Experiment No.2
Use of Sqoop tool
Date of Performance: 02/08/23
Date of Submission: 09/08/23

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 gatner
```

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).

```
$ 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
```

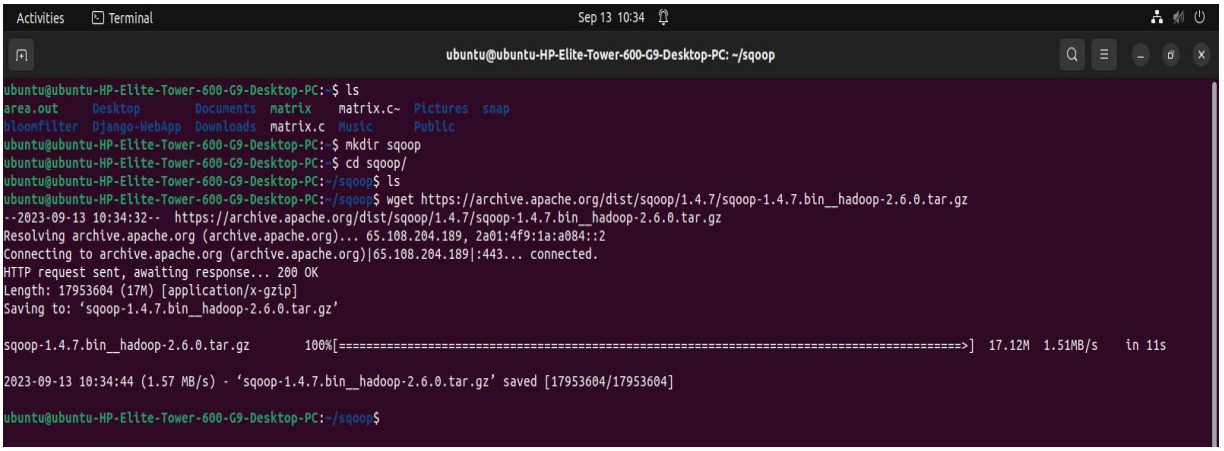
7. List Database

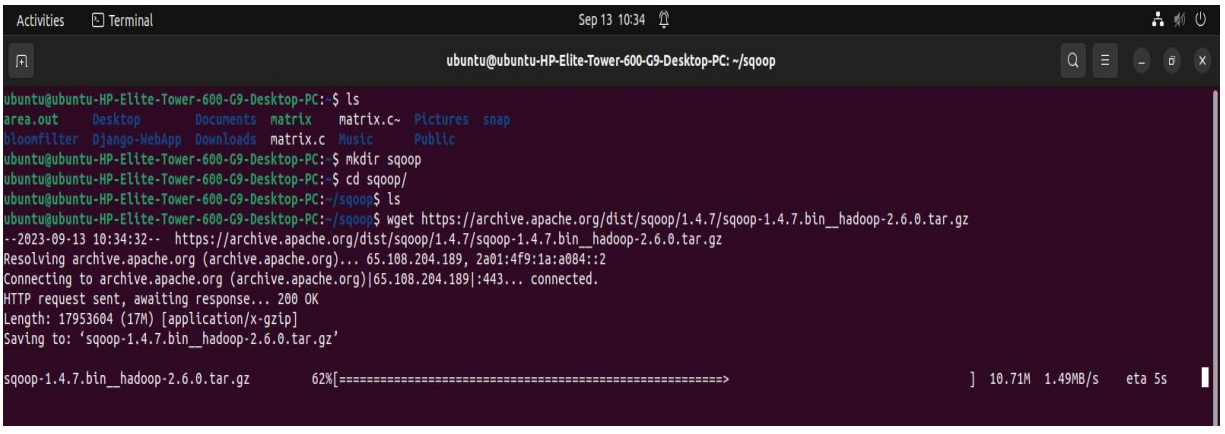
This Sqoop command lists all the available databases in the RDBMS server.

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

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

Activities
Firefox Web Browser
Sep 13 10:35
Namenode information
Set Up Containerize and
Document shared with
Installing Apache Sqoop
Index of /dist/sqoop/1.4.7
https://archive.apache.org/dist/sqoop/1.4.7/
Index of /dist/sqoop/1.4.7
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


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 summary, our findings confirm that Apache Sqoop is a potent and flexible solution for the seamless transfer of data between Hadoop and relational databases. The installation process is relatively straightforward, requiring the setup of essential dependencies and the configuration of connection parameters. Once deployed, Sqoop provides a user-friendly command-line interface and can be seamlessly integrated into data workflows, rendering it a valuable tool for data engineers and analysts alike. Whether the goal is to import data from a database into Hadoop or export data from Hadoop to a relational database, Sqoop streamlines the process, ensuring the smooth flow of data and contributing to more efficient big data processing and analysis.