

Lab - Ingesting RDBMS Data to Hadoop using Sqoop with hive import.  
=====

CopyRight - Big Data Trunk LLC  
www.BigDataTrunk.com  
Twitter - @BigDataTrunk

=====

Use this command reference file to copy and paste text for your lab.

=====

Instructions:

1. The Ingesting RDBMS Data to Hadoop using Sqoop lab is to get familiar with Sqoop commands
2. Imports tables from My SQL to Hadoop using Sqoop import and other commands
3. Finally verify the output files in HDFS

=====

#Login to MySQL from terminal

#Command: \$ mysql -u root -p  
mysql -u root -p  
#Note: Password: cloudera (password will not visible but we have to type)

#Command: mysql>create database Dimensions;  
create database Dimensions;

#Command: mysql> use Dimensions;  
use Dimensions;  
#Explanation: We have to enter into a particular database for this we need to use above command.

Mysql>source /home/cloudera/Products.sql;  
#To execute sql file for creating table and then inserting data into that table.By using source command

#Command: mysql> desc Products;  
desc orders;  
#Explanation: To find table schema of a specific table in mysql.

#Command: mysql>select \* from Products;  
select \* from Products;  
#Explanation: List out all rows of a particular table in a specific database.

#Command: mysql>quit;  
quit;  
#Explanation: To Exit from mysql.

#Sqoop commands:  
#\*\*\*\*\*

#In sqoop every command starts with sqoop word.

#Command: sqoop help  
#Explanation: generate list of sqoop commands and there usage.  
help

#Command: sqoop import\  
#Explanation:Importing the Users table of Dimensions database in hadoop.

# warehouse-dir too is used create a directory with table name.  
# map-column-java tool is used to map the mysql columns with generated java class.  
# P tool is used ask password on execution time

sqoop import --connect jdbc:mysql://localhost:3306/Dimensions --username root -P --table Products --warehouse-dir /Locations --map-column-java ISBN=String,Book-Title=String,Book-Author=String,Year-Of-Publication=Integer,Publisher=String,Image-URL-S=String,Image-URL-M=String,Image-URL-L=String;

HDFS Location : /Locations/Products

#By default sqoop runs four mappers when importing and exporting data.  
#While running sqoop job,it will create a java classes and jar files.We can see that java file like this  
#location: local filesystem. /tmp/sqoop-cloudera/

=====

#Hive external table for the HDFS data:

=====

#First need to login to hive shell  
#Type hive

hadoop@hadoop:hive

#creating database in hive  
hive>create database Dimensions;

#moving into database  
#use database\_name;  
hive>use Dimensions;

#serde is main feature of hive because of serde internally converts records into columns and columns into records.  
#LazySimpleSerde will be called by default serde when creating table in hive.  
#location is where to locate hive external table.

create external table Products(ISBN string,Book\_Title string,Book\_Author string,Year\_Of\_Publication int,Publisher string,Image\_URL\_S string,Image\_URL\_M string,Image\_URL\_L string) row format delimited fields terminated by ',' location '/Locations/Products';

=====

#Directly can connect to hive through sqoop.By using --hive-import tool  
#hive-import tool is used to import data into hive warehouse (/user/hive/warehouse)and it will automatically created table in default database in hive with same mysql table name.  
#map-column-hive tool is used to map the mysql columns with generated java class,and then hive columns.  
#In the process of sqoop with hive import,first data copy into temporary directory in hdfs,and then transfer data into hive warehouse directory.  
#Hive table will be automatically created in default database with mysql table name.

With hive-import tool:

sqoop import --connect jdbc:mysql://localhost:3306/Dimensions --username root -P --table Products --hive-import --map-column-java ISBN=String,Book\_Title=String,Book\_Author=String,Year\_Of\_Publication=Integer,Publisher=String,Image\_URL\_S=String,Image\_URL\_M=String,Image\_URL\_L=String