



---

# SQOOP ASSIGNMENT-3

---

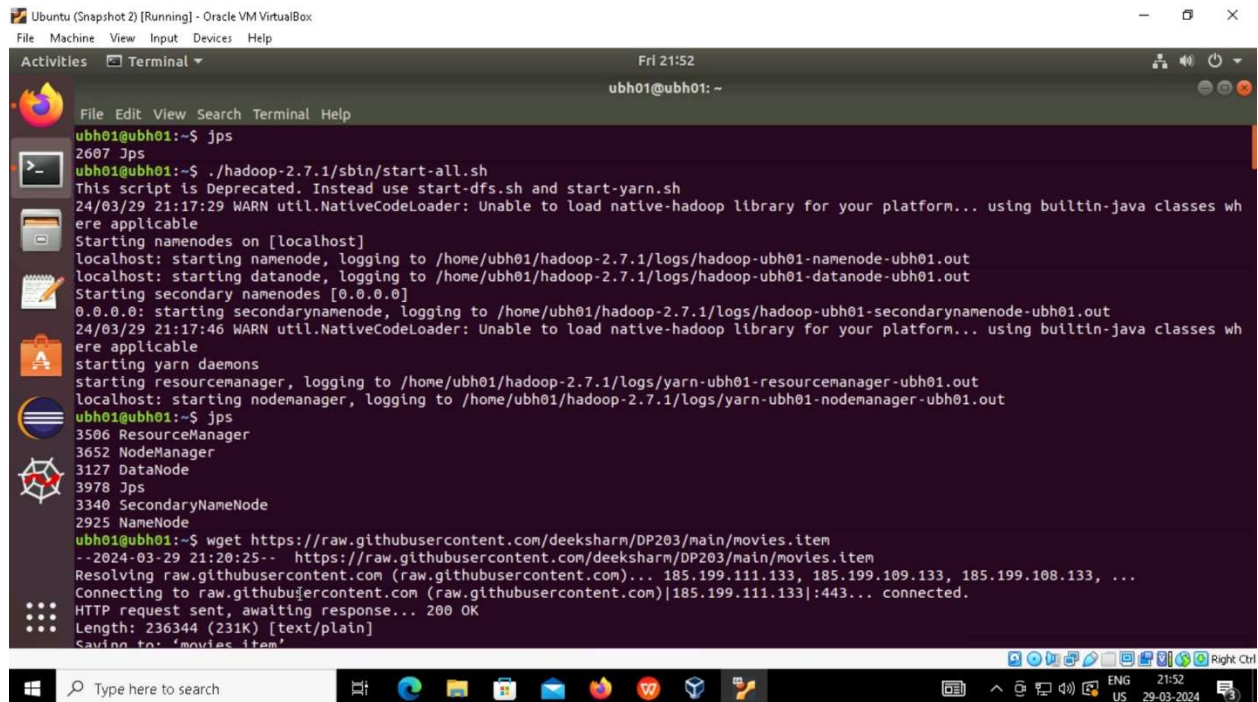


A. JOSHI SATYA VARDAN

CSDAIA24AZ003

EMP ID:2320097

## Importing from mysql to Hadoop BY USING SQOOP:



The screenshot shows a terminal window titled "Ubuntu (Snapshot 2) [Running] - Oracle VM VirtualBox". The terminal output shows the following commands and results:

```
ubh01@ubh01:~$ jps
2607 Jps
ubh01@ubh01:~$ ./hadoop-2.7.1/sbin/start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
24/03/29 21:17:29 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wh
ere applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/ubh01/hadoop-2.7.1/logs/hadoop-ubh01-namenode-ubh01.out
localhost: starting datanode, logging to /home/ubh01/hadoop-2.7.1/logs/hadoop-ubh01-datanode-ubh01.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/ubh01/hadoop-2.7.1/logs/hadoop-ubh01-secondarynamenode-ubh01.out
24/03/29 21:17:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes wh
ere applicable
starting yarn daemons
starting resourcemanager, logging to /home/ubh01/hadoop-2.7.1/logs/yarn-ubh01-resourcemanager-ubh01.out
localhost: starting nodemanager, logging to /home/ubh01/hadoop-2.7.1/logs/yarn-ubh01-nodemanager-ubh01.out
ubh01@ubh01:~$ jps
3506 ResourceManager
3652 NodeManager
3127 DataNode
3978 Jps
3340 SecondaryNameNode
2925 NameNode
ubh01@ubh01:~$ wget https://raw.githubusercontent.com/deeksharm/DP203/main/movies.item
--2024-03-29 21:20:25-- https://raw.githubusercontent.com/deeksharm/DP203/main/movies.item
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.111.133, 185.199.109.133, 185.199.108.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.111.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 236344 (231K) [text/plain]
Saving to: 'movies.item'
```

