ASSIGNMENT 6.5

1)A pig script to find the top 20 destination people travel most.

A) rel = load ‘/user/acadgild/TravelData.txt’ using PigStorage(‘ ‘) as

(citypair:chararray,fromlocation:chararray,

tolocation:chararray,Producttype:int,

adults:int,seniors:int,children:int,

youth:int,infant:int,dateoftravel:chararray,

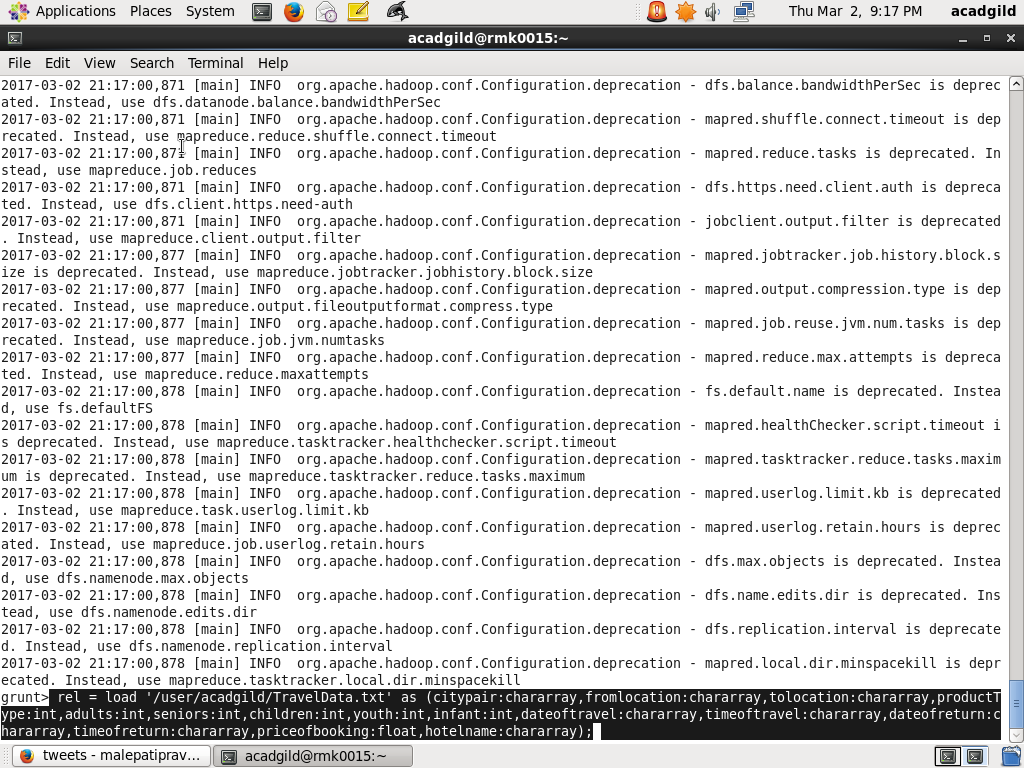
timeoftravel:chararray,dateofreturn:chararray,

