

# BDA Exp2: Sqoop

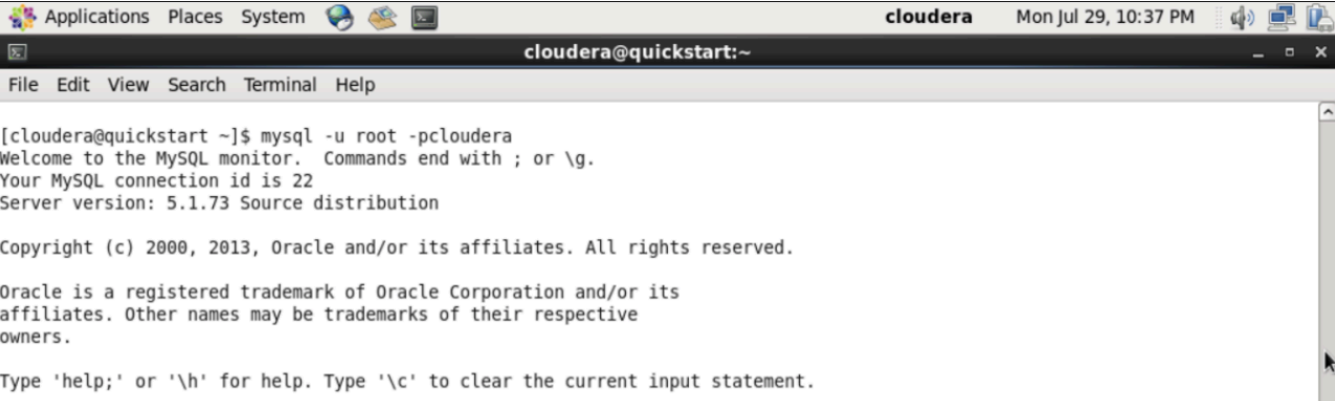
Aditya Tawade, BE4, 60

## Part 1: Using sqoop to import data from mysql to hadoop

**Step1:** Open terminal in cloudera

**Step2:** Login to mysql (username is "root" and password is "cloudera").

**Command:** `mysql -u root -pcloudera`



The screenshot shows a terminal window titled "cloudera@quickstart:~". The terminal output displays the MySQL command-line interface after running the command `mysql -u root -pcloudera`. The output includes a welcome message, connection ID (22), server version (5.1.73), and copyright information. The prompt is `[cloudera@quickstart ~]$`.

```
[cloudera@quickstart ~]$ mysql -u root -pcloudera
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 22
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.
```

**Step 3:**

(a) Create a database(db) and set it as current database.

(b) Create a table called employees and fill it up with data.

(c) Exit mysql.

**Commands:**

```
CREATE DATABASE db;
USE db;
CREATE TABLE employees (
    id INT NOT NULL PRIMARY KEY,
    name VARCHAR(20),
    deg VARCHAR(20),
    salary INT,
    dept VARCHAR(10)
);
INSERT INTO employees VALUES(1, 'Aditya Tawade', 'CEO',1000000, 'SD'),
(2, 'Suraj Jagtap', 'Janitor',2, 'SD'), (3, 'Yug Solanki', 'Senior
Developer',100000, 'SD');
EXIT;
```

```
mysql> USE db;
Database changed
mysql> CREATE TABLE employees();
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near ')' at line 1
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> INSERT INTO employees VALUES (1,'Aditya Tawade', 'CEO', 1000000, 'SD'), (2, 'Suraj Jagtap', 'Janitor', 2,'SD'),(3, 'Yug Solanki', 'Senior Developer', 100000,'SD');
Query OK, 3 rows affected (0.00 sec)
Records: 3 Duplicates: 0 Warnings: 0

mysql> EXIT;
Bye
```

