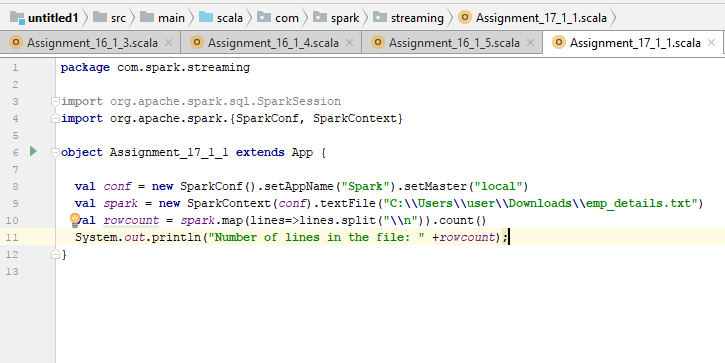
**ASSIGNMENT-17.1**

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

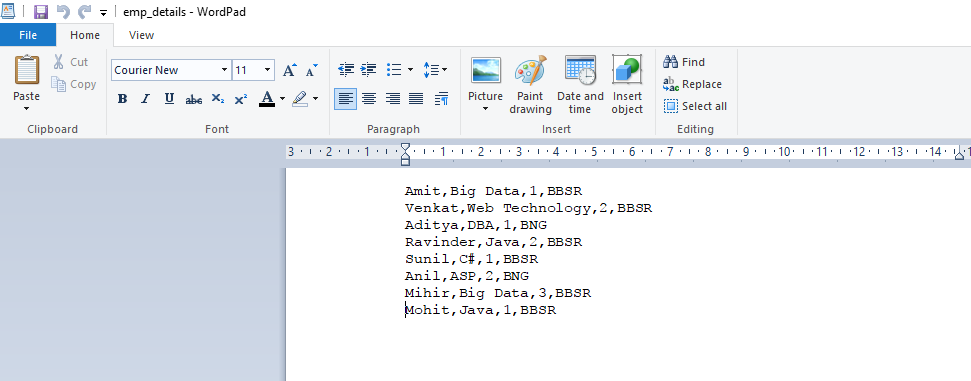
**PROGRAM:**

****

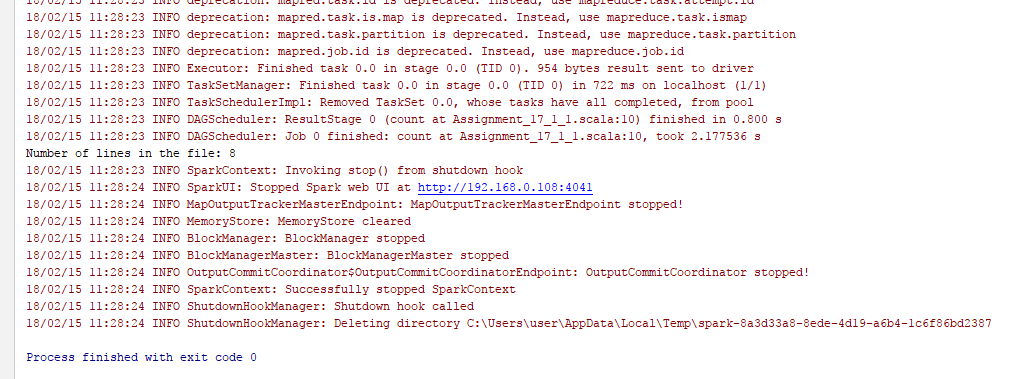
In the above program,

* Master has to be set for both local and production runs.
* textFile is passed as input to the program which is separated by commas.
* Using lines.split([\\n](\\\\n)) , total number of lines in the file are calculated to find the total rowcount.

