**Session 18**

**Assignment 1 Questions**

*Session 18: Assignment 1*

**Table of Contents**

1. Introduction

2. Problem Statement

3. Output

**1. Introduction**

This assignment will help you to consolidate the concepts learnt in the session.

**2. Problem Statement**

1) What is the distribution of the total number of air-travelers per year

2) What is the total air distance covered by each user per year

3) Which user has travelled the largest distance till date

4) What is the most preferred destination for all users.

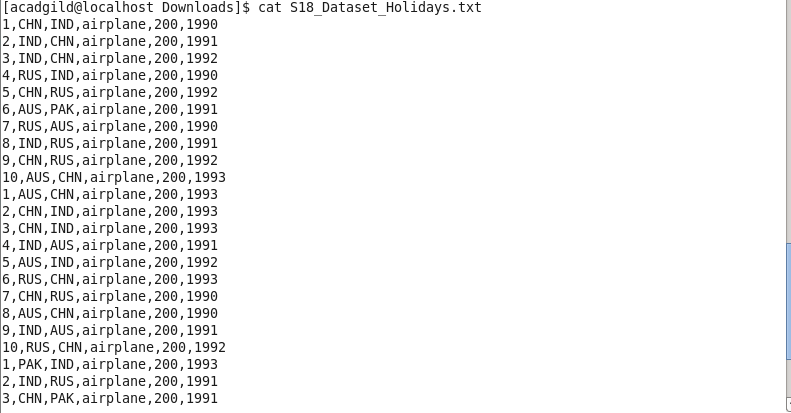
Use below link to download the dataset:

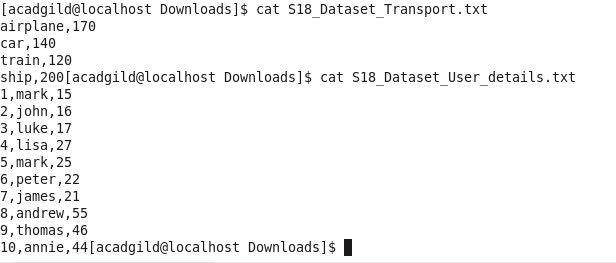
https://drive.google.com/drive/folders/0B\_P3pWagdIrrVThBaUdVSUtzbms

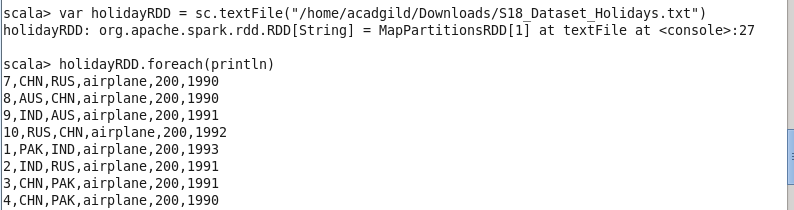
**3. Output**

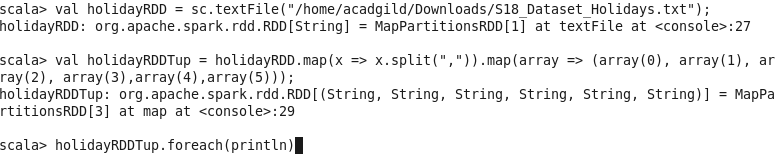
Code and Output screenshot(s).

1) What is the distribution of the total number of air-travelers per year





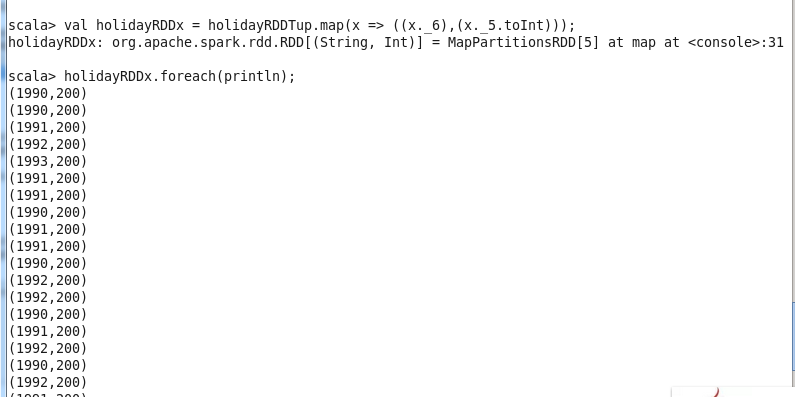




var holidayRDD = sc.textFile("/home/acadgild/Downloads/S18\_Dataset\_Holidays.txt")

val holidayRDDTup = holidayRDD.map(x => x.split(",")).map(array => (array(0), array(1), array(2), array(3),array(4),array(5)))

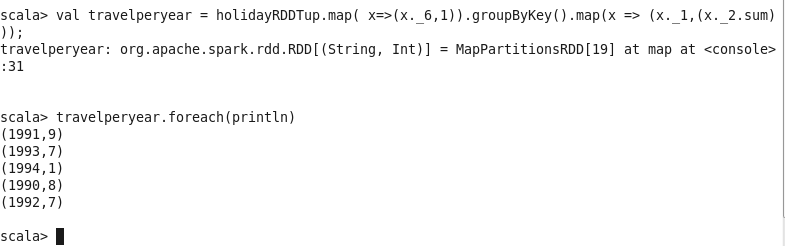
val holidayRDDx = holidayRDDTup.map(x => ((x.\_6),(x.\_5.toInt)))



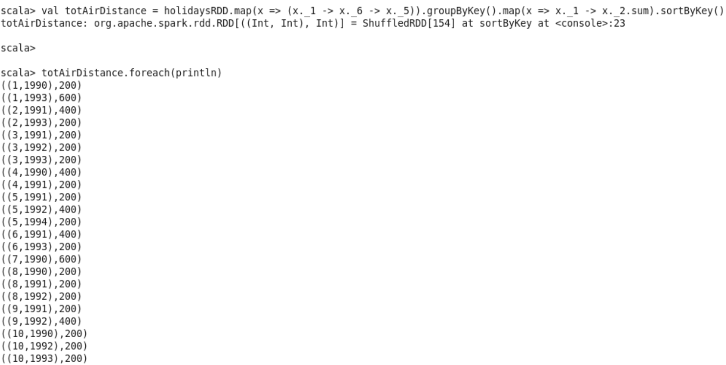
val holidayRDDx = holidayRDDTup.map(x => ((x.\_6),(x.\_5.toInt)))

1) What is the distribution of the total number of air-travelers per year

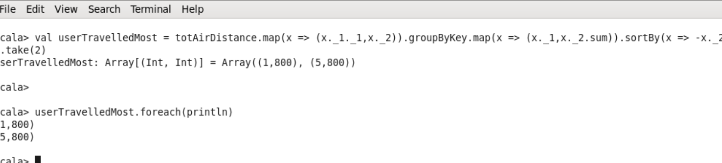
scala> val travelperyear = holidayRDDTup.map( x=>(x.\_6,1)).groupByKey().map(x => (x.\_1,(x.\_2.sum)))



2) What is the total air distance covered by each user per year



3) Which user has travelled the largest distance till date



4) What is the most preferred destination for all users

