

✔ Congratulations! You passed!

Go to next item

Grade received 100% Latest Submission Grade 100% To pass 80% or higher

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 up into words.

☐

Each word is merged into lines to be counted later.

☐

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

✔ 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.

☒

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')
```

☒

HDFS and through the "hadoop fs" command.

☐

The local file system 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

☒

HDFS and through the "hadoop fs" command.

☐

The local file system and through the “hadoop fs” command.



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



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



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.



Correct