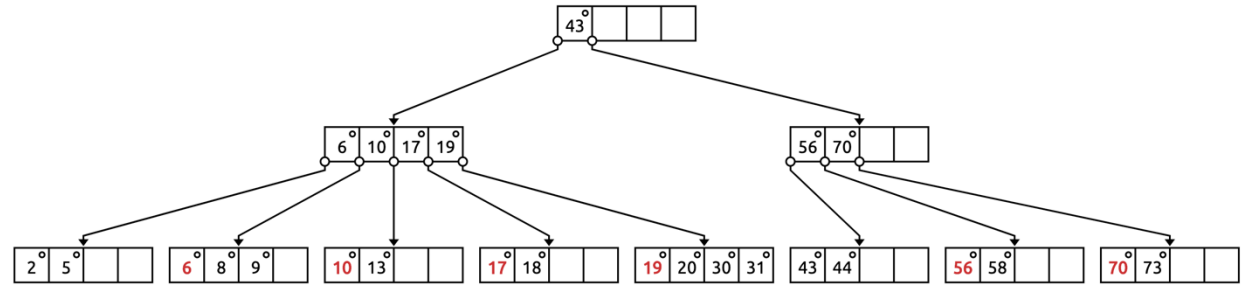
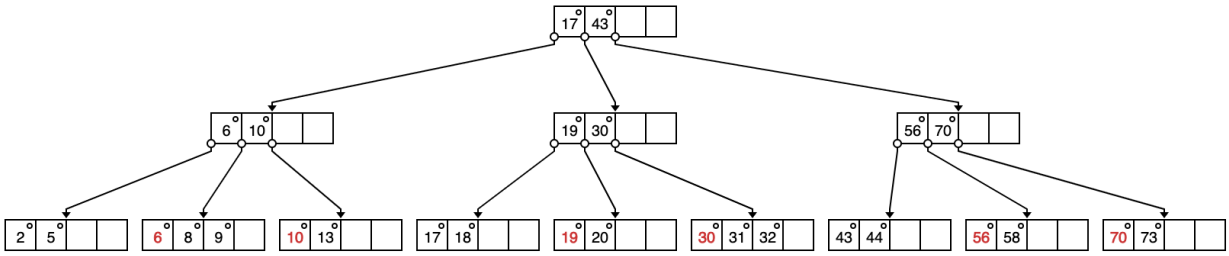
Phuong Ngo
Homework 4
INF551

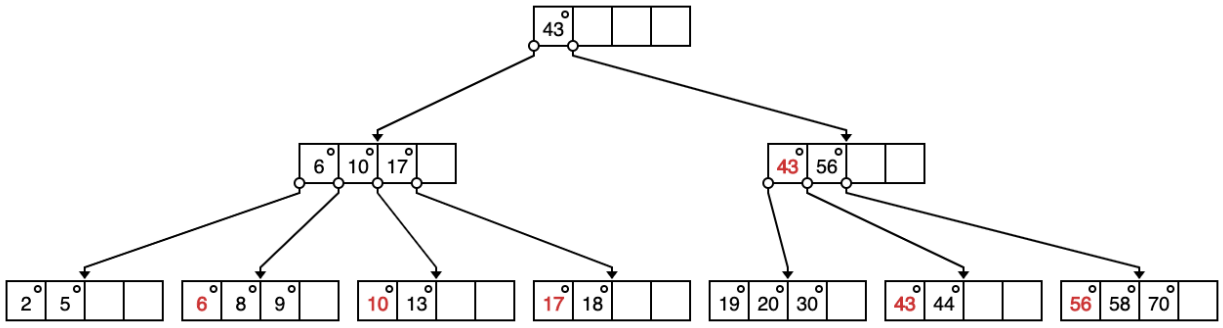
1. Answers
    a. First go to root, check its value and move to left child (1 block)
    Then find the first leaf that is greater or equal to 15, which is the right pointer of 10 (1 block) ; then traverse through all leaves on the right to find all records that satisfy the condition “age >= 15 and age <= 45”, which is 17, 18, 19, 20, 30, 43, 44 (4 blocks)
    Total I/O’s: 1 root + 5 leaves = 6 blocks.
    b. After inserting 31



After inserting 32



    c. After deleting 73 from original tree



2. Answers
    a. Input: (M-2) = (102-2) =100 blocks of R and S buffer with size of B(S) = 10000. For each block in R, iterate through every block in S; iterate through each tuple in R block and S block, if condition matches, return (r,s). Run this algorithm B(R)/(M-2) times.
    Total cost = B(R) + B(R)B(S)/(M-2) = 5000 + (5000 * 10000)/(102-2) = 505,000
    b. Input: (M-2) = (102-2)=100 blocks of S and R buffer with size of B(R) = 500. For each block in S, iterate through every block in R; iterate through each tuple in S block and R block, if condition matches, return (s,r). Run this algorithm B(S)/(M-2) times.
    Total cost = 10000 + (5000 * 10000)/(102-2) = 510,000
    c. First use 100 buffers to sort S into 100 runs with 100 blocks each. Same thing for R with 50 runs of 100 size blocks. Cost 2B(R) + 2B(S) = 30,000. Total number of runs = 150 > M-1 = 100. We can't merge them directly.
    We then merge 100 runs into 2 sorted run of size 5000 blocks each for S. 2B(S)=2*10,000=20,000
    Finally merge, cost = B(R) + B(S) = 15,000
    Total cost = 30,000 + 20,000+ 15,000 = 65,000
    d. First completely sort R and S, this cost 4B(R) + 4B(S); then read two sorted relations and match tuples, this cost B(R) + B(S).
    total cost = 5B(R) + 5B(S) = 5 * 5000 + 5 * 10000 = 75,000
    e. First hash R into M-1 = 100 buckets, each bucket has a size of 50, then perform duplicate elimination on it and write them back to disk, this cost 3B(R). Hash S using the same method into 100 buckets, each bucket has a size of 100 then perform duplicate elimination on it and write them back to disk, this cost 3B(S); then join every corresponding bucket pair.
    total cost = 3B(R) + 3B(S) = 3 * 5000 + 3 * 10000 = 45,000
    f. Iterate over R and for each tuple, fetch corresponding tuples from S.
    total cost = B(R) + T(R)B(S)/V(S,a) = 5000 + 100,000*10000/10 = 250,500

    Partition-hashed join is the most efficient