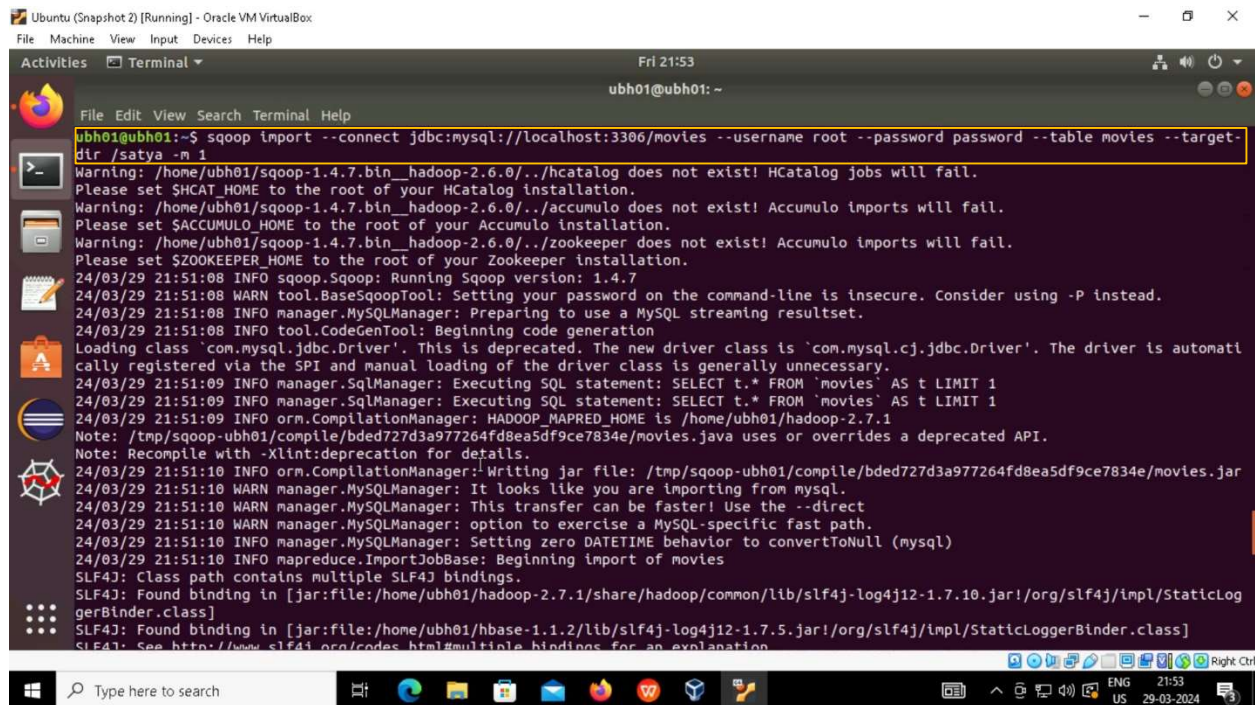
- Check whether data node and name node is on or not by using command “jps”.
- By using “/hadoop-2.7.1/sbin/start-all.sh” we are making data node and name node start.
- By using “wget <url>” fetch movies.item file from GitHub .



```
CREATE TABLE Movies (  
  movie_id INT PRIMARY KEY,  
  movie_title VARCHAR(255),  
  release_date INT,  
  video_release_date VARCHAR(255),  
  IMDb_URL VARCHAR(255),  
  unknown INT,  
  Action INT,  
  Adventure INT,  
  Animation INT,  
  Childrens INT,  
  Comedy INT,  
  Crime INT,  
  Documentary INT,  
  Drama INT,  
  Fantasy INT,  
  Film_Noir INT,  
  Horror INT,  
  Musical INT,  
  Mystery INT,  
  Romance INT,  
  Sci_Fi INT,  
  Thriller INT,  
  War INT,  
  Western INT  
);
```

