Name: _____   USC ID: _____

INF 551 – Fall 2018 (Afternoon)
Quiz 7: Query execution (10 points)
10 minutes

Consider natural-joining two relations R(A, B) and S(B, C) using the **partitioned hash join** algorithm. Suppose M = 101 pages, B(R) = 5,000 blocks, and B(S) = 10,000 blocks. Assume all memory is used in the join process.

1. [6 points] Describe the steps of the algorithm (i.e., how many passes, what are the outputs generated at each pass, and the size of the outputs).

    1st pass: hash R into 100 buckets, each bucket has a size of 50. Hash S using same hashing function into 100 buckets, each bucket has a size of 100.
    2nd pass: read a entire bucket of S into memory one by one, and read a block of R's data in the correponding bucket one by one, join and output the join result.

2. [2 points] What is the total cost (i.e., the number of block I/O's) of the algorithm?

    3B(R) + 3B(S) = 45000

3. [2 points] If the same join is performed using the **sort-merge** join algorithm, can it be done in **two** passes? Explain your answer.

    No it can not.
    Because B(R) + B(S) >M^2