

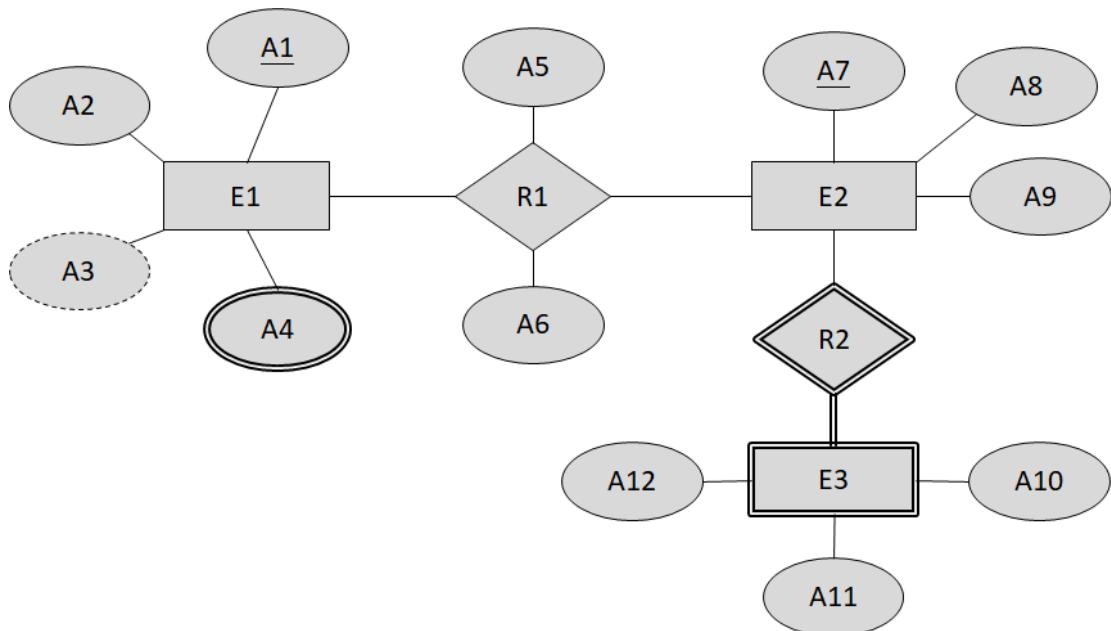
Introduction to Database Systems

2021 Final Exam

ER Model and Relational Model

1. (16%) Please provide appropriate names/terms to the ER model according to the following descriptions. (1 point for each blank)

Linda has a lighting store which sells lighting such as chandeliers, pendants, lamps, LED bulbs and more. She wants to develop a database system to manage orders from customers and keep track of products in stock. The data held on a customer is the CID (customer ID), date of birth, age derived by date of birth and multi-valued phone numbers. The CID uniquely identifies each customer. The data held on a product is the PID (product ID), name and stock level. The PID uniquely identifies each product. When a customer orders a product online, the system records the timestamp and the payment type of this ordering. Linda also wants to monitor the quantity of each product in stock. If LED bulbs are out of stock, Linda needs to purchase from the supplier. And if a chandelier is packaging and shipping to a customer, the number in stock decreases. So the purchase/ship history of a specific product should be maintained which includes data, type(purchase or shipment) and quantity.



- a. E1: _____
- b. A1: _____
- c. A2: _____
- d. A3: _____
- e. A4: _____
- f. R1: _____
- g. A5: _____
- h. A6: _____
- i. E2: _____

- j. A7: _____
- k. E3: _____
- l. A10: _____
- m. A11: _____
- n. A12: _____
- o. Participation of E3 in R2 is _____.
- p. E3 is a _____ entity set.

SQL

2. (12%) The following content is the schema of a simple social network.
- Student (ID, name, grade)
Here are students with a unique ID for each student. It is guaranteed that every student in the database will have their name and grade (you don't need to consider NULL value).
 - Friend (ID1, ID2)
The student with ID_1 is friends with the student with ID_2 . Friendship is mutual, so if (123, 456) is in the Friend table, so is (456, 123).
- Please write the SQL (MySQL) to cater the following three requirements.
- a. (4%) Find the name of all students who are friends with someone named "Gabriel".
 - b. (4%) What is the average number of friends per student?
 - c. (4%) Find the names and grades of students who only have friends in the same grade.
Return the result sorted by grade, then by name within each grade.

