

**SOLUTIONS MANUAL**



**DATABASE  
CONCEPTS**  
FOURTH EDITION



DAVID M. KROENKE    DAVID J. AUER

---

**David M. Kroenke and David J. Auer**  
**Database Concepts (4<sup>th</sup> Edition)**  
**Instructor's Manual**

---

**CHAPTER TWO**  
**THE RELATIONAL MODEL**



Prepared by  
David J. Auer  
Western Washington University

▶ **CHAPTER OBJECTIVES**

- Learn the conceptual foundation of the relational model
- Understand how relations differ from nonrelational tables
- Learn basic relational terminology
- Learn the meaning and importance of keys, foreign keys, and related terminology
- Understand how foreign keys represent relationships
- Learn the purpose and use of surrogate keys
- Learn the meaning of functional dependencies
- Learn to apply a process for normalizing relations

▶ **CHAPTER ERRATA**

- Page 74 — Figure 2-19 has an extra Department column. The correct table is:

EmployeeNumber	LastName	Email	Department	DeptPhone
100	Johnson	JJ@somewhere.com	Accounting	834-1100
200	Abernathy	MA@somewhere.com	Finance	834-2100
300	Smathers	LS@somewhere.com	Finance	834-2100
400	Caruthers	TC@somewhere.com	Accounting	834-1100
500	Jackson	TJ@somewhere.com	Production	834-4100
600	Caldera	EC@somewhere.com	Legal	834-3100
700	Bandalone	RB@somewhere.com	Legal	834-3100

- Page 74 — Figure 2-20 has an extra Department column. The correct table is:

EmployeeNumber	LastName	Email	Department
100	Johnson	JJ@somewhere.com	Accounting
200	Abernathy	MA@somewhere.com	Finance
300	Smathers	LS@somewhere.com	Finance
400	Caruthers	TC@somewhere.com	Accounting
500	Jackson	TJ@somewhere.com	Production
600	Caldera	EC@somewhere.com	Legal
700	Bandalone	RB@somewhere.com	Legal

Department	DeptPhone
Accounting	834-1100
Finance	834-2100
Legal	834-3100
Production	834-4100

▶ **THE ACCESS WORKBENCH**

All solutions for sections of The Access Workbench are in the separate The Access Workbench document.

### TEACHING SUGGESTIONS

- The Art Course database discussed in Chapter One is a good database to use for an in-class demo of the concepts in this chapter. The DBMS screenshots in Chapter Two use that database as the example database. For example, see Figures 2.7, 2.8, and 2.9. See the list, data and database files supplied, and use the following material:
  - MS Access:
    - "Art Course List" in DBC-e04-Lists-And-Data.xls
    - Art-Course-Database-CH01.accdb
  - SQL Server 2005/2008 Express Edition:
    - DBC-e04-MSSQL-Art-Course-Database-Create-Tables.sql
    - DBC-e04-MSSQL-Art-Course-Database-Insert-Data.sql
    - NOTE — Create a database diagram for the database
  - MySQL 5.X:
    - DBC-e04-MySQL-Art-Course-Database-Create-Tables.sql
    - DBC-e04-MySQL-Art-Course-Database-Insert-Data.sql
- The goal of this chapter is to present an overview of the major elements of the relational model. This includes the definition of a relation, important terminology, the use of surrogate keys, and basic design principles.
- Students often misconstrue the statement that only a single element is allowed in a cell to mean that the cells must be fixed in length. One can have a variable length memo in a cell, but that is considered, semantically, to be one thing. By the way, there are a number of reasons for this restriction. Perhaps the easiest to explain is that SQL has no means for addressing sub-elements in a cell.
- When students execute SQL SELECTs, they may generate relations with duplicate rows. Such results do not fit the definition of relations, but they are considered relations nonetheless. This is a good example of “theory vs. practice”.
- You may want to emphasize that foreign keys and the primary key which they reference need not have the same name. They must, however, have the same underlying set of values (domain). This means not just that the values look the same — it means that the values mean the same thing. A foreign key of CatName and a foreign key ValentineNickName might look the same, but they do not mean the same thing. Using ValentineNickName as a foreign key to Name in the relation CAT would return some weird results.
- Referential integrity constraints are important. You might ask the students to think of an example when a foreign key does not have a referential integrity constraint (answer: whenever a parent row is optional, say, STUDENTs need not have an ADVISER).
- We favor the use of surrogate keys. Unless there is a natural, numeric ID (like PartNumber), we almost always add a surrogate key to our database designs. Sometimes a surrogate key will be added even if there is a natural, numeric ID for consistency. Surrogate keys can cause problems (primarily patching up foreign keys) if the database imports data from other databases that either do not employ a surrogate key or use a different one. In some cases, institutions have developed policies for ensuring that surrogate keys are unique, globally. It’s probably best for the students to

