**Air Traffic Management Analysis system using Big Data**

**Analysing flights delay**: To analyze data that finds various reasons for delays and their respective percentages in it.

**Using Hive:**

**Carrier wise Delay**

Select uniquecarrier,

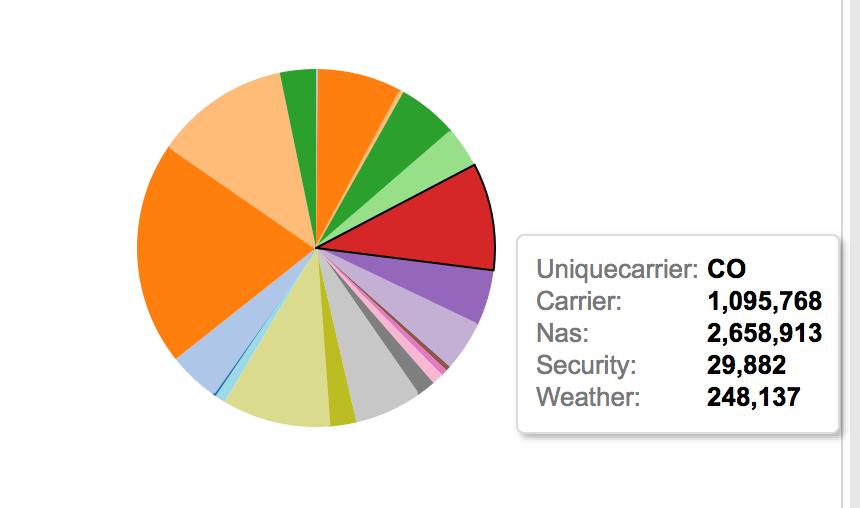
SUM(carrierdelay+weatherdelay+nasdelay+securitydelay+lateaircraftdelay) AS Delay

from flight\_details

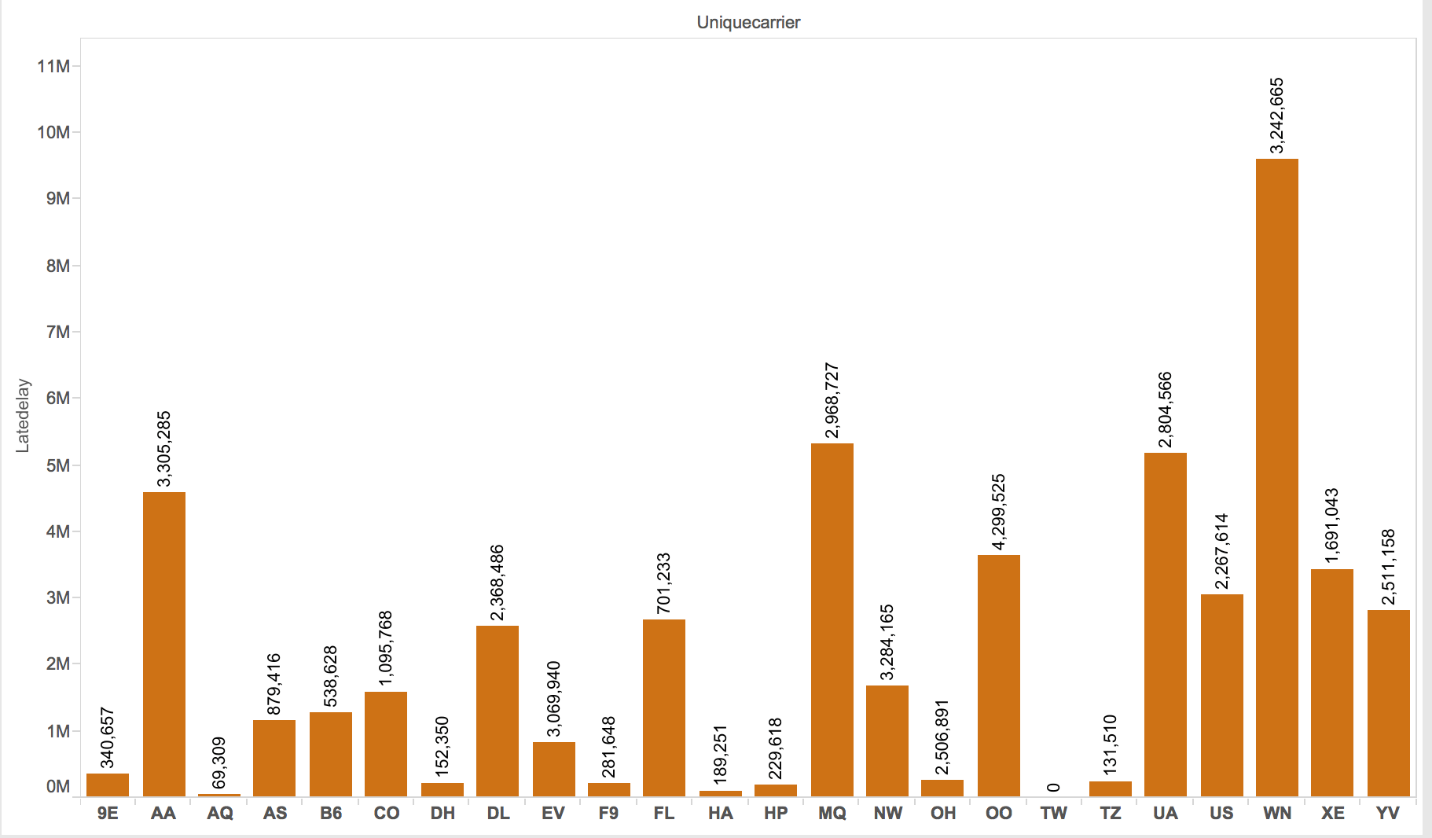
Group By uniquecarrier;



