CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Department of Computer Science and Engineering

Examination in Databases, TDA357/DIT620

Thursday 24 August 2017, 14:00-18:00

Examiner:	Graham Kemp (telephone 772 54 11, room 6475 EDIT) The examiner will visit the exam room at 14:45 and 16:00.
Results:	Will be published by 14 September at the latest.
Exam review:	See course web page for time and place: http://www.cse.chalmers.se/edu/year/2017/course/TDA357/VT2017/
Grades:	Grades for Chalmers students (TDA357) are normally determined as follows: $\geq$ 48 for grade 5; $\geq$ 36 for grade 4; $\geq$ 24 for grade 3. Grades for GU students (DIT620) are normally determined as follows: $\geq$ 42 for grade VG; $\geq$ 24 for grade G.
Help material:	One A4 sheet with hand-written notes. You may write on both sides of that sheet. If you bring a sheet, it must be handed in with your answers to the exam questions. English language dictionaries are allowed.

Specific instructions:

- Please answer in English where possible. You may clarify your answers in Swedish if you are not confident you have expressed yourself correctly in English.
- Begin the answer to each question on a new page.
- Write clearly; unreadable = wrong!
- Fewer points are given for unnecessarily complicated solutions.
- Indicate clearly if you make any assumptions that are not given in the question.
- Write the page number and question number on every page.

Question 1. Consider the following domain description.
12 p

A database will contain information about the employees of several organisations. Organisation names are unique. Each organisation consists of one or more divisions. Within an organisation, the division names are unique. However, different organisations can have divisions with the same name (e.g. "Sales"). Some organisations (universities) differ from other organisations by being allowed to award PhD degrees.

Each employee is identified by a unique personal number. Employee names are not guaranteed to be unique. Each employee can work at more than one division. Therefore, in addition to recording which employees work at each division, we also want to record what percentage of the employee's time is at the different divisions (e.g. employee "e21" could work 75% of their time in the "Sales" division and 25% of their time in the "Personnel" division of an organisation).

We want to record additional information about employees who have a PhD degree. For these employees we want to record the year when the degree was awarded, the title of their thesis, and the university that awarded their PhD.

a) Draw an E-R diagram that correctly models this domain.
(6p)

b) Translate this E-R diagram into a set of relations, clearly marking all references and keys.
(6p)

Question 2. Suppose we have relation $R(A, B, C, D, E)$ with functional dependencies $AB \to C$, $C \to B$, $C \to D$, $C \to E$, $D \to E$.
10 p

a) i) State, with reasons, which of the FDs listed above violate BCNF.
(1p)

ii) Decompose relation R to BCNF. Show each step in the normalisation process, and at each step indicate which functional dependency is being used. Indicate keys and references for the resulting relations.
(3p)

iii) What problem is caused by this decomposition?
(1p)

b) i) Decompose relation R to 3NF.
(3p)

ii) Explain whether this decomposition has the same problem as the decomposition in (a)(ii). Does it have any other problem?
(1p)

iii) Explain whether the relations in your answer to (b)(i) are in BCNF.
(1p)

Question 3. An estate agent uses a relational database to manage information about clients, properties for sale and bids made by clients for properties. This database has the following relations:
8 p

Clients(cid, name, telephone)
Properties(ref, address, area, seller, sold, saleDate)
Bids(property, client, amount)

A client has a unique client identifier (attribute *cid*). The client's name and telephone number are stored in the database. A client can be someone who is selling a property, or who is interested in buying a property or both (i.e. a seller of one property (or more than one property) and a prospective buyer for others).

Each property that the agent is trying to sell is identified by a reference code. The address of each property is stored in the database, and also the property's living area (in m^2). Attribute *seller* is the client identifier of the client who is selling he property. Attribute *sold* contains the value 'Y' or 'N' (default value is 'N') to indicate whether the property has been sold yet and, if it has been sold, attribute *saleDate* records the date of the sale.

