

# Sqoop tutorials:

## 1.start mysql in cmd:

```
[training@localhost ~]$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 27
Server version: 5.1.66 Source distribution

Copyright (c) 2000, 2012, 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;
```

---

## 2.Show table in sqoop:

```
[training@localhost Desktop]$ sqoop list-tables --connect jdbc:mysql://localhost
/movielens --username training --password training;
23/03/14 14:43:30 WARN tool.BaseSqoopTool: Setting your password on the command-
line is insecure. Consider using -P instead.
23/03/14 14:43:30 INFO manager.MySQLManager: Preparing to use a MySQL streaming
resultset.
emp
genre
movie
moviegenre
movierating
occupation
user
[training@localhost Desktop]$ █
```

---

## Import (RDBMS TO HDFS):

### 1.Show database:

```
mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| hue |
| metastore |
| movielens |
| mysql |
| test |
| training |
+-----+
7 rows in set (0.00 sec)
```

---

I used movielens database:

### 2.Show table in use database:

```
Database changed
mysql> show tables;
+-----+
| Tables_in_movielens |
+-----+
| emp                  |
| genre                |
| movie                |
| moviegenre            |
| movierating           |
| occupation            |
| user                  |
+-----+
7 rows in set (0.00 sec)
```

---

I used movie table:

Show some data in movie table

```
mysql> select * from movie limit 5;
+----+-----+-----+
| id | name                               | year |
+----+-----+-----+
| 1  | Toy Story                          | 1995 |
| 2  | Jumanji                            | 1995 |
| 3  | Grumpier Old Men                    | 1995 |
| 4  | Waiting to Exhale                   | 1995 |
| 5  | Father of the Bride Part II        | 1995 |
+----+-----+-----+
5 rows in set (0.00 sec)

mysql> █
```

---

## **To import data :**

```
[training@localhost Desktop]$ sqoop import --connect jdbc:mysql://localhost/movielens --table movie --fields-terminated-by '\t' --username training --password training;
23/03/14 15:02:24 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/03/14 15:02:24 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
23/03/14 15:02:24 INFO tool.CodeGenTool: Beginning code generation
23/03/14 15:02:25 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `movie` AS t LIMIT 1
23/03/14 15:02:25 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `movie` AS t LIMIT 1
23/03/14 15:02:25 INFO orm.CompilationManager: HADOOP_HOME is /usr/lib/hadoop
Note: /tmp/sqoop-training/compile/d5262e68f67f6b298712c32575557180/movie.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
23/03/14 15:02:28 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-training/compile/d5262e68f67f6b298712c32575557180/movie.jar
23/03/14 15:02:28 WARN manager.MySQLManager: It looks like you are importing from mysql.
23/03/14 15:02:28 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
23/03/14 15:02:28 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
23/03/14 15:02:28 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
23/03/14 15:02:28 INFO mapreduce.ImportJobBase: Beginning import of movie
23/03/14 15:02:30 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
23/03/14 15:02:31 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN(`id`), MAX(`id`) FROM `movie`
23/03/14 15:02:32 INFO mapred.JobClient: Running job: job_202303141318_0004
23/03/14 15:02:33 INFO mapred.JobClient: map 0% reduce 0%
23/03/14 15:02:46 INFO mapred.JobClient: map 50% reduce 0%
23/03/14 15:02:53 INFO mapred.JobClient: map 75% reduce 0%
23/03/14 15:02:54 INFO mapred.JobClient: map 100% reduce 0%
23/03/14 15:02:57 INFO mapred.JobClient: Job complete: job_202303141318_0004
23/03/14 15:02:57 INFO mapred.JobClient: Counters: 23
23/03/14 15:02:57 INFO mapred.JobClient: File System Counters
23/03/14 15:02:57 INFO mapred.JobClient: FILE: Number of bytes read=0
23/03/14 15:02:57 INFO mapred.JobClient: FILE: Number of bytes written=796616
23/03/14 15:02:57 INFO mapred.JobClient: FILE: Number of read operations=0
23/03/14 15:02:57 INFO mapred.JobClient: FILE: Number of large read operations=0
23/03/14 15:02:57 INFO mapred.JobClient: FILE: Number of write operations=0
23/03/14 15:02:57 INFO mapred.JobClient: HDFS: Number of bytes read=412
23/03/14 15:02:57 INFO mapred.JobClient: HDFS: Number of bytes written=102052
23/03/14 15:02:57 INFO mapred.JobClient: HDFS: Number of read operations=4
```

