Carnegie Mellon University

15-415 Database Applications

Spring 2012, Faloutsos

Assignment 4: Query Processing

Due: 3/1, 1:30 pm, in class – **hard copy**

Reminders

- Weight: **5%** of the homework grade.
- The points of this homework add up to **100**.
- Lead TA: Bin Fu (binf@andrew.cmu.edu).
- Rough time-estimates: **1~3 hours**.
- Justifying your answer is optional, and will only be used to your advantage, that is: Full score for the correct final answer. Any justification will only be used for partial credit, if the final answer is wrong.
- Drawing and equations may be hand-drawn, but make sure they are neat and legible. Please **type** everything else.
- The textbook referred to in the homework is Database Management Systems by Ramakrishnan and Gehrke, 3rd edition.

Question 1: ISAM and B+ Trees [30 points]

For the following 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) in the textbook (Figure 10.27, p. 365, R+G, 3rd edition).

[Q1.1] Assume the structure in Figure 10.27 is not a B+ tree, but an ISAM structure. Show the new structure after inserting keys 4 and 3. [**10 points**]

[Q1.2] For Q1.2 and Q1.3 consider the Figure 10.27 as a B+ tree. Starting from the original B+ tree in Figure 10.27, show the new B+ tree after inserting key 0. Show the results on the following two insertion strategies: a) No redistribution. b) Redistribution with a sibling.
[**10 points**]

[Q1.3] Starting from the original B+ tree in Figure 10.27, show the new B+ tree after deleting key 10. Show the results on the following two deletion strategies: a) Merging with its right sibling. b) Redistribution with the left sibling. [**10 points**]

Question 2: Hashing [40 points]

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

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

[Q2.1] Consider an extendible hashing structure where buckets can hold up to three records. Initially the structure is empty (only one empty bucket). Show the extendible hashing structure after these records (in the order shown above) have been inserted. Assume that as mentioned in the textbook, the directory doubles in size at each overflow. [**20 points**]

[Q2.2] For the same records above, consider a linear hashing structure where buckets can hold up to three records. Initially the structure is empty (only one empty bucket). Show the linear hashing structure after these records (in the order shown above) have been inserted. Assume that we split whenever the new key goes into a full bucket (which may or may not be the split bucket). [**20 points**]

Question 3: External Sorting [30 points]

In this problem you will estimate the running time to sort a file with $N = 10^6$ pages. Assume that we have a disk with an average seek time of 10ms, average rotation delay of 5ms and a transfer time of 1ms for each page. Assume the cost of reading/writing a page is the sum of those values (i.e. 16ms) for Q3.1 and Q3.2.

[Q3.1] Find the total running time using the two-way external merge sort (with three buffer pages). Follow the analysis and formulas in chapter 13.2 (p.423). [**5 points**]

[Q3.2] Find the total running time using 255 input buffer pages and 1 output buffer page (B = 256). Follow the analysis and formulas in chapter 13.3 (p.424). [**10 points**]

[Q3.3] The basic idea of **blocked I/O** is to read/write multiple pages (which we call a *buffer block*) at a time while performing sorting. In this case, the cost of reading/writing a buffer block of b pages is 1 random disk access for the first, and b-1 sequential disk accesses for the rest. Thus, the cost is the average seek time plus the average rotation delay (to locate the first page) plus 1ms per page (data transfer). Assume each buffer block consists of 32 pages (b = 32), and we use 7 input buffer blocks and 1 output buffer block (so still B = 256), find the total running time. Follow the analysis and formulas in chapter 13.4.1 (p.430). [**15 points**]