Assignment 17.1

1. **Write a program to read a text file and print the number of rows of data in the document.**

```
scala> val rdd = sc.textFile("/home/acadgild/sample.txt")
rdd: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[5] at textFile at <cons
ole>:27

scala> rdd.count()
res4: Long = 3
```

2. Write a program to read a text file and print the number of words in the document.

```
scala> val wordcount = rdd.flatMap(line => line.split("-")).map(word => (word,1)
)
wordcount: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[7] at map
at <console>:29

scala> wordcount.collect
res5: Array[(String, Int)] = Array((This,1), (is,1), (my,1), (first,1), (assignm
ent.,1), (It,1), (will,1), (count,1), (the,1), (number,1), (of,1), (lines,1), (i
n,1), (this,1), (document.,1), (The,1), (total,1), (number,1), (of,1), (lines,1)
, (is,1), (3,1))
```

3. We have a document where the word separator is -, instead of space. Write a spark code, to obtain the count of the total number of words present in the document.

```
scala> val wordcount = rdd.flatMap(line => line.split("-")).map(word => (word,1)
)
wordcount: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[17] at map
 at <console>:29

scala> wordcount.count()
res12: Long = 22

scala>
```