Assignment 10.5

1.Write a pig script to find the top 20 destination people travel the most: Based on the given data

Code :

A = LOAD ‘home/acadgild/Traveldata.txt’ as (citypair:chararray,fromlocation:chararray, tolocation:chararray,producttype:int, AdultsTraveling:int, SeniorsTraveling:int, Childrentraveling:int, YouthTraveling:int, Infanttraveling:int, DateOfTravel:chararray, TimeOfTravel:chararray, DateOfReturn:chararray, TimeOfreturn:chararray, PriceOfBooking:float, Hotelname:chararray);

B = GROUP A by tolocation;

C = FOREACH B generate group,COUNT(A.tolocation) as bt;

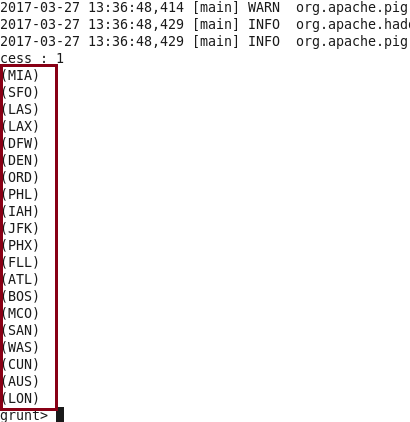
D = ORDER C by bt DESC,group;

E = FOREACH D generate group;

F = LIMIT E 20;

DUMP F;

OUTPUT:



2. Write a pig script to find the top 20 locations from where people travel the most.

CODE :

A = LOAD ‘home/acadgild/Traveldata.txt’ as (citypair:chararray,fromlocation:chararray, tolocation:chararray,producttype:int, AdultsTraveling:int, SeniorsTraveling:int, Childrentraveling:int, YouthTraveling:int, Infanttraveling:int, DateOfTravel:chararray, TimeOfTravel:chararray, DateOfReturn:chararray, TimeOfreturn:chararray, PriceOfBooking:float, Hotelname:chararray);

B = GROUP A by fromlocation;

C = FOREACH B generate group,COUNT(A.fromlocation) as bt;

D = ORDER C by bt DESC,group;

E = FOREACH D generate group;

F = LIMIT E 20;

DUMP F;

OUTPUT :

