**Session 11: Sqoop Flume**

Assignment 11.2

Student Name: Abarajithan SA

Course: Big Data Hadoop & Spark Training

Start Date:  2017-09-09

End Date:  2017-11-26

**Assignment 11.2**–

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.

Contents

[Introduction 2](#_Toc500348880)

[Problem Statement 2](#_Toc500348881)

[Prerequisite 2](#_Toc500348882)

[Task 3](#_Toc500348883)

[Incremental import to HDFS and mapping Hive table to sqoop's target-dir. 5](#_Toc500348884)

# Introduction

In this assignment, we are going perform incremental load in HIVE using Sqoop.

# Problem Statement

Perform incremental load in Hive

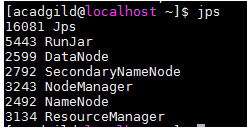
Read from MySQL Table and load it in Hive table.

Create hive table if it does not exist.

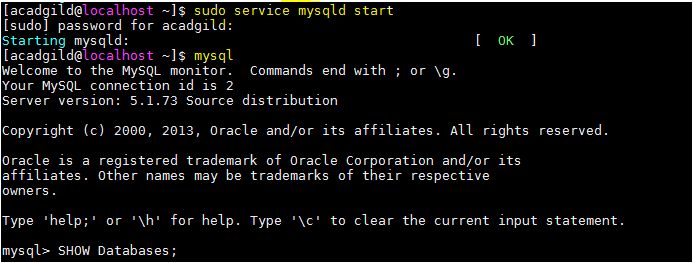
If it exists, perform the incremental load.

# Prerequisite

1. Make sure all the hadoop daemons are started,

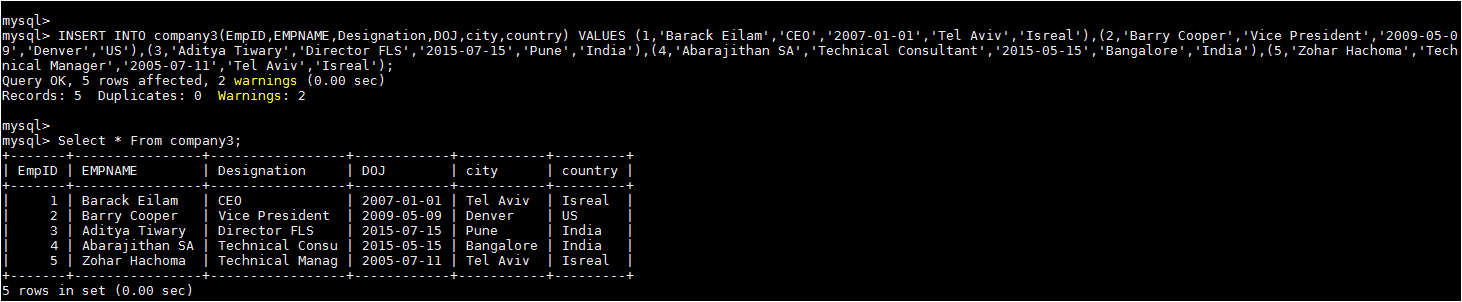
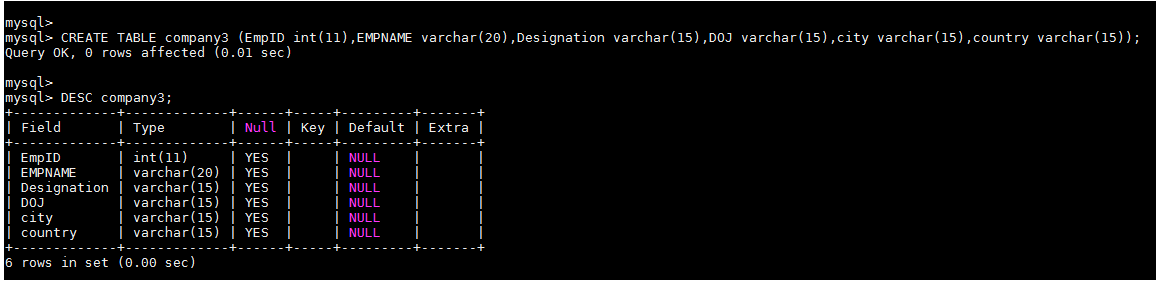