timeofreturn:chararray,priceofbooking:float,

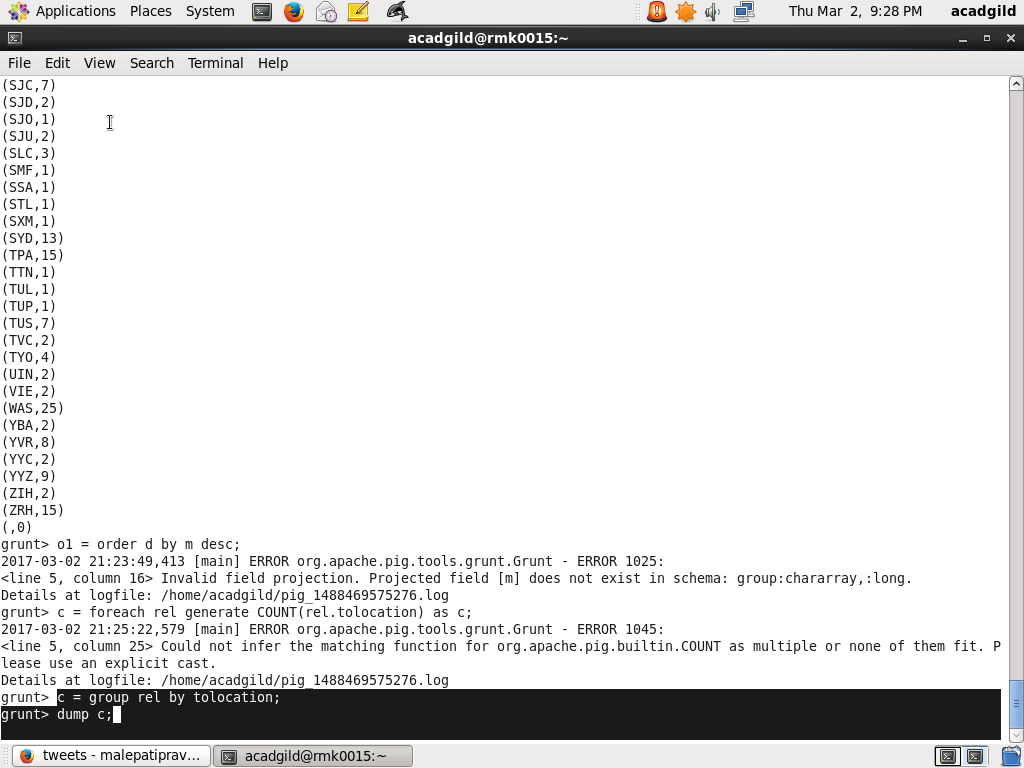
hotelname:chararray);

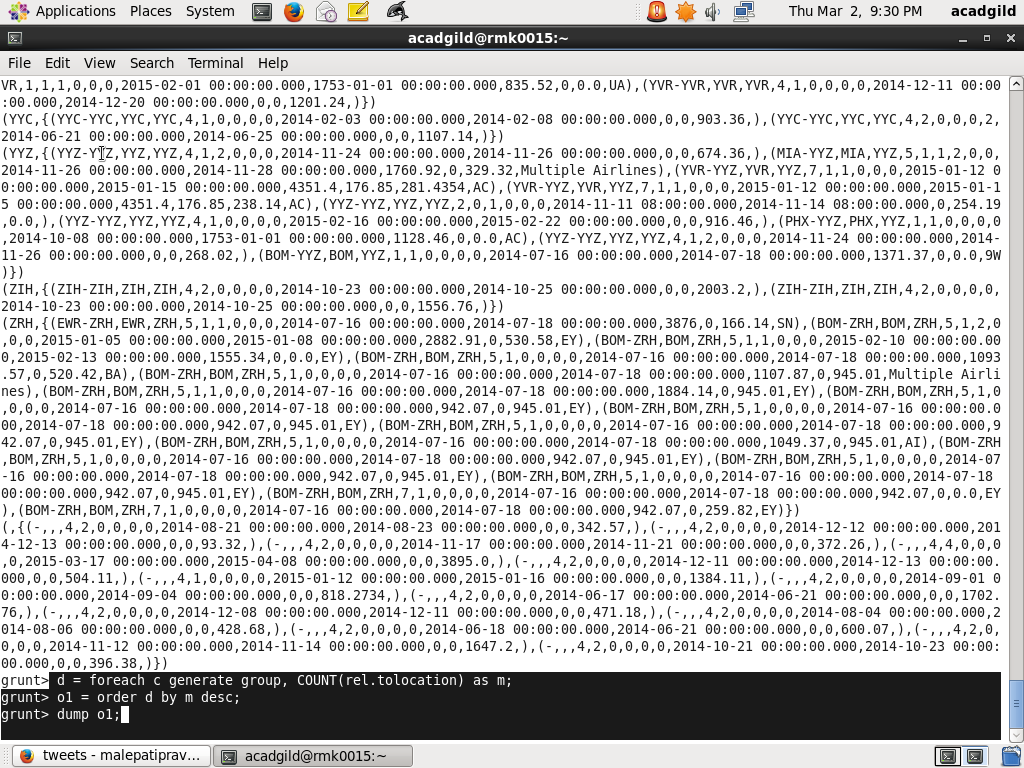
1.LOAD:

