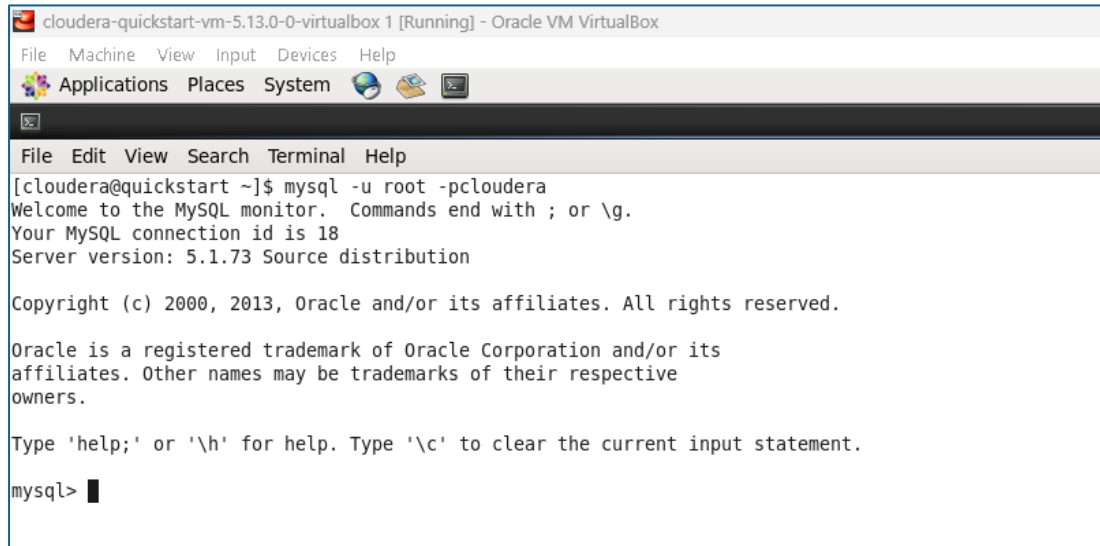


BDA EXPERIMENT-2

Name: Sahil Dinesh Patil	Class: BE3
Roll no: 38	Batch: B

- **Part1 : Using sqoop to import data from mysql to hadoop**
- `mysql -u root -pcloudera`



The screenshot shows a terminal window titled "cloudera-quickstart-vm-5.13.0-0-virtualbox 1 [Running] - Oracle VM VirtualBox". The terminal displays the output of the command `mysql -u root -pcloudera`. The output includes a welcome message, connection details, and the MySQL prompt `mysql>`.

```
cloudera-quickstart-vm-5.13.0-0-virtualbox 1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
[mysql]
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 18
Server version: 5.1.73 Source distribution

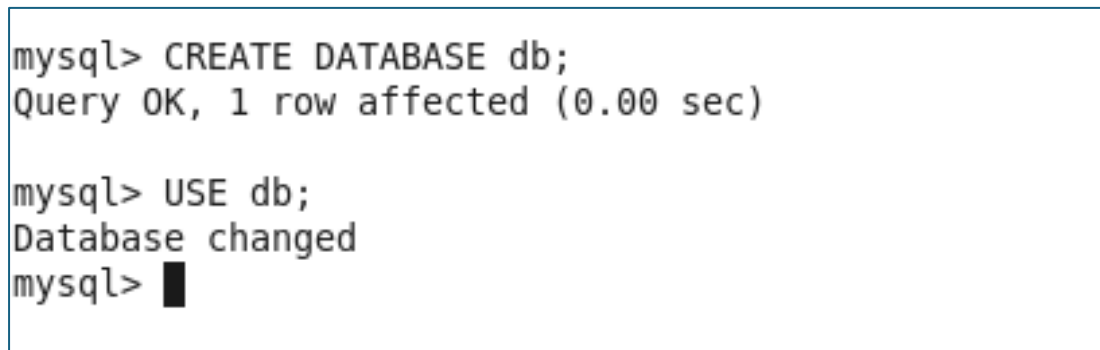
Copyright (c) 2000, 2013, Oracle and/or its affiliates. All rights reserved.

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>
```

