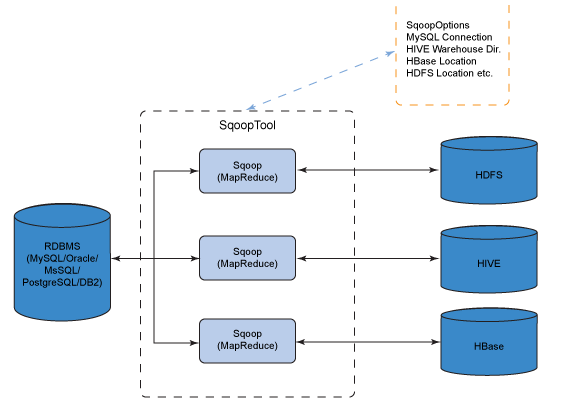
● The workflow of Sqoop and its Benefits

Sqoop is a command-line interface application for transferring data between relational databases and Hadoop.

It supports incremental loads of a single table or a free form SQL query as well as saved jobs which can be run multiple times to import updates made to a database since the last import.Using Sqoop, Data can be moved into HDFS/hive/hbase from MySQL/ PostgreSQL/Oracle/SQL Server/DB2 and vise versa.



sqoop TOOL PROPERTY\_ARGS SQOOP\_ARGS [-- EXTRA\_ARGS]

* TOOL indicates the operation eg: "import", "export".
* PROPERTY\_ARGS are Java properties in the format "-Dname=value"
* SQOOP\_ARGS mention various Sqoop parameters
* EXTRA\_ARGS are for specialized connectors, separated from the SQOOP\_ARGS with a "--"
* -m 1 specifies one mapper for each table.

Eg. % sqoop import --connect jdbc:mysql://localhost/hadoopguide --table widgets -m 1

When you have decided to move data from RDBMS to HDFS then the first product which comes into use is Apache Sqoop. When you request to bring the data to HDFS then the following things happen.

1. Sqoop asks for metadata information from Relation DB.
2. Relational DB returns the required request.
3. Based on metadata information Sqoop generates java classes.
4. Based on primary id partitioning happens in table as multiple mappers will importing data as the same time.

sqoop import --connect "jdbc:mysql://localhost/training" --username cloudera -P --table

country -target-dir /user/country\_imported

Using where clause:

sqoop import --connect "jdbc:mysql://localhost/training" --username training -P --table cityByCountry --target-dir /user/where\_clause --where "state = 'Alaska'" -m 1

To list all databases and tables present on a mysql server using a "list-databases" tool.

$ sqoop list-databases --connect "jdbc:mysql://localhost" --username cloudera --password cloudera

$ sqoop list-tables --connect "jdbc:mysql://localhost/training" --username cloudera -P

Import data to Hive:

sqoop import --connect "jdbc:mysql://localhost/training" --username training -P --table cityByCountry --target-dir /user/where\_clause --where "state = 'Alaska'" --import -hive -m 1

**Export**

Before performing export, Sqoop fetches table metadata from MySQL database. Thus we first need to create a table with required metadata.

mysql>Create table table\_name( column\_name column\_type )

Export query

sqoop export --connect jdbc:mysql://localhost/cloudera --username cloudera -P  --table exported  --export-dir /user/country\_imported/part-m-00000

# Sqoop – Job

 Sqoop job creates and saves the import and export commands. It specifies parameters to identify and recall the saved job. This re-calling or re-executing is used in the incremental import, which can import the updated rows from RDBMS table to HDFS.

Following are the commands to work with sqoop job

* To create:

sqoop job --create myjob \

--import \

--connect jdbc:mysql://localhost/db \

--username root \

--table employee --m 1

* To get list of created jobs

$ sqoop job –list

* To inspect job:

sqoop job --show myjob

* Execute

$ sqoop job --exec myjob