

## Assignement 19

Write a pgm to read a text file and print the number of rows.

use the command `val textFileLocalTest =  
sc.textFile("file://19_Dataset.txt.Docx");`

To find the no of rows use the command `textFileFromHDFS.count`

number of rows is 3

write a pgm to print words in a document.

use the commads as below-

```
scala> inputRDD.collect()
16/08/11 13:28:15 WARN LoadSnappy: Snappy native library not loaded
res0: Array[String] = Array(This is sample data to show word count program in Apache Spark in scala shell., And word count can also be done in p
ython shell.)

scala> val inputWords = inputRDD.flatMap(x => x.split(" "))
inputWords: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at flatMap at <console>:29

scala> for( x<-inputWordCount.collect()){
  | println (x)
  | }
(= 3 = 3)
```

when the word seperator is '-' instead of space

use the command

```
val res = data.flatMap(a => a.split("-") ).map(w=> (w,1)).cache()
```

now counting the words by using the command -

```
res.reduceBy( _ + _ ).take (10)
```