北京邮电大学 2010——2011 学年第 2 学期

《 数据库系统原理 》期末考试试题（B）

考试注意事项	一、学生参加考试须带学生证或学院证明，未带者不准进入考场。学生必须按照监考教师指定座位就坐。 二、书本、参考资料、书包等物品一律放到考场指定位置。 三、学生不得另行携带、使用稿纸，要遵守《北京邮电大学考场规则》，有考场违纪或作弊行为者，按相应规定严肃处理。 四、学生必须将答题内容做在试题答卷上，做在草稿纸上一律无效。 五、学生的姓名、班级、学号、班内序号等信息由教材中心统一印制。

考试课程	数据库系统原理			考试时间			2011 年 6 月 14 日				
题号	一	二	三	四	五	六	七	八	九	十	总分
满分	10	6	12	8	16	12	6	12	10	8	
得分	8	5	9	6	15	6	5	10	7	8	79
阅卷教师											

8

1. Fill in blanks. (1×10 points)

(1) The collection of information stored in the database at a particular moment is called an instance of the database.

(2) The six fundamental operations in the relational algebra are select, project , union , set difference, with , and rename.

(3) For the entity sets **A** and **B** and the relationship set **R** among them, if the cardinality limit of A is 2….*, and the cardinality limit of B is 0….1, then the mapping cardinality form *A* to *B*

is one-to-many.

(4) The decompositions of a relation schema must be lossless-join decompositions and should preferably be dependency preserving

(5) Given relations R(A,B,C) and S(A,B,C), also the SQL statement :

```
Select A,B,C
From   R
Where not exists
    (Select A, B, C
     From S
     Where R.A=S.A and R.B=S.B and R.C=S.C)
```

the result of this SQL query is equal to the following relational algebra expression:

$\Pi_{A,B,C}\sigma_{R.A\neq S.A \text{ or } R.B\neq S.B \text{ or } R.C\neq S.C}(R\times S)$

~~$\Pi_{A,B,C}$ [illegible] [$\Pi_{A,B,C}\sigma_{R.A=S.A, R.B=S.B, R.C=S.C}$ [illegible]]~~

(6) There are two basic kinds of indices, ordered indices and hash indices. Among ordered indices, dense-clustering index is an index whose search key also defines the sequential order of the data file.

(7) According to the equivalence rules, the expression

$$\sigma_{\theta1\wedge\theta2}(E_1 \infty_\theta E_2)$$

can be transformed into $\sigma_{\theta_1}E_1 \bowtie \sigma_{\theta_2}E_2$

, supposing that θ_1 involves only the attributes of E_1, and θ_2 involves only the attributes of E_2.

The basic steps involved in processing a query are parsing and translation, optimization, and evaluation.

(8) The transaction has the property called atomic property which means either all operations of the transaction are reflected properly in the database, or none are.

(9) Several concurrency-control schemes are used to ensure schedule serializability, among which locks, ~~[illegible]~~ ~~[illegible]~~ time-stamp and multiversion schemes are the most common ones.

5

2. (6 points) Give a brief answer to each of the following questions.

(1)What is Data Manipulation Language (DML)? (3 points)

DML is the basic Language for operating the data in database

It is a total name of many popular database operating lar

It is the ~~res~~ source of command for the database systems.

(2)What does data dictionary contain? (3 points)

It contains all the detail information except tuples of the entire datebase, including metadata of relations, attributes and logs.

9 **3.** （12 points）There are two data objects **sailors** and **boats** in real worlds. A sailor can reserve a boat with some color at some day, and he can also reserve several boats. The sailors are classified into different ratings (or groups) according to their skills, and each sailor has a different name. The characteristics of and the associations among these two objects are described by the following relations:

sailors(sid: integer, sname: string, rating: integer, age: real)

boats(bid: integer, bname: string, color: string)

reserves(sid: integer, bid: integer, day: date)

Use SQL statements to implement the following operations:

(1) For the table sailors, define the primary key and candidate key constraints (2 points); Add a new attribute birthplace with the string data type to it. (2 points)

```
① primary key (sid).
   candidate key (~~sid~~ sname).
                                   Begin
② alter table sailors add
      birthplace string[50];
   End;
```

-1

(2) It is required that, if a sailor's rating is below 10 or his age is lower than 25, he is not permitted to reserve a boat with white color. Define this requirement as a constraint in the database. (4 points)

```
      table
alter reserves Begin add
   check ( sid in select sid
                  from sailors
                  where rating >= 10 or age >= 25 )
   when bid in(select bid
                 from boats
                 where color = "white")
   End;
```

-1

(3) For each sailor's rating in which there are at least two sailors who have the rights to vote, i.e. whose age are not lower than 18, find out the average age of the sailors having the rights to cast votes. (4 points)

```
select ave (age)
 from sailors
 where ~~sid~~ age >= 18;
```

-1

4. (8 points) Suppose there are the following relations:

sailors(sid: integer, sname: string, rating: integer, age: real)

boats(bid: integer, bname: string, color: string)

reserves(sid: integer, bid: integer, day: date);

Use relational algebra to write the following queries.

