

CompSci 316: Intro to Databases

Lab-2: Index

Score: 100

Posted on: Thu, 03/05/2020 (to be started in class)

Due Date: **Tue, 03/17/2020, 11:59 pm**

10% extra credit for finishing all questions correctly in class (last submission timestamp for all questions has to be by 4:20 pm).

- Consider a relation $R(\underline{id}, name, age)$. Id is the key.
- Consider a B+ tree index for R on “age”.
- Suppose each B+ tree node can hold 4 index entries (non-leaves) or data entries (leaves, store pointers to data tuples), i.e., max fan out $f = 5$.
- Suppose the height of the B+-tree is 3 (root, a level of non-leaf nodes, then the leaves).
- Suppose each node of B+ tree is stored in one page of disk.
- Suppose one data page for R can contain 8 tuples from R.
- Assume all the index pages (pages of the B+tree) and all data pages (pages to store tuples of R) are initially on disk.
- Remember that cost is I/O cost to read pages from disk to memory and to write to disk from memory if a page is edited. Each read/write = 1 unit of cost.

- Consider a query Q
SELECT *
FROM R
WHERE age ≥ 25 AND age ≤ 50

- Suppose R has 90 tuples, and 50 tuples satisfy the range of age between 25 and 50 in the above query Q.

Q1. What is the maximum number of index entries this B+tree index can hold at the non-leaf nodes (including the root)? Give the exact number. No explanations needed. (10 points)

Q2. How many index nodes can appear in the 2nd level? Give the exact number. No explanations needed.(10 points)

Q3: How many nodes can appear in the leaf level? Give the exact number. No explanations needed.(10 points)

Q4. What is the maximum number of data entries this B+tree index can hold at the leaf nodes? Give the exact number. No explanations needed.(10 points)

Q5. What is the cost of accessing the index pages to answer this query using this B+ tree? Briefly explain. (20 points)

Q6. What is the cost of accessing the data pages (storing R tuples) if the index is unclustered? Briefly explain. (20 points)

Q7. What is the cost of accessing the data pages (storing R tuples) if the index is clustered on age? Briefly explain. (20 points)