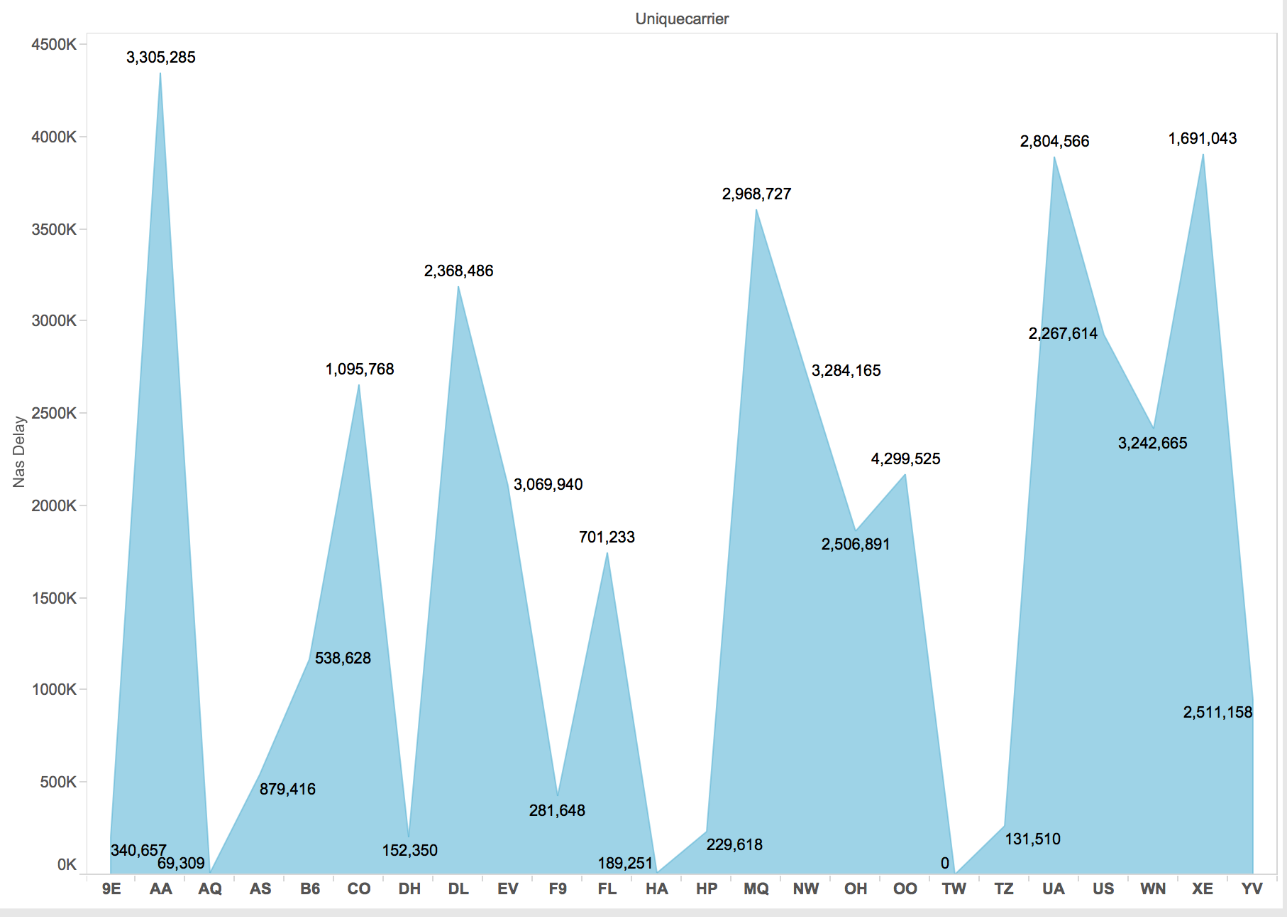
**All Delays according to carrier**:



