Exercise 1 “Implementation of Databases”

Due until 31 Oct, 2007 (before exercises)  
WS 07/08

Exercise 1.1: [Terminology]  
(5 pt.)
Define shortly in your own words the following terms:

- Database, Database Management System, and Database System
- Physical Data Independence and Logical Data Independence
- Block and Page
- Asynchronous I/O, Blocking I/O, and Block I/O
- Access Path
- Query
- Transaction

Exercise 1.2: [B⁺-tree]

1. Let’s assume for the rest of this sub-task the branch factor (the number of children) of any internal node is fixed to $b$, including the root node.

   (a) What is the time complexity of search in $B⁺$-tree, considering only reading of internal nodes?  
      (1 pt.)

   (b) Suppose database perform I/O in the unit of block, we can choose an optimal branch factor so that the I/O efficiency of key search in the index is maximized. Given the block size $B$, key size $K$, pointer size $P$, how do you determine the optimal branch factor $b$?  
      (3 pt.)

   (c) What is the difference between B-tree and $B⁺$-tree? Show that $B⁺$-tree is more efficient.  
      (1 pt.)

2. Let $k=1$ and $k*=2$. Given the $B⁺$-tree as in Figure 1, Perform the listed operations on the $B⁺$-tree and redraw the resulting $B⁺$-tree in a template as in Figure 2 after each step:  
   (4 pt.)
   (a) insert 14  
   (b) insert 20  
   (c) insert 25  
   (d) insert 40  
   (e) insert 21  
   (f) insert 36  
   (g) remove 5  
   (h) remove 14  
   (i) remove 25  
   (j) remove 21  
   (k) insert 4  
   (l) insert 15  
   (m) remove 40
Exercise 1.3:[Buffer Management] Disk access is magnitudes slower than memory access, which makes it the common bottleneck for database applications. However, utilizing buffer pools can alleviate this problem, with asynchronous I/O, prefetch and growing size of available memory. For instance, writing to disk can be totally absorbed by using buffer.

1. Modern operating systems all possess the ability to do buffering. But almost all DBMS's implement their own buffering. Give a reason why it is good to let DBMS in charge of these tasks. (2 pt.)

2. While all I/O’s go through a buffer, long processing threads usually copy their own pages out of the buffer pool. Explain the benefits in one sentence. (2 pt.)

3. If the average I/O consuming rate of DBMS is much larger than disk reading rate, buffering cannot help much. What would you do in such case? (2 pt.)