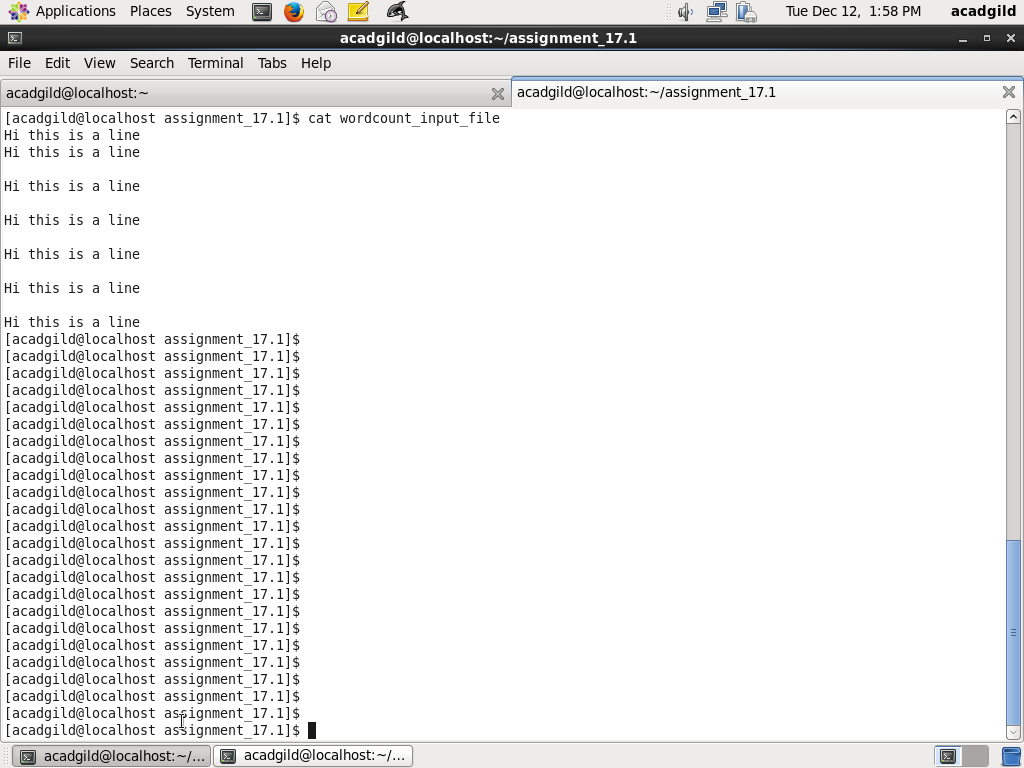
Task1: Write a program to read a text file and count number of rows

Step1:

First copy the wordcount\_input\_file. Contents of file is displayed below



Step2:

Create a RDD using the text file. Count number of lines in rdd and put to line\_count. Print line\_count.

