Assignment 11.2

**Problem Statement**

**2. Perform incremental load in Hive**

**Read from MySQL Table and load it in Hive table.**

**Create hive table if it does not exist.**

**If it exists, perform the incremental load.**

**Solution:**

**Incremental load in Hive**

Sqoop is the JDBC-based utility for integrating with traditional databases. A Sqoop Import allows for the movement of data into either HDFS (a delimited format can be defined as part of the Import definition) or directly into a Hive table.

Sqoop supports two types of incremental imports: **append** and **lastmodified**. You can use the –incremental argument to specify the type of incremental import to perform.

You should specify the **append** mode when importing a table, where new rows are continually added with increasing row id values. You must specify the column containing the row’s id with –check-column. Sqoop imports rows where the check column has a value greater than the one specified with –last-value.

An alternate table update strategy supported by Sqoop is called **lastmodified** mode*. This should be used when rows of the source table is updated, and each such update will set the value of a last-modified column to the current timestamp*. Rows where the check column holds a timestamp more recent than the timestamp specified with –last-value are imported.

At the end of an incremental import, the value which should be specified as –last-value for a subsequent import is printed to the screen. When running a subsequent import, you should specify –last-value in this way to ensure you import only the new or updated data. This is handled automatically by creating an incremental import as a saved job, which is the preferred mechanism for performing a recurring incremental import.

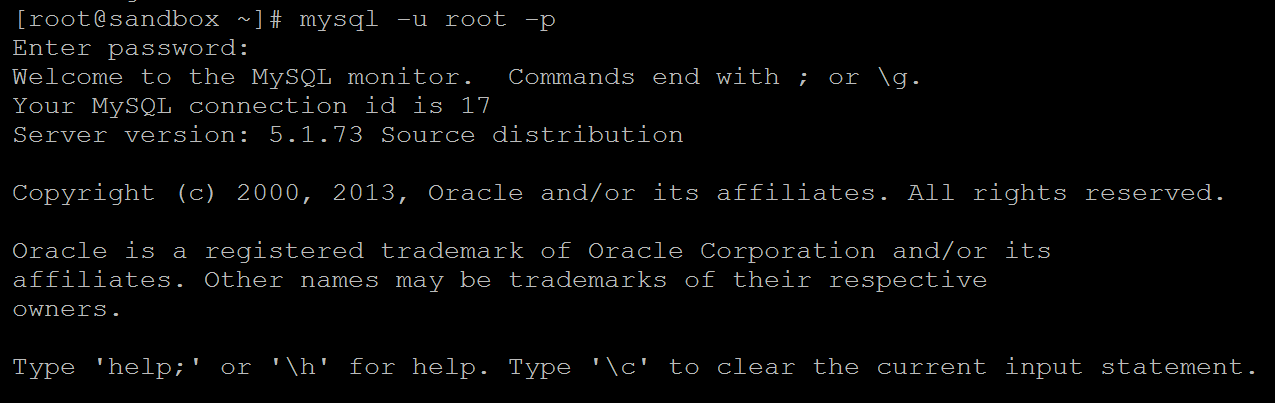
**Step:1**

Start the MySQL service with the below command:

sudo service mysqld start

And enter MySQL shell using the below command:

mysql -u root -p



**Step:2**

list database if already existing:

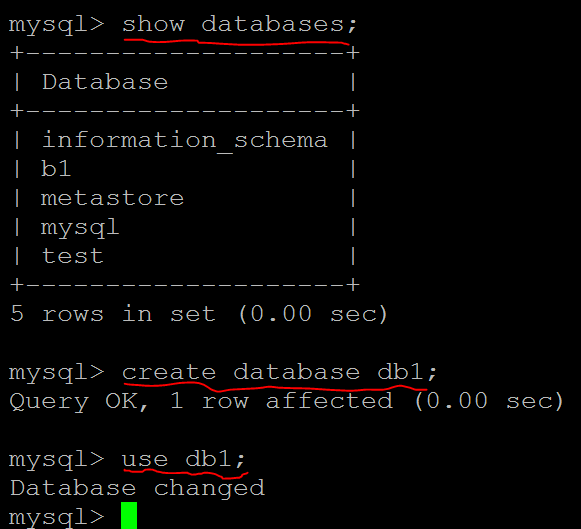
show databases;

create a new database:

create database db1;

use the database:

use db1;



**Step:3**

Also creating table, inserting values inside table is done using the following syntax.

create table acad(emp\_id INT NOT NULL AUTO\_INCREMENT,

emp\_name VARCHAR(100),emp\_sal INT,PRIMARY KEY (emp\_id) );

insert into acad values(5,"sanam",50000);

insert into acad values(6,"opra",600000);

