Database Systems, CSCI 4380-01
Homework # 7
Due Tuesday April 17, 2018 at 11 PM

Homework Statement.

This homework is worth 6% of your total grade. If you choose to skip it, Final Exam will be worth 6% more.

This homework is on B-trees and query processing. Please type your answers in a PDF document and submit on Gradescope. No handwritten homeworks please.

Question 1 (15 points). Suppose the usable space on a disk page is 8,000 bytes for indexing. This is the space filled with key values and pointers. A key value and pointer pair is called an entry. The pointers in a leaf node point to tuples in a relation. Pointers in the upper levels point to index nodes at levels below. Each index node is mapped to a disk page and all pointers are the same size: 16 bytes.

You are creating the following indices: index `i1` on `R(A)` and index `i2` on `R(A, B, C)` where attribute A is 4 bytes long, attribute B is 10 bytes long and attribute C is 8 bytes long.

First, compute the capacity of the nodes: maximum number of entries that can be stored in a node/disk page. Recall that each index node (internal or leaf) contains the same type of information and has the same capacity. You may have one extra pointer in each node, but we assume this is part of the header info that is not part of this computation. So you can safely disregard it. Assume each page contains about 75% of maximum number of entries (the root node or one of the nodes in each level may contain fewer nodes).

Given that `Tuples(R)=500,000`, compute the size of indices i1 and i2 (number of nodes at each level). Show your work.

Question 2 (35 points). You are given the following indices and information.

Index	Columns Indexed	Height	#nodes at leaf level
i1	R(A)	2	1,000
i2	R(A,B)	2	2,000
i3	R(C)	2	4,000

`Tuples(R)=800,000`, `Pages(R)=12,000.`

Query	Number of tuples that match
`Q1: select * from R where A=5`	200
`Q2: select * from R where B <= 20`	80
`Q3: select * from R where C='Pickle Rick'`	2,000
`Q4: select * from R where A=5 and B <= 20`	10
`Q5: select * from R where A=5 and C='Pickle Rick'`	120
`Q6: select * from R where B <= 20 and C='Pickle Rick'`	1
`Q7: select * from R where A=5 and B <= 20 and C='Pickle Rick'`	1

For each query, find the cost of answering the query in terms of the number of pages and total number of seeks for each and every one of the index combinations from the list below:

(1) i1 only (2) i2 only (3) i3 only (4) i1 and i3 together (5) i2 and i3 together.

Assume that all leaf nodes can be read sequentially. Internal nodes are read using random I/O. Assume the worst case scenario when matching tuples can be in many different pages. Relation R spans a total of 200 different cylinders, so reading R sequentially would require a total of 200 seeks.

You can exclude an index combination if it does not apply to this question (for example index on A will not help query Q3). Show your work.

For the remaining questions in this homework, you will compute query costs in terms of number of pages, disregarding whether it is random or sequential I/O. However, as an independent exercises, think about whether the disk operations involved in these operation are more likely to be random or more likely to be sequential I/O.

Question 3 (15 points). What is the cost of sorting relation `R` given `Pages(R)=12,000` if:

(a) `M=200`

(b) `M=40`

(c) `M=10`

Question 4 (15 points). What is the cost of block-nested loop join of `R` and `S` given `Pages(R)=12,000` and `Pages(S)=2,000` if:

(a). `M=201`

(b). `M=1,001`

(c). `M=2,001`

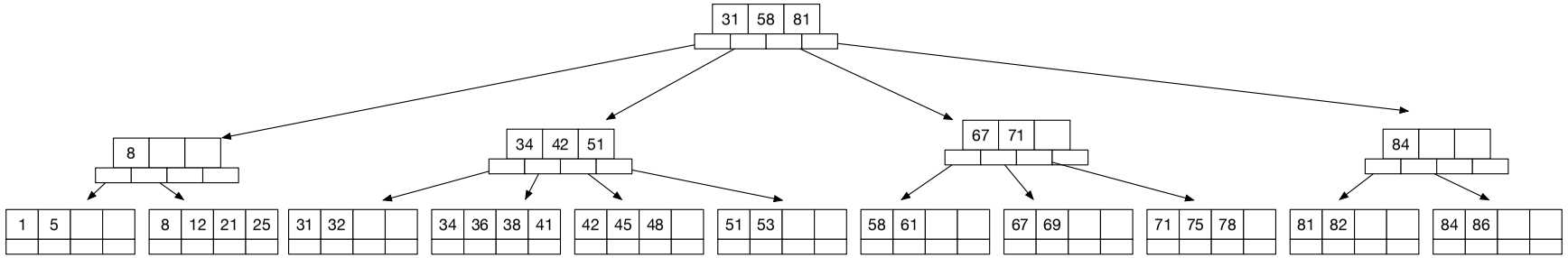
Question 5 (20 points). You are given the following B-tree of capacity 4 for an integer value that takes unique values. In this case, each leaf node stores between 2 to 4 key value and tuple pointer pairs. Each internal node stores between 2-4 pointers to nodes below, with key values to distinguish the division points.

Show the resulting B-tree after

(a) adding key values: 33, 15, 37 in this order to the original B-tree

(b) deleting key values: 69, 61, 84 in this order from the original B-tree

Draw the resulting B-tree. You can use the PDF for the given B-tree and make changes by drawing on it using a PDF source, or you can draw the changes and scanning it. We will only allow hand written components in this part of the homework.


31 58 81
8
34 42 51
67 71
84
1 5
8 12 21 25
31 32
34 36 38 41
42 45 48
51 53
58 61
67 69
71 75 78
81 82
84 86