CMPSCI 445 — Homework 4

95 Points

Due November 3rd, at the beginning of class.
Typed solutions preferred. If hand-written, solutions must be **legible**.

Database Indexing

1. (15 Points) Consider the following relation:

$$\text{Emp}(\underline{\text{eid: integer}}, \text{ sal: integer, age: real, did: integer})$$

   There is a clustered index on eid and an unclustered index on age.

   (a) Give an example of an update that is definitely speeded up because of the available indexes. (English description is sufficient.)

   (b) Give an example of an update that is definitely slowed down because of the indexes. (English description is sufficient.)

   (c) Can you give an example of an update that is neither speeded up nor slowed down by the indexes?

2. (20 points) Consider the page format for variable-length records that uses a slot directory.

   (a) One approach to managing the slot directory is to use a maximum size (i.e., a maximum number of slots) and allocate the directory array when the page is created. Discuss the pros and cons of this approach with respect to the approach discussed in the text.

   (b) Suggest a modification to this page format that would allow us to sort records (according to the value in some field) without moving records and without changing the record ids.

3. (40 points) Consider the B+ tree index shown in Figure 1, which uses Alternative (1) for data entries. Each intermediate node can hold up to five pointers and four key values. Each leaf can hold up to four records, and leaf nodes are doubly linked as usual, although these links are not shown in the figure. Answer the following questions:

   (a) Name all the tree nodes that must be fetched to answer the following query: "Get all records with search key greater than 38."

   (b) Show the B+ tree that would result from inserting a record with search key 109 into the tree.

   (c) Show the B+ tree that would result from deleting the record with search key 81 from the original tree.

   (d) Name a search key value such that inserting it into the (original) tree would cause an increase in the height of the tree.

   (e) Note that subtrees A, B, and C are not fully specified. Nonetheless, what can you infer about the contents and the shape of these trees?

4. (20 points) Suppose that a page can contain at most four data values and that all data values are integers. Using only B+ trees of order 2, give examples of each of the following:

   (a) A B+ tree whose height changes from 2 to 3 when the value 25 is inserted. Show your structure before and after the insertion.

   (b) A B+ tree in which the deletion of the value 25 leads to a redistribution. Show your structure before and after the deletion.
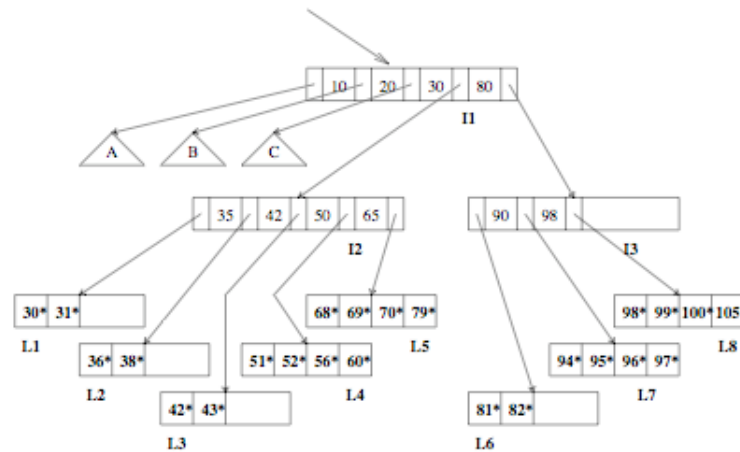

Figure 1: