CSCI 4380 Spring 2018 Quiz 5 Answers

Question (25 points). Suppose you are given the following B-tree indices, each indexing a relation `R(A,B,C,D,E,F)` with 200,000 tuples. In each index, assume that the leaf nodes are ordered consecutively on disk so they can be read sequentially from left to right following sibling pointers. In all other levels, you can assume nodes are accessed with random I/O.

Index	Columns Indexed	Height	#nodes at leaf level
i1	R(A)	3	1000
i2	R(B)	3	2000
i3	R(A,B)	4	3000

You are also given the following information, PAGES(R)=8,000, and:

Tuples that match the condition `A>20`: 20
Tuples that match the condition `B=5`: 100
Tuples that match the condition `A>20 and B=5`: 3

Question. What is the cost of answering the following query:

`Q1: select * from where A >20 and B=5`?

Please answer in the next page and show your work. You can assume that you have sufficient memory allocated to answer this query in one pass.

Answer. So, let's process the info first. B-trees map each node to a disk page. So,

index i1 has 1,000 pages at leaf, each storing 200,000/1,000=200 tuples/entries per page.

index i2 has 2,000 pages at leaf, each storing 100 tuples/entries per page.

index i3 has 3,000 pages at leaf, each storing 66 tuples/entries per page.

Question	Number of index pages read (and how many of them will be a random I/O costing a seek operation)	Number of relation `R` pages read (assume worst case scenario where each matching tuple is in a different disk page)
1(a) Answer `Q1` by scanning `R` only	0 (no index scan)	8,000 pages
1(b) Answer `Q1` scanning i1 first and read matching tuples from `R`	Scan 3 internal nodes/pages, and leaf nodes/pages for 20 tuples (1 or 2 nodes max, 20/200). Total=4 or 5 nodes/pages. Total seeks: 4.	Read matching 20 tuples, worst case 20 pages.
1(c) Answer `Q1` scanning i2 first and read matching tuples from `R`	Scan 3 internal nodes/pages, and leaf nodes/pages for 100 tuples (1 or 2 nodes max (100/100)). Total=4 or 5 nodes/pages. Total seeks: 4.	Read matching 100 tuples, worst case 100 pages.
1(d) Answer `Q1` scanning i3 first and read matching tuples from `R`	Scan 4 internal nodes and all leaf nodes for $A > 20$ (20 tuples which fit in 1-2 leaf nodes) , total: 4+1 (or 2) nodes/pages, seeks: 5 (assuming all of leaf can be read in a single sequential read.	Read matching 3 tuples from disk, worst case 3 pages.
1(e) Answer `Q1` scanning i1&i2, intersect matching tuple ids in memory, and read matching tuples from `R`	i1: 4-5 nodes, 4 seeks. i2: 4-5 nodes, 4 seeks (see parts (b),(c) above). Total 8-10 nodes/pages, 8 seeks.	Read matching 3 tuples from disk, worst case 3 pages.