CS–A1150 Databases

Exam May 22nd, 2019 **No calculators or extra material allowed**

Students whose mother tongue is not Finnish may use a dictionary, if it does not contain any markings. Those students may also obtain both Finnish and English/Swedish exam sheet, if they want.

Please write your name, your student ID, the name and the code of this course, the date, and the total number of the papers you submit on top of each paper you submit.

1. a) (8 p) Based on the description below, construct a UML diagram for a database of a movie theater. 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 theater consists of several auditoriums, which have a unique name. For each auditorium, the number of its seats is known, too. The theater arranges shows in the auditoriums. For each show, the information contains the name of the auditorium of the show, the date and the time, and the movie which is shown. Each movie has a unique ID. In addtion, the name of the movie, the name of its director and the length (in minutes) of the movie are stored. The same movie may be shown in several shows in different auditoriums at the same time.

   The database contains information about customers who can purchase tickets through the webstore. For the sake of simplicity, we assume that the tickets do not have seat numbers. The customers have a unique user id. In addition, the name, the e-mail address and the phone number of the customer are stored. For each purchase made by a customer, the database contains a unique id of the purchase, information about which show the ticket is for, time of the purchase, the number of tickets purchased (the customer can purchase simultaneously several tickets for the same show for him/herself and his/her friends – no information about the friends is stored anywhere), information about which extra service (food and drinks, for example) are included and the total price of the purchase. Part of the tickets are sold directly at the box office of the theater. Those purchases are stored otherwise in the same way in the database, but no information about the customer is stored. Possible extra services included in the purchase are food and drinks, for example. One purchase may contain, for example, four tickets for the same show, three large soft drinks, one small portion of popcorn and two portions of ice cream. The data stored about extra services contains a unique ID of the service, the name of the service, and the size of the service (small or large, for example).

   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 departments, employees, and customers of the firm and the orders it has got from customer firms. The job of the employees is to get orders from the customer firms.

```
Employees(number, name, salary, departmentId)
Departments(id, name, directorNumber)
Customers(customerId, firmName, address)
Orders(orderNo, EmpoyeeNumber, customerId, date, value, description)
```

   Relation `Employees` tells the number, name and the monthly salary of an employee. It also tells the id of the department at which the person works. Relation `Departments` tells the id, name and the employee number of the director of the department. Relation `Customers` tells the customer id, name and address of a customer firm. Relation `Orders` contains information of all orders which the employees of the firm have got. It tells the order number, number of the employee who got the order, id of the customer firm, date of the order, value of the order in euros and a short description of the order.

   The salaries of the employees and values of the sales are decimal numbers. Other attribute values are strings.You may assume that the tuples of the relations do not contain NULL values.

   Write the following SQL queries:

   a) (2 p) The customer ids and names of the customers which have made an order with value of over 1000 euros (the value of a single order).

   b) (2 p) The numbers and names of the employees whose salary is less than 3000 euros, but who have got an order from the firm with name Top Games.

   c) (2 p) The numbers and names of the employees whose salary is over 5000 euros, but who have never got an order from the firm with name Top Games.

d) (2 p) Find the departments whose employees have got at least 30 orders with the value of over 5000 euros (i.e. for each department, we calculate the total number of orders with the value of over 5000 euros got by all employees of this deparment). For those departments, the query must list the id, the name and the total number of orders with the value of over 5000 euros.

Write the following queries as expressions of the relational algebra:

e) (2 p) The names and ids of the departments which have a director whose salary is less than 5500 euros.

f) (2 p) The numbers, names, and salaries of the employees who work in a department which has a director whose name is Pekka Puupaa. If there are several departments directed by an employee with name Pekka Puupaa, the query must produce the employees of all of those departments.

3. a) (1 p) Answer to the SQL posttest and final query in A+ system. The deadline is May 29th.

   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 firm has about 1000 employees who work in 10 various departments such that the number of employees in each department is almost the same. Table `Employees` occupies 100 disk pages and the employees are stored in the table in random order. Following queries are very often executed in the database:

   - Look for the employees having a given name, for example

     ```
     SELECT *
     FROM Employees
     WHERE name = 'Tiina Teekkari';
     ```

   - Look for the employees working in a given department, for example

     ```
     SELECT *
     FROM Employees
     WHERE departmentID = 'D14';
     ```

   The given names and departments vary and they are not always the same as ones given in the examples above. Is it profitable to create an index on attribute `name` for Table `Employees`? How about to create an index on attribute `departmentID` for Table `Employees`? Justify your answer with a few sentences.

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

   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. (7 p) What are the four important properties of the transactions in a database managament system? Use a few sentences to explain each of these properties. Just listing the names of the properties and writing 1–2 sentences about each are not enough for 7 points.

Please fill the course feedback form before Jun 4th. The link has been sent to the registered students by e-mail on May 15th. Note that this is not the same questionary as mentioned in Problem 3a.