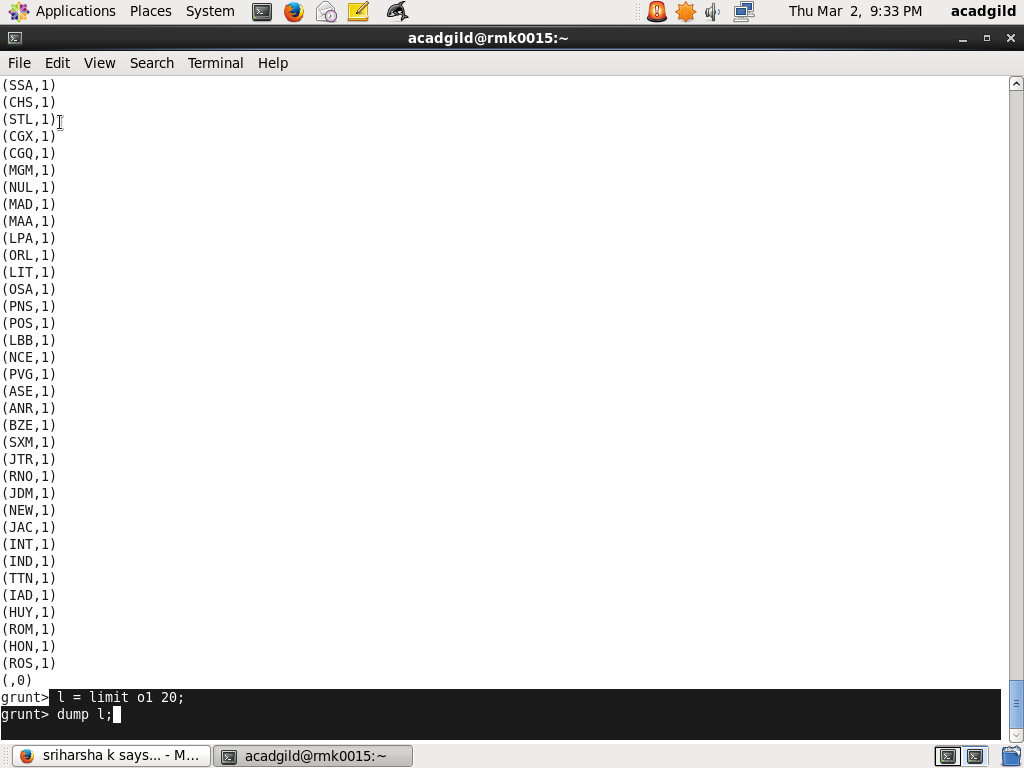
Load the travel set data into the HDFS using the load command.



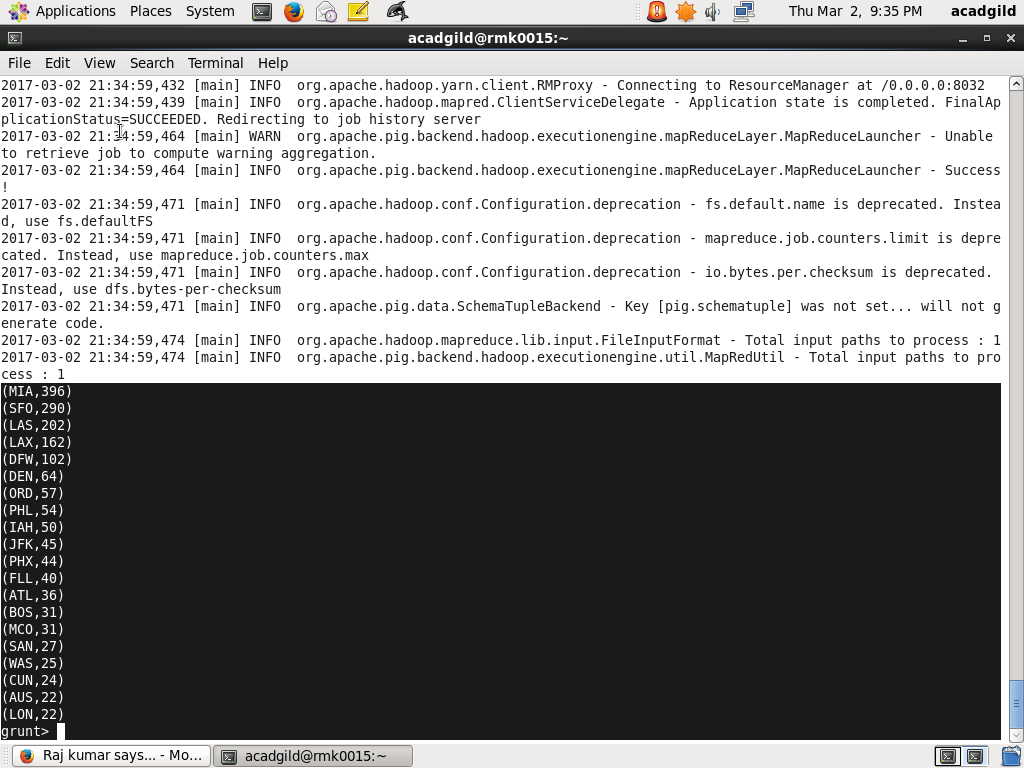
2.Group the relation rel by tolocation.so we get all the same tolocations grouped together.

