show databases;

use default;

show tables;

**Create Managed Table:**

create table storesales(RowID int,OrderID string,OrderDate date,ShipDate date,ShipMode string,CustomerID string,CustomerName string,Segment string,Country string,City string,State string,PostalCode int,Region string,ProductID string,Category string,SubCategory string,ProductName string,Sales double) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS TEXTFILE;

describe storesales;

describe formatted storesales;

**Load data from local:**

load data local inpath '/home/ganesh/Analytics/dataset' into table storesales;

Select \* from storesales limit 15;

drop table storesales;

**Create External Table:**

create external table storesales(RowID int,OrderID string,OrderDate date,ShipDate date,ShipMode string,CustomerID string,CustomerName string,Segment string,Country string,City string,State string,PostalCode int,Region string,ProductID string,Category string,SubCategory string,ProductName string,Sales double) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS TEXTFILE location '/home/ganesh/Analytics/dataset';

Describe storesales;

Describe formatted storesales;

**Load data from local:**

load data local inpath '/home/skiganesh/dataset/' into table storesales;

hdfs dfs -mkdir /user/hdfs/superstore

hdfs dfs -copyFromLocal superstore.csv /user/hdfs/superstore

create external table superstoresales(RowID int,OrderID string,OrderDate date,ShipDate date,ShipMode string,CustomerID string,CustomerName string,Segment string,Country string,City string,State string,PostalCode int,Region string,ProductID string,Category string,SubCategory string,ProductName string,Sales double) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS TEXTFILE location '/user/hdfs/superstore/';

select \* From superstoresales limit 5;

**Load data from HDFS:**

create external table superstoresales\_inpath(RowID int,OrderID string,OrderDate string,ShipDate string,ShipMode string,CustomerID string,CustomerName string,Segment string,Country string,City string,State string,PostalCode int,Region string,ProductID string,Category string,SubCategory string,ProductName string,Sales double) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS TEXTFILE location '/user/hdfs/superstore\_inpath/';

load data inpath '/user/hdfs/superstore' into table superstoresales\_inpath;

**Dynamic Partition:**

create external table salepartition(RowID int,OrderID string,OrderDate string,ShipDate string,ShipMode string,CustomerID string,CustomerName string,Segment string,Country string,PostalCode int,Region string,ProductID string,Category string,SubCategory string,ProductName string,Sales double,City string) partitioned by (State string) stored as orc;

set hive.exec.dynamic.partition=true;

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

Set hive.exec.max.dynamic.partitions=105;

Insert into table salepartition partition(State) Select RowID,OrderID,OrderDate,ShipDate,ShipMode,CustomerID,CustomerName,Segment,Country,PostalCode,Region,ProductID,Category,SubCategory,ProductName,Sales,City,State from superstoresales\_inpath;

**Static Partition:**

**Pyspark commands:**

**spark.sparkContext.setLogLevel("ERROR");**

**spark.sql("use default");**

**californiadf=spark.sql("select \* from salepartition where state='California'");**

**coloradodf=spark.sql("select \* from salepartition where state='Colorado'");**

**californiadf.coalesce(1).write.csv('file:////home/hdfs/storesale/california', header=True);**

**coloradodf.coalesce(1).write.csv('file:////home/hdfs/storesale/colorado', header=True);**

**Hive Queries:**

create table salepartitionstatic(RowID int,OrderID string,OrderDate string,ShipDate string,ShipMode string,CustomerID string,CustomerName string,Segment string,Country string,PostalCode int,Region string,ProductID string,Category string,SubCategory string,ProductName string,Sales double,City string) partitioned by (State string) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS TEXTFILE;

Load data local inpath '/home/hdfs/storesale/california/' into table salepartitionstatic partition(state='california');

Load data local inpath '/home/hdfs/storesale/colorado/' into table salepartitionstatic partition(state=’colorado’);

**Bucketed Table:**

**Create Bucketed Table:**

create external table storesales\_bucket(RowID int,OrderID string,OrderDate date,ShipDate date,ShipMode string,CustomerID string,CustomerName string,Segment string,Country string,City string,State string,PostalCode int,Region string,ProductID string,Category string,SubCategory string,ProductName string,Sales double) CLUSTERED BY (State) SORTED BY (City) INTO 50 BUCKETS ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS TEXTFILE location '/user/hdfs/superstore';

load data local inpath '/home/ganesh/Analytics/dataset' into table storesales\_bucket;