Final Exam

Question 1 (18 marks)

(a) (10 marks) Consider the following library database:

- Each book is uniquely identified by its book ID. For each book, we also record these attributes: title, authors, language, price and status. Each book can have multiple authors.
- Each book has exactly one publisher that is uniquely identified by its ID. We also record the name of a publisher. In this database, each publisher has published at least one or more books.
- Each publisher has exactly one contact person, who is uniquely identified by his/her name and the publisher's Id. Each contact person is only associated with exactly one publisher. In addition to the name, we also record the office phone number of each contact person.
- A bookshelf is uniquely identified by its bookshelf ID. For each bookshelf, we also record its location. There are zero or more books put on each bookshelf. In this library the quantity of each book is different. The same books can be put on multiple bookshelves. We also record the quantity of each book on each bookshelf, and the total number of books on each bookshelf is needed.
- A user is uniquely identified by the user ID. For each user, we also record his/her name, phone number, email and address.
- To borrow books, an order will be created by exactly one user. Each order is uniquely identified by its order ID. A user can have zero or more orders. For each order, it must contain one or more books, and we record the due date, comment and the order time. Multiple orders can contain the same book, and not all books will be borrowed. There will be no more than one copy of each book in an order. We also want to record the total number of books in each order.

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 2 (18 marks)

(18 marks) Consider the relational schema R(A, B, C, D, E, G, H, I) and a set of functional dependencies $F = \{B \rightarrow CH, BCD \rightarrow HI, EI \rightarrow H, H \rightarrow AB, I \rightarrow E\}$. Note that A, B, C, D, E, G, H, I 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. (3 marks)
3. Regarding F, is the decomposition $R_1 = \{ABCD\}$, $R_2 = \{DEGHI\}$ of R lossless-join? Please justify your answer. (4 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. (5 marks)

Question 3 (20 marks)

(16 marks) Consider the following relational schema for park-visit records:

- Visitor (vID, name, age, gender),
- Park (pID, location),
- Visit (pID, vID)

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 location. pID is the primary key.
- Visit: For each visit, we record the pID of park and the vID of visitor. The combination of pID and vID is the primary key.

Use relational algebra to answer the following questions:

1. List the pID of parks that are only visited by senior visitors (age>=65) or only visited by junior visitors (age<=24). (4 marks)
2. List the vID of visitors who have visited all the parks. (4 marks)
3. List the pID and location of all parks that are visited by either 'Daniel' or 'James', but not both of them. (4 marks)
4. Find the vID of the oldest visitors that have visited the 'Hyde park'. (Hint: You may use $A.b$ to refer the attribute b in the relation A and may need immediate relations to solve this question. E.g. R1 $\leftarrow \pi_{(pID)}(Park)$.) (8 marks)

Question 4 (12 marks)

(a) (8 marks) Consider the schedule below. Here, R(·) and W(·) 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		W(Y)			R(X)				
T_2						R(Y)	W(Z)		
T_3				R(Z)				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 providing the equivalent serial schedule. If your answer is "no", convert it to a conflict seriazable schedule with the minimum change in the scheduling while preserving the given ordering of operations in each transaction. (3 marks)

(b) (8 marks) Consider the lock request sequence given below. RL(·) and WL(·) 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			WL(A)		WL(C)			
T_3	WL(C)							RL(B)
T_4		RL(B)					WL(C)	

1. Determine whether there exists a deadlock in the lock requests and justify your answer. (3 marks)
2. For the four transactions, construct a serializable but non-serial schedule using two-phase locking mechanism. If this is impossible, please justify your answer. (3 marks)

Question 5 (14 marks)

(a) (6 marks) Consider a relation R(a,b,c,d,e,f,g,h) containing 10,000,000 records, where each data page of the relation holds 10 records. R is organized as a sorted file with the search key R.b. Assume that R.b is a candidate key of R, with values lying in the range 0 to 99,999,999. For the relational algebra, state which of the following approaches (or combination thereof) is the most likely to be the cheapest and justify your answer:

1. Access the sorted file for R directly.
2. Use a clustered B+ tree index on attribute R.a.
3. Use a clustered B+ tree index on attribute R.b.
4. Use a linear hashed index on attribute R.a.
5. Use a clustered B+ tree index on attributes (R.a,R.b).
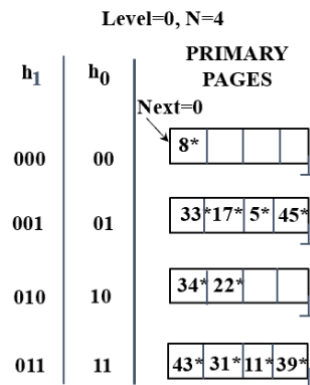6. Use a linear hashed index on attribute s (R.a,R.b).