get into the habit of using them and consider not using them as an exception. Professional opinions vary on this, however.

- If you're using Oracle, then you'll need to teach the use of sequences to implement surrogate keys. Sequences are an awkward solution to this problem, however, and may be why surrogate keys are less used in the Oracle-world. Maybe there will be a better solution to them from Oracle in the future.
- The discussion of functional dependencies is critical — maybe the most important in the book. If students can understand that all tables do is record “data points” of functional dependencies, then normalization will be easier and seem more natural.
- In physics, because there are formulae like  $F = ma$ , we need not store tables and tables of data recording data points for force, mass, and acceleration. The formula suffices for all data points. However, there is no formula for computing how much a customer of, say, American Airlines, owes for his or her ticket from New York to Houston. If we could say the cost of an airline ticket was \$.05 per mile, then we could compute the cost of a ticket, and tables of airline flight prices would be unnecessary. But, we cannot — it all depends on ... So, we store the data points for functional dependencies in tables.
- This chapter presents the design principle that every determinant should be a candidate key. This is, of course, the definition of Boyce-Codd Normal Form. This leaves out 4NF, 5NF, and domain/key normal form. At this level, we do not think those omissions are critical. See the normalization discussion in Chapter Five for a bit more on this topic, however.
- If we use domain/key normal form as the ultimate, then, insofar as functional dependencies are concerned, the domain/key definition that “every constraint is a logical consequence of domains and keys,” comes down to Boyce-Codd Normal Form. Therefore, we proceed on good theoretical ground with the discussion as presented in this chapter.
- Students should understand three ambiguities in a null value. This understanding will help them comprehend the issues addressed by INNER and OUTER joins in the next chapter.
- Exercises 2.40 and 2.41 deal with multivalued dependencies and fourth normal form (4NF). They are instructive to show students how to deal with situations where the value of one column in a table is associated with several values of another attribute in (at least initially) the same table. This is an important concept, and after BCNF the next important concept students need to understand about normalization.

### ▶ ANSWERS TO REVIEW QUESTIONS

2.1 *Why is the relational model important?*

It is the single most important standard in database processing and is used for the design and implementation of almost every commercial database worldwide.

2.2 *Define the term entity and give an example.*

Entity is the formal name for a “thing” that is being tracked in a database, and is defined as something of importance to the user that needs to be represented in the database.

#### TEXTBOOK

2.3 *List the characteristics a table must have to be considered a relation.*

- Rows contain data about an entity
- Columns contain data about attributes of the entity
- Cells of the table hold a single value
- All entries in a column are of the same kind
- Each column has a unique name
- The order of the columns is unimportant
- The order of the rows is unimportant

2.4 *Give an example of a relation (other than the one in this chapter).*

TEXTBOOK (ISBN, Title, Publisher, Copyright)

2.5 *Give an example of a table that is not a relation (other than one from this chapter).*

TEXTBOOK (ISBN, Title, Publisher, Copyright, Authors)

where there are multiple author names in the Authors column.

2.6 *Under what circumstances can an attribute of a relation be of variable length?*

If that attribute is considered to be a single thing like a memo or other variable length data item.

2.7 *Explain the use of the terms file, record, and field.*

These terms are synonyms for table, row, column. These terms, however, generally refer to pre-relational bases.

