| 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

**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 any time. They take options like (–create,–delete,–show,–exit).

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

**6.codegen:**

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.

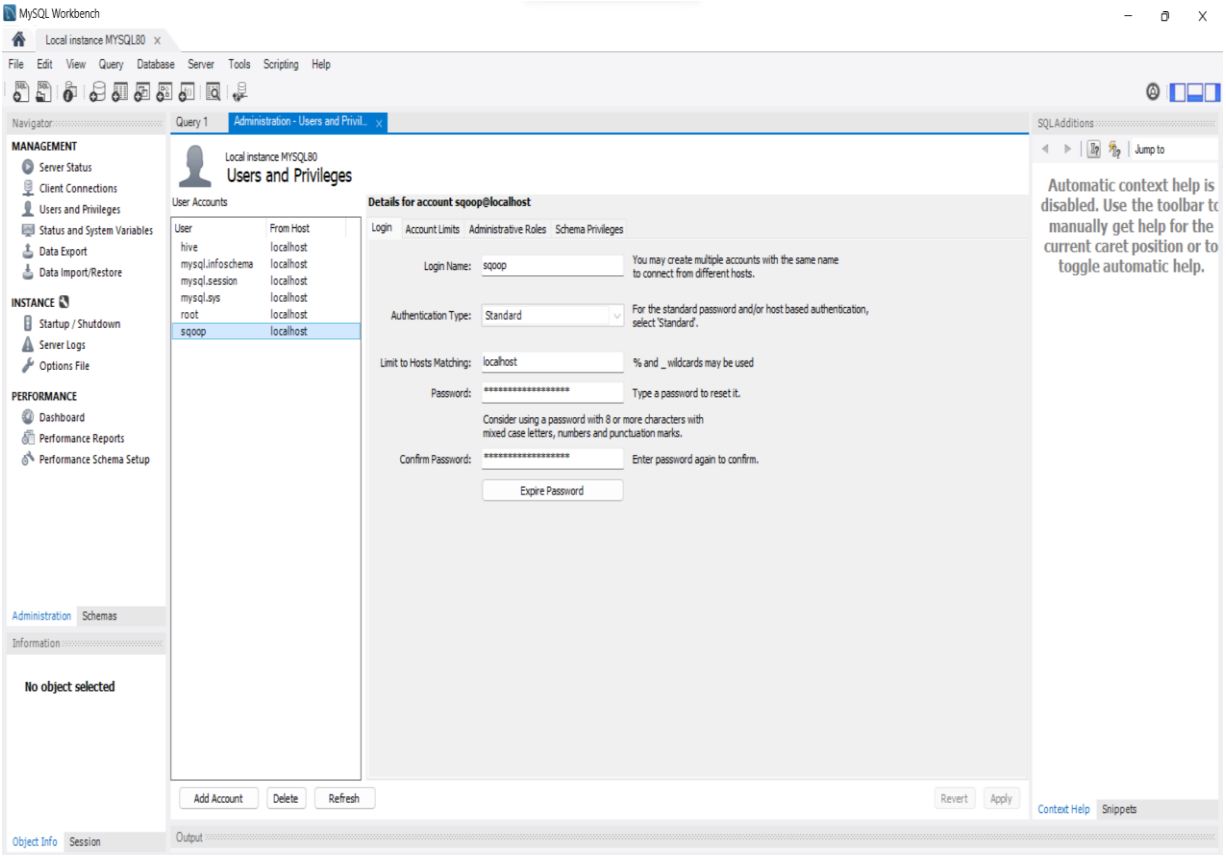
```
$ sqoop codegen --connect -table
```

