**CAS CS 460: Introduction to Database Systems – Fall 2019 – Written Assignment #3**

Due: 11/17 11:59pm in gradescope

Problem 1[25 pts]

1.What is the most important difference between a disk and a tape?

2.Explain the terms seek time, rotational delay and transfer time. Make sure that your answer shows the difference between them.

3.Both disks and main memory support direct access to any desired location (page). On average, main memory accesses are faster, of course. What is the other important difference (from the perspective of the time required to access a desired page)?

4.Explain what the buffer manager must do to process a read request for a page. What happens if the requested page is in the pool but not pinned?

Problem 2[25 pts]
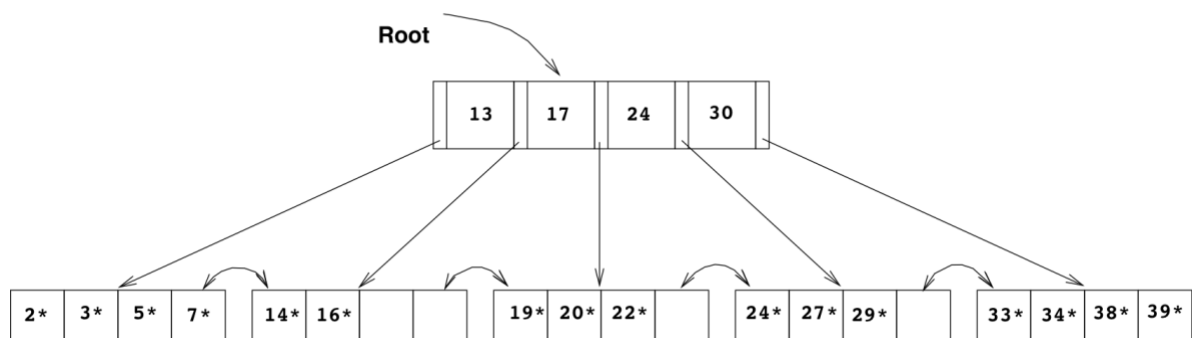
1. Based on the given B+ tree, Identify a list of five data entries such that:

   (a) Inserting the entries in the order shown and then deleting them in the opposite order (e.g., insert a, insert b, delete b, delete a) results in the original tree.

   (b) Inserting the entries in the order shown and then deleting them in the opposite order (e.g., insert a, insert b, delete b, delete a) results in a different tree.

2. What is the minimum number of insertions of data entries with distinct keys that will cause the height of the (original) tree to change from its current value (of 1) to 3?
3. Would the minimum number of insertions that will cause the original tree to increase to height 3 change if you were allowed to insert duplicates (multiple

data entries with the same key), assuming that overflow pages are not used for handling duplicates?

Problem 3. [25 pts]
Suppose that we are using extensible hashing on a file that contains records with the following search-key values:

(449, 124, 654, 831, 1016, 176, 285, 468, 615, 340, 331, 135, 667, 818, 117, 429)

Load these values into a file in the given order using extensible hashing. Assume that every block (bucket) of the hash index can store up to four (4) values.
Show the structure of the hash index after every 4 insertions, and the global and local depths. Use the hash function: $h(K) = K \bmod 128$ and then apply the extensible hashing technique. Using this function, every number is mapped first to a number between 0 and 127 and then we take its binary representation. Then, the extensible hashing technique is applied on the binary representation. Furthermore, initially, you start with a single bucket and a single pointer and the global and local depths are zero (0).

Problem 4[25 pts]

Briefly answer the following questions.
1. In the context of query optimization, what is an SQL query block?

2. Define the term reduction factor.

3. Describe a situation in which projection should precede selection in processing a project- select query and describe a situation where the opposite processing order is better. (Assume that duplicate elimination for projection is done via sorting.)

4. If there are dense, unclustered (secondary) B+ tree indexes on both R.a and S.b, the join $R \bowtie_{a=b} S$ could be processed by doing a sort-merge type of join—without doing any sorting—by using these indexes.

    (a) Would this be a good idea if R and S each have only one tuple per page, or would it be better to ignore the indexes and sort R and S? Explain.

    (b) What if R and S each have many tuples per page? Again, explain.

5. Why does the System R optimizer consider only left-deep join trees? Give an example of a plan that would not be considered because of this restriction.

6. Explain the role of interesting orders in the System R optimizer.