2.8 *Explain the use of the terms relation, tuple, and attribute.*

These terms are synonyms for table, row, column. These terms, however, are the ones used in relational database theory.

2.9 Under what circumstances can a relation have duplicate rows?

When manipulating a relation with a DBMS we may end up with duplicate rows. While in theory we should eliminate the duplicates, in practice this is often not done.

2.10 Define the term unique key and give an example.

A column whose values identify one and only one row.

**TEXTBOOK** (ISBN, Title, Publisher, Copyright)

Where **ISBN** is a unique identifier.

2.11 Define the term nonunique key and give an example.

A column, used to identify that potentially identifies many rows.

**TEXTBOOK** (ISBN, Title, Publisher, Copyright)

Where **Publisher** is a nonunique identifier.

2.12 Give an example of a relation with a unique composite key.

**APARTMENT** (BuildingNumber, ApartmentNumber, NumBedrooms, Rent)

where (BuildingNumber, ApartmentNumber) is a unique composite key.

2.13 Explain the difference between a primary key and a candidate key.

Both are unique identifiers. One is chosen to be the identifier for the relation and for foreign keys based on the relation. The other could be chosen as well, but since it is not, it is called a candidate.

2.14 Describe four uses of a primary key.

- Identify a row
- Represent the row in foreign keys
- Organize storage for the relation
- Basis for indexes and other structures to facilitate searching in storage

2.15 What is a surrogate key, and under what circumstances would you use one?

A surrogate key is a unique, numeric identifier that is appended to a relation to serve as the primary key. It is used when the relation does not have a key that is short, numeric, and with values that never change.

2.16 How do surrogate keys obtain their values?

They are supplied automatically by the DBMS.

## Chapter Two — The Relational Model

---

2.17 *Why are the values of surrogate keys normally hidden from users on forms, queries, and reports?*

Surrogate keys are normally hidden because they usually have no meaning to the users.

2.18 *Explain the term foreign key and give an example.*

A foreign key is a key value that corresponds to a primary key in a relation other than itself.

**TEXTBOOK** (ISBN, *Title*, *Publisher*, *Copyright*)

**PUBLISHER** (Name, *Street*, *City*, *State*, *Zip*)

**Publisher** in TEXTBOOK is a foreign key that references **Name** in PUBLISHER.

2.19 *Explain how primary keys and foreign keys are denoted in this book.*

Primary keys are underlined and foreign keys are in italics.

2.20 *Define the term referential integrity constraint and give an example of one.*

A rule which specifies that every value of a foreign key matches a value of the primary key.

**Publisher in TEXTBOOK must exist in Name in PUBLISHER.**

2.21 *Explain three possible interpretations of a null value.*

- Value not appropriate
- Value known to be blank
- Value appropriate and unknown

2.22 *Give an example of a null value (other than one from this chapter), and explain each of the three possible interpretations for that value.*

A null value for the attribute DeceasedDate in the table SUBSCRIBER.

- The subscriber may be a corporation and a value is inappropriate.
- The subscriber may be alive, and the value is known to be blank.
- The subscriber may be dead, but the date of death is unknown, and the value is appropriate, but not none.



- 2.23 Define the terms functional dependency and determinant using an example not in this book

A functional dependency is a logical relationship in which the value of one item in the relationship can be determined by knowing the value of the other item. For example:

**ISBN → Title**

This means that if the ISBN (of a textbook) is known, then we will also know (can determine) the title. The item on the left—the one whose value is known—is called the determinant.

- 2.24 In the following equation, name the functional dependency and identify the determinant(s).

**Area = Length X Width**

The functional dependency is:

**(Length, Width) → Area**

(Length, Width) is the determinant.

Note this is different than saying "Length and Width are the determinants".

- 2.25 Explain the meaning of the following expression.

**A → (B, C)**

Given this expression, tell if it is also true that:

**A → B**

and:

**A → C**

The functional dependency:

**A → (B, C)**

means that a value of A determines the value of both B and C.

Yes, it is true that.

