同济大学课程考核试卷（A 卷）
2012 — 2013 学年第一学期

课号：10014501 课名：数据库系统原理（双语） 考试考查：考试

此卷选为：期中考试(✓)、期终考试( )、重考( ) 试卷

年级_____专业____________________学号__________姓名_____________得分__________

Ⅰ. Multiple choice (20 marks, 2 marks each)

（C ）1. Five basic relational algebra operations are _____, others can be derived from these operations.

A. ∩，－，π，σ，× B. ∪，－，π，σ，⋈

C. ∪，－，π，σ，× D. ∩，÷，π，σ，⋈

（ABD）2. The following aggregation function(s) _____ will neglect null value.

A. SUM B. MAX

C. COUNT D. AVG

（A. ）3. Given R<U, F>, U={A,B,C}, F={B→C}, a decomposition of R is ρ={AB, BC}, and the decomposition is:

A. lossless-join, dependency preserving B. lossless-join, not dependency preserving

C. lossy-join, dependency preserving D. lossy-join, not dependency preserving

（BD ）4. When we generate relational schemas from an E-R diagram, the rules for relationship sets are:

A. for a binary 1: n relationship set, translate it into a relation, and the primary key of the relationship set is the primary key of the “1” side entity set;

B. for a binary 1: n relationship set, translate it into a relation, and the primary key of the relationship set is the primary key of the “n” side entity set;

C. a binary 1: n relationship set can be united with the “1” side entity set, and translated into one relation;

D. a binary 1: n relationship set can be united with the “n” side entity set, and translated into one relation;

（ABC）5. If R∈BCNF, then:

A. non-attributes are entirely functional dependent on non-key attributes;

B. all key attributes are entirely functional dependent on each candidate key that does not contain them;

C. all partial dependencies and transitive dependencies are removed for any attributes;

D. all anomalies are removed within any dependency scope.

（ AD）6. For PRIMARY KEY definition in CREATE TABLE, when a user wants to insert a tuple into the table, the RDBMS will:

A. check for the primary key value, and deny the insertion if it is not unique;

B. check for all attributes, and deny the insertion if there is one null value;

C. check for all candidate key attributes, and deny the insertion if there is one null value;

D. check for all primary key attributes, and deny the insertion if there is one null value

（ B. ）7. Transforming ER diagram into relational data model belongs to the step of the following database design:_______________.

A. Requirement analysis B. Logical design

C. Conceptual design D. Physical design

（ABD）8. When dealing with null value in SQL statements, the following correct is(are) _______________.

A. AGE IS NULL B. AGE IS NOT NULL

C. AGE = NULL D. NOT (AGE IS NULL)

（ C ）9. The global constraints in SQL statements are check clause based on tuples and _______________.

A. NOT NULL constraint B. Domain constraint

C. Assertion D. Foreign Key clause

（ACDE）10. A DBA's duties are:

A. database restoration and recovery;

B. transaction processing and concurrency control;

C. security and integrity control;

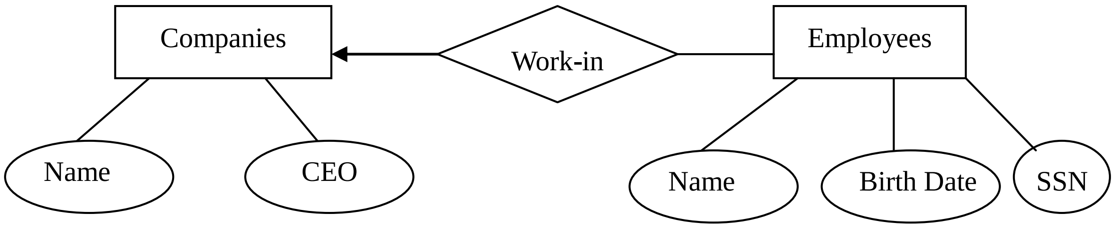
D. performance monitoring, analyzing, and tuning;

E. database reforming and restructuring.

Ⅱ. Analysis (12 marks, 6 marks each)

