Sum of delays:

delays = LOAD '/tmp/Cloud/pig5.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;

cancellation:

Dept = LOAD '/tmp/Cloud/pig5.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;

Diversion by org dest:

diversions = LOAD '/tmp/Cloud/pig5.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;

Density departure

delays = LOAD '/tmp/Cloud/pig5.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 == 2004 AND Month == 1 AND $4 == 'SYR');

dump c;