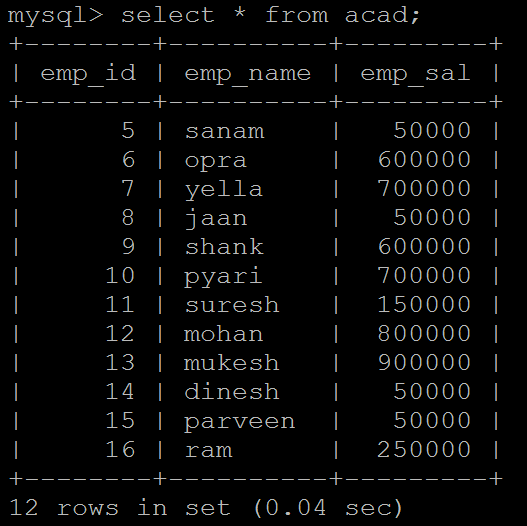
insert into acad values(7,"yella",700000);

insert into acad values(8,"jaan",50000);

insert into acad values(9,"shank",600000);

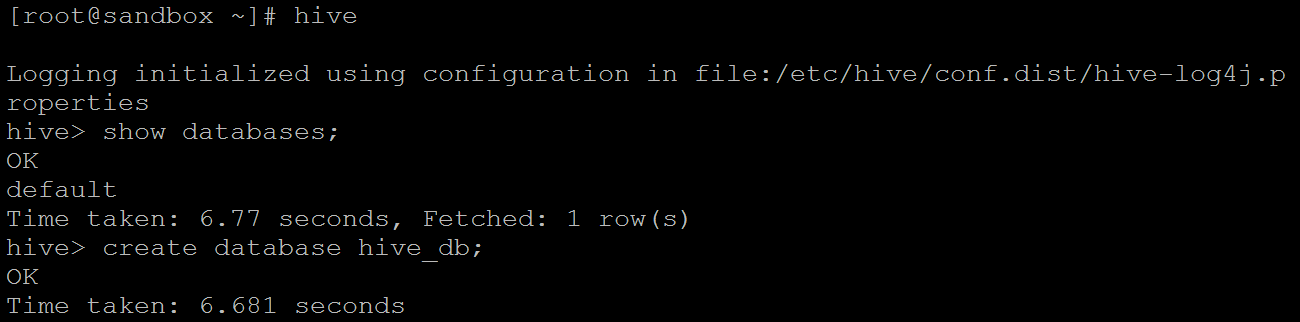
insert into acad values(10,"pyari",700000);

select \* from acad;