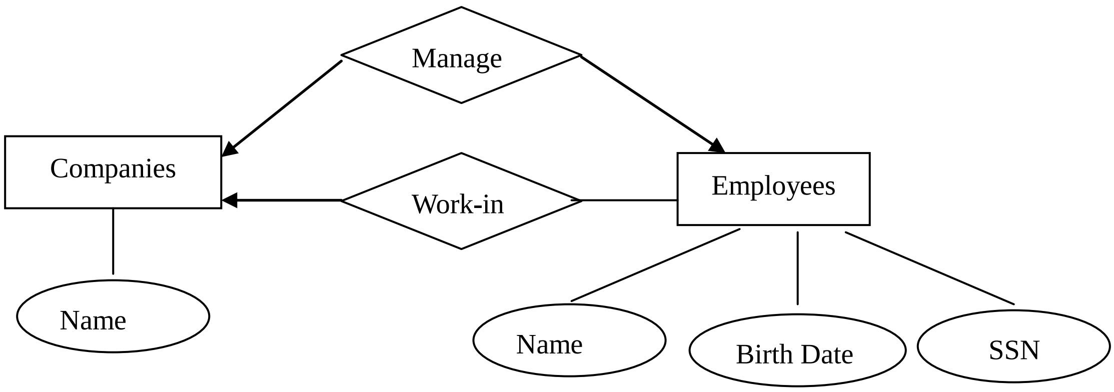
Are the following E/R diagrams good designs? Justify your answer and try to give better designs if they exist. If you feel they work well under some requirements but not under others, give specific examples. To simplify, keys are not underlined in the diagrams.

a. Below is an E/R diagram for companies and employees. One constraint for the E/R diagram below is that each employee can work in only one company.



Solution:

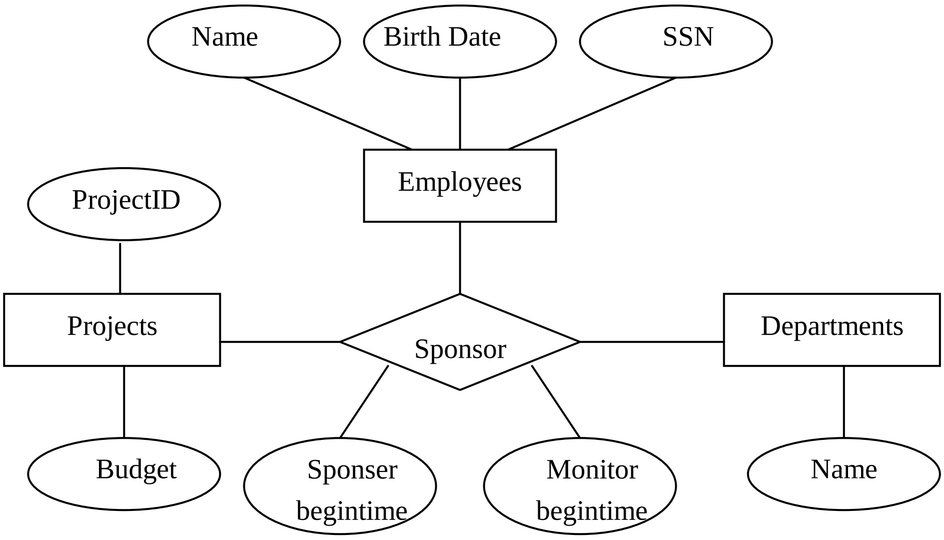
By making "manager" an attribute, the diagram maintains the one-to-one relationship. However, it does not guarantee that the manager is one of the employees.