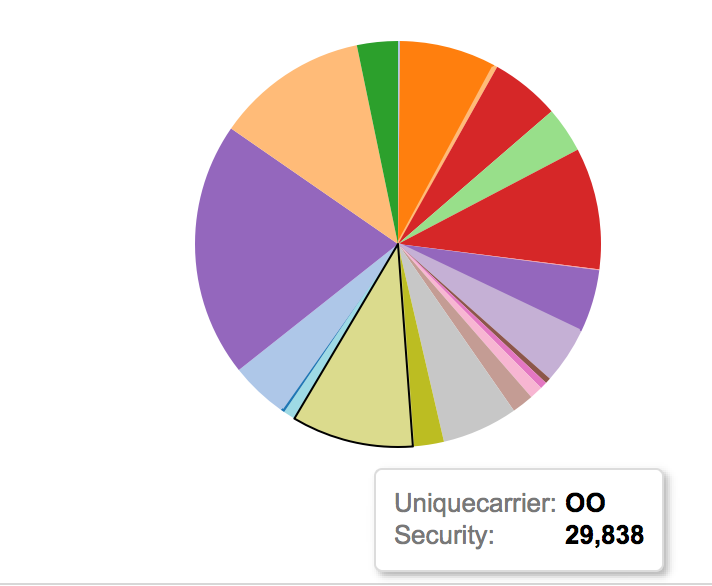
**Late Delay**



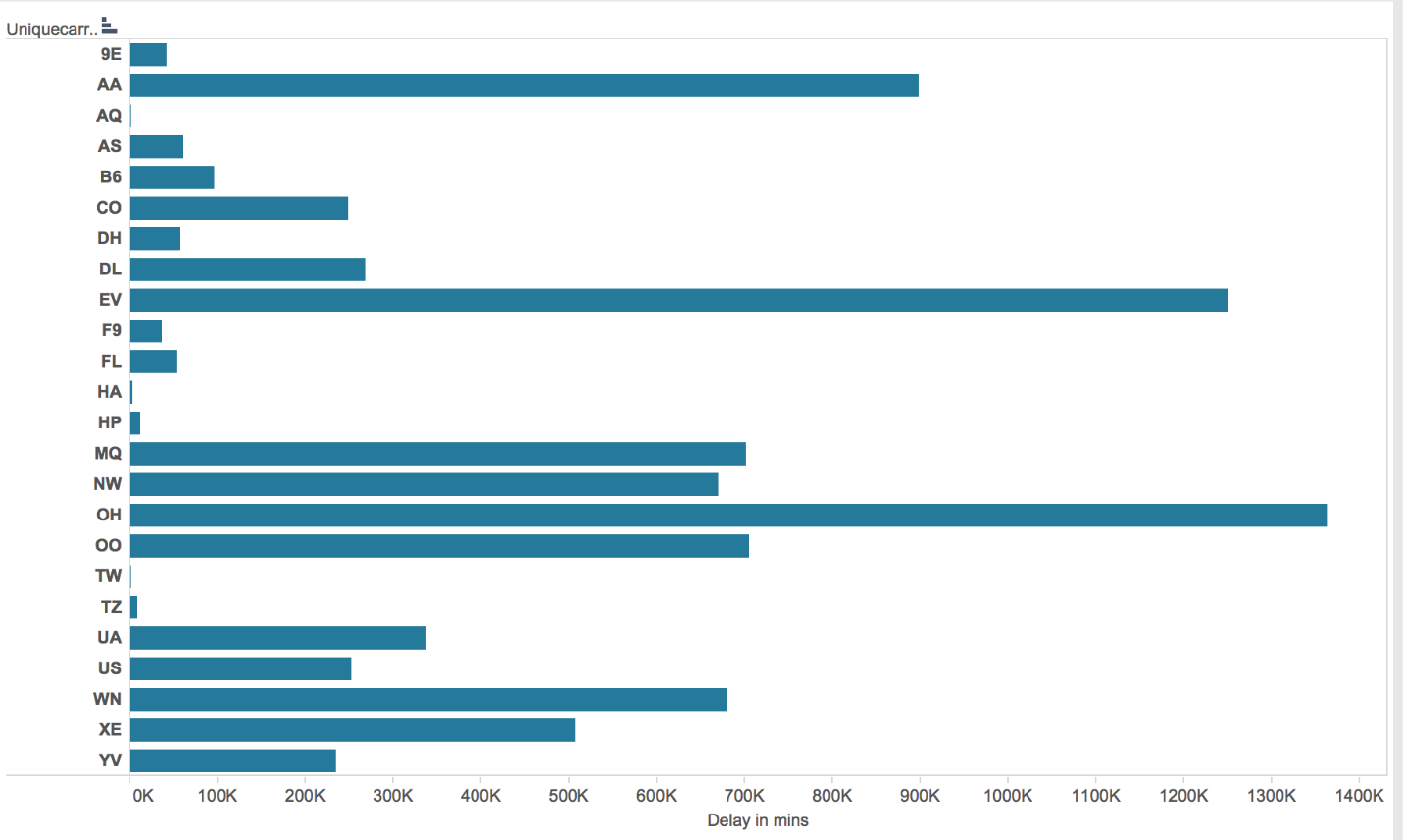
**NAS Delay**



**Security Delay**

****

**Weather Delay**



**Year wise Delay**

Select year,

