**BDP ICP-4 HIVE**

**1. Petrol Dataset:**

**1) Creating Tables:**

create table petroldata (distributer\_id STRING, distributer\_name STRING, amt\_IN STRING, amt\_OUT STRING, vol\_IN INT, vol\_OUT INT,year INT) row format delimited fields terminated by ',' stored as textfile;

**2) Load Data:**

load data local inpath '/home/cloudera/Desktop/BDP/ICP4/petrol.txt' into table petroldata;

**3) List all distributors who have the difference between Vol-in & vol-out > 500, along with the year and the difference which they have in that year.**

**Group BY**: select distributer\_id,year,(vol\_IN-vol\_OUT) as difference from petroldata where (vol\_IN-vol\_OUT)>500 group by distributer\_id;

**Distribute By:** select distributer\_id,year,(vol\_IN-vol\_OUT) as difference from petroldata where (vol\_IN-vol\_OUT)>300 distribute by distributer\_id limit 20;

**Cluster By:** select distributer\_id,year,(vol\_IN-vol\_OUT) as difference from petroldata where (vol\_IN-vol\_OUT)>300 cluster by distributer\_id limit 20;

**Sort By:** select distributer\_id,year,(vol\_IN-vol\_OUT) as difference from petroldata where (vol\_IN-vol\_OUT)>300 sort by distributer\_id limit 20;

**2. Olympic Dataset:**

**1) Creating Tables:**

create table olympicsdata (athelete STRING,age INT,country STRING,year STRING,closing STRING,sport STRING, gold INT, silver INT, bronze INT,total INT) row format delimited fields terminated by '\t' stored as textfile;

**2) Load Data:**

load data local inpath '/home/cloudera/Desktop/BDP/ICP4/olympic\_data.csv' into table olympicsdata;

**3) Country got medals for shooting - Year wise:**

select country, year from olympicsdata where sport='Shooting' group by country, year order by year limit 10;

**3. Movie Dataset:**

**1) Creating Tables:**

**Movie Table:** create table movies(id int, title string,genre string) row format delimited fields terminated by ',' stored as textfile;

**Ratings Table:** create table ratings (userId int, movieId int, rating int, timestamp string) row format delimited fields terminated by ',' stored as textfile;

**Users Table:** create table users(userid int,gender string,id int,ratingsgiven int,zip string) > row format delimited > fields terminated by ',' > stored as textfile;

**2) Loading Data to Tables:**

**Movie Data**: load data local inpath '/home/cloudera/Desktop/BDP/ICP4/Movies/movies.csv' into table movies;

**Rating Data:** load data local inpath '/home/cloudera/Desktop/BDP/ICP4/Movies/ratings.csv' into table ratings;

**User Data:** load data local inpath '/home/cloudera/Desktop/BDP/ICP4/Movies/users.txt' into table users;

**3) Selecting Movies with genre Action & Drama:**

select \* from movies where genre like '%Action%' and genre like '%Drama%';

**4) Selecting Ratings with a value equal to 5:**

select \* from ratings where rating=5;

**5) Top 11 average rated "Action" movies with descending order of rating:**

select m.id, m.title, avg(r.rating) as avg\_rating from movies m join ratings r on (m.id=r.movieId) where m.genre like '%Action%' group by m.id, m.title order by avg\_rating desc limit 11;