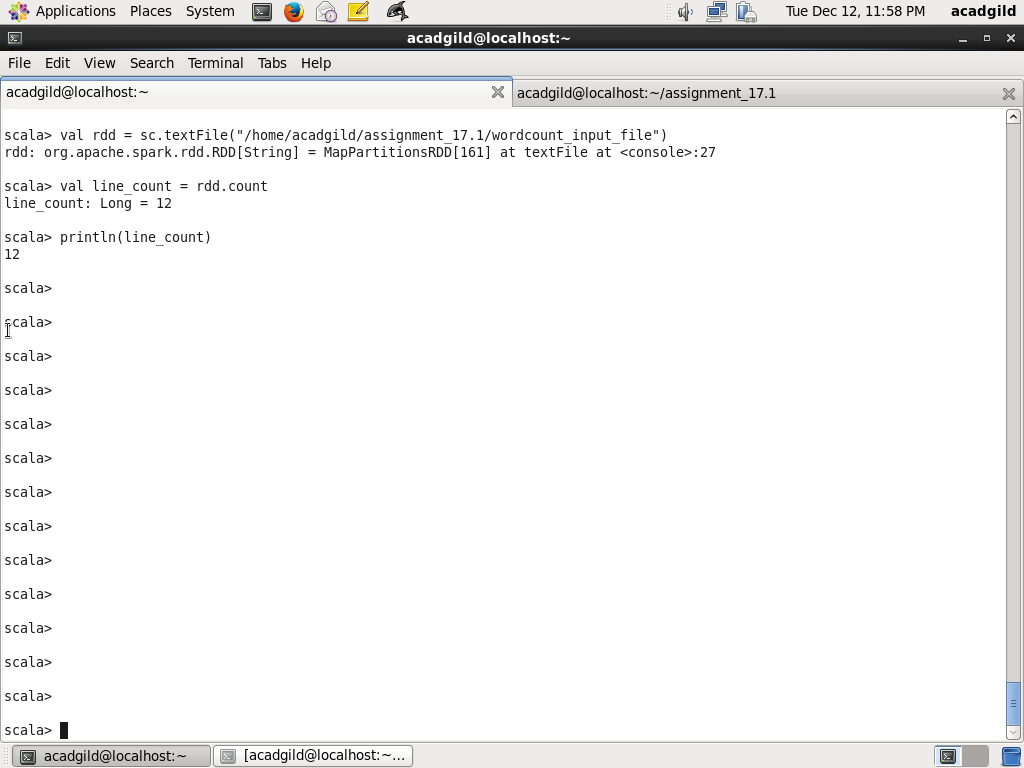
The code is as below:

val rdd = sc.textFile("/home/acadgild/assignment\_17.1/wordcount\_input\_file")

val line\_count = rdd.count

println(line\_count)

Screenshot is as below:



Task2 : Write a program to read a text file and count number of words

Create a rdd from the input file wordcount\_input\_file. Create a rdd\_words by splitting line based on blank space and using flatMap on rdd. Count the number of words using count method and assign to word\_count. Print word\_count

Code is as below:

val rdd = sc.textFile("/home/acadgild/assignment\_17.1/wordcount\_input\_file")

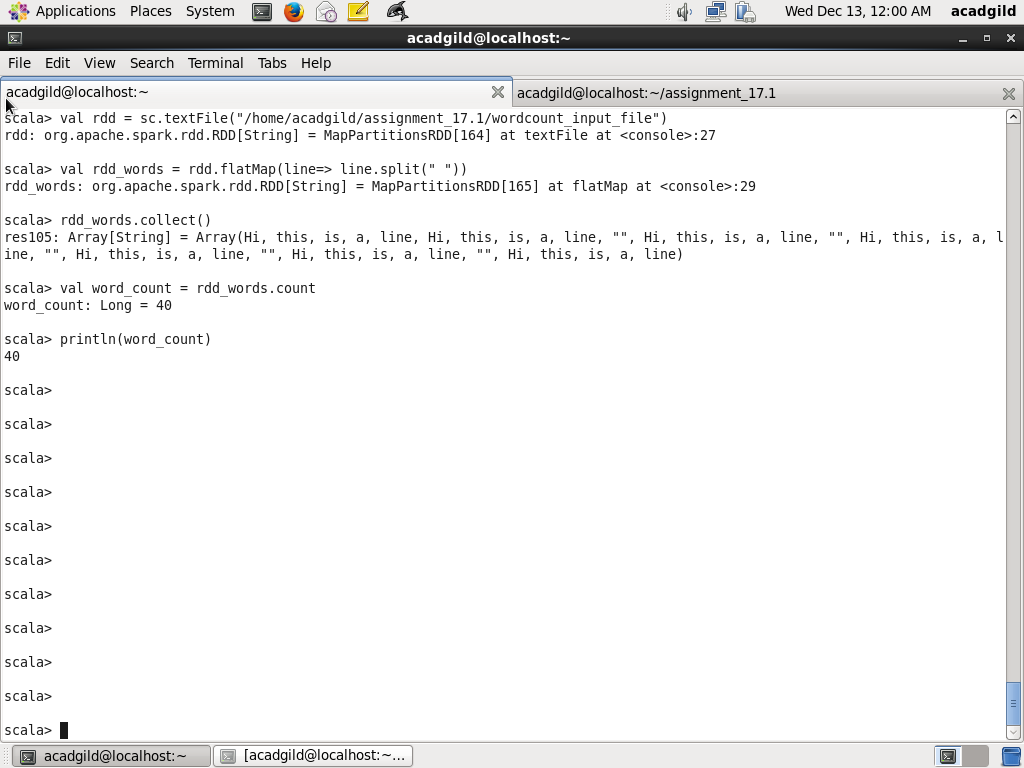
val rdd\_words = rdd.flatMap(line=> line.split(" "))

rdd\_words.collect()

val word\_count = rdd\_words.count

println(word\_count)

