MAS 201, Winter 2016, Homework 2b

1 Algebra

Consider three relations R(A;B;C), S(A;D), W(B;E) and the query
SELECT *
FROM R, S, W
WHERE R.A=S.A AND R.B=W.B AND R.C=1

Consider an optimizer that produces all join expressions, where the selection $\sigma_{R.C=1}$ is pushed down and applied directly on the table R. Plans cannot have cartesian products or trivial natural joins that are equivalent to cartesian products.

Write all possible logical query plans (i.e., algebraic expressions) that this optimizer will produce but do not write twice expressions that can be derived from each other just by using the commutativity of the join.

2 Size Estimation

Assume that the optimizer selects the best logical query plan by estimating the size of each intermediate result (i.e., of each subexpression) and selecting the plan that has the smallest sum of intermediate result sizes. For each of the logical query plans you provided earlier, calculate the size of each intermediate result and decide which is the best plan. Note that you do not have to estimate the size of the end result since it should always be the same.

Use the following data for your estimates. Write whether your estimate is the precise number, i.e., whether the number will be the same for all possible databases that conform to the following characteristics.

- B is a key of R and W.B is a non-null foreign key that references R.B
- A is a key of S and R.A is a non-null foreign key that references S.A
- T(R) = 10^6
- T(S) = 10^5
- T(W) = 10^3
- V (R.C) = 10^2

3 Rewritings

For each of the following declare whether the rewriting is true for all R and S. If not, provide counterexample. If yes, you do *not* need to provide proof. You lose 6 points for each wrong answer, you gain 5 points for each correct answer. So, do not gamble on the answers!

1. **5 points for correct, -6 for wrong**

$$\delta\gamma_{A;sum(B)\mapsto M}R = \gamma_{A;sum(B)\mapsto M}\delta R$$

2. **5 points for correct, -5 points for wrong**

$$R - (S \cup T) = (R - S) - T$$