**Session 11: Sqoop Flume**

Assignment 11.1

Student Name: Abarajithan SA

Course: Big Data Hadoop & Spark Training

Start Date:  2017-09-09

End Date:  2017-11-26

**Assignment 11.1**– Perform and explain the code flow and the associated result for the below tasks. Candidates should create and use their own employee dataset for the same. Share the screenshot of the commands used and its associated result.

Contents

[Introduction 2](#_Toc500254555)

[Problem Statement 2](#_Toc500254556)

[Prerequisite 2](#_Toc500254557)

[Task 1:Transfer data between **Mysql** and **HDFS** (Import and Export) using Sqoop 3](#_Toc500254558)

[Importing a Table from RDBMS to HDFS 3](#_Toc500254559)

[Sqoop import step, 4](#_Toc500254560)

[Expected Output 4](#_Toc500254561)

[Export Data from HDFS to RDBMS 5](#_Toc500254562)

[Sqoop Export step 6](#_Toc500254563)

[Expected output 7](#_Toc500254564)

[Task2: Transfer data between **Mysql** and **Hive** (Import and Export only selected columns) using Sqoop. 7](#_Toc500254565)

[A Sqoop command is used to transfer selected columns from MySQL to Hive 7](#_Toc500254566)

[Sqoop import step 8](#_Toc500254567)

[Expected Output 9](#_Toc500254568)

[Export command for transferring the selected columns from Hive to MySQL 9](#_Toc500254569)

[Sqoop Export step 9](#_Toc500254570)

[Expected Output 10](#_Toc500254571)

# Introduction

In this assignment, we are going perform Sqoop import/export operations from mysql-HDFS and the HIVE-mysql.

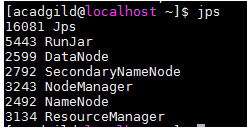
# Problem Statement

Perform and explain the code flow and the associated result for the below tasks. Candidates should create and use their own employee dataset for the same. Share the screenshot of the commands used and its associated result.

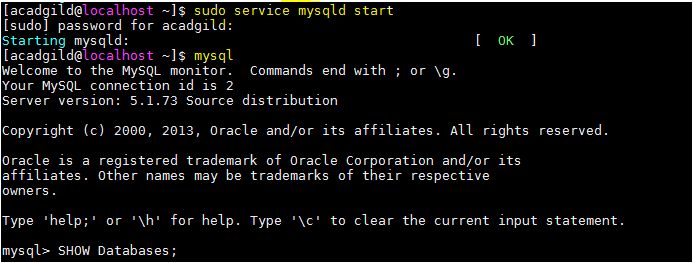
* Transfer data between Mysql and HDFS (Import and Export) using Sqoop.
* Transfer data between Mysql and Hive (Import and Export only selected columns) using Sqoop.