**A → B and A → C**

2.26 Explain the meaning of the following expression:

$$(D, E) \rightarrow F$$

Given this expression, tell if it is also true that:

$$D \rightarrow F$$

and

$$E \rightarrow F$$

The functional dependency:

$$(D, E) \rightarrow F$$

means that the values of the pair of attributes (D,E) determine the value of F.

No, it is not true that.

$$D \rightarrow F \text{ and } E \rightarrow F$$

2.27 Explain the differences in your answers to questions 2.25 and 2.26.

$A \rightarrow (B, C)$  is just shorthand for  $A \rightarrow B$  and  $A \rightarrow C$

However,  $(D, E) \rightarrow F$  means that the composite, as a whole, identifies F.

For example:

$$\text{EmployeeNumber} \rightarrow (\text{FirstName}, \text{LastName})$$

This means that  $\text{EmployeeNumber} \rightarrow \text{FirstName}$  and that  $\text{EmployeeNumber} \rightarrow \text{LastName}$ .

But:

$$(\text{FirstName}, \text{LastName}) \rightarrow \text{HireDate}$$

does not mean that  $\text{FirstName} \rightarrow \text{HireDate}$  — there could be lots of employees named "Bob".

2.28 Define the term primary key in terms of functional dependencies.

A primary key is one or more attributes that functionally determine all of the other attributes.

2.29 If we assume that a relation has no duplicate data, how do we know there is always at least one primary key?

Because the collection of all the attributes in the relation can identify a unique row.

## Chapter Two — The Relational Model

---

- 2.30 *How does your answer to question 2.29 change if we allow a relation to have duplicate data?*

It doesn't work—such tables do not have a primary key.

- 2.31 *Using your own words, describe the nature and purpose of the normalization process.*

The purpose of the normalization process is to prevent update problems in the tables (relations) in the database. The nature of the normalization process is that we break up relations as necessary to ensure that every determinant is a candidate key.

- 2.32 *Examine the data in the Veterinary Office List in Figure 1-26 (see page 50), and state assumptions about functional dependencies in this table. What is the danger of making such conclusions on the basis of sample data?*

**PetName → (Type, Breed, DOB, OwnerLastName, OwnerFirstName, OwnerPhone, OwnerEmail)**

**OwnerEmail → (OwnerLastName, OwnerFirstName, OwnerPhone)**

**OwnerPhone → (OwnerLastName, OwnerFirstName, OwnerEmail)**

The danger is that there may be possibilities not apparent from sample data. For example, two owners might have pets with the same name.

- 2.33 *Using the assumptions you stated in your answer to question 2.32, what are the determinants of this relation? What attribute(s) can be the primary key of this relation?*

Attributes that can be the primary key are called candidate keys.

**Determinants:            PetName, OwnerEmail, OwnerPhone**

**Candidate keys:        PetName**

- 2.34 *Describe a modification problem when changing data in the relation in question 2.32 and a second modification problem when deleting data in this relation.*

Changes to owner data may need to be made in several rows.

Deleting data for the last pet of an owner deletes owner data as well.

- 2.35 *Examine the data in the Veterinary Office List—Version Two in Figure 1-27 (see page 50), and state assumptions about functional dependencies in this table.*

**PetName → (Type, Breed, DOB, OwnerLastName, OwnerFirstName, OwnerPhone, OwnerEmail)**

**OwnerEmail → (OwnerLastName, OwnerFirstName, OwnerPhone)**

**OwnerPhone → (OwnerLastName, OwnerFirstName, OwnerEmail)**

**(PetName, Date) → (Service, Charge)**

## Chapter Two — The Relational Model

---

The last functional dependency assumes a pet is seen at most on one day and that there is no standard charge for a service.

- 2.36 *Using the assumptions you stated in your answer to question 2.35, what are the determinants of this relation? What attribute(s) can be the primary key of this relation?*

**Determinants:**            **PetName, OwnerEmail, OwnerPhone, (PetName, Date)**

**Candidate keys:**        **(PetName, Date)**

- 2.37 *Explain a modification problem when changing data in the relation in question 2.35 and a second modification problem when deleting data in this relation.*

Same as 2.34:

Changes to owner data may need to be made in several rows.