- By using command

```
“ LOAD DATA INFILE 'movies.item'  
  INTO TABLE Movies  
  FIELDS TERMINATED BY '|' '  
  LINES TERMINATED BY '\n' ”;
```



```
Ubuntu (Snapshot 2) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal
Fri 21:53
ubh01@ubh01: ~

File Edit View Search Terminal Help
ubh01@ubh01:~$ sqoop import --connect jdbc:mysql://localhost:3306/movies --username root --password password --table movies --target-dir /satya -m 1
Warning: /home/ubh01/sqoop-1.4.7.bin__hadoop-2.6.0/./hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /home/ubh01/sqoop-1.4.7.bin__hadoop-2.6.0/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /home/ubh01/sqoop-1.4.7.bin__hadoop-2.6.0/./zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
24/03/29 21:51:08 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
24/03/29 21:51:08 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
24/03/29 21:51:08 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
24/03/29 21:51:08 INFO tool.CodeGenTool: Beginning code generation
Loading class 'com.mysql.jdbc.Driver'. This is deprecated. The new driver class is 'com.mysql.cj.jdbc.Driver'. The driver is automatically registered via the SPI and manual loading of the driver class is generally unnecessary.
24/03/29 21:51:09 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'movies' AS t LIMIT 1
24/03/29 21:51:09 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'movies' AS t LIMIT 1
24/03/29 21:51:09 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /home/ubh01/hadoop-2.7.1
Note: /tmp/sqoop-ubh01/compile/bded727d3a977264fd8ea5df9ce7834e/movies.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
24/03/29 21:51:10 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-ubh01/compile/bded727d3a977264fd8ea5df9ce7834e/movies.jar
24/03/29 21:51:10 WARN manager.MySQLManager: It looks like you are importing from mysql.
24/03/29 21:51:10 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
24/03/29 21:51:10 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
24/03/29 21:51:10 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
24/03/29 21:51:10 INFO mapreduce.ImportJobBase: Beginning import of movies
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/ubh01/hadoop-2.7.1/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/ubh01/hbase-1.1.2/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
```

- By using command

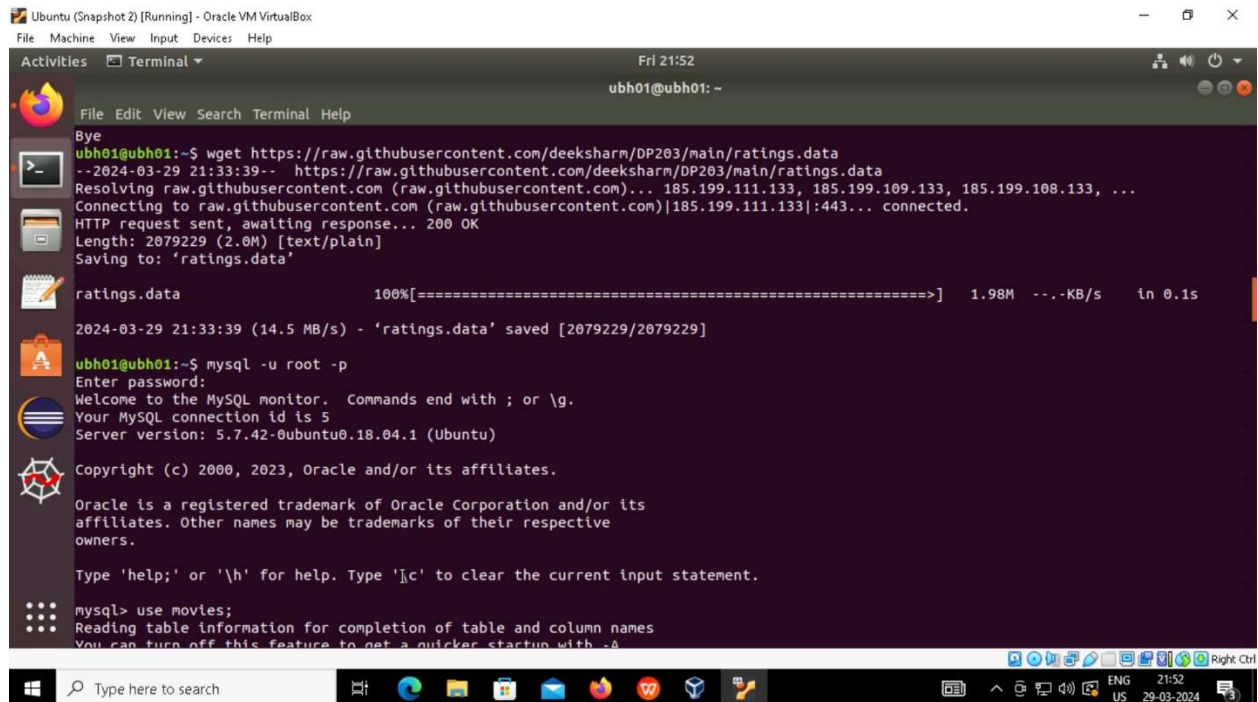
**“sqoop import --connect jdbc:mysql://your\_mysql\_host:port/movies  
--username root  
--password password  
--table Movies  
--target-dir movies.item  
--m 1”**