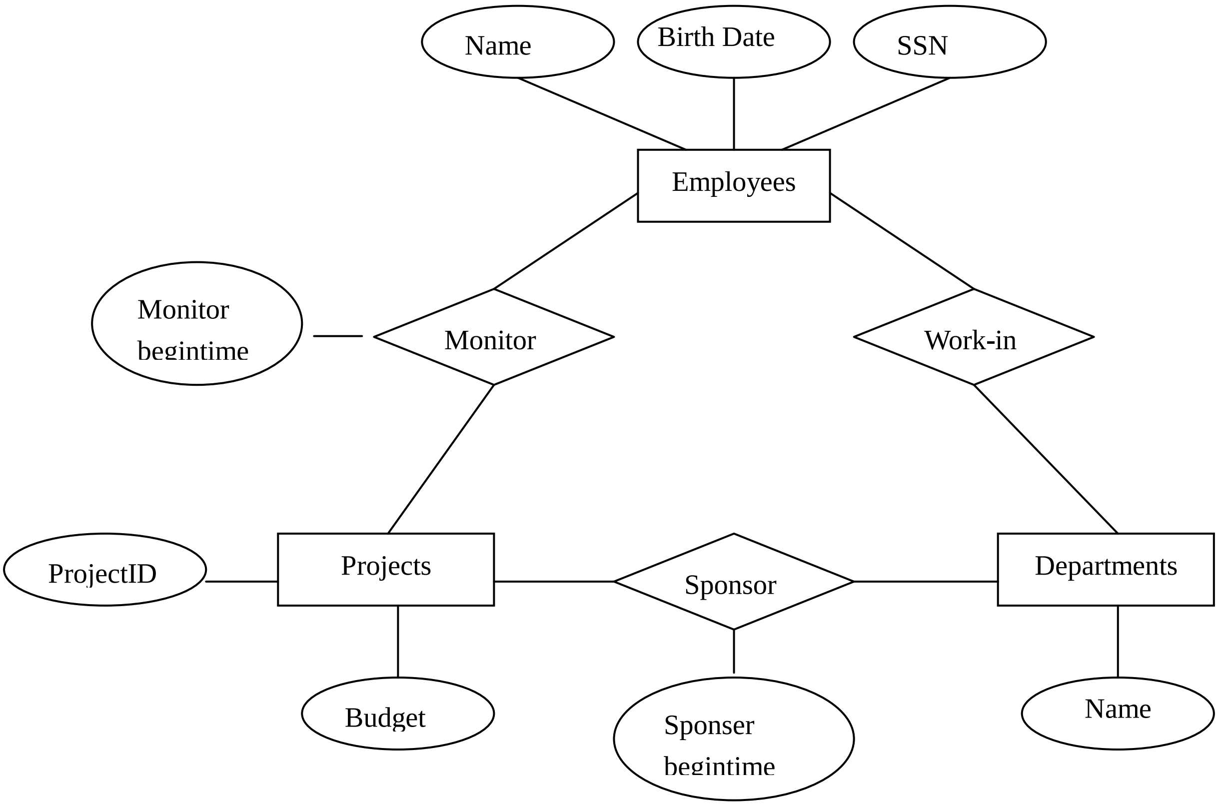
b. Below is an E/R diagram for a company database involving projects, departments, and employees. Each project can be sponsored by several departments and monitored by several employees in the company. Each department can sponsor several projects. Each employee can monitor several projects and may work in several departments. An employee can monitor a project that is *not* in his department.

Solution:

In this case a trade-off needs to be made between binary relationship and ternary relationship. The same information can be modeled more naturally using the E/R diagram as below. It's not a good idea to blend the concept of "sponsor" and "monitor" together. Also, in the original diagram, we lost the information of which employee works in which company.



Solution:



Ⅲ. Relational Database Design and Query (50 marks)

1. Assume four relations P, Q, R, and S containing the following tuples, respectively. Give out the results of the following expressions. **(6 marks, 2 marks for each)**

P: A B C D   Q: B C E   R: B C E   S: C D

a	b	c	d
a	b	e	f
b	c	e	f
b	d	c	d
b	d	e	f

b	d	e
b	c	f
d	c	a
d	e	b

d	c	a
b	d	e
b	c	e

c	d
e	f

a. $P \bowtie Q$ b. $P \div S$ c. $\sigma_{1=5}(Q \times S)$

a.

A	B	C	D	E
a	b	c	d	f
b	d	c	d	a
b	d	e	f	b

b.

A	B
a	b
b	d

c.

B	C	E	C	D
d	c	a	c	d
d	e	b	c	d

2. Consider a relation schema R<U, F>, where U={A,B,C,D,E,G}, and the set of functional dependencies F={AB→C, C→A, BC→D, ACD→B, D→EG, BE→C, CG→BD, CE→AG}. **(6 marks, 3 marks for each)**

    a. Compute the attribute closure $(AB)_F^+$.

    b. Find a minimal cover for F.

a. $(AB)_F^+$ = ABCDEG

b. A minimal cover for F is:

{AB→C, C→A, BC→D, CD→B, D→E, D→G, BE→C, CG→D, CE→G}

3. Consider the relation schema R<U, F>, where U={A,B,C,D}, and F={A→B, C→A, C→D, D→B, B→C}. Suppose that ρ={R_1, R_2, R_3} is a decomposition of R, where R_1={A, B}, R_2={B, C}, and R_3={C, D}. Is the decomposition ρ of <U, F> lossless? **(5 marks)**

The decomposition ρ of <U, F> is lossless.

4. Given the relational schema R<U, F>, U={A,B,C,D,E }, F={AC→BD, B→C, C→D, B→E};
   a. Use Armstrong axioms and related rules to proof the functional dependency AC→E; **(3 marks)**
   b. Find a canonical cover Fc of F, and find all candidate keys, then point out R is in which normal form; **(5 marks)**
   c. Decompose R into 3NF, which the decomposition is lossless-join and dependency preserving. **(5 marks)**