SUM(carrierdelay+weatherdelay+nasdelay+securitydelay+lateaircraftdelay) AS Delay

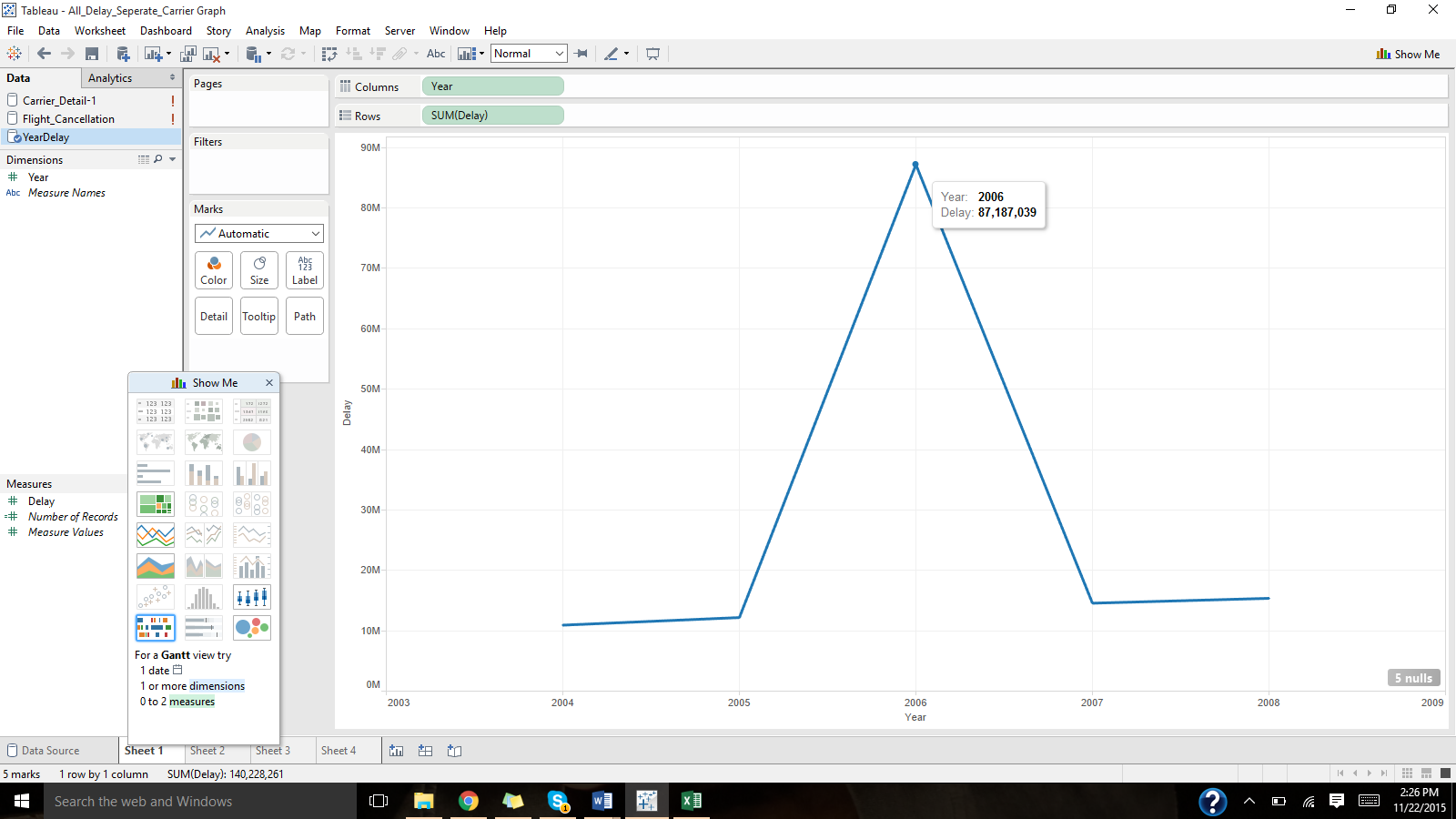
from flight\_details

Where year is NOT NULL and year>1000

Group By year

Order By year;