importing data from mysql to hdfs by using sqoop.





## Export from Hadoop to MySQL BY USING SQOOP:



```
Ubuntu (Snapshot 2) [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

Activities Terminal Fri 21:52
ubh01@ubh01: ~

File Edit View Search Terminal Help
Bye
ubh01@ubh01:~$ wget https://raw.githubusercontent.com/deeksharm/DP203/main/ratings.data
--2024-03-29 21:33:39-- https://raw.githubusercontent.com/deeksharm/DP203/main/ratings.data
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.111.133, 185.199.109.133, 185.199.108.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.111.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2079229 (2.0M) [text/plain]
Saving to: 'ratings.data'

ratings.data                               100%[=====] 1.98M  ---KB/s  in 0.1s

2024-03-29 21:33:39 (14.5 MB/s) - 'ratings.data' saved [2079229/2079229]

ubh01@ubh01:~$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 5
Server version: 5.7.42-0ubuntu0.18.04.1 (Ubuntu)

Copyright (c) 2000, 2023, Oracle and/or its affiliates.

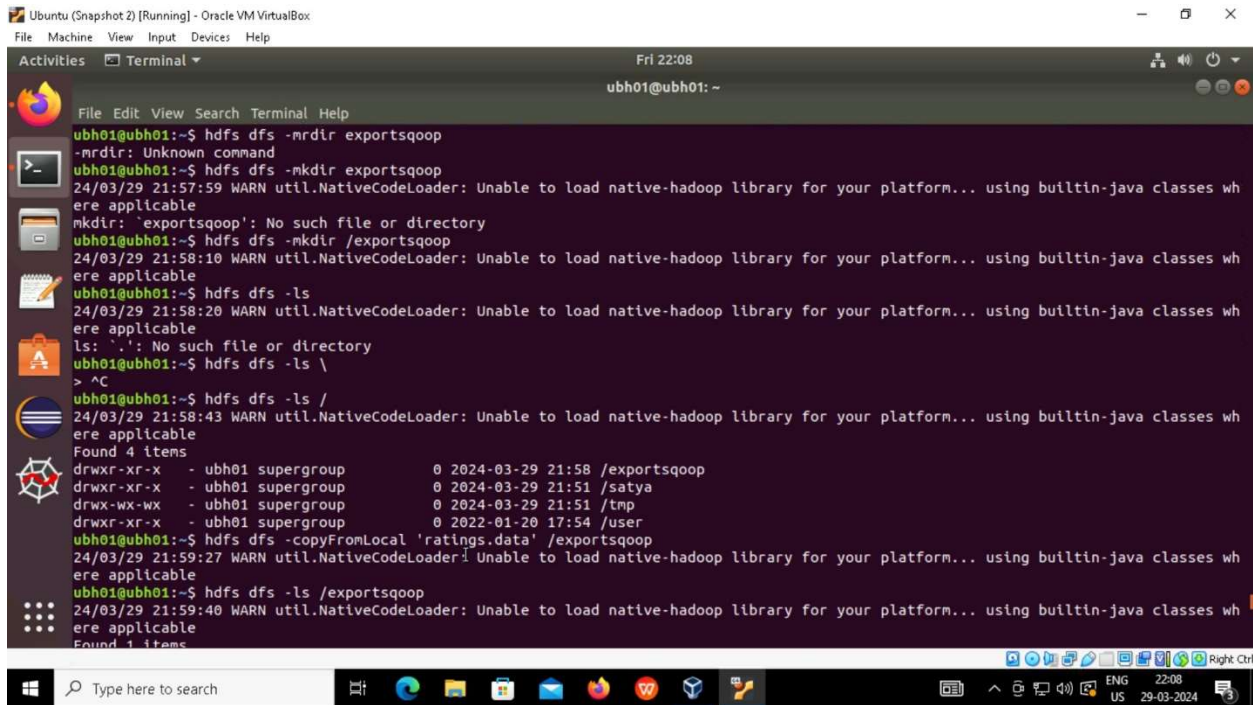
Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> use movies;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A
```

