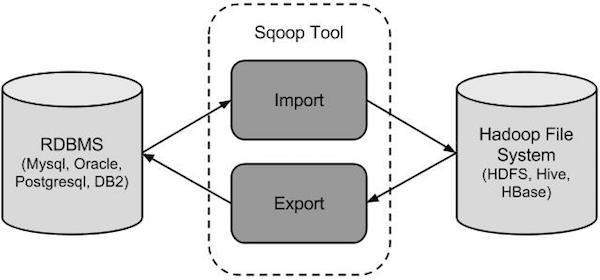
Sqoop - Introduction

**Sqoop:** “SQL to Hadoop and Hadoop to SQL”

Sqoop is a tool designed to transfer data between Hadoop and relational database servers. It is used to import data from relational databases such as MySQL, Oracle to Hadoop HDFS, and export from Hadoop file system to relational databases. It is provided by the Apache Software Foundation.

## **How Sqoop Works?**



## **Sqoop Import**

The import tool imports individual tables from RDBMS to HDFS. Each row in a table is treated as a record in HDFS. All records are stored as text data in text files or as binary data in Avro and Sequence files.

## **Sqoop Export**

The export tool exports a set of files from HDFS back to an RDBMS. The files given as input to Sqoop contain records, which are called as rows in table. Those are read and parsed into a set of records and delimited with user-specified delimiter.

# Sqoop - Installation

## **Step 1: Verifying JAVA Installation**

You need to have Java installed on your system before installing Sqoop. Let us verify Java installation using the following command:

$ java –version

## **Step 2: Verifying Hadoop Installation**

Hadoop must be installed on your system before installing Sqoop. Let us verify the Hadoop installation using the following command:

$ hadoop version

### **Step 3: Downloading Sqoop**

We can download the latest version of Sqoop from the following [link](http://mirrors.ibiblio.org/apache/sqoop/1.4.5/) For this tutorial, we are using version 1.4.5, that is, **sqoop-1.4.5.bin\_\_hadoop-2.0.4-alpha.tar.gz**.

### **Step 4: Installing Sqoop**

The following commands are used to extract the Sqoop tar ball and move it to “/usr/lib/sqoop” directory.

$tar -xvf sqoop-1.4.4.bin\_\_hadoop-2.0.4-alpha.tar.gz

$ su

password:

# mv sqoop-1.4.4.bin\_\_hadoop-2.0.4-alpha /usr/lib/sqoop

#exit

### **Step 5: Configuring bashrc**

You have to set up the Sqoop environment by appending the following lines to ~/**.bashrc** file:

#Sqoop

export SQOOP\_HOME=/usr/lib/sqoop export PATH=$PATH:$SQOOP\_HOME/bin

The following command is used to execute ~/**.bashrc** file.

$ source ~/.bashrc

### **Step 6: Configuring Sqoop**

To configure Sqoop with Hadoop, you need to edit the **sqoop-env.sh** file, which is placed in the **$SQOOP\_HOME/conf** directory. First of all, Redirect to Sqoop config directory and copy the template file using the following command:

$ cd $SQOOP\_HOME/conf

$ mv sqoop-env-template.sh sqoop-env.sh

Open **sqoop-env.sh** and edit the following lines:

export HADOOP\_COMMON\_HOME=/usr/local/hadoop

export HADOOP\_MAPRED\_HOME=/usr/local/hadoop

### **Step 7: Download and Configure mysql-connector-java**

We can download **mysql-connector-java-5.1.30.tar.gz** file from the following [link](http://ftp.ntu.edu.tw/MySQL/Downloads/Connector-J/).

The following commands are used to extract mysql-connector-java tarball and move **mysql-connector-java-5.1.30-bin.jar** to /usr/lib/sqoop/lib directory.

$ tar -zxf mysql-connector-java-5.1.30.tar.gz

$ su

password:

# cd mysql-connector-java-5.1.30

# mv mysql-connector-java-5.1.30-bin.jar /usr/lib/sqoop/lib

### **Step 8: Verifying Sqoop**

The following command is used to verify the Sqoop version.

$ cd $SQOOP\_HOME/bin

$ sqoop-version

Expected output:

14/12/17 14:52:32 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5

Sqoop 1.4.5 git commit id 5b34accaca7de251fc91161733f906af2eddbe83

Compiled by abe on Fri Aug 1 11:19:26 PDT 2014

Sqoop installation is complete.