**Quiz 4, 5, 6, 7, 8 Fall 18**

*Test Directions:* On your honor, you agree to neither give assistance nor receive assistance throughout the exam. Any suspicion of cheating will be dealt with swiftly and harshly. Answer each question to the best of your ability. For all coding questions, partial credit is available. Therefore it is in your best interest to attempt to answer every question.

There are no intentional syntax mistakes in the code.

Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Note:** We will be taking random attendance 5 times for the rest of the semester. Failing to be present in any of those days will deduct 100 points as if you are not in class to take the quiz.﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿﻿

1. (14.31) Consider the following relation for published books:

BOOK (Book\_title, Authorname, Book\_type, Listprice, Author\_affil, Publisher)

Author\_affil referes to the affiliation of the author. Suppose the following dependencies exist:

Book\_title -> Publisher, Book\_type

Book\_type -> Listprice

Author\_name -> Author-affil

(a) What normal form is the relation in? Explain your answer.

(b) Apply normalization until you cannot decompose the relations further. State the reasons behind each decomposition.

1. (14.33) Consider the following relation: R (Doctor#, Patient#, Date, Diagnosis, Treat\_code, Charge)

In this relation, a tuple describes a visit of a patient to a doctor along with a treatment code and daily charge. Assume that diagnosis is determined (uniquely) for each patient by a doctor. Assume that each treatment code has a fixed charge (regardless of patient). Is this relation in 2NF? Justify your answer and decompose if necessary. Then argue whether further normalization to 3NF is necessary, and if so, perform it.

1. (Section 20.3) List and briefly explain the ACID properties provided by a Database Management System.
2. (Chapter 8) Describe briefly what is a query tree and what are its components?
3. (Chapter 20) Briefly describe the basic Two-Phase Locking (2PL) protocol.
4. (16.34) Consider a disk with the following characteristics (these are not parameters of any particular disk unit): block size B=512 bytes, interblock gap size G=128 bytes, number of blocks per track=20, number of tracks per surface=400. A disk pack consists of 15 double-sided disks.
   1. What is the total capacity of a track and what is its useful capacity (excluding interblock gaps)?
   2. How many cylinders are there?
   3. What is the total capacity and the useful capacity of a cylinder?
   4. What is the total capacity and the useful capacity of a disk pack?
   5. Suppose the disk drive rotates the disk pack at a speed of 2400 rpm (revolutions per minute); what is the transfer rate in bytes/msec and the block transfer time btt in msec? What is the average rotational delay rd in msec? What is the bulk transfer rate (see Appendix B)?
   6. Suppose the average seek time is 30 msec. How much time does it take (on the average) in msec to locate and transfer a single block given its block address?
   7. Calculate the average time it would take to transfer 20 random blocks and compare it with the time it would take to transfer 20 consecutive blocks using double buffering to save seek time and rotational delay.
5. (17.18) Consider a disk with block size B=512 bytes. A block pointer is P=6 bytes long, and a record pointer is P R =7 bytes long. A file has r=30,000 EMPLOYEE records of fixed-length. Each record has the following fields: NAME (30 bytes), SSN (9 bytes), DEPARTMENTCODE (9 bytes), ADDRESS (40 bytes), PHONE (9 bytes), BIRTHDATE (8 bytes), SEX (1 byte), JOBCODE (4 bytes), SALARY (4 bytes, real number). An additional byte is used as a deletion marker.
   1. Calculate the record size R in bytes.
   2. Calculate the blocking factor bfr and the number of file blocks b assuming an unspanned organization.
   3. Suppose the file is ordered by the key field SSN and we want to construct a primary index on SSN. Calculate (i) the index blocking factor bfr i (which is also the index fan-out fo); (ii) the number of first-level index entries and the number of first-level index blocks; (iii) the number of levels needed if we make it into a multi-level index; (iv) the total number of blocks required by the multi-level index; and (v) the number of block accesses needed to search for and retrieve a record from the file--given its SSN value--using the primary index.
6. (14.19) Suppose we have the following requirements for a university database that is used to keep track of students’ transcripts:
   1. The university keeps track of each student's name (SNAME), student number (SNUM), social security number (SSSN), current address (SCADDR) and phone (SCPHONE), permanent address (SPADDR) and phone (SPPHONE), birthdate (BDATE), sex (SEX), class (CLASS) (freshman, sophomore, ..., graduate), major department (MAJORDEPTCODE), minor department (MINORDEPTCODE) (if any), and degree program (PROG) (B.A., B.S., ..., Ph.D.). Both ssn and student number have unique values for each student.
   2. Each department is described by a name (DEPTNAME), department code (DEPTCODE), office number (DEPTOFFICE), office phone (DEPTPHONE), and college (DEPTCOLLEGE). Both name and code have unique values for each department.
   3. Each course has a course name (CNAME), description (CDESC), code number (CNUM), number of semester hours (CREDIT), level (LEVEL), and offering department (CDEPT). The value of code number is unique for each course.
   4. Each section has an instructor (INSTUCTORNAME), semester (SEMESTER), year (YEAR), course (SECCOURSE), and section number (SECNUM). Section numbers distinguish different sections of the same course that are taught during the same semester/year; its values are 1, 2, 3, ...; up to the number of sections taught during each semester.
   5. A grade record refers to a student (Ssn), refers to a particular section, and grade (GRADE).

