Name: _____

CSE 4502/5717 Big Data Analytics

Exam I; September 26, 2024

Note: You are supposed to give proofs to the time and processor bounds of your algorithms. Read the questions carefully before attempting to solve them.

1. (17 points) Input are two sequences A and B of length n each. The elements in A are distinct, the elements in B are distinct, B is in sorted order and A may not be in sorted order. It is known that A and B have $n^{2/7}$ common elements. The problem is to output an element that is common to A and B. Show how to do this in $\widetilde{O}(n^{5/7}\log^2 n)$ time. Prove the runtime of your algorithm. (You could use the fact that $(1-x)^{1/x} \leq \frac{1}{e}$, for any $0 < x < 1$).

2. (16 points) Input is an array $A[1:n]$. It can only be one of the following two types: 1) A has $\frac{n}{2}$ copies of 0 and $\frac{n}{2}$ copies of 1; 2) A has $\frac{n}{3}$ copies of 1 and $\frac{2}{3}n$ copies of 2. Present a Monte Carlo algorithm to identify the type of a given array that has a run time of $O(\log n)$. Prove that the output of your algorithm will be correct with a high probability.

3. (16 points) Input are two $n \times n$ matrices A and B. The problem is to multiply these two matrices. Present an algorithm for this problem that takes $O(\log n)$ time. You can use up to $\frac{n^3}{\log n}$ CREW PRAM processors.

4. (17 points) Input is a sequence $X = b_1, b_2, \ldots, b_n$ of bits. The problem is to find the position of the leftmost 1. For example, if $X = 0, 0, 0, 1, 0, 0, 1, 1$, the answer is 4. Present an $O(\log n)$ time algorithm to solve this problem. You can use up to $\frac{n}{\log n}$ CREW PRAM processors.

5. (17 points) Input are two sorted sequences X and Y of length n each. Both X and Y are residing in a disk. The elements in X are distinct and the elements in Y are distinct. The problem is to compute $Z = X \cap Y$ and output Z in the disk. Show how to do this in $O\left(\frac{n}{B}\right)$ I/O operations. M is the core memory size and B is the block size.

6. (17 points) Input is a (not necessarily sorted) sequence $X = k_1, k_2, \ldots, k_n$ of real numbers residing in a disk. Input also is another real number q. Assume that the elements in X are distinct. The problem is to check if there are two elements k' and k'' in X whose sum is q. Present an algorithm to solve this problem whose I/O complexity is $O\left(\frac{n}{B}\frac{\log(n/M)}{\log(M/B)}\right)$.