**Big Data on Hadoop**

In this project, you will be creating a MySQL table named ‘customers’ which has the customer details of a leading e-commerce company. You will now push this data into Hive using SQOOP. Once you have imported the data you can answer the questions below. The steps to create a table in MySQL along with the import commands are given below

**Step 1: Login into SQL in your VM**

mysql -u root -p password pwd

**Step 2: Create a Database**

create database antwak;

use antwak;

**Step 3: Create Table in Mysql**

CREATE TABLE customers (customer\_id INT NOT NULL, customer\_fname VARCHAR(255) NOT NULL, customer\_lname VARCHAR(255) NOT NULL, customer\_street VARCHAR(255) NOT NULL,customer\_city VARCHAR(255) NOT NULL,customer\_state VARCHAR(255) NOT NULL,PRIMARY KEY (customer\_id));

**Step 4: Load data into the table in Mysql**

LOAD DATA LOCAL INFILE '/home/cd/Desktop/customers.csv' INTO TABLE customers FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' (customer\_id, customer\_fname, customer\_lname, customer\_street, customer\_city, customer\_state);

**Step 5: Now load the Mysql table to Hive with the help of Sqoop and answer the following questions**

**Command To Load Data in HIVE from MYSQL using SQOOP:-**

sqoop import --connect jdbc:mysql://mysql\_server/antwak --username antwak --password antwak --table customers --hive-import --hive-table antwak.customer

1. How many records are present?

Select Count(\*) From customers;

Ans. 12435

2. How many customers stay in the state TX?

Select Count(\*) From customers where customer\_state = ‘TX’;

Ans. 635

3.Which state has the lowest customer count?

Select customer\_state , count(\*) as customer\_count from customers group by customer\_state order by customer\_count ASC limit 1;

Ans. State = AL , CustomerCount = 3

4. Which state has the highest customer count?

Select customer\_state , count(\*) as customer\_count from customers group by customer\_state order by customer\_count DESC limit 1;

Ans. State = PR , CustomerCount = 4771

5.How many customers are present with the First Name 'Mary’?

Select count(\*) as customer\_count from customers where customer\_fname = ‘Mary’

Ans. 4741

**Extra Information:-**

In this project first I created a table called customers in Mysql and inserted data into that table from a .csv file using load command.

After that I used Apache Sqoop to import the customers table from Mysql to Apache Hive using below command.

**Command To Load Data in HIVE from MYSQL using SQOOP:-**

sqoop import --connect jdbc:mysql:// mysql\_server /antwak --username antwak --password antwak --table customers --hive-import --hive-table antwak.customer

Note: For big data analysis Apache Hive is more suitable so I am using Apache Hive instead of querying the data in Mysql.

**Important Commands:-**

**Connecting through beeline**

beeline -u jdbc:hive2://

**To List Databases**

sqoop-list-databases \

--connect "jdbc:mysql://mysql\_server:3306" \

--username root \

--password pwd

**To list the tables in a DB**

sqoop-list-tables \

--connect "jdbc:mysql:// mysql\_server:3306/retail\_db" \

--username root \

--password pwd

**To Evaluate a Table**

sqoop-eval \

--connect "jdbc:mysql:// mysql\_server:3306" \

--username root \

--password pwd \

--query "describe retail\_db.orders"

**To Import a table to a Target directory**

sqoop import \

--connect "jdbc:mysql:// mysql\_server:3306/retail\_db" \

--username root \

--password pwd \

--table orders \

--target-dir /queryresult\_new