WEEK 7 BDA LAB
USN:1BM19CS109
NAME: P PREM SAI

Spark shell word count demo

```
scala> val test=sc.textFile("/home/hdoop/spark_word_count.txt")
test: org.apache.spark.rdd.RDD[String] = /home/hdoop/spark_word_count.txt Map
rtitionsRDD[1] at textFile at <console>:23

scala> test.collect;
[Stage 0:>                                                          (0 + 0) /
[Stage 0:>                                                          (0 + 2) /

res4: Array[String] = Array(This is a test, This is an evaluation, do you wan
a test, why do you want a test)

scala> val count=test.flatMap(line=>line.split(" "))
count: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at flatMap at <
nsole>:23

scala> count.collect
res5: Array[String] = Array(This, is, a, test, This, is, an, evaluation, do,
u, want, a, test, why, do, you, want, a, test)

scala> val map_frequency=count.map(entry=>(entry,1))
map_frequency: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[3]
 map at <console>:23

scala> map_frequency.collect
res6: Array[(String, Int)] = Array((This,1), (is,1), (a,1), (test,1), (This,1
 (is,1), (an,1), (evaluation,1), (do,1), (you,1), (want,1), (a,1), (test,1),
hy,1), (do,1), (you,1), (want,1), (a,1), (test,1))
```

```
scala> map_frequency.reduceByKey(_+_)
res7: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[4] at reduceByKey
t <console>:24

scala> map_frequency.collect
res8: Array[(String, Int)] = Array((This,1), (is,1), (a,1), (test,1), (This,1
 (is,1), (an,1), (evaluation,1), (do,1), (you,1), (want,1), (a,1), (test,1),
hy,1), (do,1), (you,1), (want,1), (a,1), (test,1))

scala>

scala> val final_output=map_frequency.reduceByKey(_+_)
final_output: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[5] at red
eByKey at <console>:23

scala> final_output.collect
[Stage 4:>                                                          (0 + 2) /

res9: Array[(String, Int)] = Array((is,2), (evaluation,1), (This,2), (why,1),
want,2), (test,3), (you,2), (a,3), (do,2), (an,1))
```