COMP7640
Written Assignment #2

Due: 11:59 PM 27 Mar (Thursday), 2025

1. Given a disk with the following characteristics:

average seek time	20 ms
track-to-track seek time	4 ms
rotational delay	4 ms
max transfer rate	20 ms per track
#bytes per sector	256
#sectors per track	50
#tracks per cylinder	10
#tracks per surface	1,000

Suppose that we are given a file containing 500 256-byte records.

(1) How many seconds will it take to transfer all the records from this disk into the main memory, when adopting the "random" strategy? (**10 marks**)

(2) How many seconds will it take to transfer all the records from this disk into the main memory, when adopting the "next" strategy? (**20 marks**)

Solution:

(1) The access time for each record is 20+4+20/50 = 24.4 ms
500*24.4=12.2 secs

(2) Each track contains 256*50=12,800 bytes
500 256-byte records require 256*500/12800=10 tracks
Since 10 tracks per cylinder, we need to read 1 cylinder.
20+4+10*20=224 ms

Or since 256 bytes per sector and 256 bytes per record, each sector can hold 1 record. For transfer rate: 20ms per track/50 sectors per track = 0.4 ms per record. For the 1st record, we need 20+4+0.4=24.4ms. For the remaining 499 records: 0.4*499 = 199.6 ms. In total: 24.4ms + 199.6ms = 224ms.

2. Suppose that we are given a relation R containing 1,000,000 records, each of which takes 64 bytes. Each disk block (page) can hold 4,000 bytes of data. If we build an ISAM index on relation R in which each entry in the leaf/non-leaf pages of the index takes 4 bytes, estimate the search I/O cost of using the ISAM index and the I/O cost of using binary search on relation R. (**35 marks**)

Solution:

(1) Relation R needs 1,000,000*64/4000=16,000 pages. Namely, #data pages B=1,6000. (**7 marks**)
(**When a record cannot span two pages, the following answer is also correct**:

Each disk page can store 4,000/64=62 records at most. Relation R requires B=1,000,000/62 = 16130 pages)

(2) Since each entry in the index needs 4 bytes, each leaf/non-leaf page can hold 4000/4=1000 entries. The fan-out F is then 1001. (**7 marks**)

(3) Note that each data entry corresponds to a record. So, we need 1,000,000 data entries in the leaf pages. The total number of leaf pages is then 1,000,000/1001=1000 pages. Namely, #primary leaf pages N=1000 pages. (**7 marks**)

(4) Search I/O cost for ISAM: 1+ceil($\log_{1001}1000$)=2 I/Os (**7 marks**)
I/O cost for binary search: ceil($\log_2 16000$)=14 I/Os (**7 marks**)
(**The following answer is also correct**: I/O cost for binary search: ceil($\log_2 16130$)=14 I/Os)

3. Given a B+ tree index as shown in Figure 1 (all key values are integers), please answer the following questions

(1) If we are given a query with the condition 12< key<23, how many I/Os in total (including reading the records) are needed for this query when using this B+ tree index? (**10 marks**)

(2) Draw the B+ tree after inserting 4* and 6* into the original B+ tree; (**10 marks**)

(3) Draw the B+ tree after deleting 37* from the original B+ tree. (**15 marks**)

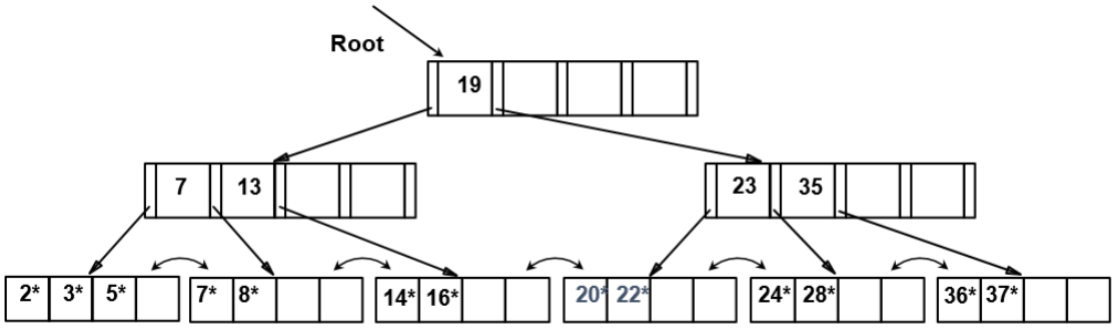


Figure 1

Solution:

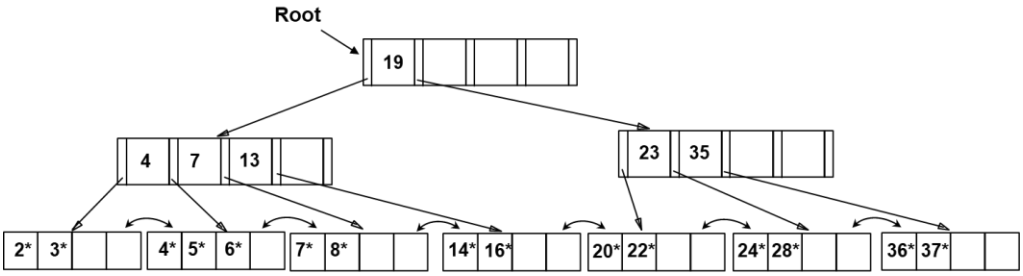
(1) When all the key values are integers,

9 I/Os in total.

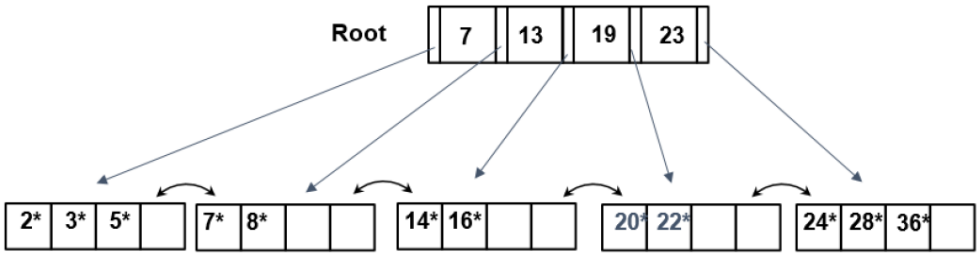
5 I/Os for B+ tree index (**5 marks**).

4 I/Os for 4 records (**5 marks**)

(2) Below please see the updated index



(3) Below please see the updated index



❖

Late Penalty: Late submissions will be marked down **50% for each day it is late**. Any exceptions to this rule must be made prior to when the assignment is due and the excuse needs to be a good one - just too busy won't cut it. Individual exceptions are unfair to other students and hence they won't be made unless the circumstances are truly exceptional.

Plagiarism: All work submitted by you should be your own. **Copying or sharing of assignments would constitute plagiarism. Penalty will be given to those involved in plagiarism. This will be no exception once plagiarism is caught – being copied without notice won't excuse it; you should simply keep your assignment well.**

Submission: Please submit your assignment via BUMoodle.