- By using “**wget <url>**” we are fetching ratings.data file from github.
- By using “**mysql -u root -p password**” we are going to enter into mysql.
- By creating table ratings

```
“create table(
user id int,
item id int,
rating int,
timestamp varchar(255)
);”
```



Ubuntu (Snapshot 2) [Running] - Oracle VM VirtualBox

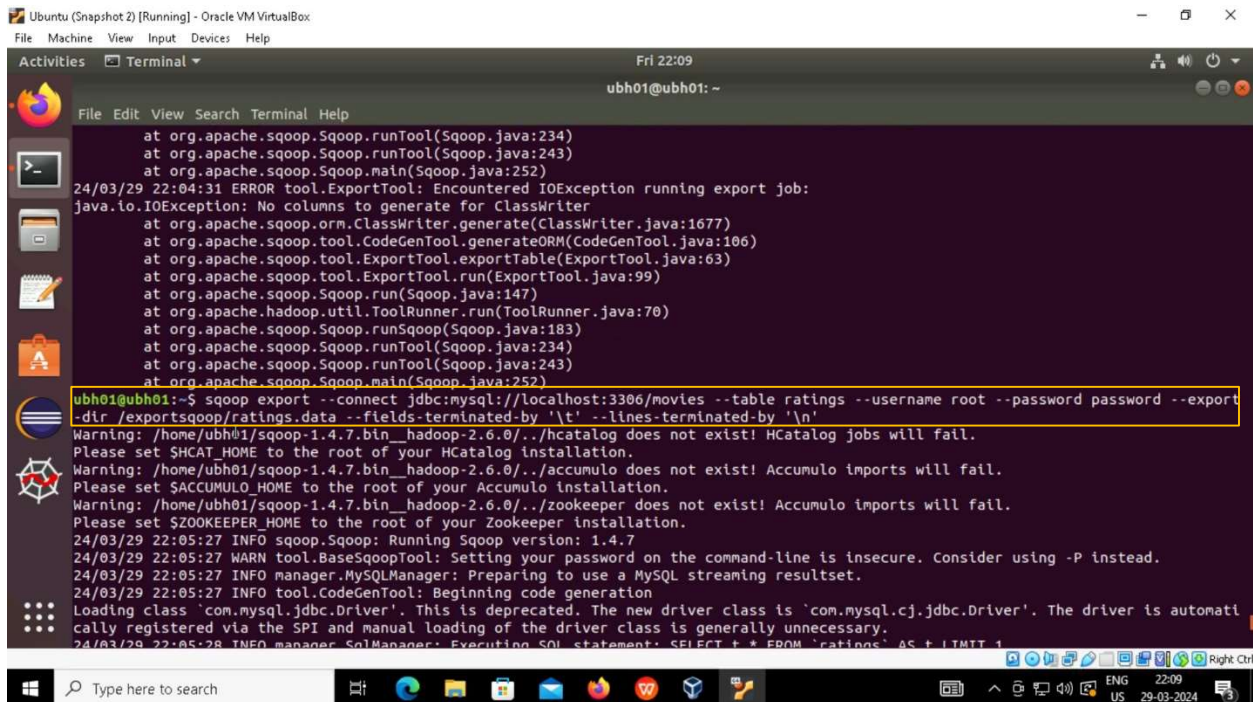
File Machine View Input Devices Help

Activities Fri 22:08

ubh01@ubh01: ~

```
File Edit View Search Terminal Help
ubh01@ubh01:~$ hdfs dfs -mrdir exportsqoop
-mrdir: Unknown command
ubh01@ubh01:~$ hdfs dfs -mkdir exportsqoop
24/03/29 21:57:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
mkdir: 'exportsqoop': No such file or directory
ubh01@ubh01:~$ hdfs dfs -mkdir /exportsqoop
24/03/29 21:58:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ubh01@ubh01:~$ hdfs dfs -ls
24/03/29 21:58:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ls: '.': No such file or directory
ubh01@ubh01:~$ hdfs dfs -ls \
> ^C
ubh01@ubh01:~$ hdfs dfs -ls /
24/03/29 21:58:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 4 items
drwxr-xr-x - ubh01 supergroup          0 2024-03-29 21:58 /exportsqoop
drwxr-xr-x - ubh01 supergroup          0 2024-03-29 21:51 /satya
drwx-wx-wx - ubh01 supergroup          0 2024-03-29 21:51 /tmp
drwxr-xr-x - ubh01 supergroup          0 2022-01-20 17:54 /user
ubh01@ubh01:~$ hdfs dfs -copyFromLocal 'ratings.data' /exportsqoop
24/03/29 21:59:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
ubh01@ubh01:~$ hdfs dfs -ls /exportsqoop
24/03/29 21:59:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
```

- creating directory in hadoop “**hdfs dfs -mkdir <exportsqoop>**”.
- Load data from local into hdfs by using “**copyFromLocal**” command.
- Checking file get inserted into created directory or not “**hdfs dfs -ls /exportsqoop**”



Ubuntu (Snapshot 2) [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Activities Fri 22:09