Schema Refinement and Normal Forms

3. (15%) Consider the relation R, which has attributes that hold schedules of courses and sections at a university; $R = \{\text{Course_no}, \text{Sec_no}, \text{Offering_dept}, \text{Credit_hours}, \text{Course_level}, \text{Instructor_ssn}, \text{Semester}, \text{Year}, \text{Days_hours}, \text{Room_no}, \text{No_of_students}\}$. Suppose that the following functional dependencies hold on R:
- $$\{\text{Course_no}\} \rightarrow \{\text{Offering_dept}, \text{Credit_hours}, \text{Course_level}\}$$
- $$\{\text{Course_no}, \text{Sec_no}, \text{Semester}, \text{Year}\} \rightarrow \{\text{Days_hours}, \text{Room_no}, \text{No_of_students}, \text{Instructor_ssn}\}$$
- $$\{\text{Room_no}, \text{Days_hours}, \text{Semester}, \text{Year}\} \rightarrow \{\text{Instructor_ssn}, \text{Course_no}, \text{Sec_no}\}$$
- a. (7%) Try to determine which sets of attributes form keys of R.
 - b. (8%) How would you normalize this relation?

Storage and Indexes

4. (16%) Consider the following relation:
Emp(eid: integer, *ename*: varchar, *sal*: integer, *age*: integer, *did*: integer), where the primary key is underlined. Salaries range from \$50,000 to \$500,000, ages vary from 20 to 60. Assume uniform distributions of values.
- For each of the following queries, which of the listed index choices would you choose to speed up the query? Explain briefly.
- a. (8%) Query: Print the names of employees whose ages are between 35 and 40.
 1. Clustered B+-tree index on the search key <*age*>
 2. Unclustered B+-tree index on the search key <*age*>
 3. Clustered hash index on the search key <*age*>
 4. No index.

b. (8%) Query: Print the names of employees whose salaries are between \$100,000 and \$500,000.

1. Unclustered B+-tree index on the search key <sal>
2. Clustered hash index on the search key <sal>
3. Unclustered hash index on the search key <sal>
4. No index.

Indexing Structures

5. (9%) Construct B+-tree with its order 2 for the following set of key values: (2, 3, 5, 7, 11, 17, 19, 23, 29). Assume that the tree is initially empty and values are added in ascending order. Please write down the **constructing progress** instead of answering only the result.

Transactions

6. (10%) Is the following schedule conflict serializable? Use precedence graph to support your answer and explain. Write down the equivalent serial schedule if it is also serializable.
 $S = [T1:R(A), T2:R(A), T2:W(A), T3:R(B), T1:R(B), T3:R(A), T3:W(B), T4:W(B), T4:R(A)]$, where $Ti:R(j)$ and $Ti:W(j)$ indicates a *read* and *write* operation by transaction i on data object j .

Concurrency Control

7. (9%) Given the following schedule without concurrency control mechanism:
 $S = [T1:R(A), T2:R(B), T3:R(C), T2:R(C), T3:R(B), T1:W(A), T2:W(C), T3:W(B)]$. Describe what would happen under a) the two-phase locking protocol (3%), b) the wait-die protocol (3%), and c) the timestamp-based protocol (3%).

Query Processing

8. (13%) Let $r1(A, B, C)$ and $r2(C, D, E)$ have the following properties: $r1$ has 20,000 tuples, $r2$ has 45,000 tuples, 25 tuples of $r1$ fit on one block, and 30 tuples of $r2$ fit on one block. Assume that we implement Nest-loop join and the outer (respectively, inner) relation is $r1$ (respectively, $r2$). Estimate the **number of block transfers** and **seeks** required under the following scenarios, also you need to **explain how you get these number**:

- a. (4%) There are only 3 memory blocks available and both $r1$ and $r2$ are sorted.
- b. (4%) If the minimal block transferred is required, what is the outer relation that we should pick from $r1$ and $r2$? (Under the assumption that there are three memory blocks)
- c. (5%) Assume that we have 900 memory blocks. Please describe how to implement nest-loop join efficiently (i.e. the minimal block transferred)

Discussion

TAs had opened a channel **Exam 討論區** on New E3 forum of the course, you can post questions about the homework on the forum. TAs will answer questions as soon as possible.

Discussion rules:

1. Do not ask for the answer to the homework.
2. Check if someone has asked the same question before asking.

Since we have this discussion forum, do not send email to ask questions about the homework unless the questions are personal and you do not want to ask publicly.

Submission

1. The deadline of this homework is **5/23 (Sun.) 23:55:00, no late submission accepted this time.**
2. You only need to submit one `pdf` file named as “**Exam_XXXXXXX.pdf**” where XXXXXX is your student ID. Wrong file or naming format causes **-10** points to your score.
3. If there is anything you are not sure about submission, ask in the discussion forum.