1. Start the mysql shell,



# Task

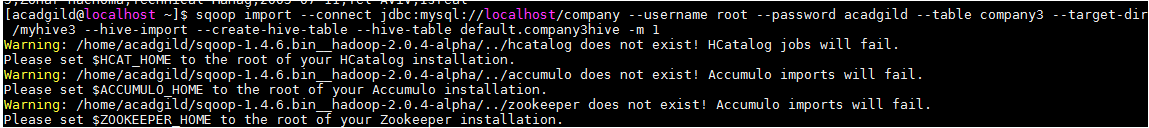
1. We are using the existing DB ‘**company’**
2. Create a table **‘company3’** and insert data into it.

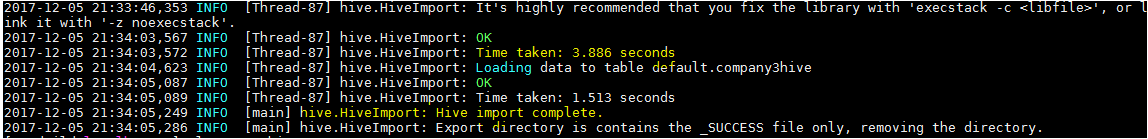


1. Since the data is present in table of MySQL and Sqoop is up and running, import the

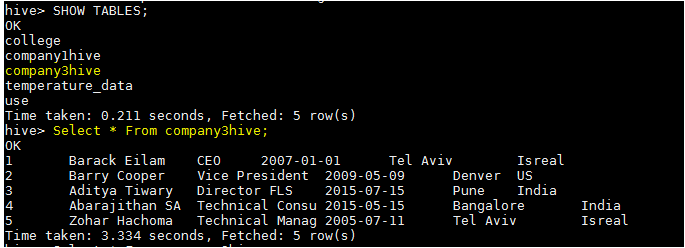
data into HIVE using the command

***sqoop import --connect jdbc:mysql://localhost/company --username root --password acadgild --table company3 --target-dir /myhive3 --hive-import --create-hive-table --hive-table default.company3hive -m 1***



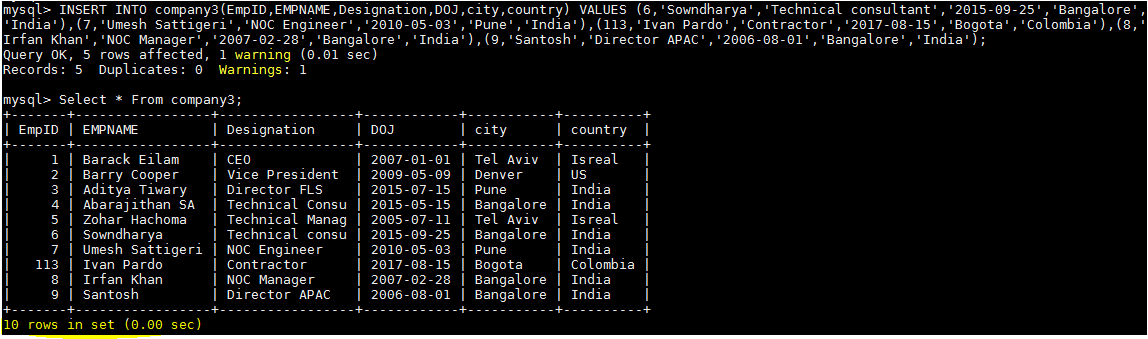


1. Let’s check the data in the hive table,



1. Inserted new values to the table employee.

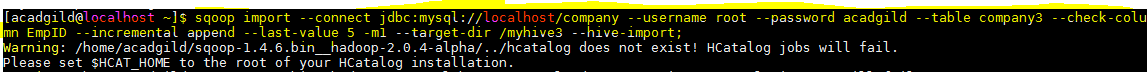
***INSERT INTO company3(EmpID,EMPNAME,Designation,DOJ,city,country) VALUES (6,'Sowndharya','Technical consultant','2015-09-25','Bangalore','India'),(7,'Umesh Sattigeri','NOC Engineer','2010-05-03','Pune','India'),(113,'Ivan Pardo','Contractor','2017-08-15','Bogota','Colombia'),(8,'Irfan Khan','NOC Manager','2007-02-28','Bangalore','India'),(9,'Santosh','Director APAC','2006-08-01','Bangalore','India');***