#### Step 4: Use sqoop to import the mysql table to hadoop

##### Commands:

```
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.
24/07/29 22:28:01 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0
24/07/29 22:28:01 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
24/07/29 22:28:01 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
24/07/29 22:28:01 INFO tool.CodeGenTool: Beginning code generation
24/07/29 22:28:02 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employees` AS t LIMIT 1
24/07/29 22:28:02 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employees` AS t LIMIT 1
24/07/29 22:28:02 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/97100ab3c7b5a03de2641baa0fbafca6/employees.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
24/07/29 22:28:04 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/97100ab3c7b5a03de2641baa0fbafca6/employees.jar
24/07/29 22:28:04 WARN manager.MySQLManager: It looks like you are importing from mysql.
24/07/29 22:28:04 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
24/07/29 22:28:04 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
24/07/29 22:28:04 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
24/07/29 22:28:04 INFO mapreduce.ImportJobBase: Beginning import of employees
24/07/29 22:28:04 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
24/07/29 22:28:04 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
24/07/29 22:28:05 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
24/07/29 22:28:05 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
24/07/29 22:28:12 INFO db.DBInputFormat: Using read committed transaction isolation
24/07/29 22:28:12 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN(`id`), MAX(`id`) FROM `employees`
24/07/29 22:28:12 INFO db.IntegerSplitter: Split size: 0; Num splits: 4 from: 1 to: 3
24/07/29 22:28:12 INFO mapreduce.JobSubmitter: number of splits:3
24/07/29 22:28:13 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1721668005163_0001
24/07/29 22:28:13 INFO impl.YarnClientImpl: Submitted application application_1721668005163_0001
24/07/29 22:28:13 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1721668005163_0001/
24/07/29 22:28:13 INFO mapreduce.Job: Running job: job_1721668005163_0001
24/07/29 22:28:24 INFO mapreduce.Job: Job job_1721668005163_0001 running in uber mode : false
24/07/29 22:28:24 INFO mapreduce.Job: map 0% reduce 0%
24/07/29 22:28:36 INFO mapreduce.Job: map 100% reduce 0%
24/07/29 22:28:38 INFO mapreduce.Job: Job job_1721668005163_0001 completed successfully
24/07/29 22:28:38 INFO mapreduce.Job: Counters: 30
File System Counters
```

**Step 5:** Check whether the table is successfully imported to hadoop.

**Commands:**

```
hadoop fs -ls /user/cloudera/employees
hadoop fs -ls /user/cloudera/employees/part-m-00000 (might be different)
hadoop fs -ls /user/cloudera/employees/part-m-00001 (might be different)
hadoop fs -ls /user/cloudera/employees/part-m-00002 (might be different)
```

## Part 2: Using sqoop to export table from hadoop to mysql.

Now the data is present in hadoop, lets use scoop to export table from hadoop to mysql.