# Prerequisite

1. Make sure all the hadoop daemons are started,



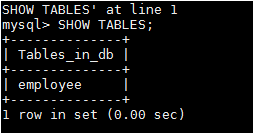
1. Start the mysql shell,



# Task 1:Transfer data between **Mysql** and **HDFS** (Import and Export) using Sqoop

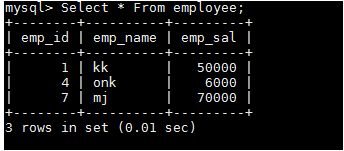
## Importing a Table from RDBMS to HDFS

1. Create Database db; (we are using an existing DB and Table)
2. **SHOW TABLES:**

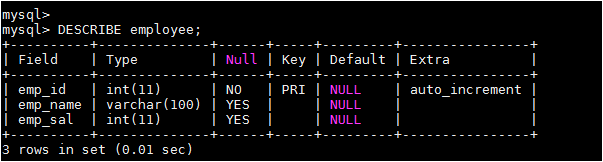


1. Data available in the table ‘**employee’**

**Select \* From employee;**

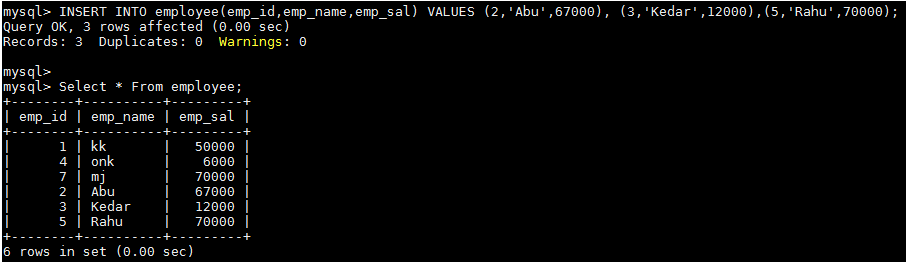


1. DESCRIBE employee;



1. Insert some more values into the existing table,

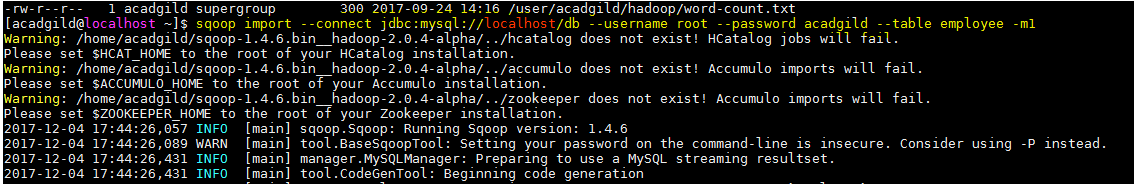
***INSERT INTO employee(emp\_id,emp\_name,emp\_sal) VALUES (2,'Abu',67000), (3,'Kedar',12000),(5,'Rahu',70000);***



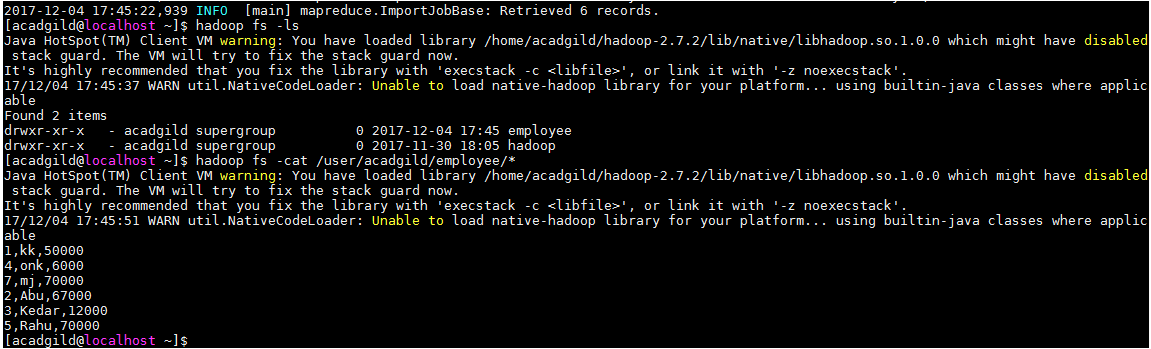
The above is the data available in the table ‘**employee’** in the database ‘**db’**

### Sqoop import step,

***sqoop import --connect jdbc:mysql://localhost/db --username root --password acadgild --table employee -m1***



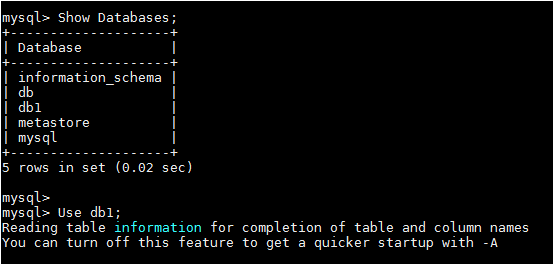
### Expected Output



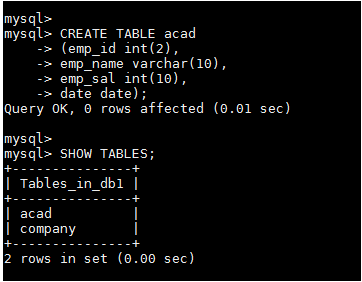
The data from employee table in the mysql has been imported using sqoop and the output is shown above in the screen shot.

