**Sqoop Import and Export**

Project 1

**Authored By**

Divya Jayaprakash (dxj160830)

BUAN 6346 – Big Data Analytics – Fall 2017

Prof. Kashif Saeed

TA - Karthik

**Part A: 1. Import table ‘accounts’ into a directory named as your NET-ID/loudacre, without creating sub directories**

**sqoop import \**

**--connect jdbc:mysql://localhost/loudacre \**

**--username training \**

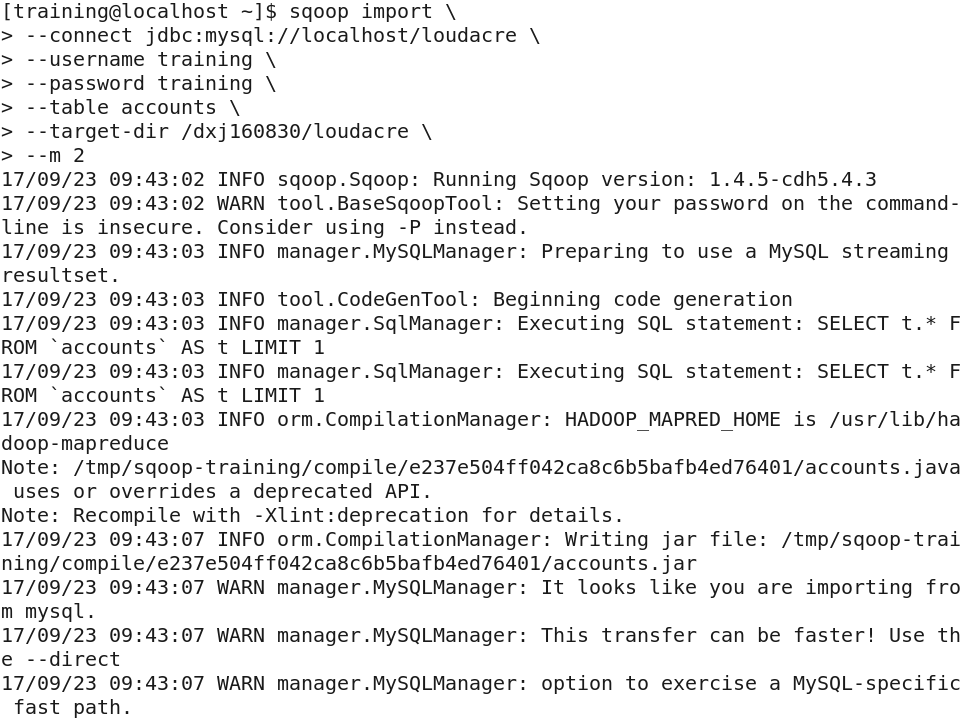
**--password training \**

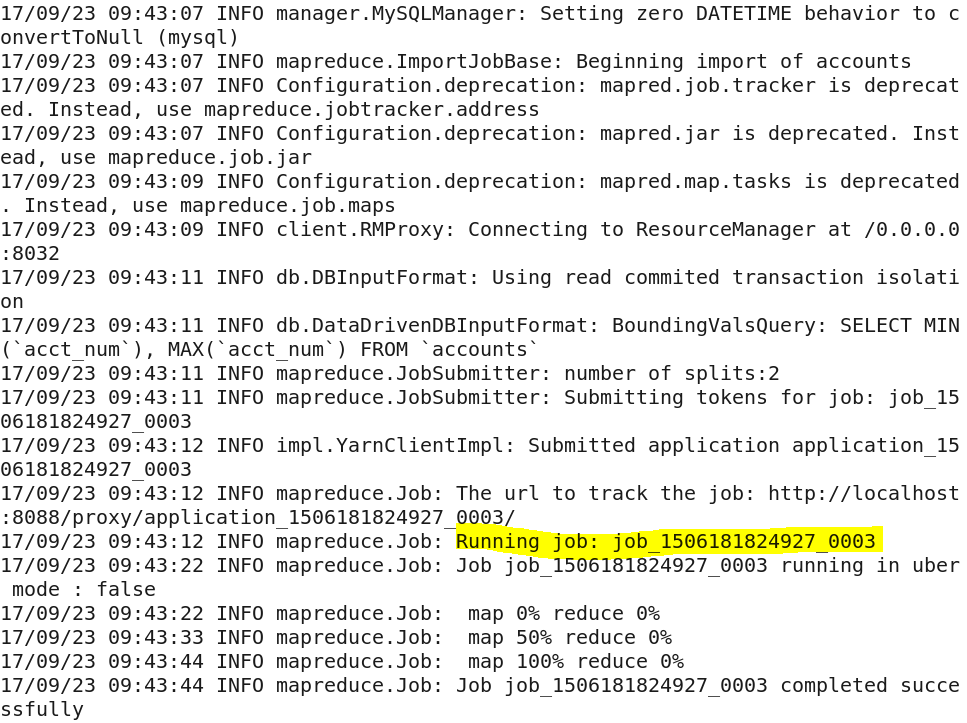
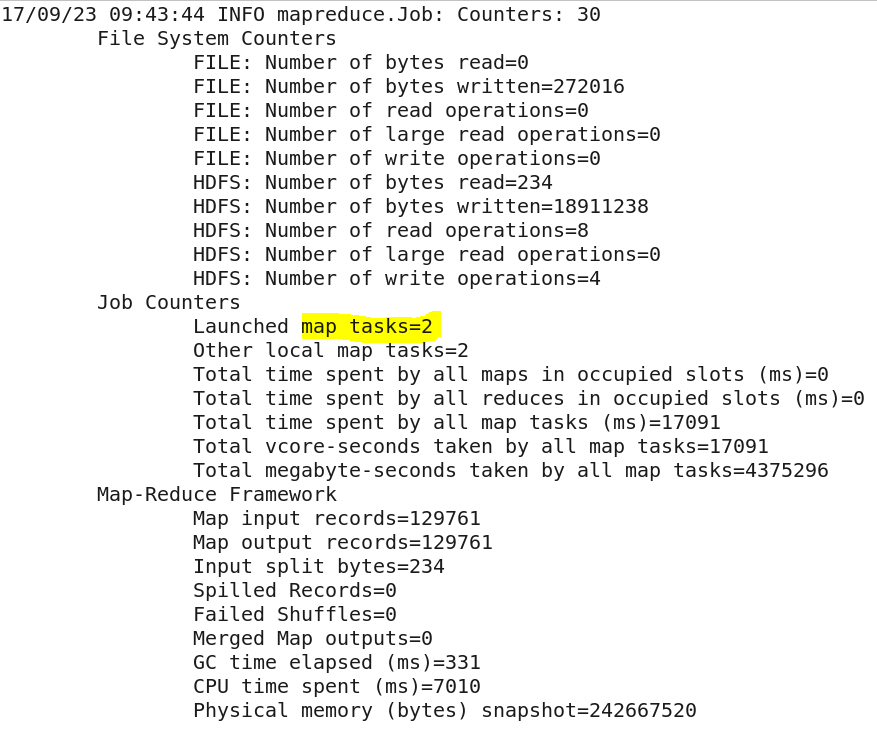
**--table accounts \**