**Carrier Percentage Delays**

SELECT

A.uniquecarrier,

A.carrier/A.Delay AS PercCarrierDelay,

A.weather/A.Delay AS PercWeatherDelay,

A.nas/A.Delay AS PercNASDelay,

A.security/A.Delay AS PercSecurityDelay,

A.lateDelay/A.Delay AS PercLateDelay

FROM

(Select uniquecarrier,

SUM(carrierdelay) AS carrier,

SUM(weatherdelay) AS weather,

SUM(nasdelay) AS nas,

SUM(securitydelay) AS security,

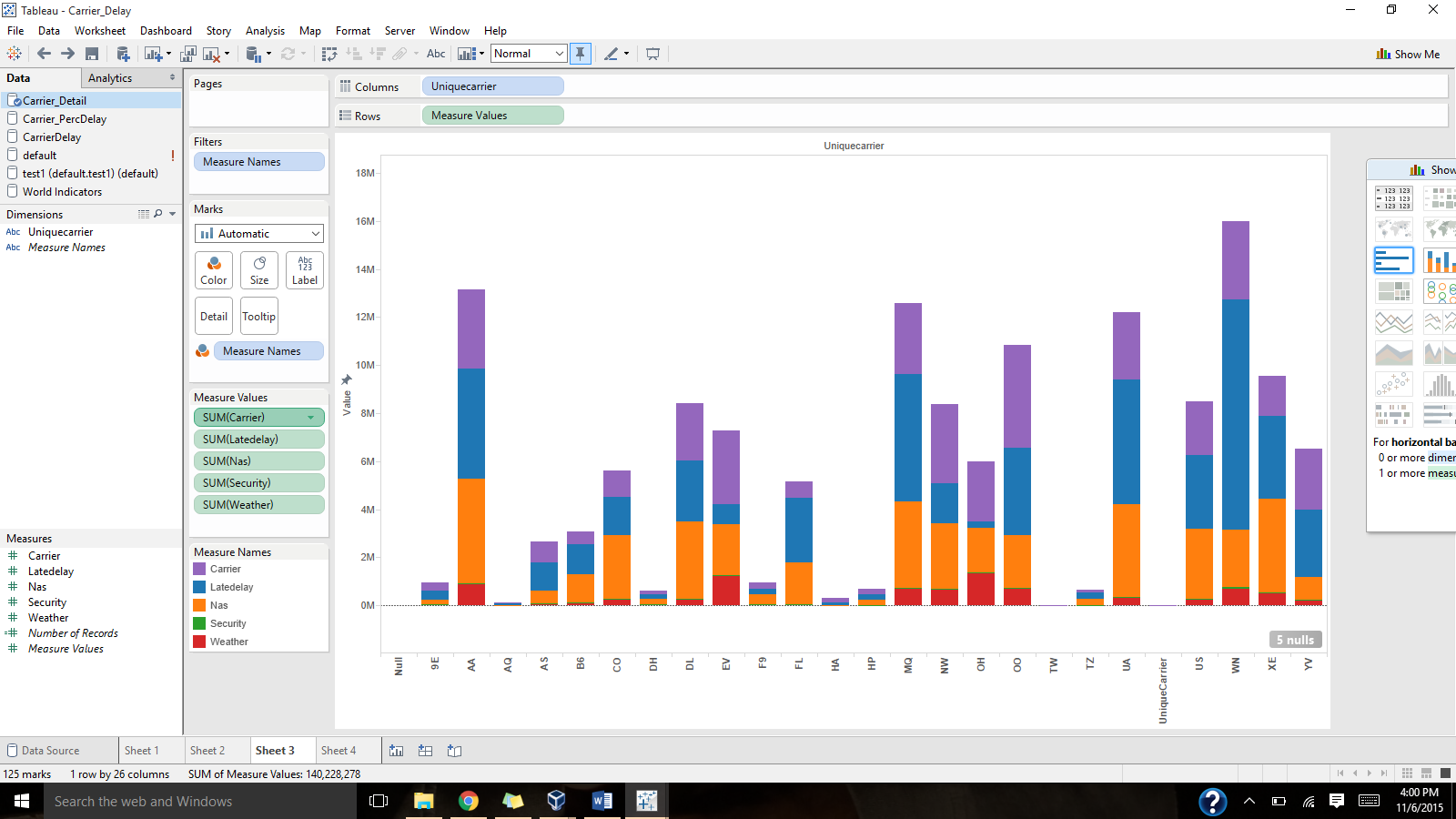
SUM(lateaircraftdelay) AS lateDelay,

SUM(carrierdelay+weatherdelay+nasdelay+securitydelay+lateaircraftdelay) AS Delay

from flight\_details

Group By uniquecarrier)A;





**Using PIG**:

delays = LOAD '/user/hue/airflight.csv' using PigStorage(',') AS (Year:int,Month:int,DayofMonth:int,

DayofWeek:int,DepTime:chararray,CRSDepTime:int,ArrTime:chararray,CRSArrTime:int,UniqueCarrier:chararray,

FlightNum:int,TailNum:chararray,ActualElapsedTime:int,CRSElapsedTime:int,AirTime:chararray,

ArrDelay:chararray,DepDelay:chararray,Origin:chararray,Dest:chararray,Distance:int,TaxiIn:int,

TaxiOut:Int,Cancelled:int,CancellationCode:chararray,Diverted:int,CarrierDelay:int,

WeatherDelay:int,NASDelay:int,SecurityDelay:int,LateAircraftDelay:int);

b = FOREACH delays GENERATE $0,$24,$25,$26,$27,$28;

c = FILTER b BY ($1>0 OR $2>0 OR $3>0 OR $4>0 OR $5>0);

g = GROUP c BY $0;

d = FOREACH c GENERATE $0,($1+$2+$3+$4+$5) AS SumD;

dump d;



**Analysing flights cancellations**: To analyze data that finds various reasons for cancellations and their respective percentages in it.