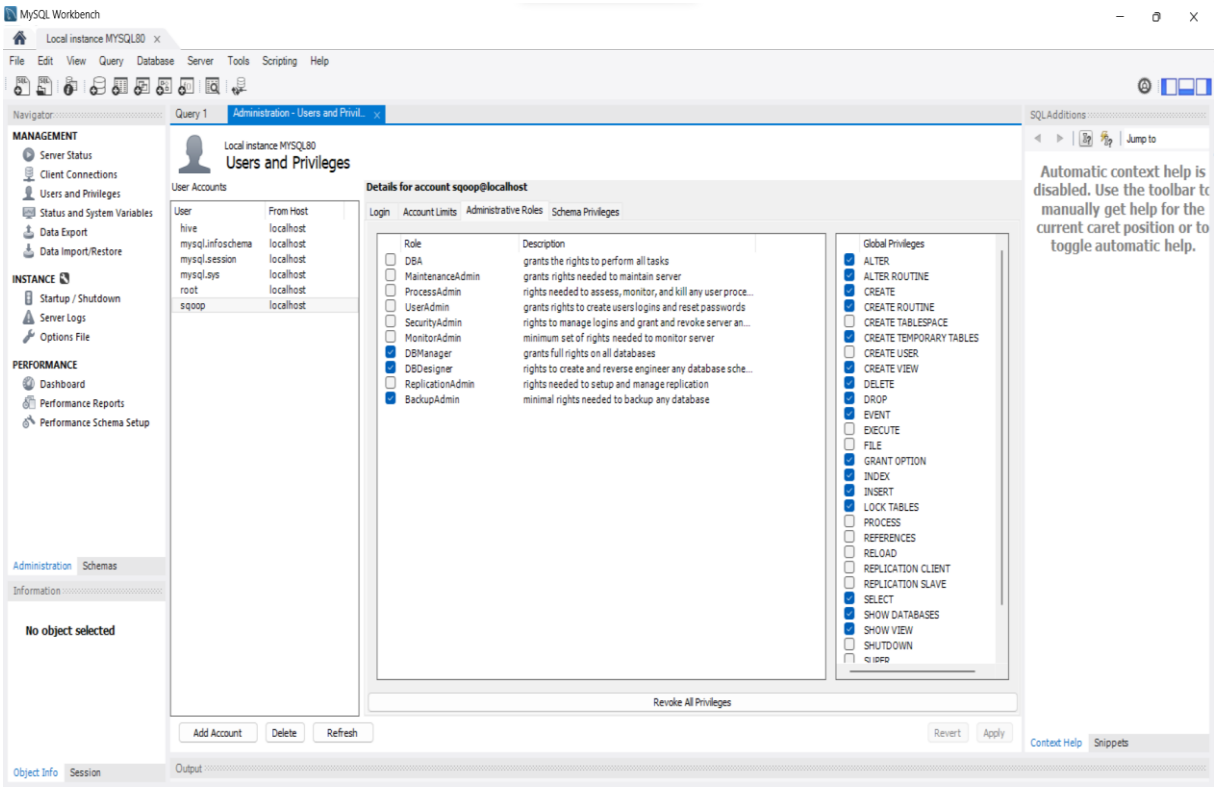
7. List Database

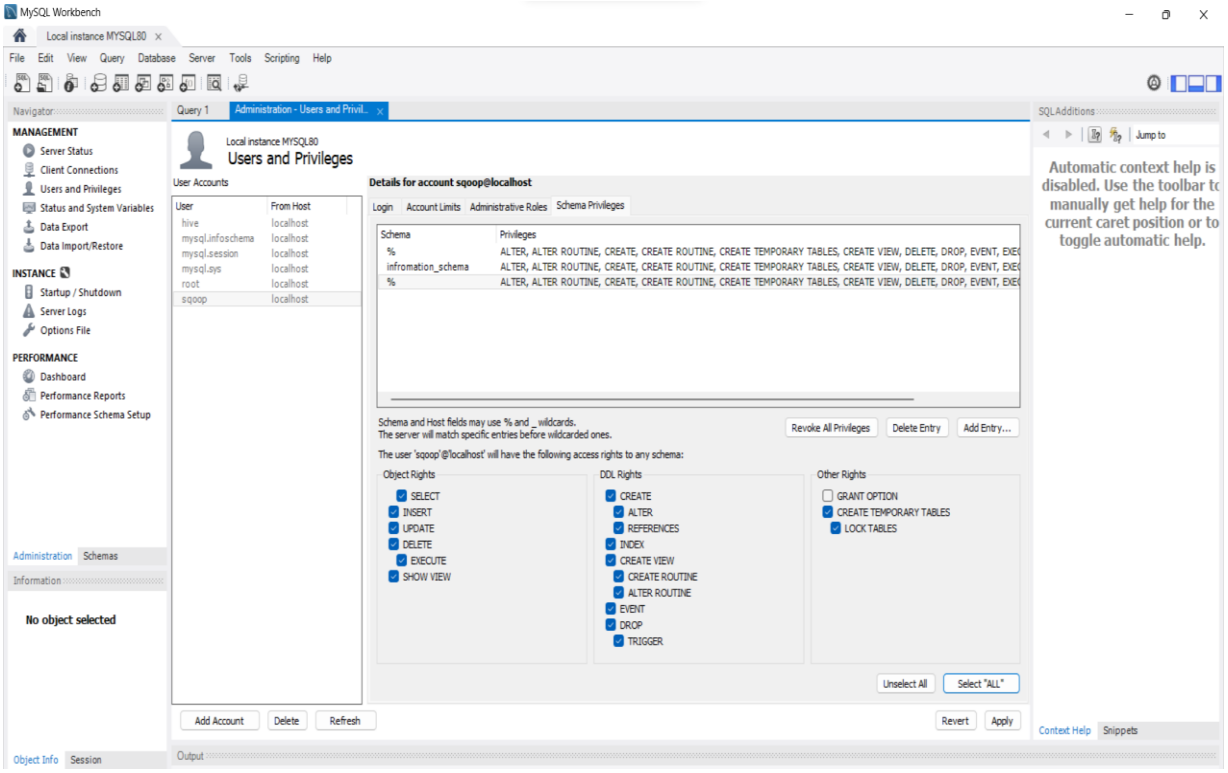
This Sqoop command lists have all the available database in the RDBMS server.

