Fetched: 8 row(s)
hive> select * from employee_hive;
OK
NULL    empName NULL
1       Akhil   70000
2       Harsha  60000
3       Vishnu  50000
4       Varun   30000
5       Karthik 40000
Time taken: 2.855 seconds, Fetched: 6 row(s)
```

- Then we display the tables present in hive using show tables command.

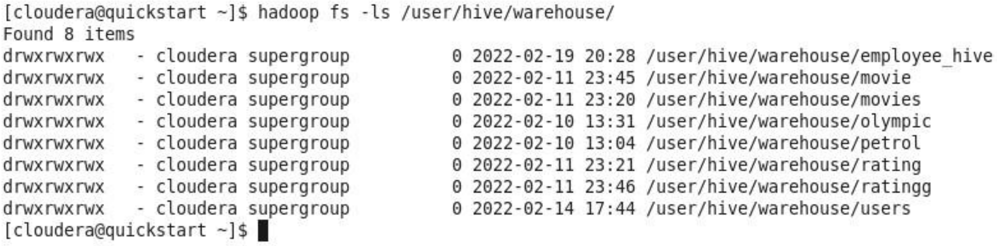
- Then we output the table using select command.

Step 2 : Create table in MySql



```
mysql> create table employee_hiveexp (id INT NOT NULL AUTO_INCREMENT, name VARCHAR(100), salary INT, PRIMARY KEY(id));
Query OK, 0 rows affected (0.39 sec)
```

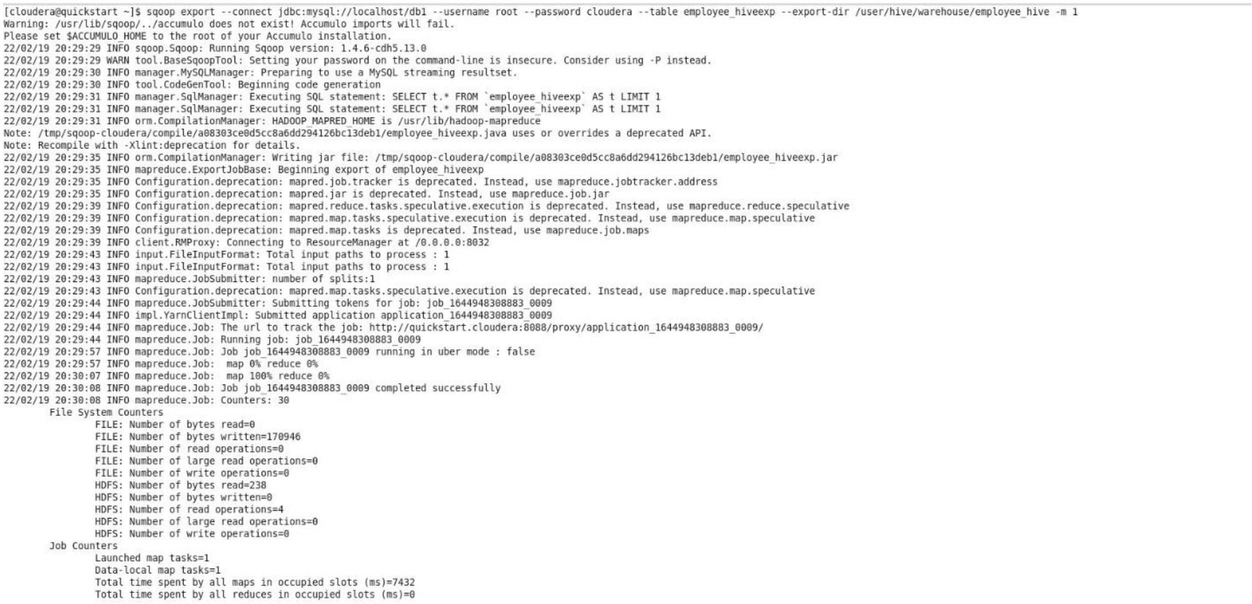
- Then we create table in MySql to which the hive table will be exported.
- We create table named employee_hiveexp with columns id, name and salary and specifying id as primary key.