- `CREATE DATABASE db;`
`USE db;`



The screenshot shows a terminal window with the MySQL prompt `mysql>`. The user enters the command `CREATE DATABASE db;`, and the output is `Query OK, 1 row affected (0.00 sec)`. Then, the user enters `USE db;`, and the output is `Database changed`. The prompt `mysql>` is shown again.

```
mysql> CREATE DATABASE db;
Query OK, 1 row affected (0.00 sec)

mysql> USE db;
Database changed
mysql>
```

- CREATE TABLE employees (
 id INT NOT NULL PRIMARY KEY,
 name VARCHAR (20),
 deg VARCHAR (20),
 salary INT,
 dept VARCHAR (10)
);

```
mysql> USE db;  
Database changed  
mysql> CREATE TABLE employees(  
    -> id INT NOT NULL PRIMARY KEY,  
    -> name VARCHAR(20),  
    -> deg VARCHAR(20),  
    -> salary INT,  
    -> dept VARCHAR(10)  
    -> );  
Query OK, 0 rows affected (0.03 sec)  
  
mysql> █
```

- INSERT INTO employees VALUES (1, 'Archit Lokhande', 'SDE1', 100000, 'SD'),
(2, 'Om Hedau', 'Data Analyst', 75000, 'DW'),
(3, 'Shail Patil', 'Data Scientist', 120000, 'DW'),
(4, 'Mihir Khatri', 'Cyber Expert', 105000, 'ST');

```
mysql> INSERT INTO employees VALUES(1,'Archit Lokhande','SDE1',100000,'SD'),  
    -> (2,'Om Hedau','Data Analyst',75000,'DW'),  
    -> (3,'Shail Patil','Data Scientist',120000,'DW'),  
    -> (4,'Mihir Khatri','Cyber Expert',105000,'ST');  
Query OK, 4 rows affected (0.00 sec)  
Records: 4 Duplicates: 0 Warnings: 0  
  
mysql> █
```

- Exit

```
mysql> INSERT INTO employees VALUES(1,'Archit Lokhande','SDE1',100000,'SD'),
-> (2,'Om Hedau','Data Analyst',75000,'DW'),
-> (3,'Sahil Patil','Data Scientist',120000,'DW'),
-> (4,'Mihir Khatri','Cyber Expert',105000,'ST');
Query OK, 4 rows affected (0.00 sec)
Records: 4 Duplicates: 0 Warnings: 0

mysql> EXIT
Bye
[cloudera@quickstart ~]$ █
```

- sqoop import --connect jdbc:mysql://localhost/db --username root --table employees --password cloudera

```
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/db --username root --table employees --password cloudera
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
24/08/04 22:21:25 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
24/08/04 22:21:25 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
24/08/04 22:21:25 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
24/08/04 22:21:25 INFO tool.CodeGenTool: Beginning code generation
24/08/04 22:21:26 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employees` AS t LIMIT 1
24/08/04 22:21:26 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employees` AS t LIMIT 1
24/08/04 22:21:26 INFO orm.CompilationManager: HADOOP MAPRED HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/bd67765cd0f206fd9ca3c60b9ad53f6e/employees.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
24/08/04 22:21:27 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/bd67765cd0f206fd9ca3c60b9ad53f6e/employees.jar
24/08/04 22:21:27 WARN manager.MySQLManager: It looks like you are importing from mysql.
24/08/04 22:21:27 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
24/08/04 22:21:27 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
24/08/04 22:21:27 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
24/08/04 22:21:27 INFO mapreduce.ImportJobBase: Beginning import of employees
24/08/04 22:21:27 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
24/08/04 22:21:27 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
24/08/04 22:21:28 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
24/08/04 22:21:28 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
24/08/04 22:21:28 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)
24/08/04 22:21:28 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$DataStreamer.run(DFSOutputStream.java:894)
24/08/04 22:21:28 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$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)
24/08/04 22:21:29 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
```

```

24/08/04 22:21:30 INFO mapreduce.Job: Running job: job_1722834117673_0001
24/08/04 22:21:37 INFO mapreduce.Job: Job job_1722834117673_0001 running in uber mode : false
24/08/04 22:21:37 INFO mapreduce.Job: map 0% reduce 0%
24/08/04 22:21:47 INFO mapreduce.Job: map 25% reduce 0%
24/08/04 22:21:50 INFO mapreduce.Job: map 50% reduce 0%
24/08/04 22:21:51 INFO mapreduce.Job: map 100% reduce 0%
24/08/04 22:21:51 INFO mapreduce.Job: Job job_1722834117673_0001 completed successfully
24/08/04 22:21:52 INFO mapreduce.Job: Counters: 31
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=684500
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=393
    HDFS: Number of bytes written=143
    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=4
    Total time spent by all maps in occupied slots (ms)=37725
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=37725
    Total vcore-milliseconds taken by all map tasks=37725
    Total megabyte-milliseconds taken by all map tasks=38630400
  Map-Reduce Framework
    Map input records=4
    Map output records=4
    Input split bytes=393
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=424
    CPU time spent (ms)=2180
    Physical memory (bytes) snapshot=524550144
    Virtual memory (bytes) snapshot=6040702976
    Total committed heap usage (bytes)=243007488
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=143
24/08/04 22:21:52 INFO mapreduce.ImportJobBase: Transferred 143 bytes in 23.7813 seconds (6.0131 bytes/sec)
24/08/04 22:21:52 INFO mapreduce.ImportJobBase: Retrieved 4 records.
[cloudera@quickstart ~]$