Solution:

a. 由 AC→BD 得 AC→B（分解规则）；再由 B→E，则有 AC→E（传递规则）

b. 对 R<U, F>中的函数依赖集 F 进行极小化处理，得最小依赖集 Fm={AC→B, B→C, C→D, B→E}，仍记为 F；

c. 判定 R 属于第几范式：

R 的候选码有：AC、AB；主属性为 A、B、C；

由 C→D 可见，非主属性 D 对码 AC 为部分函数依赖，故 R∉2NF, R∈1NF。

e.将关系模式 R 分解为 3NF：

全部属性均在 F 中出现了；不存在 X→A∈F，且 XA=U。

则对 F 按相同左部原则分组，有

U1={A,B,C}, F1={AC→B, B→C }

U2={B,C,E}, F2={B→C, B→E }

U3={C,D},　F3={C→D }

ρ={R1<U1,F1>, R2<U2,F2>, R3<U3,F3>}为保持函数依赖的分解。

由于码 AC、AB 都包含在 U1 中，因此，由检测算法可以找到相应表中的一行可以成为 a1,a2,a3,a4,a5，则 ρ 同时也具有无损连接性。

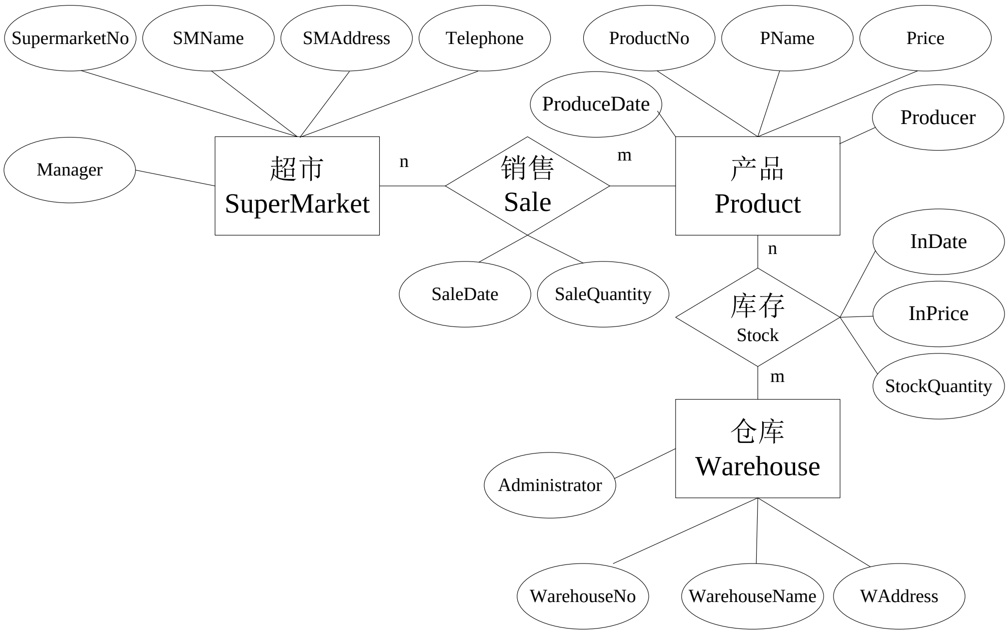
5. For a simplified application on Supermarket Management, there are three entity sets: Supermarket（超市）, Product（产品）, and Warehouse（仓库）. The attributes of entity set Supermarket includes: SupermarketNo, SMName, SMAddress, Telephone, Manager; and the attributes of Product contains: ProductNo, PName, Price, Producer, ProduceDate; and the attributes of Warehouse contains: WarehouseNo, WarehouseName, WAddress, Adminstrator. And two relationship sets: Supermarket and Product related through a binary relationship set Sale（销售）; and Product and Warehouse related through a binary relationship set Stock（库存）. The two relationship sets have the following attributes respectively: Sale has attributes: SaleDate and SaleQuantity; Stock has attributes: InDate, InPrice, and StockQuantity.

   a. Please give the corresponding ER diagram. **(5 marks)**
   b. Create the corresponding relational schemas, and point out the primary keys and the foreign keys. **(5 marks)**
   c. Use SQL statements to define Product table, and give proper integrity constraints. **(5 marks)**
   d. Use relational algebra expression and SQL statement for the following query:

Find the supermarket's name and quantity of all Chocolate manufactured by GODIVA which sale price is between 100 and 500 RMB, and its stock in the warehouse is less than 20 items. **(5 marks)**

Solution:

a.



b. Supermarket=(SupermarketNo, SMName, SMAddress, Telephone, Manager)

Product=(ProductNo, PName, Price, Producer, ProduceDate)

Warehouse=(WarehouseNo, WarehouseName, Waddress, Administratior)

Sale=(SupermarketNo, ProductNo, SaleDate, SaleQuantity)