Deleting data for the last pet of an owner deletes owner data as well.

▶ **ANSWERS TO EXERCISES**

- 2.38 Apply the normalization process to the Veterinary Office List relation shown in Figure 1-26 to develop a set of normalized relations. Show the results of each of the steps in the normalization process.

**STEP ONE:**

**PET-AND-OWNER** (PetName, Type, Breed, DOB, OwnerLastName, OwnerFirstName, OwnerPhone, OwnerEmail)

**Functional Dependencies:**

PetName → (Type, Breed, DOB, OwnerLastName, OwnerFirstName, OwnerPhone, OwnerEmail)

OwnerEmail → (OwnerLastName, OwnerFirstName, OwnerPhone)

OwnerPhone → (OwnerLastName, OwnerFirstName, OwnerEmail)

**PET-AND-OWNER Candidate Keys:** PetName

*Is every determinant a candidate key?*

NO — OwnerEmail and OwnerPhone are NOT candidate keys.

**STEP TWO:**

Break into two relations: OWNER and PET

**OWNER** (OwnerLastName, OwnerFirstName, OwnerPhone, OwnerEmail)

**PET** (PetName, Type, Breed, DOB, [Foreign Key ???])

**FOR OWNER:**

**Functional Dependencies:**

OwnerEmail → (OwnerLastName, OwnerFirstName, OwnerPhone)

OwnerPhone → (OwnerLastName, OwnerFirstName, OwnerEmail)

**OWNER Candidate Keys:** OwnerPhone, OwnerEmail

*Is every determinant a candidate key?*

YES — OwnerEmail and OwnerPhone are candidate keys — Normalization complete!

We can choose either candidate key as primary key.

## Chapter Two — The Relational Model

---

(A) IF WE USE OwnerPhone as primary key, THEN:

OWNER (OwnerPhone, OwnerLastName, OwnerFirstName, OwnerEmail)

PET (PetName, Type, Breed, DOB, OwnerPhone)

*Functional Dependencies:*

PetName → (Type, Breed, DOB, OwnerPhone)

*PET Candidate Keys:* PetName

*Is every determinant a candidate key?*

YES — PetName is a candidate key — Normalization complete!

**FINAL NORMALIZED REALTIONS:**

OWNER (OwnerPhone, OwnerLastName, OwnerFirstName, OwnerEmail)

PET (PetName, Type, Breed, DOB, OwnerPhone)

(B) IF WE USE OwnerEmail as primary key, THEN:

OWNER (OwnerPhone, OwnerLastName, OwnerFirstName, OwnerEmail)

PET (PetName, Type, Breed, DOB, OwnerEmail)

*Functional Dependencies:*

PetName → (Type, Breed, DOB, OwnerEmail)

*PET Candidate Keys:* PetName

*Is every determinant a candidate key?*

YES — PetName is a candidate key — Normalization complete!

**FINAL NORMALIZED REALTIONS:**

OWNER (OwnerPhone, OwnerLastName, OwnerFirstName, OwnerEmail)

PET (PetName, Type, Breed, DOB, OwnerEmail)

## Chapter Two — The Relational Model

---

- 2.39 Apply the normalization process to the Veterinary Office List—Version Two relation shown in Figure 1-27 to develop a set of normalized relations. Show the results of each of the steps in the normalization process.

### **STEP ONE:**

**PET-AND-OWNER** (PetName, Type, Breed, DOB, OwnerLastName, OwnerFirstName, OwnerPhone, OwnerEmail, Service, Date, Charge)

#### **Functional Dependencies:**

**PetName** → (Type, Breed, DOB, OwnerLastName, OwnerFirstName, OwnerPhone, OwnerEmail)

**OwnerEmail** → (Owner, OwnerPhone)

**OwnerPhone** → (Owner, OwnerEmail)

**(PetName, Date)** → (Service, Charge)

The last functional dependency assumes a pet is seen at most on one day and that there is no standard charge for a service.

