UNIVERSITY OF LONDON

BSc EXAMINATION 2018

For Internal Students of
Royal Holloway

DO NOT TURN OVER UNTIL TOLD TO BEGIN

CS2855: Databases

Time Allowed: **2 hours**

Answer ALL questions

Calculators are not permitted

1. The computer science department at Royal Holloway has granted you the opportunity to design its database. The department first sets out the E-R model for the database.

   In this E-R model there will be a record for each student, listing his or her student-ID, full name, programme (e.g., AI, InfoSec, CS), email address, and starting academic-year (e.g., 201617). There will also be records for modules (i.e., courses), listing the module-name, module-ID, academic-year, number-of-assignments. The same module can run in different years, and these should be treated thus as two different entities. For each student, the department should record the modules he or she has taken together with the attempt number for each module (i.e., whether it is the first attempt for taking the module, the second attempt, and so forth).

   (a) Draw an appropriate E-R diagram for the above setting. Use only the information provided in the textual description above. Make sure to identify correctly all the constraints in the diagram. Explain very briefly your design decisions. [8 marks]

   (b) Convert the E-R diagram from part (a) to *tables/relations in the Relational-Data model*. In other words, show all the relevant *table-schemas* written as a sequence of attributes, specifying their integrity constraints (e.g., primary and foreign keys), while also aiming not to add unnecessary tables and attributes. [8 marks]

   (c) The department now wishes to record specific assignments for each module, listing the number of the assignment with respect to the module (e.g., assignment2 in Databases module). The E-R model needs to allow the department to record the assignment of each student in each module (in each year), together with the mark.

   Draw the E-R diagram for this part of the database taking into account the constraints mentioned. You do **NOT** need to draw the diagram related to entities and relations that are not affected by the above consideration.

   [8 marks]

2. Consider the following relational tables:

student = (studentID, acYear, name, programme)

module = (moduleID, acYear, moduleName, numAssignments)

studentModule = (studentID, moduleID, acYear, finalMark)

Foreign Key: studentID referencing studentID in student table.

In what follows please **do not** use in your SQL queries an INNER JOIN operation. You can use JOIN (whose semantics is the natural join operation).

(a) Write the SQL definition of the relation student (including integrity constraints); in other words, create the table in **SQL**. Assume that student IDs are integers, and programmes are alphanumeric of max eight characters. [8 marks]

(b) Express the following query in **SQL**: "What is the name of the student(s) who achieved the top final mark in 'Databases' module (whose moduleID is 'cs2855') in the academic year 201718?" [10 marks]

(c) Express the following query in **Relational Algebra**:

"What are the names of the students who took the same module in the two consecutive years 2016 and 2017, together with the name of the module they repeated?" [10 marks]

(d) Consider the following instance for the table studentModule:

studentID	moduleID	acYear	finalMark
101	3	201718	70
199	1	201718	80
199	2	201718	60
201	1	201718	100

What will be the answer retreived when running the SQL query below? Explain briefly your answer (by describing in plain English what the query below asks for).

```
SELECT z.moduleID, z.avrg
FROM (SELECT    moduleID, AVG(finalMark) as avrg
        FROM    studentModule
        GROUP BY moduleID) AS z
WHERE avrg =  (SELECT MAX(y.avrg)
                      FROM
                      (SELECT   moduleID, AVG(finalMark) as avrg
                        FROM    studentModule
                        GROUP BY moduleID) AS y
                )
```

[10 marks]

3. (a) Define formally the notion of a lossless decomposition of a relation R.

[10 marks]

(b) Given below is the set F of functional dependencies for the relational schema: $R = \{A, B, C, D, E, F, G\}$.

$$F = \{AB \to C, BC \to D, G \to F, AE \to FG\}$$

i. Find a candidate key for R with respect to F.

ii. Show two nontrivial functional dependencies that follow from F (but are not in F).

iii. Decompose the relation into a collection of relations that are in BCNF.

