单选没3？

Question 2. ERD, Table Design, and Normalization **[20 marks]**

[1] Suppose you are given a relation R with four attributes A, B, C, D. For each of the following sets of FDs, assuming those are the only dependencies that hold for R: (1) Identify the candidate key(s) for R. (2) Identify the best normal form that R satisfies (1NF, 2NF, 3NF, or BCNF). (3) If R is not in 3NF, decompose it into a set of 3NF relations with lossless-join decomposition. **[6 marks]**

(a) $C \rightarrow D, B \rightarrow C$
(b) $A \rightarrow B, BC \rightarrow D, A \rightarrow C$

(a). (1) {AB}
(2). 1NF.
(3). R(B, C, D.)
→ 2NF: B→C, C→D.
R(A, B).
3NF: R(A, B)
R(B, C)
B→C
R(C, D)
C→D.

(b). (1). {A}
(2). 2NF
(3) 3NF:
R(A, B, C)
A→B, A→C
R(B, C, D)
BC→D.

[2] Design a database for keeping the records of book purchases in a bookstore. The requirements are below. **[12 marks]**

- Every book is identified by an ISBN. Also, a book has a title, year, and publisher. Every book is written by one or more employees of the store.
- Each employee is identified by a staff number, and other information includes name, date of birth, and a pseudonym (under which they may write additional books). For working in the store, an employee must have authored at least one book.
- The customer information includes first name, phone number, gender, and birthday. A customer can be identified by name and phone number together.
- Customers buy books from employees. Each sales transaction has a date and time, books and their prices, staff number, and the total sum which is written on the sales receipt.

Design an ER model based on the above description and transform the model into a set of tables. For the ERD, indicate the primary keys, cardinalities, types of entities, and attributes. For the tables, indicate the foreign keys also.

ER:

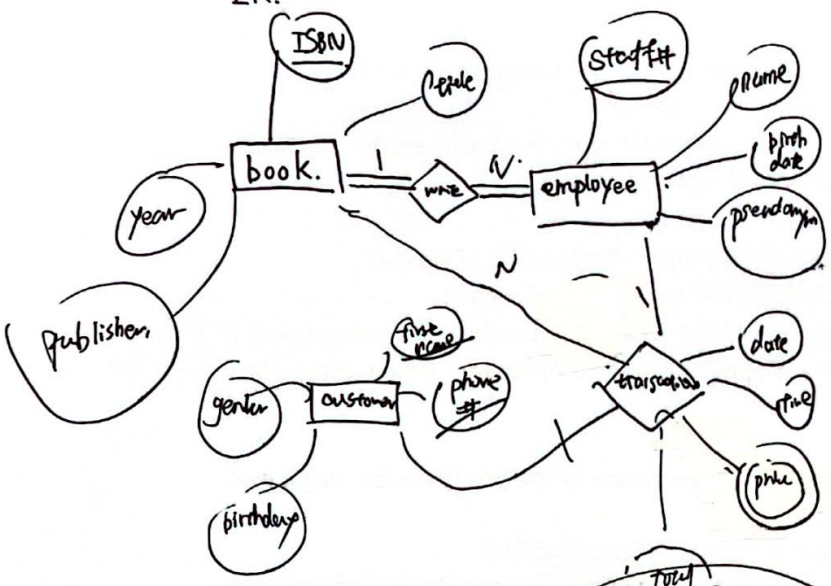


Table:

Book (ISBN, title, yr, publisher).

Employee (staff#, name, birth date, pse-; ISBN)

Customer (1st name, phone #, gender, birthday)

Transaction. (1st pushone, cusphow, employee, book)

[3] Modify the answer to part [2] to capture the following constraint: A book can only be sold by the employee who authored it. If it cannot be captured, give a brief explanation. **[2 marks]**

can't, we can only use sql, programm or other way.

Simple ER and Relation model can't do this.

Question 3. SQL and Relational Algebra **[10 marks]**

A department has a database application that keeps track of the software programs used. The database has tables below.

- Program (ProgramID, name)
  - ProgramID: software program ID; name: name of the program
- Computer (ComputerID, name, roomNumber, ownerID)
  - ComputerID: computer ID; roomNumber: the room number where the computer is located; ownerID: the owner/staff of the computer
- Staff (ID, name, officeNumber)
  - ID: staff ID; name: staff name; officeNumber: room number of the staff
- ProgramInstallation (ProgramID, ComputerID, ownerID, roomNumber)
  - ProgramID: software program ID; ComputerID: ID of computer in which program is installed; ownerID: ID of the computer's owner; roomNumber: room number where the computer is located
- LicenseType (typeCode, name)
  - typeCode: type of program licenses; name: name of the type (freeware, single user, group license, self-developed, etc.)
- LicenseStatus (ProgramID, ComputerID, typeCode, licenseVerified)
  - ProgramID: ID of software program, ComputerID: ID of computer in which program is installed; typeCode: type of license on the program; licenseVerified: a boolean variable (T/F) indicating whether the required license for the program has been verified

[1] Write a SQL statement to list the IDs of the computers that have more than 10 software with licenses not verified. **[3 marks]**

太难了

```
SELECT ComputerID.
FROM LicenseStatus
WHERE LicenseVerified = F
GROUP By ComputerID
HAVING: COUNT(*) ≥ 10
```

真的可以吗?

