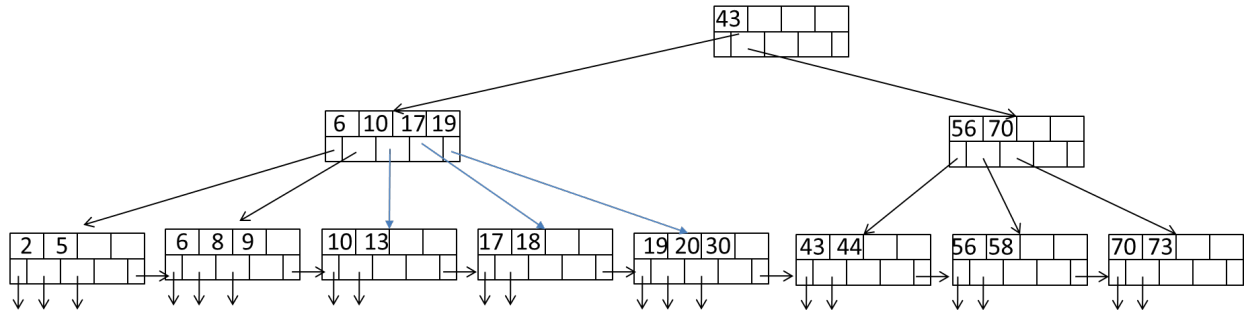# Homework #4

Due: April 20, Monday
100 points

1. [40 points] Consider the following B+tree for the search key "age. Suppose the degree d of the tree = 2, that is, each node (except for root) must have at least two keys and at most 4 keys.

a. Describe the process of finding keys for the query condition "age >= 15 and age <= 45". How many blocks I/O's are needed for the process?
b. Draw the B+tree after inserting 31 and 32 into the tree.
c. Draw the tree after deleting 73 from the original tree.

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 and 100,000 tuples
  ii. S is a clustered relation with 10,000 blocks and 200,000 tuples
  iii. S has a clustered index on the join attribute a
  iv. V(S, a) = 10 (recall that V(S, a) is the number of distinct values of a in S
  v. 102 pages available in main memory for the join.
  vi. 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 (including input, output, and their sizes at each step) for each of the following join algorithms. What is the total number of block I/O's needed for each algorithm? Which algorithm is most efficient?

a. Nested-loop join with R as the outer relation
b. Nested-loop join with S as the outer relation
c. Sort-merge join (assume only 100 pages used for sorting and 101 pages for merging). When the number of runs of a relation is too large for merging, the runs will be further merged first. Select the relation with a larger number of runs for further merging if both have too many runs.
d. Simple sort-based join (using the same assumption on # of pages used for sorting and merging as above)
e. Partitioned-hash join (assume 101 pages used in partitioning of relations and no hash table used for lookup in joining buckets)

f.   Index join (ignore the cost of index lookup)

1. Please conclude your answers in a pdf file with the naming convention:

  <Firstname>_<Lastname>.pdf.

2. For question 1.b and 1.c, you can draw them and take a picture, but make sure your photo is clear enough (Drawing by computer is preferred. You can use online tool like this https://projects.calebevans.me/b-sketcher/). For other questions, please type your answers, which should be searchable and copyable, not an image.

3. Your answers should be in the same order as the order of questions. For each question, if you provide multiple answers, we will only grade the first one.