Screenshot is as below:

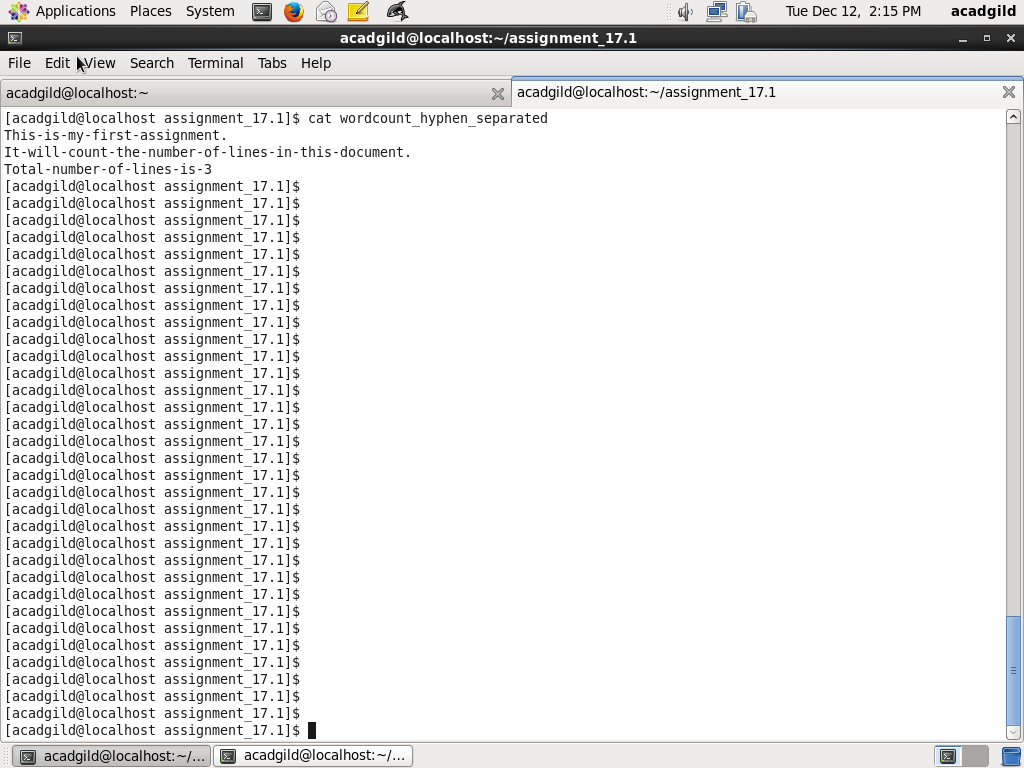


Task3: Write a program to read a text and count number of words when separator is hyphen (-)

Step1:

Copy the file wordcount\_hyphen\_separated to /home/acadgild/assignment\_17.1

Contents are as below:



Step2:

Create a RDD rdd\_hyphen from the wordcount\_hyphen\_separated file. Split each line based on separator hyphen (-) and use flatMap to combine the results. Use count method to get the count of works and assign to word\_hyphen\_separator\_count and print word\_hyphen\_separator\_count

Code is as below:

val rdd\_hyphen = sc.textFile("/home/acadgild/assignment\_17.1/wordcount\_hyphen\_separated")

val rdd\_hyphen\_words = rdd\_hyphen.flatMap(line=> line.split("-"))

rdd\_hyphen\_words.collect

val word\_hyphen\_separator\_count = rdd\_hyphen\_words.count

println(word\_hyphen\_separator\_count)

Screenshot is as below:

