WordCount in Spark

2/4 points (50%)

Quiz, 4 questions

✖ **Try again once you are ready.**

Required to pass: 80% or higher

You can retake this quiz up to 3 times every 8 hours.

| Back to Week 4 |
| Retake |

✔

1/1 points

1. What does the following line of code do?

words = lines.flatMap(lambda line: line.split(" "))

○ Each word is merged into lines to be counted later.

○ Each word in each line is counted.

◉ Each line in the document is split up into words.

▲

Correct

○ Each line in the document is split into various Spark partitions.

✖

0/1 points

2. What does the following line of code imply about the state of partitions before the action is performed?

words = lines.flatMap(lambda line: line.split(" "))

○ There is only one single partition containing the full document.

○ Each Spark partition corresponds to a word in the document.

This should not be selected

○ Each Spark partition corresponds to a line in the document.

✖

3.

0 / 1 points

When the following command is executed, where is the file written and how can it be accessed?

counts.coalesce(1).saveAsTextFile('hdfs:/user/cloudera/wordcount/outputDir')

○ HDFS and through the system directory with the "cd" terminal command.

○ HDFS and through the "hadoop fs" command.

○ The local file system and through the directory with the "cd" terminal command.

○ The local file system and through the "hadoop fs" command.

This should not be selected

✔

4.

1 / 1 points

What does the number one (1) allow us to do in the following line of code?

tuples = words.map(lambda word: (word,1))

○ The number represents the number of partitions in charge of keeping track of each word.

○ The number represents the number of partitions in charge of counting each line.

○ None, completely arbitrary in order to apply an algorithm that requires a tuple.

◉ Treat each word with a weight of one during the counting process.

Correct