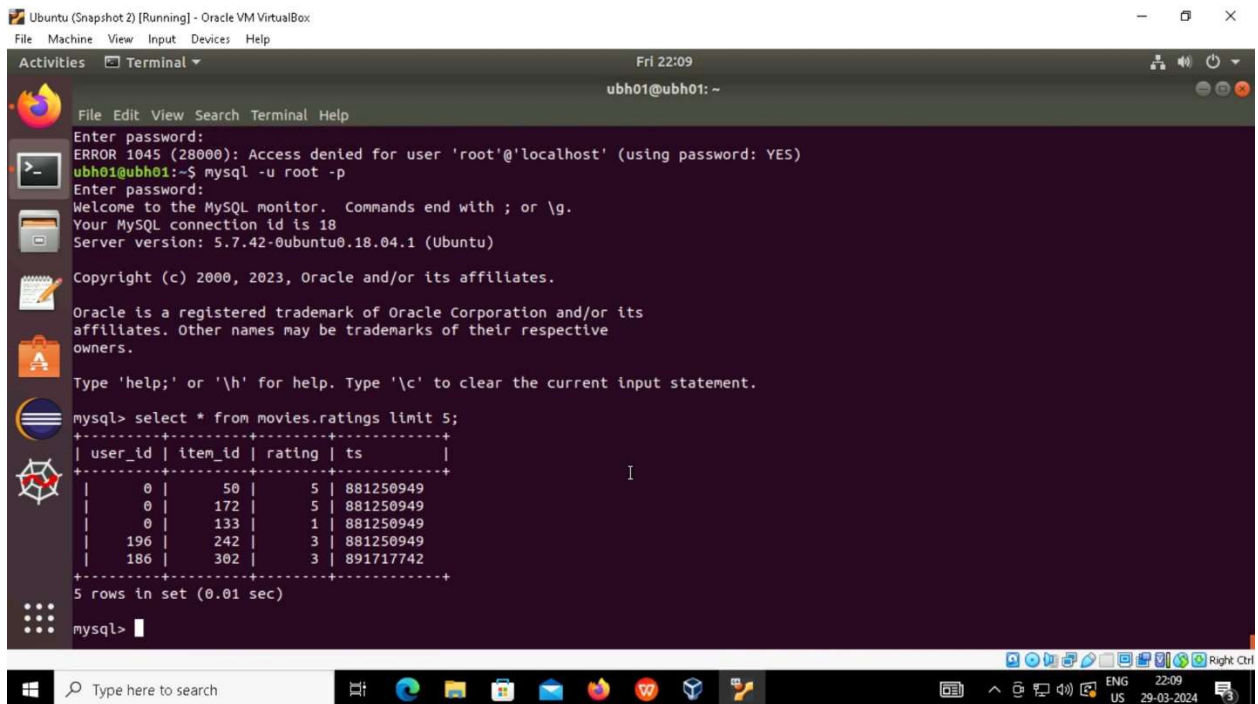
ubh01@ubh01: ~

```
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)
24/03/29 22:04:31 ERROR tool.ExportTool: Encountered IOException running export job:
java.io.IOException: No columns to generate for ClassWriter
at org.apache.sqoop.orm.ClassWriter.generate(ClassWriter.java:1677)
at org.apache.sqoop.tool.CodeGenTool.generateORM(CodeGenTool.java:106)
at org.apache.sqoop.tool.ExportTool.exportTable(ExportTool.java:63)
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.main(Sqoop.java:252)
ubh01@ubh01:~$ sqoop export --connect jdbc:mysql://localhost:3306/movies --table ratings --username root --password password --export
-dir /exportsqoop/ratings.data --fields-terminated-by '\t' --lines-terminated-by '\n'
Warning: /home/ubh01/sqoop-1.4.7.bin__hadoop-2.6.0/./hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /home/ubh01/sqoop-1.4.7.bin__hadoop-2.6.0/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /home/ubh01/sqoop-1.4.7.bin__hadoop-2.6.0/./zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
24/03/29 22:05:27 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
24/03/29 22:05:27 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
24/03/29 22:05:27 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
24/03/29 22:05:27 INFO tool.CodeGenTool: Beginning code generation
Loading class 'com.mysql.jdbc.Driver'. This is deprecated. The new driver class is 'com.mysql.cj.jdbc.Driver'. The driver is automatically
