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

**case class Holidays(userid:String,source:String,destination:String,transportationmode:String,distance:Int,year:Int)**

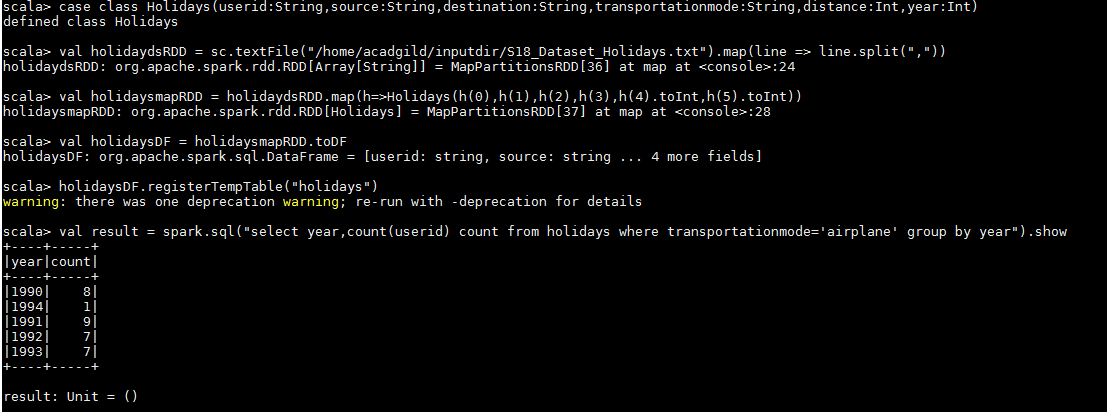
**val holidaydsRDD = sc.textFile("/home/acadgild/inputdir/S18\_Dataset\_Holidays.txt").map(line => line.split(","))**

**val holidaysmapRDD = holidaydsRDD.map(h=>Holidays(h(0),h(1),h(2),h(3),h(4).toInt,h(5).toInt))**

**val holidaysDF = holidaysmapRDD.toDF**

**holidaysDF.registerTempTable("holidays")**

**val result = spark.sql("select year,count(userid) count from holidays where transportationmode='airplane' group by year").show**



1. What is the total air distance covered by each user per year

**case class Holidays(userid:String,source:String,destination:String,transportationmode:String,distance:Int,year:Int)**

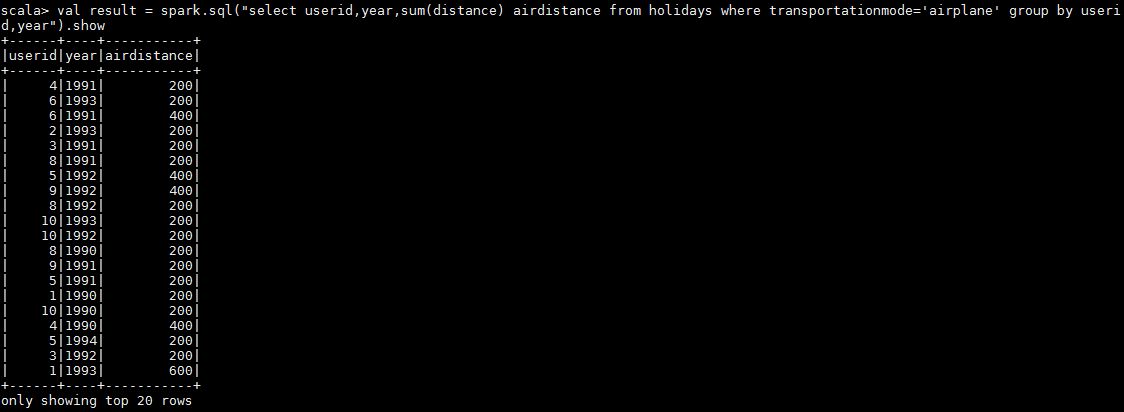
**val holidaydsRDD = sc.textFile("/home/acadgild/inputdir/S18\_Dataset\_Holidays.txt").map(line => line.split(","))**

**val holidaysmapRDD = holidaydsRDD.map(h=>Holidays(h(0),h(1),h(2),h(3),h(4).toInt,h(5).toInt))**

**val holidaysDF = holidaysmapRDD.toDF**

**holidaysDF.registerTempTable("holidays")**

**val result = spark.sql("select userid,year,sum(distance) airdistance from holidays where transportationmode='airplane' group by userid,year").show**



1. Which user has travelled the largest distance till date

**case class Holidays(userid:String,source:String,destination:String,transportationmode:String,distance:Int,year:Int)**

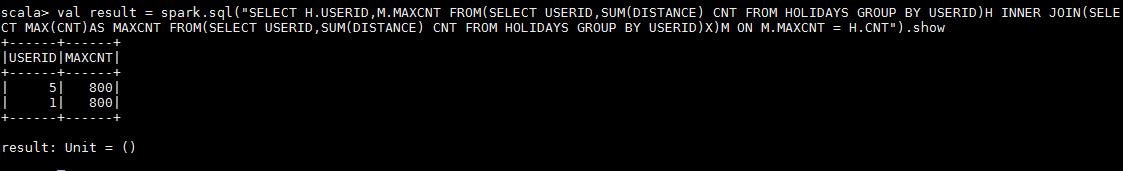
**val holidaydsRDD = sc.textFile("/home/acadgild/inputdir/S18\_Dataset\_Holidays.txt").map(line => line.split(","))**

**val holidaysmapRDD = holidaydsRDD.map(h=>Holidays(h(0),h(1),h(2),h(3),h(4).toInt,h(5).toInt))**

**val holidaysDF = holidaysmapRDD.toDF**

**holidaysDF.registerTempTable("holidays")**

**val result = spark.sql("SELECT H.USERID,M.MAXCNT FROM(SELECT USERID,SUM(DISTANCE) CNT FROM HOLIDAYS GROUP BY USERID)H INNER JOIN(SELECT MAX(CNT)AS MAXCNT FROM(SELECT USERID,SUM(DISTANCE) CNT FROM HOLIDAYS GROUP BY USERID)X)M ON M.MAXCNT = H.CNT").show**



1. What is the most preferred destination for all users.

**case class Holidays(userid:String,source:String,destination:String,transportationmode:String,distance:Int,year:Int)**

**val holidaydsRDD = sc.textFile("/home/acadgild/inputdir/S18\_Dataset\_Holidays.txt").map(line => line.split(","))**

**val holidaysmapRDD = holidaydsRDD.map(h=>Holidays(h(0),h(1),h(2),h(3),h(4).toInt,h(5).toInt))**

**val holidaysDF = holidaysmapRDD.toDF**

**holidaysDF.registerTempTable("holidays")**

**val result = spark.sql("SELECT H.DESTINATION,M.MAXCNT FROM(SELECT DESTINATION,COUNT(DESTINATION) CNT FROM HOLIDAYS GROUP BY DESTINATION)H INNER JOIN(SELECT MAX(CNT)AS MAXCNT FROM(SELECT DESTINATION,COUNT(DESTINATION) CNT FROM HOLIDAYS GROUP BY DESTINATION)X)M ON M.MAXCNT = H.CNT").show**

