

Name: \_\_\_\_\_ , \_\_\_\_\_  
(Family name) (Given name)

Student ID: \_\_\_\_\_

THE UNIVERSITY OF NEW SOUTH WALES  
Final Exam

**COMP9311**  
**Database Systems**

**TERM 3, 2020**

- 
- Time allowed: **2 hours**
  - Reading time: **10 minutes**
  - Total number of questions: **5**
  - Total number of marks: **100**
  - Answer **all** questions.
  - You can answer the questions in any order.
  - Start each question on a **new page**.
  - Answers must be written in ink.
  - Answer these questions in the script book provided.
  - Do **not** write your answer in this exam paper.
  - If you use more than one script book, fill in your details on the front of *each* book.
  - You may **not** take this question paper out of the exam.
-

## Question 1

(20 marks)

For each of the following statements, indicate whether it is true or false. (Answer true or false only; You will get 2 marks for each correct answer, -1 mark for each wrong answer and 0 mark for no answer.)

- 1) Self defined data type is not allowed in PostgreSQL.
- 2) In SQL, using the command *AS* will change the database.
- 3) Any relation in BCNF is also in 2NF.
- 4) A primary key must consist of only one attribute.
- 5) A lossless and dependency-preserving decomposition into 3NF is not always possible.
- 6) Both using OS file system to manage disk space and keeping track of free blocks can improve disk access.
- 7) SQL cannot control sequences of database operations.
- 8) A hash index is suitable for range searches.
- 9) ISAM is a dynamic structure and uses leaf pages to store data.
- 10) As a concurrency control method, optimistic control is a good option if there is a lot of interaction between transactions.

## Question 2

(18 marks)

(a) (10 marks) Consider the following course-enrolment database:

- A course is uniquely identified by its course ID. For each course, we also record these attributes: course name, lecturer name, and total capacity. We are interested in the total number of students who enrol in each course as well.
- Each course can have zero or more labs. A lab is uniquely identified by a lab ID. Each lab contains these attributes: time, location and weeks. Lab location is composed of a building name and a room number, and a lab can be held in multiple weeks.
- A student is uniquely identified by his/her zID. Each student also has the attributes: name, contact number, email and current degree.
- A lab tutor is uniquely identified by his/her ID. Each lab tutor also has the attributes: name, contact number, email and salary.
- A student can enrol in zero or more courses and can enrol in zero or more labs.
- Each lab must belong to exact one course and must have at least one tutor on duty.
- A lab tutor can work for zero or more courses and zero or more labs.
- A course or a lab must have at least one student. A course can have multiple lab tutors but is not compulsory to have lab tutors.

Draw an ER diagram to represent this scenario and clearly state the assumptions you make if any. Note: please use the drawing conventions in the lecture notes.

(b) (8 marks) Translate the ER diagram of the above question into a relational schema. Note: please use the drawing conventions in the lecture notes.

### Question 3

(30 marks)

- (a) (18 marks) Consider the relational schema  $R(A, B, C, D, E, G, H, I, J)$  and a set of functional dependencies  $F = \{BD \rightarrow CH, BC \rightarrow HI, EI \rightarrow H, H \rightarrow AB, I \rightarrow E, EJ \rightarrow I\}$ . Note that  $A, B, C, D, E, G, H, I, J$  are attributes.
- Find a minimal cover  $F_m$  for  $F$ . (3 marks)
  - Compute all the candidate keys of  $R$  with respect to  $F$ . (4 marks)
  - Regarding  $F$ , is the decomposition  $R_1 = \{ABCDH\}$ ,  $R_2 = \{EGHIJ\}$  of  $R$  lossless-join? Please justify your answer. (5 marks)
  - Determine the highest normal form of  $R$ . If  $R$  is not in the BCNF, decompose  $R$  into BCNF and indicate a key in each of the result tables by underlining the attributes. (6 marks)
- (b) (12 marks) Consider the following relational schema for a restaurant-review website:
- Customer(cID, name, age, gender, email),
  - Restaurant(rID, location),
  - Review(rID, cID, content)

Below are detailed descriptions for the fields in each schema:

- Customer: For each customer, we record cID, name, age, gender and email. cID is the primary key.
- Restaurant: For each restaurant, we record rID and its location. rID is the primary key.
- Review: For each review, we record the rID of restaurant, the cID of customer, and the content of the review. The combination of rID and cID is the primary key.

