Name: _____ , _____
(Family name) (Given name)

Student ID: _____

THE UNIVERSITY OF NEW SOUTH WALES
Final Exam

COMP9311
Database Systems

TERM 1, 2022

- Time allowed: **3 hours (2PM to 5PM, Sydney time. 7th May)**
- Total number of questions: **6**
- Total number of marks: **100**
- Answer **all** questions.
- You can answer the questions in any order.
- Start each question on a **new page**.
- We accept any format: directly answering using word, or handwriting and convert to word or pdf. We only require the file to be clear and in **.doc** or **.pdf**.
- Submit your answer file via Moodle. You may submit multiple copies and we will mark the LAST one.

Question 1 (20 marks)

(a) (20 marks) For each of the following statements, explain your answer. The explanations should be concise descriptions of your understanding. Greater marks will be awarded for answers that are simple, short, and concrete than for answers that are convoluted. Marks will be lost for giving information that is irrelevant to a question.

i. Describe the difference between simple attributes and composite attributes. Justify the need for composite attribute types. *(4 marks)*

ii. Discuss the following SQL statements. Explain whether they will always produce different results (hint: consider cases of relation schema R1 and R2). *(4 marks)*

```
select A from R1 natural join R2;
select A from R1, R2;
```

iii. Give an example and a counterexample for the atomic property in ACID. *(4 marks)*

iv. Briefly explain the how the requirements of 3NF enforce the 2NF (hint: you may frame your discussion by the types of functional dependencies 3NF disallows) *(4 marks)*

v. Suppose that we need to find the maximum A attribute value of a relation R (A, B). Consider the following two SQL queries.

```
select A from R order by A desc;
select max(A) from R;
```

where "desc" means for descending order such that the max values will appear at the top. Which is the most appropriate way to answer the query, (a) the former, (b) the latter, (c) both? Give your answer with justification. *(4 marks)*

Question 2 (16 marks)

(a) (8 marks) Draw an ER diagram to represent the following set of application requirements (A museum database). You must use the drawing conventions in the lecture notes. Clearly state any additional reasonable assumptions if you make any.

- Each painting is uniquely identified by its painting ID. For each painting, we also record its title and rarity. All paintings are on display in a gallery. We also want to keep the names of all the painting's artists.
- Each supporting document exist for exactly one painting and contains information such as rating, condition, and the year of the report. Each document has a document number that is only unique between documents belonging to the same painting. Paintings can have multiple additional supporting documentation.
- A gallery is a room within a museum, it is uniquely identified by a gallery ID. For each gallery, we record the floor which it is at, its capacity, and its area code. Paintings are displayed in galleries: a gallery can have zero or more paintings on display. A gallery must be looked after by at least one museum staff member. A gallery can have multiple equipment.
- Each external restoration artist is uniquely identified by his/her restoration artist ID. For each restoration artist, we need to record his/her name, phone number, and email. (An external restoration artist is not considered a museum staff)
- Paintings are restored through tickets. Each ticket must be assigned to at least one restoration artist, each ticket is uniquely identified by its ticket ID, and issue date. A restoration artist can have zero or more tickets. For each ticket, the ticket must contain one painting, and we record the budget, and we also want the number of days elapsed since the issue date.
- We also want to store information about the internal museum staff, we want to keep their working hours. Each museum staff is uniquely identified by their ID, and we also record their name. Each museum staff must look after one or more galleries.
- We also want to keep information for museum equipment's equipment ID, name, dimensions (height, width, length), and status (i.e., whether in use or not). Each equipment must be stored in a gallery and must be managed by one or more museum staff.

(b) (8 marks) Translate the following ER diagram into a relational model. You must use the drawing conventions in the lecture notes.