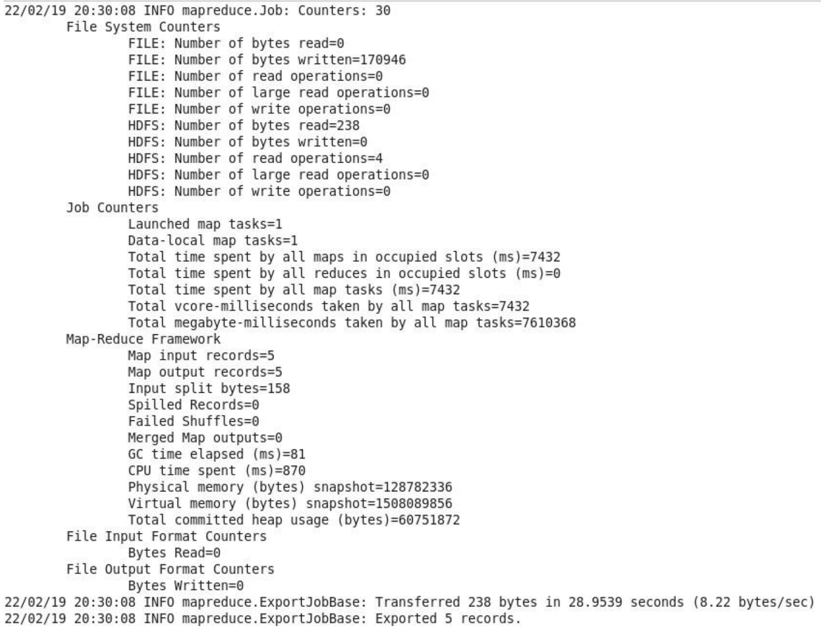
```
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/
Found 8 items
drwxrwxrwx   - cloudera supergroup          0 2022-02-19 20:28 /user/hive/warehouse/employee_hive
drwxrwxrwx   - cloudera supergroup          0 2022-02-11 23:45 /user/hive/warehouse/movie
drwxrwxrwx   - cloudera supergroup          0 2022-02-11 23:20 /user/hive/warehouse/movies
drwxrwxrwx   - cloudera supergroup          0 2022-02-10 13:31 /user/hive/warehouse/olympic
drwxrwxrwx   - cloudera supergroup          0 2022-02-10 13:04 /user/hive/warehouse/petrol
drwxrwxrwx   - cloudera supergroup          0 2022-02-11 23:21 /user/hive/warehouse/rating
drwxrwxrwx   - cloudera supergroup          0 2022-02-11 23:46 /user/hive/warehouse/ratingg
drwxrwxrwx   - cloudera supergroup          0 2022-02-14 17:44 /user/hive/warehouse/users
[cloudera@quickstart ~]$ █
```

- Then we display the files in hive using -ls command by specifying the path.

Step 3 : Exporting hive table to MySQL

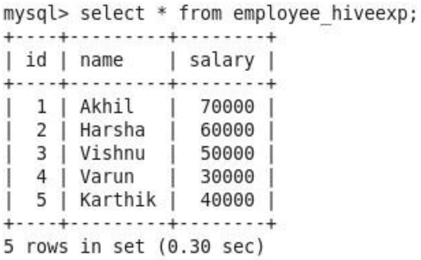


```
[cloudera@quickstart ~]$ sqoop export --connect jdbc:mysql://localhost/db1 --username root --password cloudera --table employee_hiveexp --export-dir /user/hive/warehouse/employee_hive -m 1
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
22/02/19 20:29:29 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
22/02/19 20:29:29 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
22/02/19 20:29:30 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
22/02/19 20:29:30 INFO tool.CodeGenTool: Beginning code generation
22/02/19 20:29:31 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee_hiveexp` AS t LIMIT 1
22/02/19 20:29:31 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee_hiveexp` AS t LIMIT 1
22/02/19 20:29:31 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/a08303ce0d5cc8a6dd294126bc13deb1/employee_hiveexp.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
22/02/19 20:29:35 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/a08303ce0d5cc8a6dd294126bc13deb1/employee_hiveexp.jar
22/02/19 20:29:35 INFO mapreduce.ExportJobBase: Beginning export of employee_hiveexp
22/02/19 20:29:35 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
22/02/19 20:29:35 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
22/02/19 20:29:39 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
22/02/19 20:29:39 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
22/02/19 20:29:39 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
22/02/19 20:29:39 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/02/19 20:29:43 INFO input.FileInputFormat: Total input paths to process : 1
22/02/19 20:29:43 INFO input.FileInputFormat: Total input paths to process : 1
22/02/19 20:29:43 INFO mapreduce.JobSubmitter: number of splits:1
22/02/19 20:29:43 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
22/02/19 20:29:44 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1644948308883_0009
22/02/19 20:29:44 INFO impl.YarnClientImpl: Submitted application application_1644948308883_0009
22/02/19 20:29:44 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1644948308883_0009/
22/02/19 20:29:44 INFO mapreduce.Job: Running job: job_1644948308883_0009
22/02/19 20:29:57 INFO mapreduce.Job: Job job_1644948308883_0009 running in uber mode : false
22/02/19 20:29:57 INFO mapreduce.Job:  map 0% reduce 0%
22/02/19 20:30:07 INFO mapreduce.Job:  map 100% reduce 0%
22/02/19 20:30:08 INFO mapreduce.Job: Job job_1644948308883_0009 completed successfully
22/02/19 20:30:08 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=170946
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=238
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=0
        Job Counters
                Launched map tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=7432
                Total time spent by all reduces in occupied slots (ms)=0
```



```
22/02/19 20:30:08 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=170946
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=238
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=0
        Job Counters
                Launched map tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=7432
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=7432
                Total vcore-milliseconds taken by all map tasks=7432
                Total megabyte-milliseconds taken by all map tasks=7610368
        Map-Reduce Framework
                Map input records=5
                Map output records=5
                Input split bytes=158
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=81
                CPU time spent (ms)=870
                Physical memory (bytes) snapshot=128782336
                Virtual memory (bytes) snapshot=1508089856
                Total committed heap usage (bytes)=60751872
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
22/02/19 20:30:08 INFO mapreduce.ExportJobBase: Transferred 238 bytes in 28.9539 seconds (8.22 bytes/sec)
22/02/19 20:30:08 INFO mapreduce.ExportJobBase: Exported 5 records.
