**Consumer complaint Analysis**

**Data Description:**

The dataset consists of consumer complaints details. In the process consumer files the complaints to consumer forum and then consumer forum forwards the complaint to respective company.

It contains the following fields

Date received,Product,Sub-product,Issue,Sub-issue,Consumer complaint narrative,Company public response,Company,State,ZIP code,Submitted via,Date sent to company,Company response to consumer,Timely response?,Consumer disputed?,Complaint ID,

**sample data**

10/16/2015,Debt collection,"Other (phone, health club, etc.)",Cont'd attempts collect debt not

owed,Debt was discharged in bankruptcy,,,"Convergent Resources, Inc.",OH,438XX,Web,10/16/2015,Closed with explanation,Yes,,1612132

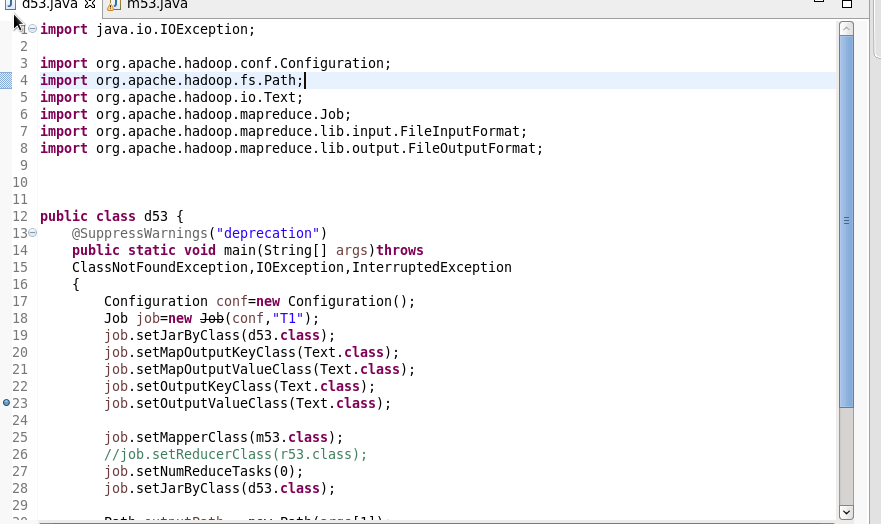
This entire column "Other (phone, health club, etc.)"

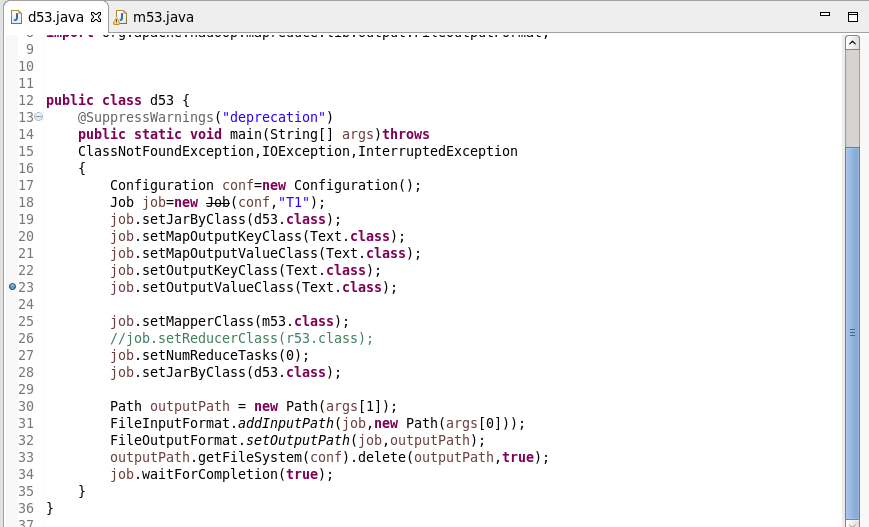
Should be product but, if we split this file based on comma then this column will be splitted into 3 columns which will result in wrong outputs.