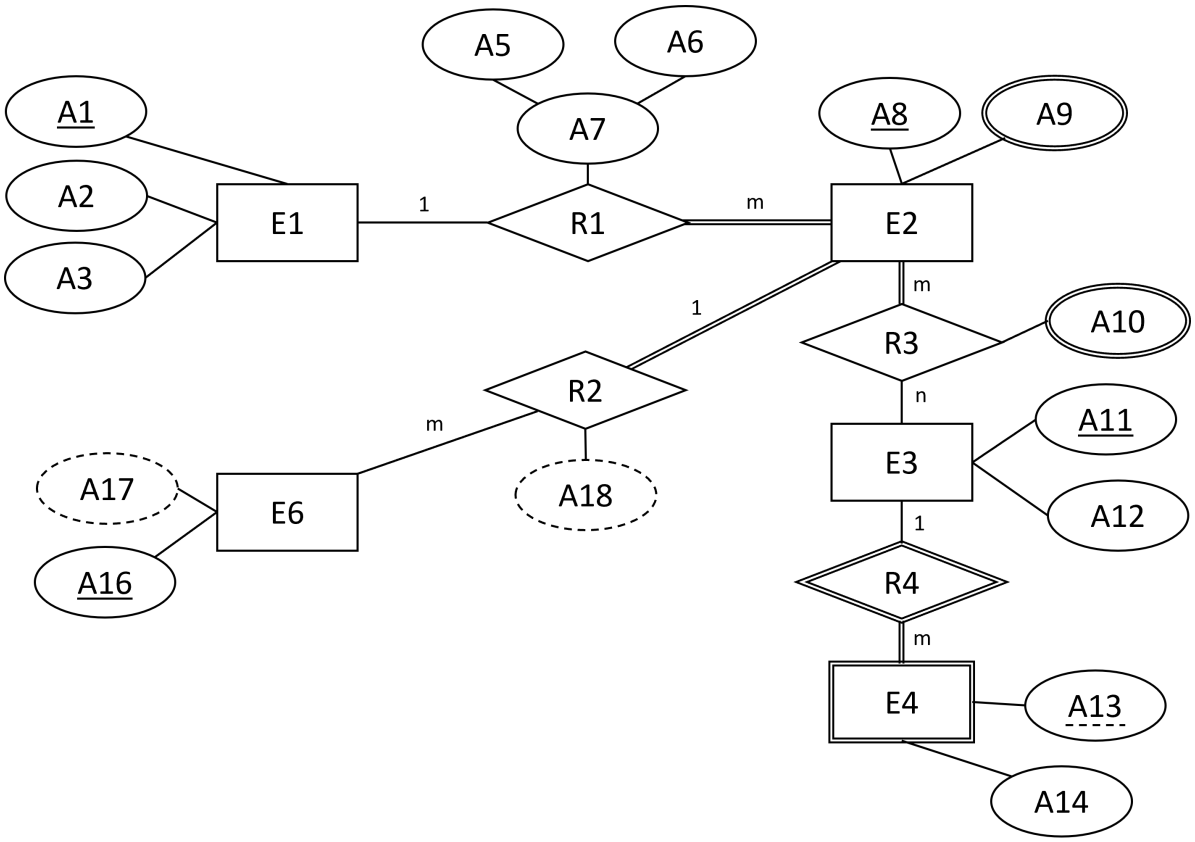

Figure 1: ER Diagram of Q2(b).

Question 3 (20 marks)

(a) (4 marks) Let a, b be a set of attributes, $\sigma_a(\prod_b(R)) = \prod_b(\sigma_a(R))$. Give an example where this is true, and an example where this is false.

(b) (16 marks) Consider the following relational database schema for a cinema service. The database schema consists of 3 relation schemas, the names and their attributes are shown below. The underlined attribute names in relation show that the combination of their values for that relationship is unique.

- customer (<u>cid</u>, name, age),
- movie (<u>mid</u>, name),
- watched (<u>cid, mid</u>, year)

Answer the following five queries by

1. express the queries using SQL (you can define auxiliary views to help breakdown the queries), and
2. express the queries using relational algebra.

(If not possible, provide a brief explanation)

i. Show the distinct names of customers who have watched the movie titled "Lorem Ipsum". *(4 marks)*

ii. Show the distinct IDs of movies with the greatest number of views out of movies that are only watched by a demographic aged 30 or above. *(4 marks)*

iii. Show the distinct IDs of customers who have never watched any movie or have watched all the movies. *(4 marks)*

iv. Show the distinct IDs of customers who have watched movies with the same name at least two times. *(4 marks)*

Question 4 (20 marks)

(a) (2 marks) Consider $F = A \rightarrow G, AG \rightarrow DE, E \rightarrow H, EC \rightarrow B, A \rightarrow D$, determine if $AC \rightarrow H$ can be inferred from F (justify your answer)

(b) (6 marks) Consider the relational schema $R(A, B, C, D, E, G, H, I)$ and a set of functional dependencies $F = H \rightarrow D, B \rightarrow AI, I \rightarrow BC, DH \rightarrow I, ABG \rightarrow HI$. Note that A, B, C, D, E, G, H, and I are attributes.

i. Regarding F, is the decomposition $R1 = ABCG$, $R2 = DEGHI$ of R lossless join? Please justify your answer. *(3 marks)*

ii. Regarding F, is the decomposition $R1 = ADEG$, $R2 = BCE$, $R3 = EH$ and $R4 = AC$ of R lossless join? Please justify your answer. *(3 marks)*

(c) (12 marks) Consider the relational schema $R(A, B, C, D, E, G, H, I)$ and the set of functional dependencies $F = AD \rightarrow BC, CD \rightarrow EGI, ACG \rightarrow H$. Note that A, B, C, D, E, G, H, and I are attributes. Justify your answer to each question.

i. Find the one and only candidate key. *(2 mark)*

ii. Find a minimal cover F_m for F. *(3 marks)*

iii. From your answer in ii., decompose R into 3NF. *(3 marks)*

iv. From your answer in iii., decompose any relations that are not in BCNF into BCNF that preserves functional dependencies of the minimal cover. *(4 marks)*

Question 5 (18 marks)

(a) (4 marks) Consider the schedule below. Here, $R(\cdot)$ and $W(\cdot)$ stands for 'Read' and 'Write', respectively. T_i represents a transaction, and t_i represents the time slot.

