University of St Andrews

DECEMBER 2020
48 HOUR ASSESSMENT
SCHOOL OF COMPUTER SCIENCE

MODULE CODE:	**IS5102**
MODULE TITLE:	Database management systems
TIME TO HAND IN:	48 hours
EXAM INSTRUCTIONS	a. Answer all questions b. Each question indicates the number of marks it carries. The paper carries a total of 60 marks.

This assessment consists of exam-style questions and you should answer as you would in an exam. As such, citations of sources are not expected, but your answers should be from your own memory and understanding and significant stretches of text should not be taken verbatim from sources. Any illustrations or diagrams you include should be original (hand or computer drawn). You may word-process your answers, or hand-write and scan them. In either case, please return your answers as a single PDF. If you handwrite, please make sure the pages are legible, the right way up and in the right order. Your submission should be your own unaided work. While you are encouraged to work with your peers to understand the content of the course while revising, once you have seen the questions you should avoid any further discussion until you have submitted your results. You must submit your completed assessment on MMS within 48 hours of it being sent to you. Assuming you have revised the module contents beforehand, answering the questions should take no more than three hours.

Some question may have word limits. These will be stated at the start of the question (or part question) and may be mandatory or advisory. Answers which exceed a **mandatory** word limit may be penalised at the rate of 5% of the available marks for being overlength and a further 5% for each 10% over the word limit. So if the limit on a 20 mark question was 1000 words, an answer of 1201 words would attract a 3 mark penalty. An **advisory** word limit is a guide to the level of detail and amount of information expected in an answer. Longer answers may lose marks for including large amounts of irrelevant material, or for failing to state arguments clearly and concisely.

1. **Database modelling:**

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

The consulting company "Research Frontiers" wants to create a database of scientific publications. Each publication is written by one or more authors, has a title, and is published in certain journal, where it can be located by the issue of the journal (given by a combination of its volume and number), page numbers for the first and last pages of the publication, and the date of publication. A publication is also identified by its unique DOI (Digital Object Identifier). A journal has a name and a website. Author information consists of their title, name, email, and their unique ORCID (Open Researcher and Contributor ID) code. It is also necessary to keep track of author affiliation, recording the name, address and a website of their institution.

"Research Frontiers" wants to provide an information service, which will help researchers to learn about new publications. For this purpose, they will recruit volunteer reviewers to write brief abstracts of new papers and submit them to the database. For each reviewer, they need to record their title, name, email and affiliation, and also their field(s) of expertise, which will be used to match the field(s) of publications when allocating them to reviewers. For each paper, there will be only one reviewer allocated at a time, and once they submit a review of a particular paper, no further reviews will be needed for this paper.

Assigning reviewers to publications will be managed by staff at "Research Frontiers". For the fair allocation process, it should be possible to check, for example the number of reviews submitted by a given reviewer within the last year. Also, an author should not be asked to review their own publication.

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 necessary primary key, foreign key, and non-null constraints. [6 marks]

(c) Give SQL queries over SQL tables corresponding to your relational schema from part (b) for the following queries (minor syntax errors will not be penalised, provided it is clear what you meant):

(i) List names and websites for all institutions in the database, ordered alphabetically by name. [2 marks]

(ii) List names and emails of all reviewers with expertise in "DBMS". [2 marks]

(iii) List titles and reviews of all articles written by at least one author from the University of St Andrews. [2 marks]

[Total marks 20]

2. **SQL:**

A wildlife conservation project needs to organise their data about observations of ringed birds in the UK. The following relational schemas have been defined:

```
bird (ring_id, species, date_ringed, age_ringed)

location (loc_id, loc_name, latitude, longitude, postcode, county)

staff (staff_id, name, address, phone)

observation (ring_id, loc_id, date, staff_id, weight, length, wingspan)
```

It is assumed that it is possible to record a sighting of a ringed bird with a known ring ID without taking any measurements.

For questions 2(a), 2(b) and 2(c) below minor syntax errors will not be penalised, provided it is clear what you meant.

(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 coordinates of all locations in the database situated in Fife, ordered by their postcode [2 marks]

(ii) Names of different bird species observed in the UK, ordered by the total number of their recorded observations [2 marks]

(iii) The number of swan observations in each county in 2019 [2 marks]

(c) Write SQL statement(s) to modify the `location` table that has been already defined in part (a), by adding the new attribute `country` and setting its value to "UK" for all locations contained in this table. [3 marks]

(d) (Advisory word limit of the answer for 2(d) is 200 words.) How would you use views and authorisation to allow the user of a database to explore the data about birds without being able to access the staff ID and further personal details of staff members recording the observations?

[3 marks]

[Total marks 20]

3. **Normalisation:**

This question is about the following unnormalized data about feeding animals temporarily left by their owners in a pets hotel:

Owner's name	Owner's phone	Species	Pet Id	Pet's name	Food	Food Storage Box No	Times per day/ Amount
Jerry Swan	3265	Gerbil	2	Mario	Muesli	2	1 x 10 g
					Seeds	7	1 x 5 g
			3	Luigi	Muesli	2	1 x 10 g
					Seeds	7	1 x 5 g
Aiko Suzuki	2466	Rabbit	9	Yoshi	Hay	8	2 x 500g
Diego Garcia	4677	Guinea pig	5	Fluffy	Hay	8	2 x 100g
					Seeds	3	2 x 20g
Tom Katz	7824	Cat	6	Larry	GoCat	1	3 x 100g

(a) (Advisory word limit of the answer for 3(a) is 200 words.) Explain why this data is unnormalized. 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) Are 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 2NF, and then to 3NF, making clear each step. [7 marks]

(e) (Advisory word limit of the answer for 3(e) is 200 words.) What are two reasons we might prefer denormalised data? Illustrate by possible denormalisations of your answer from part (d). [3 marks]

[Total marks 20]

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