---

```

23/03/14 15:02:57 INFO mapred.JobClient: Job Counters
23/03/14 15:02:57 INFO mapred.JobClient:   Launched map tasks=4
23/03/14 15:02:57 INFO mapred.JobClient:   Total time spent by all maps in occupied slots (ms)=36791
23/03/14 15:02:57 INFO mapred.JobClient:   Total time spent by all reduces in occupied slots (ms)=0
23/03/14 15:02:57 INFO mapred.JobClient:   Total time spent by all maps waiting after reserving slots (ms)=0
23/03/14 15:02:57 INFO mapred.JobClient:   Total time spent by all reduces waiting after reserving slots (ms)=0
23/03/14 15:02:57 INFO mapred.JobClient: Map-Reduce Framework
23/03/14 15:02:57 INFO mapred.JobClient:   Map input records=3881
23/03/14 15:02:57 INFO mapred.JobClient:   Map output records=3881
23/03/14 15:02:57 INFO mapred.JobClient:   Input split bytes=412
23/03/14 15:02:57 INFO mapred.JobClient:   Spilled Records=0
23/03/14 15:02:57 INFO mapred.JobClient:   CPU time spent (ms)=6940
23/03/14 15:02:57 INFO mapred.JobClient:   Physical memory (bytes) snapshot=198975488
23/03/14 15:02:57 INFO mapred.JobClient:   Virtual memory (bytes) snapshot=1554759680
23/03/14 15:02:57 INFO mapred.JobClient:   Total committed heap usage (bytes)=63963136
23/03/14 15:02:57 INFO mapreduce.ImportJobBase: Transferred 0 bytes in 28.6136 seconds (0 bytes/sec)
23/03/14 15:02:57 INFO mapreduce.ImportJobBase: Retrieved 3881 records.
[training@localhost Desktop]$ █

```

---

## **Show folder in hdfs :**

### Movie folder

```

drwxr-xr-x - training supergroup 0 2023-02-13 07:36 averageinput
drwxr-xr-x - training supergroup 0 2023-02-13 07:40 averageoutput
drwxr-xr-x - training supergroup 0 2023-02-11 11:48 distinctinput
drwxr-xr-x - training supergroup 0 2023-02-11 11:50 distinctoutput
drwxr-xr-x - training supergroup 0 2023-02-11 11:39 evenoddinput
drwxr-xr-x - training supergroup 0 2023-02-11 11:41 evenoddoutput
drwxr-xr-x - training supergroup 0 2023-03-07 19:18 hivework1
drwxr-xr-x - training supergroup 0 2023-03-14 15:02 movie
drwxr-xr-x - training supergroup 0 2023-03-07 20:50 part

```

---

```

[training@localhost Desktop]$ hadoop fs -ls movie
Found 6 items
-rw-r--r-- 1 training supergroup 0 2023-03-14 15:02 movie/_SUCCESS
drwxr-xr-x - training supergroup 0 2023-03-14 15:02 movie/_logs
-rw-r--r-- 1 training supergroup 24596 2023-03-14 15:02 movie/part-m-00000
-rw-r--r-- 1 training supergroup 24740 2023-03-14 15:02 movie/part-m-00001
-rw-r--r-- 1 training supergroup 26371 2023-03-14 15:02 movie/part-m-00002
-rw-r--r-- 1 training supergroup 26345 2023-03-14 15:02 movie/part-m-00003
[training@localhost Desktop]$ █

```

---

## EXPORT(HDFS TO RDBMS):

### To create a database:

```
mysql> create table student(id int, firstname varchar(15), lastname varchar(15))
;
Query OK, 0 rows affected (0.07 sec)

mysql> describe table;
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 'table
' at line 1
mysql> describe student;
+-----+-----+-----+-----+-----+-----+
| Field      | Type          | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| id         | int(11)       | YES  |     | NULL    |       |
| firstname  | varchar(15)   | YES  |     | NULL    |       |
| lastname   | varchar(15)   | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
3 rows in set (0.00 sec)
```

### Sample file: in hdfs

```
[training@localhost ~]$ hadoop fs -put sample export
[training@localhost ~]$ hadoop fs -cat export/sample
1,aditya,anal
2,sandeep,kumar
3,rahul,singh
4,ajay,singh
[training@localhost ~]$ █
```

## **Export to rdbs:**

```
[training@localhost aditya]$ sqoop export --connect jdbc:mysql://localhost/aditya --username training --password training --table student --export-dir 'export/sample/' --fields-terminated-by ',';
23/03/14 16:24:29 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/03/14 16:24:29 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
23/03/14 16:24:29 INFO tool.CodeGenTool: Beginning code generation
23/03/14 16:24:30 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `student` AS t LIMIT 1
23/03/14 16:24:30 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `student` AS t LIMIT 1
23/03/14 16:24:30 INFO orm.CompilationManager: HADOOP HOME is /usr/lib/hadoop
Note: /tmp/sqoop-training/compile/cledac7ec51cff9c0d2b3c6c394aaceb/student.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
23/03/14 16:24:33 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-training/compile/cledac7ec51cff9c0d2b3c6c394aaceb/student.jar
23/03/14 16:24:33 INFO mapreduce.ExportJobBase: Beginning export of student
23/03/14 16:24:36 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
23/03/14 16:24:38 INFO input.FileInputFormat: Total input paths to process : 1
23/03/14 16:24:38 INFO input.FileInputFormat: Total input paths to process : 1
23/03/14 16:24:38 INFO mapred.JobClient: Running job: job_202303141318_0008
23/03/14 16:24:39 INFO mapred.JobClient: map 0% reduce 0%
23/03/14 16:24:49 INFO mapred.JobClient: map 100% reduce 0%
23/03/14 16:24:53 INFO mapred.JobClient: Job complete: job_202303141318_0008
23/03/14 16:24:53 INFO mapred.JobClient: Counters: 24
23/03/14 16:24:53 INFO mapred.JobClient:   File System Counters
23/03/14 16:24:53 INFO mapred.JobClient:     FILE: Number of bytes read=0
23/03/14 16:24:53 INFO mapred.JobClient:     FILE: Number of bytes written=197577
23/03/14 16:24:53 INFO mapred.JobClient:     FILE: Number of read operations=0
23/03/14 16:24:53 INFO mapred.JobClient:     FILE: Number of large read operations=0
23/03/14 16:24:53 INFO mapred.JobClient:     FILE: Number of write operations=0
23/03/14 16:24:53 INFO mapred.JobClient:     HDFS: Number of bytes read=172
23/03/14 16:24:53 INFO mapred.JobClient:     HDFS: Number of bytes written=0
23/03/14 16:24:53 INFO mapred.JobClient:     HDFS: Number of read operations=6
23/03/14 16:24:53 INFO mapred.JobClient:     HDFS: Number of large read operations=0
23/03/14 16:24:53 INFO mapred.JobClient:     HDFS: Number of write operations=0
23/03/14 16:24:53 INFO mapred.JobClient:   Job Counters
```

