

CMU - SCS
15-415/15-615 Database Applications
Spring 2013, C. Faloutsos
Homework 4: Query processing
Released: Thursday, 02/21/2013
Deadline: Thursday, 02/28/2013

Reminders - IMPORTANT:

- Like all homework, it has to be done **individually**.
- Please submit a **hard copy of your typeset answers, in class**, on Thursday, 02/28/2013 1:30pm.
- **Separate your answers:** As before, for ease of grading, please print each of the three questions on a **separate** page, i.e., three pages in total for this homework. If you need more pages for one problem, please staple them together. Type your **name and andrew ID on each** of the three pages. FYI, we will have three labeled piles at the front of the classroom, one for each problem.

Reminders - FYI:

- Weight: 5% of homework grade.
- The points of this homework add up to 100.
- *No* need to justify your answers, unless explicitly required.
- Rough time estimates: 1-3 hours.

Question 1. B+ trees (40 points) [Kate] SUBMIT ON A SEPARATE PAGE

For the following sub-questions, consider the B+ tree structure with order $d = 2$ (i.e. there are at most 4 keys per node, and at most 5 pointers to children) from the textbook - Figure 10.29, p. 367, R+G, 3rd edition¹, as shown in Figure 1.

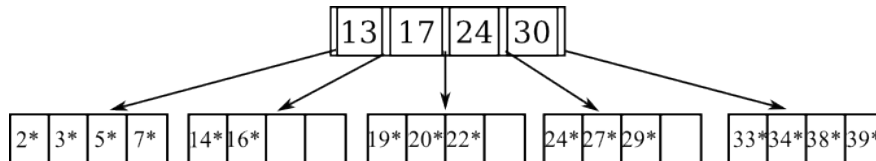
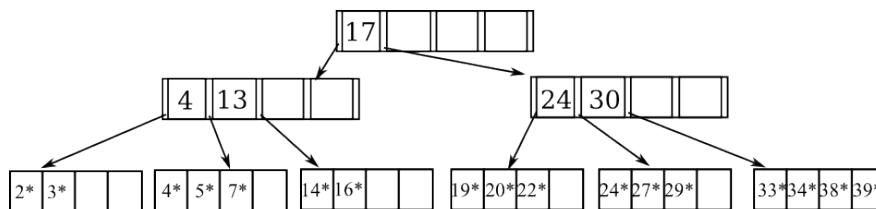
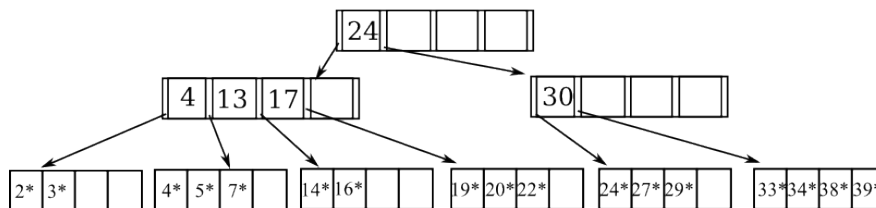


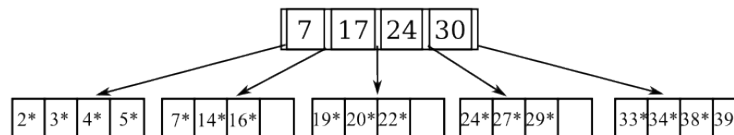
Figure 1: Copy of Figure 10.29, p 367.



Tree A.



Tree B.



Tree C.

Figure 2: Figure for questions 1.2 and 1.3.

¹the solutions to the odd-numbered questions at the end of each chapter are publicly available here: <http://pages.cs.wisc.edu/~dbbook/openAccess/thirdEdition/solutions/ans3ed-oddone.pdf>