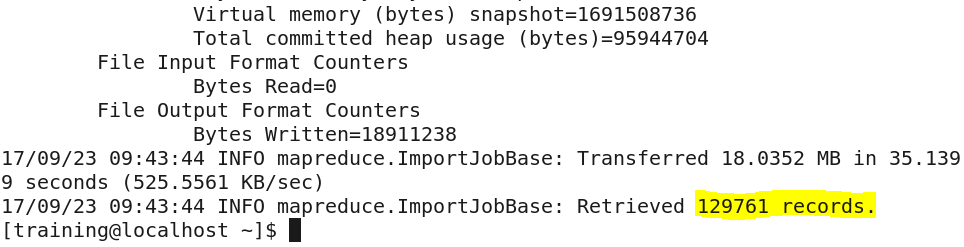
**--target-dir /dxj160830/loudacre \**

**--m 2**

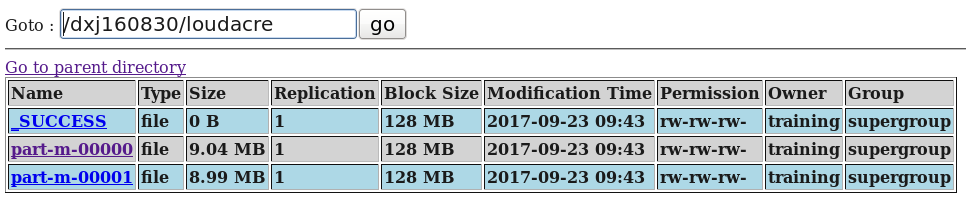
**Sqoop:**



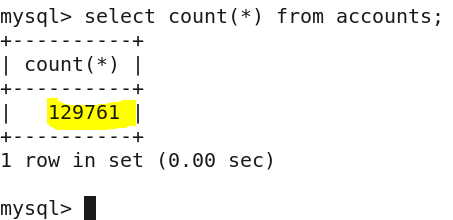
 



**Output:**



**Verification: mysql**



**Answer:**

* Tool Name: import
* Tool Options: table, target-dir
* To import the table accounts into the directory /dxj160830/loudacre **without creating sub directories,** I used target-dir tool option
* The number of records in mysql and hdfs are matching
* Trials by changing the number of mappers:

