Examination in Databases

TDA357 (Chalmers), DIT620 (University of Gothenburg)
Department of Computer Science and Engineering

5^{th} June 2018, 14:00-18:00

Course responsible: Pablo Picazo-Sanchez. Markus and Evgeny will visit the exam rooms around 15:00 and 17:00.

Results: Will be published by the end of June 2018 at the latest.

Exam review: Individual agreement with the examiner.

Grades: Grades for Chalmers students (TDA357) are normally determined as follows:
Chalmers: 24 for 3, 36 for 4, 48 for 5.

Grades for GU students (DIT620) are normally determined as follows:
GU: 24 for G, 42 for VG.

Help material: One cheat sheet, which is an 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. One English language dictionary is also allowed.

Specific instructions: Answer questions in English. Begin the answer to each question (numbers 1 to 5) on a new page. The a,b,c,... parts with the same number can be on the same page.

Specific Instructions: Unreadable = wrong! Fewer points are given for unnecessarily complicated solutions. Indicate clearly if you make any assumptions that are not given in the question. In SQL questions, use standard SQL or PostgreSQL. If you use any other variant (such as Oracle or MySQL), say this; but full points are not guaranteed since this may change the nature of the question.

Multiple Choice Marks right answer = +1, bad answer = -0,5; blank = 0. The minimum score that you can get for question 1 is zero — if the total points for question 1 turns out to be negative, then a value of zero will be given for this question.

1 Multiple Choice Questions (12p)

For each question, it is sufficient to write the letter corresponding to the correct answer.

1) Which of the following terms does refer to the correctness and completeness of the data in a database?

a) Data security

b) Data constraint

c) Data independence

d) Data integrity

D

---

2) Given the relation instructor(ID, dept_name, salary) and the following SQL query:

```
SELECT dept_name, ID, avg (salary)
FROM instructor
GROUP BY dept_name;
```

This statement IS syntactically erroneous because:

a) Avg(salary) should not be selected

b) Group by clause must contain both dept_name and ID

c) Misplaced group by clause

d) Group by clause cannot be used with foreign keys

B

---

3) The ..... operator takes the results of two queries and returns only rows that appear in both result sets.

a) Union

b) Difference

c) Intersection

d) Join

C

---

4) Given the following XML:

```
<book category="Web">
  <bname>XML Tutorials</bname>
  <pages>100</pages>
  <price>$300.00</price>
</book>
```

How many nodes and attributes are there in the document?

a) 4 Nodes and 1 Attribute

b) 3 Nodes and 1 Attribute

c) 2 Nodes and 2 Attribute

d) 4 Nodes and 2 Attribute

A

---

5) The result which operation contains all pairs of tuples from the two relations, regardless of whether their attribute values match:

a) Outer Join

b) Cartesian product

c) Union

d) Set difference

B

---

6) How can all rows in the relation teaches be deleted?

a) Drop table teaches;

b) Purge table teaches;

c) Delete from teaches where Id ='Null';

d) Delete from teaches;

D

---

7) Considerer the following actions:

```
TRANSACTION...
Commit;
ROLLBACK;
```

What does ROLLBACK do?

a) Undoes the statements before commit

b) Redoes the statements before commit

c) Clears all transactions

d) No action

D

---

8) Given these SQL statements:

```
CREATE TABLE Employee(
  Emp_id  INTEGER PRIMARY KEY NOT NULL,
  name   VARCHAR(100),
  dept_name  VARCHAR(100),
  Salary  INTEGER,
  UNIQUE(Emp_id,name));

INSERT INTO Employee VALUES(102, 'Markus', 'CSE', 1000);
INSERT INTO Employee VALUES(106, 'Ana', 'Finance', NULL);
INSERT INTO Employee VALUES(103, 'Steven', 'Sales', 2000);
```

What will be the final result?

a) All statements will be executed

b) Error in INSERT INTO Employee VALUES(103, 'Steven', 'Sales', 2000);

c) Error in CREATE statement

d) Error in INSERT INTO Employee VALUES(106, 'Ana', 'Finance', NULL);

A

---

9)

```
CREATE TABLE course(
. . .
FOREIGN KEY (dept_name) REFERENCES department
. . . );
```

Which one of the following statements is used to delete referenced rows in table department when a row is deleted from table course?