```

- `hadoop fs -ls /user/cloudera/employees`

```

[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/employees
Found 5 items
-rw-r--r-- 1 cloudera cloudera      0 2024-08-04 22:21 /user/cloudera/employees/_SUCCESS
-rw-r--r-- 1 cloudera cloudera    33 2024-08-04 22:21 /user/cloudera/employees/part-m-00000
-rw-r--r-- 1 cloudera cloudera    33 2024-08-04 22:21 /user/cloudera/employees/part-m-00001
-rw-r--r-- 1 cloudera cloudera    39 2024-08-04 22:21 /user/cloudera/employees/part-m-00002
-rw-r--r-- 1 cloudera cloudera    38 2024-08-04 22:21 /user/cloudera/employees/part-m-00003
[cloudera@quickstart ~]$

```

- `hadoop fs -ls /user/cloudera/employees/part-m-00000`
- `hadoop fs -ls /user/cloudera/employees/part-m-00001`
- `hadoop fs -ls /user/cloudera/employees/part-m-00002`

```

File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/employees/part-m-00000
-rw-r--r-- 1 cloudera cloudera    33 2024-08-04 22:21 /user/cloudera/employees/part-m-00000
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/employees/part-m-00001
-rw-r--r-- 1 cloudera cloudera    33 2024-08-04 22:21 /user/cloudera/employees/part-m-00001
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/employees/part-m-00002
-rw-r--r-- 1 cloudera cloudera    39 2024-08-04 22:21 /user/cloudera/employees/part-m-00002
[cloudera@quickstart ~]$ hadoop fs -ls /user/cloudera/employees/part-m-00003
-rw-r--r-- 1 cloudera cloudera    38 2024-08-04 22:21 /user/cloudera/employees/part-m-00003

```

- **Part2 : Using sqoop to export table from Hadoop to mysql**

mysql -u root -pcloudera

Use db;

DROP TABLE employees;

```
[cloudera@quickstart ~]$ mysql -u root -pcloudera
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 27
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 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> DROP TABLE employees;
Query OK, 0 rows affected (0.01 sec)

mysql> █
```

- CREATE TABLE employees (
id INT NULL PRIMARY KEY;
name VARCHAR (20),
deg VARCHAR (20),
salary INT,
dept VARCHAR (10)
);
EXIT;

```
[cloudera@quickstart ~]$ mysql -u root -pcloudera
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 29
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 db;
Database changed
mysql> CREATE TABLE employees(
-> id INT NOT NULL PRIMARY KEY,
-> name VARCHAR(20),
-> deg VARCHAR(20),
-> salary INT,
-> dept VARCHAR(10)
-> );
Query OK, 0 rows affected (0.02 sec)

mysql> █
```

