**Assignment 17.1**

**Problem Statement**

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

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

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 :

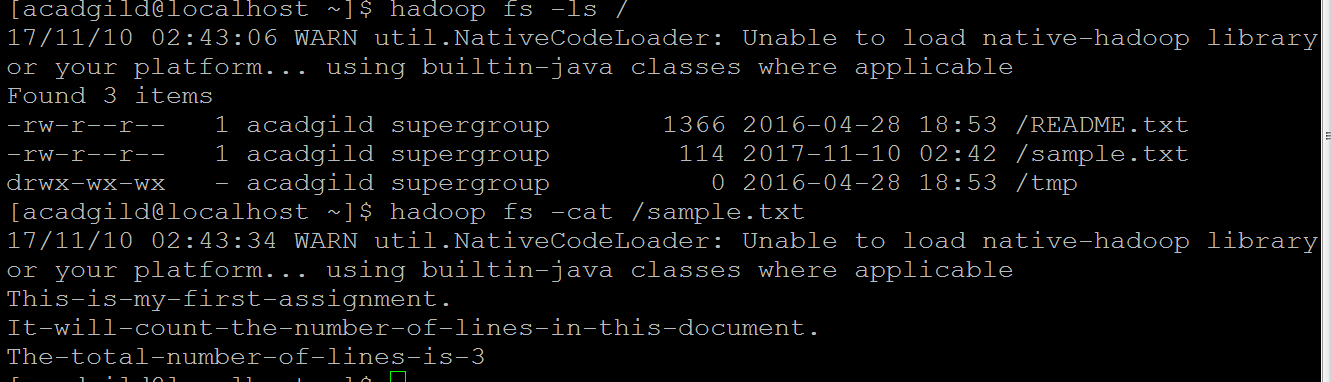
This-is-my-first-assignment.

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

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

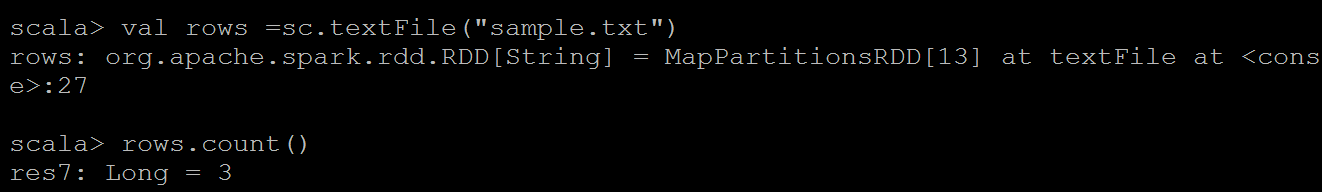
Solution:

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

****

scala>val rows =sc.textFile("sample.txt")

rows.count()

****

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

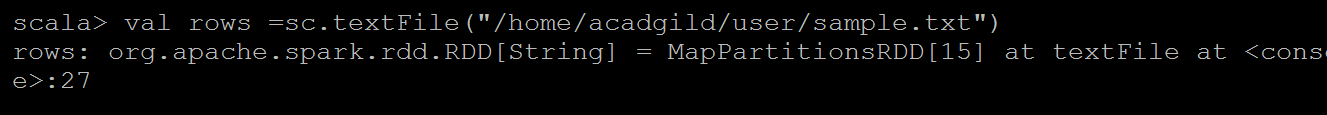
val rows =sc.textFile("*("/home/acadgild/user/*sample.txt")

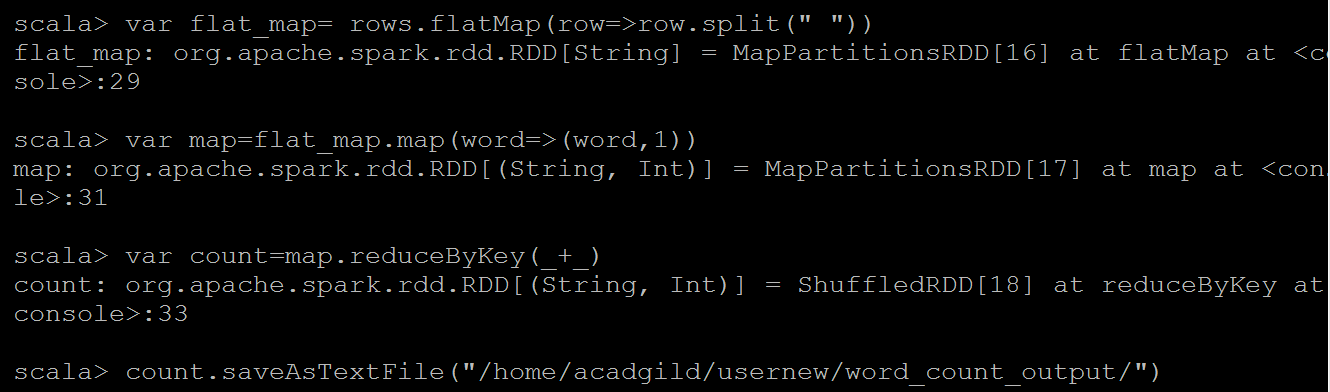
var flat\_map= rows.flatMap(row=>row.split(" "))

