



Vidyavardhini's College of Engineering & Technology

Department of Artificial Intelligence and Data Science

Experiment No. 2
Use of Sqoop tool
Date of Performance: 25/07/23
Date of Submission: 01/08/23



AIM: To install SQOOP and execute basic commands of Hadoop ecosystem component Sqoop.

THEORY:

Installation and configuration of SQOOP

1) Download SQOOP from <https://sqoop.apache.org>

2) Unzip and Install SQOOP

After Downloading the SQOOP, we need to Unzip the sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz file.

3) Create a folder and move the final extracted file in it.

4) Set up the environment variables

a. Set SQOOP_HOME

b. Set up path variable

5) Configure SQOOP

Basic SQOOP commands:

1. List Table

This command lists the particular table of the database in MYSQL server.

```
sqoop list - tables --connect jdbc:mysql://localhost/payment --username gartner
```

2. Target directory

This command import table in a specific directory in HDFS. -m denotes mapper argument. They have an integer value.

```
$ sqoop import --connect jdbc:mysql://localhost/inventory --username jony -table inventory --m 1 --target-dir/inv
```

3. sqoop-eval

This command runs quickly SQL queries of the respective database.

```
$ sqoop eval --connect --query "SQLQuery"
```



4. sqoop – version

This command displays version of the sqoop.

```
$ sqoop version      sqoop {revnumber}
```

5. sqoop-job

This command allows us to create a job, the parameters that are created can be invoked at any time. They take options like (–create,–delete,–show,–exit).

```
$ sqoop job --create --import --connect --table
```

6. code gen

This Sqoop command creates java class files which encapsulate the imported records. All the java files are recreated, and new versions of a class are generated. They generate code to interact with database records. Retrieves a list of all the columns and their datatypes.

```
$ sqoop codegen --connect -table
```

7. List Database

This Sqoop command lists have all the available database in the RDBMS server.

```
>$ sqoop list - database -- connect
```

CONCLUSION:

In conclusion, Sqoop is a vital tool for any organization that aims to harness the power of Big Data and Hadoop. Its robust features, user-friendliness, and support for multiple databases make it an invaluable asset for data engineers and administrators seeking to integrate and manage data across different platforms. By utilizing Sqoop, businesses can enhance their data processing capabilities, ultimately leading to better-informed decision-making and improved operational efficiency.