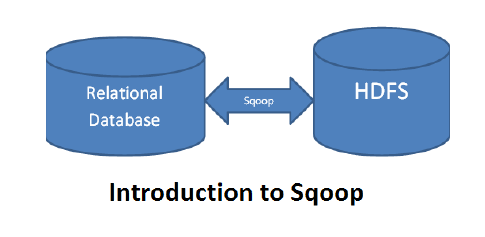
**SQOOP**

Apache Sqoop is a tool designed for efficiently transferring bulk data between Apache

Hadoop and structured datastores such as relational databases.

Apache Sqoop is a tool designed for efficiently transferring bulk data between Apache

Hadoop and structured datastores such as relational databases



**Features:-**

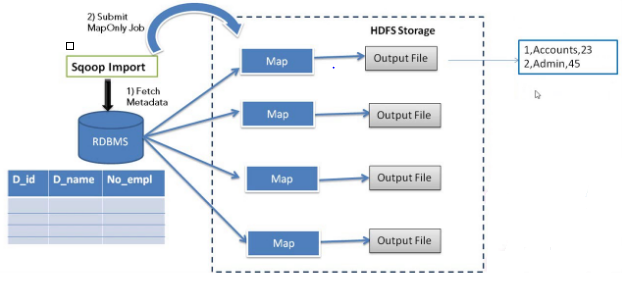
import :- Import a table from a Source system database to HDFS

Import-all-tables :- Import tables from a Source system database to HDFS

List-databases :- List availadatabases on a Source system server

List-tables :-List available tables in a database in Source system

Export :- Export a tables from HDFS/Hive to RDBMS system



**Apache Sqoop Tutorial: Flume vs Sqoop**

The major difference between Flume and Sqoop is that:

* Flume only ingests unstructured data or semi-structured data into HDFS.
* While Sqoop can import as well as export structured data from RDBMS or Enterprise data warehouses to HDFS or vice versa.

**sqoop-import**

• The import tool imports an individual table from an RDBMS to HDFS. Each row from a

table is represented as a separate record in HDFS. Records can be stored as text files

(one record per line), or in binary representation as Avro or SequenceFiles.

**Syntax:**

$ sqoop import (generic-args) (import-args)

**Sqoop Export:-**

The export tool exports a set of files from HDFS back to an RDBMS. The target table must already exist in the database. The input files are read and parsed into a set of records according to the user-specified delimiters..

The default operation is to transform these into a set of INSERT statements that inject there records into the database.

**Syntax:**

– $ sqoop export (generic-args) (export-args)

– $ sqoop-export (generic-args) (export-args)

**Sqoop Create a Hive tables:-**

The create-hive-table tool populates a Hive metastore with a definition for a table based on a database table previously imported to HDFS, or one planned to be imported. This effectively performs the “--hive-import" step of sqoop-import without running the preceding import.

**Syntax:**

$ sqoop create-hive-table (generic-args) (create-hive-table-args)

**Example:**

$ sqoop create-hive-table --connect Jdbc:mysql://db.example.com/corp \ --table

employees --hive-table emps

**Sqoop File Format:-**

Sqoop supports three different file formats; one of these is text, and the other two are binary. The binary formats are Avro and Hadoop’s SequenceFile. You can enable import into SequenceFile using the --as-sequencefile parameter:

sqoop import --connect jdbc:mysql://localhost/vel --username root --password root -- table cities –as-sequencefile

Avro can be enabled by specifying the --as-avrodatafile parameter:

sqoop import --connect jdbc:mysql://localhost/vel --username root --password root --table cities --as-avrodatafile

**Sqoop Compression :-**

To compress Importing Data Use the parameter --compress to enable compression

sqoop import --connect jdbc:mysql://localhost/vel --username root --table cities --compress

By default - GZip codec and all files will end up with a .gz extension.

Choose codec - --compression-codec parameter.

**Splitable**  **Not Splitable**

BZip2,LZO GZip,Snappy

##################################################################################