SQOOP COMMANDS AND USAGE

1. Install MariaDB database in your system and enable the service through the commands:

   # yum groupinstall -y mariadb
   (Set the password when asked)
   # systemctl start mariadb
   #systemctl enable mariadb
   #mysql_secure-installation
   #firewall-cmd --permanent --add-service=mysql
   #firewall-cmd --reload

2. Login to the database :

```
$mysql -u root -p $rootpassword

mysql> create database db;
mysql> use db;
mysql> create table employee(emp_id INT NOT NULL PRIMARY KEY
-auto_increament,
first_name VARCHAR(50) NOT NULL,
last_name VARCHAR(50), designation VARCHAR(50) NOT NULL,
salary INT(10))
mysql> Insert into employee (emp_id, first_name, last_name,
designation, salary)
values
(1001, “Mike” , “Ross”, “Associate”,70000) ,
(1002, “Hariett” ,“Specter”,”Secretary”,80000),
(1003, “Michelle” , “Ross” , “Paralegal”,50000) ,
(1004, “Jessica”, “Pearson”, “Managing Partner”, 15000000),
(1005, “Robert” , “Zane”, “Managing Partner”, 15000000);
mysql> show tables;
mysql> exit
```

3. The following command is used to import the **employee** table from MySQL database server to HDFS.

```
$ sqoop import \
--connect jdbc:mysql://localhost/db1 \
--username root \ --password redhat
--table employee \
--import-dir /Employee --m 1
```

4. To verify the imported data in HDFS, use the following command.

```
$ hdfs dfs -ls /Employee
```

```
$ hdfs dfs -cat /Employee/part-m-*
```

INCREMENTAL IMPORT

Incremental import is a technique that imports only the newly added rows in a table. The following syntax is used for the incremental option in Sqoop import command.

```
--incremental <mode>
--check-column <column name>
--last value <last check column value>
```

Let us assume the newly added data into employee table is as follows –

```
1006, “Katrina”, “Bennett”, “Junior Associate”, 60000
```

To perform incremental import :

```
$ sqoop import \
--connect jdbc:mysql://localhost/db1 \
--username root –password redhat \
--table employee \
--m 1 \
--incremental append \
--check-column emp_id \
--last-value 1005
```

The following command is used to verify the imported data from employee table to HDFS /Employee directory.

```
$ hdfs dfs -ls /Employee
$ hdfs dfs -cat /Employee/part-m-*

1001, Mike, Ross, Associate, 70000
1002, Hariett, Specter, Secretary, 80000
1003, Michelle, Ross, Paralegal, 50000
1004, Jessica, Pearson, Managing Partner,15000000
1005, Robert, Zane, Managing Partner,15000000
1006, Katrina, Bennett, Junior Associate, 60000
```

To check only the newly incremented entry we can use the command :

```
$ hdfs dfs -cat /Employee/part-m-*1

1006, Katrina, Bennett, Junior Associate, 60000
```

SQOOP EXPORT –

This command exports data back from the HDFS to the RDBMS. The target table must exist in the target database.

```
$sqoop export (generic-args) (export-args)
```

Let us take an example of the employee data in file, in HDFS. The employee data is available in '/Employee' directory in HDFS.

The following query is used to create the table 'employee1' in mysql command line.

```
$ mysql
mysql> create db1;
mysql> USE db1;
mysql> CREATE TABLE employee1 ( emp_id INT NOT NULL PRIMARY KEY -
auto increment,
first_name VARCHAR(50) NOT NULL,
last_name VARCHAR(50),
designation VARCHAR(50) NOT NULL,
salary INT(10));
```

The following command is used to export the table data to the employee1 table in db1 database of Mysql database server.

```
$ sqoop export \
--connect jdbc:mysql://localhost/db1 \
--username root –password redhat \
--table employee1 \
--export-dir /Employee
```

To verify the table in mysql command line :

```
mysql>select * from employee1;
```

The output must show the table imported from the HDFS.