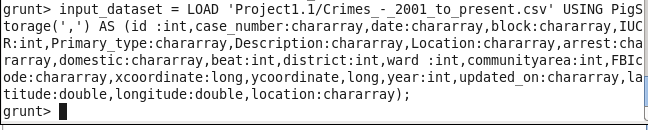
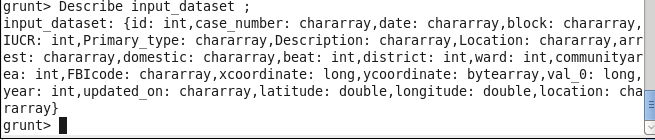
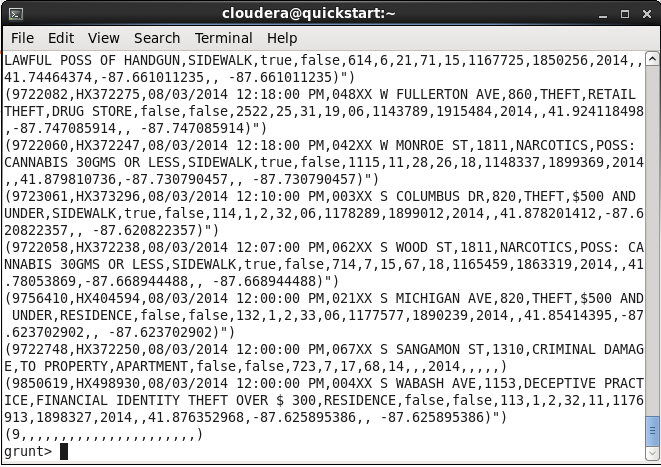
**Project 1.1 (USA Crime Analysis)**

1. Write a MapReduce/Pig program to calculate the number of cases investigated under each FBI Code
2. First Load the data into HDFS as shown below.

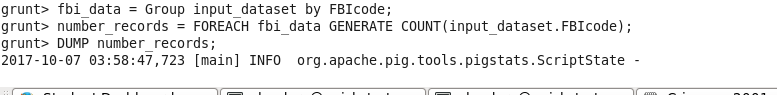






1. THEN Group The data BY FBICODE

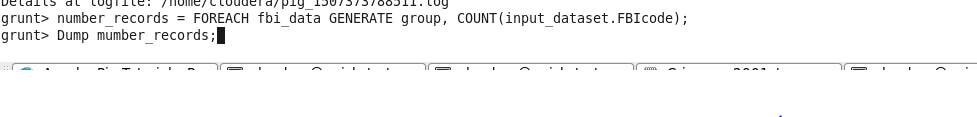
Fbi\_data = GROUP input\_dataset by FBIcode;



1. After that we will count the number of records under each FBI code as shown below.

Num\_records = FOREACH fbi\_data GENERATE group,COUNT(input\_dataset.FBIcode);

After that we dump the records;



**OUTPUT**

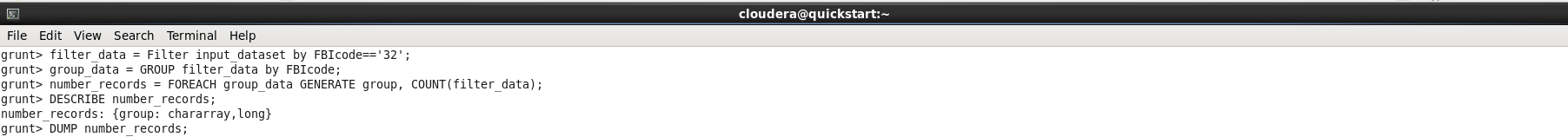




1. **Write a MapReduce/Pig program to calculate the number of cases investigated under FBI**

**code 32**.

1. First Load the data into HDFS as we do in Earlier question
2. After that we will filter thae data as our requirement
3. After that we will group the data by FBIcode
4. After that we will aggregation function



OUTPUT





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

1. First Load the data into HDFS
2. After that we will filter the data using Command

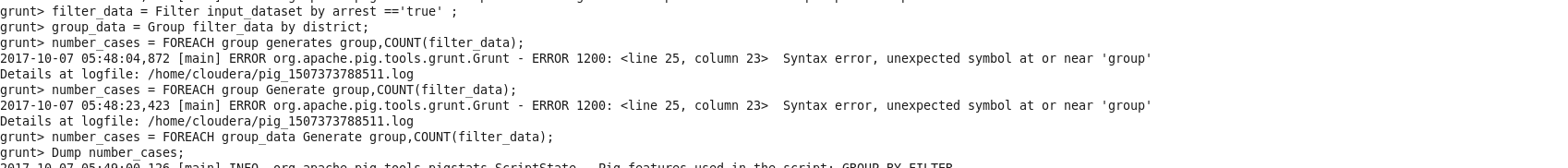
Filter\_data = FILTER input\_dataset by arrest ==’TRUE’

