Fall 2020 – Quiz 7 (External Sorting)

10 points, 15 minutes

Monday afternoon section

Change B(R) = 103 and M = 5.

Consider external-sorting a table R which contains 103 blocks of data, using 5 pages of memory buffer. That is, B(R) = 103 and M = 5. Note: use all available memory for sorting and merging.

1. [8 points] For each pass (sorting and merging), state the number of runs and the size of runs generated by the pass.
   Pass 0 (use all pages to sort): Generate 21 runs, 5 blocks/run*20 and 3 blocks/run*1.
   Pass 1 (4-way merge): Generate 6 runs, 20 blocks/run*5 and 3 blocks/run*1.
   Pass 2 (4-way merge): Generate 2 runs, 80 blocks/run*1 and 23 blocks/run*1.
   Pass 3 (2-way merge): Generate 1 sorted run of 103 blocks

2. [2 points] What is the total cost (measured by the number of block I/O's) of this external-sorting?
   Total cost = (# of passes) × 2 × (# of blocks) = 4 × 2 × 103 = 824
   Or Total cost = $\left(1 + \left\lceil \log_4 \left\lceil \frac{103}{5} \right\rceil \right\rceil\right) \times 2 \times 10 = 824$

MakeUP

Consider external-sorting a table R which contains 110 blocks of data, using 4 pages of memory buffer. That is, B(R) = 110 and M = 4. Note: use all available memory for sorting and merging.

1. [8 points] For each pass (sorting and merging), state the number of runs and the size of runs generated by the pass.

   pass0, 28 runs, 27 runs of 4 blocks, 1 run of 2 blocks.
   pass1, 10 runs, 9 runs of 12 blocks , 1 run of 2 blocks.
   pass2, 4 runs, 3 runs of 36 blocks , 1 run of 2 blocks.
   pass3, 2 runs, 1 run of 108 blocks, 1 run of 2 blocks.
   pass4, 1 run , 1 run of 110 blocks.

2. [2 points] What is the total cost (measured by the number of block I/O's) of this external-sorting?

2 * B(R) * # of pass = 2*110*5 = 1100