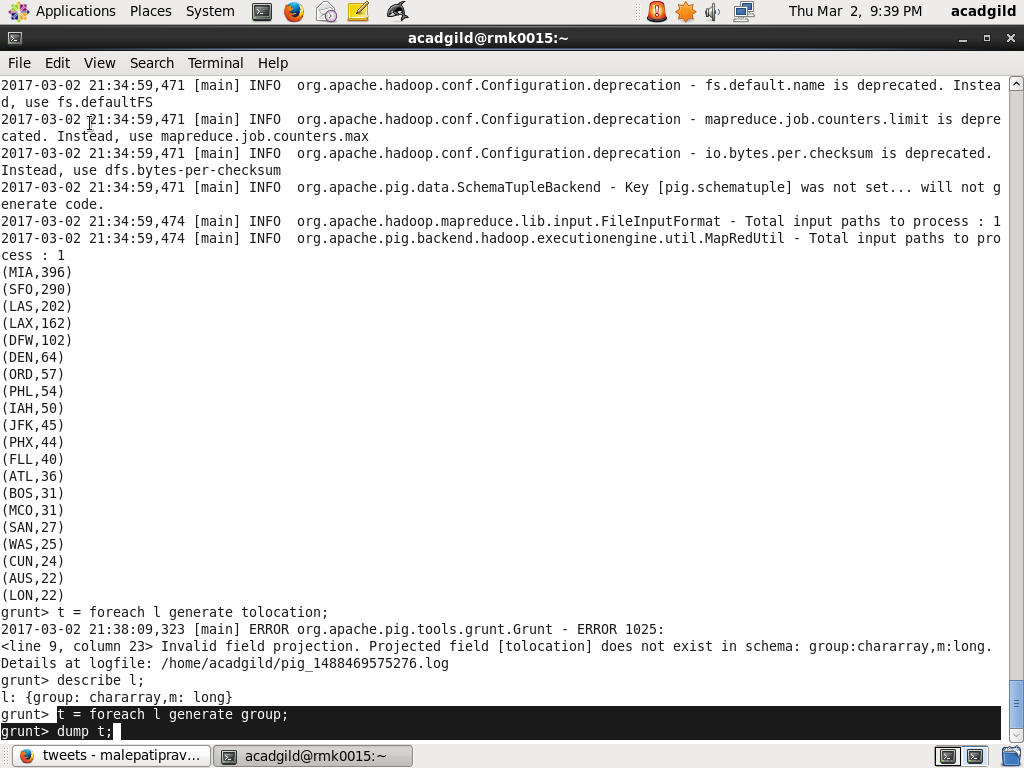


****

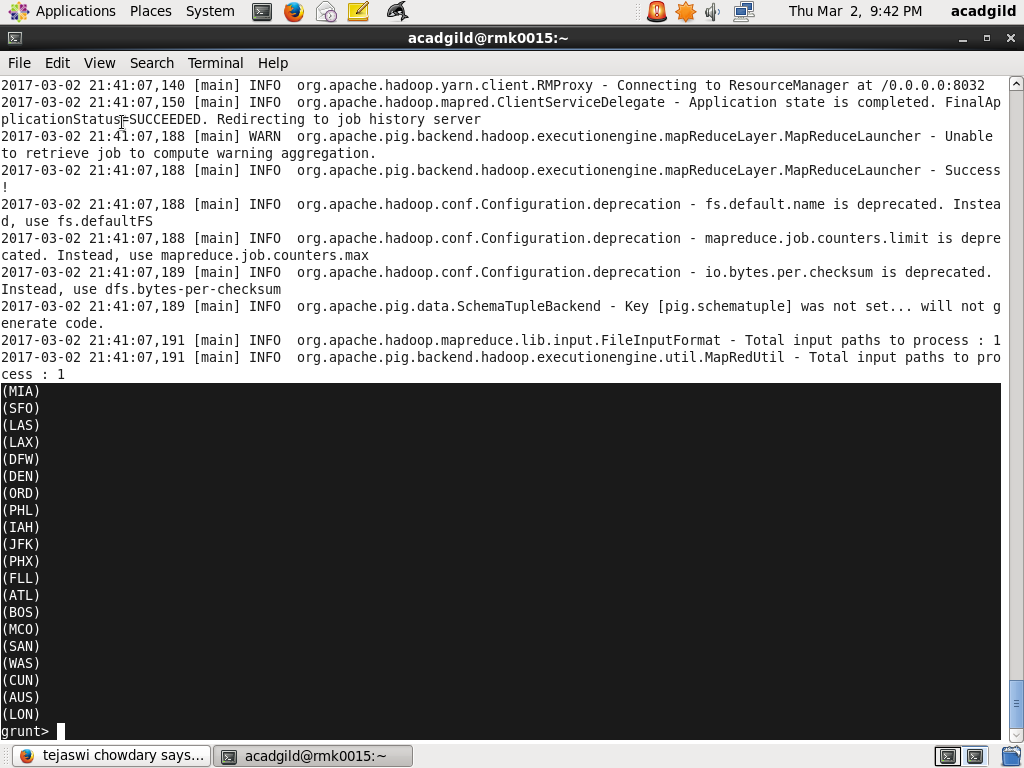


The Top 20 destinations alomg with how many