Design a relational database schema for this database application. First show all the functional dependencies that should hold among the attributes. Then, design relation schemas for the database that are each in 3NF or BCNF. Specify the key attributes of each relation. Note any unspecified requirements, and make appropriate assumptions to make the specification complete.

1. (20.14) Change transaction T 2 in Figure 21.2b to read:

read\_item(X);

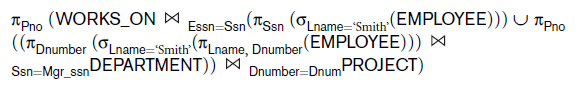
X:= X+M;

if X > 90 then exit

else write\_item(X);

Discuss the final result of the different schedules in Figure 21.3 (a) and (b), where M = 2 and N = 2, with respect to the following questions. Does adding the above condition change the final outcome? Does the outcome obey the implied consistency rule (that the capacity of X is 90)?

1. (22.21) Suppose that the system crashes before the [read\_item,T3,A] entry is written to the log in Figure 23.1 (b); will that make any difference in the recovery process?
2. (22.24) Suppose that we use the deferred update protocol for the example in Figure 23.6. Show how the log would be different in the case of deferred update by removing the unnecessary log entries; then describe the recovery process, using your modified log. Assume that only redo operations are applied, and specify which operations in the log are redone and which are ignored.
3. Describe the meaning of a Table Scan
4. Why are the table scans considered to be a problem
5. How are the table scans avoided?
6. Describe the relationship
   1. Schema and Relation
   2. Relation and Attribute
   3. Relation and Tuples
7. Translate the following set of relational operations into a single SQL query:



1. Which of following languages is used to define the application’s queries?
   1. DDL
   2. DQL
   3. DML
   4. Metadata
2. What does DDL define?
3. What mechanism is used to authorize user?
4. How do we eliminate duplicates from a select’s result set?
5. Which of the following describe the act of making a transaction’s changes to the database state permanent?
   1. The SAVE operation
   2. The ROLLBACK operation
   3. The COMPLETE operation
   4. The COMMIT operation
6. What best describes the referential integrity?
   1. Primary to Foreign Key
   2. Primary to Candidate Key
   3. Foreign to Primary Key
7. Which of the clauses is only useful when used in SELECT statements containing aggregation function? e.g. min(), max(), sum(), etc.?
   1. Group By
   2. Having
   3. Order By
   4. In
8. Which of the following is not true of views?
   1. Serves to decouple clients from the internal schema
   2. Can be used to cache the results from complex aggregated queries
   3. Implements CRUD operations across joined tables
   4. Can be used to provide clients access to otherwise protected tables
9. Describe a records blocking factor.
10. What are the two strategies employed by the query processor when performing searches on two attributes from a single table.
    1. Consider the case where one attribute is indexed and the other is not
    2. Consider the case when both attributes are separately indexed

Use the example of searching for a student with last-name and phone number

1. Name and describe the four levels of Transaction Isolation provided by SQL clients
2. How does 2 Phase Locking guarantee the schedules when all transactions are following 2 phase locking?
3. What activities identify the layout of the table files and index files on the target file system?
4. What describes the general solution to normalizing relations that are not in second and third normal form?
5. Is this true that functional dependencies are symmetric?
6. A relation in First Normal Form must be in \_\_\_\_\_\_\_\_\_\_\_\_\_\_ form.
7. Fundamentally what do indexing and other optimization techniques minimize?
8. What is the disadvantage of Unspanned Records?
9. What is one of the characteristics of Unordered Data Files?
10. What is the main purpose of using an index?
11. Under what circumstances a secondary index is required?
12. Is the following schedule serializable? R1(Y); R1(X); W1(X);R2(X);W1(Y);C1;W2(X);
13. Which of the locks allow shared access to an object?
    1. Write Locks
    2. Read Locks
    3. Read-Write Locks
    4. Binary Locks
14. Create an EER based on the description.

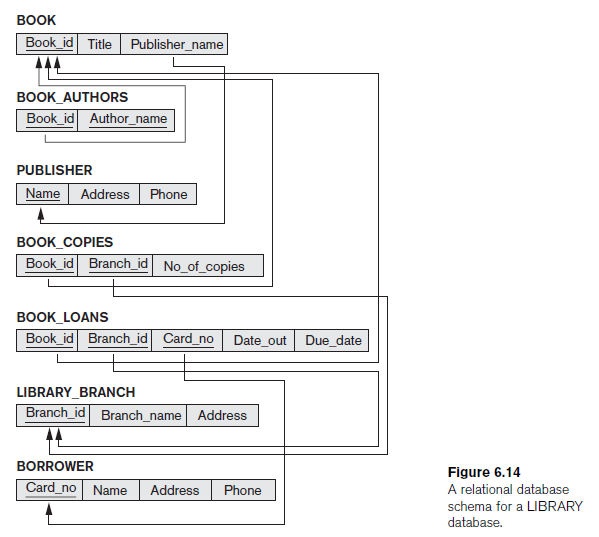
The problem:

Buying and selling artwork is a very high business, and anything that is a big business wants its data to be handled neatly and cleanly, because a fake painting sold for millions is an embarrassment. In that respect, the Marcus Art Gallery has hired you to create a database to track their pictures.

