Experiment No. 2
Use of Sqoop tool
Date of Performance:
Date of Submission:

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.
4) Set up the environment variables
   a. Set SQOOP_HOME
   b. Set up path variable
5) Configure SQOOP

OUTPUT:

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 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 anytime. 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 thejava 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:

Explain basic Scoop commands

Sqoop (SQL-to-Hadoop) is a tool designed to transfer data between relational databases and Hadoop. It allows importing data from relational databases like MySQL, Oracle, SQL Server, etc., into Hadoop's HDFS, Hive, or HBase and exporting data back to relational databases. Here are some basic Sqoop commands:

1. **Import Command**
   - **Purpose**: Imports a table from a relational database into HDFS.
   - **Syntax**:
     sqoop import \
     --connect jdbc:mysql://localhost/database_name \
     --username your_username \
     --password your_password \
     --table table_name \
     --target-dir /user/hadoop/target_directory
   - **Explanation**:
     - --connect: Specifies the JDBC connection string for the database.
     - --username: The username for the database.
     - --password: The password for the database.
     - --table: The name of the table to import.
     - --target-dir: The HDFS directory where the imported data will be stored.
2. **Export Command**
   - **Purpose**: Exports data from HDFS to a relational database table.
   - **Syntax**:
     sqoop export \
     --connect jdbc:mysql://localhost/database_name \
     --username your_username \
     --password your_password \
     --table table_name \
     --export-dir /user/hadoop/export_directory
   - **Explanation**:
     - --table: The target table in the relational database where data will be exported.
     - --export-dir: The HDFS directory containing the data to export.
3. **List Databases**

- **Purpose**: Lists all databases available in the connected RDBMS.
- **Syntax**:
  sqoop list-databases \
  --connect jdbc:mysql://localhost/ \
  --username your_username \

  --password your_password
- **Explanation**: This command shows all the databases available on the specified RDBMS server.

4. **List Tables**
   - **Purpose**: Lists all tables in a specified database.
   - **Syntax**:
     sqoop list-tables \
     --connect jdbc:mysql://localhost/database_name \
     --username your_username \
     --password your_password
   - **Explanation**: This command will list all the tables within the specified database.
5. **Importing a Specific Column**
   - **Purpose**: Imports only specified columns from a table.
   - **Syntax**:
     sqoop import \
     --connect jdbc:mysql://localhost/database_name \
     --username your_username \
     --password your_password \
     --table table_name \
     --columns "column1,column2,column3" \
     --target-dir /user/hadoop/target_directory
   - **Explanation**:
     - --columns: Specifies which columns to import from the table.
6. **Incremental Import**
   - **Purpose**: Imports only new data added to the source table since the last import.
   - **Syntax**:
     sqoop import \
     --connect jdbc:mysql://localhost/database_name \
     --username your_username \
     --password your_password \
     --table table_name \
     --target-dir /user/hadoop/target_directory \
     --incremental append \
     --check-column id_column \
     --last-value last_imported_value
   - **Explanation**:
     - --incremental append: Specifies that only new rows will be imported.

- --check-column: Specifies the column to check for new data (e.g., a timestamp or auto-incremented ID).
- --last-value: The last imported value of the check-column.