Experiment No.2
Use of Sqoop tool
Name:- 9 Amruta Poojary
Date of Performance:17/8/23
Date of Submission:24/8/23

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.

CONCLUSION :

In summary, we determine that Apache Sqoop's installation and utilization serve as a potent and adaptable tool for the efficient transfer of data between Hadoop and relational databases. Its installation procedure is relatively uncomplicated, involving the setup of necessary dependencies and the configuration of connection parameters. After it's set up, Sqoop provides a user-friendly command-line interface and can be seamlessly integrated into data workflows, making it a valuable resource for data engineers and analysts. Whether your objective 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 movement of data and contributing to more effective big data processing and analysis.