1. After inserting the new values in the table in MySQL shell, we shall import the updated

values into HIVE by using the INCREMENTAL command

***sqoop import --connect jdbc:mysql://localhost/company --username root --password acadgild --table company3 --check-column EmpID --incremental append --last-value 5 -m1 --target-dir /myhive3 --hive-import;***



In this command we are updating the new values by comparing the existing values of the table with the column EmpID*.*

Unfortunately, we get the below information,



**Append mode for hive imports is not yet supported. Please remove the parameter --append-mode**

Hence I tried the below method,

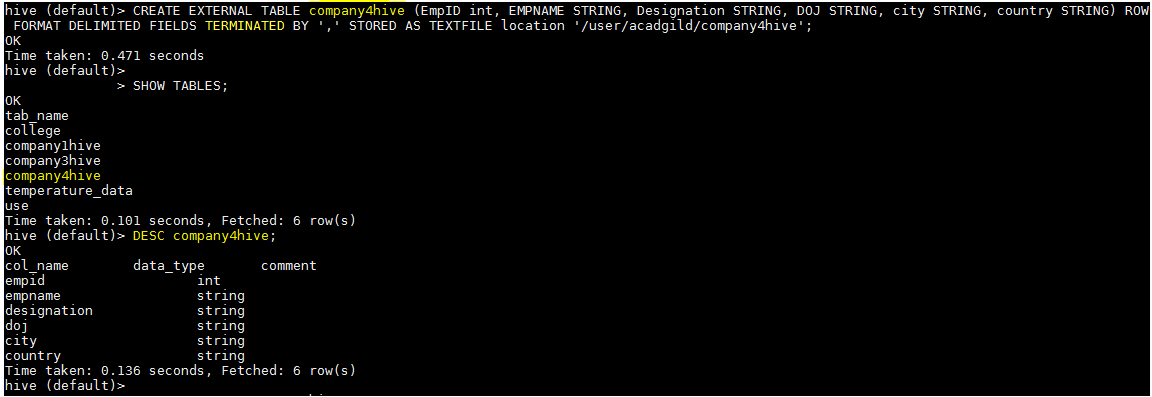
# Incremental import to HDFS and mapping Hive table to sqoop's target-dir.

A full example is shown below,

1. Define your Hive table as external table, a new external table **company4hive** is going to be created

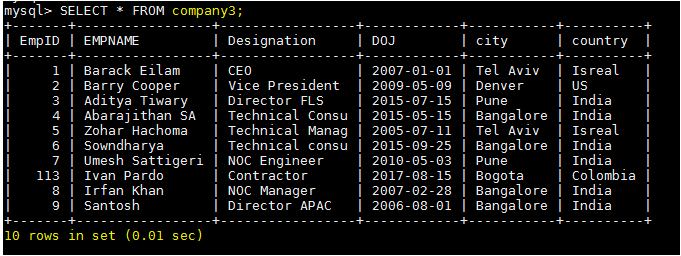
**Hive command**

***CREATE EXTERNAL TABLE company4hive (EmpID int, EMPNAME STRING, Designation STRING, DOJ STRING, city STRING, country STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE location '/user/acadgild/company4hive';***



1. Initially it's empty. Do the first sqoop, importing 10 entries from MySql table **company3** (having id's set to 1-9), setting target-dir to the location of our external table ***/user/acadgild/company4hive***

**Mysql>SELECT \* FROM company3;**



**Sqoop Command**

***sqoop import --connect jdbc:mysql://localhost/company --username root --password acadgild --table company3 --target-dir /user/acadgild/company4hive -m 1 --incremental append -check-column EmpID***



Now check the data in the hive table **‘company4hive’**. We have the data imported from the mysql as we see the below screen shot, it have fetched 10 rows.

