Loading dataset in to hive table can be done in two ways

1. From local file system
2. From Hadoop file system

Loading dataset from Hadoop file system command

hive> load data inpath < file path> into table vehilcle\_collision

Showing tables:

Show tables;

Showing table schema:

describe tablename;

Queries:

1. Query to find number of pedestrians killed?

Select pedestrian\_killed,

Count(\*) as num\_pedestrian\_killed\_count

From vehicle\_collision

group by pedestrian\_killed;

1. Finding out how many pedestrians have been killed in an accident occurred in a particular zip code using the attribute zip\_code?

Select count(num\_pedestrian\_killed), zip\_code

from vehilcle\_collisions,

group by zip\_code,

having count(num\_pedestrians\_killed) >1;

1. Finding out the location in which number of persons injured is maximum and grouping it by location?

Select location, sum(num\_persons\_injured) from

vehicle\_collision group by location;

1. Creating table vehicle\_collision-1 and displaying zip code and location and using substr to find the particular reason for the accident like passenger distraction using attribute contributing factor?

Create table vehicle\_collision\_1 as

Select zip\_code(contributing\_factor, 1, 10)

= ‘passenger distraction’

From vehicle\_collision

group by zip\_code

order by location;

5)Finding out the minimum number of persons injured, killed and minimum number of pedestrians killed and injured by creating table?

Create table vehicle\_collision\_1 as select zip-code, location, date , min(number\_of\_persons\_injured), min(number\_of\_persons\_killed), min(number\_of\_pedestrians\_injured), min(number\_of\_perdestrians\_killed), from vehicle\_collision group by zip\_code, location, order by cast([date] as date\_time) asc;

6)Finding out the number of persons killed at particular date and grouping it by zip code and location?

Create table vehicle\_collision as

Select zip\_code, location

Substr(date, 4, 4) = ‘18’ as date1

Sum(number\_persons\_killed) as persons\_killed

Sum(number\_pedestrians\_killed) as pedestrians\_killed

From vehilcle\_collision

Group by zip\_code, substr(date, 4, 4) = ‘18’

Order by location;

7) finding out number of vehicle collisions in particular month?

Create table vehicle\_collision\_1 as

Select location

From vehicle\_collision

Where substr(date, 4, 4) = ‘18’;