Homework 2 – Answer

17.a. Assume a page size of 4K bytes and a page access time of t ms. The data file occupies 5000 pages and all pages will have to be scanned. Hence, the query will require $5000 \times t$ ms. For t = 10ms, this will give 50 sec.

b. Hash index without overflow page means that we have at most one bucket for the answer. The answer depends on whether the index is clustered and the number of index entries in each bucket.

- If the index is clustered, we need only one page transfer to output all the rows; this, together with the cost for accessing the index (1.2) gives 2.2 page transfer, i.e., 22 ms.
- If the index is unclustered, we need $1.2 + t \times$ #rows in the answer.

c. If we assume that each entry in the index occupies 100 bytes then an index page can hold 40 entries. Since the data file occupies 5000 pages, the leaf level of the tree must contain at least 5000/40, or 125 pages. Then the number of levels in the tree (assuming page 75% occupancy in the index) is $\lceil \log_{30} 125 \rceil + 1 = 3$. Assume that the index is clustered, not integrated with the data file and that all matching entries are in a single page, 4 I/O operations and $t \times 4$ ms are required to retrieve all matching records.

18. a. $(\log_2 5000 + 1) \times t$ ms (binary search plus one I/O for output)
b. 2 page transfers are needed (finding and inserting, assuming that the bucket still have space); so the cost is $2 \times t$ ms
c. 4 page transfers will be required to find the leaf node (where to insert) then one for reading the data page and one for write back, total is 6. So, $6 \times t$ ms

19. For updating the search key value, we need to locate the row, delete it (key is different), create the update row, and then write-back.

a. $(\log_2 5000 + 1) \times t$ ms (binary search plus one I/O for output) to find
$1 \times t$ ms (for deleting)
$(\log_2 5000 + 1) \times t$ ms (locate new page and update)

Total: $(2(\log_2 5000 + 1) + 1) \times t$ ms

b. Similarly, the cost will double comparing to 18 b: $2 \times t$ ms

c. Same as with b, the cost will double comparing to 18 c: $12 \times t$ ms