Final Exam

Please make sure that you always use notations consistent with lecture notes. Different notations will not be accepted. The deadline for final exam is:
Thu 3, Dec 8:30 PM Sydney Time

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) In ER, a relationship type can have no key.
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) Frequent checkpoints can reduce the amount of data lost if a disk is destroyed.
9) The time-stamp ordering method never produces deadlocks.
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 hotel database:

- Each hotel is uniquely identified by its hotel ID. For each hotel, we also record these attributes: name, e-mail, contact number, owner and address. The Address is composed of a suburb and a street. Each hotel can have multiple owners.
- There are a lot of staff working at the hotels. A staff is uniquely identified by his/her ID. For each staff, we also record his/her name, phone number and email. Each staff works in one or more hotels.
- Each hotel owns one room or more rooms, and each room is identified by the room number and the hotel ID. We also record the phone number and the description of each room.
- A customer is uniquely identified by his/her email. For each customer, we also record his/her name, phone number and gender.
- A reservation is uniquely identified by a reservation ID. A reservation must be created by exact one customer. A customer can have zero or more reservations. A reservation must involve one hotel, and a customer cannot reserve the rooms from multiple hotels in one reservation. For each reservation, it must contain one or more rooms. We record the number of reserved rooms, total price, the date of arrival and date of departure. Besides, each hotel and room can be included in zero or more reservations.

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 model. 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 → CH,BC → HI,EI → H,H → AB,I → E, EJ→ I}. Note that A, B, C, D, E, G, H, I, J are attributes.

1. Find a minimal cover F_m for F. (3 marks)
2. Compute all the candidate keys of R with respect to F. (4 marks)
3. Regarding F, is the decomposition R_1 = {ABCDH}, R_2 = {EGHIJ} of R lossless-join? Please justify your answer. (5 marks)
4. 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 park-visit records:

- Visitor (<u>vID</u>, name, age, gender),
- Park (<u>pID</u>, name),
- Visit (<u>pID, vID, date</u>)

Below are detailed descriptions for the fields in each schema:

- Visitor: For each visitor, we record vID, name, age and gender. vID is the primary key.
- Park: For each park, we record pID and its name. pID is the primary key.
- Visit: For each visit, we record the pID of park, the vID of visitor and visit date. The combination of pID, vID and date is the primary key.

Use relational algebra to answer the following questions:

1. List the total number of the visitors who have visited all the parks. (2 marks)
2. List the name of all parks that are only visited by senior visitors (age>=65) or only visited by junior visitors (age<=24). (3 marks)
3. List the name of all parks that 'Daniel' has visited at least two more times than 'James' or 'James' has visited at least two more times than 'Daniel'. (Note that, you need to include the parks that A has visited for more than two times while the parks are never visited by B). (3 marks)
4. Find the names of the oldest visitors that have visited all the parks. Note that, in this question, you are **not allowed to use division** (i.e., only the following operators can be used in question 4: Select, Project, Union, Intersection, Difference, Cartesian Product, Join and Aggregation) (4 marks)

Question 4 (18 marks)

(a) (9 marks) Consider the schedule below. Here, R(·) and W(·) stand for 'Read' and 'Write', respectively. T_i represents a transaction 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		W(Y)		R(Y)	R(X)				
T_2						R(Y)		W(Z)	
T_3							R(Y)		R(X)
T_4	W(Y)			W(X)					

1. Give the precedence graph of this schedule. (3 marks)
2. Is this schedule conflict serializable? If your answer is "yes", please provide the equivalent serial schedule. If your answer is "no", convert it to a conflict serializable schedule with the minimum change in the scheduling while preserving the given ordering of operations in each transaction. (3 marks)
3. Given following two new transactions, please provide a schedule of T_5 and T_6 that is serializable but not conflict serializable and give the reasons. (3 marks)

   T_5 : R(A); W(B); R(B)

   T_6 : R(A); W(A); W(B)

(b) (5 marks) Consider the lock request sequence given below. RL(·), WL(·) and UL(·) stand for "read lock", "write lock" and "unlock", 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(A)	UL(A)			RL(A)		
T_2	WL(C)				WL(A)			
T_3	RL(B)	UL(B)		WL(A)				RL(B)
T_4			RL(B)				WL(C)	

1. Give the wait-for graph for the given lock requests. (3 marks)
2. Determine whether there exists a deadlock in the lock requests and briefly explain why. (2 marks)

(c) (4 marks) Given a schedule:

$$w_1(Y);\ w_2(X);\ r_3(Z);\ r_1(Y);\ r_4(X);\ w_4(X);\ r_2(Z);$$

Could the schedule be created by the 2PL protocol? If yes, augment the schedule with lock (read/write) and unlock actions to show it can be generated by 2PL. If not, explain briefly why.

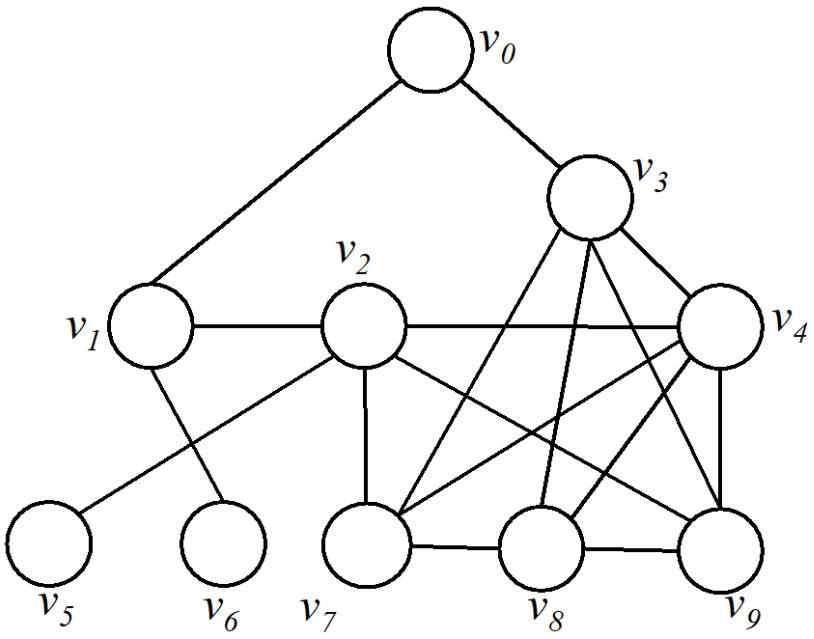
Question 5 (14 marks)

(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 better than 'Most Recently Used' but worse than 'Least Recently Used'. (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 neighbours and each edge in H is contained in at least d triangles. Based on the definition of (k; d)-core, draw the (2; 2)-core of the above graph G. (4 marks)

Final Exam Submission

- Students must submit an electronic copy of their answers to the above questions to the course website in Moodle.
- Only **.doc** or **.pdf** file is accepted. The file name should be final_**studentID**.doc or final_**studentID**.pdf (e.g., **final_z5100000.doc** or **final_z5100000.pdf**).

Note:

1. For any problems in submissions, please email to comp9311unsw@gmail.com
2. All submissions will be checked for plagiarism.
3. We do not accept e-mail submissions.

Late Submission Penalty

0 mark.