**HIVE**

**Tables:**

**CREATE TABLE user27dw.bike\_teams (**

event\_type STRING,

event\_id INT,

team\_id INT,

team\_name STRING,

team\_captain\_contact\_id INT,

team\_division STRING,

company STRING ,

number\_of\_participants INT ,

total\_fees\_paid INT,

team\_total\_confirmed INT ,

total\_online\_gifts INT,

total\_offline\_confirmed\_gifts INT,

total\_offline\_unconfirmed\_gifts INT,

team\_goal INT,

confirmed\_gifts\_history INT, previous\_event\_confirmed\_gifts INT, previous\_event\_team\_members INT,

fiscal\_year INT,

ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' TBLPROPERTIES ( 'skip.header.line.count'='1');

**Query to Load the data into the table**

**LOAD DATA LOCAL INPATH '/home/user27/dwproject/DW1-Bike\_teams.csv' OVERWRITE INTO TABLE user27dw.Bike\_teams;**

**Create table user27dw.Participants**(

Fiscal\_Year INT,

Internal\_Event\_Name STRING, Participation\_Type\_Name STRING,

Team\_Name STRING,

Team\_Division STRING,

Team\_ID INT,

Member\_ID INT,

Registration\_Status STRING,

Is\_Team\_Captain STRING,

Is\_Prior\_Participant STRING, Total\_of\_All\_Confirmed\_Gifts INT, Total\_From\_Participant INT, Total\_Not\_From\_Participant INT,

Number\_from\_Participant INT, Number\_Not\_From\_Participant INT,

Employer STRING,

Occupation STRING,

Participant\_Connection\_to\_MS STRING,

State STRING,

County STRING,

City STRING,

Zip\_Code STRING,

Registration\_Type STRING,

Event\_ID INT,

Participant\_Gender STRING,

Participant\_Goal INT)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' TBLPROPERTIES ( 'skip.header.line.count'='1');

**Load data local inpath '/home/user27/dwproject/DW1-Participants1.csv' OVERWRITE INTO TABLE user27dw.Participants;**

**Create table user27dw.Events**(

Fiscal\_Year INT,

Event\_ID INT,

Event\_Goal INT,

Active\_Registrations INT,

Inactive\_Registrations INT,

Total\_Fees\_Paid INT,

Total\_of\_All\_Confirmed\_Gifts INT,

Total\_Online\_Gifts INT,

Teams INT,

Captains INT,

Average\_Team\_Size INT, Total\_Offline\_Confirmed\_Gifts INT,

Self\_Donors INT,

Total\_From\_Participant INT,

Non\_self\_Donors INT,

Total\_Not\_From\_Participant INT,

Total\_Team\_Gifts INT,

Total\_Event\_Gifts INT, Total\_Offline\_Unconfirmed\_Gifts INT,

Address STRING,

City STRING,

State STRING,

Zip\_Code STRING,

Previous\_Event\_ID INT)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' TBLPROPERTIES ( 'skip.header.line.count'='1');

**Load data local inpath '/home/user27/dwproject/DW1-Events.csv' OVERWRITE INTO TABLE user27dw.Events;**

**CREATE TABLE USER27DW.Donation**(

Event\_ID INT,

Public\_Event\_Name STRING,

Fiscal\_Year INT,

Campaign\_Title STRING,

Campaign\_ID INT,

Gift\_Amount INT,

Gift\_Type STRING,

Gift\_Payment\_Method STRING,

Offline\_Status STRING,

Is\_Registration STRING,

Donor\_ConsID INT,

Donor\_Member\_ID INT,

Donor\_Gender STRING,

Donor\_Email\_Status STRING,

Donor\_City STRING,

Donor\_State STRING,

Donor\_County STRING,

Donor\_ZIP STRING,

Donor\_Employer STRING,

Donor\_Connection\_to\_MS STRING,

Participant\_Contact\_ID INT,

Participant\_Member\_ID INT,

Registration\_Active\_Status STRING,

Participant\_Goal INT,I

s\_Prior\_Participant STRING,

Is\_Team\_Captain STRING,

Additional\_Gift\_Amount INT,

Team\_Name STRING,

Team STRING,

Original\_Value\_Transacted INT,

Net\_Transaction\_Amount INT,

Ledger\_Transaction\_Amount INT)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' TBLPROPERTIES ( 'skip.header.line.count'='1');

**LOAD DATA LOCAL INPATH '/home/user27/dwproject/DW1-Donations1.csv' OVERWRITE INTO TABLE user27dw.Donation;**

**We run the same queries as in PostgreSQL**

**Q1.**select sum(gift\_amount), is\_prior\_participant, fiscal\_year from Donation Group by is\_prior\_participant, Fiscal\_year, gift\_amount;

**Q2.** select max(gift\_amount), donor\_state, donor\_city from donation group by donor\_state, donor\_city ;

**Q3**. select avg(gift\_amount), fiscal\_year, event\_id from donation group by fiscal\_year,event\_id

**Q4.**select max(sum), max(average),donor\_gender,fiscal\_year from public.gender group by fiscal\_year,donor\_gender limit 1000;

**Q5.**Select select a.team\_name,a.number\_of\_participants,b.gift\_amount from bike\_teams a left join donation b on a.event\_id=b.event\_id limit 100;

**3 new queries in Hive**

**H1.What is the total amount of givings done by each team and their respective names in each state ?**

select donations,donorid,year,sum(givings)

over(partition by donations.donorid order by year)

from donations,donors

where donations.donorid=donors.donorid and

donors.gender=’female’

group by donations,donorid,year,givings;

**H2.Retrieve female donors and total amount of givings per year.**

select donations,donorid,year,sum(givings)

over(partition by donations.donorid order by year)

from donations,donors

where donations.donorid=donors.donorid and

donors.gender=’female’

group by donations,donorid,year,givings;

**H3.Retrieve the employer who have donated through credit card.**

select disctinct employer

from donors

where payment method = ’credit card’ ;