中 山 大 学 软 件 学 院 **2007** 级 软 件 工 程 专 业 (2009 秋季学期)

《 数据库系统原理》 期 末 试 题 (B 卷)

(考 试 形 式 ： 开 卷　　考 试 时 间 ： 2 小 时)

警 示

《中山大学授予学士学位工作细则》第六条

考 试 作 弊 不 授 予 学 士 学 位

方向：_____　　姓名：_____　　学号：_____

Question 1　Multiple Choices (10 points)

(a) Comparing DBMS and files for storing your data, which of the following are **true**?

A. DBMS is faster than files in data retrieval.

B. DBMS requires less application programming effort.

C. DBMS takes less disk space than files.

D. Data in files can only be accessed sequentially.

(b) Which of the following operation may change the relation schema?

A. Select.

B. Set Difference.

C. Union.

D. Project.

E. Join

(c) Which of the following statements is **false** in the relational data model?

A. Two relations can be joined only by their primary and foreign keys.

B. Tuples in a relation cannot be identical in all attributes.

C. Tuples in a relation are not ordered.

D. Tuples in a relation have identical values under the foreign key.

(d) Which of the following is **not true** about the weak entity?

A. A weak entity has a key.

B. A weak entity must imply an existence dependency.

C. An existence dependency needs not to imply a weak entity.

D. A weak entity can be identified with its own attributes.

(e) Which of the following statement about indexing is **true**?

A. Ordered indices are generally better at answering range search compared to hashing.

B. A file can have more than one indices.

C. Secondary indices have to be sparse.

D. When a file is modified, every index on the file must be updated.

E. The search key of a primary index must be the primary key of the relation.

Question 2　Database Design (16 points)

You are required to build a database to manage the student societies in a university.

- For each **society**, you need to record its name, contact email, and the web site. We can reasonably assume that the society name can uniquely identify a society.
- Society issues **membership card** to member students. Each card has a unique card id, the holder's name, holder's student id, issue date and membership expiry date.
- **Student** applies for a membership card of a society in order to join the society. For every student, information like student ID, name, email, gender and department are needed to be stored.
- Each society must elect an **executive committee** every semester. Information associated with the executive committee includes the semester, the year and the office location. Members in the executive committee are students, who manage the society. A student can join in the executive committee of some society by taking some position (like chairperson, secretary, manager and etc), and he/she should work for the executive committee for the whole semester.

(a) Draw an ER diagram for the database and **write down any assumptions** you have made.
(b) Construct appropriate relation schemas (**with primary keys underlined**) for the ER diagram in (a).

Question 3 RA and SQL (18 points)

Suppose a bookstore has the following five relational tables:

BOOK (BID, TITLE, *AID*, SUBJECT, QUANTITY-IN-STOCK)
AUTHOR (AID, FIRST-NAME, LAST-NAME)
CUSTOMER (CID, FIRST-NAME, LAST-NAME)
ORDER-DETAILS (*OID, BID*, QUANTITY)
ORDER (OID, *CID*, ORDER-YEAR)

ASSUMPTIONS:
Keys are underlined and foreign keys are in *italics*.
Each author has authored at least one book in the store.
Each book has exactly one author.
Each order is made by exactly one customer and has one or more associated record in ORDER-DETAILS (e.g., an order may contain different books).

Express the following queries in (a) to (c) using (i) the **relational algebra (RA)**, and (ii) **SQL** expressions.
(a) Find all book titles of the author whose last name is "Sipser"
(b) Find the IDs of customers who have not bought any book.

Question 4 Constraints and Normalization (16 points)

Consider the following database schema for answering (a) to (c):

```
Create table customer
( customer-name      char(20) not null,
customer-street   char(30),
customer-city     char(30),
primary key (customer-name))

create table branch
( branch-name        char(15) not null,
branch-city       char(30),
assets            integer,
primary key (branch-name))

create table account
( branch-name        char(15),
account-number  char(10) not null,
balance       integer,
primary key(account-number),
foreign key (branch-name) references branch)

create table depositor
( customer-name char(20) not null,
account-number  char(10) not null,
primary key (customer-name, account-number),
foreign key (account-number) references account,
foreign key (customer-name) references customer)
```

(a) Write down all non-trivial FDs that have only *one* attribute on the left hand side and *two* attributes on the right hand side.(E.g., A →BC, where A, B, C are attributes)

(b) Explain what the database system should check in order to preserve the constraints in the given schema in the following two cases:

(Case 1) Insert a new tuple in the account table.

(Case 2) Delete an old tuple in the account table.

(c) Assume we also have the FD: branch-city → assets, in addition to the constraints of schema branch, justify whether the schema branch is in 3NF? If not, show how to obtain a lossless-join decomposition into 3NF of schema.

Question 5 Query Optimization (16 points)

Consider a database consisting of the following three relation schemas:

CUSTOMERS (CID, Cname, job, age, street, city)
BRANCHES (BID, Bname, city)
ACCOUNTS (*CID, BID*, balance)

- The meaning of the attributes in the above schemas is self-explanatory. For example, Bname is the name of the branch. Keys are underlined and foreign keys are in *italics*.
- There is no index on all the relations.
- The relation CUSTOMERS has 10,000 tuples and 20 tuples of CUSTOMERS fit into one block.
- The relation BRANCHES has 2,000 tuples and 40 tuples of CUSTOMERS fit into one block.
- The relation ACCOUNTS has 12,000 tuples and 30 tuples of ACCOUNTS fit on one block.

(a) Assume that we use block nested-loop join to compute CUSTOMERS ⋈ ACCOUNTS using CUSTOMERS as the outer relation. Estimate the cost of the join in terms of block accesses under the following three memory buffer settings:

(i) The memory buffer has no restriction in size.
(ii) The memory buffer has 3 blocks.
(iii) The memory buffer has 22 blocks.

(b) Consider CUSTOMERS ⋈ BRANCHES ⋈ ACCOUNTS.
Consider the following two strategies for computing the join:

Strategy 1: (CUSTOMERS ⋈ BRANCHES) ⋈ ACCOUNTS
Strategy 2: (CUSTOMERS ⋈ ACCOUNTS) ⋈ BRANCHES

Which strategy is better? Explain the reason(s) of your choice.

Question 6 Indexing and Storage structure (16 points)

(a) Create a B+tree where each node can hold at most 3 pointers and 2 keys when the following keys are inserted in the following order: 1, 10, 2, 11, 3, 4.
(b) Using the date given in (a), draw the B+Tree created via bulk loading. And explain the advantage of bulk loading algorithm.

Question 7 Transaction Management (8 points)

Consider the schedule S for transations T_i, i=1,2,3: $W_3(A)$; $R_1(A)$; $W_1(B)$; $R_2(B)$; $R_3(C)$; $W_2(C)$; where W_i and R_i are actions for T_i. Answer the following questions:
(a) Draw the precedence graph for the schedule S, and determine whether it is conflict-serializable or not.
(b) If you find that S is **serializable** in (a), write down all equivalent serial schedules of S.