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. If you submission is late, it will be awarded 0 points.

- 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.

- **Please do not submit hand-drawn graphs(B+ trees, hash tables, etc.) for this HW; otherwise, we will deduct half the points from what you get.**

1 Indexing (10 pts)

1. [7] Suppose we want to build an index on a relation R which has a total of x records, with each block capable of holding either y records or z key-pointer pairs. Assuming x is divisible by y, please answer the following questions (if your value evaluates to a fraction, use ceiling $\lceil \quad \rceil$ or floor $\lfloor \quad \rfloor$ as appropriate):

   (a) [3] Suppose you construct a simple single level index, and that index is dense. How many index blocks are required to access all of the records of R?

   (b) [4] Suppose the index built is sparse. If the index stores a pointer to the lowest search key in each block, and the index is a simple single level index, how many data blocks do we need? How many index blocks do we need?

2. [3] True/False question - In order to use a dense index, you will have to have the data file sorted by the search key; otherwise, you will need to use a sparse index. Explain your reasoning.

2   B+ tree (30 points)

Consider a B+ tree of degree 2 shown below:
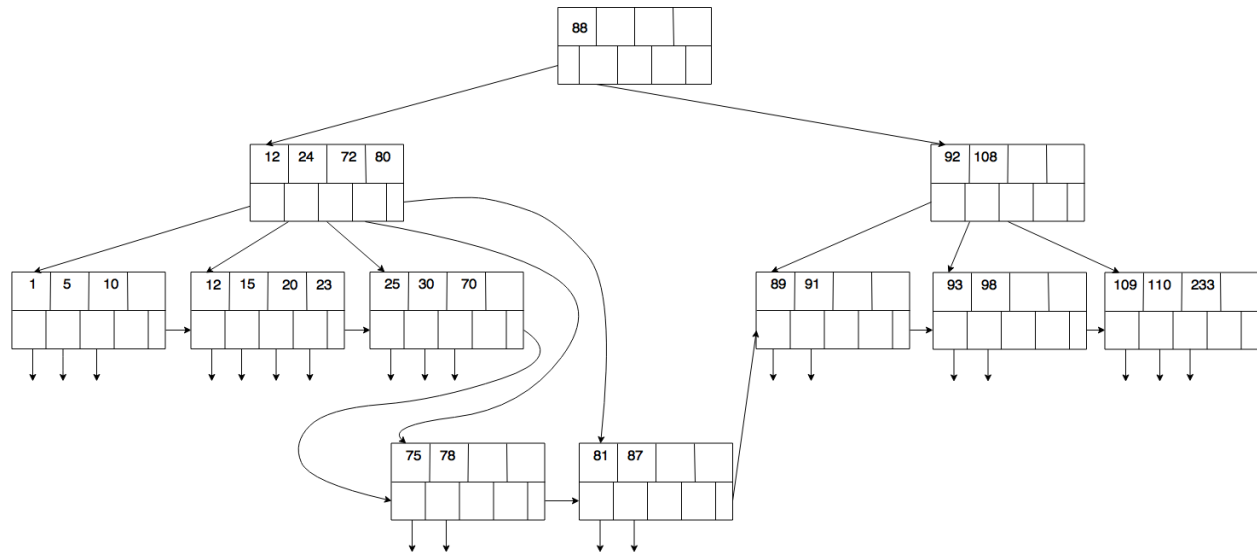
1. [10] Draw the B+ tree that would result from inserting a data entry with key 13.

2. [10] Based on the B+ tree that you drew in the previous question, draw the B+ tree that would result from deleting the data entry with key 75.

3. [10] Based on the B+ tree that you drew in the previous question, draw the B+ tree that would result from deleting the data entry with key 89.

3   Extensible Hashing (30 pts)

Assume you have a extensible hash table with hash function h(k) = k mod 13, expressed as a binary string of size 4, and data block of size 2 (i.e., it can accommodate two tuples). You are asked to index the following key values in order: 25, 13, 23, 21.

1. [20] Draw the extensible hash table which obeys the above constraints after the four keys are inserted.

2. [10] Using your solution to the previous question, now consider insertion of keys 18 and 20 into the hash table, and draw the resulting hash table.

4 Linear Hashing (30 pts)

Consider a linear hash table with $r \leq 1.76n$ with each data block capable of holding 2 records (that is, the average number of record per bucket should not exceed 88% of the total number of records per block):
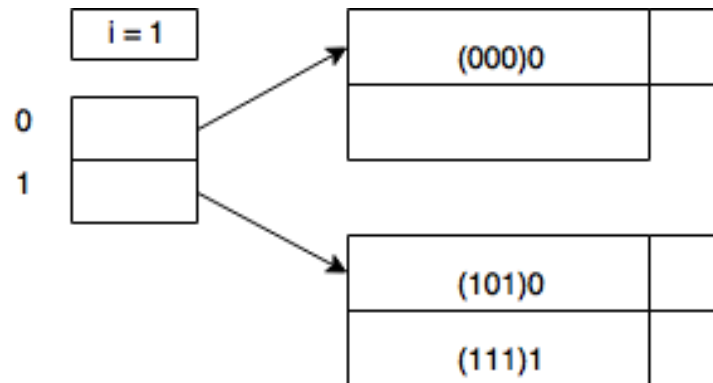
i = 1		(000)0	
0			
1		(101)0	
		(111)1	

1. [10] Insert 1001 and draw the resulting table.

2. [20] With your solution from the previous question, insert 1101, 1110, 0001 incrementally and draw the final table; that is, insert one at a time, check the condition, and move to the next one.