```

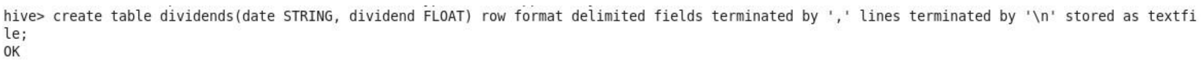
- Then we export the hive table named employee_hive into MySql table named employee_hiveexp by specifying the path and also we gain access to localhost using username and password.



```
mysql> select * from employee_hiveexp;
+----+---------+--------+
| id | name    | salary |
+----+---------+--------+
|  1 | Akhil   |  70000 |
|  2 | Harsha  |  60000 |
|  3 | Vishnu  |  50000 |
|  4 | Varun   |  30000 |
|  5 | Karthik |  40000 |
+----+---------+--------+
5 rows in set (0.30 sec)
```

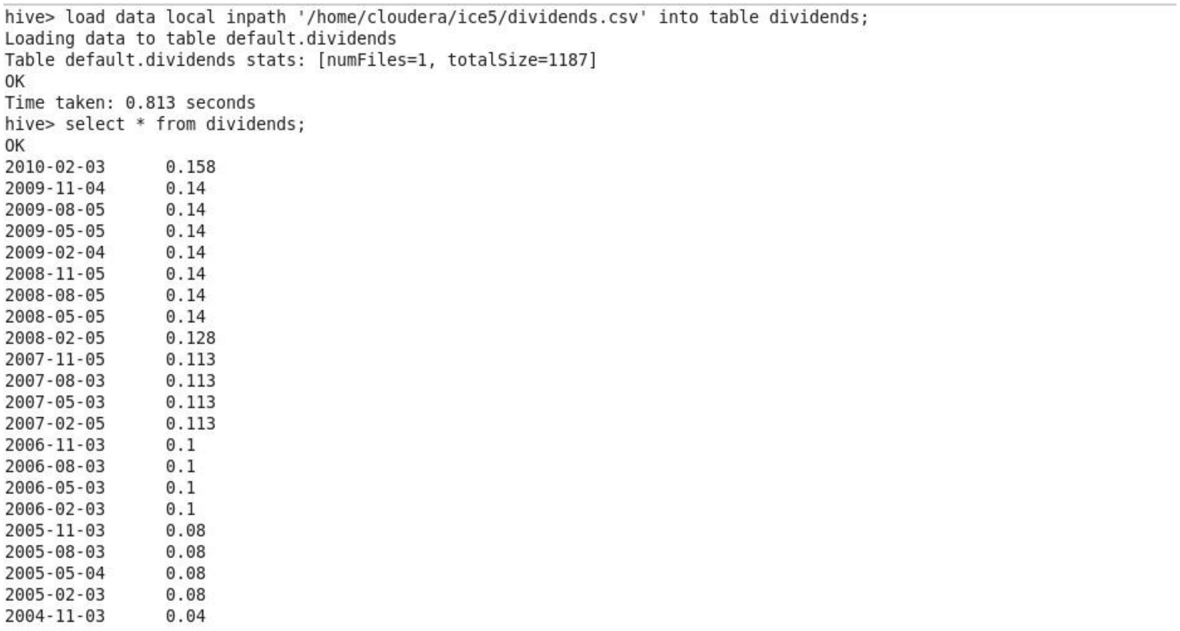
- Then we check whether the table is exported or not using select command.
- The hive table is successfully exported to Sql.