## Export Data from HDFS to RDBMS

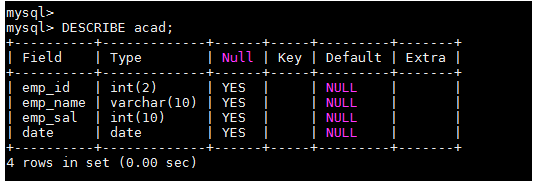
1. We are using the existing the database **‘db1’**



1. Create a table ‘**acad’** in the ‘**db1’** like below,



1. Table ‘**acad’** Format,



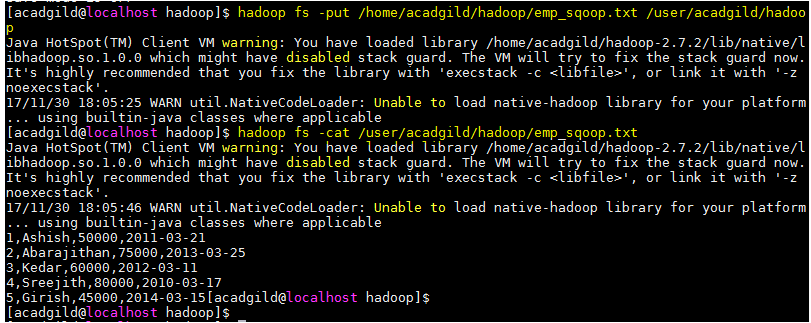
1. The data in this table are yet to be imported from the HDFS, hence it will be shown empty set,



1. We are going to use the data set **‘emp\_sqoop.txt’,**  the **‘emp\_sqoop.txt’** should be transferred from the local system to the HDFS first and read the data using **cat** command,

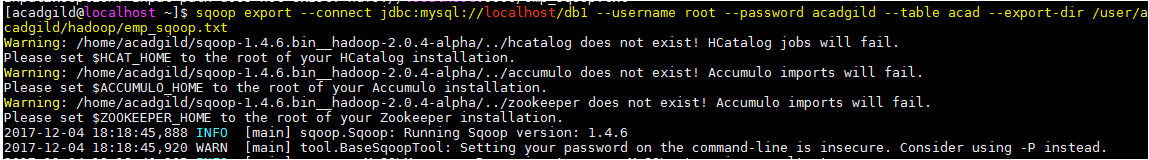
***hadoop fs -put /home/acadgild/hadoop/emp\_sqoop.txt /user/acadgild/hadoop***

***hadoop fs -cat /user/acadgild/hadoop/emp\_sqoop.txt***

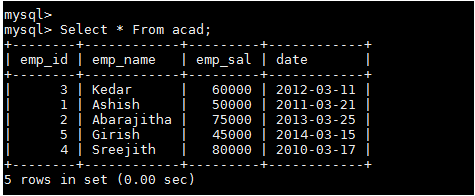


### Sqoop Export step

***sqoop export --connect jdbc:mysql://localhost/db1 --username root --password acadgild --table acad --export-dir /user/acadgild/hadoop/emp\_sqoop.txt***



### Expected output

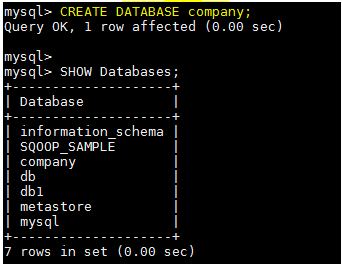


The data from the **‘emp\_sqoop.txt’** which was in the HDFS has been exported to the table ‘**acad’** to the database **‘db1’.**

# Task2: Transfer data between **Mysql** and **Hive** (Import and Export only selected columns) using Sqoop.

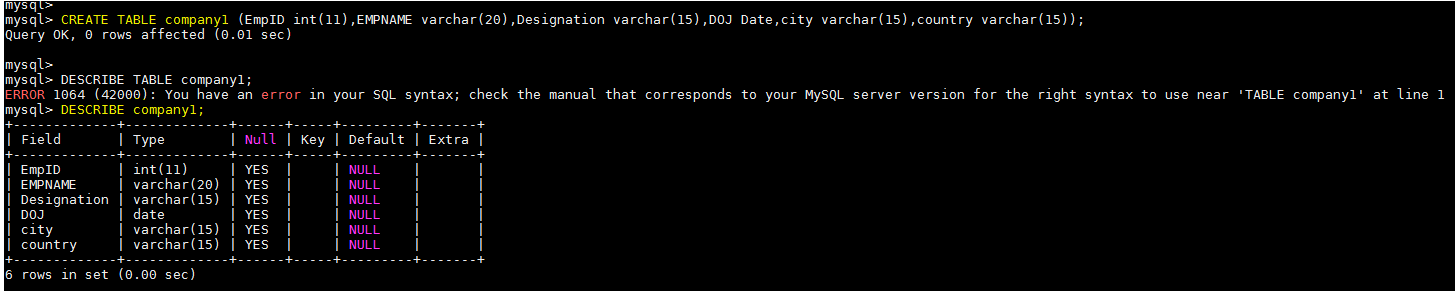
## A Sqoop command is used to transfer selected columns from MySQL to Hive