m 2 = 33 seconds

m 1 = 63 seconds

The import was faster when two mappers was used

**Part A: 2. Import table ‘accounts’ into hive with the following conditions:**

**a. Nulls are represented as /N**

**b. Fields terminated by ‘,’**

**c. Accounts that are active and from the state of California(CA)**

**sqoop import**

**--connect jdbc:mysql://localhost/loudacre \**

**--username training \**

**--password training \**

**--table accounts \**

**--hive-import \**

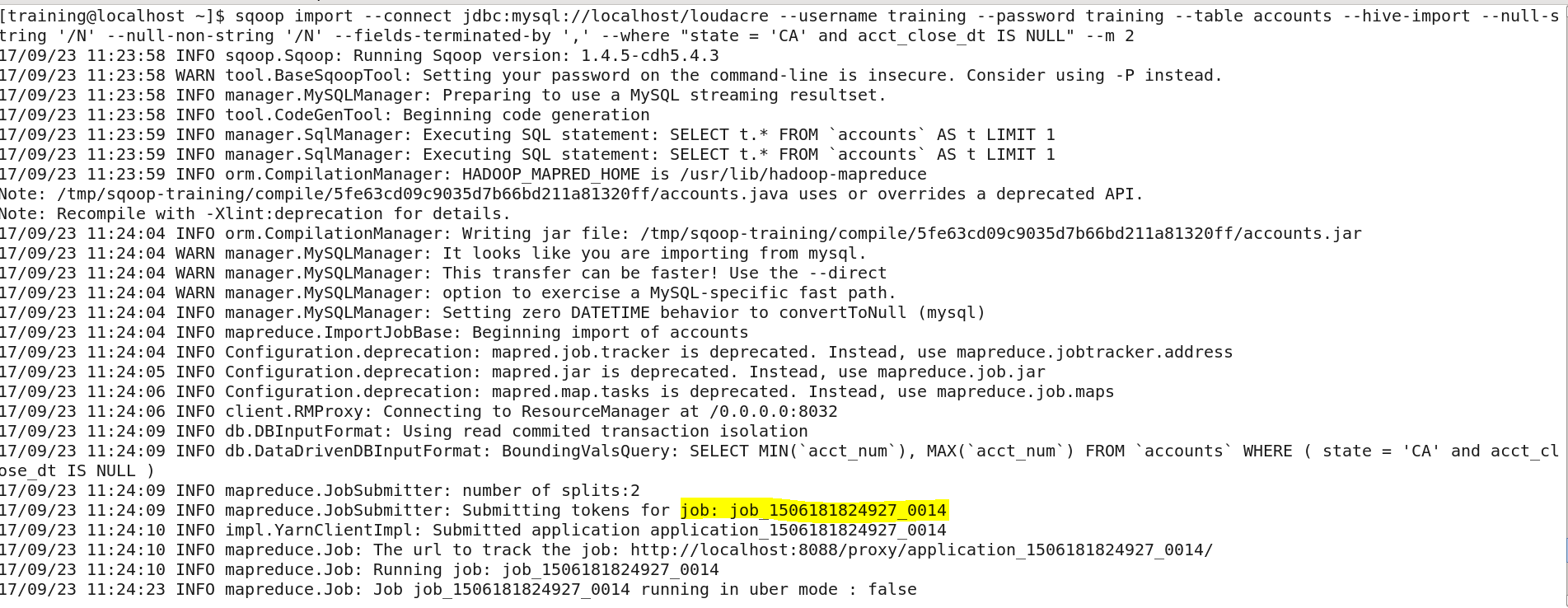
**--null-string '/N' \**

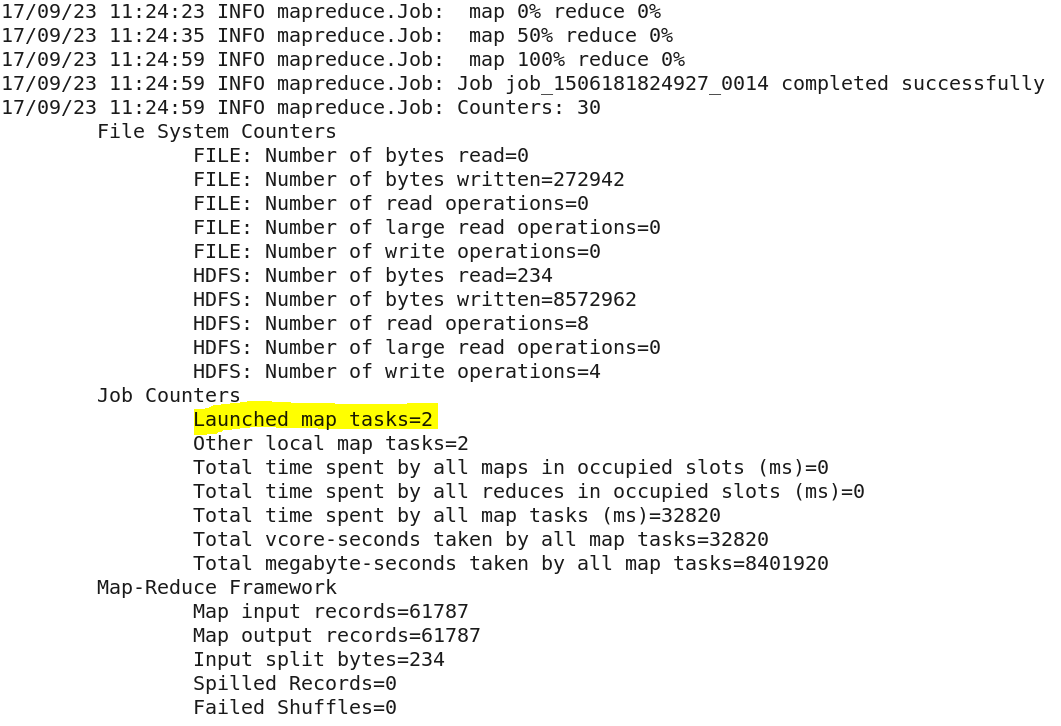
**--null-non-string '/N' \**

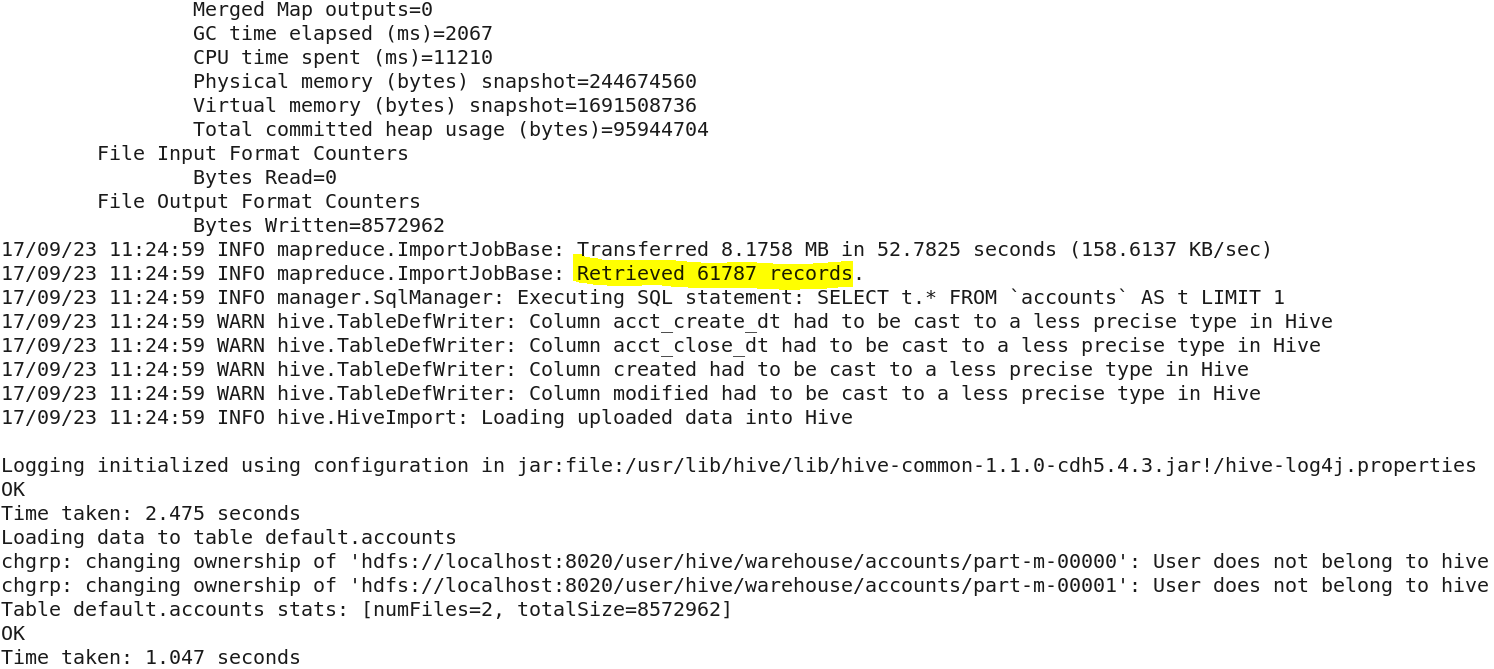
**--fields-terminated-by ',' \**

**--where "state = 'CA' and acct\_close\_dt IS NULL" \**

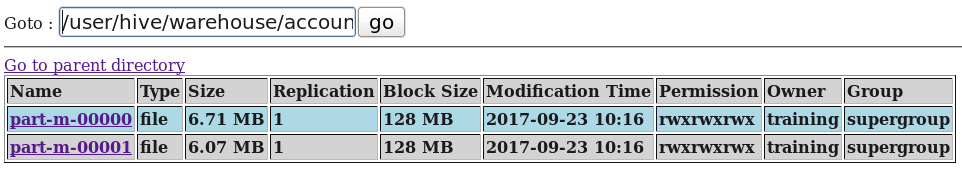
**--m 2**

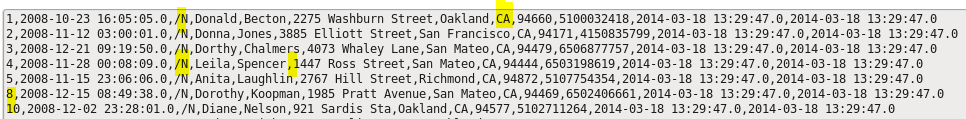






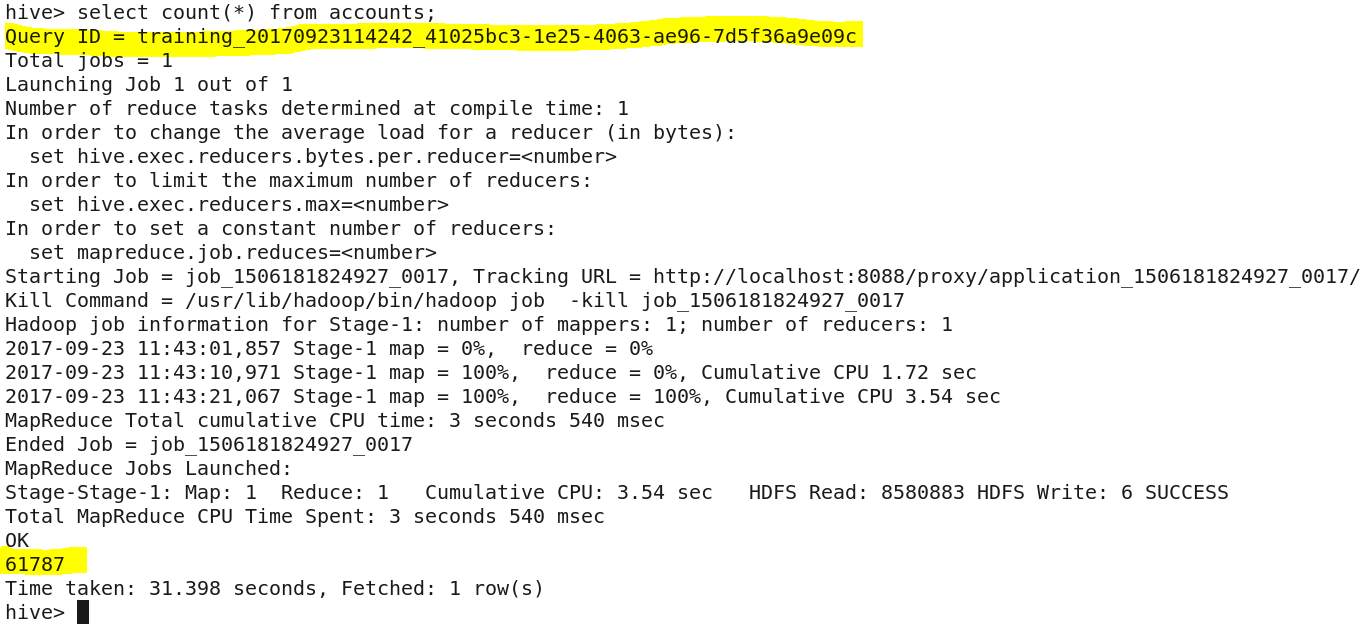
**Hive Output:**



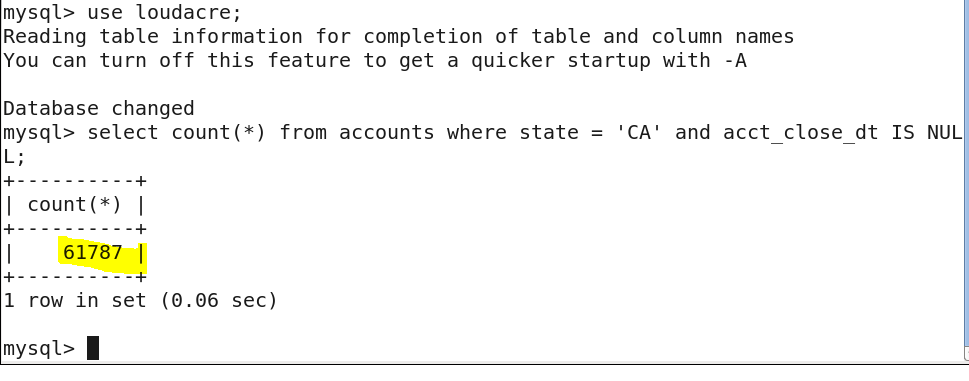