Creating Dividends table in Hive



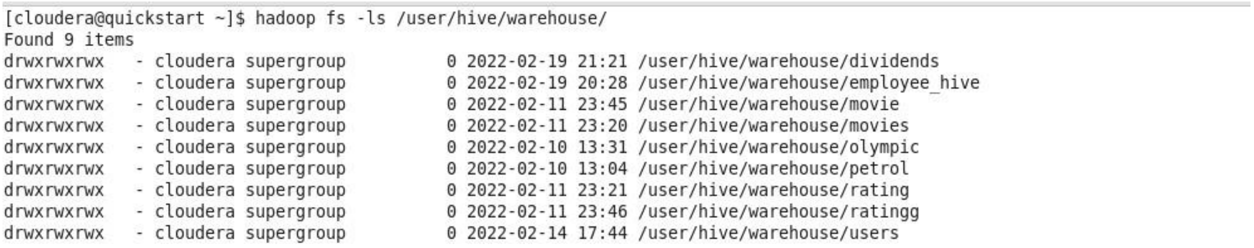
```
hive> create table dividends(date STRING, dividend FLOAT) row format delimited fields terminated by ',' lines terminated by '\n' stored as textfi
le;
OK
```

- Firstly we create a table named dividends in hive and we specify columns as date and dividend with their respective datatypes.



```
hive> load data local inpath '/home/cloudera/ice5/dividends.csv' into table dividends;
Loading data to table default.dividends
Table default.dividends stats: [numFiles=1, totalSize=1187]
OK
Time taken: 0.813 seconds
hive> select * from dividends;
OK
2010-02-03      0.158
2009-11-04      0.14
2009-08-05      0.14
2009-05-05      0.14
2009-02-04      0.14
2008-11-05      0.14
2008-08-05      0.14
2008-05-05      0.14
2008-02-05      0.128
2007-11-05      0.113
2007-08-03      0.113
2007-05-03      0.113
2007-02-05      0.113
2006-11-03      0.1
2006-08-03      0.1
2006-05-03      0.1
2006-02-03      0.1
2005-11-03      0.08
2005-08-03      0.08
2005-05-04      0.08
2005-02-03      0.08
2004-11-03      0.04
```

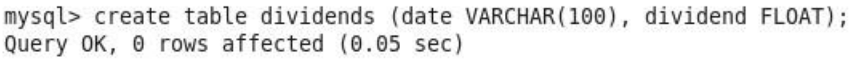
- Then we load the data dividends.csv into dividends table using load data command.
- Then we display the data in table using select command.



```
[cloudera@quickstart ~]$ hadoop fs -ls /user/hive/warehouse/
Found 9 items
drwxrwxrwx   - cloudera supergroup          0 2022-02-19 21:21 /user/hive/warehouse/dividends
drwxrwxrwx   - cloudera supergroup          0 2022-02-19 20:28 /user/hive/warehouse/employee_hive
drwxrwxrwx   - cloudera supergroup          0 2022-02-11 23:45 /user/hive/warehouse/movie
drwxrwxrwx   - cloudera supergroup          0 2022-02-11 23:20 /user/hive/warehouse/movies
drwxrwxrwx   - cloudera supergroup          0 2022-02-10 13:31 /user/hive/warehouse/olympic
drwxrwxrwx   - cloudera supergroup          0 2022-02-10 13:04 /user/hive/warehouse/petrol
drwxrwxrwx   - cloudera supergroup          0 2022-02-11 23:21 /user/hive/warehouse/rating
drwxrwxrwx   - cloudera supergroup          0 2022-02-11 23:46 /user/hive/warehouse/ratingg
drwxrwxrwx   - cloudera supergroup          0 2022-02-14 17:44 /user/hive/warehouse/users
```

- Then we check the items present in hive using -ls command.

