UNIVERSITY OF LONDON

BSc EXAMINATION 2023

For Internal Students of
Royal Holloway

DO NOT TURN OVER UNTIL TOLD TO BEGIN

CS2855/CS2855R: Databases – for FIRSTSIT/RESIT CANDIDATES

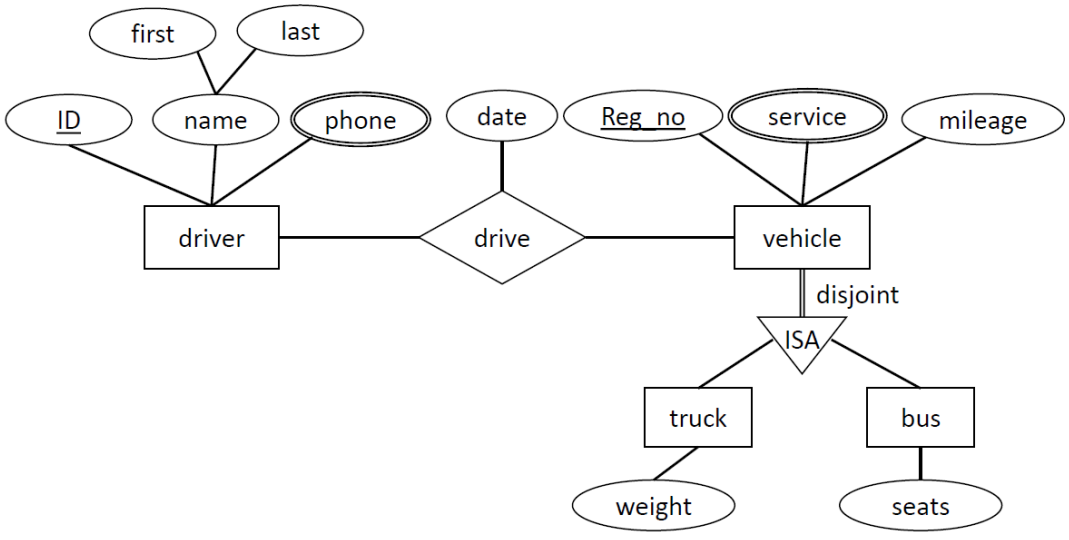
Time Allowed: **TWO hours**

Please answer **ALL** questions

1. (a) The department of Computer Science needs a database for monitoring the Final Year Projects of the students and their supervisors. Its desired design is described below.
   - For every staff member the system stores their username, which is unique, and a list of emails.
   - For every student the system stores their academic year and their email address.
   - Each student is associated with two staff members: one staff member acts as the supervisor, another staff member acts as a marker. There can exist staff members that are not supervising or marking any students.
   - In addition, the system stores the available topics for the Final Year Projects. For every topic, the system stores its ID and its description. Finally, the system stores the topic of the Final Year Project of every student. There can exist topics that no student studies. Every student studies exactly one topic.

   Draw an E-R diagram according to the above design. Remember to include all constraints. [15 marks]

   (b) Convert the following E-R diagram into a relational model. Write down the relational model, including primary key and foreign key constraints. Minimize as far as possible.



[20 marks]

NEXT PAGE

2. The following is a collection of relations that store information about an online coach ticket-booking system. Users book coach tickets and gain points for each trip that they make. Every attribute which shares its name with the primary key of a different relation is a foreign key.

trip			
tri_id	operator	miles	duration
N1	North Line	403	360
N2	North Line	212	210
E1	Eastern	95	100
W1	Fast West	103	120

user		
user_id	name	points
1226	Kate	1504
1023	John	2207
3322	John	3218
5017	Alex	985

booking		
tri_id	user_id	seats
N2	1226	1
E1	3322	2
E1	1226	1
W1	3322	3

(a) Evaluate and give the results of the following relational algebra expressions:

i. $\pi_{\text{user\_id, seats}}(\text{booking})$ [2 marks]

ii. $\text{trip} \bowtie \text{booking}$ [3 marks]

(b) Translate the following into equivalent expressions in Relational Algebra.

i. Find the user_id of all users who have booked tickets with both North Line and Eastern. [5 marks]

ii. Find the user_id and the name of all users with more than 1000 points. [5 marks]

(c) Write SQL statements for the following tasks, over the above relations. Your statements should be correct for general instances of the above schema and may not depend on the specific contents of the example tables above.

i. Delete all trips without any tickets booked. [5 marks]

ii. Increase the points of every user with less than or equal to 1500 points by 10% and the points of every user with more than 1500 points by 15%. [5 marks]

(d) Write SQL statements for the following tasks, over the above relations. Your statements should be correct for general instances of the above schema and may not depend on the specific contents of the example tables above.

i. Output the name and the points of every user that has more than 2000 points. The output should be ordered by points in ascending order. [5 marks]

ii. Output the tri_id alongside the total number of booked seats. E.g., for the tables above the output of the query should be

```
tri_id  seats
=============
N2      1
E1      3
W1      3
```

[5 marks]

iii. Give a list of operators and show for each operator the number of trips it offers, ordered by the number of trips in descending order. E.g., for the tables above the output of the query should be

```
North Line  2
Eastern     1
Fast West   1
```

[5 marks]

iv. Output the user_id of every user that has booked tickets associated with at least five different operators. [5 marks]

v. The company wants to keep track of the delays for the trips. Define a relation for the following table
delay=(delay_id, tri_id, date, delay);
delay_id consists of at most 6 letters and the delay is measured in minutes. [5 marks]

3. (a) Consider the following schedule of transactions T1 and T2. Is the schedule serialisable? If yes, provide an equivalent serial schedule. If no, briefly explain why. [5 marks]

```
T1       | T2
-----------------
         | Read(B)
Read(A)  |
         | Read(C)
         | Write(C)
Write(A)|
         | Write(B)
Read(B)  |
Write(B)|
```

(b) Consider the relation $R = (A, B, C, D)$ and the set of functional dependencies $F = \{A \rightarrow B, C \rightarrow D\}$. Is the decomposition of R to $R1 = (A, B)$ and $R2 = (A, C, D)$ lossless? Explain why. [5 marks]

(c) Write the definition of 3NF (Third Normal Form). [5 marks]

END