

ADVANCED DATA ENGINEERING: ASSIGNMENT 5

NGUYEN T. Hoang - SID: 15M54097

Fall 2015, W831 Tue. Period 5-6

Due date: 2015/11/24

Problem

Assuming:

- Average seek time of the disk is 2ms.
- Rotation Speed of the disk is 12,000 rpm.
- Data transfer bandwidth is 20 MB/s = 20,971,520 Bytes/s.
- A page size 4KB (4096 Bytes).
- Effective storage capacity (B) is 4000 Bytes.
- $k = 20$ Bytes, $t = 100$ Bytes, $p = 4$ Bytes, $u = 0.6$

Question 1

Calculate disk page access time for the disk.

Answer:

Assuming there is no controller overhead, and the disk is idle at the time of calculation (no queuing delay), also the transfer bandwidth is the actual data throughput. We have the following formula for the page access time:

$$\begin{aligned} T_{Page\ access} &= T_{seek\ time} + T_{rotational\ latency} + T_{data\ transfer} \\ &= 2 + \frac{60 \times 1000}{12,000 \times 2} + \frac{4096 \times 1000}{20,971,520} \ (ms) \\ &= 2 + 2.5 + 0.195 \ (ms) \\ &\approx 4.7 \ (ms) \end{aligned}$$

Question 2

Derive the maximum number of tuples for a B-Tree storing TIDs, of which height is 2 and 3.

Answer:

From the information given in the question, I derive the average number of entries in an index node F^* and the average number of entries in a leaf node C^* . In this case, F^* and C^* have the same value since B-Tree stores TIDs.

$$\begin{aligned} F^* = C^* &= \left\lfloor \frac{B}{k+p} \times u \right\rfloor \\ &= \left\lfloor \frac{4000}{20+4} \times 0.6 \right\rfloor \text{ (entries)} \\ &= \lfloor 99.6 \rfloor = 99 \text{ (entries)} \end{aligned}$$

Tree height is 2 (H = 2) Using the formula in the lecture note, I derive the maximum number of tuples in the database:

$$\begin{aligned} N_{H=2} &= C^* \times F^{*(H-1)} \\ &= 99 \times 99^{2-1} \text{ (entries)} \\ &= 9,801 \text{ (entries)} \end{aligned}$$

Tree height is 3 (H = 3) The maximum number of tuples in the database:

$$\begin{aligned} N_{H=3} &= C^* \times F^{*(H-1)} \\ &= 99 \times 99^{3-1} \text{ (entries)} \\ &= 970,299 \text{ (entries)} \end{aligned}$$

Question 3

Calculate access time to derive a TID of a tuple by the B-Tree stored into the disk for 500,000 tuples.

Answer:

The given number of tuples is $N = 500,000$. Therefore, the height of tree needed in this case is:

$$\begin{aligned} H &= 1 + \left\lceil \log_{F^*} \left(\left\lceil \frac{N}{C^*} \right\rceil \right) \right\rceil \\ &= 1 + \left\lceil \log_{99} \left(\left\lceil \frac{500,000}{99} \right\rceil \right) \right\rceil \\ &= 3 \end{aligned}$$

Since we store every node in a disk page, so a key search only need to access 1 node per 1 level of the tree. In this case, for a key search, we need 3 disk reads. In this calculation,

I do not take into account the time it needs for key-searching within a disk page. I also assume that every disk page read takes average page read time (calculated in question 1), although normally individual records are grouped together in a disk block, which reduces page read time in sequential read. The access time to derive a TID of a tuple is:

$$\begin{aligned}T_{TID} &= T_{Page\ access} \times H \\&= 4.7 \times 3\ (ms) \\&= 14.1\ (ms)\end{aligned}$$