Shengtao Tony Hou

DSCI551

HW4

November 20th

Question 1:

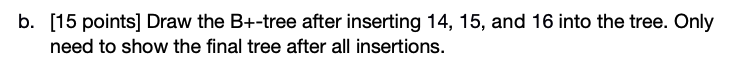
Diagram

Description automatically generated

Answer :

Read the root and since 43 is >= 35 we move to left child internal node. We will then be moving to fifth leaf using the fifth pointer since 35 >=19. Since the largest value in fifth leaf is 30 which is less than 35, we will return to root and move down to the right child internal node to continue searching. Since 35 is less than 56, we will be moving to the first leaf node under. Because 35 <= 43 we know that 43 will be the first value within our query condition. We will then read 43,44 and use the pointer to go to second leaf node, where we read 56,58, since they are all less than 65. We then use the pointer to move to third leaf node under right child internal node and found ourselves value 70 which is larger than our upper bound of 65. Therefore the process of finding keys stop and we have our results as 43,44,56,58

Total blocks I/O needed are 6 blocks.



Answer:

Diagram, engineering drawing

Description automatically generated



Answer:

Diagram, engineering drawing

Description automatically generated

2. [60 points] Consider natural-joining tables R(a, b) and S(a,c). Suppose we have the following scenario.

i. R is a clustered relation with 5,000 blocks. ii. S is a clustered relation with 20,000 blocks.

iii. 102 pages available in main memory for the join.  
iv. Assume the output of join is given to the next operator in the query

execution plan (instead of writing to the disk) and thus the cost of writing the output is ignored.

Describe the steps for each of the following join algorithms. For sorting and hashing- based algorithms, also indicate the sizes of output from each step. What is the total number of block I/O’s needed for each algorithm? Which algorithm is most efficient in terms of block’s I/O?

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