*var map=flat\_map.map(word=>(word,1))*

*var count=map.reduceByKey(\_+\_)*

*count.saveAsTextFile("/home/acadgild/usernew/word\_count\_output/")*

****

****

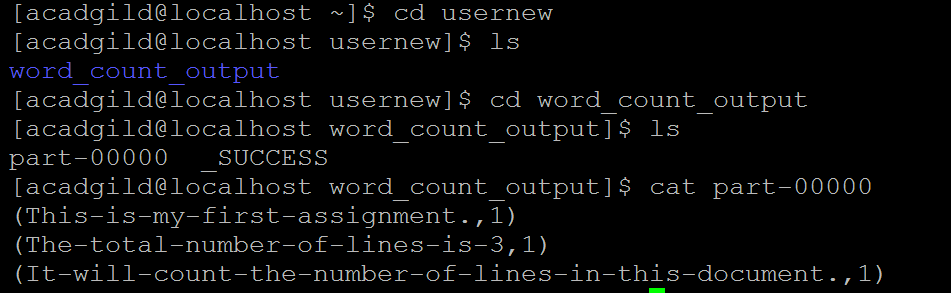
cd usernew

ls

cd *word\_count\_output*

*ls*

*cat part-00000*

****

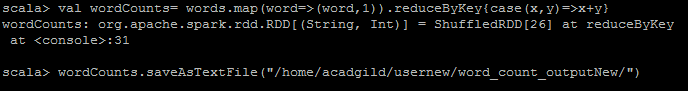
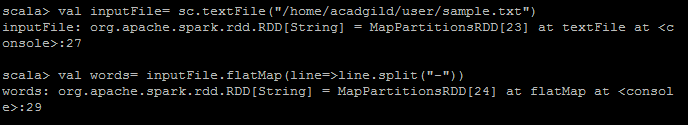
3. We have a document where the word separator is -, instead of space. Write a spark code, to obtain the count of the total number of words present in the document.

val inputFile= sc.textFile("/home/acadgild/user/sample.txt")

val words= inputFile.flatMap(line=>line.split("-"))

val wordCounts= words.map(word=>(word,1)).reduceByKey{case(x,y)=>x+y}

wordCounts.saveAsTextFile("/home/acadgild/usernew/word\_count\_outputNew/")

****

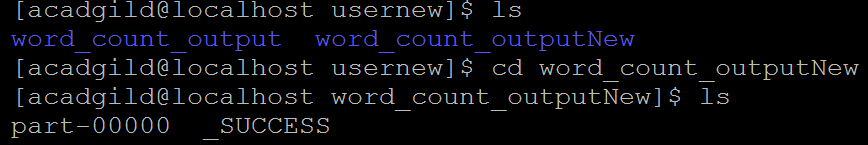
cd usernew

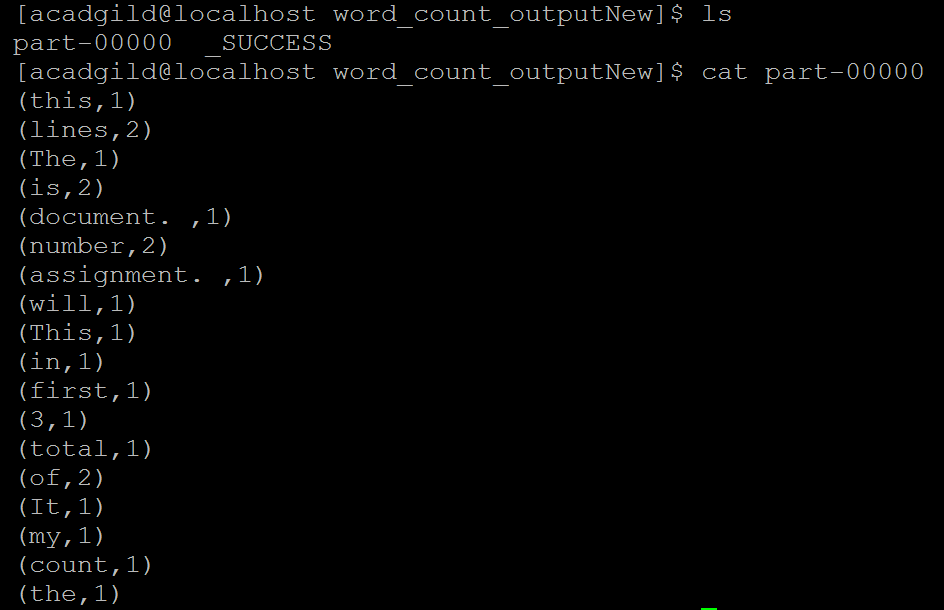
ls

cd word\_count\_outputNew

ls

cat part-00000

****

****