Congratulations! You passed!
TO PASS 80% or higher

Keep Learning
Retake the assignment in **7h 58m**

GRADE
100%

WordCount in Spark

LATEST SUBMISSION GRADE
100%

1. What does the following line of code do? **1 / 1 point**

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

- Each word in each line is counted.
- Each line in the document is split into various Spark partitions.
- Each word is merged into lines to be counted later.
- (selected) Each line in the document is split up into words.

Correct

2. What does the following line of code imply about the state of partitions before the action is performed? **1 / 1 point**

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

- Each Spark partition corresponds to a word in the document.
- (selected) Each Spark partition corresponds to a line in the document.
- There is only one single partition containing the full document.

Correct

3. When the following command is executed, where is the file written and how can it be accessed? **1 / 1 point**

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

- The local file system and through the "hadoop fs" command.
- (selected) HDFS and through the "hadoop fs" command.
- HDFS and through the system directory with the "cd" terminal command.
- The local file system and through the directory with the "cd" terminal command.

Correct

4. What does the number one (1) allow us to do in the following line of code? **1 / 1 point**

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

- (selected) Treat each word with a weight of one during the counting process.

- 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.

✓ **Correct**