1. ***CREATE DATABASE company;***



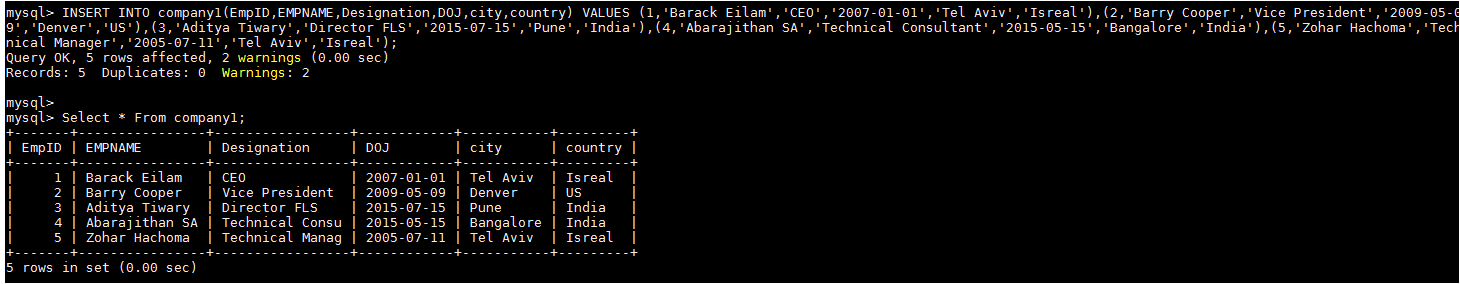
1. Create a table ***‘company1’***

***CREATE TABLE company1 (EmpID int(11),EMPNAME varchar(20),Designation varchar(15),DOJ varchar(15),city varchar(15),country varchar(15));***



1. Add some data into the newly created table ‘company1’

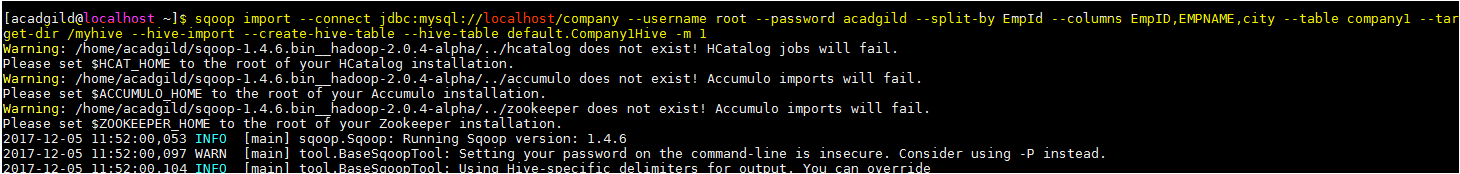
***INSERT INTO company1(EmpID,EMPNAME,Designation,DOJ,city,country) VALUES (1,'Barack Eilam','CEO','2007-01-01','Tel Aviv','Isreal'),(2,'Barry Cooper','Vice President','2009-05-09','Denver','US'),(3,'Aditya Tiwary','Director FLS','2015-07-15','Pune','India'),(4,'Abarajithan SA','Technical Consultant','2015-05-15','Bangalore','India'),(5,'Zohar Hachoma','Technical Manager','2005-07-11','Tel Aviv','Isreal');***



### Sqoop import step

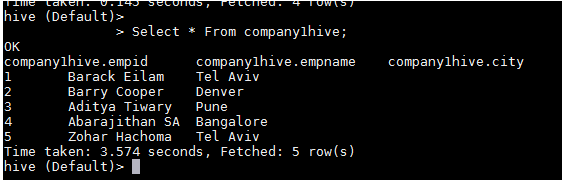
***sqoop import --connect jdbc:mysql://localhost/company --username root --password acadgild --split-by EmpId --columns EmpID,EMPNAME,city --table company1 --target-dir /myhive --hive-import --create-hive-table --hive-table default.Company1Hive -m 1***

the above command will import the data from the table ‘**company1’** to the new table ‘**Company1Hive’** in to the database ‘**default’** in the HIVE,