Stock=(ProductNo, WarehouseNo, InDate, InPrice, StockQuantity)

c. create table Product(

ProductNo char(10),

PName char(30) not null,

Price numeric(10,2),

Producer char(30),

ProduceDate date,

primary key (ProductNo),

check Price>0)

d. Relational algebra expression:

π SMName, SaleQuantity + StockQuantity(σ PName=' Chocalate' and Producer=' GODIVA' and (Price is between 100 and 500) and StockQuantity<20 (Supermarket ⋈ Sale ⋈ Product ⋈ Stock))

SQL statement:

```
Select SMName, SaleQuantity + StockQuantity as Quantity
From Supermarket S, Product P, Sale, Stock
Where Pname=' Chocalate'
      and Producer=' GODIVA'
      and Price is between 100 and 500
      and StockQuantity < 20
      and S.SupermarketNo = Sale.SupermarketNo
      and Sale.ProductNo = P.ProductNo
      and P.ProductNo = Stock.ProductNo
```

Ⅳ. Choose THREE questions (or items) of the following and please give your explanations respectively. (18 marks, 6 marks each)

(1) DB (database) and DBMS (database management system)

(2) Data independence

(3) Database Normalization and the differences between 3NF and BCNF

(4) How to design a database? What is a good design for relational database? How to achieve a good design?

(5) How many Turing Award winners in database area? Who are they, and for what contributions did they gain the Turing Award?

(6) What are the Challenges and Opportunities for Database Research?

Solution:

(1) A database is a structured collection of data. The data are typically organized to model relevant aspects of reality (for example, the availability of rooms in hotels), in a way that supports processes requiring this information (for example, finding a hotel with vacancies).

Database Management System (DBMS) is a set of programs that enables you to store, modify, and extract information from a database, it also provides users with tools to add, delete, access, modify, and analyze data stored in one location. A group can access the data by using query and reporting tools that are part of the DBMS or by using application programs specifically written to access the data. DBMS's also provide the method for maintaining the integrity of stored data, running security and users access, and recovering information if the system fails. The information from a database can be presented in a variety of formats.

(2) Data independence is the type of data transparency that matters for a centralized DBMS. It refers to the immunity of user applications to make changes in the definition and organization of data.

Physical data independence deals with hiding the details of the storage structure from user applications. The application should not be involved with these issues, since there is no difference in the operation carried out against the data.

(3) Database normalization is the process of organizing the fields and tables of a relational database to minimize redundancy and dependency. Normalization usually involves dividing large tables into smaller (and less redundant) tables and defining relationships between them. The objective is to isolate data so that additions, deletions, and modifications of a field can be made in just one table and then propagated through the rest of the database via the defined relationships. A 3NF definition that is equivalent to Codd's, but expressed differently, was given by Carlo Zaniolo in 1982. This definition states that a table is in 3NF if and only if, for each of its functional dependencies X → A, at least one of the following conditions holds:

- X contains A (that is, X → A is trivial functional dependency), or
- X is a super key, or

Every element of A-X, the set difference between A and X, is a prime attribute

If a relational scheme is in BCNF then all redundancy based on functional dependency has been removed, although other types of redundancy may still exist. A relational schema R is in Boyce–Codd normal form if and only if for every one of its dependencies X → Y, at least one of the following conditions hold:[4]

- X → Y is a trivial functional dependency (Y ⊆ X)
- X is a super key for schema R

3NF tables are not meeting BCNF. Only in rare cases does a 3NF table not meet the requirements of BCNF. A 3NF table which does not have multiple overlapping candidate keys is guaranteed to be in BCNF.[5]Depending on what its functional dependencies are, a 3NF table with two or more overlapping candidate keys may or may not be in BCNF

(4) Database design is the process of producing a detailed data model of a database. This logical data model contains all the needed logical and physical design

choices and physical storage parameters needed to generate a design in a Data Definition Language, which can then be used to create a database. A fully attributed data model contains detailed attributes for each entity.

The term database design can be used to describe many different parts of the design of an overall database system. Principally, and most correctly, it can be thought of as the logical design of the base data structures used to store the data. In the relational model these are the tables and views. In an object database the entities and relationships map directly to object classes and named relationships. However, the term database design could also be used to apply to the overall process of designing, not just the base data structures, but also the forms and queries used as part of the overall database application within the database management system (DBMS).[1]

The process of doing database design generally consists of a number of steps which will be carried out by the database designer. Usually, the designer must:

- Determine the relationships between the different data elements.
- Superimpose a logical structure upon the data on the basis of these relationships.

(5) Barchman、EF Codd、Jim Gray

(6) Big data etc