Name: _____   USC ID: _____

Quiz 10: Hadoop MapReduce (10 points), 10 minutes

Consider the WordCount program you have seen in class (with TokenizerMapper and IntSumReducer). Assume that the **Combiner** is used which executes the same reduce function as the Reducer. Suppose that the program takes a directory with two input files: file1 and file2 whose contents are as follows. Both file1 and file2 have two lines of text.

File1	File2
here	here or there
here and there	there

1. [1 point] How many Map tasks will Hadoop create for the above input directory?

    2 map tasks

2. [5 points] What are the key-value pairs (please specify the content of both key and value) that each Map task will handle? What key-value pairs will it output?

    Mapper 1
    input: : [(0, "here"), (5, "here and there")]
    output: [("here", 1), ("here", 1), ("and", 1), ("there", 1)]

    Mapper 2
    input: : [(0, "here or there"), (14, "there")]
    output: [("here", 1), ("or", 1), ("there", 1), ("there", 1)]

3. [2 points] What are the key-value pairs output by each combiner?

    Combiner 1: [("here", 2), ("and", 1), ("there", 1)]
    Combiner 2: [("here", 1), ("or", 1), ("there", 2)]

4. [2 points] What are the final key-value pairs output by the Reducer, assuming there is only one reducer?

    [("here", 3), ("and", 1), ("or", 1), ("there", 3)]