**Using HIVE**:

SELECT

cancellationcode,

count(cancellationcode)

FROM

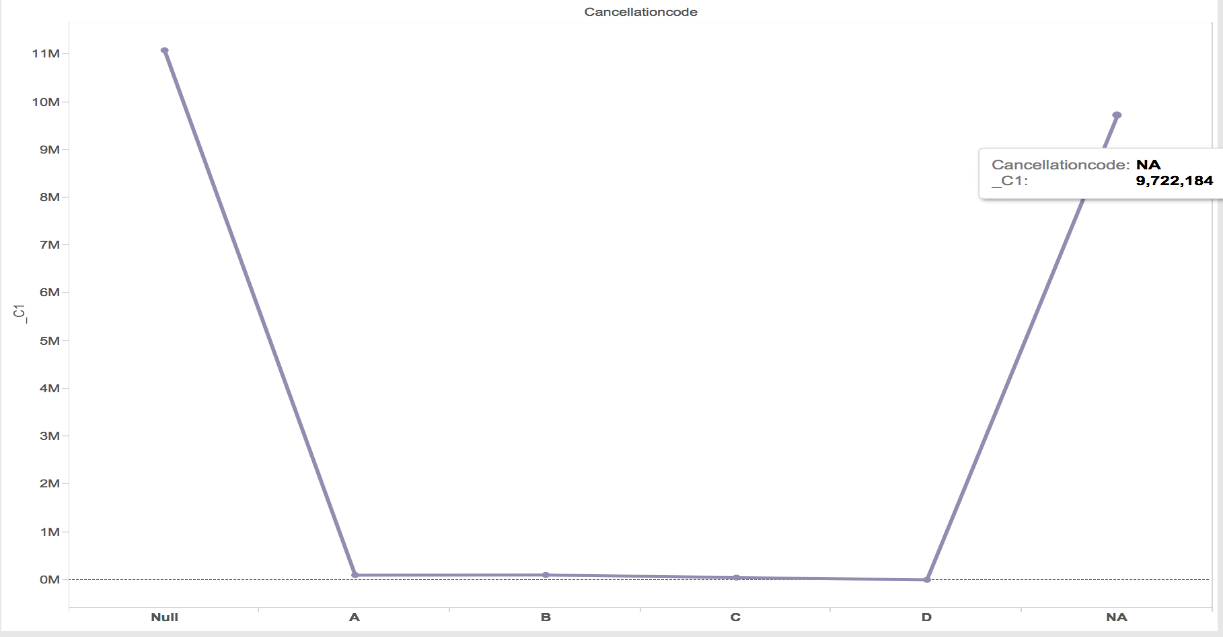
flight\_details

WHERE

cancellationcode<>'NA' OR cancellationcode IS NOT NULL

Group by cancellationcode;





**Using Pig** -

Dept = LOAD '/user/hue/airflight.csv' using PigStorage(',') AS (Year:int,Month:int,DayofMonth:int,

DayofWeek:int,DepTime:chararray,CRSDepTime:int,ArrTime:chararray,CRSArrTime:int,UniqueCarrier:chararray,

FlightNum:int,TailNum:chararray,ActualElapsedTime:int,CRSElapsedTime:int,AirTime:chararray,

ArrDelay:chararray,DepDelay:chararray,Origin:chararray,Dest:chararray,Distance:int,TaxiIn:int,

TaxiOut:Int,Cancelled:int,CancellationCode:chararray,Diverted:int,CarrierDelay:int,

WeatherDelay:int,NASDelay:int,SecurityDelay:int,LateAircraftDelay:int);

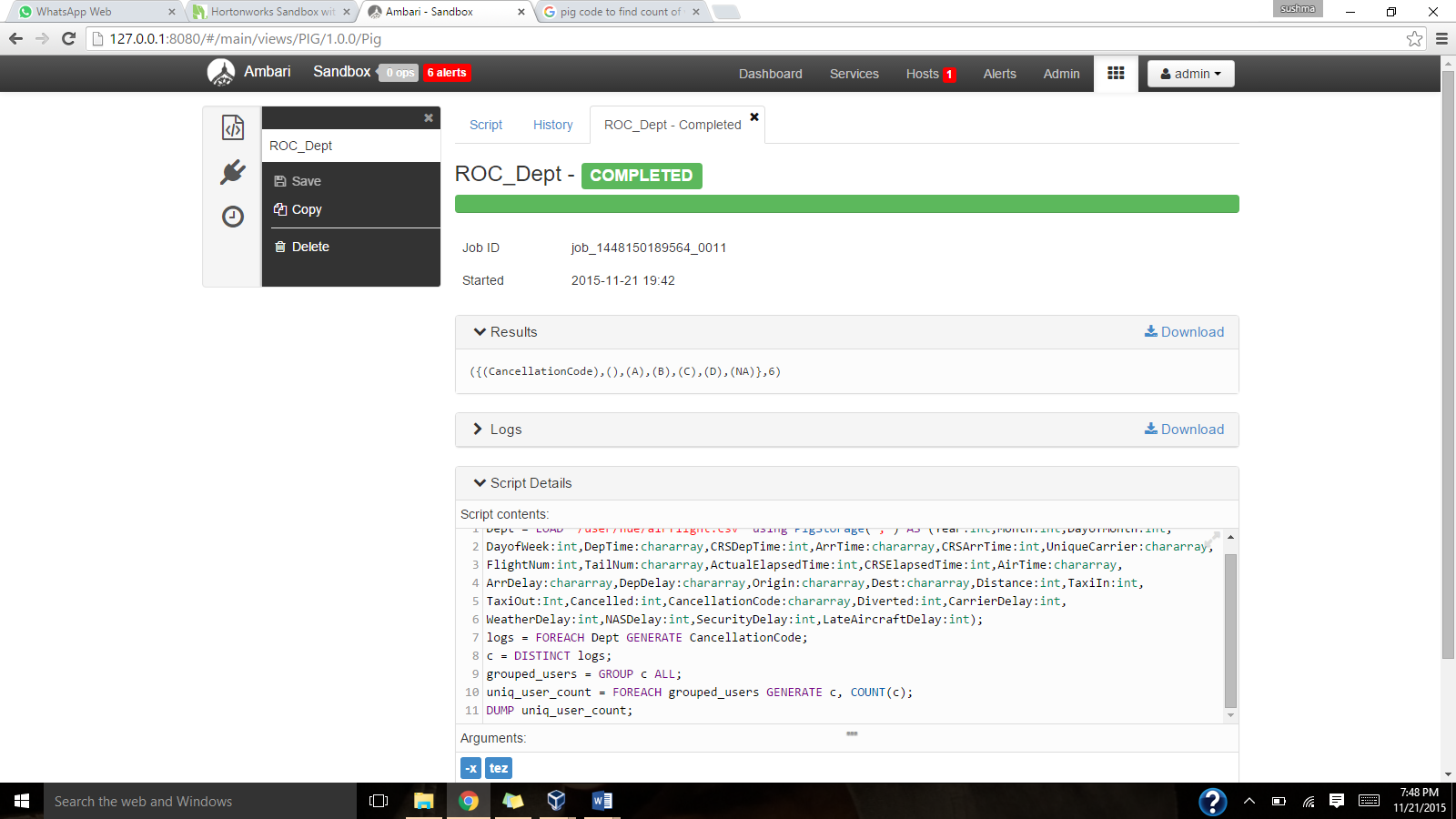
logs = FOREACH Dept GENERATE CancellationCode;

