

Assignment no. 13

Code:

```
scala> val input = sc.textFile("passage.txt")
```

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

```
scala> val words = input.flatMap(x => x.split(" "))
```

```
words: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at flatMap at <console>:25
```

```
scala> val counts = words.map(x => (x, 1))
```

```
counts: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[3] at map at <console>:26
```

```
scala> val reducedCounts = counts.reduceByKey((x, y) => x + y)
```

```
reducedCounts: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[4] at reduceByKey at <console>:27
```

```
scala> reducedCounts.saveAsTextFile("output.txt")
```

```
scala> reducedCounts.foreach(println)
```

Output:

```
$ spark-shell
```

```
(spark,1)
```

```
(world,1)
```

```
(hello,2)
```