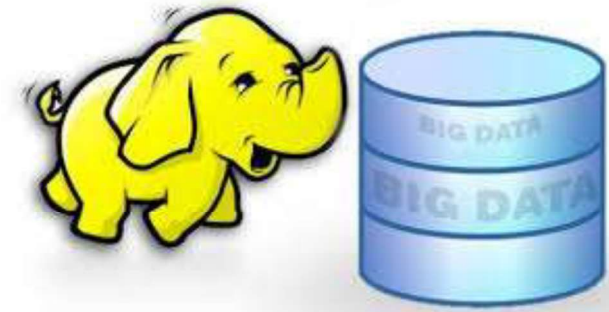




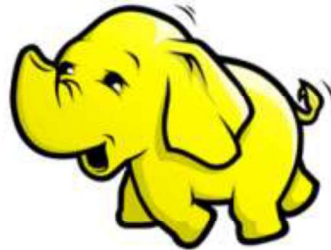
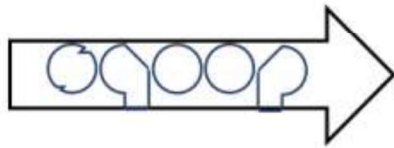
रा.इ.सू.प्रौ.सं
NIELIT



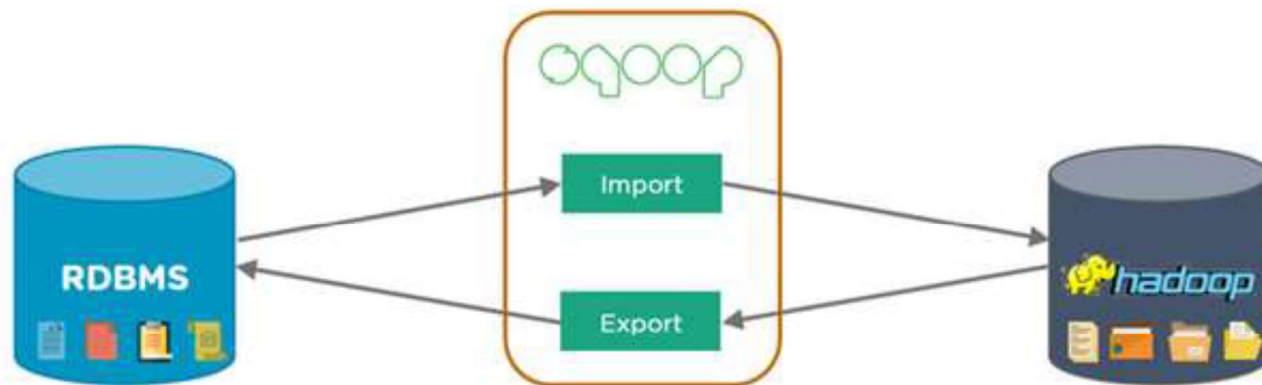
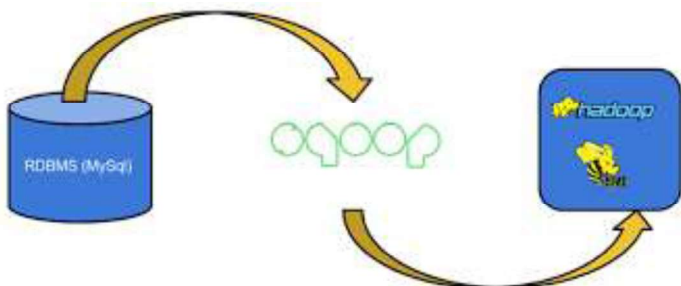
National Institute of Electronics and Information Technology

Big Data and Hadoop

Module 4-Exploring Sqoop




Apache Sqoop



Motivation

1. In traditional application management system, the interaction of **applications with relational database using RDBMS, is one of the sources that generate Big Data.**
2. As more organizations deploy Hadoop to analyze vast streams of information, they find the need to transfer large amount of data between Hadoop and their existing databases.
3. Loading bulk data into Hadoop or accessing it from map-reduce applications is a challenging task. Transferring data using script is an inefficient and time-consuming task
4. Such Big Data, generated by RDBMS, is stored in Relational Database Servers in the relational database structure.
5. When Big Data storages and analyzers such as **MapReduce, Hive, Hbase, Pig, etc. of the Hadoop ecosystem** came into picture, they required a tool to interact with the relational database servers **for importing and exporting the Big Data residing in them.**
6. Here, **Sqoop occupies a place in the Hadoop ecosystem to provide feasible interaction between relational database server and Hadoop's HDFS.**

Advantage of Sqoop

1. Sqoop: “SQL to Hadoop and Hadoop to SQL” Sqoop is a tool designed to transfer data between Hadoop and relational database servers. It is provided by the Apache Software Foundation.
2. As per Hadoop Developers, the real work starts when the data is loaded into HDFS. It is possible to write the custom coding or scripts for transferring data to and from Hadoop, but Apache Sqoop is an alternative that lets you do the same tasks in a simple yet efficient manner.
3. It allows easy import and export of data from structured datastores. Relational Database, Enterprise data warehouses and NoSQL Systems.
4. Sqoop lets you automate the process and depending on the database it is used to **describe the schema** of the data to be imported.
5. Sqoop works on the **MapReduce framework** for importing and exporting data, providing a parallel fault-tolerant mechanism. However, here, developers need to understand the database authentication, source, destination, and such other information.
6. Sqoop can also be deployed for exporting data into external datastores, when necessary. The Sqoop architecture is a **connector-based architecture** that can support plugins, which provides connectivity to new external sources.

Installing Sqoop

1. We can download the latest version of Sqoop from the link <http://archive.apache.org/dist/> For this tutorial, we are using version 1.4.7, that is, **sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz**.
2. The following commands are used to extract the Sqoop tar ball and move it to “/usr/local/sqoop” directory.

```
$ tar -xvf sqoop-1.4.7.bin__hadoop-2.6.0.tar.gz  
$ sudo su  
password:  
# mv sqoop-1.4.7.bin__hadoop-2.6.0 /usr/local/sqoop  
# exit
```

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

```
export SQOOP_HOME=/usr/local/sqoop  
export PATH=$PATH:$SQOOP_HOME/bin
```

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

```
$ source ~/.bashrc
```

Installing Sqoop

5. 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
```

6. Open sqoop-env.sh and edit the following lines –

```
export HADOOP_COMMON_HOME=${HADOOP_HOME}  
export HADOOP_MAPRED_HOME=${HADOOP_HOME}
```

7. We can download mysql-connector-java-5.1.49.tar.gz file from the following link <http://ftp.ntu.edu.tw/MySQL/Downloads/Connector-J/>
8. The following commands are used to extract mysql-connector-java tarball and move mysql-connector-java-5.1.49-bin.jar to /usr/local/sqoop/lib directory.

```
$ tar -zxf mysql-connector-java-5.1.49.tar.gz  
$ su  
password:  
# cd mysql-connector-java-5.1.49  
# mv mysql-connector-java-5.1.49-bin.jar /usr/local/sqoop/lib
```

9. Verify Sqoop

```
$ sqoop-version
```