Exercises and hand-ins

Advanced database technology

March 20, 2003

Hand-in

To be handed in at the latest March 27 at 1.00 PM.

Suppose we have relations R_1, R_2 and R_3 with common attributes A (appearing in R_1 and R_2) and B (appearing in R_2 and R_3). Tuples in all relations have fixed length, and attribute A occupies 5% of the total tuple length. The relations occupy $B(R_1) = 8.000$, $B(R_2) = 10.000$ and $B(R_3) = 90.000$ blocks on disk, respectively. Values in attribute A are uniformly distributed in the domain $\{1, \ldots, 20\}$, and values in attribute B are uniformly distributed in $\{1, \ldots, 30\}$. Consider the following relational algebra expression:

$$\delta(\pi_A((\sigma_{A \geq 11}(R_1)) \bowtie (\sigma_{B=4}(R_3)) \bowtie (\sigma_{A \leq 15}(R_2)))).$$

1. Estimate the sizes of all subexpressions using the formulas from the lecture. (Ignore that you might be able to make a better estimate!)
2. Using these estimates, apply dynamic programming (Selinger-Style Optimization, see GUW page 845) to find the best physical query plan:
   - Determine the order of joins.
   - Determine the algorithms used for all operations. Assume that there is memory for either a two-pass sorting based join using $5(B(R_i) + B(R_j))$ I/Os to join R_i and R_j, or a two-pass hash join using $3(B(R_i) + B(R_j))$ I/Os to join R_i and R_j.
   - Determine where to use pipelining. Assume that there are 10 extra memory buffers available for pipelining purposes.
3. Suppose that $R_1 = R_2 = R_3$ (forgetting the information above on their different attribute sets and sizes). Suggest a better algebraic expression for $\delta(\pi_A((\sigma_{A \geq 11}(R_1)) \bowtie (\sigma_{B=4}(R_3)) \bowtie (\sigma_{A \leq 15}(R_2))))$. Can you state the general rewriting rule you used?

Exercises for discussion on March 27

Last week's exercises and hand-in will be discussed from 11.00 to 12.00.