1. Go hive shell, **Select \* From Company1Hive;**

### Expected Output

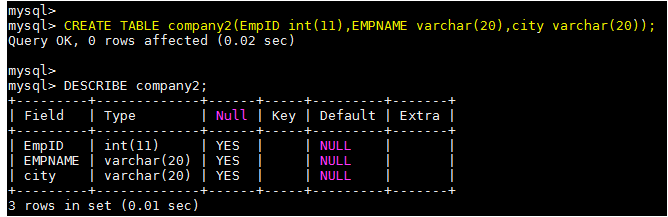


Hence, the data from the ‘**company1’** table has been imported into the hive table ‘**Company1hive’**

## Export command for transferring the selected columns from Hive to MySQL

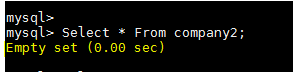
1. In order to export data from the HIVE table to mysql, we need create a table in the mysql prior,

***CREATE TABLE company2(EmpID int(11),EMPNAME varchar(20),city varchar(20));***



1. The data in this table must be empty,

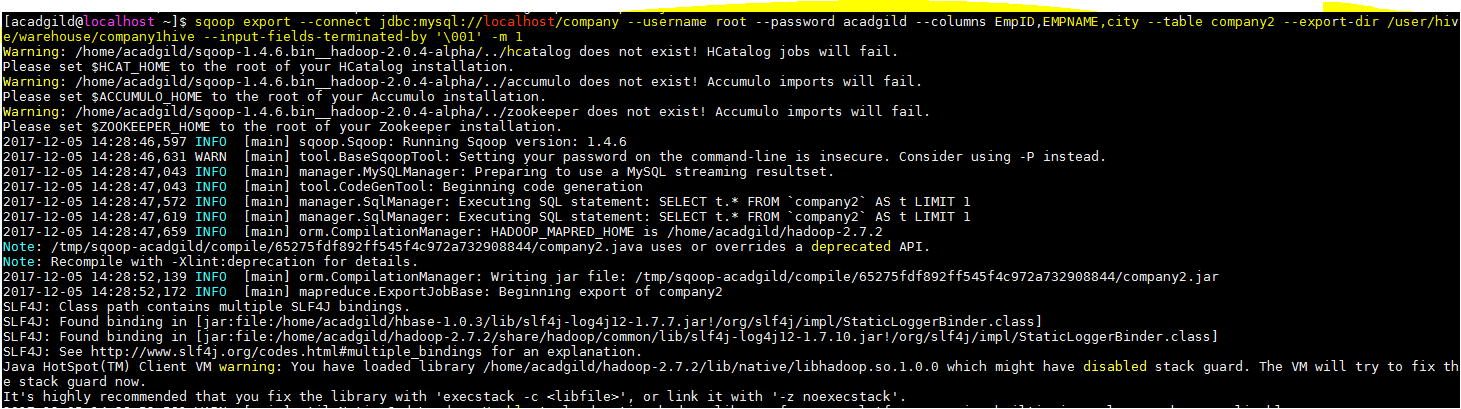
***Select \* From company2;***

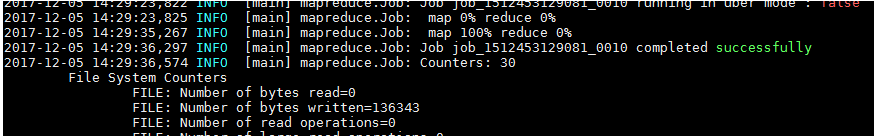


### Sqoop Export step

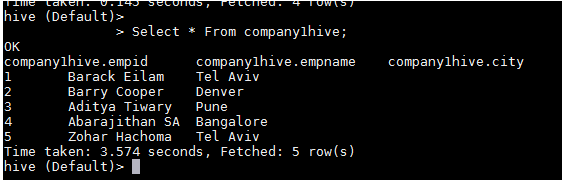
***sqoop export --connect jdbc:mysql://localhost/company --username root --password acadgild --columns EmpID,EMPNAME,city --table company2 --export-dir /user/hive/warehouse/company1hive --input-fields-terminated-by '\001' -m 1***

The above command will export the data from the hive table ‘**company1hive’** into the newly created table **‘company2’** in the mysql,





Company1hive data,



### Expected Output