**hive (default)>**

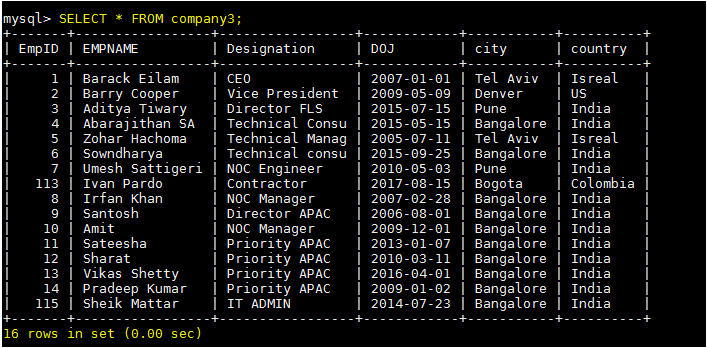
**> SELECT \* FROM company4hive;**



1. Now append new 6 entries to MySql table **company3**, 9<=EmpID and do incremental append import setting last-value to 9.

***INSERT INTO company3(EmpID,EMPNAME,Designation,DOJ,city,country) VALUES (10,'Amit','NOC Manager','2009-12-01','Bangalore','India'),(11,'Sateesha','Priority APAC','2013-01-07','Bangalore','India'),(12,'Sharat','Priority APAC','2010-03-11','Bangalore','India'),(13,'Vikas Shetty','Priority APAC','2016-04-01','Bangalore','India'),(14,'Pradeep Kumar','Priority APAC','2009-01-02','Bangalore','India'),(115,'Sheik Mattar','IT ADMIN','2014-07-23','Bangalore','India');***

***mysql> SELECT \* FROM company3;***



Sqoop Command:

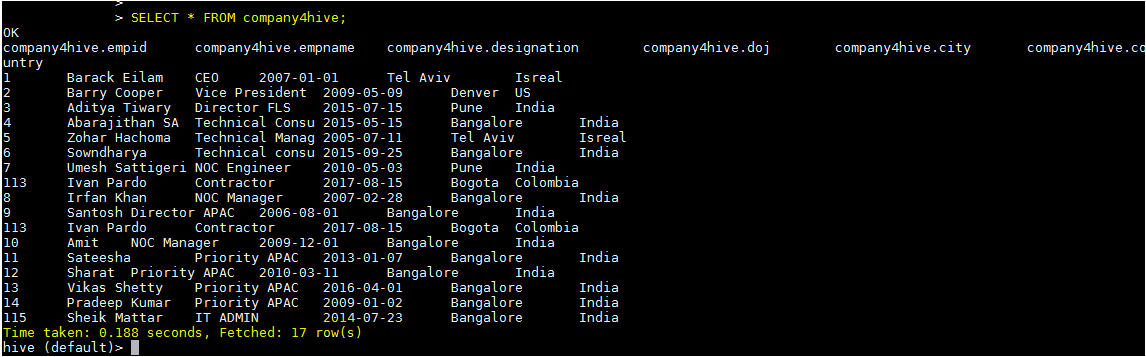
***sqoop import --connect jdbc:mysql://localhost/company --username root --password acadgild --table company3 --target-dir /user/acadgild/company4hive -m 1 --incremental append -check-column EmpID --last-value 9***



Check it in the hive table,

***hive (default)>***

***> SELECT \* FROM company4hive;***



Hence, the incremental append from mysql to hive has been succeeded by performing ***“Incremental import to HDFS and mapping Hive table to sqoop's target-dir”***

Note:- in the hive table we have got 17 rows instead of 16 rows, this is due to the contractor **EmpID** **113**which is greater than the key value 9, hence it is repeated 2 times.