- `sqoop export --connect jdbc:mysql://localhost/db-table employees --export-dir /user/cloudera/employees --input-fields-terminated-by ',' --username root --password cloudera`

```
[cloudera@quickstart ~]$ sqoop export --connect jdbc:mysql://localhost/db --table employees --export-dir /user/cloudera/employees --input-fields-terminated-by ',' --username root --password cloudera
Warning: /usr/lib/sqoop/.../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
24/08/04 22:45:23 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
24/08/04 22:45:23 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
24/08/04 22:45:23 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
24/08/04 22:45:23 INFO tool.CodeGenTool: Beginning code generation
24/08/04 22:45:24 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employees` AS t LIMIT 1
24/08/04 22:45:24 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employees` AS t LIMIT 1
24/08/04 22:45:24 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: Recompile with -Xlint:deprecation for details.
24/08/04 22:45:25 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/84e877b4d1ed923b4daedcf9b62e5203/employees.jar
24/08/04 22:45:25 INFO mapreduce.ExportJobBase: Beginning export of employees
24/08/04 22:45:25 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.job.tracker.address
24/08/04 22:45:25 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
24/08/04 22:45:26 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
24/08/04 22:45:26 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
24/08/04 22:45:26 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.map.maps
24/08/04 22:45:26 INFO client.HMProxy: Connecting to ResourceManager at /0.0.0.0:8032
24/08/04 22:45:27 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:785)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
24/08/04 22:45:27 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:785)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
24/08/04 22:45:27 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:785)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
24/08/04 22:45:27 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:785)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
24/08/04 22:45:27 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:785)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
24/08/04 22:45:43 INFO mapreduce.Job: Counters: 12
Job Counters
  Failed map tasks=1
  Killed map tasks=2
  Launched map tasks=3
  Data-local map tasks=3
  Total time spent by all maps in occupied slots (ms)=17012
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=17012
  Total vcore-milliseconds taken by all map tasks=17012
  Total megabyte-milliseconds taken by all map tasks=17420288
Map-Reduce Framework
  CPU time spent (ms)=0
  Physical memory (bytes) snapshot=0
  Virtual memory (bytes) snapshot=0
24/08/04 22:45:43 WARN mapreduce.Counters: Group FileSystemCounters is deprecated. Use org.apache.hadoop.mapreduce.FileSystemCounter instead
24/08/04 22:45:43 INFO mapreduce.ExportJobBase: Transferred 0 bytes in 16.2906 seconds (0 bytes/sec)
24/08/04 22:45:43 INFO mapreduce.ExportJobBase: Exported 0 records.
24/08/04 22:45:43 ERROR tool.ExportTool: Error during export:
Export job failed!
    at org.apache.sqoop.mapreduce.ExportJobBase.runExport(ExportJobBase.java:439)
    at org.apache.sqoop.manager.SqlManager.exportTable(SqlManager.java:931)
    at org.apache.sqoop.tool.ExportTool.exportTable(ExportTool.java:80)
    at org.apache.sqoop.tool.ExportTool.run(ExportTool.java:99)
    at org.apache.sqoop.Sqoop.run(Sqoop.java:147)
    at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:70)
    at org.apache.sqoop.Sqoop.runSqoop(Sqoop.java:183)
    at org.apache.sqoop.Sqoop.runTool(Sqoop.java:234)
    at org.apache.sqoop.Sqoop.runTool(Sqoop.java:243)
    at org.apache.sqoop.Sqoop.main(Sqoop.java:252)
[cloudera@quickstart ~]$
```

```
mysql -u root -pcloudera
```

```
USE db;
```

```
SELECT * FROM employees;
```

```
[cloudera@quickstart ~]$ mysql -u root -pcloudera
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 40
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 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> SELECT *FROM employees;
+-----+-----+-----+-----+-----+
| id | name          | deg          | salary | dept |
+-----+-----+-----+-----+-----+
| 1 | Archit Lokhande | SDE1        | 100000 | SD   |
| 2 | Om Hedau       | Data Scientist | 150000 | DW   |
| 3 | Sahil Patil    | Data Analyst  | 150000 | DW   |
| 4 | Mihir Khatri   | Software tester | 85000  | SD   |
+-----+-----+-----+-----+-----+
4 rows in set (0.00 sec)

mysql> █
```