In order to tackle this we should remove the commas present only inside double quotes

**MAP REDUCE TO REMOVE ALL COMMAS WITHIN DOUBLE QUOTES**

**DRIVER CLASS**



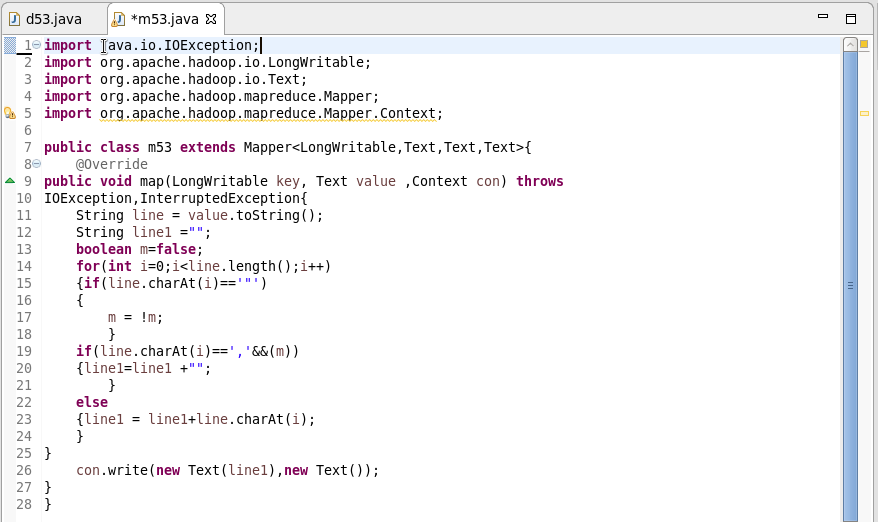


**MAPPER CLASS**

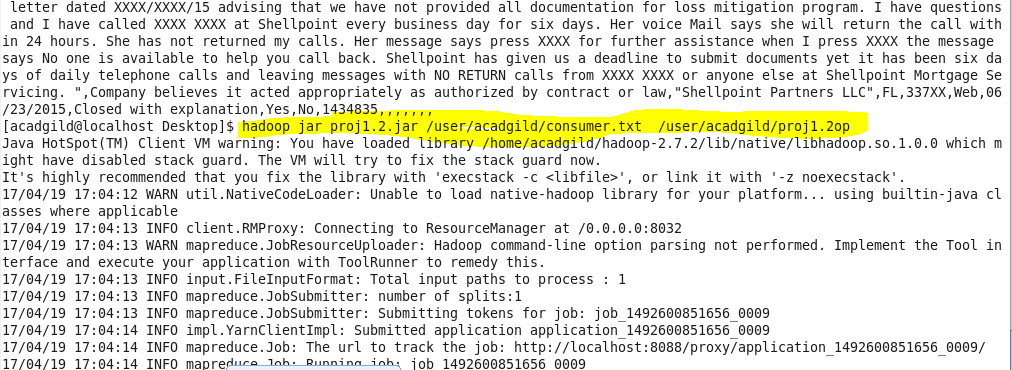
**LOGIC: I WILL SET A BOOLEAN AS FALSE AND IT WILL REVERSES IT STATE WHEN EVER THERE IS A DOUBLE QUOTE**

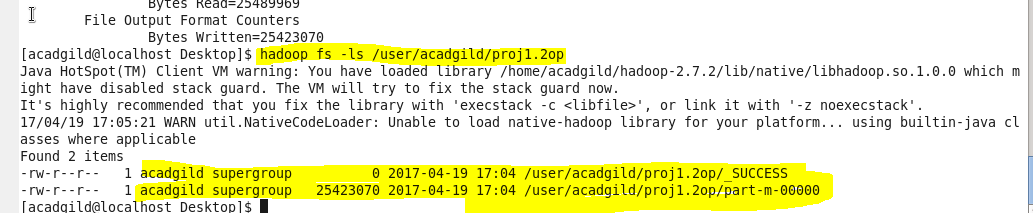
**STEP 2: WHENEVER IT ENCOUNTERS A COMMA AND IS WITHIN DOULE QUOTES IT REPLACES THE COMMA AND IT STORE OR ELSE THE CHARACTER IS WRITTEN WITH THE HELF OF If CONDITION AS HIGHLIGHTED**