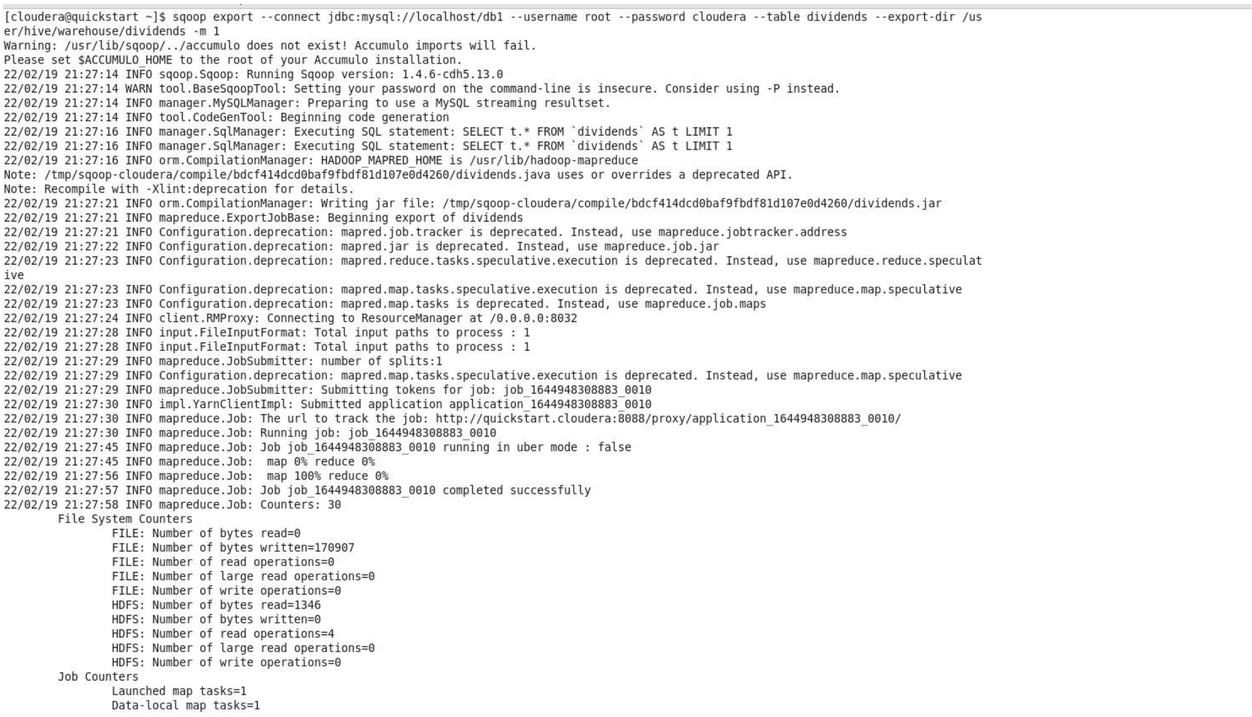
Step 2 : Create table in MySQL



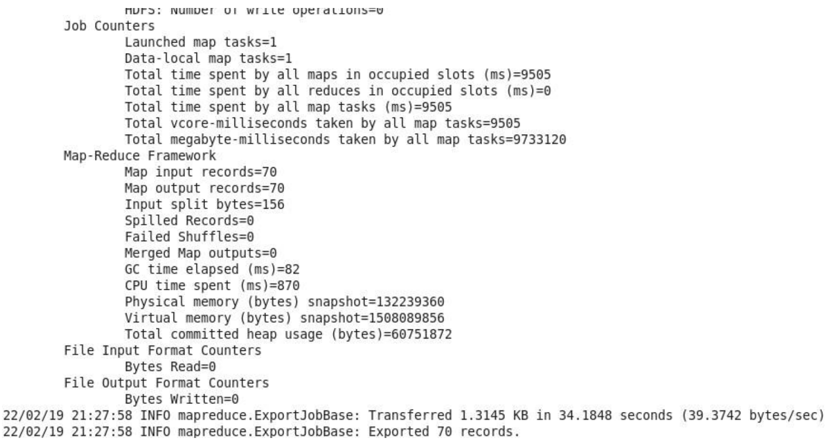
```
mysql> create table dividends (date VARCHAR(100), dividend FLOAT);
Query OK, 0 rows affected (0.05 sec)
```

- Then we create table named dividends in MySql to which the hive table will be exported.

