**BDP ICP 5 – SQOOP**

1. **Importing and Exporting Data to/from Mysql from/to Hadoop:**
2. Starting the MySQL service: sudo service mysqld start
3. Enter MySQL shell by: mysql -u root - pcloudera
4. Create a new database: create database icp5;
5. Now to create a table in icp5 database we need to use the same database

use icp5;

1. Now create a table called "**employees**" with three columns "**empId**", “**emp\_name**" and “**emp\_sal**” with **empId** as primary key.

create table employees(empId INT NOT NULL AUTO\_INCREMENT, emp\_name varchar(100), emp\_sal INT PRINARY KEY (empId));

1. Insert values into table

Insert into employees(1, “Ravi”, 10000), (2, “Ajay”, 20000)

1. Now using the sqoop import command import the data from mysql to Hadoop and we can specify the target directory as an option also.

sqoop import --connect jdbc:mysql://localhost/icp5 --username root --password cloudera --table employees --m 1;

sqoop import --connect jdbc:mysql://localhost/icp5 --username root --password cloudera --table employees –m 1 --target-dir queryResult;

we can test the output using the Hadoop ls command. **hdfs dfs -ls** will list the folder employees imported from MySQL.

1. To export the data into MySql from HDFS we can use the sqoop export command. We are using the same data and moving into new table persons in mysql.

create table persons(id int NOT NULL AUTO\_INCREMENT, name varchar(100), salary INT, primary key (id));

sqoop export --connect jdbc:mysql://localhost/icp5 --username root --password cloudera --table persons --export-dir queryResult/part-m-00000

**2. Import Hive tables to/from MySQL:**

1. create a table “emp” in hive and load data into the table.

create table emp(empId INT, emp\_name STRING) row format delimited fields terminated by ‘,’ lines terminated by ‘\n’ stored as textfile;

load data inpath ‘employees’ into table emp;

2. Now Export the data from hive table into Mysql using sqoop export command.

In order to export the data into sql, we need to create the table manually in mysql.

create table emp\_hive (id INT NOT NULL AUTO\_INCREMENT, name VARCHAR(100), primary key(id));

sqoop export –connect jdbc:mysql://localhost/icp5 --username root --password cloudera –table emp\_hive --export-dir /user/hive/warehouse/emp/ -m 1

3. We can import the date from MySql to Hive using the sqoop import command.

create table companies (id INT NOT NULL AUTO\_INCREMENT, name VARCHAR(100), place VARCHAR(100), primary key(id));

insert into companies (1, “Google”, “USA”), (1, “Facebook”, “USA”), (1, “Amazon”, “USA”)

sqoop import --connect jdbc:mysql://localhost/icp5 --username root --password cloudera --table comapnies --target-dir /companies --hive-import --create-hive-table --hive-table default.CompaniesHive -m 1

**3. Import and export using Sqoop:**

1. Created a Hive table using the dataset “Dividends.csv”.

create table dividends(date DATETIME, dividend float) row format delimited fields terminated by ‘,’ lines terminated by ‘\n’ stored as textfile;

load data local inpath ‘/home/cloudera/Downloads/dividends.csv’ into table dividends;

2. Exporting the data from Hive to MySql.

create table dividends (date DATETIME, dividend float);

sqoop export –connect jdbc:mysql://localhost/icp5 --username root --password cloudera –table dividends --export-dir /user/hive/warehouse/dividends -m 1

4.Queries to Analyze the data in Dividends table

**Statistics of Table:** analyze table dividends compute statistics;

**Count of Months:**

select date\_format(date, “MMM”), count(\*) from dividends group by date\_format(date, “MMM”);

**Count of Months with dividend less than 0.01:**

select date\_format(date, “MMM”), count(\*) from dividends where dividend <=0.01 group by date\_format(date, “MMM”);