太难.

[2] Write the relational algebra expression(s) to list the ID of the staff who own computers in more than one room. **[3 marks]**

感觉做不了 (我猜现场改网了 ≥1 分)

[3] Write a SQL statement to list the names of the computers that have all programs installed. **[4 marks]**

```
SELECT C.name
FROM. computer C.
WHERE NOT EXISTS (
    SELECT *
    FROM ProgramInstallation P
    WHERE P.ComputerID = C.ComputerID ),
```

Question 4. File Organization and Indexing **[15 marks]**

[1] An EMP relation contains the following fields for each employee record:

```
EID: integer, EName: string, Age: integer, Salary: integer
```

Assume that the maximum length of the EName field is 25 characters. An integer occupies 4 bytes; a character occupies 1 byte. Consider the INSERT statement below:

```
INSERT INTO EMP VALUES (123456, 'Kathleen', 32, 89000);
```

How many bytes are required to store this record using the following record organization respectively? Show your calculation process and result. **[2 marks]**

(a) Fixed-length record organization $4 \times 3 + 25 \times 1 = 37$.

(b) Variable-length record organization using a special character as the separator

(b) $4 \times 3 + (8+1) \times 1 = 21$.

[2] Suppose the EMP relation contains 8000 fixed-length records. The EMP file is ordered by the primary key EID, and stored on a hard disk with the following configuration:

- Block size = 500 bytes
- Block pointer size = 6 bytes

A primary index is constructed on the EID field. How many block accesses are needed to retrieve the EMP record with EID = 123456 using the primary index? Show your calculation process and result. **[5 marks]**

entry size = 6 + 4 = 10

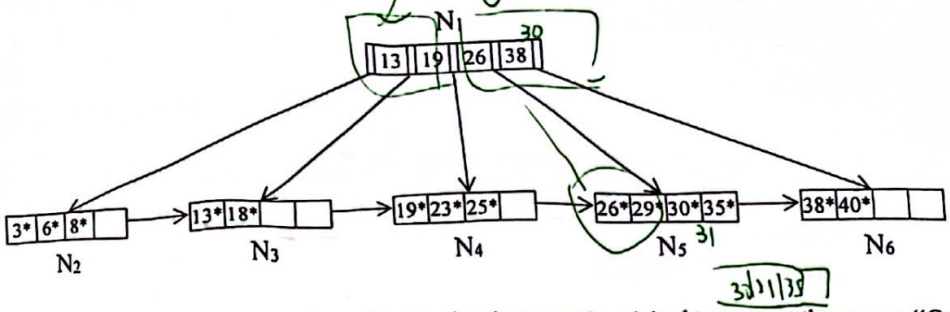
$bfr_{index} = \lfloor \frac{500}{10} \rfloor = 50$. $num_{block} = \lceil \frac{8000}{13} \rceil = 616$

$bfr = \lfloor \frac{500}{37} \rfloor = 13$.

number of index block $= \lceil \frac{616}{50} \rceil = 13$.

$ans = \lceil \log_2 13 \rceil + 1 = 5$.

[3] Consider a B+ tree as shown in the figure below, where the tree nodes are labeled as N_1, N_2, ..., N_6. Assume the following rule applies for redistributing keys after a leaf node split: Two keys stay in the old leaf node, and the remaining keys move to a new leaf node.



(a) What is the minimum number of tree nodes that must be visited to answer the query: "Get all records with keys greater than 20"? List all the visited tree nodes. **[3 marks]**

N_1 N_4 N_5 N_6.

(b) Show the updated B+ tree after inserting an entry with key "31". **[5 marks]**

Question 5. Query Optimization **[20 marks]**

[1] Cost-based "selection" operator evaluation estimates the cost (i.e., number of block accesses) of each algorithm (i.e., linear scan, binary search, and index search) and chooses the one with the minimum cost. Consider the following Student relation:

```
Student (SID, SName, Email, Age, Gender, Department)
```

SID is the primary key of the Student relation. The relation contains 20000 records, with each record occupying 100 bytes. The file is unsorted and stored on a hard disk with the following configuration:

- Block size = 1000 bytes
- Block pointer size = 6 bytes

A B+ tree index has been created on the SID field (Integer, 4 bytes), and each tree node is 60% full on average. Given the following SQL query:

```
SELECT * FROM Student WHERE SID > 18000;
```

Assume that 2000 records are generated as the result of this SQL query. Which algorithm will be used to answer this query based on cost-based "selection" operator evaluation? Justify your answer by comparing the algorithm cost. Show your calculation process and result. **[7 marks]**

Linear: $\lfloor \frac{1000}{100} \rfloor = 10$ bfr

$\lceil \frac{2 \cdot 10^4}{10} \rceil = 2000$ ☆

binary X.

