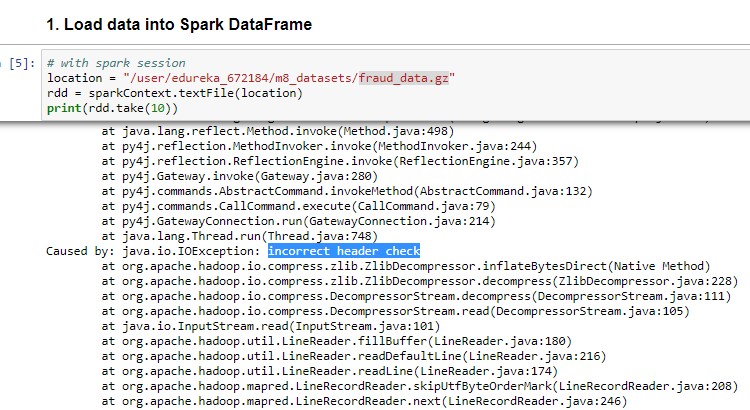
I couldn’t load the data. I created several tickets but none helped.

1. The gz file is corrupted so cannot load it with sparkContext



1. I tried to extract the gz file, put onto hdfs and load it with sparkContext. Spark shows Name quota exceeded

**tar xvf fraud\_data.gz**

**hdfs dfs -put maildir /user/edureka\_672184/m8\_datasets/maildir.gz**

1. I tried to zip extracted files into a zip and put it onto hdfs

zip -r fraud.zip maildir

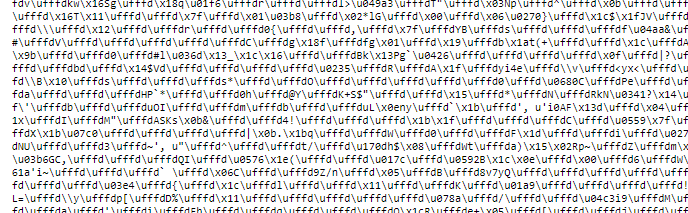
**hdfs dfs -put fraud.zip /user/edureka\_672184/m8\_datasets/maildir.gz**

1. Now I was able to read the zip file with spark context but I couldn’t find the 150 users and 0.5 M messages. It shows unreadble text

**location = "/user/edureka\_672184/m8\_datasets/fraud.zip"**

**rdd = sparkContext.textFile(location)**

**print(rdd.take(200))**

****