Step 3 : Exporting hive table to MySQL

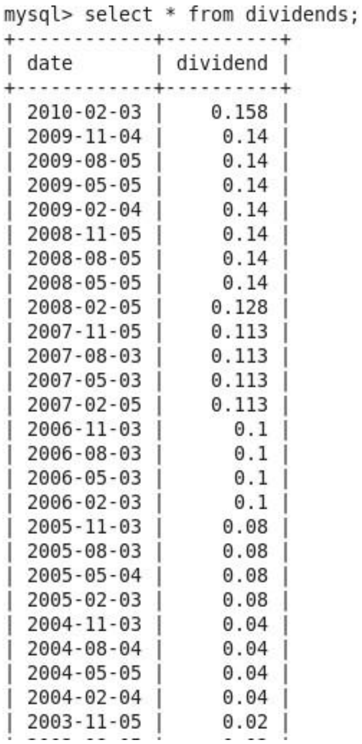


```
[cloudera@quickstart ~]$ sqoop export --connect jdbc:mysql://localhost/db1 --username root --password cloudera --table dividends --export-dir /us
er/hive/warehouse/dividends -m 1
Warning: /usr/lib/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
22/02/19 21:27:14 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
22/02/19 21:27:14 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
22/02/19 21:27:14 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
22/02/19 21:27:14 INFO tool.CodeGenTool: Beginning code generation
22/02/19 21:27:16 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `dividends` AS t LIMIT 1
22/02/19 21:27:16 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `dividends` AS t LIMIT 1
22/02/19 21:27:16 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/bdcf414dcd0baf9fbdf81d107e0d4260/dividends.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
22/02/19 21:27:21 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/bdcf414dcd0baf9fbdf81d107e0d4260/dividends.jar
22/02/19 21:27:21 INFO mapreduce.ExportJobBase: Beginning export of dividends
22/02/19 21:27:21 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
22/02/19 21:27:22 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
22/02/19 21:27:23 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculat
ive
22/02/19 21:27:23 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
22/02/19 21:27:23 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
22/02/19 21:27:24 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/02/19 21:27:28 INFO input.FileInputFormat: Total input paths to process : 1
22/02/19 21:27:28 INFO input.FileInputFormat: Total input paths to process : 1
22/02/19 21:27:29 INFO mapreduce.JobSubmitter: number of splits:1
22/02/19 21:27:29 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
22/02/19 21:27:29 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1644948308883_0010
22/02/19 21:27:30 INFO impl.YarnClientImpl: Submitted application application_1644948308883_0010
22/02/19 21:27:30 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1644948308883_0010/
22/02/19 21:27:30 INFO mapreduce.Job: Running job: job_1644948308883_0010
22/02/19 21:27:45 INFO mapreduce.Job: Job job_1644948308883_0010 running in uber mode : false
22/02/19 21:27:45 INFO mapreduce.Job:  map 0% reduce 0%
22/02/19 21:27:56 INFO mapreduce.Job:  map 100% reduce 0%
22/02/19 21:27:57 INFO mapreduce.Job: Job job_1644948308883_0010 completed successfully
22/02/19 21:27:58 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=170907
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1346
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=0
        Job Counters
                Launched map tasks=1
                Data-local map tasks=1
```



```
                HDFS: Number of write operations=0
        Job Counters
                Launched map tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=9505
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=9505
                Total vcore-milliseconds taken by all map tasks=9505
                Total megabyte-milliseconds taken by all map tasks=9733120
        Map-Reduce Framework
                Map input records=70
                Map output records=70
                Input split bytes=156
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=82
                CPU time spent (ms)=870
                Physical memory (bytes) snapshot=132239360
                Virtual memory (bytes) snapshot=1508089856
                Total committed heap usage (bytes)=60751872
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
22/02/19 21:27:58 INFO mapreduce.ExportJobBase: Transferred 1.3145 KB in 34.1848 seconds (39.3742 bytes/sec)
22/02/19 21:27:58 INFO mapreduce.ExportJobBase: Exported 70 records.
```

- Then we export the table present in hive named dividends to the table present in MySql named dividends using sqoop export command by specifying the username and password of localhost and also specify the path of files and mapper as 1.

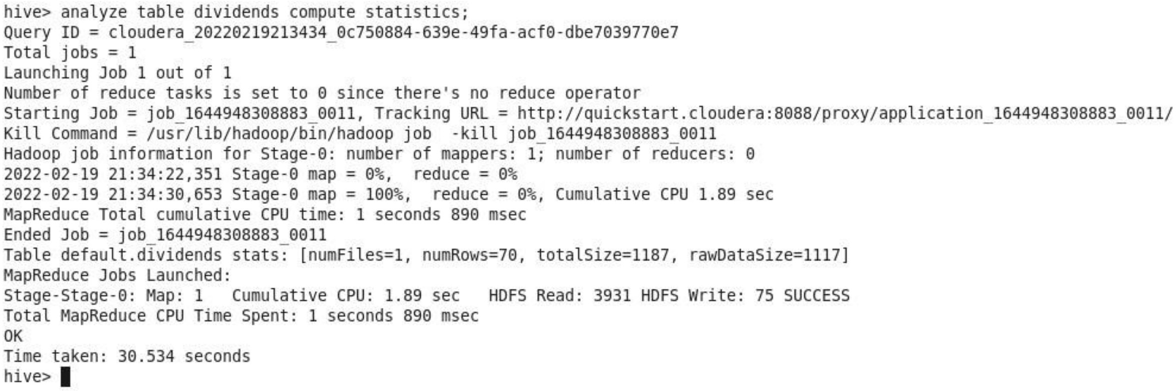


```
mysql> select * from dividends;
+------------+----------+
| date       | dividend |
+------------+----------+
| 2010-02-03 |    0.158 |
| 2009-11-04 |     0.14 |
| 2009-08-05 |     0.14 |
| 2009-05-05 |     0.14 |
| 2009-02-04 |     0.14 |
| 2008-11-05 |     0.14 |
| 2008-08-05 |     0.14 |
| 2008-05-05 |     0.14 |
| 2008-02-05 |    0.128 |
| 2007-11-05 |    0.113 |
| 2007-08-03 |    0.113 |
| 2007-05-03 |    0.113 |
| 2007-02-05 |    0.113 |
| 2006-11-03 |      0.1 |
| 2006-08-03 |      0.1 |
| 2006-05-03 |      0.1 |
| 2006-02-03 |      0.1 |
| 2005-11-03 |     0.08 |
| 2005-08-03 |     0.08 |
| 2005-05-04 |     0.08 |
| 2005-02-03 |     0.08 |
| 2004-11-03 |     0.04 |
| 2004-08-04 |     0.04 |
| 2004-05-05 |     0.04 |
| 2004-02-04 |     0.04 |
| 2003-11-05 |     0.02 |
```

- Then we check and display the data in dividends table in MySql using select command.