B+: ~~key~~ p: fo bfr index $\lfloor \frac{1000}{4+6} \rfloor = 100$

$\lceil \log_{fo} 20000 \rceil = \lceil \log_{100} 20000 \rceil = 3.$

cost = $3 + \lceil \frac{2000}{100} \rceil$

$= 3 + 20 = 23.$ ☆

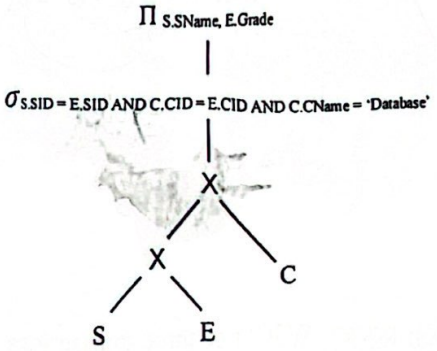
[2] Consider the following relations in a University database:

```
Student (SID, SName, Email, Age, Gender, Department)
Course (CID, CName, Lecturer)
Enrollment (SID, CID, Semester, Grade)
```

In the Enrollment relation, SID is a foreign key referencing the Student relation and CID is a foreign key referencing the Course relation. CName is a unique field in the Course relation. Given the following SQL query:

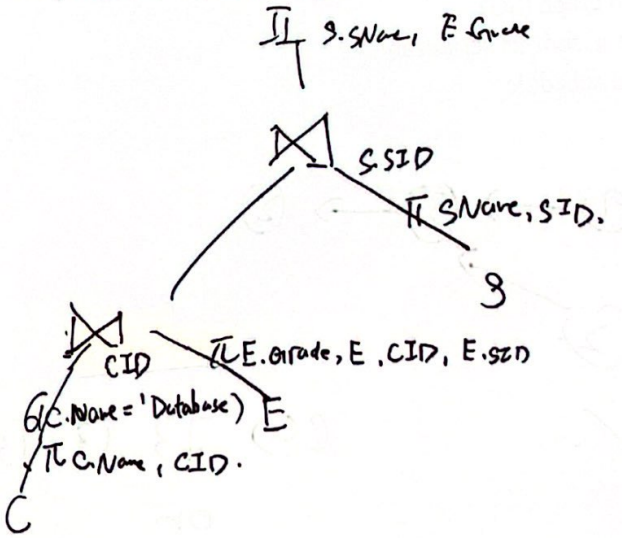
```
SELECT S.SName, E.Grade
FROM Student S, Course C, Enrollment E
WHERE S.SID = E.SID AND C.CID = E.CID AND C.CName = 'Database';
```

The initial query tree is illustrated as follows:



共3.

Show the most efficient query tree after applying all the five steps of heuristic-based query tree optimization. **[13 marks]**

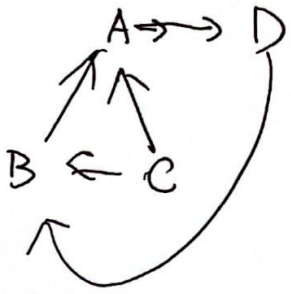


A → B. 時 A在等lock

Question 6. Concurrency Control and Database Recovery **[15 marks]**

[1] Construct a wait-for-graph (WFG) based on the following lock table. Is there any deadlock in the system? Justify your answer. **[7 marks]**

Data Item	Locks Granted	Queued Requests
A	<T1, read_lock>, <T5, read_lock>	<T4, write_lock>
B	<T2, write_lock>	<T1, read_lock>
C	<T3, write_lock>	<T1, write_lock>, <T2, read_lock>
D	<T4, read_lock>	<T2, write_lock>



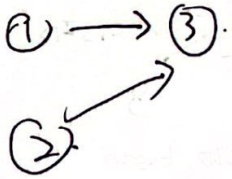
Yes. B→A→D.

cycle.

[2] Given the schedule S: R1(X); W1(X); R3(X); R2(X); W3(X) of three transactions T1, T2, and T3, answer each of the following questions. **[4 marks]**

(a) Construct the serialization graph (SG)
(b) Determine if the schedule is conflict serializable
(c) Show the equivalent serial schedule

(a)



(b) ✓.

(c) T1, T2, T3
or
T2, T1, T3

[3] Assume that the steal/no-force buffer management strategy is adopted. The recovery engine periodically adds a checkpoint in the database log indicating that all the modified buffers have been flushed to disk at that point. Given the following database log, what recovery operations are needed for transactions T1, T2, T3, and T4, respectively? Justify your answer. **[4 marks]**

[start_transaction, T1]
[write_item, T1, D, 10, 20]
[commit, T1]
[checkpoint]
[start_transaction, T2]
[write_item, T2, B, 2, 15]
[abort, T2]
[start_transaction, T3]
[write_item, T3, C, 1, 5]
[start_transaction, T4]
[write_item, T4, D, 20, 25]
[commit, T4]
[write_item, T3, A, 10, 5] ← SYSTEM CRASH

** END **