

**Winter Examination Period 2020 — January — Semester A****ECS519U      Database Systems****Duration: 3 hours**

This is a 3-hour open-book exam, which must be started within a 24-hour period.

You **MUST** submit your answers within 3 hours of the time that you started the exam.

Follow all instructions on the download page.

You can refer to textbooks, notes and online materials to facilitate your working, but normal referencing and plagiarism rules apply, and you must cite any sources used.

Answers are to be submitted via the exam system (questionnaire) for entering the answers.

Calculators are permitted in this examination.

**Answer FOUR questions.**

**You MUST adhere to the word limits, where specified in the questions. Failure to do so will lead to those answers not being marked.**

**YOU MUST COMPLETE THE EXAM ON YOUR OWN, WITHOUT CONSULTING OTHERS.**

**Examiners:**

Dr. Thomas Roelleke and Dr. Mustafa Bozkurt

**Question 1****Database (DB) History and General Concepts**

- (a) DB history: The relational model was a major milestone in DB history.
- (i) Explain what type of model(s) the relational model replaced.  
In your explanation, discuss why the relational model was viewed as superior to the model it replaced.
  - (ii) Object-oriented databases emerged in the early 90s. They were believed to replace relational DB technology.  
Discuss the reasons why relational databases are still dominating over OO databases.

**[5 marks — word limit 100]**

- (b) DB applications and users:
- (i) List five of the main DB application domains. For each domain, explain what requirement it implied for DB technology.
  - (ii) For five of the user groups of DB technology, explain what type of DB knowledge the users of the respective group need.

**[10 marks — word limit 200]**

- (c) DBMS architecture and abstraction:
- (i) Explain the layered architecture and the concept of data abstraction. Refer to the three schemas in data abstraction, and to the three layers of the DB architecture.
  - (ii) Explain in your own words what connects data abstraction and the DB architecture, and where the difference is.

**[10 marks — word limit 200]**

**Question 2****Database Design and Entity-Relationship Model (ERM)**

Scenario: The English tennis association wishes to offer to local tennis clubs a DB system that can manage several tasks in one database.

There are players (members) registered with local clubs. Address and contact details (phone number, email) are to be stored. Regarding membership, there are different types: full membership (no restrictions), and part-time membership (can play Mon to Fri during day time, but not after 1700, and not on week-ends; cannot participate in tournaments). The DB system must enforce that a member can only have one membership (full or part-time).

There are numerous activities the players participate in. For example, there are activities like the autumn tournament, and the Christmas tournament.

There are several league tables associated with each year: Women single, Men single, Men doubles, Women doubles, and Mixed doubles. Only club members can participate in league table activities.

There are other activities where non-club-members can participate. For example, summer camps for kids. Parents have to register themselves and their kids, and this registration does not mean membership.

The clubs have facilities (e.g. club house, coffee machine, parking). This information needs to be stored since only if there are some facilities, the club may host matches between the home-club and a guest-club.

- (a) (i) Explain briefly the mapping of a conceptual model to a logical model. The explanation should be general, i.e. not specific to the scenario to be modelled.
- (ii) Design a conceptual model (ER model) for the tennis scenario. Show your model in textual form (as shown on lecture slides and practised in the course work). Mention the cardinality and participation constraints you assume.  
Note that this exam does not require to submit an ER diagram.
- (iii) Map the conceptual model to the relational model. Explain each step, that is, for each table, explain its schema. In particular, explain where key migration occurs. Show the relational model before and after normalisation (normalise to 3rd NF).

**[15 marks — word limit 300]**

## (b) (i) Normalisation:

Given the following schema:

person(Id, Name, YearOfBirth, Age, City, PostCode, Country)

Primary key: person.Id

student(Id, Name)

Primary key: student.Id

student.Id refers to person.Id

We assume the schema is in 1st NF; we consider all attributes being atomic, even though post-code could be considered as a composite.

Show where the tables are not in 2nd NF or 3rd NF, and propose how to normalise the schema.

