**Sqoop**

• Sqoop allows easy import and export of data from structured data stores such as

relational databases, enterprise data warehouses, and NoSQL systems.

• Using Sqoop, we can provision the data from external system on to HDFS, and populate

tables in Hive and HBase.

• Sqoop integrates with Oozie, allowing you to schedule and automate import and export

tasks.

• Sqoop uses a connector based architecture which supports plugins that provide

connectivity to new external systems.

• The following command is used to import all data from a table called ORDERS from a

MySQL database:

---

sqoop import --connect jdbc:mysql://localhost/acmedb --table ORDERS --username test --

password \*\*\*\*

• In this command the various options specified are as follows:

import:

• This is the sub-command that instructs Sqoop to initiate an import.

--connect <connect string>,

• Connection String contains type of RDBMS Server, RDBMS Sever IP Address, port,

database name and other optional arguments

--username <user name>,

--password <password>:

• These are connection parameters that are used to connect with the database. This is no

different from the connection parameters that you use when connecting to the database

via a JDBC connection.

--table <table name>:

• This parameter specifies the table which will be imported.

• The import is done in two steps:

• In the first Step Sqoop introspects the database to gather the necessary metadata for

the data being imported.

• The second step is a map-only Hadoop job that Sqoop submits to the cluster.

• It is this job that does the actual data transfer using the metadata captured in the

previous step.