During the lifetime of the database, the same client may sell more than one property, and the same property could be sold several times (each time a property is sold it will be assigned a new reference code).

a) The following SQL statements attempt to create these tables, but there are three errors in these statements. Find the errors and correct the statements.

```
CREATE TABLE Clients (
    cid         INT PRIMARY KEY,
    name        VARCHAR(30),
    telephone   VARCHAR(15)
);

CREATE TABLE Properties (
    ref         VARCHAR(10) PRIMARY KEY,
    address     VARCHAR(30),
    area        INT,
    seller      VARCHAR(30),
    sold        CHAR(1) 'N' CHECK (sold IN ('Y','N')),
    saleDate    DATE,
    FOREIGN KEY (seller) REFERENCES Clients(cid)
);

CREATE TABLE Bids (
    property    VARCHAR(10) PRIMARY KEY,
    client      INT PRIMARY KEY,
    amount      INT PRIMARY KEY,
    FOREIGN KEY (property) REFERENCES Properties(ref),
    FOREIGN KEY (client) REFERENCES Clients(cid)
);
```

(4p)

b) A client is not allowed to bid on a property which they are selling.
Write trigger code that enforces this constraint.
(4p)

Question 4. Assume the same relations as in Question 3:

6 p

Clients(cid, name, telephone)
Properties(ref, address, area, seller, sold, saleDate)
Bids(property, client, amount)

a) Write a relational algebra expression that finds the seller's name and the address of properties with an area greater than $300m^2$. The results should be sorted by area.
(3p)

b) Write a relational algebra expression that finds the addresses of properties that received more than three bids over 4000000.
(3p)

Question 5. Assume the same relations as in Question 3:

11 p

Clients(cid, name, telephone)
Properties(ref, address, area, seller, sold, saleDate)
Bids(property, client, amount)

a) Write an SQL query that finds the names and telephone numbers of all clients who are selling properties that are currently unsold.
(3p)

b) Write an SQL query that finds for each property: the property identifier, property address and the amount of the highest bid on that property.
(4p)

c) Write an SQL query that finds the the client identifiers of clients who bid for a property that was sold in 2011, but did not bid for any properties that were sold in 2012.
(4p)

Question 6. An airline's check-in system has a transaction T with the following three steps:

4 p

T_1: list available seats
T_2: select a seat from the list
T_3: confirm the selection that was made in step T_2

Suppose transaction T is run with isolation level `READ UNCOMMITTED`. Explain how *dirty reads* can occur. Refer to transaction T in your answer.

Discuss whether it would be a good idea to run transaction T with some other isolation level.

(4p)

Question 7. Consider the following piece of XML:

9 p

```
<University>
  <Departments>
    <Department name="CSE">
      <Programme code="MPALG" type="Masters" />
      <Programme code="MPCSN" type="Masters" />
    </Department>
    <Department name="Maths">
      <Programme code="MPENM" type="Masters" />
    </Department>
  </Departments>
  <People>
    <Person name="Andersson">
      <Teacher position="Professor" dept="CSE" />
    </Person>
    <Person name="Johansson" address="21 Main Street">
      <Teacher position="Lecturer" dept="Maths" />
    </Person>
    <Person name="Karlsson">
      <Student prog="MPALG">
        <Registered>DAT620</Registered>
        <Registered>TIN093</Registered>
      </Student>
    </Person>
    <Person name="Nilsson">
      <Student prog="MPCSN">
        <Registered>EDA093</Registered>
        <Registered>EDA387</Registered>
      </Student>
    </Person>
    <Person name="Eriksson" />
  </People>
</University>
```

a) Write a Document Type Definition (DTD) for the XML that is given above.
(4p)

b) Write an XPath expression that finds Person elements where the person is registered for course DAT620.
(2p)

c) Draw an Entity-Relationship diagram for this domain.
(3p)