CS411: Database Systems **Spring 2017**

Assignment 4

Due: Monday, April 10 at 11:59pm

General Instructions

- Feel free to talk to other members of the class in doing the homework. You should, however, write down your solutions yourself. *List the names of everyone you worked with at the top of your submission.*
- Keep your solutions brief and clear.
- Please use Piazza if you have questions about the homework but do not post answers. Feel free to use private posts or come to the office hours.

Homework Submission

- We DO NOT accept late homework submissions.
- We will be using Compass for collecting the homework assignments. Please submit your answers via Compass. Hard copies are not accepted.
- Contact the TAs if you are having technical difficulties in submitting the assignment; attempt to submit well in advance of the due date/time.
- The homework must be submitted in **pdf** format. Scanned handwritten and/or hand-drawn pictures in your documents won't be accepted.
- Please do not zip the answer document (PDF) so that the graders can read it directly on Compass. You need to submit one answer document, named as **hw4_netid.pdf**.
- Please see the assignments page for more details. In particular, we will be announcing errata, if any, on this page.

Question 1. Indexing (10 points)

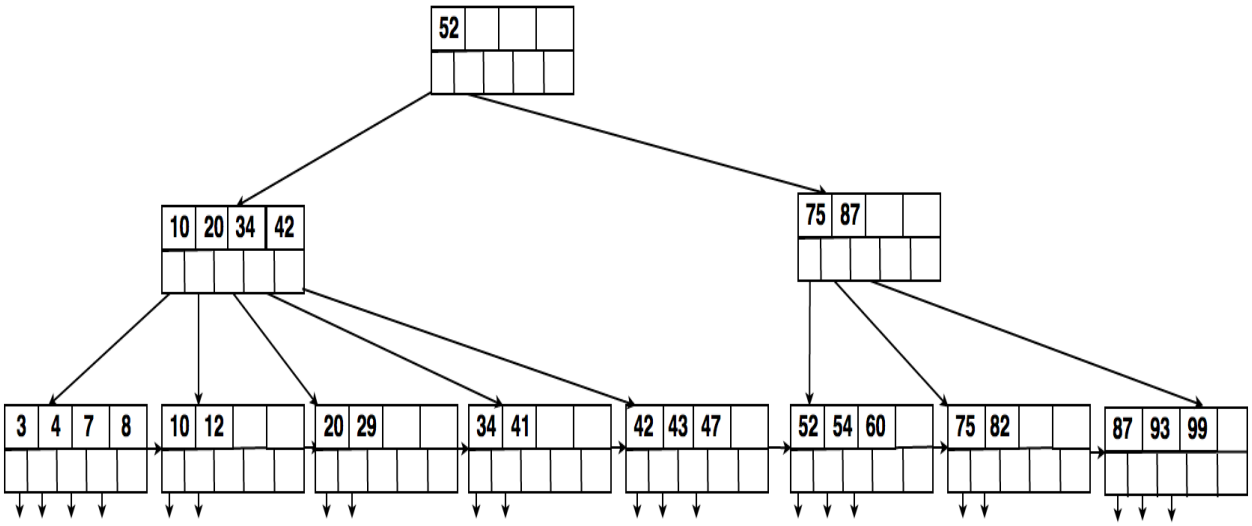
Assume our DBMS is implemented on a system with block size of 32kb. We want to build an index on a relation R with a total of 1,000,000 records, where each record is 256 bytes and each key/pointer pair requires 8 bytes.

1. Suppose the index is dense and unclustered. How many blocks are required to store all of the records of R (data file and index file)?

2. Suppose the index is sparse and clustered (one key-pointer pair/block). How many blocks are required to store all of the records of R (data file and index file)?

Question 2. B+ tree (30 points)

Consider the B+ tree of degree 2 (i.e., d = 2, which means each index node can hold at most 2d = 4 keys and 2d + 1 = 5 pointers) shown below:



1. Draw the B+ tree that would result from inserting a data entry with key 5.

2. Draw the B+ tree that would result from deleting the data entry with key 10 **from the original tree**.

3. Draw the final B+ tree that would result from successively deleting the data entries with keys 34, 41, 42, 43, 47, 52 and 75 **from the original tree**. For partial credit, show the intermediate steps.

Note: *To spare you the trouble of redrawing this diagram, you can use the given BTree.xml file. You can open the file here: http://www.draw.io and modify it to produce your solution.*

Question 3. Extensible Hashing (20 points)

Consider indexing key values using an extensible hash table, in the following order: 1, 25, 44, 58, 71, 80. We are given a hash function h(n) for key n, h(n) = n mod 16; i.e., the hash function is the remainder after the key value is divided by 16. The hash function outputs a 4-bit value. A record is assigned to a bucket based on the first i bits of its hashed key as is typical in extensible hashing. Assume that each data block can hold upto 2 entries.

1. Draw the extensible hash table which contains both the array of pointers and the data blocks after all the six keys are inserted. Show the keys along with their hash values in the buckets. Be sure to specify the variable i mentioned in the lecture slides. Also indicate the bit counter of each block.

2. Now insert record with key 9 in the above hash table and re-draw the resulting hash table.

Question 4. Linear Hashing (40 points)

Let us suppose keys are hashed to 4-bit values and each data block can hold 3 data items. Starting with the hash table composed of two empty buckets (corresponding to 0 and 1), insert records with the hashed keys from

$$0000, 0001, \ldots 1111$$

A record is assigned to a bucket based on the last i bits of its hashed key as is typical in linear hashing. Show the organization after the insertion of every 4th record, for instance after insertion of 0011, after insertion of 0111 and so on. You may assume a linear hashing scheme with capacity threshold of 100%, i.e., you trigger a reorganization if capacity $> 100\%$. For partial credit, please show/explain your steps.