**Question1)**

a = load '/user/pig/full\_text.txt' AS (id:chararray, ts:chararray, location:chararray, lat:float, lon:float, tweet:chararray);

b = filter a by ToDate(ts)==ToDate(2010-03-06);

c = foreach b generate GetHour(ToDate(ts)) as h1;

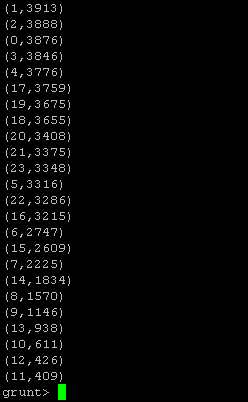
d = group c by h1;

e = foreach d generate group, COUNT(C) as cnt;

f = order e by cnt desc;

g = limit f 24;

Dump g;



**Question2:**

a = load '/user/pig/full\_text.txt' AS (id:chararray, ts:chararray, location:chararray, lat:float, lon:float, tweet:chararray);

b = foreach a generate FLATTEN(TOKENIZE (tweet)) as tweet1;

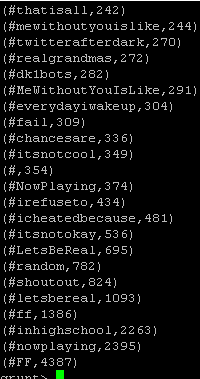
c = group b by tweet1;

d = foreach c generate group as tweet1, COUNT(b) as cnt;

e = order d by cnt asc;

f = filter e by tweet1 matches ‘#\\s\*(\\w+)’;

dump f ;



Question 3:

a =  load '/user/pig/full\_text.txt' AS (id:chararray, ts:chararray, location:chararray, lat:float, lon:float, tweet:chararray);

b =  foreach a generate FLATTEN(TOKENIZE (tweet)) as tweet1;

c =  group b by tweet1;

d =  foreach c generate group as tweet1, COUNT(b) as cnt;

e = order d by cnt asc;

f = filter e by tweet1 matches ‘@.\*’;

dump f ;

