ACD\_BDD2.3\_Session\_17\_Assignment\_1

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

val data= SparkContext.textFile(“inputfile.txt”)

data.count()

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

Program

val data= SparkContext.textFile(“inputfile.txt”)

val words = data.flatMap(\_.split(" "))

val wordCounts = words.map(x => (x, 1)).reduceByKey(\_ + \_)

wordCounts.print()

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.

Sample document :

This-is-my-first-assignment.

It-will-count-the-number-of-lines-in-this-document.

The-total-number-of-lines-is-3

Program:-

val data= SparkContext.textFile(“inputfile.txt”)

val words = data.flatMap(\_.split("-"))

val wordCounts = words.map(x => (x, 1)).reduceByKey(\_ + \_)

wordCounts.print()