**Assignment 17.1 Spark**

Problem Statement

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:

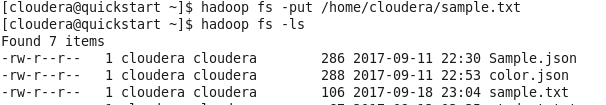
This-is-my-first-assignment.

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

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

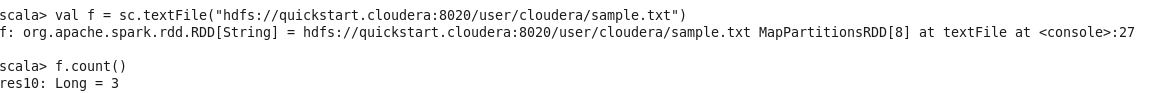
1. Write a program to read a text file and print the number of rows of data in the document.

**Solution:**



val f = sc.textFile("hdfs://quickstart.cloudera:8020/user/cloudera/sample.txt")

f.count();



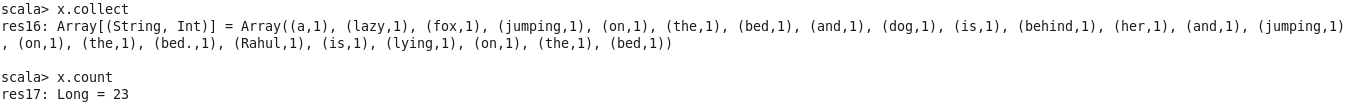
2. Write a program to read a text file and print the number of words in the document.

val f = sc.textFile("hdfs://quickstart.cloudera:8020/user/cloudera/sample.txt")

val w = f.flatMap(l => l.split(" "))

val x = w.map(word => (word, 1)).cache()

x.count()  
val z = x.reduceByKey(\_+\_).collect()



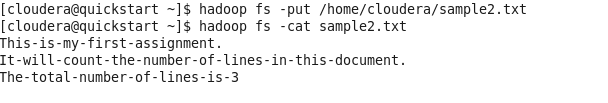
3. We have a document where the word separator is -, instead of space. Write a spark code, to obtain the count of the total number of words present in the document.

Sample document:

This-is-my-first-assignment.

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

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



val f = sc.textFile("hdfs://quickstart.cloudera:8020/user/cloudera/sample2.txt")

val w = f.flatMap(l => l.split("-"))

val x = w.map(word => (word, 1)).cache()

x.count()

val z = x.reduceByKey(\_+\_).collect()