(1) Find the Sailor_IDs of Sailors with age over 25 who have not reserved a red boat. (4 points)

$\pi_{sid}\,\sigma_{age>25 \text{ and boats.color is not "red"}}(\text{Sailors} \bowtie \text{boats} \bowtie \text{reserve}$

(2) Find the names of sailors who have reserved all boats. (4 points)

$\pi_{sname}\,\sigma_{count(reserves.bid)=count(boats.bid)}(\text{boats} \bowtie \text{reserve}$

-2

5. (16 points)A library borrow management database should provide the following services:

(1)Can look for the current books' kinds, quantities and storage locations at any time. All kinds of books can be identified uniquely by its book_id.

(2) At any time the information about the book's borrowing and returning can be searched for, including the borrower's units, name, borrowing card_id, and the borrowing date and

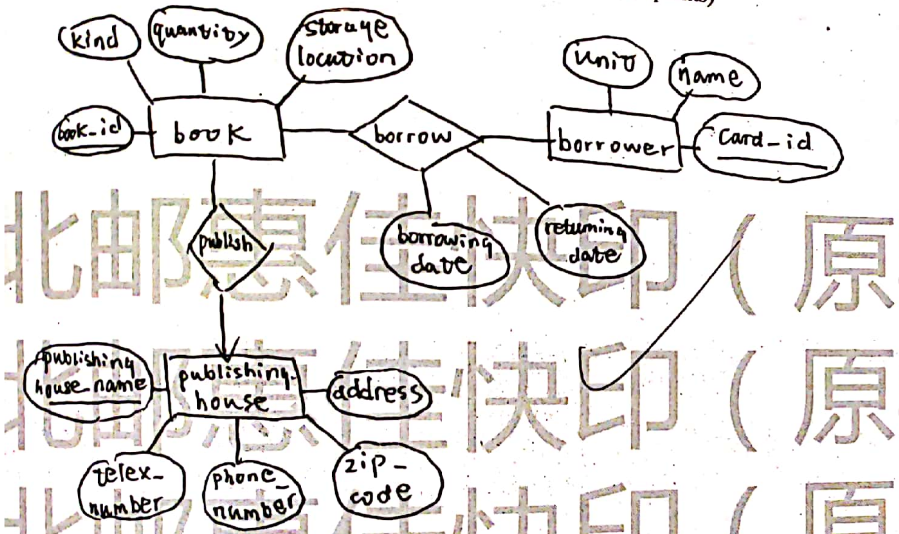
returning date.

We assume that one can borrow many kinds of books, and any book can be borrowed by more than one people, and the borrowing card_id is unique.

(3)The database saves the information about the publishing house in order to buy more books, such as telex_number, phone_number, Zip_code and address. We assume that a publishing house can publish many kinds of books, and one book can be published by only one publishing house and the publishing house_name is unique.

According to the above conditions and assumptions, answer the following questions:

(1) Design the E-R Diagram on basis of the information above. (8 points)



(2) Convert the E-R diagram to the proper relation schemas. (8 points)

book (book-id, kind, quantity, storagelocation).

borrower (card-id, unit, name).

publishinghouse (publishinghons_name, telex-number, phone-number zip-code, address)

borrow (book_id, card_id, borrowing_date, returning_date)

publish (book_id, publishinghouse_name)

6

6. (12 points)The following set F of functional dependencies is for relation schema R = (A, B, C, D,E,G): A→D, E→D,D→B,BC→D,DC→A

(1) List all the candidate keys of R. (2 points)

(A, C, E)

(D, C, E)

(2) Compute $(AE)^+$ (2 points)

AE result = AE

AE A→D result = ADE

ADE D→B result = ABDE ✓

ABDE is the final result.

(3) What is the highest normal form of R? Why? (4 points)

3NF is the highest normal form of R. -4

Fc = (A→D, E→D, D→B, BC→D, DC→A).

