CS–A1150 Databases
Exam May 20th, 2020

1. a) (8 p) Based on the description below, construct a UML diagram for a database of a delivery company. Use the notation used in the course text book and mark the key attributes like they are marked in the course text book (and in the lecture slides, too).

   The delivery company has a number of offices where the customers can bring parcels to be delivered to another office and pick them up. One transport is always from one office to another office. Thus, the same parcel may belong to several transports before it has been delivered from the origin to the final destination.

   Information about offices contains their id (unique), name and address. The transports are run on certain permanent routes. For each route, the information contains the starting office, the destination office and a unique route id. The route is always between exact two offices (not more). Information about future transports on each route is stored in the database. In addition, the database stores information about transports already run in the past. For each transport, the information contains the route, planned or actual starting time and the date and the type of the vehicle used.

   For each parcel, the information contains a unique parcel id, contact information of the sender and the receiver, volume, weight, delivery fee, the original office where the parcel left for delivery, and the final destination office. In addition, the database contains the information about every transport to which the parcel has belonged and the current status of the parcel (for example in transport, waiting for the next transport of waiting for pick up).

   The sender may choose a number of extra services with the delivery (for example an insurance, extra storage time in the destination, cautious processing, or extra notification to the receiver). The database contains information about the extra services available (service type id, description, fee) and for each parcel information about which extra services have been chosen for this parcel.

   b) (2 p) Convert the UML diagram from part (a) into relations. Write the relation schemas and underline the names of the key attributes.

2. Consider the following database schema which presents artists (painters), paintings, art museums and information about which museum owns which painting.

   Dababase schema:

   ```
   Artists(name, born, died, country)
   Paintings(id, name, painter, year, type, value)
   Museums(name, city, webpage)
   Owns(museumName, paintingID, since, until)
   ```

   Relation `Artists` contains information about the artists in the database. It tells the name, year of birth, year of death and the home country of the artist. If the artist is still alive, the year of death is NULL. We assume that there are no two artists with the same name.

   Relation `Paintings` contains information about the paintings in the database. Each painting has an id (unique), a name, the painter's name, the year the work was painted, the type (for example oil or watercolor) and the value of the painting as euros.

   Relation `Museums` gives for each museum its name, city and address of its webpage. We assume that the names of the museums are unique.

   The `Owns` relation tells which painting is owned by which museum. Because the relation contains also historical data from paintings which were sold from one museum to another, it also contains the year the museum received the painting and the year the museum gave up the painting. (The value of the latter attribute is NULL if the museum still owns the painting.) We assume that the same museum has not owned the same painting more than once.

   The values of the paintings are decimal numbers and the years integers. All other attribute values are strings. You may assume that the tuples of the relations do not contain NULL values except for attributes `died` and `until`.

   Write the following SQL queries:
   A hint: You can examine whether the value of attribute `until` is NULL by writing
   `WHERE until IS NULL`
   or something other than NULL by writing
   `WHERE until IS NOT NULL`

a) (2 p) The name, painter's name and year the work was painted of all those paintings which have the painter who was born after year 1990 and the value of which is over 10000 euros.

b) (2 p) The names and cities of the art museums which own or have owned a painting whose value is over million euros.

c) (2 p) The names and cities of the museums which own or have owned a painting painted by Pekka Halonen, but which have never owned any painting painted after year 2000.

d) (2 p) Find the museums which currently own at least four paintings painted by Vincent van Gogh. For each such museum, the query must list the name of the museum, the number of paintings by Vincent van Gogh currently owned by the museum and the total value of those paintings (the number and the total value of the paintings is calculated separately for each museum based on only paintings owned by this museum).