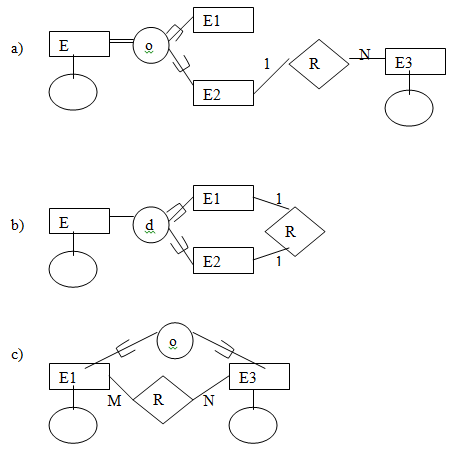
The Description:

* Every painting has a unique catalog number, a location, and a text description
* Every artist has a name, a date of birth, and a unique library reference number
* Each artist can pain many paintings, but no painting is created by multiple artists. Every painting has been painted by an artist, and that relationship has a date painted
* Three subclasses of artist exist
  + The first is the genuine master painter, which includes information about their country of origin and the century they painted.
  + The second subclass is the copyist someone under license creates reproduction of original painting, which includes information about their current business address, their telephone number, and the name of their agent
  + The third class of artist is the forger who is a known faker of art and the gallery needs to be aware of their activities, which include if the forger is active, in prison, and which country the faker currently resides
* Two independent individuals are a dealer and a buyer
* A dealer is a person with a name, a contact number, and a certain level of reputation
  + Their name and contact number make up a unique identifying number
  + A dealer sells the painting to the gallery, and that relationship has a price attribute and a date. Many different dealers sell many different paintings
* A buyer is a person with a name, a contact number, and a shipping address
  + Their name and contact number make up a unique identifying number
  + A buyer buys the painting from the gallery, and that relationship has a price attribute and a date and a commission. Many different buyers buy many different paintings

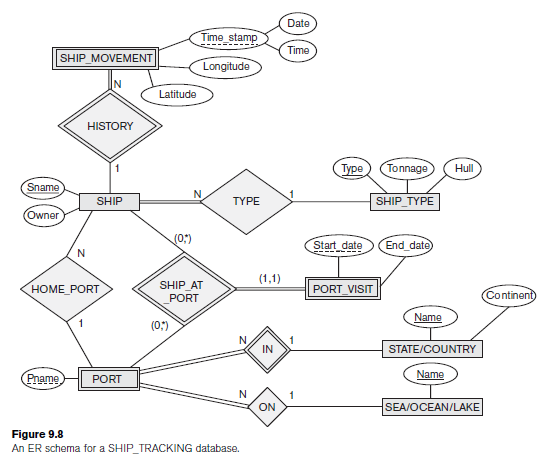
1. Try to map the relational schema of Figure 6.14 into an ER schema. This is part of a process known as reverse engineering, where a conceptual schema is created for an existing implemented database. State any assumptions you make.



1. Which of the following EER diagram(s) is/are incorrect and why?

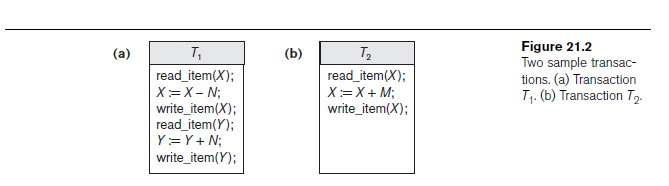


1. Figure 9.8 shows an ER schema for a database that may be used to keep track of transport ships and their locations for maritime authorities. Map this schema into a relational schema, and specify all primary keys and foreign keys.

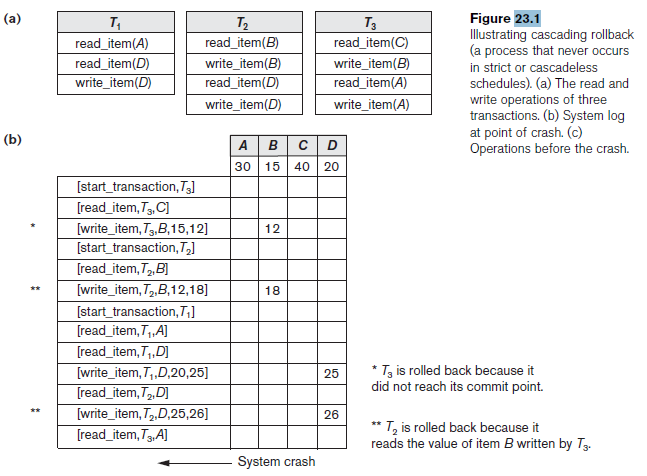


21. What act or operation makes transaction's changes to the database state permanent?

9. Fig 21.2



10. Fig 23.1b



11. 23.6

