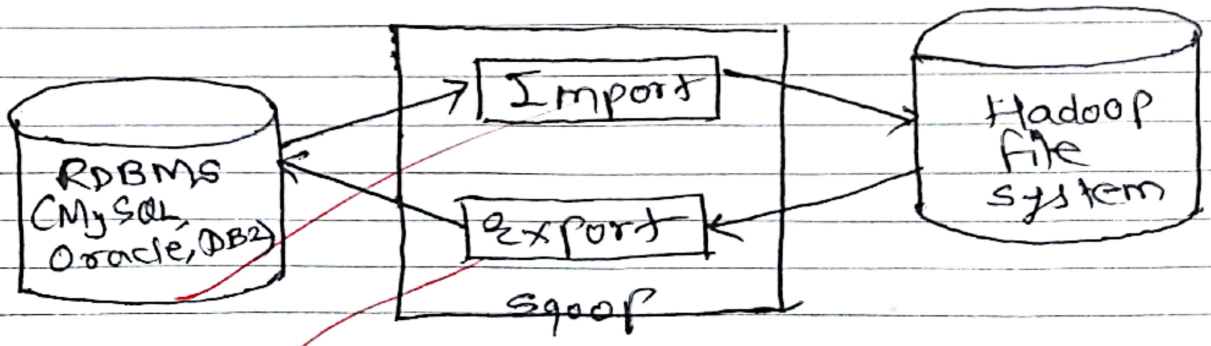
Experiment No : 2

Aim :- Use of sqoop tool to transfer the data between Hadoop and relational database servers
a) Sqoop - installation
b) To execute Basic Commands of sqoop

System Software : Ubuntu 16.04, Hadoop, sqoop

Theory / Working Principle
- Sqoop is a data Ingestion tool, specifically it is a command line interface application for transferring data between relational database and Hadoop
- Simple as user specifies the "what" and leave the "how" to the underlying processing Engine
- Rapid development
- No JAVA Knowledge required



Installation of Sqoop and Theory
(1) Apache Sqoop is a tool designed for efficiently transferring bulk data

between Apache Hadoop and structured datastores such as relational databases For example: MySQL, Oracle, Microsoft SQL server. Sqoop ships as one binary package that incorporates two seperate parts - client and server

Steps of installation

(1) Download Sqoop using wget
(2) Extract Sqoop tar file
(3) Move the Sqoop Directory
(4) Set Hadoop and Sqoop Environment variables
(5) Copy Required Jar Files to Sqoop Server lib Directory
(6) Edit core-site.xml
(7) Initialize Metadata Repository
(8) Start sqoop Server
(9) start Sqoop client
(10) Download RDBMS Connectors
(11) Set an Environment Variable to use RDBMS Connectors

Conclusion: We have successfully setup Apache Sqoop on Ubuntu 16.04. Now we are ready to import/export data using Sqoop. The next step is to use any of the RDBMS connectors and import/export data from RDBMS to HDFS