

# Exercises and hand-ins

## Advanced database technology

March 4, 2004

1. Consider a database that receives a sequence  $x_1, x_2, \dots, x_N$  of integers, and has to compute *statistics*, i.e., compute how many occurrences of each integer it has seen. For example, after seeing the sequence (11, 87, 33, 87, 22, 11, 11) it should be able to report the counts (11, 3), (22, 1), (33, 1), (87, 2). The goal of this problem is to investigate the I/O performance of different approaches with respect to:
  - The average processing time for the  $N$  integers.
  - How long it takes to report the counts when the sequence stops.

Use  $Z$  to denote the number of distinct integers in  $x_1, x_2, \dots, x_N$ . Analyze and compare the I/O performance of the following methods:

- (a) Initially write all  $N$  integers to disk. To compute the counts, sort the integers and scan to do the counting and reporting.
  - (b) Keep a sorted list of counts on the disk. New integers are initially stored in internal memory, but whenever we have seen  $M$  new integers (and thus filled memory), or if the sequence has ended, we update the list of counts to include the  $M$  new integers (by scanning). Reporting is trivially done by scanning.
  - (c) Insert the integers one by one in a B-tree, where leaves store integers and their current counts. At the end traverse the leaves to report the counts.
2. Suppose we want to search among  $N$  keys, that internal memory can hold  $M$  keys/pointers, and that a disk block can hold  $n$  keys/pointers. Further, suppose that the *only way* of accessing disk blocks is by following pointers. Show that a search takes at least  $\lceil \log_n(N/M) \rceil$  I/Os in the worst case. **Hint:** Consider the size of the set of blocks that can be accessed in at most  $t$  I/Os.

**The following exercise is to be handed in at the latest March 18 at 10.00 AM.**

Problem 1 from the ADBT exam, June 2003. (See News section on home page.)

Hand in before the lecture on that day, or earlier in our mail boxes in the information office.