**INPUT FILE:**

****

**RESULT:**

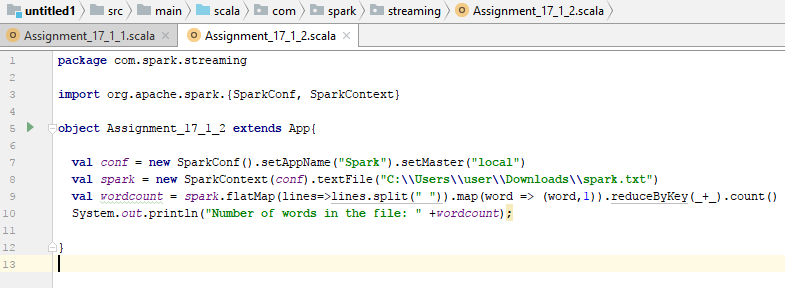
****

*Below are the files attached for reference:*

**

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

**PROGRAM:**

****

In the above program,

* Master has to be set for both local and production runs.
* textFile is passed as input to the program which is separated by commas.
* Using lines.split(“ “) , total number of words in the file are calculated to find the total wordcount.
* After words have been split as, for example:

(spark,1)

(is,1)

(spark,1)

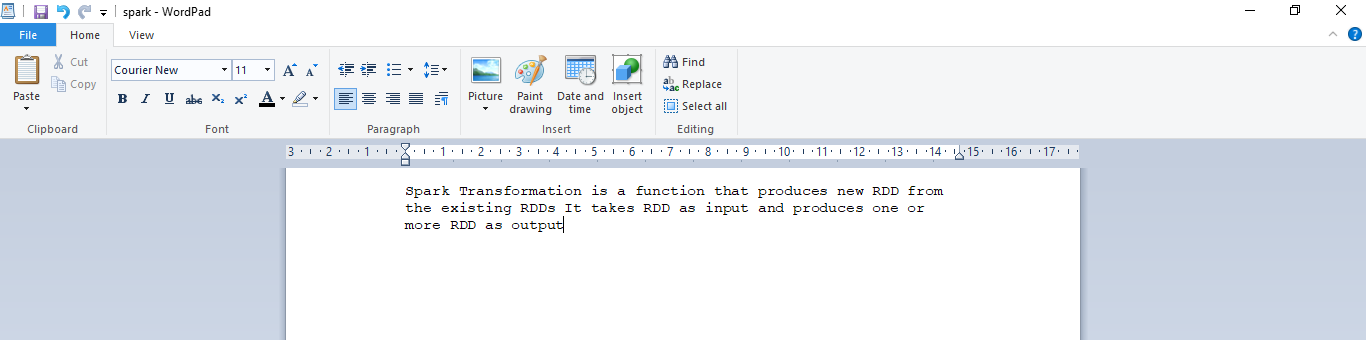
The above needs to be reduced by key to give the total count

For example:

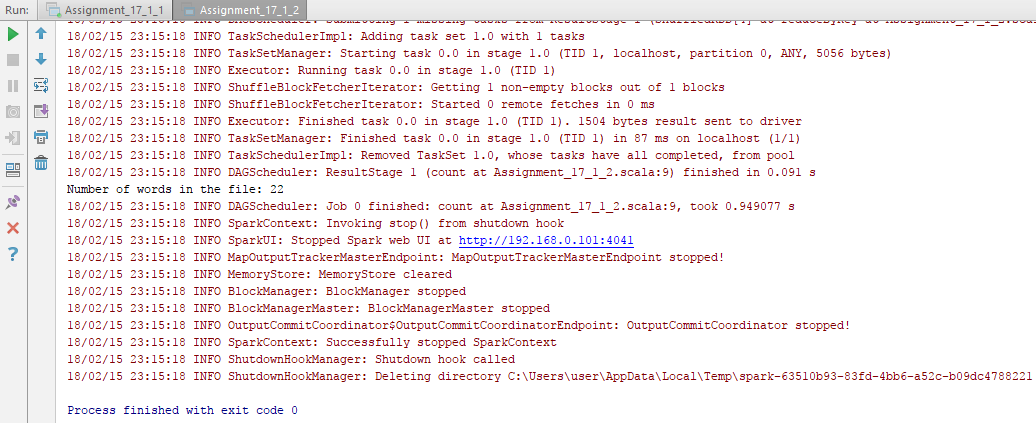
(spark,2)

(is,1)

**INPUT FILE:**

****

**RESULT:**

****

*Below are the files for reference:*

**

1. 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.

