My SQL Installation:

- Install on ubuntu:

  sudo apt-get update

  sudo apt-get install mysql-server

  Ref: https://www.digitalocean.com/community/tutorials/how-to-install-mysql-on-ubuntu-16-04

- Install JDBC driver

  sudo apt-get install libmysql-java

  Ref: https://help.ubuntu.com/community/JDBCAndMySQL

- Create new user in mysql

  mysql -u root -p

  GRANT ALL PRIVILEGES ON *.* TO 'username'@'localhost' IDENTIFIED BY 'password';

  Ref:
  https://www.a2hosting.com/kb/developer-corner/mysql/managing-mysql-databases-and-users-from-the-command-line

- Login MySQL:

  mysql –u retail_dba –p

- Create database

  CREATE DATABASE dbname;

- Create Table in Database mydb

  USE mydb;

  CREATE TABLE Test (id INT NOT NULL PRIMARY KEY, name VARCHAR(20));

- Know host port

  select user();

  SHOW VARIABLES WHERE Variable_name = 'port';

  Ref:
  https://stackoverflow.com/questions/4093603/how-do-i-find-out-my-mysql-url-host-port-and-username

Sqoop Installation:

1. Download Sqoop from any mirror

   http://apache.mirror.rafal.ca/sqoop/1.4.7/

   Download both files, sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz

   and sqoop-1.4.7.tar.gz

2. Extract

   sudo tar -zxvf sqoop-1.4.7.tar.gz

   sudo tar -zxvf sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz

3. Copy sqoop-1.4.7.jar from sqoop-1.4.7.bin__hadoop-2.6.0 to sqoop-1.4.7
4. Modify bashrc file

   export SQOOP_HOME=$HADOOP_INSTALL/sqoop

   export PATH=$PATH:$SQOOP_HOME/bin

   Note: my sqoop files are present at $HADOOP_INSTALL/sqoop location, for you it might be different

5. Test

   sqoop version

Exercise:

Ref: https://sqoop.apache.org/docs/1.4.6/SqoopUserGuide.html

1. Sqoop Export

Create table in Hive and insert data into it:

CREATE TABlE test(id INT, name STRING) ROW FORMAT delimited fields terminated by ',' LINES TERMINATED BY '\n' STORED AS TEXTFILE;

Insert into test values(1,'name1');

Insert into test values(2,'name2');

Export Hive table to mysql:

sqoop export --connect jdbc:mysql://localhost :3306/retail_db --username retail_dba --password cloudera --table test --fields-terminated-by ',' --export-dir <HDFS DIRCTORY NAME>

See the data into mysql table:

Select * from test;

Try Out

1. Update records in RDBMS table based on data stored in HDFS

2. Sqoop Import

Import Table from RDBMS

sqoop import --connect jdbc:mysql://localhost :3306/<DATABASE NAME> --username root –password cloudera --table <TABLE NAME> --m 1 --target-dir <HDFS DIRCTORY NAME>

Import data with a condition

sqoop import --connect jdbc:mysql://localhost :3306/<DATABASE NAME> --username root -p --table <TABLE NAME> --m 1 --where "<CONDITION>" --target-dir <HDFS DIRCTORY NAME>

Try out

1. Delete target directory
2. Append imported records to existing data

3. Sqoop Jobs

Create Job

sqoop job --create myjob -- import --connect jdbc:mysql:// localhost :3306/retail_db --username retail_dba --password cloudera --table departments --target-dir <HDFS DIRCTORY NAME> --fields-terminated-by ‘,’

List all created jobs

sqoop job --list

Show details of one specific job

sqoop job --show myjob

Execute created job

sqoop job --exec myjob

4. codegen

Generate java code

sqoop codegen --connect jdbc:mysql:// localhost :3306/retail_db --username retail_dba --password cloudera --table departments

5. eval

Insert a record into RDBMS table

sqoop eval -- connect jdbc:mysql:// localhost :3306/retail_db --username retail_dba --password cloudera -e "INSERT INTO Test VALUES(999, 'name999')"