registered via the SPI and manual loading of the driver class is generally unnecessary.
24/03/29 22:05:28 INFO manager.SqlManager: Executing SQL statement: SELECT * FROM 'ratings' AS t LIMIT 1
```

- Export command by using sqoop



**--scoop export --connect jdbc:mysql://your\_mysql\_host:port/movies**  
**--username root**  
**--password password**  
**--table ratings**  
**--export-dir /exportscoop/ratings.data**  
**--fields-terminated-by '\t'**  
**--lines-terminated-by '\n' "**



```
Enter password:
ERROR 1045 (28000): Access denied for user 'root'@'localhost' (using password: YES)
ubh01@ubh01:~$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 18
Server version: 5.7.42-0ubuntu0.18.04.1 (Ubuntu)

Copyright (c) 2000, 2023, Oracle and/or its affiliates.

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> select * from movies.ratings limit 5;
+-----+-----+-----+-----+
| user_id | item_id | rating | ts          |
+-----+-----+-----+-----+
| 0       | 50      | 5      | 881250949   |
| 0       | 172     | 5      | 881250949   |
| 0       | 133     | 1      | 881250949   |
| 196     | 242     | 3      | 881250949   |
| 186     | 302     | 3      | 891717742   |
+-----+-----+-----+-----+
5 rows in set (0.01 sec)

mysql>
```

- Finally checking data whether insert into mysql ratings table by using **“select \* from movies.ratings limit 5;”**