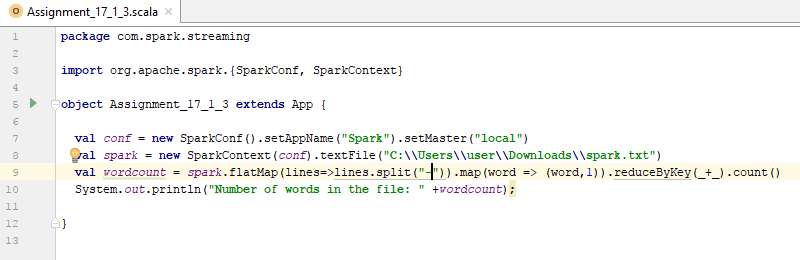
*Sample document :*

This-is-my-first-assignment.

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

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

**PROGRAM:**

****

In the above program,

* Master has to be set for both local and production runs.
* textFile is passed as input to the program which is separated by commas.
* Using lines.split(“- “) , total number of words in the file are calculated to find the total wordcount.
* After words have been split as, for example:

(spark,1)

(is,1)

(spark,1)

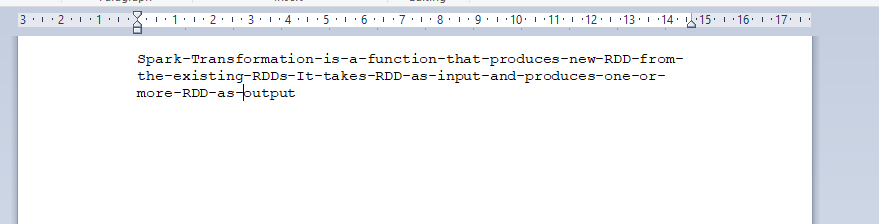
The above needs to be reduced by key to give the total count

For example:

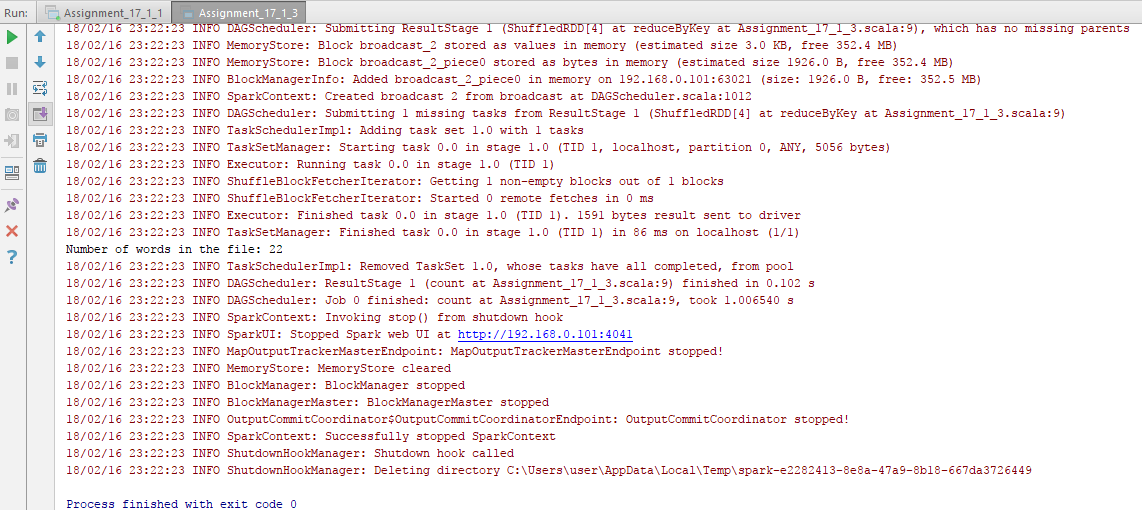
(spark,2)

(is,1)

**INPUT FILE:**

****

**RESULT:**

****

*Below are the files for reference:*

**