IN3020&IN4020 – Database Systems Exam, 2024

30 May, 9:00am–1:00pm (Inspera)

Question 1 of 12 (8%)

Consider the following relations:

Table1:

A	B	C
NULL	1	1
NULL	1	2
1	*NULL*	3
1	*NULL*	4

Table2:

D	E	F
NULL	1	1
NULL	1	2
1	*NULL*	3
1	*NULL*	4

Evaluate the following (bag) RA expression over these relations:

$$\gamma_{A,\mathsf{COUNT(B)}\to C1,\mathsf{COUNT_DISTINCT}(B)\to C2,\mathsf{COUNT}(C)\to C3,\mathsf{COUNT_DISTINCT}(C)\to C4}(\delta(\mathsf{Table1} \ltimes \mathsf{Table2}))$$

Note: you do not need to give any intermediate steps, only the final result is enough.

IN3020&IN4020 – Database Systems Exam, 2024

30 May, 9:00am–1:00pm (Inspera)

Question 2 of 12 (8%)

Consider relation schema $\mathsf{R}[A]$. Write a relational algebra expression that outputs a relation with one attribute B of type string that contains one tuple $\langle$'Empty'$\rangle$ if the input R is empty and one tuple $\langle$'NonEmpty'$\rangle$ otherwise.

Note: in this and the following questions you may introduce and then use reasonable sequences of symbols for RA operations; for example, you can first say "I use X| for right semi-join below" and then use X| instead of $\rtimes$.

IN3020&IN4020 – Database Systems Exam, 2024

30 May, 9:00am–1:00pm (Inspera)

Question 3 of 12 (8%)

Consider relations with schemas $\mathsf{Table1}[A, B]$ and $\mathsf{Table2}[C, D]$. Compile the following SQL query to a (logical) query plan:

```
SELECT A FROM Table1 WHERE
        NOT EXISTS (SELECT * FROM Table2 WHERE A = C);
```

Note: it is only necessary to write any query plan that does the same as the query; in particular, you do not need to write any intermediate steps of compilation and you do not need to follow any specific compilation algorithm.

IN3020&IN4020 – Database Systems Exam, 2024

30 May, 9:00am–1:00pm (Inspera)

Question 4 of 12 (8%)

Recall that, in our cost function, we estimated the number $Tups(\sigma_{Cond}(\mathsf{R}))$ of tuples, for a relation R with an attribute A and a condition $Cond$ of the form $A < c$ with a constant c, in two ways: $Tups(\mathsf{R})/3$ and $Tups(\mathsf{R})/2$. Explain (in at most a few sentences) the rationale under both options.

IN3020&IN4020 – Database Systems Exam, 2024

30 May, 9:00am–1:00pm (Inspera)

Question 5 of 12 (8%)

Consider a relation with only one attribute and assume that, when stored, the relation is ordered (by the only attribute). In this case, a key-value pair for an index of any type takes more space than the tuple it refers to. Can it still be beneficial to use an index of some type rather than binary search over the relation itself? (Here, 'beneficial' means that the number of block reads is smaller on average.) Explain your answer (try to be as concise and to the point as possible).

IN3020&IN4020 – Database Systems Exam, 2024

30 May, 9:00am–1:00pm (Inspera)

Question 6 of 12 (10%)

Recall that we have a rule of thumb that pushing selection through natural join (if possible) usually gives a smaller logical plan cost. Give an example when it is not the case—that is, give a condition *Cond* and relations R1 and R2 such that

- the plan $\sigma_{Cond}(\mathsf{R1} \bowtie \mathsf{R2})$ is equivalent to the plan $\sigma_{Cond}(\mathsf{R1}) \bowtie \mathsf{R2}$;
- the cost (as introduced in the lectures) of the first plan is less than the cost of the second plan.

IN3020&IN4020 – Database Systems Exam, 2024

30 May, 9:00am–1:00pm (Inspera)

Question 7 of 12 (8%)