**Description:**

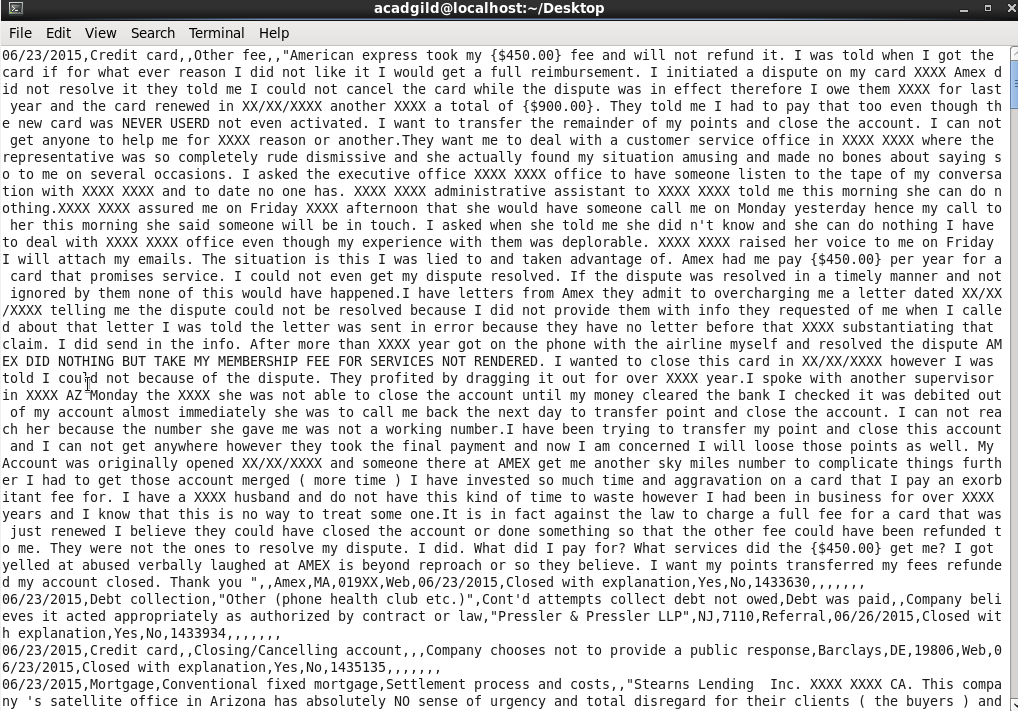


RUNNING JAR

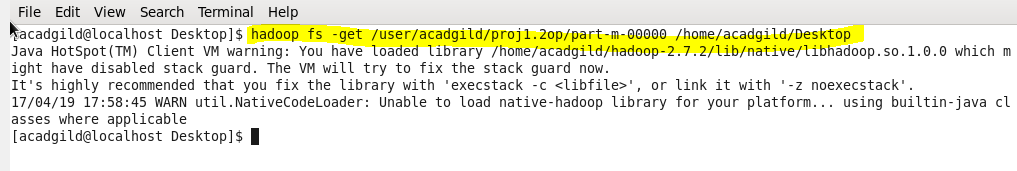




OUTPUT



**USED GET COMMAND TO COPY TO LOCAL FOR DOING PIG**

****

**PIG COMMANDS**

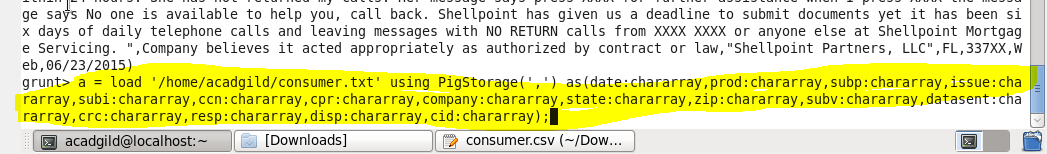
1. **Write a pig script to find no of complaints which got timely response**

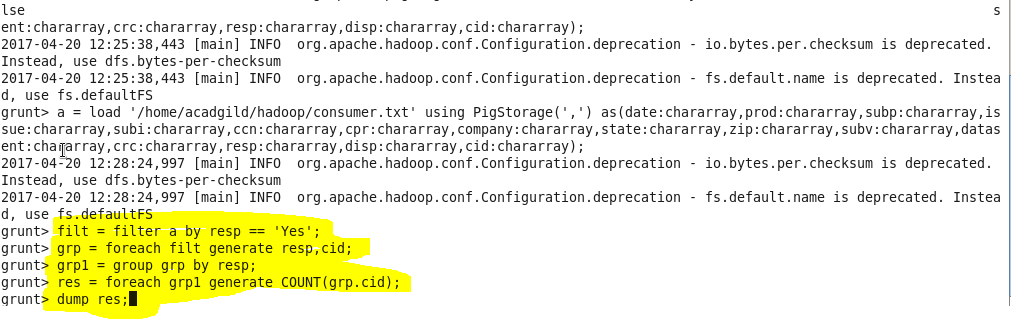
**LOGIC:**

**Step 1:loading cleared data**

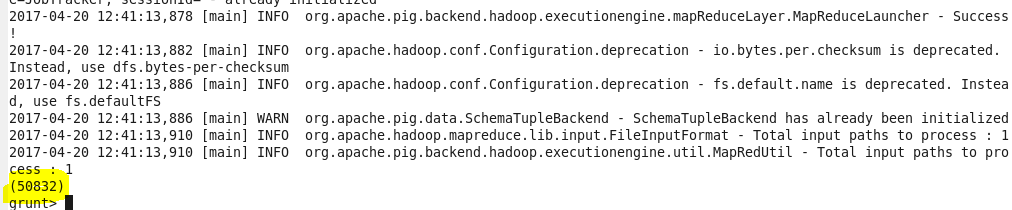
**Step2: filter complaints where timely response is true**