c = DISTINCT logs;

grouped\_users = GROUP c ALL;

uniq\_user\_count = FOREACH grouped\_users GENERATE c, COUNT(c);

DUMP uniq\_user\_count;



**To study flight diversions**: A built in functionality will be used for displaying list of flights that were diverted.

**Using Hive**:

**According to origin and destination** -

SELECT

flightnum,

origin,

dest

FROM

flight\_details

where diverted=1;



**According to date** -

SELECT

flightnum,

concat(month,'/',dayofmonth,'/', year) AS diverted\_date,

origin,

dest

FROM

flight\_details

where diverted=1

order by flightnum;



**Using PIG**:

diversions = LOAD '/user/hue/airflight.csv' using PigStorage(',') AS (Year:int,Month:int,DayofMonth:int,

DayofWeek:int,DepTime:chararray,CRSDepTime:int,ArrTime:chararray,CRSArrTime:int,UniqueCarrier:chararray,

FlightNum:int,TailNum:chararray,ActualElapsedTime:int,CRSElapsedTime:int,AirTime:chararray,

ArrDelay:chararray,DepDelay:chararray,Origin:chararray,Dest:chararray,Distance:int,TaxiIn:int,

TaxiOut:Int,Cancelled:int,CancellationCode:chararray,Diverted:int,CarrierDelay:chararray,

WeatherDelay:chararray,NASDelay:chararray,SecurityDelay:chararray,LateAircraftDelay:chararray);