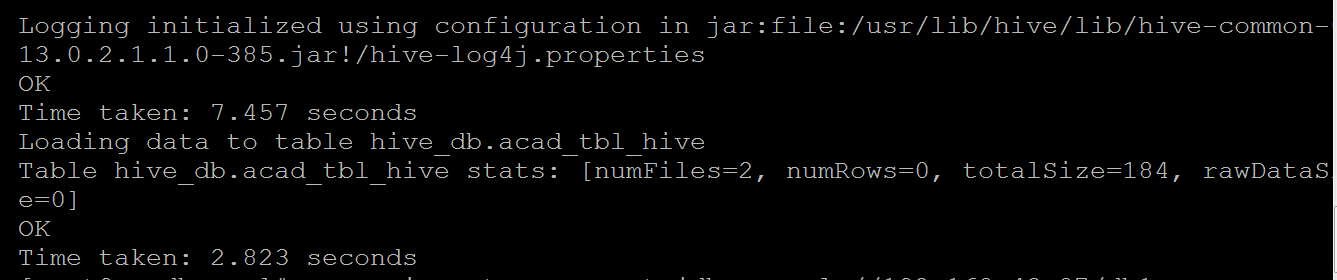
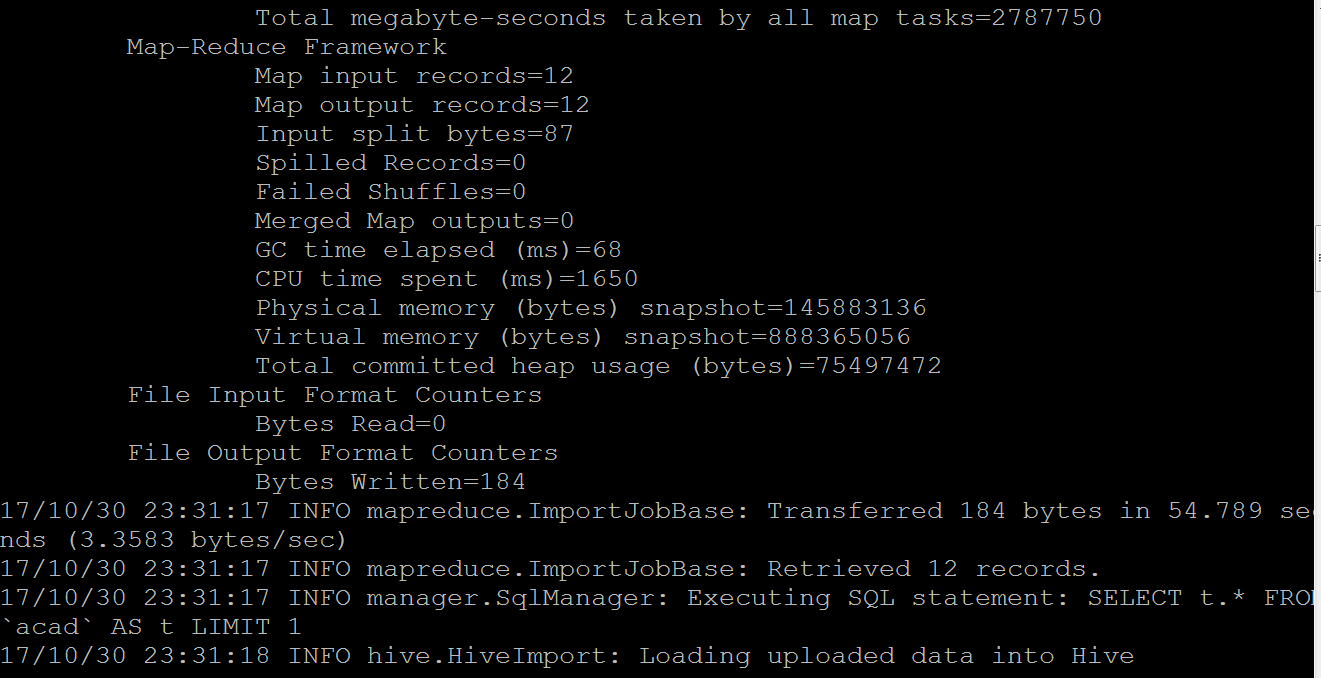
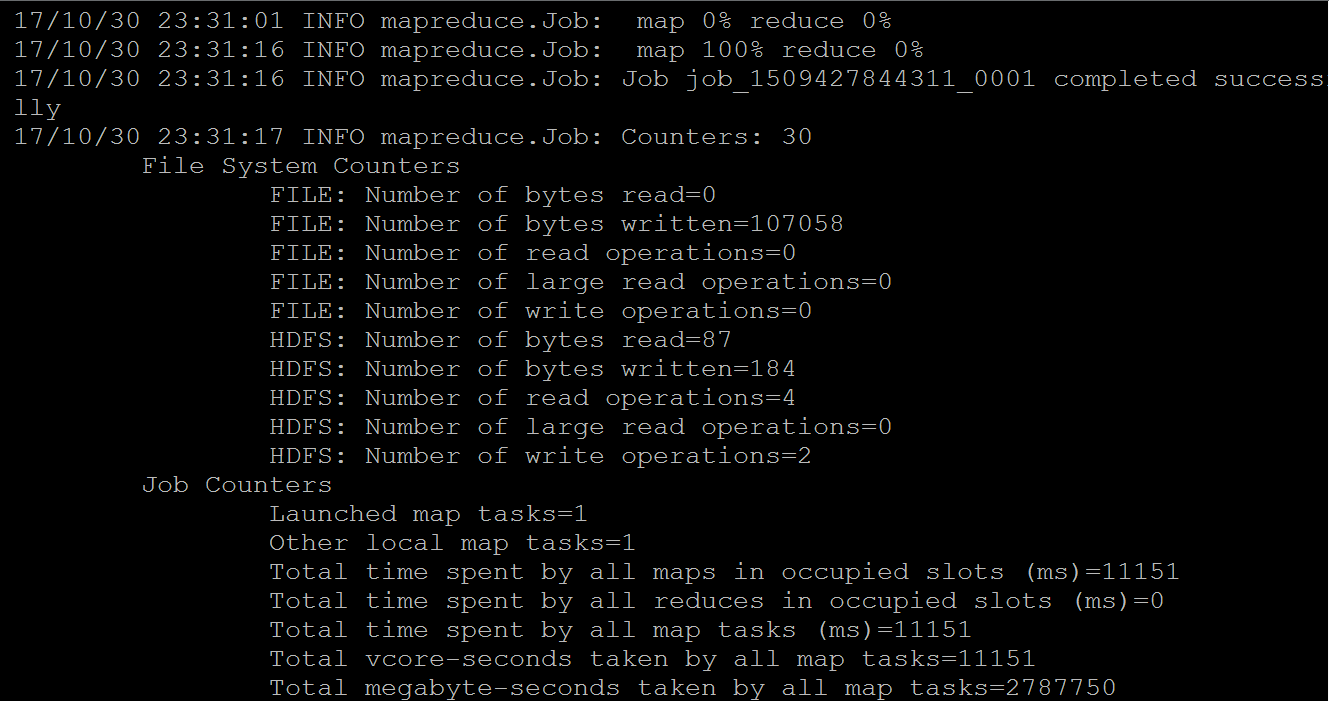