times the place has been travelled.



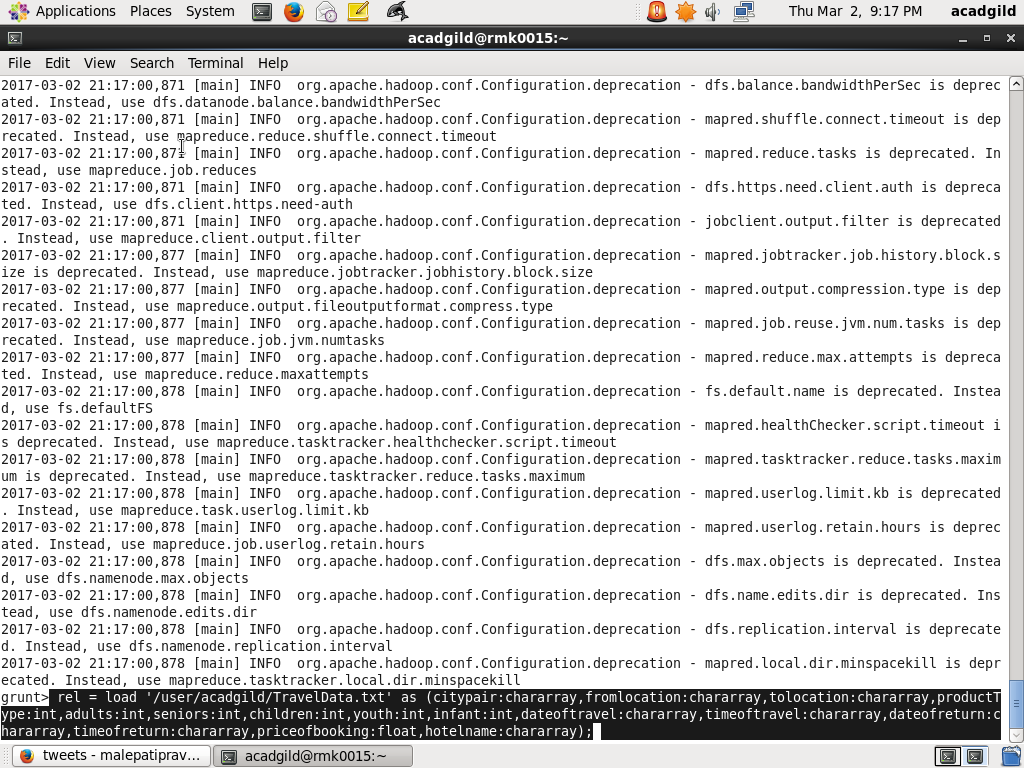


OUTPUT: Top 20 destinations people travel most are:

