

Implementation of DBMS

Exercise Sheet 11, Solutions

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 $\pi_A(R \cup S) \neq \pi_A(R) \cup \pi_A(S)$

Solution:

We can use the different handling of duplicates by the two operators to create an example in which the two expressions produce different result relations. Consider:

R:

A	B
1	2

S:

A	B
1	3

Then we have

$R \cup S$:

A	B
1	2
1	3

$\pi_A(R \cup S)$:

A
1
1

On the other hand we get:

$\pi_A(R) \cup \pi_A(S)$:

A
1

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.

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).

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

Solution:

This law does not hold. This results from the fact that tuples which do not fulfil the condition of a selection are simply gone whereas the outerjoin will produce tuples even if we do not have a matching tuple in the other relation. Consider the following example with the selection condition C being $X = Y$:

R:

X	Y
1	2

S:

Y	Z
2	3

Then we have

$R \bowtie S$:

X	Y	Z
1	2	3

$\sigma_{X=Y}(R \bowtie S)$:

X	Y	Z
no tuple		

On the other hand we get:

$\sigma_{X=Y}(R)$:

X	Y
no tuple	

$\sigma_{X=Y}(R) \bowtie S$:

X	Y	Z
null	2	3

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

Solution: This law does not hold, either. Consider:

R:

A	B
1	1

S:

C	D
2	2

T:

B	C
3	3

Then we have

$R \bowtie S$:

A	B	C	D
1	1	2	2

$(R \bowtie S) \bowtie T$

A	B	C	D
1	1	2	2
null	3	3	null

On the other hand we get:

$(S \bowtie T)$:

B	C	D
null	2	2
3	3	null

$R \bowtie (S \bowtie T)$:

A	B	C	D
1	1	null	null
null	null	2	2
null	3	3	null

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) $\sigma_{d>10}(Z)$

Solutions:

- a) $T(W) / V(W, a) = 100 / 20 = 5$
- b) $T(Y) / V(Y, c) = 300 / 50 = 6$
- c) $T(W) T(Y) = 100 * 300 = 30000$
- d) $T(Z) / 3 = 400 / 3 = 133 + 1/3$

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?

Solutions:

a) Tuples with an A-value of 7 can only be found among those tuples which have A-values in the range between 1 and 10. Therefore, we can apply our estimate to this subset of the tuples. As we assume that 7 is one of the 8 values occur in this range for attribute A, the estimate is $100 / 8 = 12.5$.

b) Similar to a), we just consider tuples in the relevant value range. However, this time, we assume that 17 is some value from the domain, i.e., an integer value between 11 and 20. Therefore our estimate is $200 / 10 = 20$.