a) DELETE

b) SET NULL

c) DELETE CASCADE

d) All of the mentioned

C

---

10) Given the relation R=$\{A, B, C, D, E\}$ and the following Functional Dependencies: FD = $\{A \rightarrow C; B \rightarrow D; A, B \rightarrow E\}$. In terms of normalization, R is in:

a) 1NF

b) 2NF

c) 3NF

d) BCNF

A

---

11) Which of the following is not a property of transactions?

a) Atomicity

b) Concurrency

c) Isolation

d) Durability

B

---

12) Given the following XML:

```
<notes>
  <note date=12/11/2007>
    <to>Tove</to>
    <from>Jani</from>
  </note>
  <note date=12/11/2007>
    <to>Tove</to>
    <from>Jani</from>
  </note>
  <note date=12/11/2007>
    <to>Tove</to>
    <from>Jani</from>
  </note>
</notes>
```

Point out the error it contains.

a) XML elements must be properly nested

b) None of these

c) XML documents must have a root element

d) XML attribute values must be quoted

D

---

2 Entity Relationship Modelling (12p)

The 2018 FIFA World Cup is taking place in Russia this month, and a group of students decided to use a database to store information about the national teams and the results during the World Cup.

Each team is identified by its name. Each team might play up to 7 matches: 3 in the group phase + Round-16 + Quarter finals + Semi finals + {Final,Third place}. Regarding matches, a match is identified by a combination of its date and the city where the match takes place. The result of the match should also be stored. The following Fig. 1 depicts the composition of the groups and the final phase of the World Cup.