[14 marks]

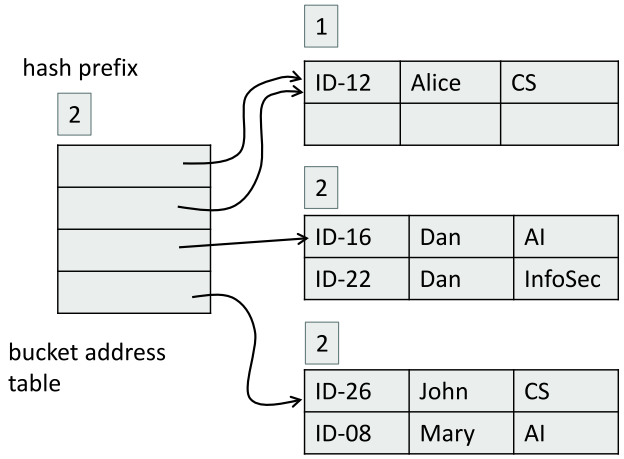
(c) Determine whether the following relation schema R is in 3NF with respect to the following set F of functional dependencies, and explain very briefly your answer:

$$R = (A, B, C, D, E)$$

$$F = (B \to CD,\ A \to E)\,.$$

[8 marks]

4. An extendible hashing table, together with the hash function (on the right), is shown in the following figure:



name	*h*(name)
Alice	0010
Dan	1010
John	1100
Mary	1111
Noel	0011

Suppose that we **insert** into the hashing table the following information

(ID-13, Mary, InfoSec) .

Which of the following figures correctly depicts the hashing table after the above insert command? Explain very briefly your answer. [6 marks]

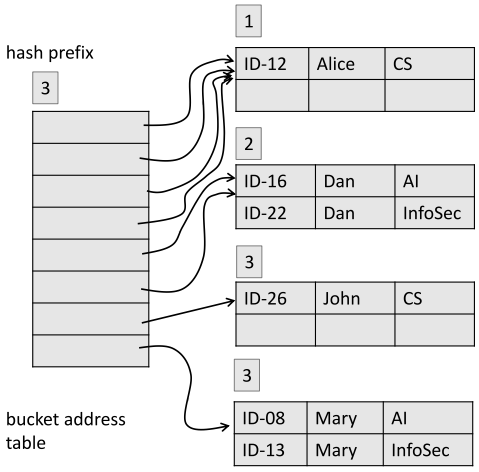


Figure 1

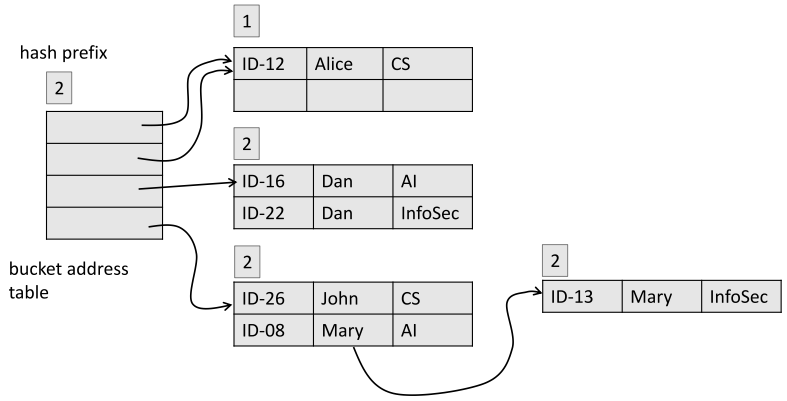


Figure 2

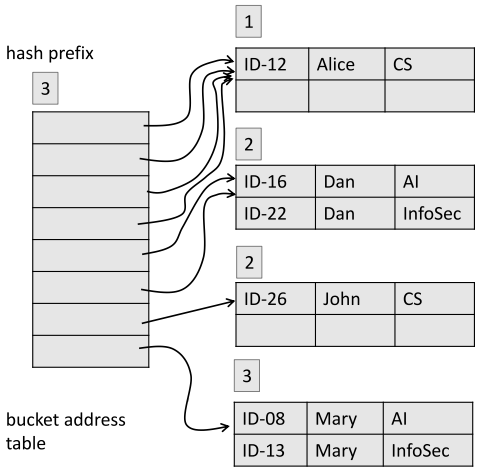


Figure 3

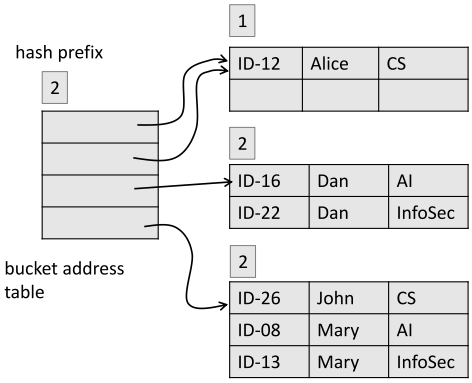


Figure 4

END