SC3020-CZ4031: Tutorial 6

Problem 1. Suppose that we have three sets of integers, denoted as S_1, S_2, and S_3, respectively. Each set is sorted and given to you in a file. Assume that each disk block can hold 2 integers, and the memory has 4 blocks. The content of S_1, S_2, and S_3 is shown below.

$S_1 : [1, 10], [20, 30], [40, 50]$
$S_2 : [5, 35], [37, 38]$
$S_3 : [45, 60]$

The notation $[x, y]$ represents a block holding integers x and y. It is clear that S_1, S_2, and S_3 occupy 3, 2, and 1 block, respectively.

Demonstrate how to merge S_1, S_2, and S_3 into one sorted file using the merging algorithm taught in the class. What is the I/O cost of the algorithm?

Problem 2. Assume that each disk block can hold 2 integers, and the memory has 3 blocks. You are given a set S stored on the disk in 9 blocks as shown below:

$S : [50, 30], [80, 20], [10, 70], [40, 60], [55, 25], [90, 5], [85, 95], [35, 15], [65, 75]$

Suppose that we execute the initial step of the external sort algorithm taught in the class. Show all the sorted runs produced by this step. What is the I/O cost of this step?

Problem 3. Continuing on Problem 2, now execute the first merging step on the sorted runs you obtained. Show the sorted runs at the end of this step. What is the I/O cost of this step?

Problem 4. Let S be a set of integers stored in 100000 blocks. If the memory has $M = 100$ blocks. How many merging steps are required to sort S?

Problem 5. Let S be a set of integers stored in 10^9 blocks. If we sort S using the external sort algorithm, what is the smallest number M of memory blocks required for the algorithm to incur only one merging step?

Problem 6 Let $R(A, B)$ be a relation with attributes A and B. No two tuples in R are identical. Explain how to sort R according to the following order: rank tuple t_1 before tuple t_2 if (i) $t_1.A < t_2.A$ or (ii) $t_1.A = t_2.A$ and $t_1.B < t_2.B$.

Problem 7 Let S be a *bag* (a.k.a. multi-set) of integers, namely, S may contain duplicate integers. Adapt the external sort algorithm to sort S.

Problem 8 Let S be a *bag* (a.k.a. multi-set) of integers, namely, S may contain duplicate integers. Adapt the external sort algorithm to output a disk file containing the *distinct* integers of S in ascending order.

Problem 9 Discuss how to process the following query on a relation $R(A, B)$:

    SELECT A, MAX(B) FROM R GROUP BY A