Name: Yijun Lin USC ID: 3689281438

10/10

INF 553 – Spring 2016

Quiz 2 (10 points)

Circle all answers to the question to get the full credit. ***No*** *partial credits will be given.*

a.b 1. Which of the following operation(s) for computing statistics of a set of integers is (are) **commutative** and **associative**? (1 point)

a. Sum
b. Count
c. Average
d. Median

(1)

C 2. Consider a MapReduce program that computes the **smallest** integer in a large set of integers. Suppose one of its Mappers outputs a key-value pair: (8, 1). Which of the following is most likely the input to the Mapper? (1 point)

a. ("8", [3, 5, 1, 8])
b. ("1", [1, 3, 5, 8])
c. ("1", [10, 8, 9, 10])
d. ("1", [7, 8, 8])

(1)

3. What is the role of **Master** in Map-Reduce? Explain! (4 points)

(4)

① Ping the workers periodically to detect failure

② Schedule a map task or a reduce task to a worker

③ Communicate between workers, like tell the reduce worker that map worker is failure.

④ Master also manage the ~~local~~ location and size of the intermediate file, and transfer the file from map to reduce.

⑤ Schedule the group-by key process to ~~convient~~ sort or shuffle the key-value pairs from map tasks to reduce tasks.

4. Fill in the logic for the **matrix multiplication** using **Two-Phases MapReduce**. (This doesn't need to be Python code—just indicate the logic needed.) (4 points)

The input to the map function will be a list of one or more matrix values, each represented as a tuple of the form (matrix, i, j, value), where matrix is a string and i, j, and value are integers. The matrix string identifies which matrix the record originates from: either "A" or "B".

The output from the reduce function will also be a row of the result matrix represented as a tuple. Each tuple will be of the form (i, j, value) where each element is an integer, $C[i,k] = \text{Sum}_j(A[i,j] * B[j,k])$.