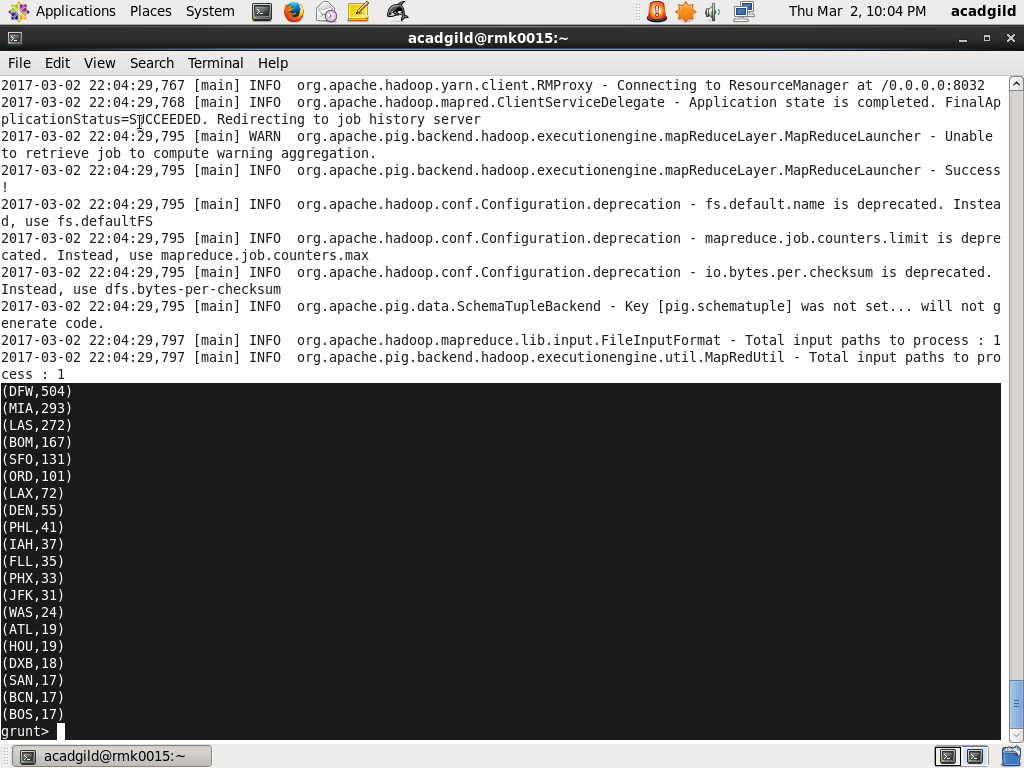
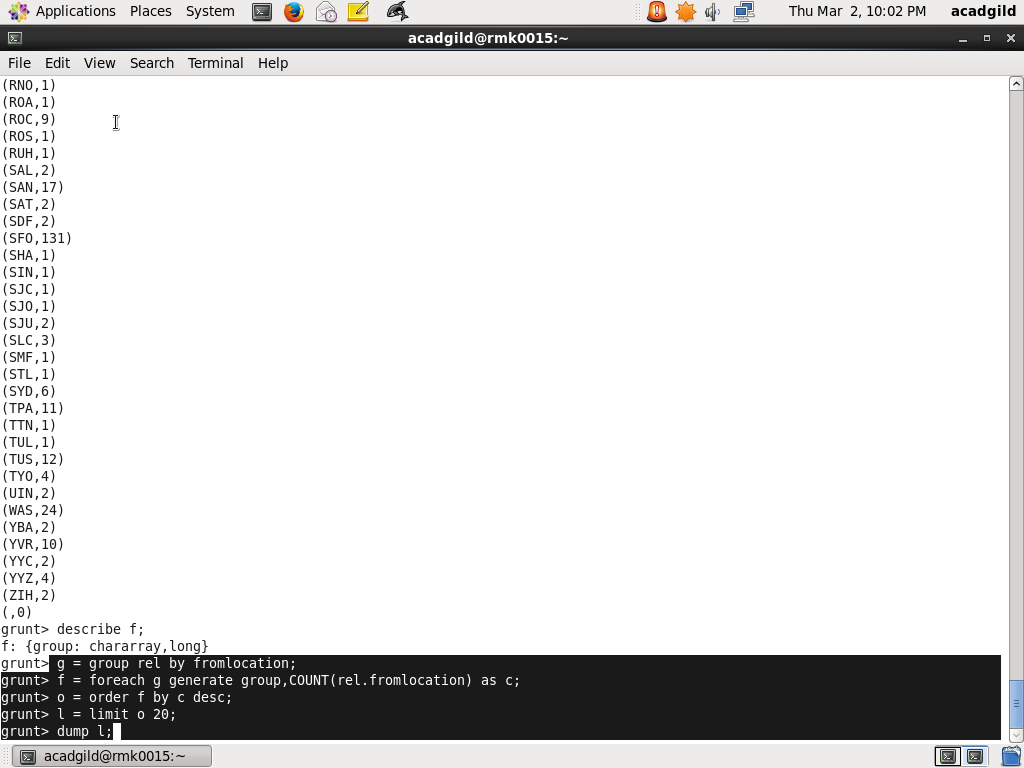