Queries

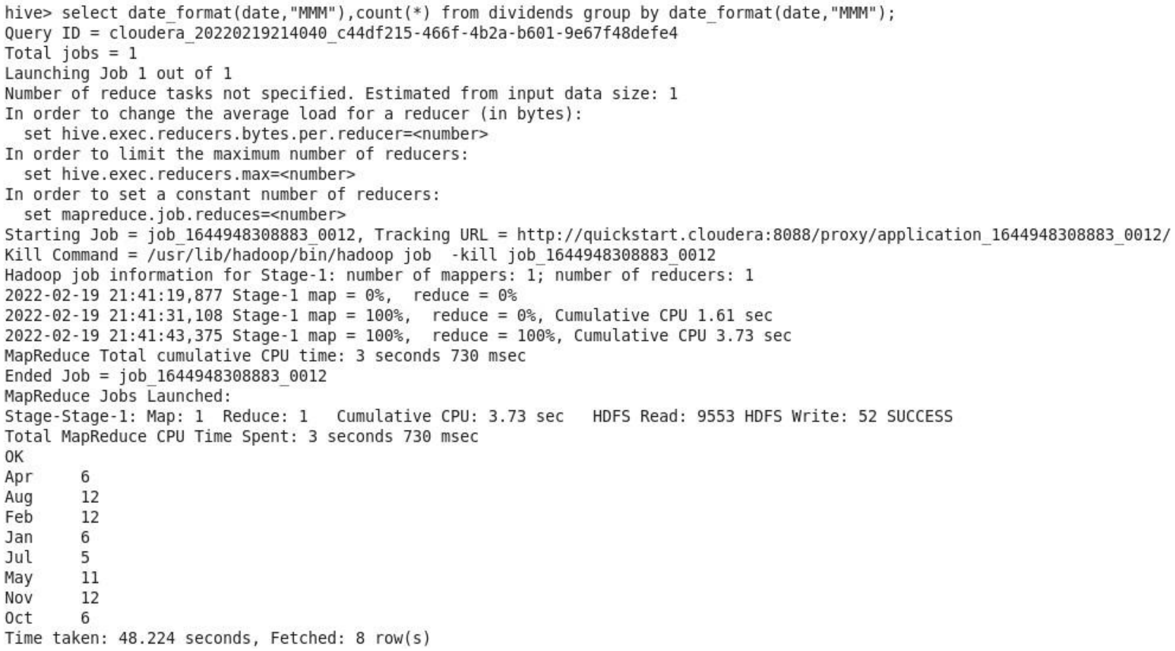
1. Write the command to analyze the statistics of dividends table.



```
hive> analyze table dividends compute statistics;
Query ID = cloudera_20220219213434_0c750884-639e-49fa-acf0-dbe7039770e7
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1644948308883_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1644948308883_0011/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1644948308883_0011
Hadoop job information for Stage-0: number of mappers: 1; number of reducers: 0
2022-02-19 21:34:22,351 Stage-0 map = 0%,  reduce = 0%
2022-02-19 21:34:30,653 Stage-0 map = 100%,  reduce = 0%, Cumulative CPU 1.89 sec
MapReduce Total cumulative CPU time: 1 seconds 890 msec
Ended Job = job_1644948308883_0011
Table default.dividends stats: [numFiles=1, numRows=70, totalSize=1187, rawDataSize=1117]
MapReduce Jobs Launched:
Stage-Stage-0: Map: 1   Cumulative CPU: 1.89 sec   HDFS Read: 3931 HDFS Write: 75 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 890 msec
OK
Time taken: 30.534 seconds
hive> █
```

- Then we analyze the statistics of the table dividends using analyze and compute command.