**Verification:**

**Hive:**



**MySQL:**



**Answer:**

The number of rows in mysql and hive are now matching.

**Part A: 3. Export the contents from hive into a new MySQL table called ‘Cali-‘ followed by your NET-ID a. Do not import the following columns: i. acct\_close\_dt ii. modified iii. created**

**In HIVE:**

**CREATE TABLE accounts\_lesscolumns(**

**acct\_num int,**

**acct\_create\_dt timestamp,**

**first\_name string,**

**last\_name string,**

**address string,**

**city string,**

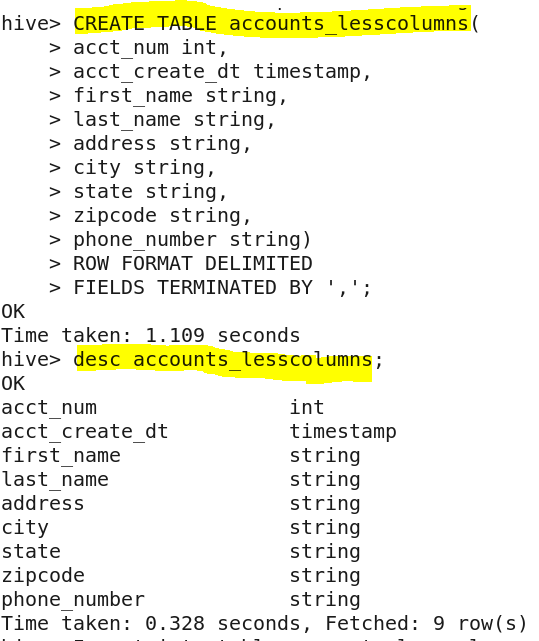
**state string,**

**zipcode string,**

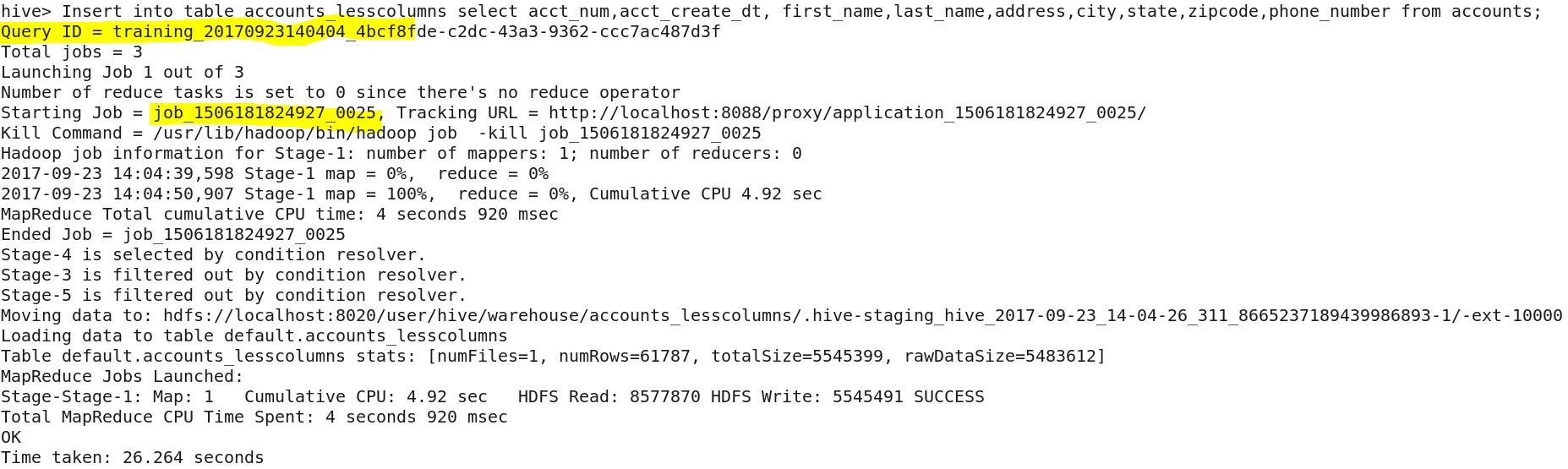
**phone\_number string)**

**ROW FORMAT DELIMITED**

**FIELDS TERMINATED BY ',';**

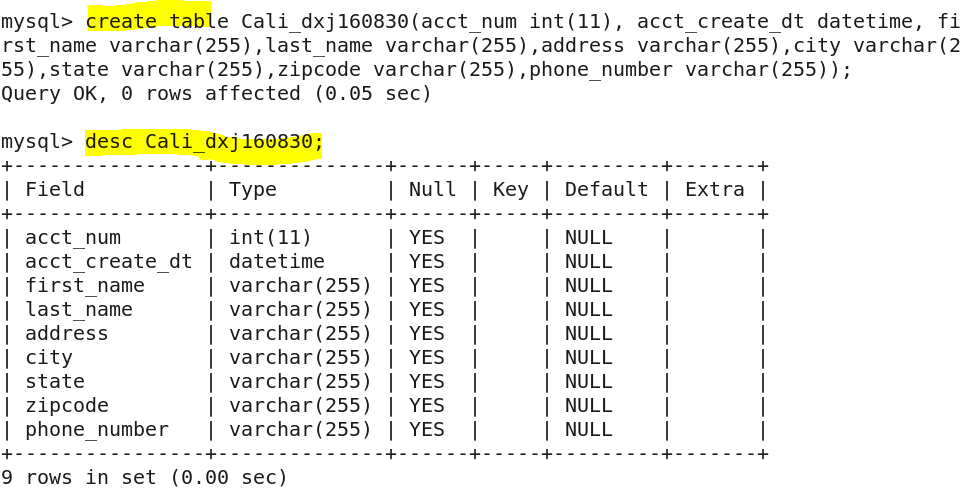


**Insert into table accounts\_lesscolumns select acct\_num,acct\_create\_dt, first\_name,last\_name,address,city,state,zipcode,phone\_number from accounts;**