Use relational algebra to answer the following questions:

1. List the rID of restaurants that are only reviewed by male customers or only reviewed by female customers. (4 marks)
2. List the cID of customers who have reviewed all the restaurants or never review any restaurant. (4 marks)
3. List the average age of the customers of each gender who have reviewed at least one restaurant. (4 marks)

## Question 4

(18 marks)

- (a) (8 marks) Consider the lock request sequence given below.  $RL(\cdot)$  and  $WL(\cdot)$  stand for “read lock” and “write lock”, respectively.  $T_1$ ,  $T_2$ ,  $T_3$  and  $T_4$  represent four transactions.

time	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$	$t_7$	$t_8$
$T_1$				WL(B)		RL(A)		
$T_2$		RL(B)					WL(C)	
$T_3$			WL(A)		WL(C)			
$T_4$	WL(C)							RL(B)

- Give the wait-for graph for the given lock requests. (4 marks)
  - Determine whether there exists a deadlock in the lock requests and briefly explain why. (4 marks)
- (b) (10 marks) Consider the schedule below. Here,  $R(\cdot)$  and  $W(\cdot)$  stand for ‘Read’ and ‘Write’, respectively.  $T_1$ ,  $T_2$ ,  $T_3$  and  $T_4$  represent four transactions and  $t_i$  represents a time slot.

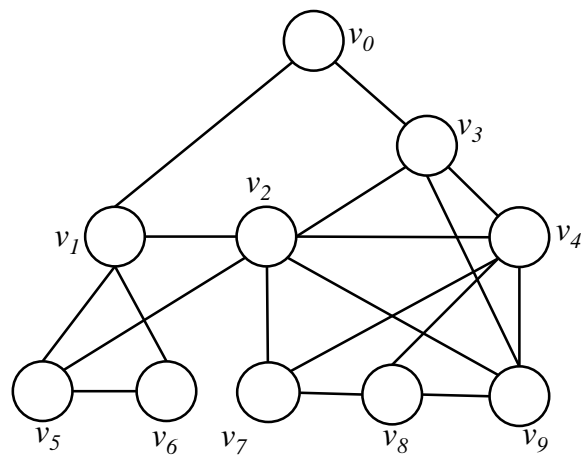
time	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$t_6$	$t_7$	$t_8$	$t_9$
$T_1$				R(Z)				R(Y)	R(X)
$T_2$	W(Y)		W(X)						
$T_3$		W(Y)			R(X)				
$T_4$						R(Y)	W(Z)		

- Give the precedence graph of this schedule. (5 marks)
- Is this schedule conflict serializable? If your answer is “yes”, please provide the equivalent serial schedule. If your answer is “no”, briefly explain why. (5 marks)

## Question 5

(14 marks)

- (a) (6 marks) Consider three buffer replacement policies: ‘Least Recently Used’, ‘Most Recently Used’ and ‘First In First Out’. Please answer the following questions:
- **Construct an example** that ‘Most Recently Used’ buffer replacement policy performs the best among these three buffer replacement policies. (3 marks)
  - **Construct an example** that ‘First In First Out’ buffer replacement policy performs the best among these three buffer replacement policies. (3 marks)
- (b) (8 marks) Consider the following graph  $G$ :



- Draw the 2-core of  $G$ . (2 marks)
- Draw the 3-core of  $G$ . (2 marks)
- In graph theory, triangle is defined as an undirected graph with exactly 3 vertices and 3 edges. In addition, given an undirected graph  $G$ , the  $(k, d)$ -core  $H$  of  $G$  is a subgraph that each vertex in  $H$  has at least  $k$  neighbors and each edge in  $H$  is contained in at least  $d$  triangles. Based on the definition of  $(k, d)$ -core, draw the  $(3, 2)$ -core of the above graph  $G$ . (4 marks)

**END OF EXAM PAPER**