**Task 1**: Write a pig script to find no of complaints which got timely response.

DATA = LOAD '/project1/Consumer\_Complaints.csv' USING TextLoader() AS (line:chararray);

B = FOREACH DATA GENERATE FLATTEN(STRSPLIT(line, ',(?=(?:[^"]\*"[^"]\*")\*[^"]\*$)'));

C = FOREACH B GENERATE $13,$15 ;

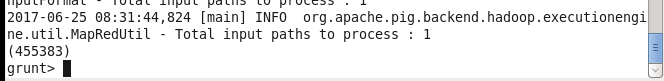
D = FILTER C BY $0 == 'Yes' ;

E = GROUP D ALL;

F = FOREACH E GENERATE COUNT(D.$1);

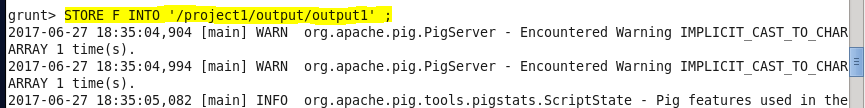
**After dumping the result**

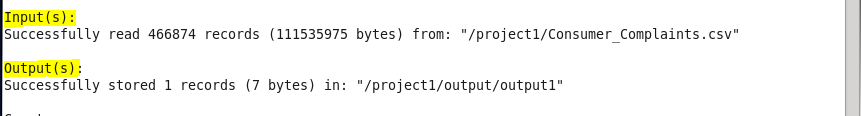
Dump F;



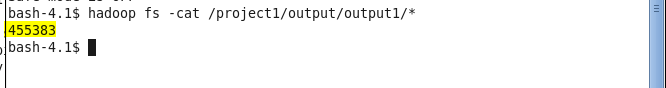
**Storing result in HDFS**

STORE F INTO '/project1/output/output1’ ;





**Output**



**After exporting to mysql**



**Task 2**: Write a pig script to find no of complaints where consumer forum forwarded the complaint same day they received to respective company.

DATA = LOAD '/project1/Consumer\_Complaints.csv' USING TextLoader() AS (line:chararray);

A = FOREACH DATA GENERATE FLATTEN(STRSPLIT(line, ',(?=(?:[^"]\*"[^"]\*")\*[^"]\*$)'));

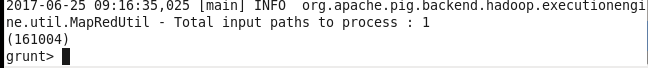
B = FILTER A BY $0 == $11 ;

C = GROUP B ALL;

D = FOREACH C GENERATE COUNT(B.$15);

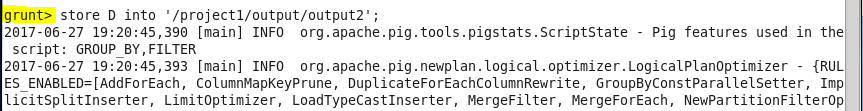
**After dumping the result**

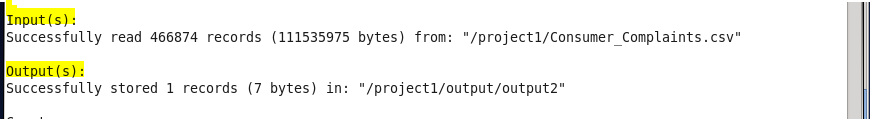
Dump D;



**Storing result in HDFS**

STORE D INTO '/project1/output/output2’ ;





**Output**



**After exporting to mysql**



**Task 3**: Write a pig script to find list of companies toping in complaint chart (companies with maximum number of complaints).

DATA = LOAD '/project1/Consumer\_Complaints.csv' USING TextLoader() AS (line:chararray);

A = FOREACH DATA GENERATE FLATTEN(STRSPLIT(line, ',(?=(?:[^"]\*"[^"]\*")\*[^"]\*$)'));

A1 = FILTER A BY $7 != 'Company' ;

B = FOREACH A1 GENERATE $7, $15 ;

C = GROUP B BY $0 ;

D = FOREACH C GENERATE group , COUNT(B.$1);

E = ORDER D BY $1 DESC ;

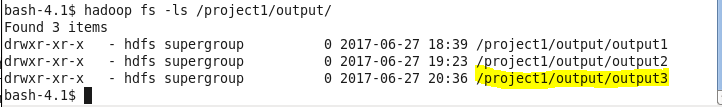
**After dumping the result**

DUMP E;

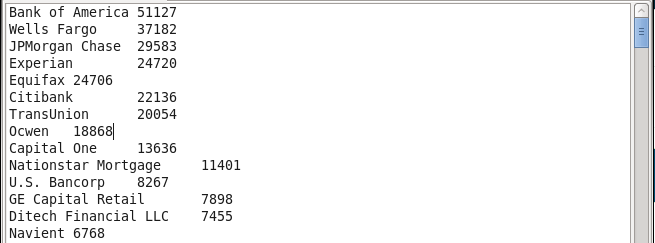


**Storing result in HDFS**

STORE E INTO '/project1/output/output3’ ;



**Output:**



**Task 4:** Write a pig script to find no of complaints filed with product type has “ Debt collection” for the year 2015.

DATA = LOAD '/project1/Consumer\_Complaints.csv' USING TextLoader() AS (line:chararray);

A = FOREACH DATA GENERATE FLATTEN(STRSPLIT(line, ',(?=(?:[^"]\*"[^"]\*")\*[^"]\*$)'));

B = FILTER A BY ( $1 == 'Debt collection' AND SUBSTRING($0,6,10) == '2015') ;

C = GROUP B ALL;

D = FOREACH C GENERATE COUNT(B.$15) ;

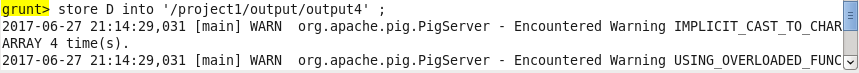
**After dumping the result**

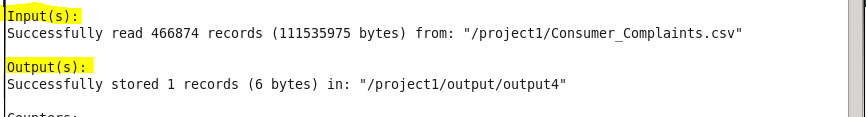
DUMP D;



**Storing result in HDFS**

STORE D INTO ' /project1/output/output4’ ;





**Output:**

****