**Sqoop**

**Aim:** To understand the concept of data ingestion tool “Sqoop”

**Objective:**

* To import structured data from MYSQL to HDFS
* To export text file (structured data) to MYSQL

**Key concept:**

* Apache Sqoop is a tool in [*Hadoop ecosystem*](https://www.edureka.co/blog/hadoop-ecosystem) which is designed to transfer data between [*HDFS*](https://www.edureka.co/blog/hdfs-tutorial) (Hadoop storage) and relational database servers like mysql, Oracle RDB, SQLite, Teradata, Netezza, Postgres etc.
* Apache Sqoop imports data from relational databases to HDFS, and exports data from HDFS to relational databases. It efficiently transfers bulk data between Hadoop and external datastores such as enterprise data warehouses, relational databases, etc.
* *This is how Sqoop got its name – “SQL to Hadoop & Hadoop to SQL”.*

**Q1: How to enter in mysql CLI in cloudera**

[cloudera@quickstart Desktop]$ mysql -uroot -pcloudera

**Q2: Create a database**

mysql> create database sqoop\_db;

Query OK, 1 row affected (0.27 sec)

**Q3: Select the database created**

mysql> use sqoop\_db;

**Q4: Create a table inside created database**

mysql> create table emp(empno int primary key, ename varchar(10),job varchar(9),mgr int,hiredate date,sal int,deptno int);

Query OK, 0 rows affected (0.08 sec)

**Q5: Insert records in the table created**

mysql> INSERT INTO emp VALUES (1,'ABC','teacher',100,2013-1-12, 50000,10);

mysql> INSERT INTO emp VALUES (2,'XYZ','teacher',200,2014-10-2,650000,20);

mysql> INSERT INTO emp VALUES (3,'PQRS','teacher',200,2010-10-2,640000,30);

**Q6: Check the contents in the table?**

mysql> select \* from emp;

+-------+--------+---------+------+------------+--------+--------+

| empno | ename | job | mgr | hiredate | sal | deptno |

+-------+--------+---------+------+------------+--------+--------+

| 1 | ABC| teacher | 100 | 2013-1-12| 50000 | 10 |

| 2 | XYZ | teacher | 200 | 2014-10-2| 650000 | 20 |

| 3 |PQRS | teacher | 200 | 2010-10-2 | 640000 | 30 |

+-------+--------+---------+------+------------+--------+--------+

3 rows in set (0.00 sec)

### **Sqoop – Import**

**Q7: Import this table in HDFS**

[cloudera@quickstart Desktop]$ sqoop import --connect jdbc:mysql://localhost/sqoop\_db --username root --password cloudera --table emp --target-dir /user/cloudera/scoop\_data

**Q8: Check if the data is imported in HDFS**

[cloudera@quickstart Desktop]$ hadoop fs -ls /user/cloudera/scoop\_data

Found 4 items

-rw-r--r-- 1 cloudera cloudera 0 2018-10-02 10:26 /user/cloudera/scoop\_data/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 34 2018-10-02 10:26 /user/cloudera/scoop\_data/part-m-00000

-rw-r--r-- 1 cloudera cloudera 36 2018-10-02 10:26 /user/cloudera/scoop\_data/part-m-00001

-rw-r--r-- 1 cloudera cloudera 35 2018-10-02 10:26 /user/cloudera/scoop\_data/part-m-00002

**Q9: Open the partitions and check for the records**

[cloudera@quickstart Desktop]$ hadoop fs -cat /user/cloudera/scoop\_data/part-m-00000

1, Veena ,teacher,100, 2013-1-12,50000,10

[cloudera@quickstart Desktop]$ hadoop fs -cat /user/cloudera/scoop\_data/part-m-00001

2,sheena,teacher,200, 2014-10-2,650000,20

[cloudera@quickstart Desktop]$ hadoop fs -cat /user/cloudera/scoop\_data/part-m-00002

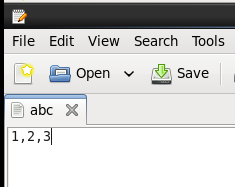
3,meena,teacher,200, 2010-10-2,640000,30

### **Sqoop – Export**

**Q10: Export data(like csv) in MYSQL in a table?**

**Create a file in local in the Desktop, suppose abc.txt**

gedit abc.txt



**Next put this abc.txt in HDFS**

[cloudera@quickstart Desktop]$ hadoop fs -put abc.txt /user/cloudera/

**Come to MYSQL prompt**

Create a table suppose abc\_table

**NOTE (you should first write use database\_name command: that means you need to tell in which database you wish to create a table, in our case i’am using database sqoop\_db)**

mysql> use sqoop\_db;

sql> create table abc\_table( a int, b int, c int);

Query OK, 0 rows affected (0.13 sec)

**In terminal type the following command to export abc.txt in abc\_table table**

[cloudera@quickstart Desktop]$ sqoop export --connect jdbc:mysql://localhost/sqoop\_db --username root --password cloudera --table abc\_table --export-dir /user/cloudera/abc.txt

**Q11: Check in MYSQL if data is exported from HDFS into abc\_table**

mysql> select \* from abc\_table;

+------+------+------+

| a | b | c |

+------+------+------+

| 1 | 2 | 3 |

+------+------+------+

1 row in set (0.01 sec)