**PET-AND-OWNER Candidate Keys:** (PetName, Date)

**Is every determinant a candidate key?**

**NO** — PetName, OwnerEmail and OwnerPhone are NOT candidate keys.

### **STEP TWO:**

**Break into two relations:** OWNER and PET-SERVICE

**OWNER** (OwnerLastName, OwnerFirstName, OwnerPhone, OwnerEmail)

**PET-SERVICE** (PetName, Type, Breed, DOB, [Foreign Key ???], Service, Date, Charge)

**FOR OWNER:**

#### **Functional Dependencies:**

**OwnerEmail** → (OwnerLastName, OwnerFirstName, OwnerPhone)

**OwnerPhone** → (OwnerLastName, OwnerFirstName, OwnerEmail)

**OWNER Candidate Keys:** OwnerPhone, OwnerEmail

**Is every determinant a candidate key?**

**YES** — OwnerEmail and OwnerPhone are candidate keys — Normalization complete!

**We can choose either candidate key as primary key. We will use OwnerPhone.**

## Chapter Two — The Relational Model

---

If a student chooses OwnerEmail, the steps will be similar as shown in Exercise 2.37.

**IF WE USE OwnerPhone as primary key, THEN:**

**OWNER** (OwnerPhone, OwnerLastName, OwnerFirstName, OwnerEmail)

**PET-SERVICE** (PetName, Type, Breed, DOB, OwnerPhone, Service, Date, Charge)

**FOR PET-SERVICE:**

**Functional Dependencies:**

**PetName** → (Type, Breed, DOB, OwnerPhone)

(**PetName**, Date) → (Service, Charge)

The last functional dependency assumes a pet is seen at most on one day and that there is no standard charge for a service.

**PET-AND-SERVICE Candidate Keys:** (**PetName**, Date)

**Is every determinant a candidate key?**

**NO** — PetName is NOT a candidate key.

**STEP THREE:**

**Break PET-SERVICE into two relations:**                      **PET and SERVICE**

**OWNER** (OwnerPhone, OwnerLastName, OwnerFirstName, OwnerEmail)

**PET** (PetName, Type, Breed, DOB, OwnerPhone)

**SERVICE** (PetName, Date, Service, Charge)

**PET Functional Dependencies:**

**PetName** → (Type, Breed, DOB, OwnerPhone)

**PET Candidate Keys:**                      **PetName**

**Is every determinant a candidate key?**

**YES** — PetName is a candidate key — Normalization complete!



## Chapter Two — The Relational Model

---

**SERVICE Functional Dependencies:**

**(PetName, Date) → (Service, Charge)**

The functional dependency assumes a pet is seen at most on one day and that there is no standard charge for a service.

**SERVICE Candidate Keys:** **(PetName, Date)**

*Is every determinant a candidate key?*

**YES — (PetName, Date) is a candidate key — Normalization complete!**

**FINAL NORMALIZED REALTIONS:**

**OWNER (OwnerPhone, OwnerLastName, OwnerFirstName, OwnerEmail)**

**PET (PetName, Type, Breed, DOB, OwnerPhone)**

**SERVICE (PetName, Date, Service, Charge)**

2.40 Consider the following relation.

**STUDENT (StudentNumber, StudentName, SiblingName, Major)**

Assume that the values of SiblingName are the names of all of a given student's brothers and sisters; also assume that students have at most one major.

A. Show an example of this relation for two students, one of whom has three siblings and the other of whom has only two siblings.

StudentNum	StudentName	SiblingName	Major
100	Mary Jones	Victoria	Accounting
100	Mary Jones	Slim	Accounting
100	Mary Jones	Reginald	Accounting
200	Fred Willows	Rex	Finance
200	Fred Willows	Billy	Finance

B. List the candidate keys in this relation.

**STUDENT Candidate Keys:** **(StudentNumber, SiblingName)**

This assumes that StudentName is not unique.

C. State the functional dependencies in this relation.

**StudentNumber → (StudentName, Major)**

**(StudentNumber, SiblingName) → (StudentName, Major)**

D. Explain why this relation does not meet the relational design criteria set out in this chapter (i.e., why this is not a well-formed relation).

