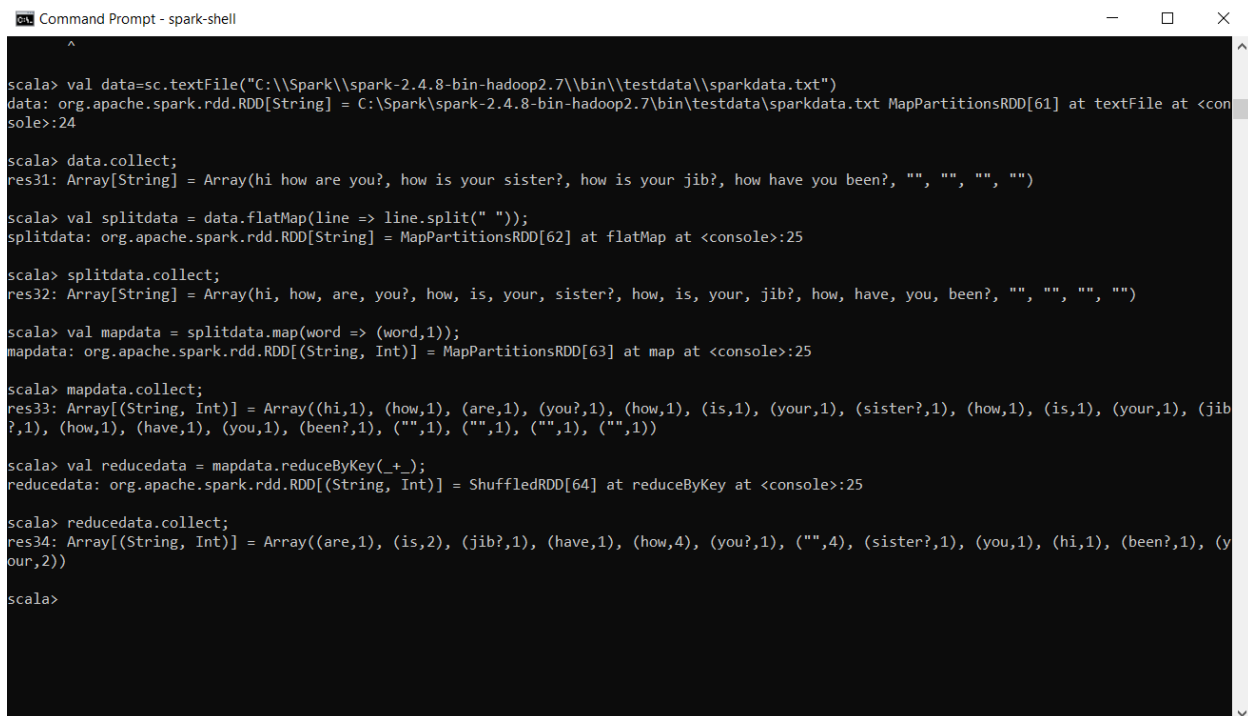
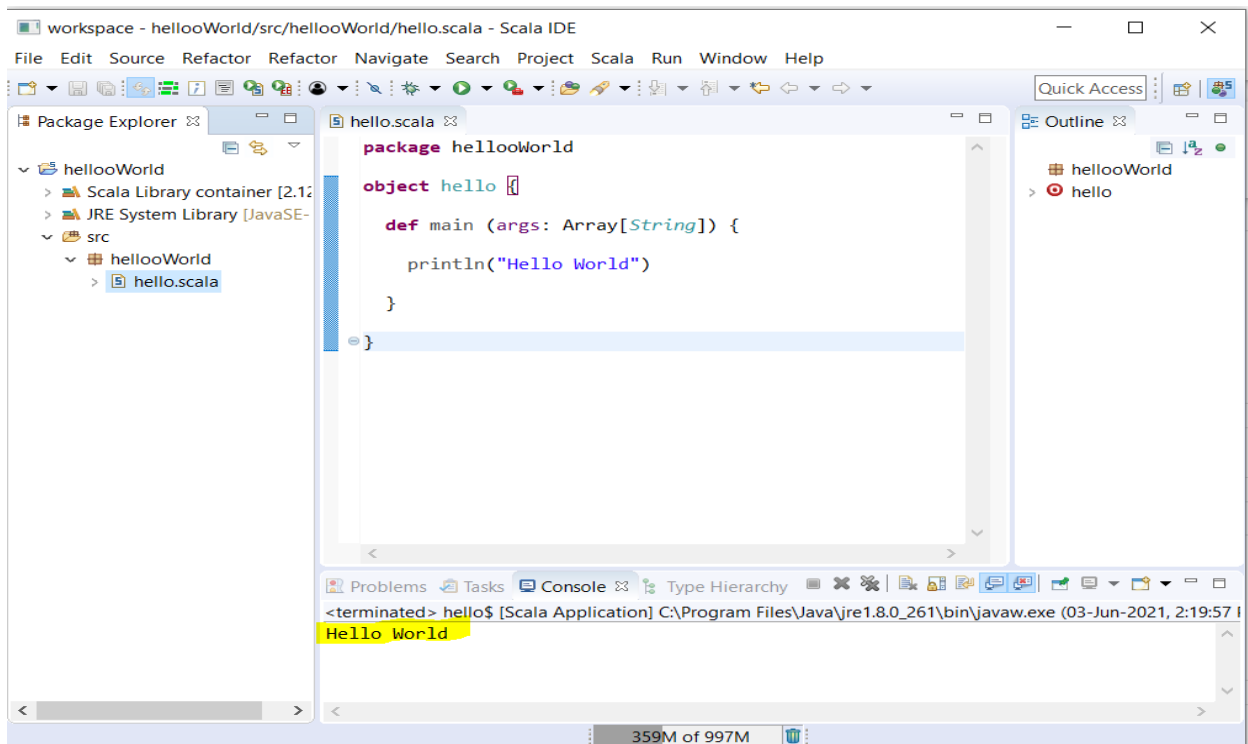


# WEEK-08



```
Command Prompt - spark-shell
scala> val textFile = sc.textFile("C:\\Spark\\spark-2.4.8-bin-hadoop2.7\\bin\\testdata\\sparkdata.txt")
textFile: org.apache.spark.rdd.RDD[String] = C:\\Spark\\spark-2.4.8-bin-hadoop2.7\\bin\\testdata\\sparkdata.txt MapPartitionsRDD[75] at textFile at
<console>:31

scala> val counts = textFile.flatMap(line => line.split(" ")).map(word => (word, 1)).reduceByKey(_ + _)
counts: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[78] at reduceByKey at <console>:32

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

scala> val sorted=ListMap(counts.collect.sortWith(_._2 > _. _2):_*)// sort in descending order based on values
sorted: scala.collection.immutable.ListMap[String,Int] = Map(how -> 5, "" -> 4, is -> 2, your -> 2, are -> 1, can -> 1, have -> 1, you? -> 1, j
ob? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, i -> 1, been? -> 1)

scala> println(sorted)
Map(how -> 5,  -> 4, is -> 2, your -> 2, are -> 1, can -> 1, have -> 1, you? -> 1, job? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, i ->
1, been? -> 1)

scala> for((k,v)<-sorted)
| {
|   | if(v>4)
|   | {
|   |   | print(k+",")
|   |   | print(v)
|   |   | println()
|   | }
| }
how,5

scala>
```