**Step 3: group by timely response and make a count of id**

****

****

**OUTPUT**

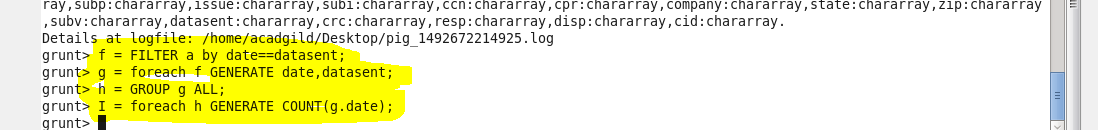
****

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

**1:loading cleared data**

**2:filter where date received =date sent to company**

**3:making COUNT AFTER GROUPIN BY ALL**

****

**OUTPUT**

****

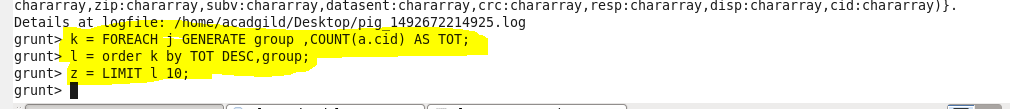
1. **Write a pig script to find list of companies topping in complaint chart (companies with maximum number of complaints)**

**1:loading cleared data**

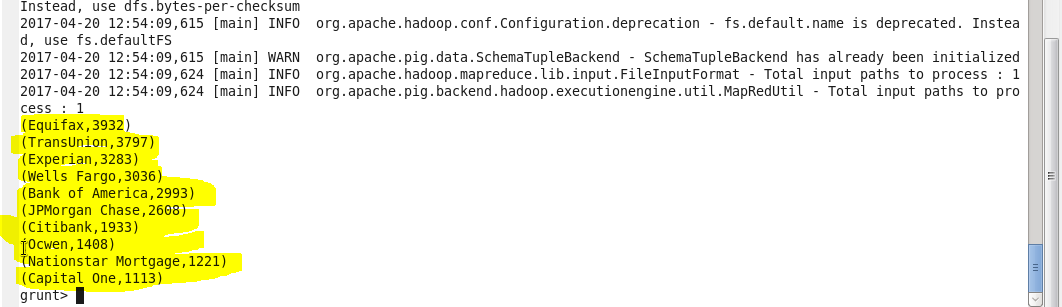
**2:group by company name**

**3:make a count of cid for Each company**

**4: order by desc order of counts and limit by 10**

****

**OUTPUT**

****

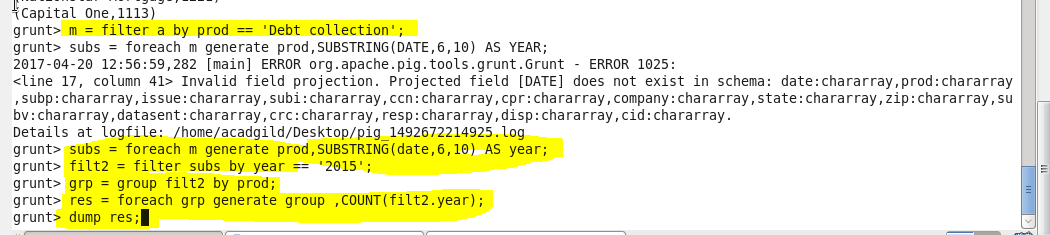
1. **Write a pig script to find no of complaints filed with product type has "Debt collection" for the year 2015**

**1:filter by product =”DEBT COLLECTION”**

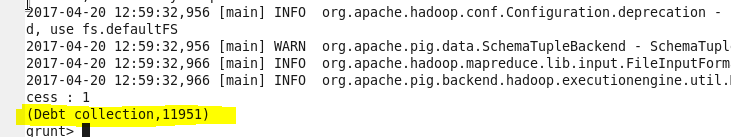
**2:USED SUBSTRING FUNCTION TO GENERATE YEAR as “DAT”**

**3:filter by DAT==2015**

**4:MADE COUNT**

****

**OUTPUT**

****