Explain which query the following expressions of Relational Algebra produce the answer (for example: "the expression produces names and years of all paintings painted by Pekka Halonen")

e) (2 p)

$$\pi_{year,country}(\texttt{Artists} \bowtie (\sigma_{value>5000}(\texttt{Paintings})))$$

f) (2 p) (the expression consists of three lines)

$$\pi_{O1.museumName}(\rho_{O1}(\texttt{Owns} \bowtie_{paintingID=ID} (\sigma_{painter='Dali'}(\texttt{Paintings})))$$
$$\bowtie_{O1.ID<O2.ID \texttt{ AND } O1.museumName=O2.museumName}$$
$$(\rho_{O2}(\texttt{Owns} \bowtie_{paintingID=ID} (\sigma_{painter='Dali'}(\texttt{Paintings})))))$$

3. a) (1 p) Answer to the SQL posttest and final query in A+ system. The deadline is May 27th. (i.e. You do not have to do it during the exam.)

b) (3 p) Assume that the database in Problem 2 is such large that it does not fit in the main memory. Assume that the database contains about 10 000 paintings and 1000 artists, which have 15 various home countries. The number of artists from each home country is almost the same. Table `Artists` occupies 30 disk pages and the artists have been ordered by their name. Table `Paintings` occupies 700 disk pages and the paintings are stored in random order.

Following queries are very often executed in the database:

- Look for the paintings painted by a certain artist, for example

```
SELECT *
FROM Paintings
WHERE painter = 'Pekka Halonen';
```

- Look for the painters from a certain country, for example

```
SELECT *
FROM Artists
WHERE country = 'Finland';
```

The names of the artists and the countries vary and they are not always the same as ones given in the examples above. Is it profitable to create an index on attribute `painter` for Table `Paintings`? How about to create an index on attribute `country` for Table `Artists`? Justify your answer with a few sentences. You do not have to give any exact calculations.

4. Consider a relation R with schema $R(A, B, C, D, E)$ and functional dependencies $A \rightarrow B$, $B \rightarrow C\ D\ E$, and $D \rightarrow E$.

a) (1 p) Explain why this relation is not in Boyce-Codd normal form (BCNF).

b) (6 p) Decompose the relation using the BCNF decomposition algorithm taught in this course and in the text book. Give a short justification for each new relation. Continue the decomposition until the final relations are in BCNF. Explain why the final relations are in BCNF.

5. Consider a database containing information about the storage of a web store. The database contains Table Products(productNo, name, count, price) (and some other tables as well). The value of attribute `productNo` is a unique id of the product, `name` its name, and `count` the number of items of this product currently in the storage, and `price` the price of this product in the web store. Assume that before the transactions described below are started, the table contains at least the following row:

```
(R-300, superball, 20, 75.0)
```

No such contraints which would affect the execution of the transactions described below has been defined.

Assume that the following transactions T1, T2, T3, and T4 are executed partly concurrently. We do not know the order in which the transactions are executed, but if a certain transaction contains several instructions, those instructions inside the same transaction are executed in the order they have been given.

T1
- Reads the value of attribute count of the product `R-300` in Table Products.
- Adds 500 to the value it just read and updates of the value of attribute count of the product `R-300` in Table Products such that the new value is the result of the addition.
- Updates the value of attribute price in of the product `R-300` in Table Products such that the new value is 80.0.

T2
- Inserts into Table Products a new row

  `(T-200, superballmachine, 50, 400.0)`

T3
- Calculates the number of the products with a price of over 100 euros in Table Products.
- Calculates the total number of items of all products with a price of over 100 euros in Table Products. (i.e. Only one sum which is the total number of the items of all these products is calculated, not separately for each product.)

T4
- Updates the value of attribute price of product R-300 in Table Products. The new value is 150.0.

a) (3 p)
Assume that all ACID properties of the transactions are valid in the database. Which of the following situations are then possible? Choose all correct alternatives.