**Step 0:** Drop the table we created in Part 1 from mysql.

**Commands:**

```
mysql -u root -pcloudera
USE db;
DROP TABLE employees;
```

**Step 1:** Create table employees.

**Commands:**

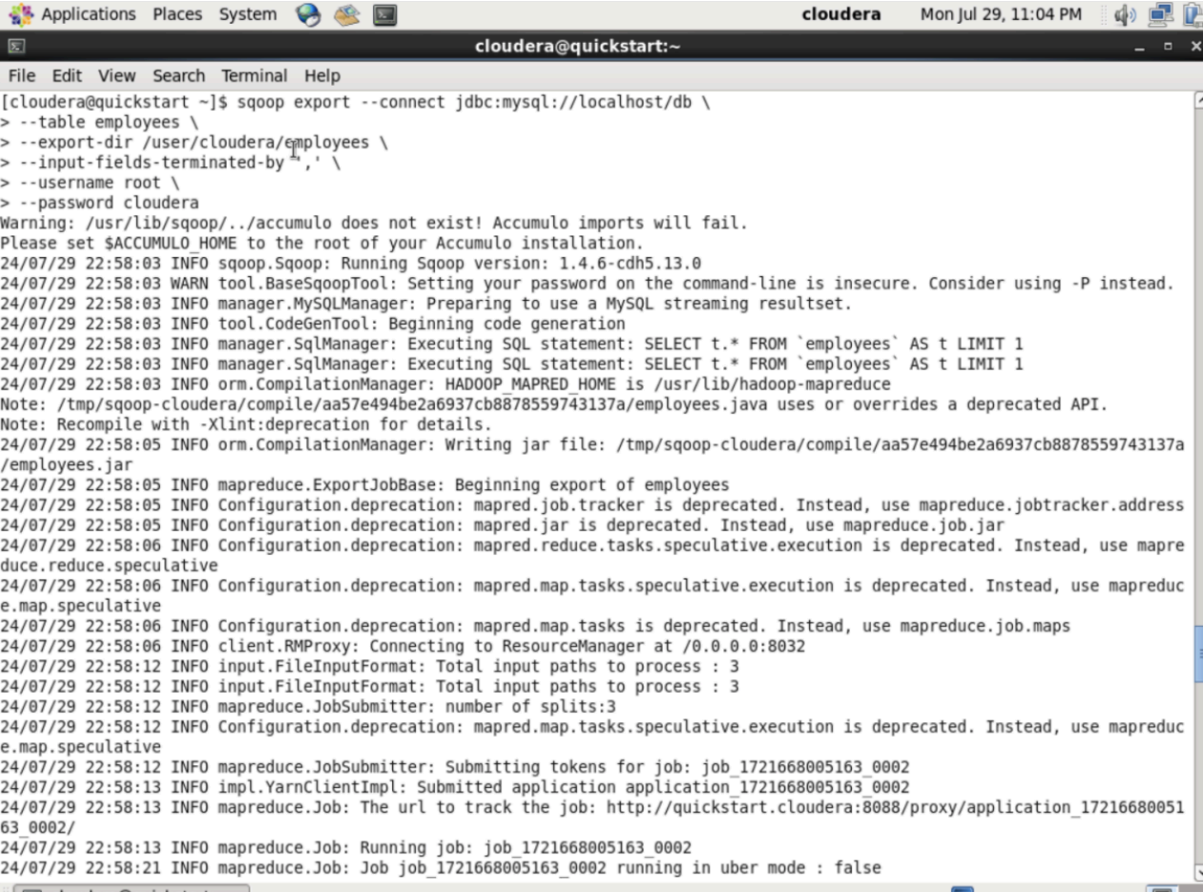
```
CREATE TABLE employees(
id INT NOT NULL PRIMARY KEY;
name VARCHAR(20),
deg VARCHAR(20),
salary INT,
dept VARCHAR(10)
);
EXIT;
```

```
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)
```

**Step 2:** Use sqoop to export table from hadoop to mysql.

**Commands:**

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



```
cloudera@quickstart:~  
[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.  
24/07/29 22:58:03 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6-cdh5.13.0  
24/07/29 22:58:03 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.  
24/07/29 22:58:03 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.  
24/07/29 22:58:03 INFO tool.CodeGenTool: Beginning code generation  
24/07/29 22:58:03 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employees` AS t LIMIT 1  
24/07/29 22:58:03 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employees` AS t LIMIT 1  
24/07/29 22:58:03 INFO orm.CompilationManager: HADOOP MAPRED HOME is /usr/lib/hadoop-mapreduce  
Note: /tmp/sqoop-cloudera/compile/aa57e494be2a6937cb8878559743137a/employees.java uses or overrides a deprecated API.  
Note: Recompile with -Xlint:deprecation for details.  
24/07/29 22:58:05 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/aa57e494be2a6937cb8878559743137a/employees.jar  
24/07/29 22:58:05 INFO mapreduce.ExportJobBase: Beginning export of employees  
24/07/29 22:58:05 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address  
24/07/29 22:58:05 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar  
24/07/29 22:58:06 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative  
24/07/29 22:58:06 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative  
24/07/29 22:58:06 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps  
24/07/29 22:58:06 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032  
24/07/29 22:58:12 INFO input.FileInputFormat: Total input paths to process : 3  
24/07/29 22:58:12 INFO input.FileInputFormat: Total input paths to process : 3  
24/07/29 22:58:12 INFO mapreduce.JobSubmitter: number of splits:3  
24/07/29 22:58:12 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative  
24/07/29 22:58:12 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1721668005163_0002  
24/07/29 22:58:13 INFO impl.YarnClientImpl: Submitted application application_1721668005163_0002  
24/07/29 22:58:13 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1721668005163_0002/  
24/07/29 22:58:13 INFO mapreduce.Job: Running job: job_1721668005163_0002  
24/07/29 22:58:21 INFO mapreduce.Job: Job job_1721668005163_0002 running in uber mode : false
```

