val baseRDD = sc.textFile("/home/acadgild/hadoop/S18\_Dataset\_Holidays.txt")

import org.apache.spark.storage.StorageLevel

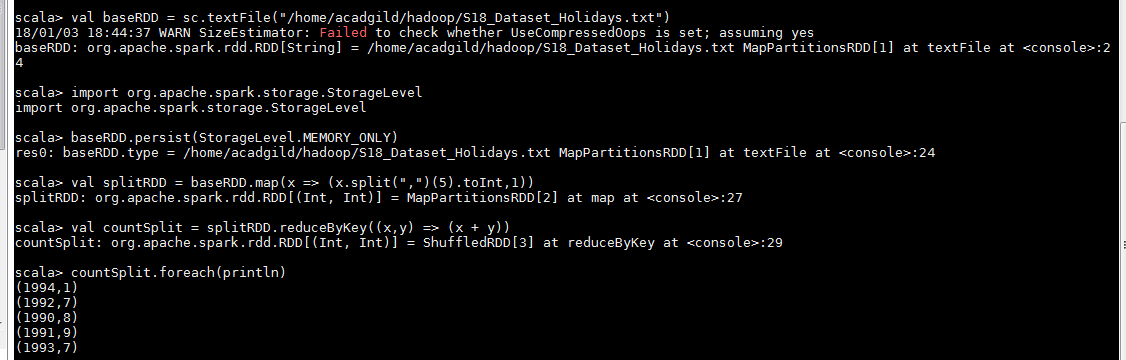
baseRDD.persist(StorageLevel.MEMORY\_ONLY)

1-

val splitRDD = baseRDD.map(x => (x.split(",")(5).toInt,1))

val countSplit = splitRDD.reduceByKey((x,y) => (x + y))

countSplit.foreach(println)

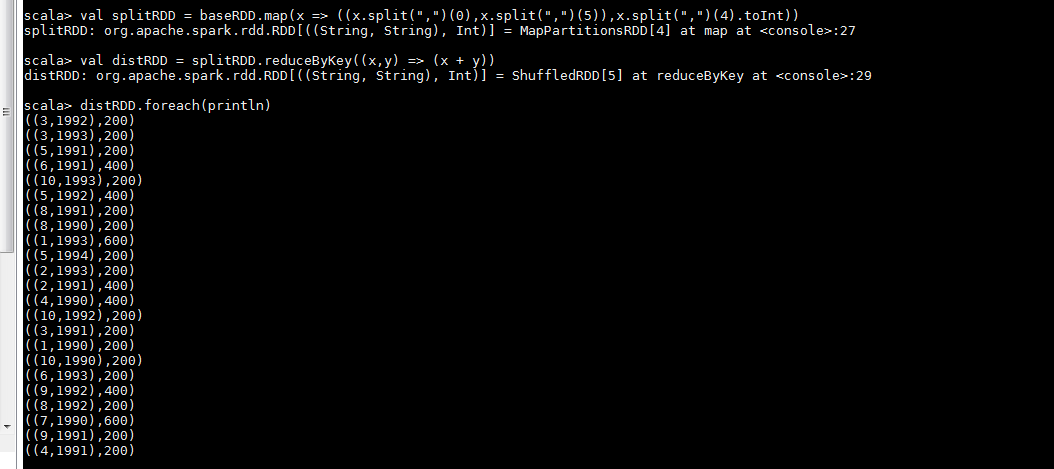


2-

val splitRDD = baseRDD.map(x => ((x.split(",")(0),x.split(",")(5)),x.split(",")(4).toInt))

val distRDD = splitRDD.reduceByKey((x,y) => (x + y))

distRDD.foreach(println)

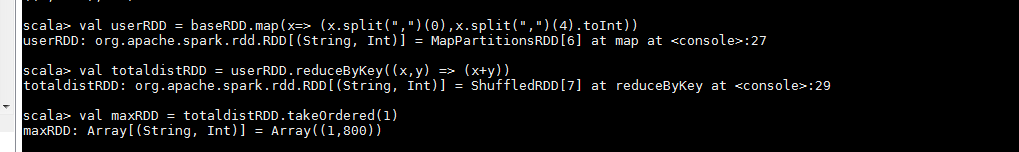


3-

val userRDD = baseRDD.map(x=> (x.split(",")(0),x.split(",")(4).toInt))

val totaldistRDD = userRDD.reduceByKey((x,y) => (x+y))

val maxRDD = totaldistRDD.takeOrdered(1)



4-

val destRDD = baseRDD.map(x => (x.split(",")(2),1))

val destreduceRDD = destRDD.reduceByKey((x,y) => (x + y))

val maxRDD = destreduceRDD.takeOrdered(1)(Ordering[Int].reverse.on(\_.\_2))

