1. How to create database in hive

a. create database database_name;

b. checked the folder created in /hive/warehouse/database_name(HDFS)

2. Select Database

Use database_name;

3. Create table for storing transactional records

Create table txnrecords (txnno INT, txndate STRING, custno INT, amount DOUBLE, category STRING, product STRING, city STRING, state STRING, spendby STRING)

row format delimited

fields terminated by ‘,’

stored as textfile;

Create external table externaltxnrecords(txnno INT, txndate STRING, custno INT, amount DOUBLE, category STRING, product STRING, city STRING, state STRING, spendby STRING)

row format delimited

fields terminated by ‘,’

location '/hivetable';

4. Load the data into table

LOAD DATA LOCAL INPATH 'home/notroot/lab/data/txns' OVERWRITE INTO TABLE txnrecords;

5. Describing metadata or schema of the table

Describe txnrecords;

6. Counting no. of records

Select count(*) from txnrecords;

7. Counting total by category of products

select category,sum(amount) from txnrecords group by category;

8. Top 10 customers

select custno, sum(amount) from txnrecords group by custno limit 10;

9. Create partitioned table

create table txnrecsByCat(txnno INT, txndate STRING, custno INT, amount DOUBLE, product STRING, city STRING,state String, spendby STRING)

partitioned by (category STRING)

clustered by (state) into 10 buckets

row format delimited

fields terminated by ‘,’

stored as textfile;

10. Configure Hive to allow partitions

set hive.exec.dynamic.partition.mode=nomstrict;

To enable dynamic partitions, by default it is false

set hive.exec.dynamic.partition=true;

set hive.enforce.bucketing=true;

11. Load data into partition table

FROM txnrecords txn INSERT OVERWRITE TABLE txnrecsbycat PARTITION (category)

SELECT txn.txnno, txn.txndate, txn.custno, txn.amount, txn,product, txn.city, txn.state,txn.spendby,

txn.category DISTRIBUTED BY category;