- Then we received the stats of the table including information like number of files, rows, totalsize and raw datasize.

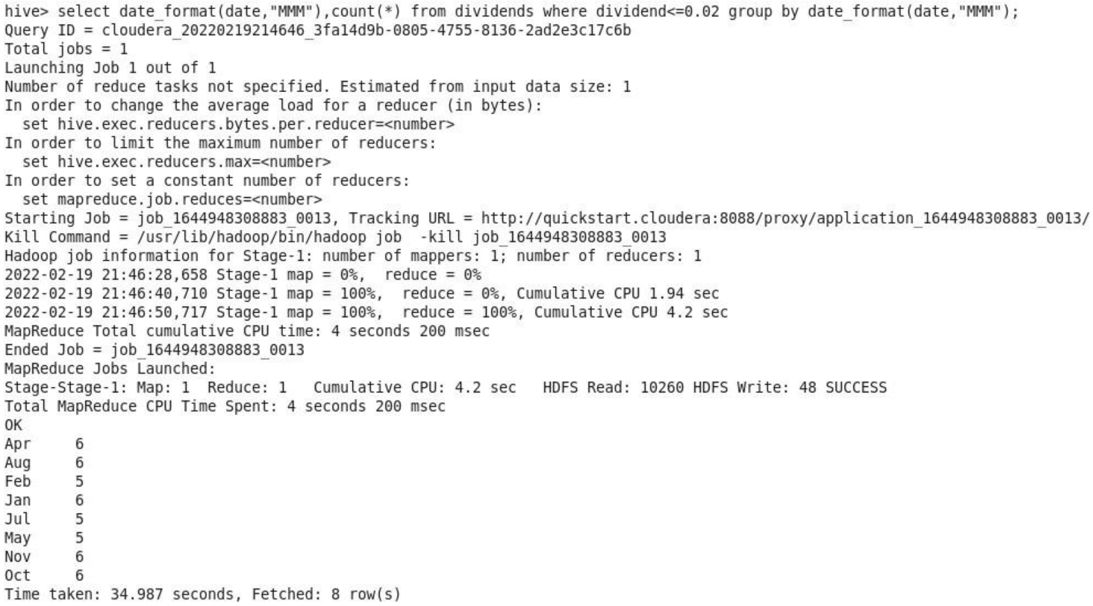
2. Write command to obtain the count of dividends in every month.



```
hive> select date_format(date,"MMM"),count(*) from dividends group by date_format(date,"MMM");
Query ID = cloudera_20220219214040_c44df215-466f-4b2a-b601-9e67f48defe4
Total jobs = 1
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_1644948308883_0012, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1644948308883_0012/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1644948308883_0012
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-02-19 21:41:19,877 Stage-1 map = 0%,  reduce = 0%
2022-02-19 21:41:31,108 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.61 sec
2022-02-19 21:41:43,375 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.73 sec
MapReduce Total cumulative CPU time: 3 seconds 730 msec
Ended Job = job_1644948308883_0012
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.73 sec   HDFS Read: 9553 HDFS Write: 52 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 730 msec
OK
Apr     6
Aug     12
Feb     12
Jan     6
Jul     5
May     11
Nov     12
Oct     6
Time taken: 48.224 seconds, Fetched: 8 row(s)
```

- Then we get the count of dividends in each month using select command which selects and displays the month and count from dividends table and we use group by to group the data according to month.

3. Write command to obtain the count of dividends in every month with condition as dividend less than 0.02.



```
hive> select date_format(date,"MMM"),count(*) from dividends where dividend<=0.02 group by date_format(date,"MMM");
Query ID = cloudera_20220219214646_3fa14d9b-0805-4755-8136-2ad2e3c17c6b
Total jobs = 1
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_1644948308883_0013, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1644948308883_0013/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1644948308883_0013
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-02-19 21:46:28,658 Stage-1 map = 0%,  reduce = 0%
2022-02-19 21:46:40,710 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.94 sec
2022-02-19 21:46:50,717 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.2 sec
MapReduce Total cumulative CPU time: 4 seconds 200 msec
Ended Job = job_1644948308883_0013
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.2 sec   HDFS Read: 10260 HDFS Write: 48 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 200 msec
OK
Apr     6
Aug     6
Feb     5
Jan     6
Jul     5
May     5
Nov     6
Oct     6
Time taken: 34.987 seconds, Fetched: 8 row(s)
```

- Then we get the count of dividends in every month where the condition is that the dividend must be less than 0.02.
- For this we use the select command to select and display month and count from dividends table and we use where command to specify the condition and group by to group the data according to month.