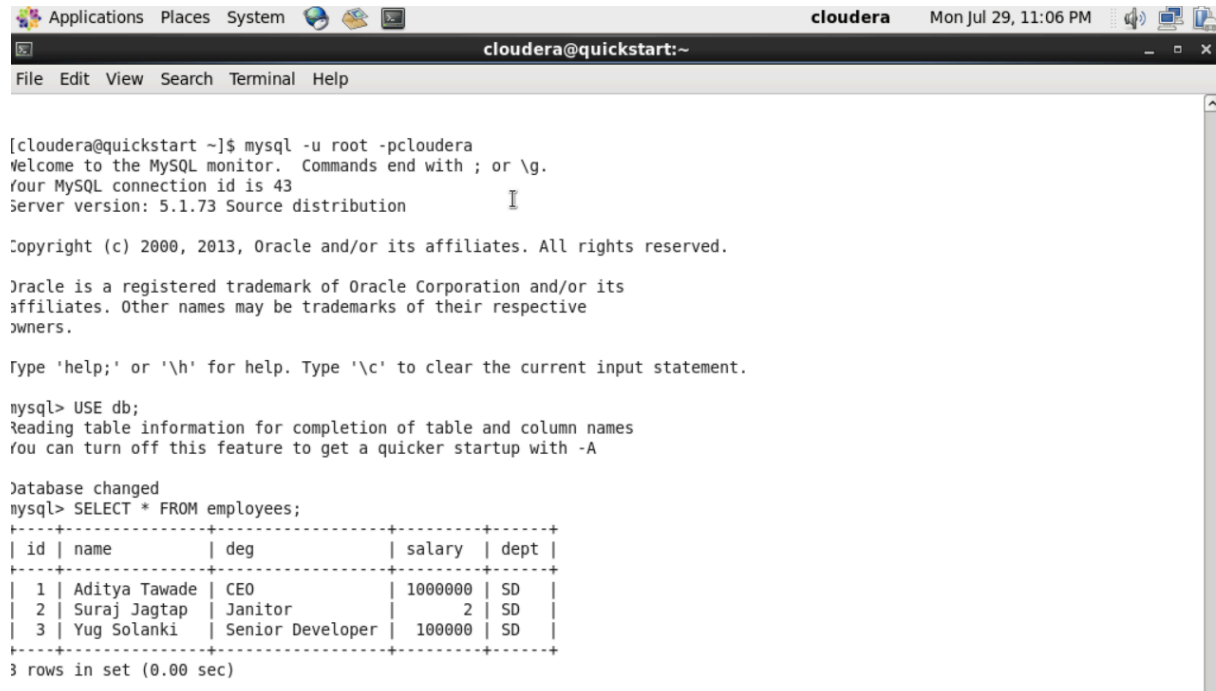
**Step 3:** Login to mysql and check if the table is exported successfully.

Commands:

Mysql -u root -pcloudera

USE db;

SELECT \* FROM employees;



The screenshot shows a terminal window titled 'cloudera@quickstart:~'. The terminal displays the following commands and output:

```
[cloudera@quickstart ~]$ mysql -u root -pcloudera
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 43
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	Aditya Tawade	CEO	1000000	SD
2	Suraj Jagtap	Janitor	2	SD
3	Yug Solanki	Senior Developer	100000	SD

```
3 rows in set (0.00 sec)
```