CMPE 180B Database Management Systems

Assignment: Index Size

Name: Sakshat Patil

Student ID: 018318287

You have a table with 2 Billion rows. The rows contain a string field called *'unique name'* and a few other fields. Given the enormous size of the database, you decide to build a B+ tree indices on the *'unique name'* field.

- Each page is exactly **4KB** (4096 Bytes)
- The size of each key in your B+ tree is **128 bytes**
- The size of each pointer in your B+ tree is **8 bytes**
- Your system has **48GB** of free RAM and **infinite** hard disk space
- The constant fan-out f of the B+ tree is **66**.

a) How many pages can we store in the first level? In the second level? In tenth level?

b) How many levels do we need in our B+ tree? Compute the space required by each index level

c) Assume that each level must either be completely on RAM or disk. Note that all data pages stay on the disk. What is the worst-case IO requirement

Answers:

1.
   a. For Level 0, one node with 65 keys and 66 pointers. Node Size = 65*128 + 66*8 Bytes = 8848 Bytes. Each page is 4KB = 4096 B, so one node needs ceil(8848/4096) = ceil(2.16) = 3 pages.
   b. For Level 1, we have 1 Node, each needing 3 pages.
   c. For Level 2, we have 66 Nodes, each needing 3 pages, totaling 198 pages.
   d. For Level 9, we have 66^9 Nodes, each needing 3 pages, for a total of 66^9 * 3 pages.
2.
   a. We need to determine how many levels our B+ tree requires to store 2 billion keys.
   b. For level L (the bottom level), the calculation works as follows:
   c. Each node contains 65 keys
   d. Level L contains 66^(L-1) nodes
   e. Total keys at level L = 66^(L-1) × 65
   f. For 2 billion keys: 66^(L-1) × 65 ≥ 2,000,000,000
   g. Solving for L: L > 5.12
   h. Therefore, we need at least 6 levels in our B+ tree to store all 2 billion keys.
3.
   a. To find a key in our 6-level B+ tree, the worst-case I/O analysis is:

b. Each node requires 3 pages (since node size is 8848 bytes and page size is 4096 bytes)
c. We need to traverse all 6 levels in the worst case
d. At each level, we must read all 3 pages of a node
e. Therefore: Worst Case I/O requirement = (6 levels × 3 pages per node) + final data I/O cost = 18 pages + I/O cost for retrieving the actual record