SQOOP ASSIGNMENT S.SADHANA-221047005

1.Grant privileges to dbs in mysql for users to connect from SQOOP
mysql>
GRANT ALL PRIVILEGES ON *.* TO 'cdc22'@'localhost' IDENTIFIED BY 'msis';

Implement the following in SQOOP

2.List the databases in mysql through sqoop
sqoop list-databases --connect jdbc:mysql://localhost?useSSL=false --username cdc22 --password msis

3.List all tables in selected databases
sqoop list-tables --connect jdbc:mysql://localhost/retail_db?useSSL=false --username cdc22 --password msis

4.Import table into HDFS using target-dir method
//to transfer one table
sqoop import --connect jdbc:mysql://localhost/retail_db?useSSL=false --username cdc22 --password msis --table customers --target-dir '/testCustomer1'

5.Import table into HDFS using warehouse-dir method
//to transfer multiple table
sqoop import --connect jdbc:mysql://localhost/retail_db?useSSL=false --username cdc22 --password msis --table orders --warehouse-dir '/retail_db'

output:
drwxr-xr-x sois supergroup 0 B Feb 28 16:06 0 0B orders

6.To import partial data from specified table
sqoop import --connect jdbc:mysql://localhost/retail_db?useSSL=false --username cdc22 --password msis --table categories --target-dir '/tempCat' --where "category_id < 30"

output:
drwxr-xr-x sois supergroup 0 B Feb 28 16:16 0 0B tempCat

7.To import data from selected columns table

sqoop import --connect jdbc:mysql://localhost/retail_db?useSSL=false --username cdc22 --password msis --columns product_id,product_name,product_price --table products --target-dir '/selected_colTransfer'

output:
drwxr-xr-x sois supergroup 0 B Feb 28 16:31 0 0B selected_colTransfer

8.To import all tables from given database
sqoop import-all-table --username cdc22 --password msis --warehouse-dir '/path'

9.To export some tables while importing all the tables

--warehouse-dir '/path'--exclude-tables table1,table2

10.To export table from HDFS to local db

Implement following using HIVE tool

1.Find unique states in customer table

```
SELECT DISTINCT state FROM customer;
```

2.Find number of customers from each state

```
SELECT state, COUNT(*) as num_customers
FROM customer
GROUP BY state;
```

3.List and count number of unique fnames from customer table

```
SELECT fname, COUNT(DISTINCT fname) as num_unique_fnames
FROM customer
GROUP BY fname;
```

4.List and count unique cities from customer table

```
SELECT COUNT(DISTINCT city) as num_unique_cities
FROM customer;
```

5.Find customer id who placed maximum orders

```
SELECT customer_id, fname, lname, num_orders
FROM (
  SELECT o.customer_id, c.fname, c.lname, COUNT(*) as num_orders,
       RANK() OVER (ORDER BY COUNT(*) DESC) as rank_orders
  FROM orders o
  JOIN customer c ON o.customer_id = c.customer_id
  GROUP BY o.customer_id, c.fname, c.lname
) ranked_orders
WHERE rank_orders = 1;
```