SESSION 17 ASSIGNEMNT 1

**Problem Statement:**

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

**Solution:**

Input file: Contents of 17.1\_datatset.txt

This-is-my-first-assignment.

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

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

Spark Code:

import org.apache.spark.sql.SparkSession

object Assignment\_17\_1 extends App {

val spark = SparkSession.builder()

.master("local")

.appName("example")

.getOrCreate()

val sc = spark.sparkContext

//Loading the text file

val csvDF1 = sc.textFile("C:/ACADGILD/Big Data/SESSION\_17/17.1\_datatset.txt")

//Splitting the data separated by -

val counts = csvDF1.flatMap(line => line.split("-"))

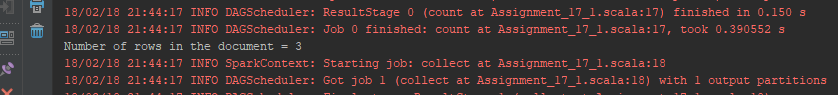
println("Number of rows in the document = " + csvDF1.count())

println("Total number of words in file = " + counts.collect.length)

}

**Output:**

number of rows of data in the document



the count of the total number of words present in the document