---

```
23/03/14 16:24:39 INFO mapred.JobClient: map 0% reduce 0%
23/03/14 16:24:49 INFO mapred.JobClient: map 100% reduce 0%
23/03/14 16:24:53 INFO mapred.JobClient: Job complete: job_202303141318_0008
23/03/14 16:24:53 INFO mapred.JobClient: Counters: 24
23/03/14 16:24:53 INFO mapred.JobClient:   File System Counters
23/03/14 16:24:53 INFO mapred.JobClient:     FILE: Number of bytes read=0
23/03/14 16:24:53 INFO mapred.JobClient:     FILE: Number of bytes written=197577
23/03/14 16:24:53 INFO mapred.JobClient:     FILE: Number of read operations=0
23/03/14 16:24:53 INFO mapred.JobClient:     FILE: Number of large read operations=0
23/03/14 16:24:53 INFO mapred.JobClient:     FILE: Number of write operations=0
23/03/14 16:24:53 INFO mapred.JobClient:     HDFS: Number of bytes read=172
23/03/14 16:24:53 INFO mapred.JobClient:     HDFS: Number of bytes written=0
23/03/14 16:24:53 INFO mapred.JobClient:     HDFS: Number of read operations=6
23/03/14 16:24:53 INFO mapred.JobClient:     HDFS: Number of large read operations=0
23/03/14 16:24:53 INFO mapred.JobClient:     HDFS: Number of write operations=0
23/03/14 16:24:53 INFO mapred.JobClient:   Job Counters
23/03/14 16:24:53 INFO mapred.JobClient:     Launched map tasks=1
23/03/14 16:24:53 INFO mapred.JobClient:     Data-local map tasks=1
23/03/14 16:24:53 INFO mapred.JobClient:     Total time spent by all maps in occupied slots (ms)=12506
23/03/14 16:24:53 INFO mapred.JobClient:     Total time spent by all reduces in occupied slots (ms)=0
23/03/14 16:24:53 INFO mapred.JobClient:     Total time spent by all maps waiting after reserving slots (ms)=0
23/03/14 16:24:53 INFO mapred.JobClient:     Total time spent by all reduces waiting after reserving slots (ms)=0
23/03/14 16:24:53 INFO mapred.JobClient:   Map-Reduce Framework
23/03/14 16:24:53 INFO mapred.JobClient:     Map input records=4
23/03/14 16:24:53 INFO mapred.JobClient:     Map output records=4
23/03/14 16:24:53 INFO mapred.JobClient:     Input split bytes=109
23/03/14 16:24:53 INFO mapred.JobClient:     Spilled Records=0
23/03/14 16:24:53 INFO mapred.JobClient:     CPU time spent (ms)=1430
23/03/14 16:24:53 INFO mapred.JobClient:     Physical memory (bytes) snapshot=49557504
23/03/14 16:24:53 INFO mapred.JobClient:     Virtual memory (bytes) snapshot=388620288
23/03/14 16:24:53 INFO mapred.JobClient:     Total committed heap usage (bytes)=15990784
23/03/14 16:24:53 INFO mapreduce.ExportJobBase: Transferred 0 bytes in 17.0849 seconds (0 bytes/sec)
23/03/14 16:24:53 INFO mapreduce.ExportJobBase: Exported 4 records.
[training@localhost aditya]$
```

---



## SEE In RDBMS TABLE:

```
mysql> select *from student;
+-----+-----+-----+
| id    | firstname | lastname |
+-----+-----+-----+
| 1     | aditya    | anal     |
| 2     | sandeep   | kumar    |
| 3     | rahul     | singh    |
| 4     | ajay      | singh    |
+-----+-----+-----+
4 rows in set (0.00 sec)

mysql> █
```

---