Create a file having below sample document data:

**This-is-my-first-assignment.**
**It-will-count-the-number-of-lines-in-this-document.**
**The-total-number-of-lines-is-3**

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

   **val docRDD = sc.textFile("/home/acadgild/inputdir/doc.txt")**

   **docRDD.count**

```
scala> val docRDD = sc.textFile("/home/acadgild/inputdir/doc.txt")
docRDD: org.apache.spark.rdd.RDD[String] = /home/acadgild/inputdir/doc.txt MapPartitionsRDD[24] at textFile at <console>:24

scala> docRDD.count
res20: Long = 3

scala>
```

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

   **val empRDD = sc.textFile("/home/acadgild/inputdir/doc.txt")**

   **val splitRDD = empRDD.flatMap(x=>x.split(","))**

   **splitRDD.count**

```
scala> val empRDD = sc.textFile("/home/acadgild/inputdir/emp.txt")
empRDD: org.apache.spark.rdd.RDD[String] = /home/acadgild/inputdir/emp.txt MapPartitionsRDD[36] at textFile at <console>:24

scala> val splitRDD = empRDD.flatMap(x=>x.split(","))
splitRDD: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[37] at flatMap at <console>:26

scala> splitRDD.count
res26: Long = 12

scala>
```

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.

**val docRDD = sc.textFile("/home/acadgild/inputdir/doc.txt")**
**val splitRDD = docRDD.flatMap(x=>x.split("-"))**
**splitRDD.count**

```
scala> val docRDD = sc.textFile("/home/acadgild/inputdir/doc.txt")
docRDD: org.apache.spark.rdd.RDD[String] = /home/acadgild/inputdir/doc.txt MapPartitionsRDD[39] at textFile at <console>:24

scala> val splitRDD = docRDD.flatMap(x=>x.split("-"))
splitRDD: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[40] at flatMap at <console>:26

scala> splitRDD.count
res27: Long = 22

scala>
```