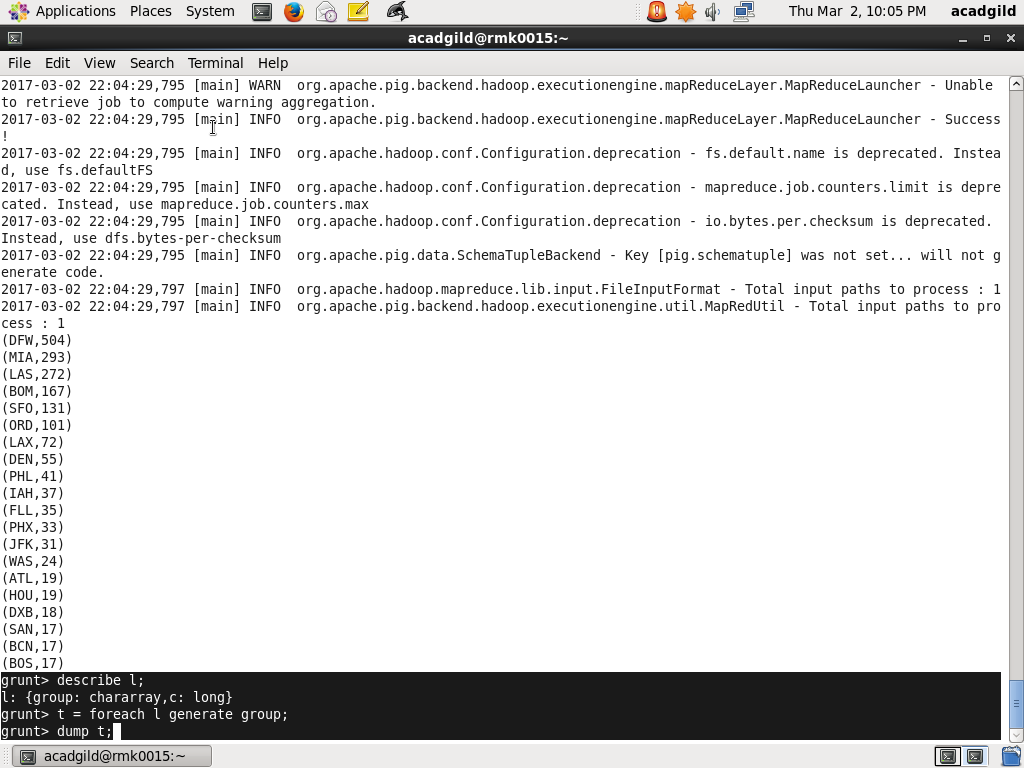
2) The top 20 locations from where the top travel most.

1.Load the travel set into the hdfs directory using the load command.



2.After loading the data first group the relation by tolocation .After grouping generate count foreach group.Order the group by count in the descending order.Using the limit command top 20 data from the table is retrieved.





OUTPUT:

The top 20 loactionf from where the people travelled most are:

