CARNEGIE MELLON UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE
15-445/645 – DATABASE SYSTEMS (FALL 2017)
PROF. ANDY PAVLO

Homework 4 (by Sivaprasad Sudhir)

Due: **Wednesday Oct 11, 2017 @ 11:59pm**

IMPORTANT:

- **Upload this PDF** with your answers to **Gradescope by 11:59pm on Wednesday Oct 11, 2017**.
- **Plagiarism**: Homework may be discussed with other students, but all homework is to be completed **individually**.
- **You have to use this PDF for all of your answers.**

For your information:

- Graded out of **100** points; **2** questions total
- Rough time estimate: $\approx$ 1 - 2 hours (0.5 - 1 hours for each question)

Revision : 2017/10/08 22:58

Question	Points	Score
Sorting	40	
Join Algorithms	60	
Total:	100	

Question 1: Sorting . . . . . [40 points]

We have a file with a million pages (N = 1,000,000 pages), and we want to sort it using external merge sort. Assume the simplest algorithm, that is, no double buffering, no blocked I/O, and quicksort for in-memory sorting. Let B denote the number of buffers.

(a) **[10 points]** What is the smallest number of buffers B, that can sort the file with N = 1,000,000 pages, in 2 passes?

☐ 32 ☐ 33 ☐ 34 ☐ 99 ☐ 100 ☐ 101 ☐ 102 ☐ 999 ☐ 1,000
☐ 1,001 ☐ 1,000,000 ☐ 1,000,001

(b) **[10 points]** What is the smallest number of buffers B, that can sort the file with N = 1,000,000 pages, in 3 passes?

☐ 32 ☐ 33 ☐ 34 ☐ 99 ☐ 100 ☐ 101 ☐ 102 ☐ 999 ☐ 1,000
☐ 1,001 ☐ 1,000,000 ☐ 1,000,001

(c) **[10 points]** How many passes are needed to sort the file with N = 1,000,000 pages with 6 buffers?

☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11

(d) **[10 points]** What is the total I/O cost to sort the file with N = 1,000,000 pages with 6 buffers?

☐ 14,000,000 ☐ 8,000,000 ☐ 18,000,000 ☐ 10,000,000 ☐ 22,000,000

Question 2: Join Algorithms . . . . . [60 points]

Consider relations R(x, y) and S(x, z) to be joined on the common attribute x. Assume that there are no indexes.

- There are B = 30 pages in the buffer
- Table R spans M = 1800 pages with 50 tuples per page
- Table S spans N = 500 pages with 100 tuples per page

What are the I/O costs for the following joins?

- Assume the simplest cost model, where pages are read and written one at a time
- Assume that you will need one buffer block to hold the evolving output block and one input block to hold the current input block of the inner relation
- Ignore the cost of the final writing of the results

(a) **[10 points]** Block nested loop join with R as the outer relation and S as the inner relation
☐ 30,000 ☐ 31,800 ☐ 32,900 ☐ 33,300 ☐ 34,300

(b) **[5 points]** Block nested loop join with S as the outer relation and R as the inner relation
☐ 29,300 ☐ 31,100 ☐ 31,800 ☐ 32,900 ☐ 34,300

(c) Hash join with S as the outer relation and R as the inner relation (Ignore recursive partitioning and partially filled blocks)

i. **[5 points]** Cost of partition phase
☐ 2,300 ☐ 4,600 ☐ 6,900 ☐ 3,600 ☐ 1,000

ii. **[5 points]** Cost of probing phase
☐ 2,300 ☐ 4,600 ☐ 6,900 ☐ 3,600 ☐ 1,000

(d) Sort-merge join

i. **[10 points]** Cost of sorting R
☐ 8966 ☐ 7934 ☐ 6578 ☐ 1828 ☐ 2204

ii. **[5 points]** Cost of sorting S
☐ 8966 ☐ 7934 ☐ 6578 ☐ 1828 ☐ 2204

iii. **[10 points]** Cost of merge assuming no duplicates in the join attribute
☐ 2,300 ☐ 4,600 ☐ 6,900 ☐ 154 ☐ 77

iv. **[10 points]** Cost of merge in the worst case
☐ 6,900 ☐ 31,800 ☐ 33,300 ☐ 900,000 ☐ 1,800,000