

# CS351

## Section 1

### QUIZ 5 SOLUTIONS

December 19, 2010

We have 50,000 student records in a B+ tree. The degree of this tree is 5.

- 1) How many data nodes exist in this B+ tree?

$$bk = \frac{n}{2 \times d \times \ln 2} \Rightarrow \frac{50,000}{2 \times 5 \times 0.7} \cong 7,143 \text{ data nodes}$$

- 2) To determine the average GPA of students, how many number of data nodes accesses are needed?

It can be approximated by  $2 \times bk$ .  $2 \times bk = 14,286$  accesses (this approximation is accepted during grading).

However, this is not a good estimate since when we access an index node in the disk, it will be kept in main memory by the operating system as long as we need it. Recall that each index node has fan out ( $f_o$ ) number of data node pointers. As a result, we are able to access  $f_o$  number of data nodes after accessing each index node. By this observation, a much more accurate estimation for the number of disk accesses is  $\frac{bk}{f_o} + bk$ .

If we continue with Salzberg's number for  $f_o$ :

$$\frac{bk}{f_o} + bk = 52 + 7,143 = 7,195 \text{ accesses are needed where } f_o = \frac{2400}{(8+4)} \times \ln 2 = 140$$

Also refer to the solution of question 5 in homework 4 from the course web site!

# CS351

## Section 2

### QUIZ 5 SOLUTIONS

December 16, 2010

We have 100,000 student records in a B+ tree. The degree of this tree is 10.

- 3) How many data nodes exist in this B+ tree?

$$bk = \frac{n}{2 \times d \times \ln 2} \Rightarrow \frac{100,000}{2 \times 10 \times 0.7} \cong 7,143 \text{ data nodes}$$

- 4) To determine the average GPA of students, how many number of data nodes accesses are needed?

It can be approximated by  $2 \times bk$ .  $2 \times bk = 14,286$  accesses (this approximation is accepted during grading).

However, this is not a good estimate since when we access an index node in the disk, it will be kept in main memory by the operating system as long as we need it. Recall that each index node has fan out ( $f_o$ ) number of data node pointers. As a result, we are able to access  $f_o$  number of data nodes after accessing each index node. By this observation, a much more accurate estimation for the number of disk accesses is  $\frac{bk}{f_o} + bk$ .

If we continue with Salzberg's number for  $f_o$ :

$$\frac{bk}{f_o} + bk = 52 + 7,143 = 7,195 \text{ accesses are needed where } f_o = \frac{2400}{(8+4)} \times \ln 2 = 140$$

Also refer to the solution of question 5 in homework 4 from the course web site!

# CS351

## Section 3

### QUIZ 5 SOLUTIONS

December 30, 2010

We have 100,000 student records in a B+ tree. The degree of this tree is 20.

- 5) How many data nodes exist in this B+ tree?

$$bk = \frac{n}{2 \times d \times \ln 2} \Rightarrow \frac{100,000}{2 \times 20 \times 0.7} \cong 3,572 \text{ data nodes}$$

- 6) To determine the average GPA of students, how many number of data nodes accesses are needed?

It can be approximated by  $2 \times bk$ .  $2 \times bk = 7,144$  accesses (this approximation is accepted during grading).

However, this is not a good estimate since when we access an index node in the disk, it will be kept in main memory by the operating system as long as we need it. Recall that each index node has fan out ( $f_o$ ) number of data node pointers. As a result, we are able to access  $f_o$  number of data nodes after accessing each index node. By this observation, a much more accurate estimation for the number of disk accesses is  $\frac{bk}{f_o} + bk$ .

If we continue with Salzberg's numbers for  $f_o$

$$\frac{bk}{f_o} + bk = 26 + 3,572 = 3,598 \text{ accesses are needed} \quad \text{where } f_o = \frac{2400}{(8+4)} \times \ln 2 = 140$$

Also refer to the solution of question 5 in homework 4 from the course web site!