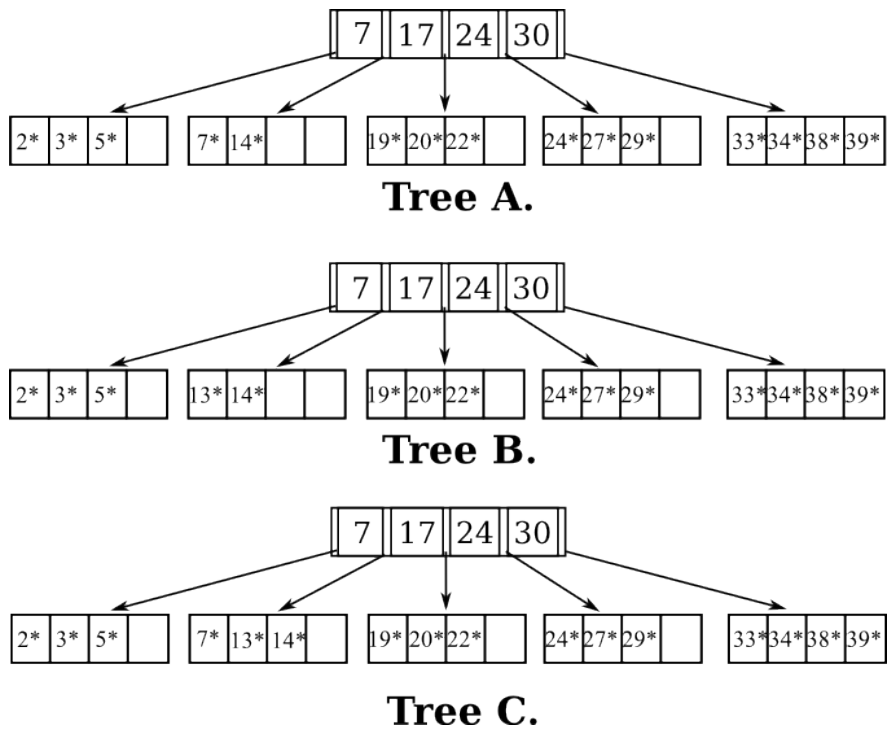


Figure 3: Figure for question 1.4.

- Q 1.1.** Assume the structure in Figure 10.29 is not a B+ tree, but an ISAM structure. Show the new structure after inserting keys 35 and 37. [10 points]
- Q 1.2.** Consider the Figure 10.29 as a B+ tree. Starting from the original B+ tree in Figure 10.29, we insert the record 4* with a key 4. From the choices in Figure 2, which is the resulting tree if we use the default “no redistribution” strategy: A, B, C, or neither? If neither, draw the correct tree. [10 points]
- Q 1.3.** Starting from the original B+ tree in Figure 10.29, which is the resulting tree if we insert the record 4* with key 4 using the B* tree “redistribution with a sibling” strategy: A, B, C, or neither? If neither, draw the correct tree. [10 points]
- Q 1.4.** Starting from the original B+ tree in Figure 10.29, we delete the record 16*. From the choices in Figure 3, which is the resulting tree if we borrow a node from the left sibling: A, B, C, or neither? If neither, draw the correct tree. [10 points]

Question 2. Extendible Hashing (35 points) [Kate] SUBMIT ON A SEPARATE PAGE

Assume we have the following records where we indicate the hashed key in parenthesis (in binary):

i [001100]
h [001100]
g [101101]
f [010010]
e [111111]
d [010010]
c [100001]
b [001100]
a [000000]

Consider an Extendible Hashing structure where buckets can hold up to three records. Initially the structure is empty (only one empty bucket). Consider the result after the records above have been inserted in the order shown, using the lower-bits for the hash function. As mentioned in the textbook, assume that the directory doubles in size at each overflow.

- Q 2.1.** What will be the global depth of the resulting directory? [5 points]
- Q 2.2.** How many buckets will we have? [5 points]
- Q 2.3.** OPTIONAL - zero points (but it may help you answer the questions): Show the resulting table in a format similar to figure 11.15 on page 376 from the textbook. [0 points]
- Q 2.4.** List all the elements in the bucket which contains the element “i.” What is the local depth of this bucket? [5 points]
- Q 2.5.** List all the elements in the bucket which contains the element “c.” What is the local depth of this bucket? [5 points]
- Q 2.6.** Do we store the number of bits to use in the hash function in: [5 points]
- a) the global depth, or
 - b) local depth?
- Q 2.7.** Consider the case that the directory just doubled. Is it true that every bucket will be split in two? (Yes/No) [5 points]
- Q 2.8.** If the local depth of a bucket is equal to the global depth of the directory, is this bucket pointed to by: [5 points]
- a) exactly one directory entry, or
 - b) multiple directory entries?

Question 3. External sorting (25 points) [Kate] SUBMIT ON A SEPARATE PAGE

Suppose you have a file with $N = 10^6$ pages and you have $B = 17$ buffer pages.

Q 3.1. How many runs will you produce in the first pass? [5 points]

Q 3.2. How many passes will it take to sort the file completely? [10 points]

Q 3.3. How many buffer pages do you need to sort the file completely in just two passes? [10 points]