- After the execution of all transactions the price of product R-300 is 75.0 euros.
- After the execution of all transactions the price of product R-300 is 80.0 euros.
- After the execution of all transactions the price of product R-300 is 150.0 euros.
- The database remains in a state where the value of attribute count of product R-300 is 520 and the value of its attribute price is 75.0.
- The database remains in a state where the value of attribute count of product R-300 is 20 and the value of its attribute price is 80.0.
- Product R-300 is not taken into account, when T3 calculates the number of products with price over 100 euros in its first instruction, but it is taken into account when T3 calculates the total number of items of products with price over 100 euros in its second instruction.
- Product T-200 is not taken into account, when T3 calculates the number of products with price over 100 euros in its first instruction, but it is taken into account when T3 calculates the total number of items of products with price over 100 euros in its second instruction.

b) (2 p) Consider the same table and transactions as before. Assume that all ACID properties of the transactions are otherwise valid in the database, but it is possible that property atomicity does not hold. Which of the following situations are then possible? Choose all correct alternatives.

- After the execution of all transactions the price of product R-300 is 75.0 euros.
- After the execution of all transactions the price of product R-300 is 80.0 euros.
- After the execution of all transactions the price of product R-300 is 150.0 euros.
- The database remains in a state where the value of attribute count of product R-300 is 520 and the value of its attribute price is 75.0.
- The database remains in a state where the value of attribute count of product R-300 is 20 and the value of its attribute price is 80.0.
- Product R-300 is not taken into account, when T3 calculates the number of products with price over 100 euros in its first instruction, but it is taken into account when T3 calculates the total number of items of products with price over 100 euros in its second instruction.
- Product T-200 is not taken into account, when T3 calculates the number of products with price over 100 euros in its first instruction, but it is taken into account when T3 calculates the total number of items of products with price over 100 euros in its second instruction.

c) (1 p) Consider the same table and transactions as before. Assume that all ACID properties of the transactions are otherwise valid in the database, but instead of serializable, the isolation level of all transactions is set to READ COMMITTED. Which of the following situations are then possible? Choose all correct alternatives.

- After the execution of all transactions the price of product R-300 is 75.0 euros.
- After the execution of all transactions the price of product R-300 is 80.0 euros.
- After the execution of all transactions the price of product R-300 is 150.0 euros.
- The database remains in a state where the value of attribute count of product R-300 is 520 and the value of its attribute price is 75.0.
- The database remains in a state where the value of attribute count of product R-300 is 20 and the value of its attribute price is 80.0.
- Product R-300 is not taken into account, when T3 calculates the number of products with price over 100 euros in its first instruction, but it is taken into account when T3 calculates the total number of items of products with price over 100 euros in its second instruction.
- Product T-200 is not taken into account, when T3 calculates the number of products with price over 100 euros in its first instruction, but it is taken into account when T3 calculates the total number of items of products with price over 100 euros in its second instruction.

d) (1 p) Consider the same table and transactions as before. Assume that all ACID properties of the transactions are otherwise valid in the database, but instead of serializable, the isolation level of all transactions is set to REPEATABLE READ. Which of the following situations are then possible? Choose all correct alternatives.

- After the execution of all transactions the price of product R-300 is 75.0 euros.
- After the execution of all transactions the price of product R-300 is 80.0 euros.
- After the execution of all transactions the price of product R-300 is 150.0 euros.
- The database remains in a state where the value of attribute count of product R-300 is 520 and the value of its attribute price is 75.0.
- The database remains in a state where the value of attribute count of product R-300 is 20 and the value of its attribute price is 80.0.
- Product R-300 is not taken into account, when T3 calculates the number of products with price over 100 euros in its first instruction, but it is taken into account when T3 calculates the total number of items of products with price over 100 euros in its second instruction.
- Product T-200 is not taken into account, when T3 calculates the number of products with price over 100 euros in its first instruction, but it is taken into account when T3 calculates the total number of items of products with price over 100 euros in its second instruction.

Please fill the course feedback form before Jun 2nd. The link has been sent to the registered students by e-mail on May 13th. Note that this is not the same questionary as mentioned in Problem 3a.