



Experiment No.2
Use of Sqoop tool
Date of Performance: 24/07/23
Date of Submission: 31/07/23

AIM : To install SQOOP and execute basic commands of Hadoop eco system 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.



- Set up the environment variables
 - Set SQOOP_HOME
- 4) 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 SQL queries of the respective database.

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

4. sqoop – version

This command displays a 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).



Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

```
$ 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 data types.

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

7. List Database

This Sqoop command lists all the available databases in the RDBMS server.

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

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

The screenshot shows a web browser window with the address bar displaying <https://archive.apache.org/dist/sqoop/1.4.7/>. The page title is "Index of /dist/sqoop/1.4.7". Below the title is a table listing files in the directory.

Name	Last modified	Size	Description
Parent Directory		-	
sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz	2020-07-06 15:19	17M	
sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz.asc	2020-07-06 15:20	819	
sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz.md5	2020-07-06 15:19	71	
sqoop-1.4.7.tar.gz	2020-07-06 15:20	1.1M	
sqoop-1.4.7.tar.gz.asc	2020-07-06 15:19	819	
sqoop-1.4.7.tar.gz.md5	2020-07-06 15:20	53	



Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

```
Activities Terminal Sep 13 10:34
ubuntu@ubuntu-HP-Elite-Tower-600-G9-Desktop-PC: ~/sqoop

ubuntu@ubuntu-HP-Elite-Tower-600-G9-Desktop-PC:~$ ls
area.out  Desktop  Documents  matrix  matrix.c~  Pictures  snap
bloomfilter  Django-WebApp  Downloads  matrix.c  Music  Public
ubuntu@ubuntu-HP-Elite-Tower-600-G9-Desktop-PC:~$ mkdir sqoop
ubuntu@ubuntu-HP-Elite-Tower-600-G9-Desktop-PC:~$ cd sqoop/
ubuntu@ubuntu-HP-Elite-Tower-600-G9-Desktop-PC:~/sqoop$ ls
ubuntu@ubuntu-HP-Elite-Tower-600-G9-Desktop-PC:~/sqoop$ wget https://archive.apache.org/dist/sqoop/1.4.7/sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz
--2023-09-13 10:34:32-- https://archive.apache.org/dist/sqoop/1.4.7/sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz
Resolving archive.apache.org (archive.apache.org)... 65.108.204.189, 2a01:4f9:1a:a084::2
Connecting to archive.apache.org (archive.apache.org)[65.108.204.189]:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 17953604 (17M) [application/x-gzip]
Saving to: 'sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz'

sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz 100%[=====] 17.12M 1.51MB/s in 11s

2023-09-13 10:34:44 (1.57 MB/s) - 'sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz' saved [17953604/17953604]

ubuntu@ubuntu-HP-Elite-Tower-600-G9-Desktop-PC:~/sqoop$
```

```
Activities Terminal Sep 13 10:34
ubuntu@ubuntu-HP-Elite-Tower-600-G9-Desktop-PC: ~/sqoop

ubuntu@ubuntu-HP-Elite-Tower-600-G9-Desktop-PC:~$ ls
area.out  Desktop  Documents  matrix  matrix.c~  Pictures  snap
bloomfilter  Django-WebApp  Downloads  matrix.c  Music  Public
ubuntu@ubuntu-HP-Elite-Tower-600-G9-Desktop-PC:~$ mkdir sqoop
ubuntu@ubuntu-HP-Elite-Tower-600-G9-Desktop-PC:~$ cd sqoop/
ubuntu@ubuntu-HP-Elite-Tower-600-G9-Desktop-PC:~/sqoop$ ls
ubuntu@ubuntu-HP-Elite-Tower-600-G9-Desktop-PC:~/sqoop$ wget https://archive.apache.org/dist/sqoop/1.4.7/sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz
--2023-09-13 10:34:32-- https://archive.apache.org/dist/sqoop/1.4.7/sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz
Resolving archive.apache.org (archive.apache.org)... 65.108.204.189, 2a01:4f9:1a:a084::2
Connecting to archive.apache.org (archive.apache.org)[65.108.204.189]:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 17953604 (17M) [application/x-gzip]
Saving to: 'sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz'

sqoop-1.4.7.bin_hadoop-2.6.0.tar.gz 62%[=====] 10.71M 1.49MB/s eta 5s
```

CONCLUSION :



In conclusion, our research demonstrates that Apache Sqoop is a strong and adaptable tool for transferring data between relational databases and Hadoop effectively. Setting up necessary dependencies and configuring connection parameters are both necessary steps in the installation procedure. Sqoop is a useful tool for both data engineers and analysts because, once implemented, it offers an easy-to-use command-line interface and can be integrated into data workflows. Sqoop simplifies the process, providing a smooth data flow and assisting in more effective big data processing and analysis, regardless of whether the objective is to import data from a database into Hadoop or export data from Hadoop to a relational database.