**Virginia Tech.** **CS 4604 – Introduction to DBMS**
**Computer Science** **Spring 2014, Prakash**

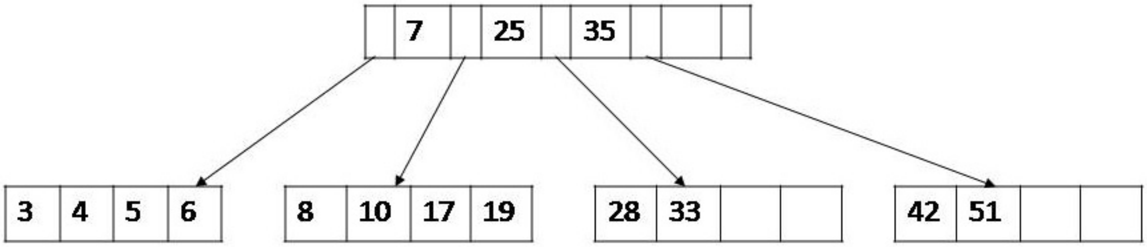
# Homework 3: B+ Trees, Hashing, Bulk Loading
## <u>(due March 4th, 2014, 3:30pm, in class—hard-copy please)</u>

*Reminders*:

a. Out of 100 points. Contains 3 pages.
b. Rough time-estimates: 2~4 hours.
c. Please type your answers. Illegible handwriting may get no points, at the discretion of the grader. Only drawings may be hand-drawn, as long as they are neat and legible.
d. There could be more than one correct answer. We shall accept them all.
e. Whenever you are making an assumption, please state it clearly.
f. Lead TA for this HW: Pranav Nakate.

## Q1. B+ Tree [25 points]

Assume the following B+ tree exists with d = 2:



Sketch the state of the B+ tree after each step in the following sequence of insertions and deletions, maintaining at least 50% occupancy at each step and overflow triggered split. In the diagram above we have not shown pointers in the leaf nodes for simplicity but remember that the leaf nodes are linked lists.

Note: Use the insertion and deletion algorithms given in the textbook section 10.5 (page 349) and 10.6 (page 353) respectively. Root node can have 1 to 2d keys. During deletion redistribute the leaf pages wherever possible.

Q1.1. (5 points) Insert 34

Q1.2. (5 points) Insert 2

Q1.3. (5 points) Insert 15

Q1.4. (5 points) Delete 28

Q1.5. (5 points) Delete 8

## Q2. Bulk Loading a B+Tree [20 points]

Consider an empty B+ Tree with order d=2, i.e. there are at most 4 keys per node, and at most 5 pointers to children. Bulk load the B+ tree with data entries with odd numbers from 1 to 50 (i.e. 1,3,5…49) so that each leaf is **at least** half full using the algorithm outlined in Section 10.8.2 (Page 360) of the textbook.

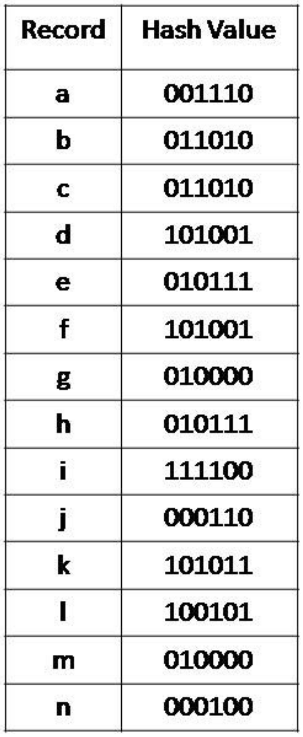
Q2.1. (10 points) Sketch the final B+ tree after bulk loading.

Q2.2. (5 points) What is the height of the tree after inserting all the above keys?

Q2.3. (5 points) What is the minimum number of keys that must be deleted so that the height of the tree decreases by 1? List these keys. Note: Please use the algorithm outlined in section 10.6 (page 353) of the textbook.

## Q3. Linear Hashing [15 points]

We have the following records with the given hash values.



| Record | Hash Value |
|---|---|
| a | 001110 |
| b | 011010 |
| c | 011010 |
| d | 101001 |
| e | 010111 |
| f | 101001 |
| g | 010000 |
| h | 010111 |
| i | 111100 |
| j | 000110 |
| k | 101011 |
| l | 100101 |
| m | 010000 |
| n | 000100 |

Table 1

Q3.1. (10 points) For the records given in Table 1 with the given hash keys, sketch the linear hash structure for the file. There are 4 buckets in the start and each bucket can hold at most 2 records. Initially the structure is empty. Assume that we split whenever a new key triggers the creation of a new overflow page.

Note: Use the linear hashing algorithm outlined in Section 11.3 (page 379) of the textbook.

Q3.2. (5 points) How many overflow pages were created by the time the last record was inserted (which were subsequently discarded)?

## Q4. Extendible Hashing [15 points]

Q4.1. (10 points) Consider an extendible hash structure where buckets can hold up to three records. Initially the structure is empty and global depth is 2. Sketch the extendible hash structure after the records given in Table 1 in Q3 above (in the same order shown above) have been inserted. Assume that as mentioned in the textbook, the directory doubles in size at each overflow.

Note: Use the extendible hashing algorithm outlined in section 11.2 of the textbook.

Q4 .2 (5 points) How many buckets have the same local depth as the global depth of the file structure?

## Q5. External Sorting [25 points]

Suppose you have a file with 30,000 pages and 6 available buffer pages. Answer the following questions using the general external sorting algorithm outlined in section 13.3 of the textbook. Please write the formula you used in calculating the answers.

Q5.1. (15 points) How many runs will you produce in the third pass using the general external sorting algorithm? List the number of pages that are sorted in each pass.

Q5.2. (5 points) How many passes will it take to sort the file completely?

Q5.3. (5 points) What is the total I/O cost of sorting the file?