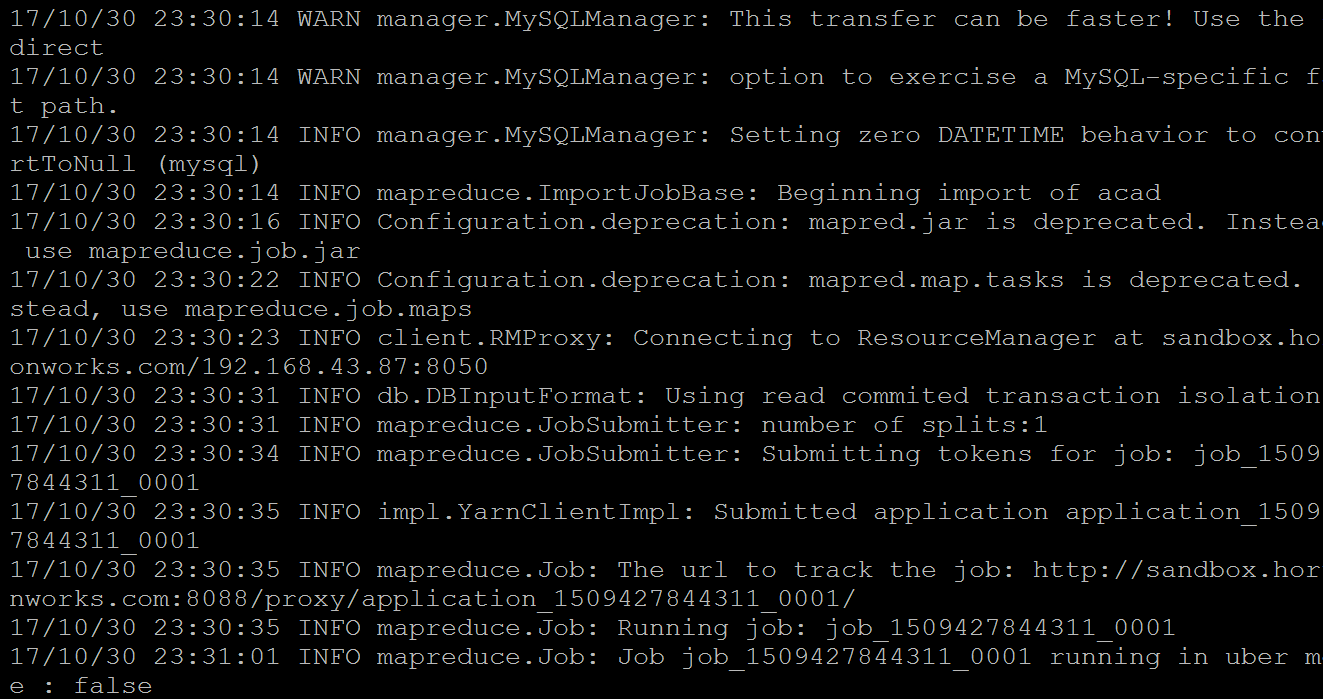
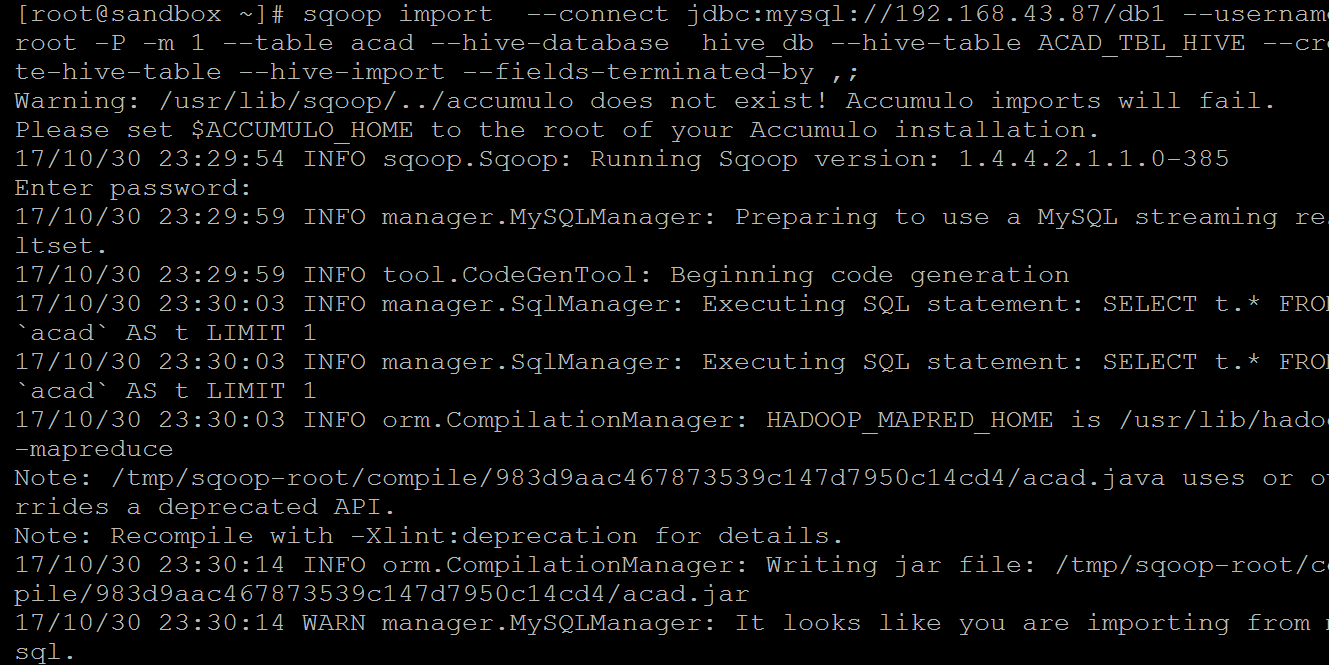
**Step:4**

