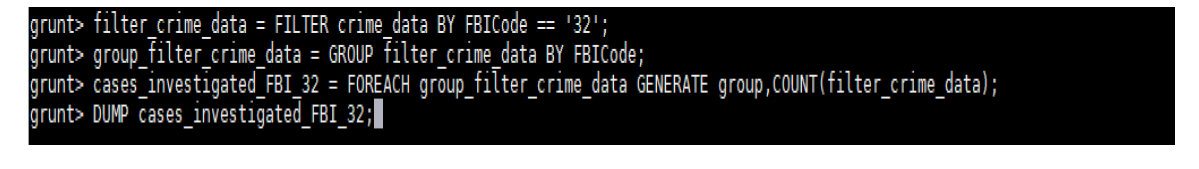


Write a MapReduce/Pig program to calculate the number of cases investigated under FBI code 32.

filter\_crime\_data = FILTER crime\_data BY FBICode == '32';

group\_filter\_crime\_data = GROUP filter\_crime\_data BY FBICode;

cases\_investigated\_FBI\_32 = FOREACH group\_filter\_crime\_data GENERATE group,COUNT(filter\_crime\_data);







hadoop fs -ls file:/tmp/temp1141537601/tmp2119968396



hadoop fs -cat file:/tmp/temp1141537601/tmp2119968396/part-r-00000



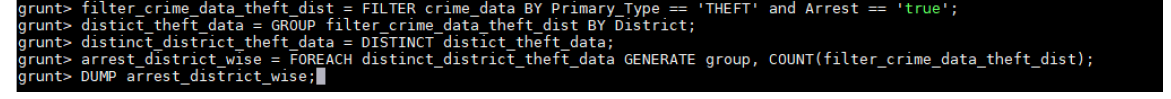
Write a MapReduce/Pig program to calculate the number of arrests in theft district wise.

filter\_crime\_data\_theft\_dist = FILTER crime\_data BY Primary\_Type == 'THEFT' and Arrest == 'true';

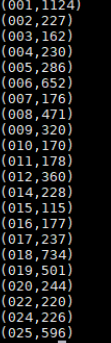
distict\_theft\_data = GROUP filter\_crime\_data\_theft\_dist BY District;

distinct\_district\_theft\_data = DISTINCT distict\_theft\_data;

arrest\_district\_wise = FOREACH distinct\_district\_theft\_data GENERATE group, COUNT(filter\_crime\_data\_theft\_dist);



Output:



Write a MapReduce/Pig program to calculate the number of arrests done between October 2014 and October 2015.

crime\_details\_between\_dates = FILTER crime\_data BY ToDate(Date, 'MM/dd/yyyy hh:mm:ss a') >= ToDate('10/01/2014 12:00:00 AM', 'MM/dd/yyyy hh:mm:ss a') AND ToDate(Date, 'MM/dd/yyyy hh:mm:ss a') <= ToDate('10/31/2015 11:59:59 PM', 'MM/dd/yyyy hh:mm:ss a');

filter\_crime\_data\_arrest = FILTER crime\_details\_between\_dates BY Arrest == 'true';

group\_filter\_crime\_data\_arrest = GROUP filter\_crime\_data\_arrest ALL;

final\_result = FOREACH group\_filter\_crime\_data\_arrest GENERATE group, COUNT(filter\_crime\_data\_arrest.Arrest);

