Name: _____ USC ID: _____

INF 551 – Fall 2016 (Morning)

Quiz 12: 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. For example, file1 contains 3 lines of text, where the first line is: cat and dog.

File1	File2
cat and dog cat dog dog or cat	cat or dog this cat

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

   2

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?

   Map Task 1:

Input	Output
(0, “cat and dog”) (12, “cat dog”) (20, “dog or cat”)	(“cat”, 1) (“cat”, 1) (“cat”, 1) (“and”, 1) (“dog”, 1) (“dog”, 1) (“dog”, 1) (“or”, 1)

   Map Task 2:

Input	Output
(0, “cat or dog”) (11, “this cat”)	(“cat”, 1) (“cat”, 1) (“this”, 1) (“dog”, 1) (“or”, 1)

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

Combiner 1:
("dog", 3)
("cat", 3)
("and", 1)
("or", 1)

Combiner 2:
("cat", 2)
("or", 1)
("dog", 1)
("this", 1)

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

("dog", 4)
("cat", 5)
("and", 1)
("or", 2)
("this", 1)