(4) Give a lossless-join, dependency-preserving decomposition into 3NF of R. (4 points)

Fc = (A→D, E→D, D→B, BC→D, DC→A).

so the result is: R1 (A, D) R4 (B, C, D) ✓

R2 (E, D) ✓ R5 (D, C, A) ✓

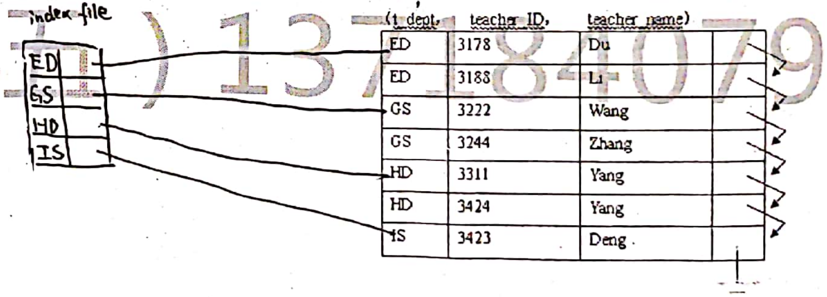
R3 (D, B) R6 (A, C, E) -1

7. (6 points) Give the data file teacher(t_dept, teacher_ID, teacher_name) as shown below, which is organized as a sequential file, taking the attribute *t_dept* as the search key,

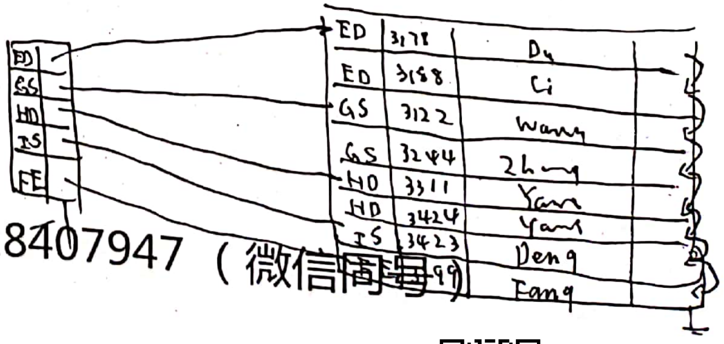
(1) define a *dense* and *clustering* index for the indexed file teacher.

It is required that the index file and index entries in the index file should be figured out. (3 points)

(2) If a tuple (FE, 3199, Fang) is inserted into the indexed file, depict the indexed file and the index file. (3 points)



t_dept	teacher_ID	teacher_name
ED	3178	Du
ED	3188	Li
GS	3222	Wang
GS	3244	Zhang
HD	3311	Yang
HD	3424	Yang
IS	3423	Deng



10

8. (12 points) Consider the following relations in banking enterprise database, where the primary keys are underlined.

branch (<u>branch-name</u>, branch-district, assets),

loan (<u>loan-number</u>, branch-name , amount)

borrower(<u>customer-name</u>, <u>loan-number</u>, borrow-date)

customer (<u>customer-name</u>, customer-street, customer-city)

account (<u>account-number</u>, branch-name, balance)

depositor (<u>customer-name</u>, <u>account-number</u> , deposit-date)

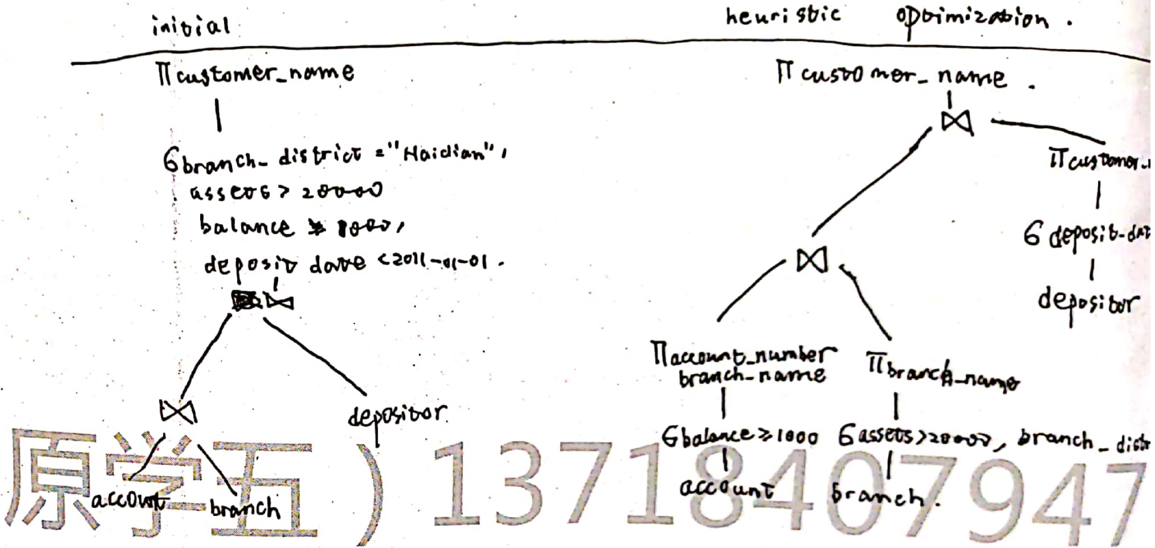
For the query " Find the names of all customers who have an account at any branch that is located in HaiDian district and has an asset more than $200,000, requiring that the account balance is not lower than $1000 and the deposit data is before 2011-01-01"

