

University of St Andrews



DECEMBER 2016 EXAMINATION DIET

SCHOOL OF COMPUTER SCIENCE

MODULE CODE: IS5102

MODULE TITLE: Database Management Systems

EXAM DURATION: 2 hours

EXAM INSTRUCTIONS: (a) Answer **three** questions.
(b) Each question carries 20 marks.
(c) Answer questions in the script book.

YOU MUST HAND IN THIS EXAM PAPER AT THE END OF THE EXAM.

**DO NOT TURN OVER THIS EXAM PAPER UNTIL
YOU ARE INSTRUCTED TO DO SO.**

1. Database Modelling

- (a) Draw an ER diagram to model the following scenario:

The Admissions department wants a database to record details of applicant students and courses they have applied for. Each student has a name, date of birth, country of origin, age, and previously held qualifications recorded. Students can be home students or international students. Home students (those whose country of origin is UK) additionally have their post code recorded, while international students (all other countries of origin) have their visa status recorded. Courses have a name (for example Information Technology, or Management), levels (a level can be MSc, BSc Honours, MA, or BA), and are offered by a school. Students can apply for multiple courses. For each course a student applies, they are assigned an advisor, who is a member of staff with a name and a telephone number. Each advisor can advise multiple students, for multiple courses.

Show clearly any cardinality or participation constraints for relationships in your diagram. State any assumptions you need to make. [8 marks]

- (b) Derive a relational schema for your ER diagram from part (a). Be sure to specify sensible attribute types, and any primary key, foreign key, and non-null constraints. [6 marks]
- (c) Give relational algebra queries over your relational schema from part (b) for the following queries:
- (i) Names of all students from Japan who have applied to study for a MSc in Information Technology. [2 marks]
 - (ii) Countries of origin of all students advised by John Thomson for courses offered by the School of Computer Science. [2 marks]
 - (iii) Courses applied for by students over the age of 30. [2 marks]

[Total marks 20]

2. SQL

Imagine you are designing a database for an estate agency. The following relations schemas have been defined:

```
property = (property_id, address, type, no_of_rooms)
owner = (owner_id, name, phone)
owns = (owner_id, property_id)
tenant = (tenant_id, name, phone)
rents = (tenant_id, property_id, rent_per_month)
```

- (a) Write suitable SQL DDL statements to create the tables as above. Include attribute types, primary and foreign key constraints. [8 marks]
- (b) Write SQL DML queries over the tables defined in part (a) to find:
 - (i) Names and phone numbers of all tenants of properties with 3 or more rooms. [3 marks]
 - (ii) The names of the top 5 owners in the order of most rent earned per month on their properties. [3 marks]
- (c) Write a SQL statement to increase the rent for all properties owned by “John Hoggan” by £ 50. [3 marks]
- (d) How would you use views and authorisation to allow a secretary at the estate agency to provide a reference for a tenant? Referees need to know tenant’s names, phone and address, but should not be able to see how much rent the tenant pays per month. [3 marks]

[Total marks 20]

3. Normalisation

This question is about the following unnormalised data:

Pa-tient Id	Patient Name	Doctor	Doctor's Phone	Med. No	Medicine Name	Dose	Units Per Day
1	James Li	Carol Knight	4268	120	Tetracycline	10mg	3
2	Nigella Dobson	Ravi Singh	3165	156	Morphine	10ml	5
				120	Tetracycline	20mg	4
				243	Codeine	5mg	6
				460	Avorstatin	10mg	2
5	Lisa Ray	Carol Knight	4268	156	Morphine	10ml	3

- (a) This data is unnormalised. Give two reasons why leaving it in this form is not useful, and convert the data to 1NF. [3 marks]
- (b) Define a functional dependency. Identify all the functional dependencies in your answer table for part (a). [4 marks]
- (c) Is there any update anomalies in your answer to part (a)? If so, give an example. If not, add another row to make a table that would have an update anomaly. [3 marks]
- (d) Convert the data from part (a) to 3NF, making clear each step. [8 marks]
- (e) What are two reasons we might prefer denormalised data? Illustrate by possible denormalisations of your answer from part (d). [2 marks]

[Total marks 20]

4. Data Mining

A book store approaches you to analyse their historical sales records and help improve their revenue and being a Data Mining expert you gladly agree.

- (a) Very briefly outline the steps that you would perform before commencing to analyse their datasets. [4 marks]
- (b) You find the dataset below which you think might be relevant for the firm.

Month Index, Sales (in pounds), Advertising (in pounds),
Number of Promotions, Number of Weekends

Briefly elaborate how you would use this dataset to the store's advantage and the method(s) that you would employ for the same. [6 marks]

- (c) You stumble upon a dataset that records customers buying behavior and realise that it has nearly two dozen attributes recorded. What techniques would you use to analyse this particular dataset? [4 marks]
- (d) Additionally, based on the given data you are also asked to group users based on their buying patterns. What method(s) would you consider to solve this problem? [2 marks]
- (e) The firm asks if it would be possible to predict the probability of a customer buying a book written by a particular author based on their demographics and their previous sales history. Do you think it is possible to do that? If so, how would you go about doing it? If not, why? [4 marks]

[Total marks 20]

***** END OF PAPER *****