create database in Hive

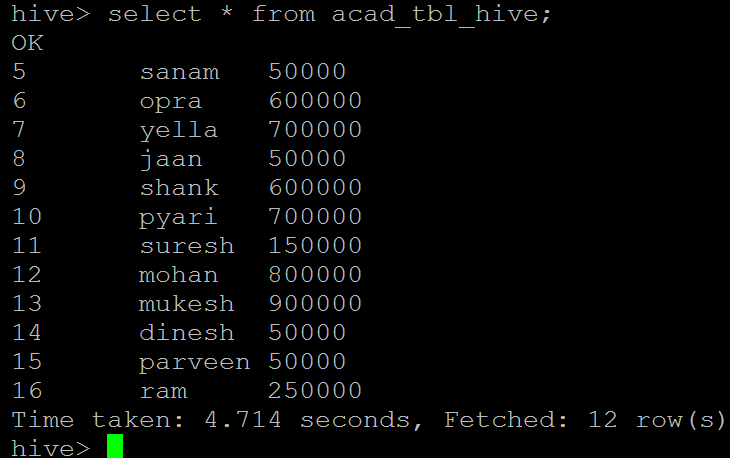


**step 5:** Initial import of data from MySQL to Hive

sqoop import --connect jdbc:mysql://192.168.43.87/db1 --username root -P -m 1 --table acad --hive-database hive\_db --hive-table ACAD\_TBL\_HIVE --create-hive-table --hive-import --fields-terminated-by ,;

****

**Step 6: Check in Hive**

****

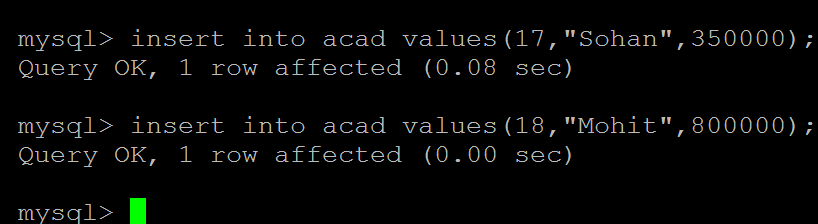
As confirmation of the result, you can see in the image, we retrieved all records.

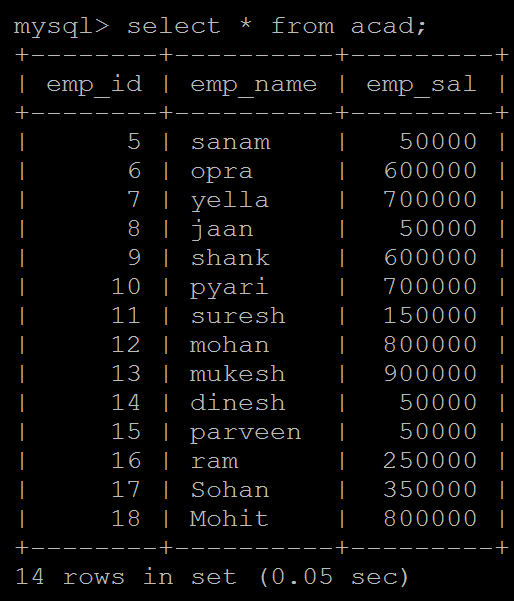
This confirms the data inside MySQL has come inside the Hive.