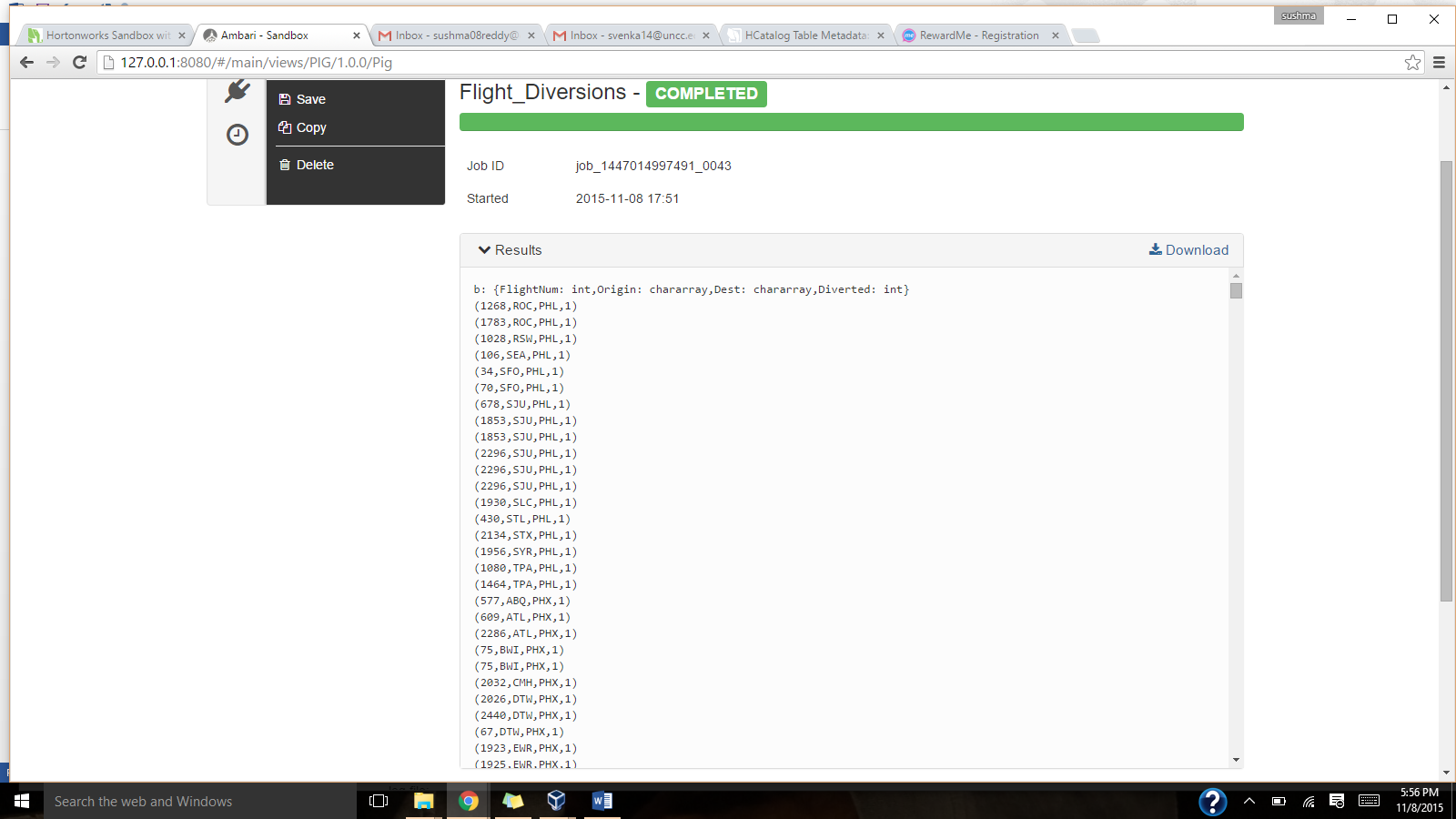
a= FOREACH diversions generate $9,$16,$17,$23;

b = FILTER a BY Diverted == 1;

describe b;

Dump b;





**Traffic Density**: This will help the airport authorities to study the traffic rates of the flights at different time periods which would help for future flight schedules.

**Using Hive**:

**Departure** -

SELECT

'ROC',

dayofmonth,

deptime

FROM

flight\_details

Where origin='ROC' and year = 1999 and month=1 and deptime <> 'NA'

Order by dayofmonth,deptime;





**Arrival** -

SELECT

'ROC',

dayofmonth,

arrtime

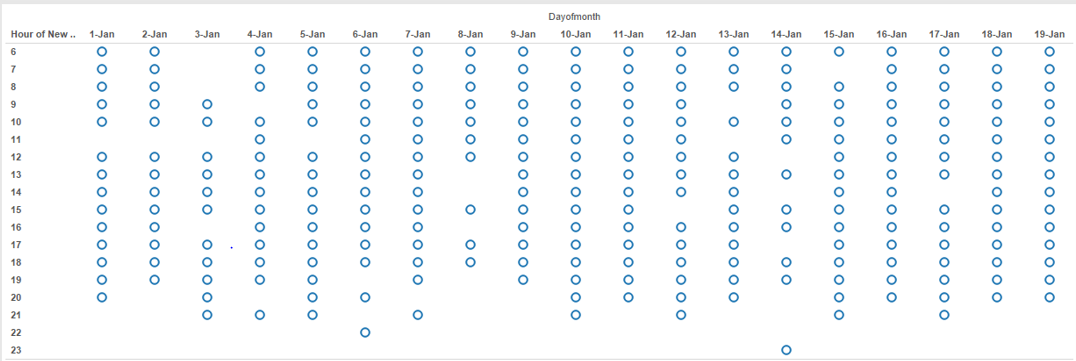
FROM

flight\_details

Where DEST='ROC' and year = 1999 and month=1 and arrtime <> 'NA'

Order by dayofmonth,arrtime;





**Using Pig**-

delays = LOAD '/user/hue/airflight.csv' using PigStorage(',') AS (Year:int,Month:int,DayofMonth:int,

DayofWeek:int,DepTime:chararray,CRSDepTime:int,ArrTime:chararray,CRSArrTime:int,UniqueCarrier:chararray,

FlightNum:int,TailNum:chararray,ActualElapsedTime:int,CRSElapsedTime:int,AirTime:chararray,

ArrDelay:chararray,DepDelay:chararray,Origin:chararray,Dest:chararray,Distance:int,TaxiIn:int,

TaxiOut:Int,Cancelled:int,CancellationCode:chararray,Diverted:int,CarrierDelay:int,

WeatherDelay:int,NASDelay:int,SecurityDelay:int,LateAircraftDelay:int);

b = FOREACH delays GENERATE $0,$1,$2,$5,$16;

c = FILTER b BY (Year == 1999 AND Month == 1 AND $4 == 'ROC');

dump c;