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
T_1	*R(Y)*		*W(X)*						
T_2		*R(X)*							
T_3						*W(Y)*		*R(Z)*	
T_4					*W(X)*		*R(Y)*		*R(X)*
T_5				*W(Y)*					

Figure 2: Schedule of Q5(a).

Give the complete precedence graph of this schedule and label the data item(s) for each edge, determine if this is a conflict serializable graph?

(b) (4 marks) Consider the following schedule of two transactions:

T_1	T_2
	R(B)
	B ← 100 – B
	W(B)
	R(A)
R(A)	
R(B)	
A ← 10A + B	
B ← 2B + A	
	A ← 30 + 7B
	W(A)
W(A)	
W(B)	

Figure 3: Schedule of Q5(b).

i. Result-serializability is when the result of a schedule is the same as some serial schedule. Given the initial values $A = 10$ and $B = 20$, show if this schedule is result-serializable or not. *(2 marks)*

ii. Show how to add locks/unlocks to make the two transactions T_1 and T_2 conflict serializable. *(2 marks)*

(c) (4 marks) Consider the lock request sequence given below. $RL(\cdot)$, $WL(\cdot)$ and $UL(\cdot)$ stand for "read lock", "write lock" and "unlock", respectively. T_1, T_2, T_3, T_4, T_5 and T_6 represent six transactions.

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}
T_1	WL(P)									UL(P)		WL(Q)	
T_2				RL(P)			RL(Q)						
T_3					WL(X)								WL(S)
T_4						WL(S)			RL(X)				
T_5			RL(S)								WL(R)		
T_6		WL(X)						WL(R)					

Figure 4: Schedule of Q5(c).

Draw the wait-for graph for the above lock requests and determine whether a deadlock exists.

(d) (6 marks) Consider the buffer replacement policies: 'Most Recently Used' and 'First in First Out':

i. Construct an example (sequence of read or write requests) that, for the same query set, the 'First in First Out' buffer replacement policy performs the worst with one buffer size and performs the best with another buffer size. Justify your answer and explain why if it does not exist. *(3 marks)*

ii. Construct an example (sequence of read or write requests) that, for the same query set, the 'Most Recent Used' buffer replacement policy performs the worst with one buffer size and performs the best with another buffer size. Justify your answer and explain why if it does not exist. *(3 marks)*

Question 6 (6 marks)

(a) (3 marks) Consider the following Graph G in a Graph Database:

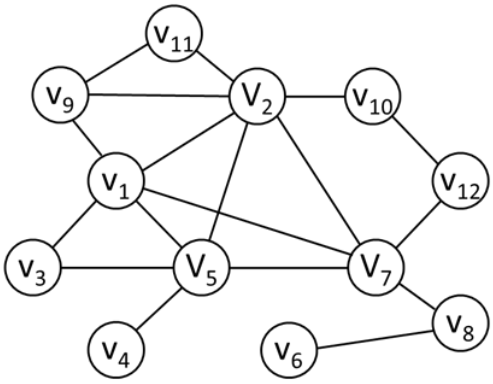

Figure 5: Graph G of Q6(a).

Find the 3-core of the above graph, provide your step-by-step process of the order in which vertices and edges are deleted in each step.

(b) (3 marks) In graph theory, a temporal graph $G(V, E)$ is a type of graph where edges record a timestamp, that timestamp specifies the time which the edge exists. Each edge $e \in E$ contains three pieces of information (u, v, t) that generally represent an event/occurrence between u and v at time t. G over a specific time-window $[t_1, t_2]$, is denoted by $G_{[t_1,t_2]}$. Formally, $G_{[t_1,t_2]}$ contains the vertex set $V_{[t_1,t_2]} = V$ and the edge set $E_{[t_1,t_2]} = \{(u, v)|(u, v, t) \in E, t \in [t_1, t_2]\}$. (To explain: it includes all vertices in G, but only the edges with timestamp between t_1 and t_2 (inclusive)).

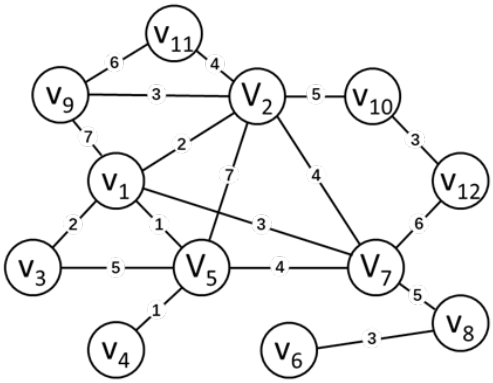

Figure 6: Temporal graph G of Q6(b).

Given the undirected temporal graph G above, give $G_{[3,6]}$ and the 2-core of $G_{[3,6]}$. For the 2-core of $G_{[3,6]}$, provide your step-by-step process of the order in which vertices and edges are deleted in each step.

END OF EXAM PAPER