## WEEK-08

```
workspace - hellooWorld/src/hellooWorld/hello.scala - Scala IDE
 File Edit Source Refactor Refactor Navigate Search Project Scala Run Window Help
 □ ▼ 🖫 📵 😘 🚅 📝 🗑 😘 🕦 (② ▼ │ ¾ ▼ ○ ▼ 💁 ▼ │ 炒 ▼ │ ∮ ▼ │ │ ▼ │ │ ▼ │ │ ▼ │ │ ▼ │ │ ▼ │ ○ ▼ □ ○ ▼
                                                                                                                                                                                                                                             Quick Access 🔡 😭 🛮 🐉
  ■ Package Explorer 🖾 📅 🗖 🖺 hello.scala 🖾
                                                                                                                                                                                                                          _ _
                                                                                                                                                                                                                                         B Outline ⊠
                                                 = $
                                                                \nabla
                                                                                 package hellooWorld
  object hello {
       > Marcon Scala Library container [2.12
                                                                                                                                                                                                                                           > • hello
        > A JRE System Library [JavaSE-
                                                                                      def main (args: Array[String]) {
       println("Hello World")
                  > 🗈 hello.scala
                                                                         🖭 Problems 🖉 Tasks 🖳 Console 🗯 Type Hierarchy 🔳 🗶 🐉 🚇 🚱 🔛 🖭 🗗 🖵 🗂 🔻 🗀
                                                                          termin<mark>ated> hell</mark>o$ [Scala Application] C:\Program Files\Java\jre1.8.0_261\bin\javaw.exe (03-Jun-2021, 2:19:57 |
                                                                         Hello World
                                                                                                                                                359M of 997M
                                                                                                                                                                                   Ŵ
 Command Prompt - spark-shell
                                                                                                                                                                                                                                                                         X
 scala> val data=sc.textFile("C:\\Spark\\spark-2.4.8-bin-hadoop2.7\\bin\\testdata\\sparkdata.txt")
data: org.apache.spark.rdd.RDD[String] = C:\Spark\spark-2.4.8-bin-hadoop2.7\bin\testdata\sparkdata.txt MapPartitionsRDD[61] at textFile at <con
sole>:24
scala> data.collect;
res31: Array[String] = Array(hi how are you?, how is your sister?, how is your jib?, how have you been?, "", "", "", "")
scala> val splitdata = data.flatMap(line => line.split(" "));
splitdata: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[62] at flatMap at <console>:25
scala> splitdata.collect;
 res32: Array[String] = Array(hi, how, are, you?, how, is, your, sister?, how, is, your, jib?, how, have, you, been?, "", "", "")
scala> val mapdata = splitdata.map(word => (word,1));
mapdata: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[63] at map at <console>:25
scala> mapdata.collect;
Scalar mapuata.collect, respectively. Scalar mapuata.collect, sister (how,1), (how,1), (how,1), (is,1), (your,1), (sister (how,1), (how,1), (is,1), (your,1), (jib (how,1), (h
scala> val reducedata = mapdata.reduceByKey(_+_);
reducedata: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[64] at reduceByKey at <console>:25
scala> reducedata.collect;
 res34: Array[(String, Int)] = Array((are,1), (is,2), (jib?,1), (have,1), (how,4), (you?,1), ("",4), (sister?,1), (you,1), (hi,1), (been?,1), (y
```

```
scalary val textFile = sc.textFile("C:\\Spark\\spark-2.4.8-bin-hadoop2.7\\bin\\textBialar\\sparkdata.txt")
textFile: org.apache.spark.rdd.RDD[String] = C:\Spark\\spark-2.4.8-bin-hadoop2.7\\bin\\textBialar\\sparkdata.txt MapPartitionsRDD[75] at textFile at
cconsoler:31
scalary 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 cconsoler:32
scalary import scala.collection.immutable.ListMap
import scala.collection.immutable.ListMap
import scala.collection.immutable.ListMap
import scala.collection.immutable.ListMap[String,Int] = Map(how -> 5, "" -> 4, is -> 2, your -> 2, are -> 1, can -> 1, have -> 1, you? -> 1, job? -> 1, help? -> 1, sister? -> 1, you? -> 1, job? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, help? -> 1, sister? -> 1, you -> 1, hi -> 1, job? -> 1, help? -> 1, help
```