# Exercises *Databases*
# Session 7: query processing

Hans Philippi, Lennart Herlaar, Ad Feelders

March 24, 2017

## query processing

### Exercise 41

We are going to take a look at the block-nested-loop algorithm to compute an equijoin for relations $R$ and $S$, with sizes $B(R)$ and $B(S)$.
Let us first consider the case where buffer memory is large enough to read at least one relation completely into main memory.
Ignore the cost of writing the result tuples.

(i) How many block IO's does it cost to calculate the join?

(ii) Could a better result be obtained by any other method to calculate the results of a join?

Now suppose that the relations do not fit together in main memory. The approach is that we start with $M-1$ blocks from $S$ and use the remaining block of the bufferspace to iterate over $R$.

(iii) Explain that the number of required block IO's is $B(S) + B(R) * \lceil \frac{B(S)}{M-1} \rceil$

(iv) Knowing the number of blocks of the two relations, do we have a preference for which relation should play the role of $S$, i.e. the outer relation?

### Exercise 42

In chapter 5.2.7, you will find the definition of an outer join.

(i) Express the outer join in SQL without making use of the [LEFT|RIGHT] OUTER JOIN construct.

(ii) Describe how the algorithm of the sort-merge join could be adapted to calculate an outer join.

**Exercise 43**

The query optimizer will make use of database statistics to estimate the costs of different algebraic operations. Interesting figures are:
$T(R)$ = the number of tuples in $R$,
$B(R)$ = the number of blocks in $R$,
$V(R, A)$ = the number of different values in the $A$-column of $R$.
One of the problems is to estimate how an algebraic operator will affect these statistics. Try to find answers for the following operations. If necessary, try to distinguish several variations of the operator.

- selection (one condition, more conditions, $A = c, A \leq c, A <> c$
- projection (set / bag)

**Exercise 44**

Give a counterexample for the commutativity of the minus.

**Exercise 45**

Describe a situation in which 'applying selections early' is a bad strategy.