

# Apache Sqoop Extra Question Solutions

## Apache Sqoop - I

Q1.

```
sqoop import --connect jdbc:mysql://<Public DNS>:3306/movielens --table  
movies --username root --password 123 --target-dir  
/user/root/MovieLens/movies -m 1
```

Q2.

```
sqoop import-all-tables --connect jdbc:mysql://<Public DNS>:3306/movielens  
--username root --password 123 --warehouse-dir  
/user/root/MovieLens/All_tables
```

Q3.

```
sqoop import --connect jdbc:mysql://<Public DNS>:3306/movielens --table  
users --username root --password 123 --target-dir  
/user/root/MovieLens/users_NULL --null-string '\\N' --null-non-string  
'\\\\\\N'
```

Q4.

```
sqoop import --connect jdbc:mysql://<Public DNS>:3306/movielens --table  
genres --username root --password 123 --target-dir  
/user/root/MovieLens/genres --split-by id -m 3
```

Retrieved 18 records

```
[root@ip-172-31-45-59 ~]# hadoop fs -ls /user/root/MovieLens/genres  
Found 4 items  
-rw-r--r--  3 root supergroup          0 2020-05-08 04:45  
/user/root/MovieLens/genres/_SUCCESS  
-rw-r--r--  3 root supergroup        63 2020-05-08 04:45  
/user/root/MovieLens/genres/part-m-00000  
-rw-r--r--  3 root supergroup        66 2020-05-08 04:45  
/user/root/MovieLens/genres/part-m-00001  
-rw-r--r--  3 root supergroup        62 2020-05-08 04:45  
/user/root/MovieLens/genres/part-m-00002
```

Q5.

```
sqoop import --connect jdbc:mysql://<Public DNS>:3306/movielens --table
users --username root --password 123 --target-dir
/user/root/MovieLens/users_parquet -m 1 --as-parquetfile

Retrieved 946 records

[root@ip-172-31-45-59 ~]# hadoop fs -ls /user/root/MovieLens/users_parquet
Found 3 items
drwxr-xr-x  - root supergroup          0 2020-05-08 04:53
/user/root/MovieLens/users_parquet/.metadata
drwxr-xr-x  - root supergroup          0 2020-05-08 04:53
/user/root/MovieLens/users_parquet/.signals
-rw-r--r--   3 root supergroup      12517 2020-05-08 04:53
/user/root/MovieLens/users_parquet/db2c6193-9cb0-4937-a396-46efb3bc76a4.par
quet

sqoop import --connect jdbc:mysql://<Public DNS>:3306/movielens --table
users --username root --password 123 --target-dir
/user/root/MovieLens/users_sequence -m 1 --as-sequencefile

Retrieved 946 records

[root@ip-172-31-45-59~]# hadoop fs -ls /user/root/MovieLens/users_sequence
Found 2 items
-rw-r--r--   3 root supergroup          0 2020-05-08 05:01
/user/root/MovieLens/users_sequence/_SUCCESS
-rw-r--r--   3 root supergroup      39109 2020-05-08 05:01
/user/root/MovieLens/users_sequence/part-m-00000
```

Q6.

```
sqoop import --connect jdbc:mysql://<Public DNS>:3306/movielens --table
movies --username root --password 123 --target-dir
/user/root/MovieLens/movies_normal -m 1

Retrieved 1682 records

[root@ip-172-31-45-59 ~]# hadoop fs -ls /user/root/MovieLens/movies_normal
Found 2 items
-rw-r--r--   3 root supergroup          0 2020-05-08 05:14
```

```
/user/root/MovieLens/movies_normal/_SUCCESS
-rw-r--r--  3 root supergroup      66927 2020-05-08 05:14
/user/root/MovieLens/movies_normal/part-m-00000

sqoop import --connect jdbc:mysql://<Public DNS>:3306/movielens --table
movies --username root --password 123 --target-dir
/user/root/MovieLens/movies_compress -m 1 --compression-codec
org.apache.hadoop.io.compress.SnappyCodec

Retrieved 1682 records

[root@ip-172-31-45-59 ~]# hadoop fs -ls
/user/root/MovieLens/movies_compress
Found 2 items
-rw-r--r--  3 root supergroup      0 2020-05-08 05:16
/user/root/MovieLens/movies_compress/_SUCCESS
-rw-r--r--  3 root supergroup    36993 2020-05-08 05:16
/user/root/MovieLens/movies_compress/part-m-00000.snappy
```

## Apache Sqoop - II

Q1.

```
sqoop import --connect jdbc:mysql://<Public DNS>:3306/movielens --table
genres_movies --username root --password 123 --target-dir
/user/root/MovieLens/genres_movies_1 --where "genre_id=1" -m 1
```

Q2.

```
sqoop import --connect jdbc:mysql://<Public DNS>:3306/movielens --username
root --password 123 --query 'select
users.id,users.occupation_id,occupations.name from users inner join
occupations on users.occupation_id=occupations.id where $CONDITIONS'
--target-dir /user/root/joined_data -m 1
```

Q3.

```
sqoop import --connect jdbc:mysql://<Public DNS>:3306/movielens --table
genres --username root --password 123 --target-dir
/user/root/MovieLens/genres_new --incremental append --check-column id
--last-value 18 -m 1
```

Retrieved 5 records

```
[root@ip-172-31-45-59 ~]# hadoop fs -ls /user/root/MovieLens/genres_new
Found 1 items
-rw-r--r--    3 root supergroup          40 2020-05-08 05:34
/user/root/MovieLens/genres_new/part-m-00000
```

Q4.

```
sqoop job --create genres_import -- import --connect jdbc:mysql://<Public
DNS>:3306/movielens --table genres_movies --username root --password-file
/user/root/pass.txt --where "genre_id=2" --target-dir
/user/root/MovieLens/genres_movies_job -m 1
```

```
[root@ip-172-31-45-59 ~]# sqoop job --list
```

Available jobs:

```
  flightsimport
  genres_import
  importjob
  myimport
sqoop job --exec genres_import
```

Retrieved 135 records