Birkbeck

(University of London)

BSc/FD EXAMINATION

Department of Computer Science and Information Systems

Database Management (COIY028H6)

CREDIT VALUE: 15 credits

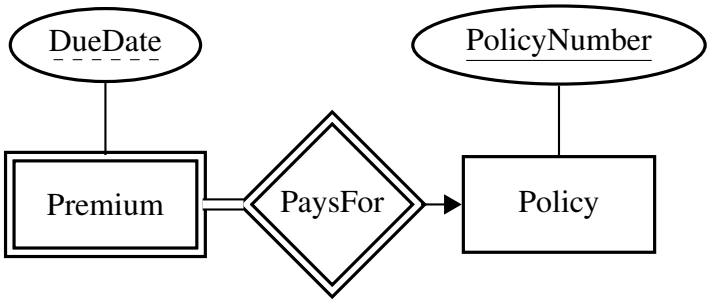
Date of examination: 30 May 2019
Duration of paper: (13:30–15:30)

There are **five** questions on this paper.
Answer only **four** of the five questions.
If you answer more than four questions, only the best four answers will count.
Each question carries **25** marks in total.
The paper is not prior-disclosed.
The use of electronic calculators is not permitted.

1. (a) Name the three components that are considered to be part of any *data model*. Comment on the extent to which the relational and entity-relationship models conform to this view.

(5 marks)

(b) Consider the following fragment of an *Entity-Relationship Diagram* representing premium payments for insurance policies:



i. What is the *cardinality* (or *multiplicity*) of the PaysFor relationship type?

(1 mark)

ii. What kind of entity type is the Premium entity type and what does this mean?

(3 marks)

iii. What kind of relationship type is the PaysFor relationship type and what purpose does it serve?

(3 marks)

iv. What *participation constraints* for the PaysFor relationship type are implied by the notation used in the above diagram?

(2 marks)

v. How would the Premium entity type be represented in the relational model (i.e., give the relation schema and any necessary constraints)? You can assume that the Policy entity type is represented by a relation schema of the same name.

(5 marks)

(c) Explain the purpose of the `outer join` operator in SQL. What is the difference between the `left`, `right` and `full outer join` operators?

(6 marks)

2. (a) Consider two relations R and S, containing attributes A and B, respectively. Attribute A is the primary key of relation R, while attribute B is a foreign key referencing A. Assume that the database employs the default handling of foreign key constraints.

   i. Explain what it means to say that "B is a foreign key referencing A".

   (3 marks)

   ii. Under what conditions would a user be allowed to delete a row from R?

   (2 marks)

   iii. Under what conditions would a user be allowed to delete a row from S?

   (2 marks)

   iv. Now consider joining R and S on attributes $R.A$ and $S.B$. If R contains n rows and S contains m rows, how many rows would you expect in the join? Explain your answer.

   (3 marks)

   (b) Explain the notion of *serialisability* with respect to transactions being executed by a database system. Use an example to illustrate a sequence of operations that is not serialisable.

   (6 marks)

   (c) Give two examples of features available in *object-relational* database systems which violate the notion of *first normal form* (1NF) from pure relational systems, explaining in what way they violate 1NF.

   (4 marks)

   (d) Relational database systems are sometimes seen as inflexible in that they require a fixed and "flat" schema for each relation in the database. Name two categories of so-called NoSQL database systems which address this criticism, and explain in what way(s) they do so.

   (5 marks)

3. (a) Consider an example relation schema Reviews, with schema(Reviews) = {UserID, Book, Rating, Author}, representing users submitting ratings for books where the author of the book is also recorded. Let a relation over Reviews be given by

UserID	Book	Rating	Author
123	Amsterdam	Good	McEwan
456	Amsterdam	Excellent	McEwan
456	Disgrace	Good	Coetzee
789	Disgrace	Average	Coetzee

We assume that (A) a user (represented by UserID) is not allowed to rate the same book more than once, and that (B) each book has only a single author.

i. Write down the above two constraints as *functional dependencies* (FDs).

(2 marks)

ii. Explain what the *redundancy problem* is in the Reviews schema.

(2 marks)

iii. Find a *key* for the Reviews schema, explaining precisely why it is a key.

(3 marks)

iv. Produce a Boyce-Codd Normal Form (BCNF) decomposition of the Reviews schema, explaining each step of the process.

(5 marks)

v. Consider decomposing Reviews into {UserID, Book, Rating} and {Rating, Author}. Give two reasons why such a decomposition would be bad. You should answer in terms of the consequences of such a decomposition, not simply state that it would not in BCNF or 3NF.

(4 marks)

(b) Let F be the following a set of *functional dependencies* (FDs):

$$C \to T,$$
$$HR \to C,$$
$$HT \to R,$$
$$CS \to G,$$
$$HS \to R$$

Use the *closure algorithm* to determine the closure of HS, showing each step of the process.

(5 marks)

(c) What are *null* values used for in the relational model? Explain how null values are treated by the SQL operators `max` and `group by`.

(4 marks)

4. (a) Consider a relation schema Course, with schema(Course) = {Code, Name, Level, Textbook}, where Code is the course code, Name its name, Level its FHEQ level (4, 5, 6, 7), and Textbook the name of a textbook recommended for the course. Assume we have established that the following functional dependencies (FDs) hold:

$$\text{Code} \rightarrow \text{Name}$$
$$\text{Code} \rightarrow \text{Level}$$
$$\text{Name} \rightarrow \text{Code}$$

i. Identify two *keys* for the Course schema, demonstrating why they are keys.

(4 marks)

ii. Decompose the Course schema into *Third Normal Form* (3NF), explaining each step of the process.

(6 marks)

(b) Consider a B-tree index in which at most 6 (key) values can be stored in any node. What is the minimum number of (key) values that would be stored in any (non-root) node? Assume that one node in the B-tree contains the values $(17, 19, 22, 23, 26, 28)$. Explain in detail what happens when the value 25 is inserted into the tree.

(8 marks)

(c) When SQL is used along with a language such as PHP, there is a possibility of using *placeholders* in the SQL. Explain the purpose of placeholders. Name the two different kinds of placeholder, demonstrating their use with examples.

(7 marks)

5. (a) Assume that we have the following three relation schemas

```
Module(m_code, m_name)
Student(s_id, s_name)
Result(s_id, m_code, grade)
```

with primary keys `m_code`, `s_id` and (`s_id`, `m_code`), respectively. Relation `Module` associates each module code with its name, relation `Student` associates each student id with a student name, and relation `Result` stores the grade achieved by each student in each module taken. Now consider the following view definition:

```
create view Transcript(student, module, mark) as
select S.s_id, M.m_code, grade
from Student S, Module M, Result R
where S.s_id=R.s_id and M.m_code=R.m_code;
```

i. What is the name given to the identifiers `S`, `M` and `R` above, and why are they needed in this view definition?

(3 marks)

ii. Rewrite the `from` and `where` parts of the view definition using SQL `join` expressions.

(3 marks)

iii. State why, from a syntactic point of view, the above view definition is not *updatable* according to the SQL-92 specification.

(2 marks)

iv. It turns out that the above view definition *can* be updated, according to more recent SQL specifications. Consider the following `insert` statement:

```
insert into Transcript
       values ('123456', 'CS123', 85);
```

Explain why there will not be a problem executing the above `insert` statement, as long as certain other conditions on the relation schemas (table definitions) are satisfied. Identify the necessary conditions.

(7 marks)

(b) Give a short explanation of what is meant by each of the following terms (2 marks each):

i. natural join

ii. data independence

iii. transaction isolation level

iv. three-valued logic

v. correlated subquery

(10 marks)