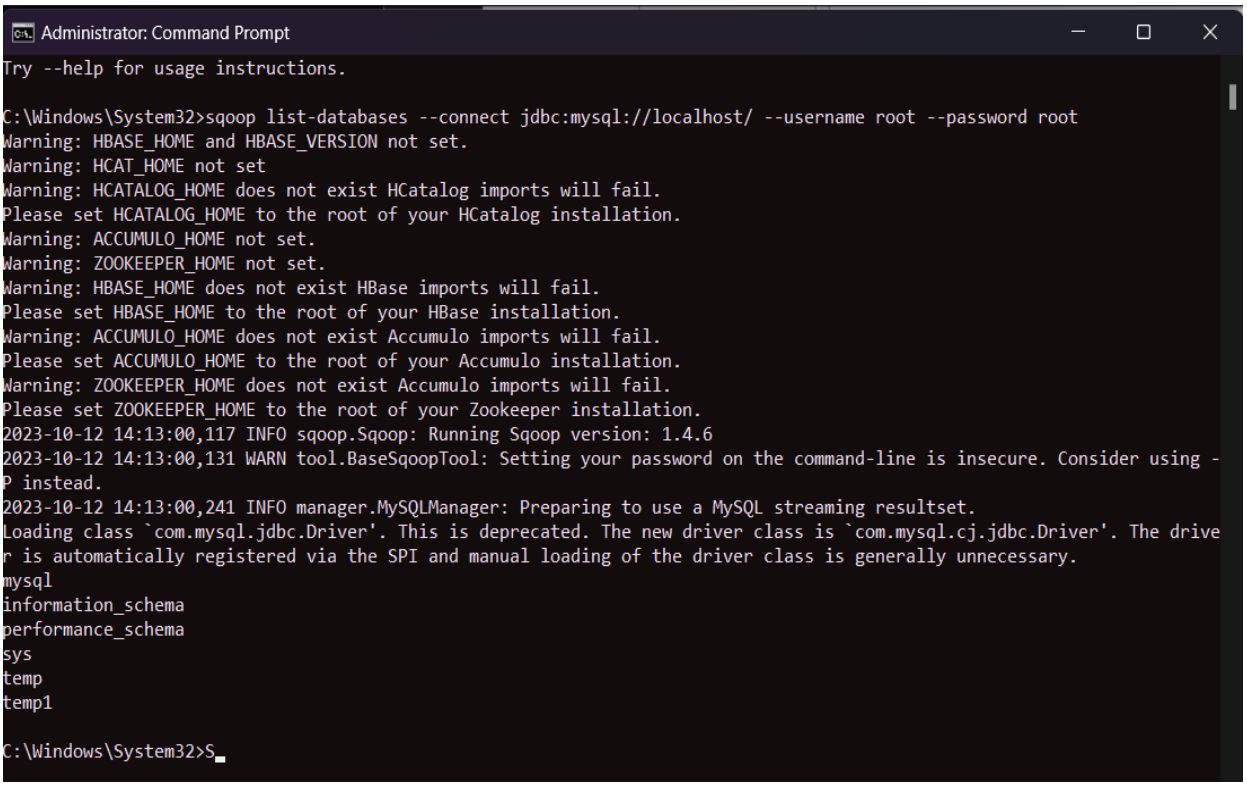
```
>$ sqoop list - database -- connect
```