**In Mysql:**

**create table Cali\_dxj160830(acct\_num int(11), acct\_create\_dt datetime, first\_name varchar(255),last\_name varchar(255),address varchar(255),city varchar(255),state varchar(255),zipcode varchar(255),phone\_number varchar(255));**



**sqoop export \**

**--connect jdbc:mysql://localhost/loudacre \**

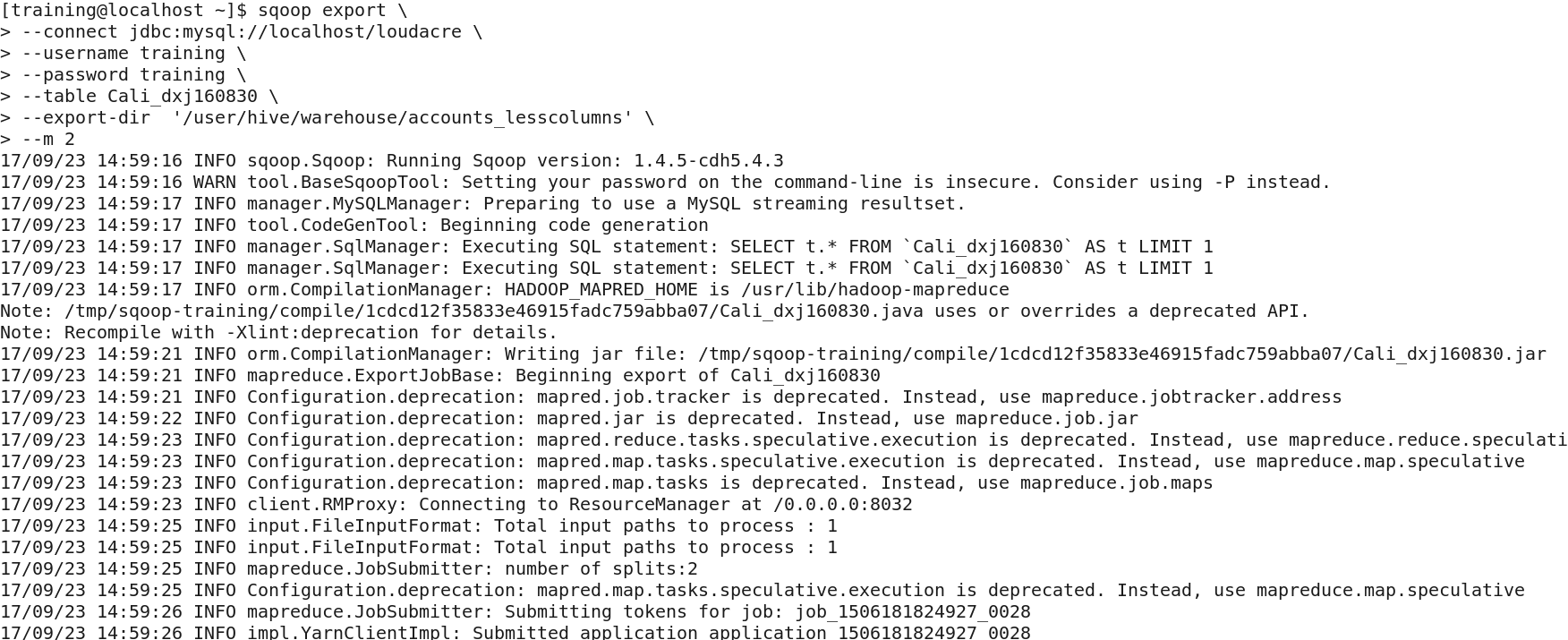
**--username training \**

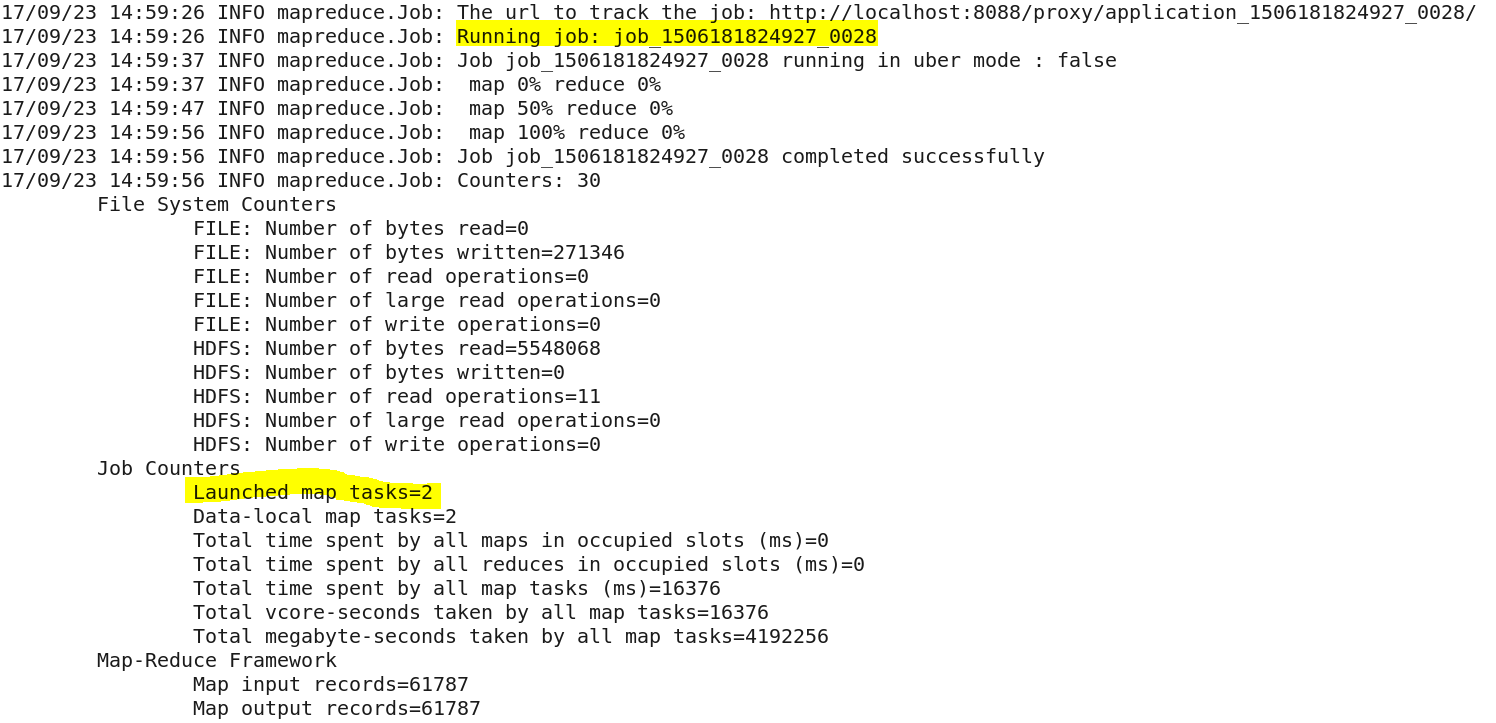
**--password training \**

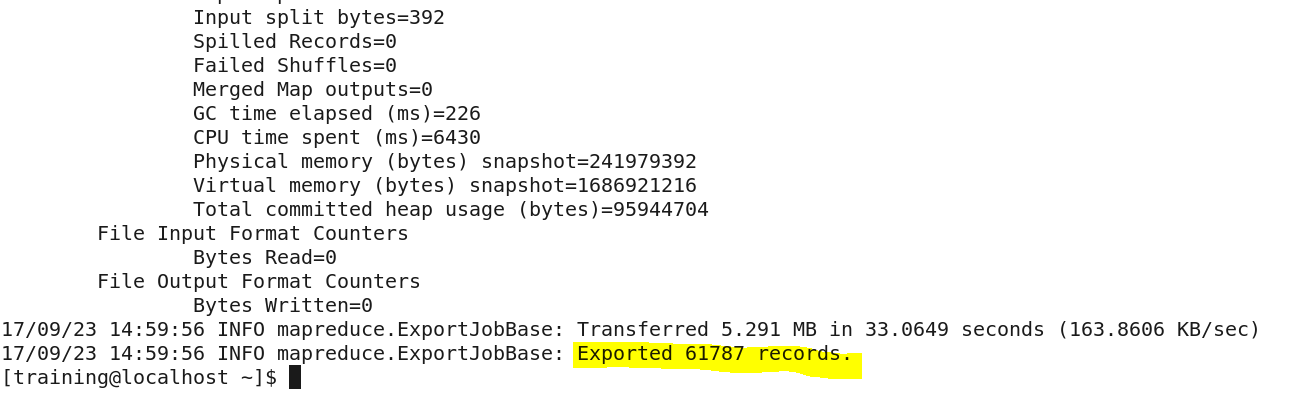
