

Execute the lab Program on Scala Shell: Using RDD and FlatMap count how many times each word appears in a file and write out a list of words whose count is strictly greater than 4 using Spark.

Commands:

```
scala> val textFile = sc.textFile("/home/hduser/Desktop/input.txt")
```

```
textFile: org.apache.spark.rdd.RDD[String] = /home/hduser/Desktop/input.txt  
MapPartitionsRDD[1] at textFile at <console>:24
```

```
scala> val counts = textFile.flatMap(line => line.split(" ")).map(word => (word,  
1)).reduceByKey(_ + _)
```

```
counts: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[4] at reduceByKey at  
<console>:25
```

```
scala> import scala.collection.immutable.ListMap
```

```
import scala.collection.immutable.ListMap
```

```
scala> val sorted=ListMap(counts.collect.sortWith(_._2 > _._2):_*)
```

```
sorted: scala.collection.immutable.ListMap[String,Int] = Map(Car -> 5, River -> 5, Deer -> 2,  
Bear -> 2)
```

```
scala> println(sorted)
```

```
Map(Car -> 5, River -> 5, Deer -> 2, Bear -> 2)
```

```
scala> for((k,v)<-sorted)
```

```
| {  
|   if(v>4)  
|   {  
|     print(k+",")  
|     print(v)  
|     println()  
|   }  
| }
```

Car,5

River,5

Output:

```
scala> val textFile = sc.textFile("/home/hduser/Desktop/input.txt")  
textFile: org.apache.spark.rdd.RDD[String] = /home/hduser/Desktop/input.txt MapPartitionsRDD[1] at textFile at <console>:24  
  
scala> val counts = textFile.flatMap(line => line.split(" ")).map(word => (word, 1)).reduceByKey(_ + _)  
counts: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[4] at reduceByKey at <console>:25  
  
scala> import scala.collection.immutable.ListMap  
import scala.collection.immutable.ListMap  
  
scala> val sorted=ListMap(counts.collect.sortWith(_._2 > _._2):_*)  
sorted: scala.collection.immutable.ListMap[String,Int] = Map(Car -> 5, River -> 5, Deer -> 2, Bear -> 2)  
  
scala> println(sorted)  
Map(Car -> 5, River -> 5, Deer -> 2, Bear -> 2)  
  
scala> for((k,v)<-sorted)  
| {  
|   if(v>4)  
|   {  
|     print(k+",")  
|     print(v)  
|     println()  
|   }  
| }  
Car,5  
River,5  
  
scala> □
```