**Problem Statement**

**Task 1**

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.

COMMAND:-

1.val textfile=sc.textFile("/dataset1.txt")

textfile.count()

2.textfile.flatMap(x=>x.split(",").map(y=>(y,1)).count()

3.textfile.flatMap(x=>x.split("-").map(y=>(y,1)).count()

Note:- Here "dataset1.txt" file is used and it is uploaded to github.

EXPLANATION:-

1. This program read the text file "dataset1.txt" from HDFS and then print the number of rows of data in the document.

2. This program to read a text file "dataset1.txt" and print the number of words in the document.

3. We have a document where the word separator is -, instead of space. This spark

code is to obtain the count of the total number of words present along with "-" separator in the document.

OUTPUT:-  