**--table Cali\_dxj160830 \**

**--export-dir '/user/hive/warehouse/accounts\_lesscolumns' \**

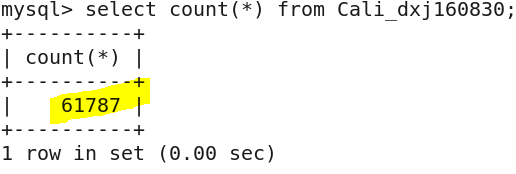
**--m 2**

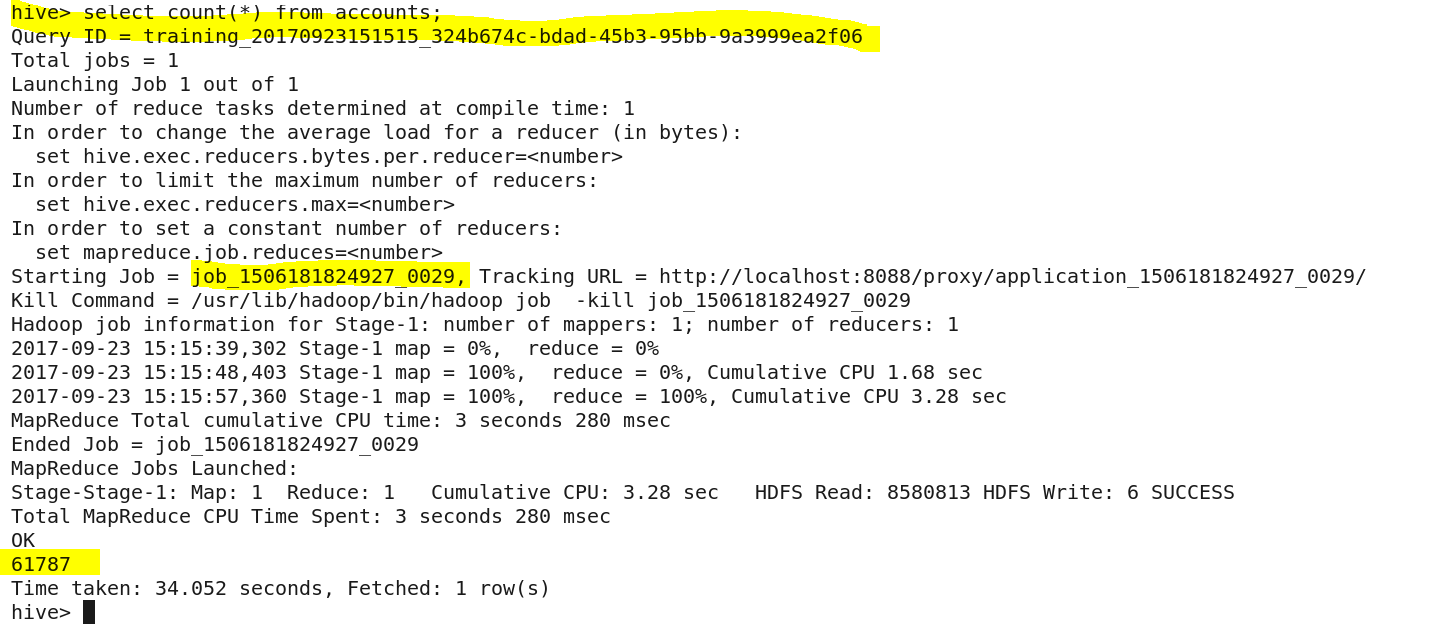




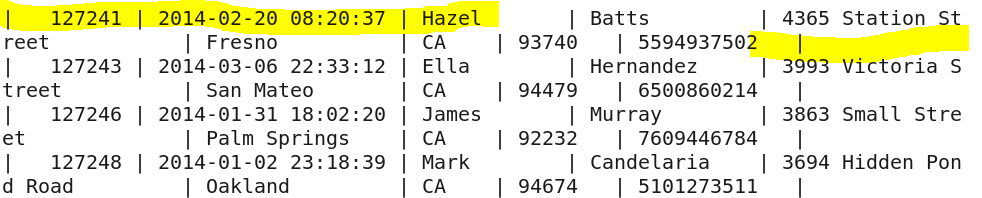


**Verification:**





**In Mysql:**



**Answer:**

**In Hive:**

Created accounts\_lesscolumns table in hive WITHOUT ’acct\_close\_dt’, ‘modified’ and ‘created’

Inserted all columns except ’acct\_close\_dt’, ‘modified’ and ‘created’ from accounts table to accounts\_lesscolumns table

**In MySQL:**

Created Cali\_dxj160830 table in mySQL WITHOUT ’acct\_close\_dt’, ‘modified’ and ‘created’