Index is a collection of data entries which can be used to quickly access the records. We assume that the database considers index-only plans. Index-only plans means that an index contains all the columns needed to answer the query without having to access the data records in the file that contain the relations in the query.

(b) (8 marks) Consider three buffer replacement policies: 'Least Recently Used', 'Most Recently Used' and 'First In First Out'. Please answer the following questions:

1. Construct an example that 'First In First Out' buffer replacement policy performs better than the other two buffer replacement policies. (4 marks)
2. Construct an example that 'Least Recently Used' buffer replacement policy performs the best, 'First In First Out' performs moderately, and 'Most Recently Used' performs the worst. (4 marks)

Question 6 (18 marks)

(a) (6 marks) Consider the Linear Hashing index shown below. Assume that a bucket split occurs whenever an overflow page is created. $h_0(x)$ takes the rightmost **2 bits** of key x as the hash value, and $h_1(x)$ takes the rightmost **3 bits** of key x as the hash value.

Level=0, N=4

h_1	h_0	PRIMARY PAGES
000	00	Next=0 → 8*
001	01	33* 17* 5* 45*
010	10	34* 22*
011	11	43* 31* 11* 39*

1. What is the largest key less than 50 whose insertion will cause a split? Please justify your answer (2 marks)
2. Show the index after inserting entries with hash values 9, 27 and 78 (plot the final hash table and update the 'Next' pointer, if needed). (4 marks)

(b) (12 marks) Consider the following B$^+$ tree:
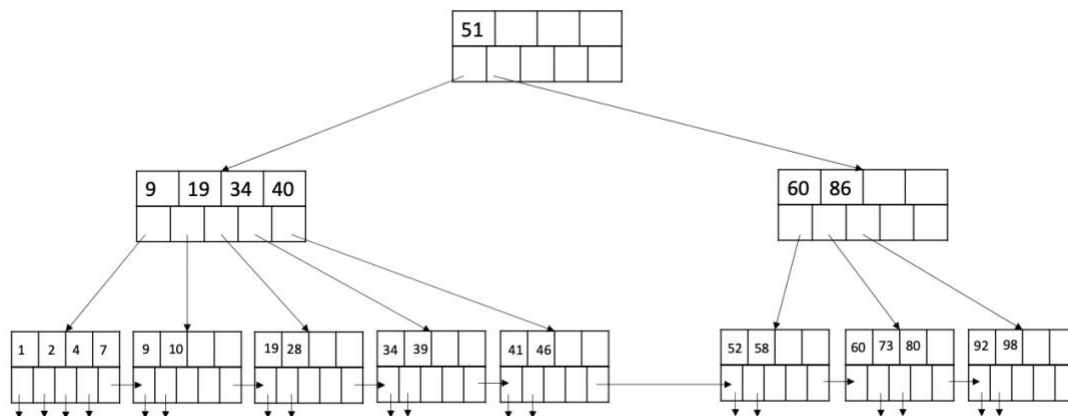
1. How many additional records are required to be added to increase its height? Give the minimum possible number and justify your answer. (4 marks)
2. Draw the B$^+$ tree after adding the data entries with keys 8, 82, 85 sequentially into the original tree. (4 marks)
3. Draw the B$^+$ tree after deleting the data entries with keys 52, 60, 92 sequentially from the original tree. (4 marks)

Final Exam Submission

We accept electronic submissions only. Please submit your assignments as follows:

- The file name should be final.pdf.
- Log into the CSE server, ensure that you are in the directory containing the file to be submitted. (note: we only accept files with .pdf extension)
- Type "give cs9311 final final.pdf" to submit.
- You can also use the web give system to submit.
- In case that the system is not working properly, you **must** take the following actions:
  1) Please keep a screen capture (including your **zid**, the submission **timestamp** and the **size** of the submitted file) for your submissions as proof. If you are not sure how, please have a look at the guidelines.
  2) Please keep a copy of your submitted file on the CSE server. If you are not sure how, please have a look at taggi.

Note:

1. If the size of your pdf file is larger than 4MB, the system will not accept the submission. If you face this problem, try converting to compress pdf.
2. If you have any problems in submissions, please email to comp9311unsw@gmail.com.
3. We do not accept e-mail submissions.

Late Submission Penalty

0 mark.