University of St Andrews

DECEMBER 2018 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.

PERMITTED MATERIALS: Non-programmable calculator

<u>**YOU MUST HAND IN THIS EXAM PAPER AT THE END OF THE EXAM.**</u>

**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 ancient coin society would like a database to catalogue their findings of ancient coins being dug up. When they make a finding, they record it with a unique ID, and record the place (nearest town or village, latitude and longitude), and time of the finding. Each finding consists of one or more coins. Each coin will have information about material, time period it is from, weight (in grams), and a rating of its condition (1 being poor and 10 being perfect). We would also like to record the designs if any on the coins, which come in two types: a picture of something or somebody, typically a Roman emperor (where we should record the description), and some words (where we should record the language and the text). Either or both types of designs may be present. For each finding, we want to record who made the finding (at minimum their name, address and email or phone). Coins are stored in trays (which have identifying numbers), which are stored in cupboards in rooms (there are now multiple rooms in the collection). Rooms have room numbers, and cupboards have numbers as well (so a cupboard location may be something like "12-22", 12 for the room, 22 for the cupboard number within the room).*

   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 SQL queries over SQL tables corresponding to your relational schema from part (b) for the following queries:

   (i) List the weight and condition for all coins.   [2 marks]

   (ii) List all coins with a picture of 'Augustus Caesar' found before 2010, and stored in trays numbered between 70 and 80.   [2 marks]

   (iii) List the place of finding and name of the finder of the 5 coins in the poorest condition.   [2 marks]

   [Total marks 20]

2. **SQL**

(a) What do the SQL operations `inner join`, `left outer join`, `right outer join`, and `full outer join` do? Illustrate your answer by showing an example where they all give different results. [4 marks]

(b) A small company seeks your help to design a database for project management. The following relational schemas have been defined:

```
employee = (employee_id, name, address, date_of_birth, salary)
project = (project_id, name, budget, start_date, end_date)
manages = (employee_id, project_id)
works_on = (employee_id, project_id)
```

Write suitable SQL DDL statements to create the tables as above. Include attribute types, primary and foreign key constraints. [6 marks]

(c) Write SQL DML queries over the tables defined in part (b) to find:

(i) Names and addresses of all employees who work on projects which are current (have not ended yet). [2 marks]

(ii) The names of managers of projects, ordered by the total number of employees they manage. [2 marks]

(iii) The names of employees who work on the project with the largest budget amount. [2 marks]

(d) Write a SQL trigger over the tables defined in part (b) to ensure a single employee does not manage more than 5 projects. Either ISO standard SQL or MariaDB syntax will be accepted. [4 marks]

[Total marks 20]

3. **Unorganised data**

The Student Advice Center has a large number of student information files on paper. Each file has the following information:

   `Matric No, Name, Address, Degree Registered For, Date of Birth`, group of associated `Contacts`.

The contacts associated with each student file is just a bunch of sheets recording when the student had contact with the Advice Center, and there are usually many such sheets in each file. Each sheet records one contact and has

   `Date of Contact, Advisor, Advisor's Phone, Chief Issue, Action Items`.

There is one file per student seen.

(a) Define the concept of a functional dependency in the relational model. What functional dependencies would you expect to have in the data in the student files above? [4 marks]

(b) Design a relational data model for the data from the student files. Make sure your answer is in 3rd Normal Form, and justify why it is. [8 marks]

(c) How might you design Documents in a Document-Oriented Database like MongoDB for storing the data from the student files? Discuss what the common queries on the database might be, and how you would denormalise the data in MongoDB to handle such queries efficiently. (You are free to use MongoDB syntax if you wish, but incorrect syntax will not be penalised as long as your data design is clear.) [8 marks]

[Total marks 20]

4. **Data Models for Analysis**

An online-only bookstore approaches you to sort through their data on website-driven sales. They hold data on site visits (referring website, date of visit, book pages visited, customer email address if user is logged in), books (title, author, publisher), and sales (date of sale, books sold, customer address and email). Note that not all site visits have email address data as the user may or may not be logged in, but all sales do have an associated customer email address.

(a) Sketch out a relational data model for this scenario. Make sure to specify your data types, primary keys and foreign key constraints. State any assumptions you need to make. [6 marks]

(b) Describe the Data Mining technique of Link Analysis, and describe how this technique can be used in this scenario. [4 marks]

(c) How would you analyse the website data to find which visits led to sales shortly afterwards? Describe the query or queries you would need over your relational model. Formal relational algebra or SQL notation is not required. [6 marks]

(d) Describe a graph model of the website data (for example in Graph Databases like Neo4J) that could be used to easily find associations between site visits and sales, and describe how such associations can be identified in your model. [4 marks]

[Total marks 20]

*** END OF PAPER ***