1. After that we will group the data by district

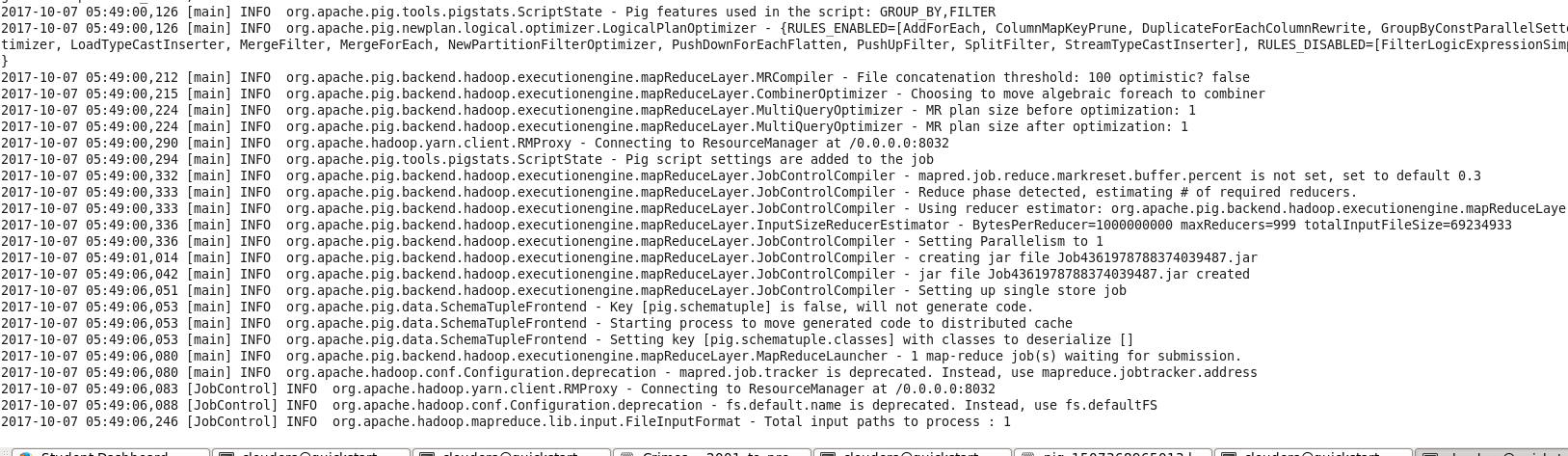
Group\_data = Group filter\_data by district;

d)After that we will aggregaye the data

number\_records = FOREACH group\_data Generate group,COUNT(filter\_data);



OUTPUT



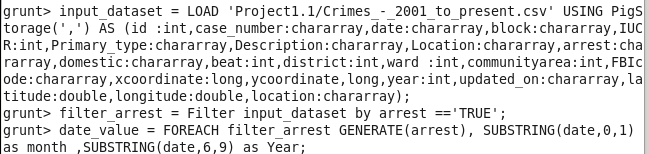


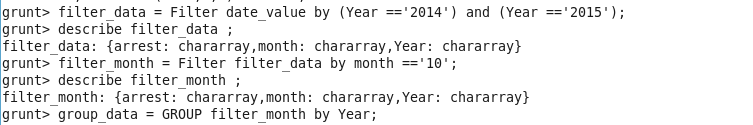


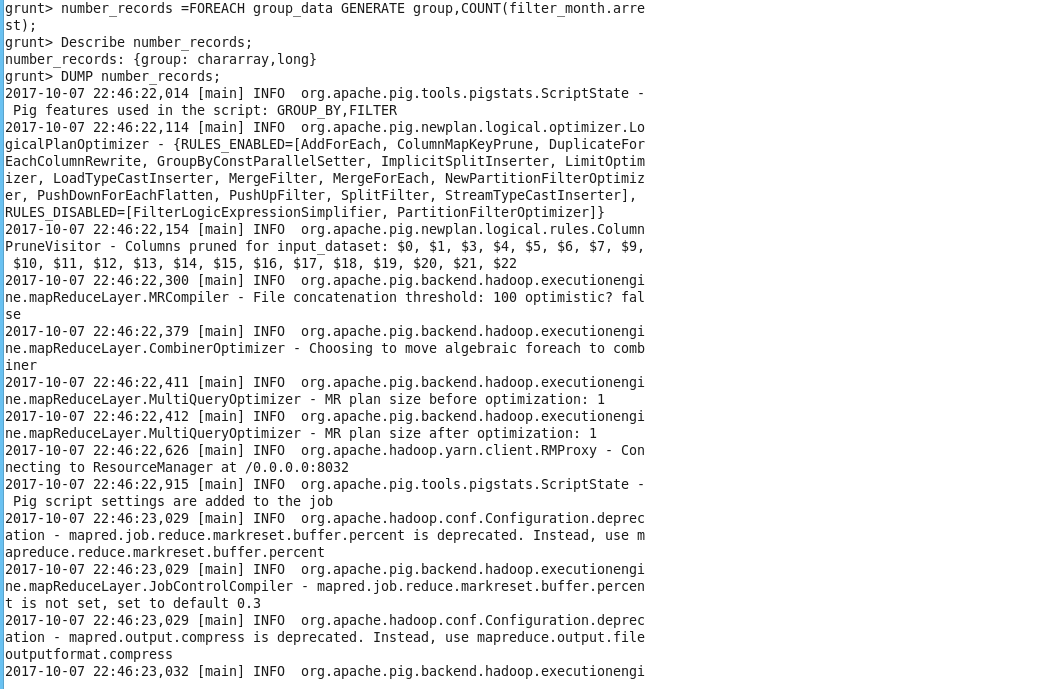
1. Write a MapReduce/Pig program to calculate the number of arrests done between October
2. October 2015.
3. First load the data into HDFS
4. After that we will filter the data where arrest == true
5. After that we will get the month and year from date string using SUBSTRING FUNCTION

date\_value = FOREACH filter\_data GENERATE(arrest),SUBSTRING(date,0,1)as month,SUBSTRING(date,6,9) as YEAR

1. After that we will filter the data by year
2. After that we will filter the data by month
3. After that we will group the data by year
4. Then we perform aggregation function and then dump the data







OUTPUT

