Implementation of DBMS
Exercise Sheet 11
Klingemann, WS 2023 / 2024

1) We assume in this task that a projection (like in SQL) does not remove duplicates. Give an example to show that projection cannot be pushed below set (no duplicates!) union. E.g., give relations R and S such that $\pi_A(R \cup S) \neq \pi_A(R) \cup \pi_A(S)$

2) Some familiar laws also apply for variants of joins, others do not. Tell, whether each of the following is true or not. Condition C involves only attributes of R. Give either a proof that the law holds or a counterexample.

a) $\sigma_C(R \bowtie S) = \sigma_C(R) \bowtie S$

b) $(R \bowtie S) \bowtie T = R \bowtie (S \bowtie T)$

Note, that $\bowtie$ means the outerjoin (similar to the ordinary inner join but we also add for each relation the tuples that do not find a match in the other relation).

3) Below are some statistics for four relations W, X, Y and Z.

W(a, b)	X(b, c)	Y(c, d)	Z(d, e)
T(W) = 100	T(X) = 200	T(Y) = 300	T(Z) = 400
V(W, a) = 20	V(X, b) = 50	V(Y, c) = 50	V(Z, d) = 40
V(W, b) = 60	V(X, c) = 100	V(Y, d) = 50	V(Z, e) = 100

Estimate the number of tuples of the relations that are the result of the following expressions:
a) $\sigma_{a=10}(W)$
b) $\sigma_{c=20}(Y)$
c) $W \times Y$
d) $\sigma_{d>10}(Z)$

4) Consider a query optimizer that uses statistical data. In particular, the following information is known about an attribute A of relation R. Attribute A is of type integer. Make the best use of the given the information.

- There are 100 tuples with A values between 1 and 10. In this range, there are 8 unique A values.
- There are 200 tuples with A values between 11 and 20. In this range, there are 5 unique A values.
- There are 300 tuples with A values between 21 and 30. In this range, there are 10 unique A values.
- There are 400 tuples with A values between 31 and 40. In this range, there are 10 unique A values.

a) Consider the query $\sigma_{A=7}(R)$. How many tuples are expected in the answer, assuming values are uniformly distributed over possible V(R, A) values?
b) Consider the query $\sigma_{A=17}(R)$. How many tuples are expected in the answer, assuming values are uniformly distributed over possible domain values?