## (ii) Participation: Explain the participation constraint. For your explanation, use the following relationship: registers(Student, Course)

Explain the effect of the constraint for each of the four options:

total-total

total-partial

partial-total

partial-partial

**[10 marks — word limit 200]**

**Question 3****SQL**

Given the following schema of a database:

# Database Schema:

# This is a schema for storing data about bicycles and manufacturers.

manufacturer(Id, Name, Address);

# Primary Key: (Id)

# Examples for Name: Giant, Kettler, Rayleigh, ...

bike(Id, Type, Manuld);

# Primary Key: (Id)

# Examples for Type: e-bike, mountain-bike, road-bike, classic-bike

# Manuld: refers to manufacturer.Id

shop(Id, Name, City, PostCode, OpeningTimes, Description);

sells(ShopId, BikeId, AskingPrice);

# Primary Key: (BikeId)

# ShopId: refers to shop.Id

# BikeId: refers to bike.Id

sold(ShopId, BikeId, SoldPrice);

# Primary Key: (BikeId)

rankings(Source, Manuld, Rank);

# Primary Key: (Source, Manuld)

# Manuld: refers to manufacturer.Id

# Example: ("BikeReviews", m123, 5)

# BikeReviews is a site ranking bicycle manufacturers.

Provide SQL expressions for the text queries (query parts (a) to (d)). For all queries, make the join conditions an explicit part of the WHERE clause. Consider using views (or nested queries) where the query formulation becomes complex.

- (a) Show the names of shops that are in the city of Cambridge and sell mountain bikes.

**[5 marks — word limit 100]**

- (b) Show a list (table) where each tuple in the list shows the name of the shop, the average asking price for the mountain bikes it sells from manufacturer 'Giant', and the difference between the average asking price and the overall average asking price (that is the average asking price for mountain bikes from manufacturer Giant in all shops). For example, for the overall average asking price being 1,500 pounds, a tuple in the result list is:

'Bike Shed Cambridge' | 1,700 | -200

**[5 marks — word limit 100]**

- (c) Show the name and address (that is, city and postcode) of the bicycle shops that sell mountain bikes priced less than 1,000 pounds and road bikes priced less than 2,000 pounds, and e-bikes priced less than 3,000 pounds.

**[5 marks — word limit 100]**

- (d) Show all shops (their name, city, and postcode) that sell bikes of the top-3 bike manufacturers (all bikes the shops sells must be from one of the top-3 manufacturers). A manufacturer is considered to be among the top-3 manufacturers if one of the rankings assigns a rank of less or equal than 3.

**[5 marks — word limit 100]**

- (e) Given the following SQL query.

```
CREATE VIEW selectedTypes AS
SELECT bike.Type AS Type
FROM bike , manufacturer M
WHERE bike.Manuld = M.Id AND M.Name = 'Giant';
```

```
SELECT shop.Name
FROM shop , sells , bike , selectedTypes
WHERE shop.Id = sells.ShopId
AND sells.Bikeld = bike.Id
AND NOT EXISTS (SELECT *
                 FROM selectedTypes
                 WHERE Type != bike.Type);
```

Express the query in relational algebra.

**[5 marks — word limit 100]**

**Question 4****Constraints, Extended ERM, Transactions, Legal & Ethical Issues, DWH, NoSQL**

- (a) Constraints: There are three approaches for enforcing the referential integrity constraint: restrict, nullify, cascade.

- (i) Explain each approach briefly.
- (ii) Which approach do you prefer and why?

**[5 marks — word limit 100]**

- (b) Extended ERM:

- (i) Explain the two main approaches for modelling super-classes and sub-classes in the relational model. Show the approaches applied to the entities person, student, employee and manager.

A person has a Name, a YearOfBirth, and a Nationality. A student is a person, and has a StudentNumber, and is registered at one Uni. An employee is a person, and has a national insurance number. A manager is an employee, and has an attribute "YearsOfExperience".

- (ii) For this scenario, which of the two approaches would you prefer and why?

**[5 marks — word limit 100]**

- (c) Transaction Management: Two-phase locking (2PL) guarantees serialisability. Explain in your own words what this means. Refer to the names of the two phases, and refer to conservative 2PL. What does conservative 2PL prevent, and why?

**[5 marks — word limit 100]**

- (d) Legal & Ethical Issues: The UK Data Protection Act contains eight principles. The 1st, 5th and 6th principle are: (1) fair and lawful, (5) kept not longer than necessary, (6) rights of data subjects.

- (i) Why does it say "fair and lawful"? Is "lawful" not sufficient?
- (ii) Explain what "not longer than necessary" implies for the following scenarios: "storing CVs of job candidates", and for "crawling data from the Internet".
- (iii) What does "rights of data subjects" mean?

**[5 marks — word limit 100]**

- (e) Datawarehousing (DWH) and NoSQL:

- (i) List the main DWH schemas and describe each schema very briefly.
- (ii) NoSQL is schema-less. How would you explain what is the difference (or the common ground) between DWH and NoSQL?

**[5 marks — word limit 100]**

---

**End of questions**