# CSE4411 fall 2007

## Assignment #2: Due November 9, 2007, 2 pm

## Weight 2.5%

This assignment involves the topic of query processing (chapters 12, 13, 14, 15 of textbook).

**Exercise 1**: do exercise 12.4, questions 1 and 2, from the textbook.

## Answer:

1.
   (a) Match. Primary conjuncts are: Sailors.sid < 50,000
   (b) Match. Primary conjuncts are: Sailors.sid = 50,000
2.
   (a) No Match. Range queries cannot be applied to hash indexes.
   (b) Match. Primary conjuncts are: Sailors.sid = 50,000

**Exercise 2**: do exercise 13.4, questions (2.a) and (2.b), from the textbook,.

## Answer:

In Pass 0, 31250 sorted runs of 320 pages each are created. For each run, we read and write 320 pages sequentially. The I/O cost per run is $2 * (10+5+1* 320) = 670$ms. Thus, the I/O cost for Pass 0 is $31250 * 670 = 20937500$ms. For each of the cases discussed below, this cost must be added to the cost of the subsequent merging passes to get the total cost. Also, the calculations below are slightly simplified by neglecting the effect of a final read/written block that is slightly smaller than the earlier blocks.

For 319-way merges, only 2 more passes are needed. The first pass will produce Ceiling(31250/319) = 98 sorted runs; these can then be merged in the next pass. Every page is read and written individually, at a cost of 16ms per read or write, in each of these two passes. The cost of these merging passes is therefore $2*(2*16)*10000000 = 640000000$ms. (The formula can be read as 'number of passes times cost of read and write per page times number of pages in file'.)

With 256-way merges, only two additional merging passes are needed. Every page in the file is read and written in each pass, but the effect of blocking is different on reads and writes. For reading, each page is read individually at a cost of 16ms. Thus, the cost of reads (over both passes) is $2 * 16 * 10000000 =$

320000000ms. For writing, pages are written out in blocks of 64 pages. The I/O cost per block is 10 + 5 + 1 ∗ 64 = 79ms. The number of blocks written out per pass is 10000000/64 = 156250, and the cost per pass is 156250∗79 = 12343750ms. The cost of writes over both merging passes is therefore 2 ∗ 12343750 = 24687500ms. The total cost of reads and writes for the two merging passes is 320000000 + 24687500 = 344687500ms.

**Exercise 3**: do exercise 14.6, questions 1, 2, from the textbook.

## Answer:

1. Basic idea is to read each page of the outer relation, and for each page scan the inner relation for matching tuples. Total cost would be #pagesinouter + (#pagesinouter ∗ #pagesininner) which is minimized by having the smaller relation be the outer relation. TotalCost = M + (M ∗ N) = 2, 000, 010, 000
The minimum number of buffer pages for this cost is 3.

2. This time read the outer relation in blocks, and for each block scan the inner relation for matching tuples. So the outer relation is still read once, but the inner relation is scanned only once for each outer block, of which there are ceiling(#pagesinouter/( B−2) ).

TotalCost = M + N ∗ ceiling(M/( B − 2)) = 2, 010, 000.
The minimum number of buffer pages for this cost is 1002.

**Exercise 4**: do exercise 15.8, question 1, from the textbook.

## Answer:

The query optimizer will need information such as what indexes exist (and what type) on: S.sid, Y.sid, Y.pid, P.pid, S.city, P.price It will also need statistics about the database such as low/high index values and distribution between fields.