## Output:




MySQL Workbench
Local instance MYSQL80
File Edit View Query Database Server Tools Scripting Help
Navigator
MANAGEMENT
Server Status
Client Connections
Users and Privileges
Status and System Variables
Data Export
Data Import/Restore
INSTANCE
Startup / Shutdown
Server Logs
Options File
PERFORMANCE
Dashboard
Performance Reports
Performance Schema Setup
Administration Schemas
Information
No object selected
Object Info Session
Query 1
Administration - Users and Privil...
Local instance MYSQL80
Users and Privileges
User Accounts
User From Host
hive localhost
mysql.infoschema localhost
mysql.session localhost
mysql.sys localhost
root localhost
sqoop localhost
Details for account sqoop@localhost
Login Account Limits Administrative Roles Schema Privileges
Role Description
DBA grants the rights to perform all tasks
MaintenanceAdmin grants rights needed to maintain server
ProcessAdmin rights needed to assess, monitor, and kill any user proce...
UserAdmin grants rights to create users logins and reset passwords
SecurityAdmin rights to manage logins and grant and revoke server an...
MonitorAdmin minimum set of rights needed to monitor server
DBManager grants full rights on all databases
DBDesigner rights to create and reverse engineer any database sche...
ReplicationAdmin rights needed to setup and manage replication
BackupAdmin minimal rights needed to backup any database
Global Privileges
ALTER
ALTER ROUTINE
CREATE
CREATE ROUTINE
CREATE TABLESPACE
CREATE TEMPORARY TABLES
CREATE USER
CREATE VIEW
DELETE
DROP
EVENT
EXECUTE
FILE
GRANT OPTION
INDEX
INSERT
LOCK TABLES
PROCESS
REFERENCES
RELOAD
REPLICATION CLIENT
REPLICATION SLAVE
SELECT
SHOW DATABASES
SHOW VIEW
SHUTDOWN
Revoke All Privileges
Add Account Delete Refresh
Revert Apply
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Context Help Snippets
Output


MySQL Workbench
Local instance MYSQL80
File Edit View Query Database Server Tools Scripting Help
Navigator
MANAGEMENT
Server Status
Client Connections
Users and Privileges
Status and System Variables
Data Export
Data Import/Restore
INSTANCE
Startup / Shutdown
Server Logs
Options File
PERFORMANCE
Dashboard
Performance Reports
Performance Schema Setup
Administration Schemas
Information
No object selected
Object Info Session
Query 1
Administration - Users and Privil...
Local instance MYSQL80
Users and Privileges
User Accounts
User From Host
hive localhost
mysql.infoschema localhost
mysql.session localhost
mysql.sys localhost
root localhost
sqoop localhost
Details for account sqoop@localhost
Login Account Limits Administrative Roles Schema Privileges
Schema Privileges
% ALTER, ALTER ROUTINE, CREATE, CREATE ROUTINE, CREATE TEMPORARY TABLES, CREATE VIEW, DELETE, DROP, EVENT, EXE
infromation_schema ALTER, ALTER ROUTINE, CREATE, CREATE ROUTINE, CREATE TEMPORARY TABLES, CREATE VIEW, DELETE, DROP, EVENT, EXE
% ALTER, ALTER ROUTINE, CREATE, CREATE ROUTINE, CREATE TEMPORARY TABLES, CREATE VIEW, DELETE, DROP, EVENT, EXE
Schema and Host fields may use % and _ wildcards.
The server will match specific entries before wildcarded ones.
Revoke All Privileges Delete Entry Add Entry...
The user 'sqoop'@'localhost' will have the following access rights to any schema:
Object Rights
SELECT
INSERT
UPDATE
DELETE
EXECUTE
SHOW VIEW
DDL Rights
CREATE
ALTER
REFERENCES
INDEX
CREATE VIEW
CREATE ROUTINE
ALTER ROUTINE
EVENT
DROP
TRIGGER
Other Rights
GRANT OPTION
CREATE TEMPORARY TABLES
LOCK TABLES
Unselect All Select "ALL"
Add Account Delete Refresh
Revert Apply
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Context Help Snippets
Output




## Conclusion:

Sqoop is a tool used to move large amounts of data between Hadoop and external datastores like relational databases (MS SQL Server, MySQL), which are external datastores. Basically operations that take place in Sqoop are usually user-friendly. Sqoop used the command-line interface to the process command of user. It helps us to perform ETL operations in a very fast and cost-effective manner. In this experiment, we successfully implemented the Installation and Configuration of Sqoop Tool Hadoop Ecosystem.