Consider the following four schedules:

$$S_1 : \mathtt{r}_1(X), \mathtt{w}_1(X), \mathtt{c}_1, \mathtt{r}_2(X), \mathtt{w}_2(Y), \mathtt{w}_2(X), \mathtt{c}_2$$
$$S_2 : \mathtt{r}_1(X), \mathtt{w}_1(X), \mathtt{r}_2(X), \mathtt{w}_2(Y), \mathtt{w}_3(Y), \mathtt{w}_2(X), \mathtt{c}_2, \mathtt{c}_1, \mathtt{c}_3$$
$$S_3 : \mathtt{r}_1(X), \mathtt{w}_1(X), \mathtt{c}_1, \mathtt{r}_2(X), \mathtt{w}_2(Y), \mathtt{w}_2(X), \mathtt{c}_2, \mathtt{w}_3(Y), \mathtt{c}_3$$
$$S_4 : \mathtt{r}_1(X), \mathtt{r}_2(X), \mathtt{w}_1(X), \mathtt{c}_1, \mathtt{w}_3(Y), \mathtt{w}_2(Y), \mathtt{c}_3, \mathtt{w}_2(X), \mathtt{c}_2$$

Which pairs of these schedules are conflict equivalent? Explain your answer (try to be as concise and concrete as possible).

IN3020&IN4020 – Database Systems Exam, 2024

30 May, 9:00am–1:00pm (Inspera)

Question 8 of 12 (8%)

Consider the following two transactions:

$$T_1 : \mathtt{r}(X), \mathtt{w}(Y), \mathtt{r}(Z), \mathtt{w}(Z)$$
$$T_2 : \mathtt{r}(X), \mathtt{w}(Z)$$

and the following partial schedule (with T_2 requesting a read-lock on X as the last operation):

$$\mathtt{b}_1, \mathtt{rl}_1(X), \mathtt{r}_1(X), \mathtt{wl}_1(Y), \mathtt{b}_2, \mathtt{rl}_2(X)$$

Assume that we are using the two-phase locking protocol with the no-waiting deadlock prevention protocol. Write down how the rest of the schedule (i.e., so that the result is a complete schedule) can be.

IN3020&IN4020 – Database Systems Exam, 2024

30 May, 9:00am–1:00pm (Inspera)

Question 9 of 12 (8%)

Assume that we have a database D with the following levels of item granularity:

1. Database
2. Table
3. Record

In other words, let the root of the granularity tree be the entire database, then let this root have one child per table in the database, and finally let each table have one child per record.

Furthermore, assume that database D has tables Table1 and Table2; that Table1 has records R1 and R2; and that Table2 has records P1, P2, and P3.

Make a schedule for the following transactions that uses the two-phase locking protocol with multiple granularity locking with the granularity described above:

$$T_1 : \mathtt{r}(\mathsf{R1}), \mathtt{w}(\mathsf{P2})$$
$$T_2 : \mathtt{r}(\mathsf{R2}), \mathtt{w}(\mathsf{R2})$$

IN3020&IN4020 – Database Systems Exam, 2024

30 May, 9:00am–1:00pm (Inspera)

Question 10 of 12 (8%)

Consider the following transactions:

$$T_1 : \mathtt{r}(X), \mathtt{w}(X)$$
$$T_2 : \mathtt{r}(X), \mathtt{w}(Y), \mathtt{w}(Z)$$
$$T_3 : \mathtt{r}(Y), \mathtt{w}(Z)$$

Write down a complete schedule for these transactions that is recoverable but not cascadeless, or explain why this is not possible.

IN3020&IN4020 – Database Systems Exam, 2024

30 May, 9:00am–1:00pm (Inspera)

Question 11 of 12 (8%)

Assume that we are using the immediate update protocol. Consider a database where item X has value 7 and item Y has value 3. Write down what the log will contain after the following (partial) schedule has been executed starting on this database:

$$\mathtt{b}_1, \mathtt{r}_1(X), \mathtt{eval}_1(X{:=}X * 2),$$
$$\mathtt{b}_2, \mathtt{r}_2(Y), \mathtt{eval}_2(X{:=}Y + 1), \mathtt{w}_2(X), \mathtt{w}_1(X), \mathtt{c}_1, \mathtt{b}_3, \mathtt{r}_3(Y)$$

In this schedule, $\mathtt{eval}_1$ and $\mathtt{eval}_2$ are (local) update operations that are evaluated within transaction T_1 and transaction T_2, respectively.

IN3020&IN4020 – Database Systems Exam, 2024

30 May, 9:00am–1:00pm (Inspera)

Question 12 of 12 (10%)

Consider the following statements in natural language:

- 'UiO is an organisation and a university, and it was established in 1811.'
- 'Carl is a person and he worked for UiO from 2017 to 2020.'

(a) Represent these statements in RDF. You can invent your own namespaces and IRIs (and you can use the rdf-prefix without defining it in your answer).

(b) Represent these statements in Neo4J (i.e., make CREATE-statements that adds the representations of the statements to a database).