**After Updating some rows in MYSQL**

**Step:1**

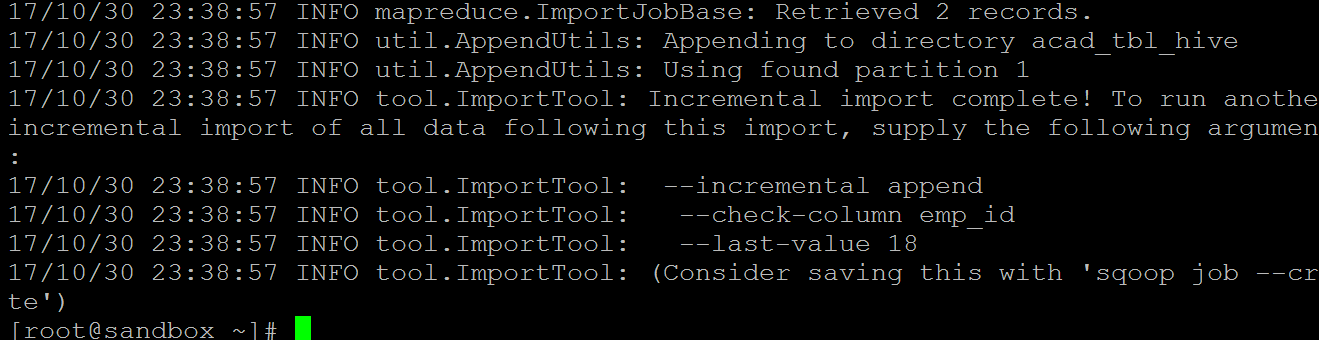
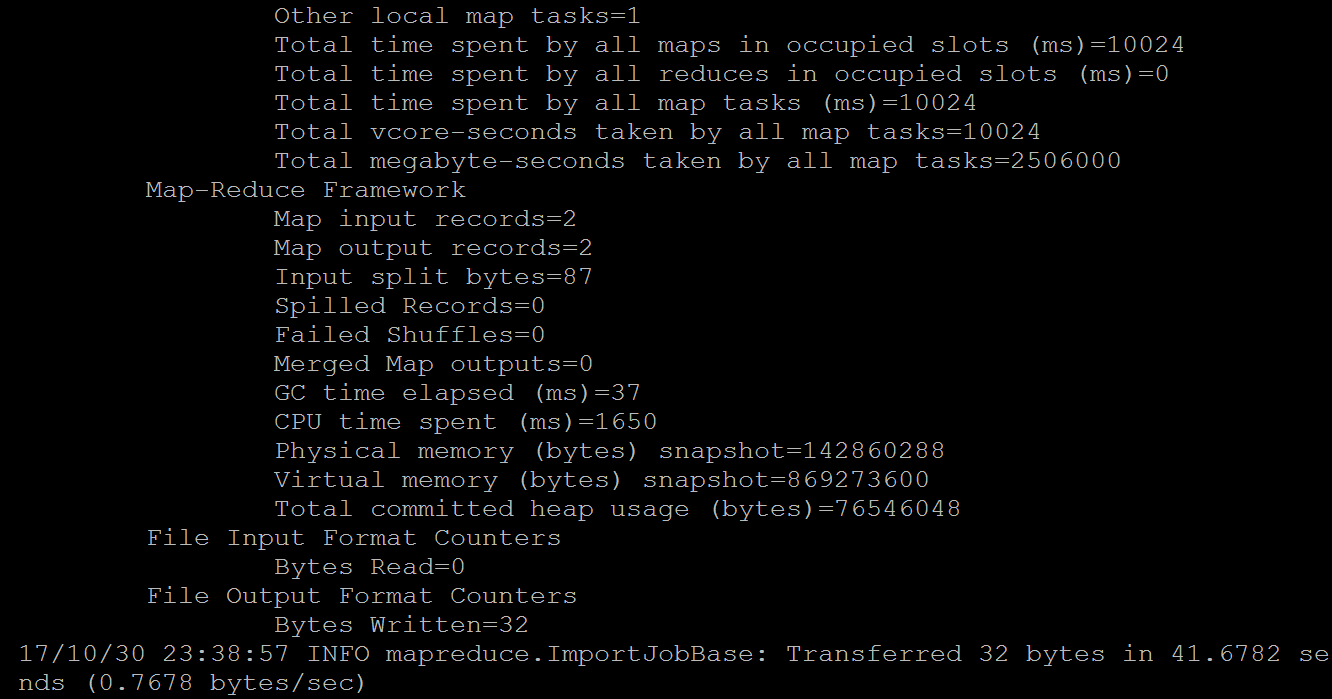
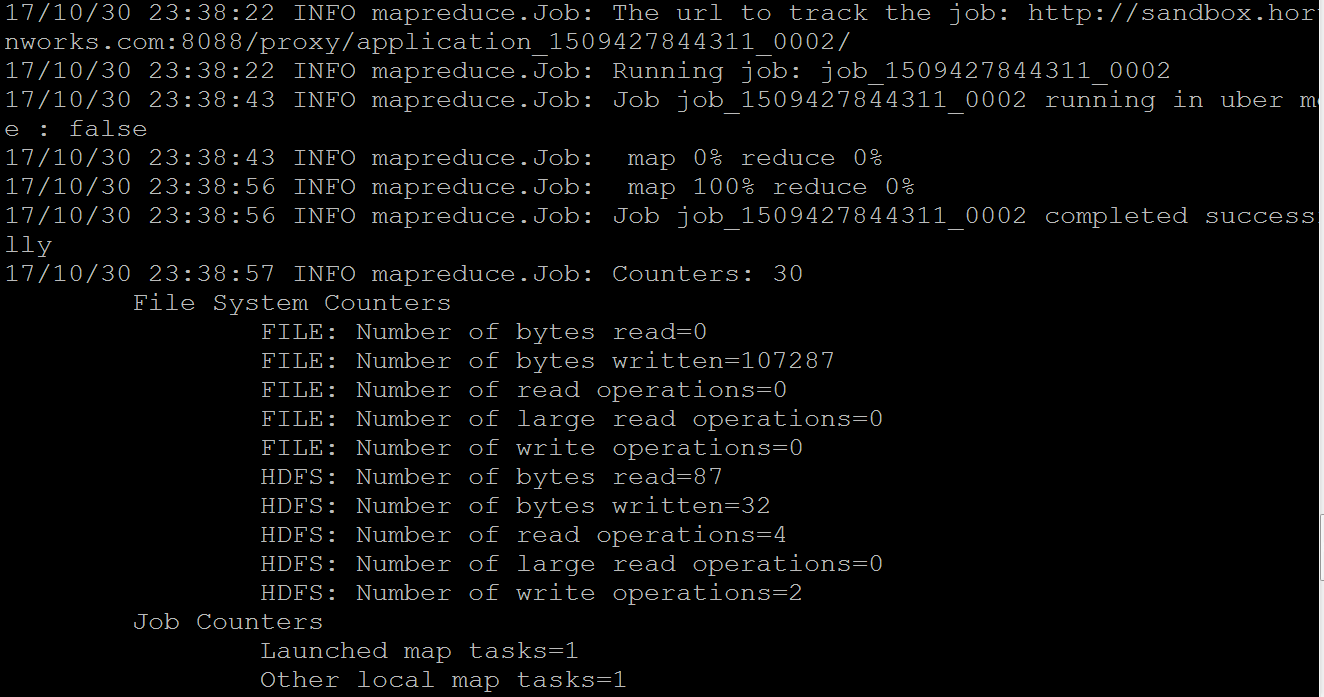
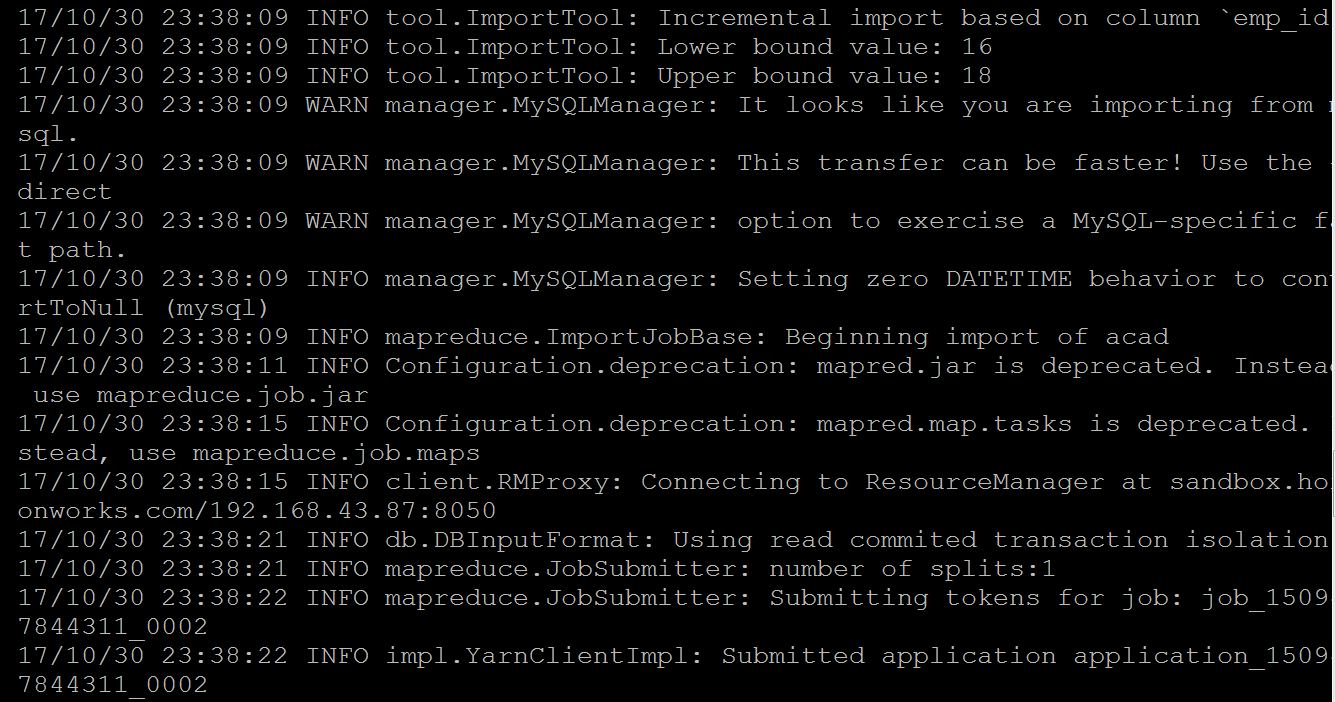
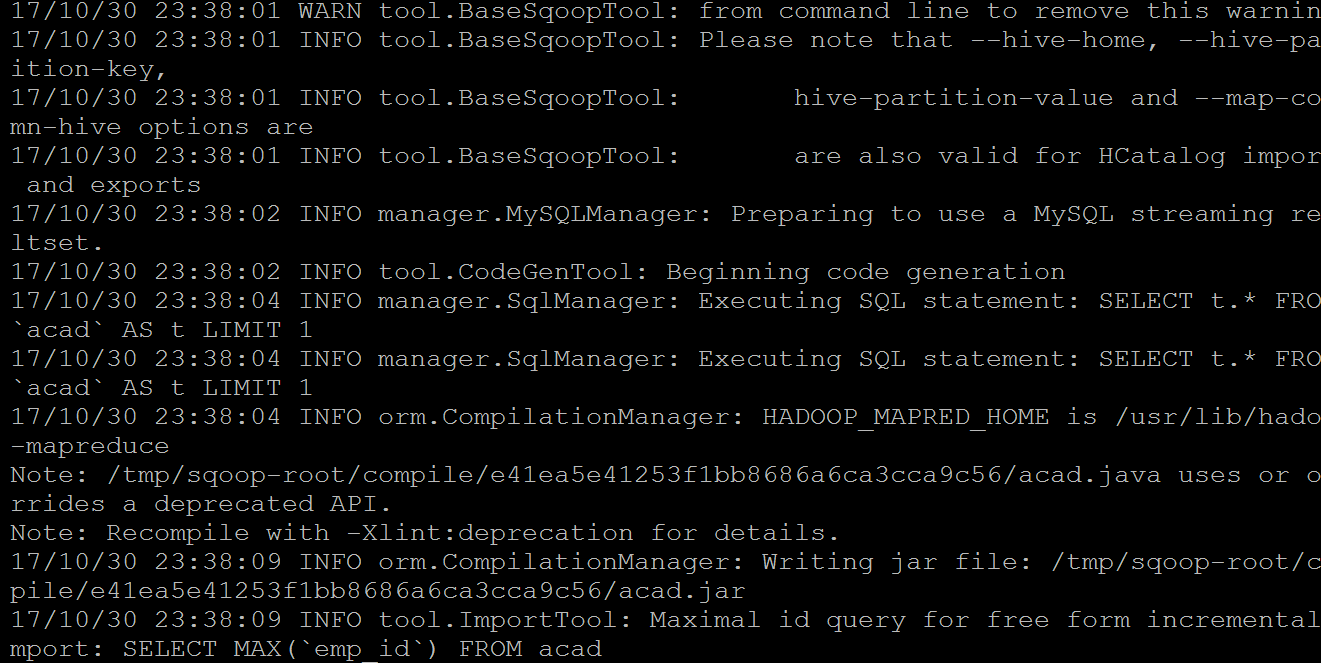
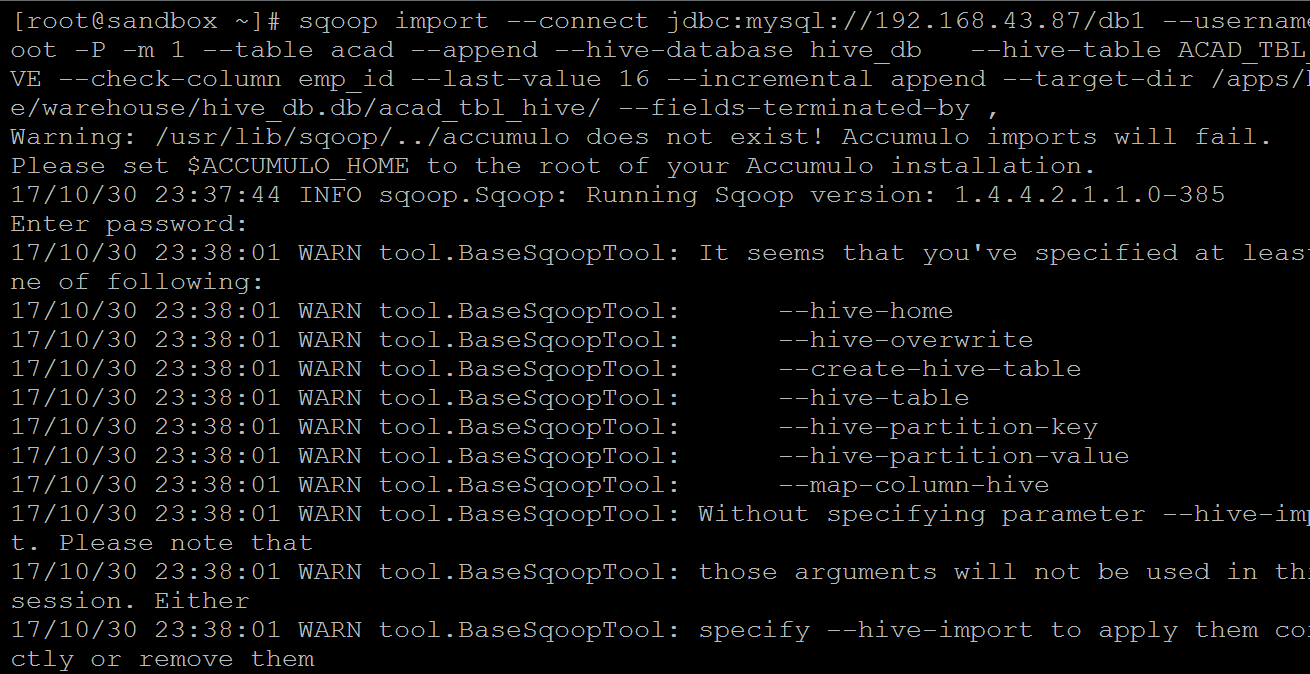
Let’s manually insert few extra values in mysql / acad table.

****



**Step:2**

sqoop import --connect jdbc:mysql://192.168.43.87/db1 --username root -P -m 1 --table acad --append --hive-database hive\_db --hive-table ACAD\_TBL\_HIVE --check-column emp\_id --last-value 16 --incremental append --target-dir /apps/hive/warehouse/hive\_db.db/acad\_tbl\_hive/ --fields-terminated-by ,



As you can see in above image, 3 more records have been retrieved and the incremental import is now complete.

Along with message for next incremental import, you need to give last value as 10.

**Step:3**

Now let’s check hive table and number of records.