Exported using sqoop from hive to mysql

**Part B: 1. Check 10 rows of content on table ‘webpage’ without importing the table (hint: use eval command in sqoop)**

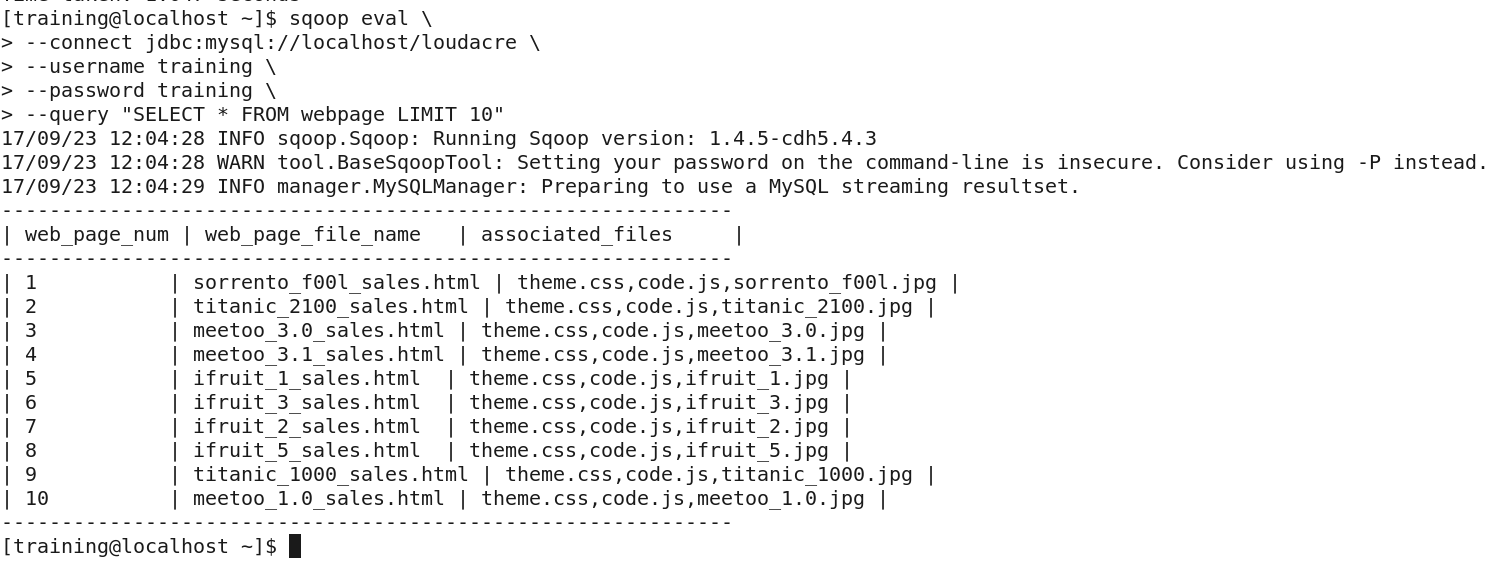
**sqoop eval \**

**--connect jdbc:mysql://localhost/loudacre \**

**--username training \**

**--password training \**

**--query "SELECT \* FROM webpage LIMIT 10"**



**Answer:**

Eval tool in Sqoop enables us to check 10 rows of in table ‘Webpage’ without importing the table into hdfs

**Part B: 2. Import table ‘webpage’ into a directory called Your NET-ID/loudacre/webpage a. Fields should be terminated by tab**