(1) Give an SQL statement for this query. (5 points)

(1)
```
select customer_name
  from ~~customer,~~ account , branch, depositor.
where depositer.account_number = account.account_number and
      account. branch_name = branch. branch_name and
      branch _ district = "HaiDian" and
      ~~balance~~ assets [illegible] and
         balance >= 1000 and
         deposit. date < 2011-01-01
```

7

(2) Given an initial query tree for the query in (1), then convert it into an optimized query tree by means of heuristic optimization. (7 points)



2

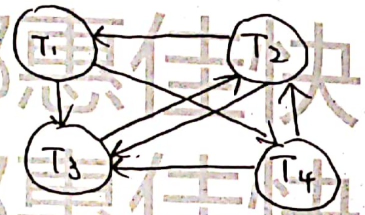
9. (10 points) Some questions about concurrent schedules.

(1) With respect to the concurrent schedule S1 on transaction set {T1, T2, T3, T4}, construct the precedence graph G(S1); Is S1 conflict serializable, and why? (4 points)

S1

T1	T2	T3	T4
read(X)			
	write(Y)		
	read(X)		
			read(Z)
		read(Y)	
write(Y)			
			write(X)
		write(X)	
		read(Z)	
	write(Z)		

(1)



-2

It is confilict serializable. Because there is ~~a~~ at least one circle in the precedence graph.

(2) Is the concurrent schedule S2 a recoverable schedule, and why? (1 points)

Is S2 a cascadeless schedule, and why? (1 points)

S2

T1	T2	T3
		write(X)
read(X)		
write(Y)		
commit		
	read(Y)	
	write(Y)	
	read(X)	
		commit
	commit	

(2) S2 is not a recoverable schedule.

-1 For T2 reads X [illegible] before T3. If crash happens before T3 commits,

T3 must commits before T2 read(X).

(3) Is the concurrent schedule S3 obeys the strict 2PL, and why? (4 points)

S3

T1	T2
Lock-S(A)	
	Lock-S(C)
Lock-X(B)	
read (A)	
	read(C)
read (B)	
Unlock (A)	
	Unlock (C)
	Lock-X(A)
	read (A)
write (B)	
	write (A)
Unlock(B)	
	Unlock (A)

(3) It does not obey 2PL. For T2 appilies for Lock-X after it has unlocked entirely, which doesn't obeys the 2PL rule.

8 **10.** (8 points) Considering the concurrent transactions T_0, T_1, T_2, T_3 and T_4, and the data items A, B, C, D modified by these transactions. It is assumed that the initial values of these data items are A=10, B=0, C=10, D=0. With respect to the log in the following figure that describes the concurrent executing of T_0, T_1, T_2, T_3 and T_4, when a failure occurs, the log-based recovery scheme consults the log to determine the recovery operations (i.e. redo, undo, ignore) done on T_0, T_1, T_2, T_3 and T_4.

After recovery operations on T_0, T_1, T_2, T_3 and T_4. are completed,

(1) What are the values of the data items A, B, C, D in the database? (4 points)

(2) Write down the redo, undo, ignore list respectively. (4 points)

<T_0 start>

<T_1 start>

<T_0, A, 10, 20>

<T_1, B, 0, 20>

<T_1 commit>

<checkpoint {T0}>

<T_2 start>

<T_2, B, 20, 100>

<T_0, C, 10, 50>

<T_3 start>

<T_3, D, 0, 60>

<checkpoint {T0,T2,T3}>

<T_4 start>

<T_0, C, 50, 70>

<T_0, A, 20, 80>

<T_2, B, 100, 200>

<T_2 commit>

<T_4, B, 200, 60>

<T_3, D , 60, 80>

Crash

(1)

A	B	C	D
10	200	0	0

(2) redo : T_2

undo : T_0 , T_3 , T_4

ignore : T_1