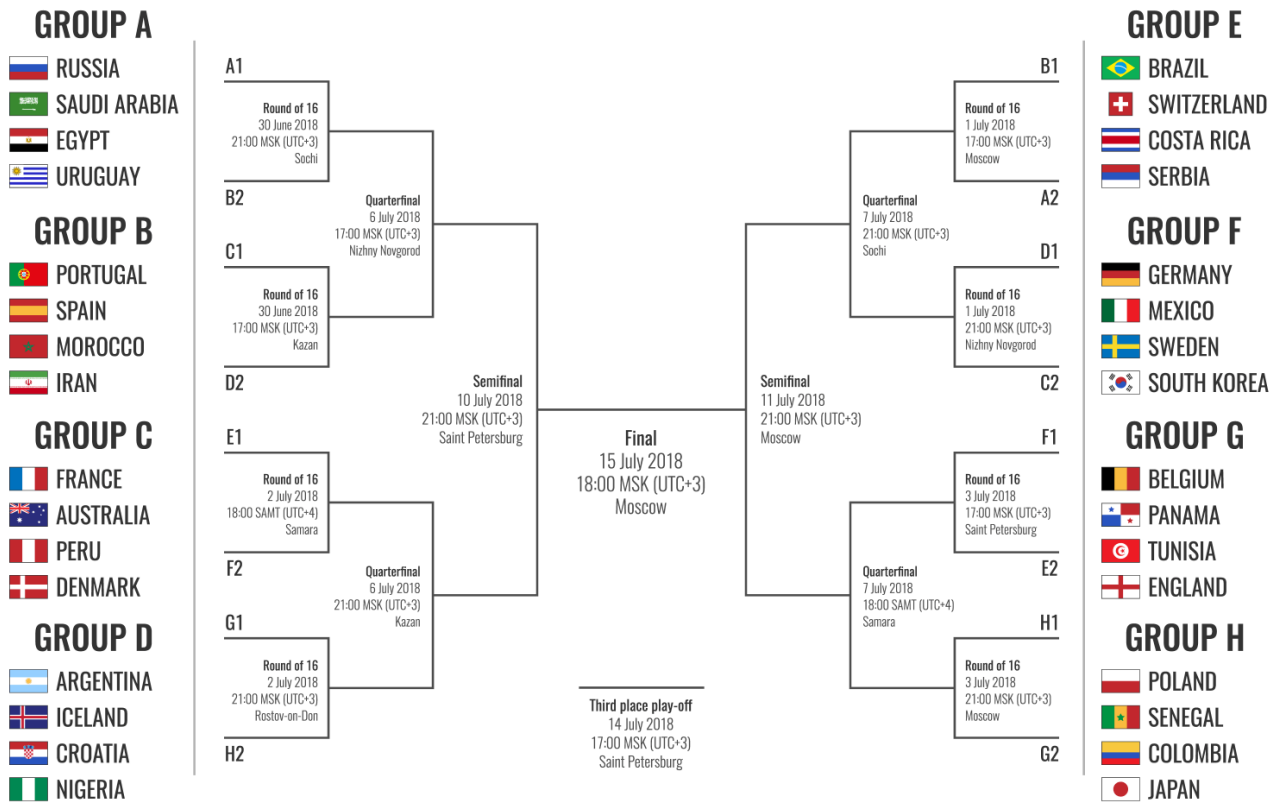
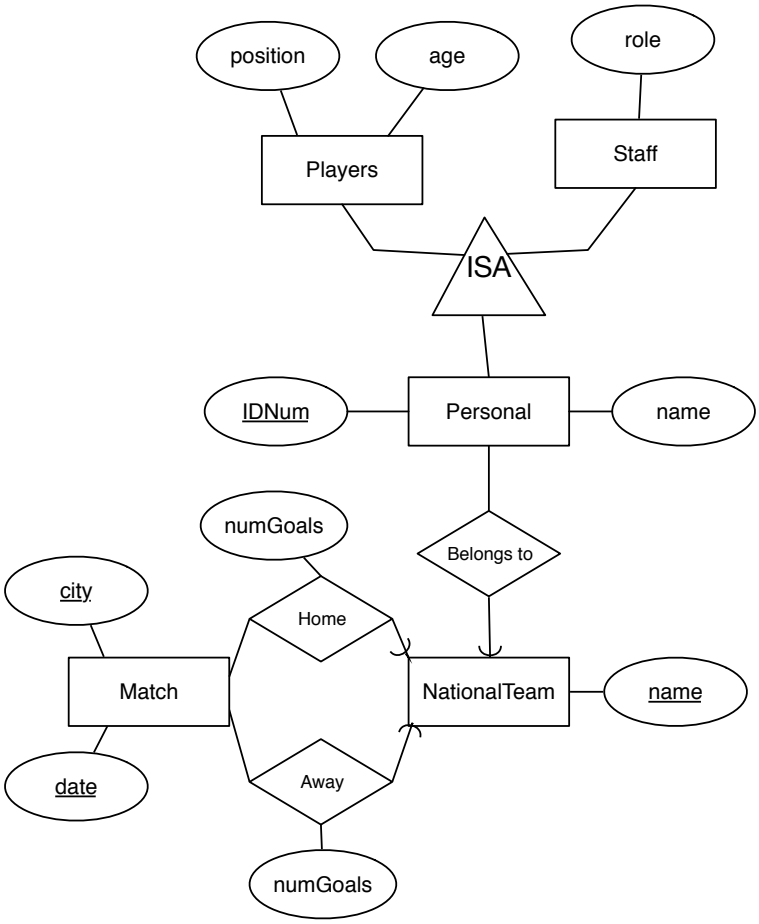


Figure 1: Russia 2018 WorldCup Wallchart.

National teams are composed of two different people: players and staff members. Players are identified by their ID number. Additionally, other information such as player's names, their positions in the field, and their age must be stored into the database. On the other hand, the staff members are also identified by their ID number and values such as their names and the role they have in the team should be stored.

2a. Draw an Entity-Relationship diagram for this domain listing any assumptions you make. (8p)

SOLUTION:



2b. Translate the ER diagram into a set of relations, clearly marking keys and references in your answer. (4p)

SOLUTION:

NationalTeam (name)
Personal (IDNum, team, name)
    team→NationalTeam(name)
Player (IDNum, position, age)
    IDNum→Personal(IDNum)
Staff (IDNum, role)
    IDNum→Personal(IDNum)
Match(date, city, home, away, anumGoals, hnumGoals)
    home→NationalTeam(name)
    away→NationalTeam(name)
    anumGoals→Away(numGoals)
    hnumGoals→Home(numGoals)

3 Functional Dependencies and Normal Forms (12p)

3a. Calculate the set of candidate keys of the relation R(A, B, C, D, E) with functional dependencies: FD={A → BC, CD → E, B → D, E → A}. (4p)

SOLUTION:

$\{A\}^+ = \{ABCDE\}$ so A is a candidate key.