**sqoop import \**

**--connect jdbc:mysql://localhost/loudacre \**

**--username training \**

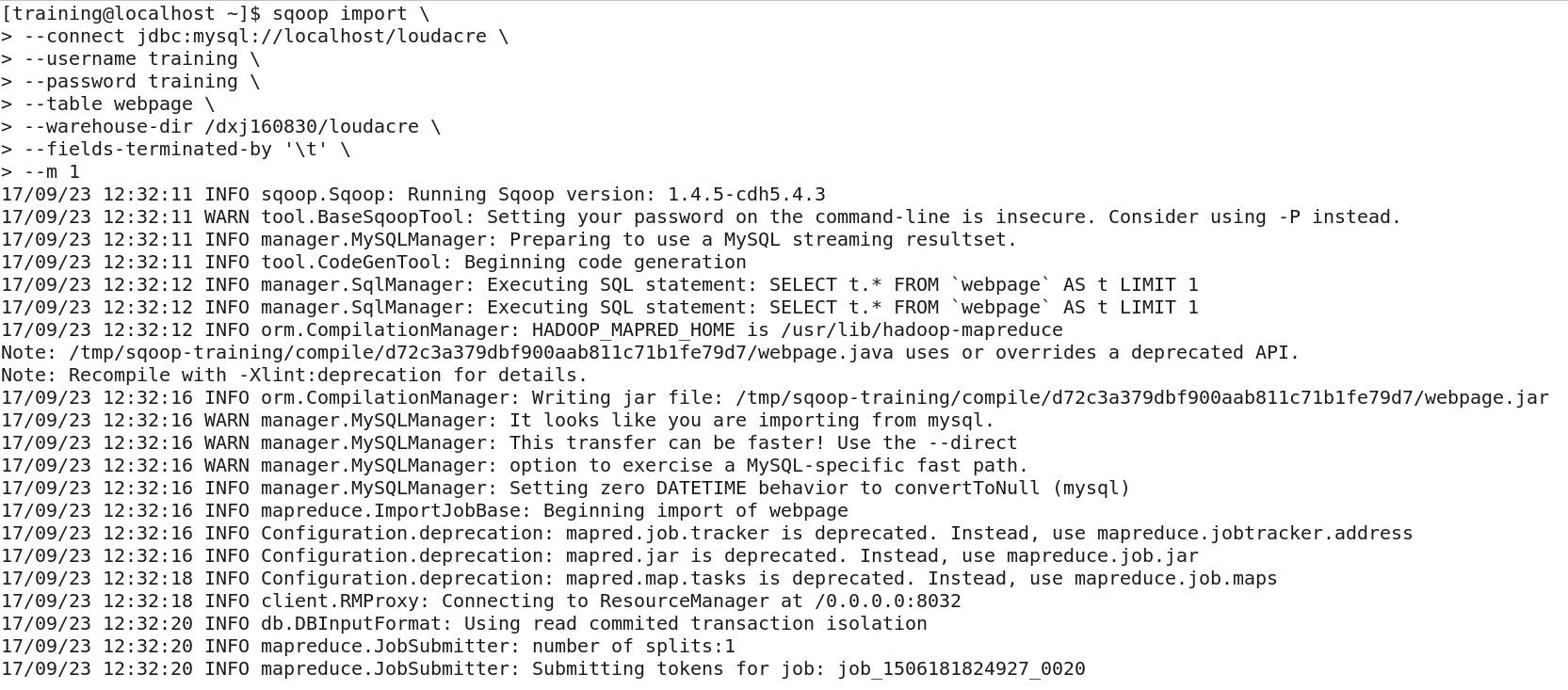
**--password training \**

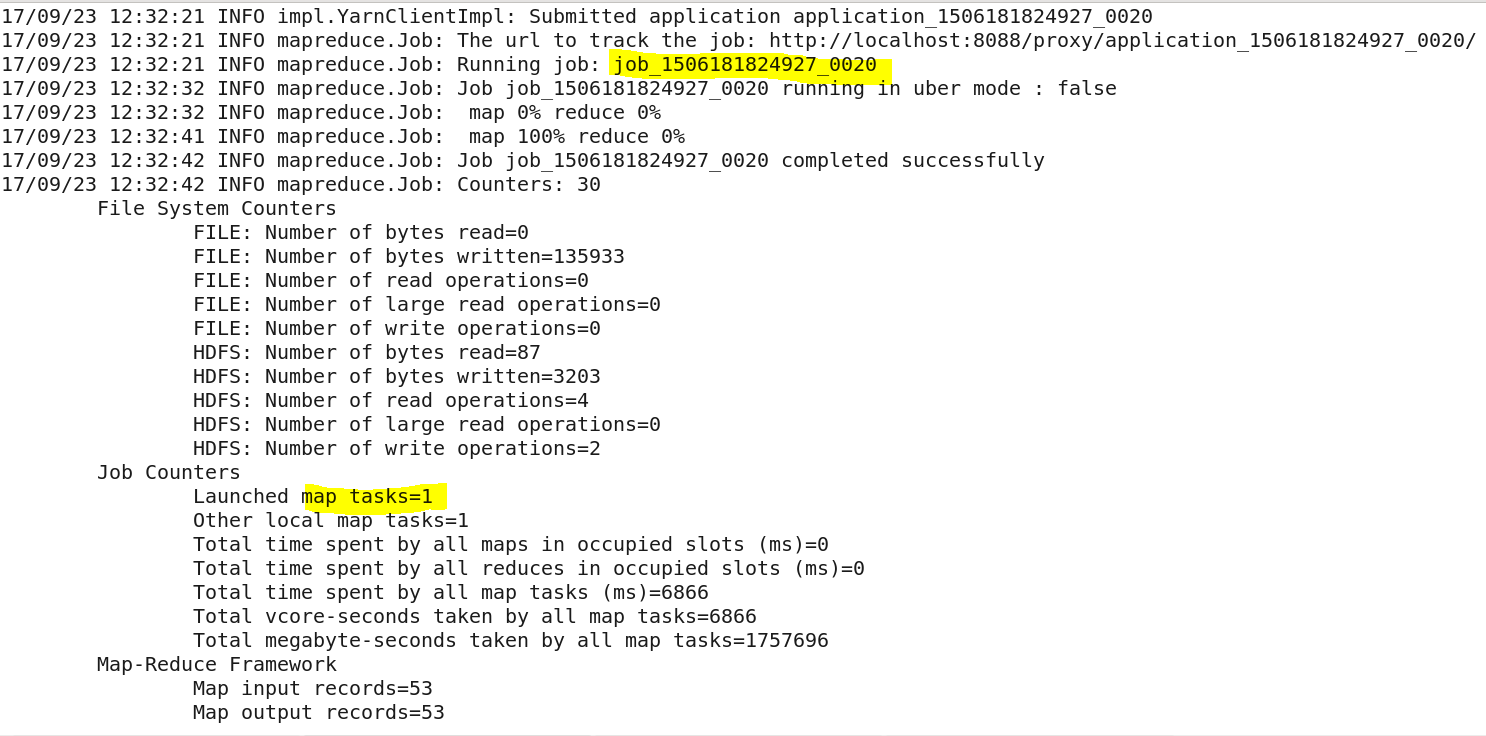
**--table webpage \**

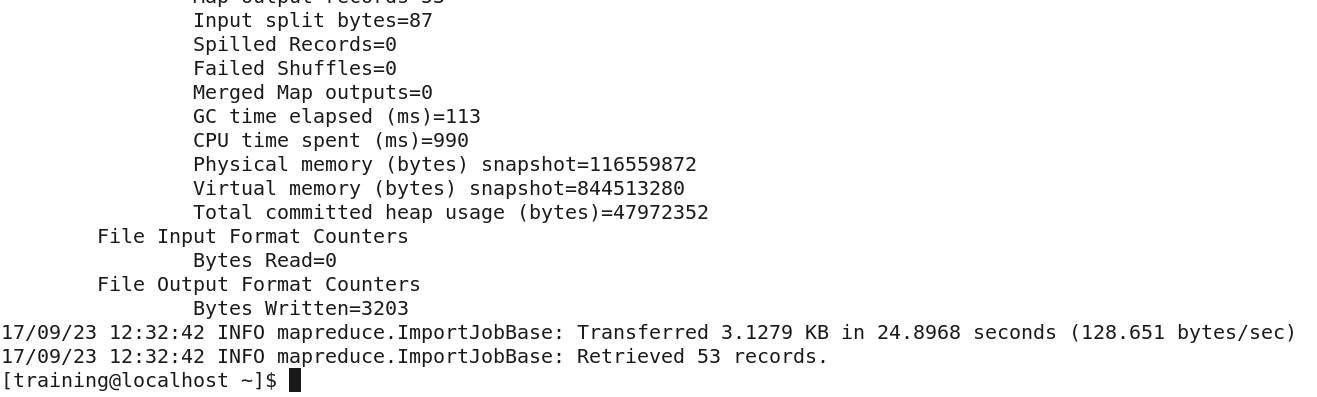
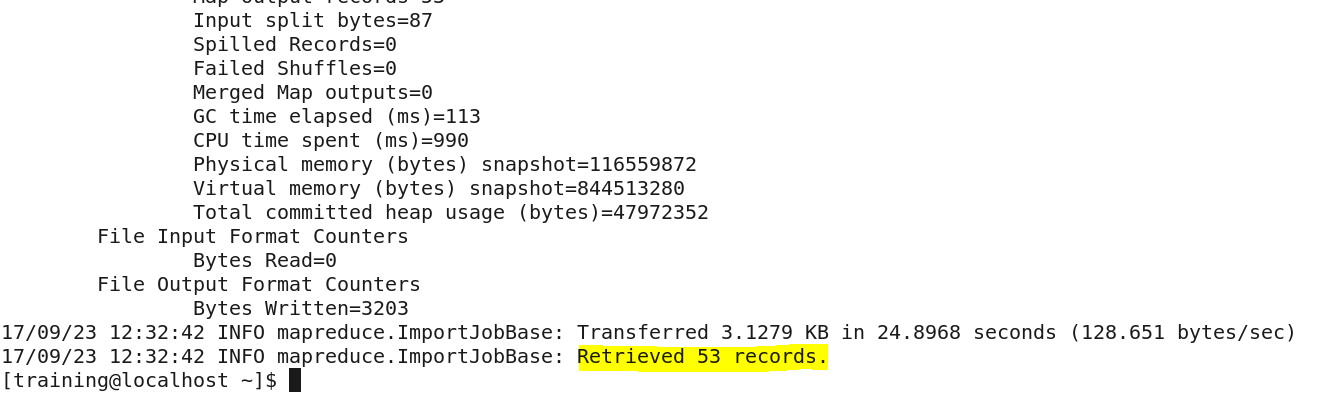
**--warehouse-dir /dxj160830/loudacre \**

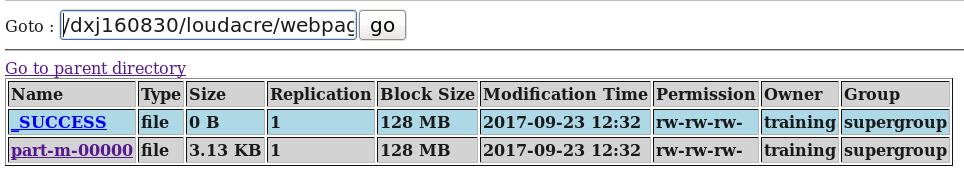
**--fields-terminated-by '\t' \**

**--m 1**

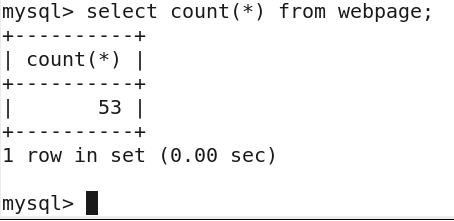




**MySql:**



* The number of records is matching in both mySQL and hdfs
* The warehouse-dir tool creates the webpage folder on its own
* Fields are terminated by tab