$\{B\}^+ = \{BD\}$.

$\{C\}^+ = \{C\}$.

$\{D\}^+ = \{D\}$.

$\{E\}^+ = \{ABCDE\}$ so E is a candidate key.

$\{BC\}^+ = \{ABCDE\}$ so BC is a candidate key.

$\{BD\}^+ = \{BD\}$.

$\{CD\}^+ = \{ABCDE\}$ so CD is a candidate key.

3b. Calculate (and justify) the minimal key of R(A, B, C, D, E, F) with the functional dependencies: FD={BD → E, CD → A, E → C, B → D}. (4p)

SOLUTION: Note that neither B nor F are determined by any other attributes, thus calculate $\{BF\}^+$ is a good starting point:

$\{BF\}^+ = \{ABCDEF\}$ which means that BF is the minimal key.

3c. Given R(A, B, C, D, E) with the functional dependencies: FD={A → BC, C → D, B → E}, justify why R is not in BCNF and decompose R to BCNF. (4p)

SOLUTION:

R is not in BCNF because C is not a key and there is a FD={C → D}. We split R into R_1=(A,B,C,E) and R_2=(C,D)

R_1 is not in BCNF because B is not a key and there is a FD={B → E}. We split R_1 into R_3=(A,B,C) and R_4=(B,E)

R_2, R_3 and R_4 are in BCNF.

4 Relational algebra (12p)

A relational database used by a dog exhibition has the following relations: Dog, Owner, Jury and Evaluation. Each dog belongs to an owner and it has a name, breed and is identified by a code. Regarding owners, they are identified by an ID and we also have their names. About the jury, they are identified by an ID and their names and surnames are stored as well. Finally, an evaluation is given by the jury to the dogs which is denoted as a grade.

We are provided with the following relational database schema:

```
Dog(did, oid, name, breed)
Owner(oid, name, age)
```

```
Jury(jid, name, surname)
Evaluation(did, jid, grade)
```

4a. Write a relational algebra expression that finds the age of owners who have a dog whose breed is "boxer". (3p)

SOLUTION: $\pi_{age}(\sigma_{breed='boxer'}(\text{Dog} \bowtie_{Dog.oid=Owner.oid} \text{Owners}))$

4b. Write a relational algebra expression that finds dogs that have not been graded yet (i.e., no members of the jury have evaluated that dog). (4p)

SOLUTION: $\pi_{did}(Dogs) - \pi_{did}(Evaluation)$

4c. Write a relational algebra expression that finds members of the jury (jid) that have given a grade ≥ 6 to at least 5 dogs. (5p)

SOLUTION: $\sigma_{total>5}(\gamma_{jid,COUNT(*)\rightarrow total}(\sigma_{grade\geq 6}(\text{Evaluation})))$

5 SQL (12p)

Assume the description provided in Question 4 and the same relational database schema:

```
Dog(did, oid, name, breed)
Owner(oid, name, age)
Jury(jid, name, surname)
Evaluation(did, jid, grade)
```

5a. Write an SQL query which returns the owners (name and age) who have dogs whose breed is "boxer". Avoid repeated information (i.e., do not return the same name, age more than once). (3p)

SOLUTION:

```
SELECT Owner.name, age
FROM Dog,Owner
WHERE breed = 'boxer' AND Dog.oid=Owner.oid
GROUP BY Owner.name, age;
```

SOLUTION 2:

```
SELECT DISTINCT Owner.name, age
FROM Dog,Owner
WHERE breed = 'boxer' AND Dog.oid=Owner.oid
```

5b. Write an SQL query which returns the members of the jury (name, surname) who have given a grade ≥ 6 to at least 5 dogs. (4p)

SOLUTION:

```
SELECT name, surname
FROM Jury
NATURAL JOIN Evaluation
GROUP BY jid
HAVING COUNT(*)>5;
```

5c. Write an SQL query that returns both the name of the dog and the name of its owner of those dogs with the best grade. (5p)

SOLUTION:

```
SELECT did, Dog.name, Owner.name
FROM Dog, Owner
WHERE did IN (SELECT did
  FROM Evaluation
  WHERE grade = (SELECT MAX(grade)
  FROM Evaluation))
AND Dog.oid = Owner.oid;
```