## Chapter Two — The Relational Model

---

Some attributes are functionally dependent on a part of the composite primary key.

- E. *Divide this relation into a set of relations that do meet the relational design criteria (i.e., that are well formed).*

**Break into two relations:      STUDENT and STUDENT-SIBLING**

**STUDENT** (StudentNumber, StudentName, Major)

**STUDENT-SIBLING** (StudentNumber, SiblingName)

**FOR STUDENT-SIBLING:**

**Functional Dependencies:**

**(StudentNumber, SiblingName) → (StudentNumber)**

**(StudentNumber, SiblingName) → (SiblingName)**

**STUDENT-SIBLING Candidate Keys:      (StudentNum, SiblingName)**

*Is every determinant a candidate key?*

**YES — (StudentNum, SiblingName) is a candidate key — Normalization complete!**

**FOR STUDENT:**

**STUDENT** (StudentNumber, StudentName, Major)

**Functional Dependencies:**

**StudentNumber → (StudentName, Major)**

**STUDENT Candidate Keys:      StudentNumber**

*Is every determinant a candidate key?*

**YES — StudentNumber is a candidate key — Normalization complete!**

**FINAL NORMALIZED REALTIONS:**

**STUDENT** (StudentNumber, StudentName, Major)

**STUDENT-SIBLING** (StudentNumber, SiblingName)

## Chapter Two — The Relational Model

2.41 Alter question 2.40 to allow students to have multiple majors. In this case, the relational structure is:

**STUDENT** (StudentNumber, StudentName, SiblingName, Major)

- A. Show an example of this relation for two students, one of whom has three siblings and the other of whom has one sibling. Assume each student has a single major.

StudentNum	StudentName	SiblingName	Major
100	Mary Jones	Victoria	Accounting
100	Mary Jones	Slim	Accounting
100	Mary Jones	Reginald	Accounting
200	Fred Willows	Rex	Finance

- B. Show the data changes necessary to add a second major only for the first student.

StudentNum	StudentName	SiblingName	Major
100	Mary Jones	Victoria	Accounting
100	Mary Jones	Slim	Accounting
100	Mary Jones	Reginald	Accounting
200	Fred Willows	Rex	Finance
100	Mary Jones	Victoria	Info Systems
100	Mary Jones	Slim	Info Systems
100	Mary Jones	Reginald	Info Systems

- C. Based on your answer to part B, show the data changes necessary to add a second major for the second student.

StudentNum	StudentName	SiblingName	Major
100	Mary Jones	Victoria	Accounting
100	Mary Jones	Slim	Accounting
100	Mary Jones	Reginald	Accounting
200	Fred Willows	Rex	Finance
100	Mary Jones	Victoria	Info Systems
100	Mary Jones	Slim	Info Systems
100	Mary Jones	Reginald	Info Systems
200	Fred Willows	Rex	Accounting

- D. Explain the differences in your answers to questions parts B and C. Comment on the desirability of this situation.

We had to add three rows in the first case — one major for each of the siblings of the student. If we didn't do that, it would appear the student has a sibling with one major, but doesn't have the sibling as a second major. This is nuts!

## Chapter Two — The Relational Model

---

E. Divide this relation into a set of well-formed relations.

If we split STUDENT into two relations, STUDENT and STUDENT-SIBLING, then we get:

**STUDENT** (StudentNumber, StudentName, Major)

**STUDENT-SIBLING** (StudentNumber, SiblingName)

The relation is identical to the STUDENT-SIBS relation in 2.38 above, and is properly normalized. Now we need to check STUDENT-MAJOR.

**FOR STUDENT:**

**STUDENT** (StudentNumber, StudentName, Major)

**Functional Dependencies:**

StudentNumber → (StudentName)

(StudentNumber, Major) → StudentName

**STUDENT Candidate Keys:** (StudentNumber, Major)

*Is every determinant a candidate key?*

**NO** — StudentNum is NOT a candidate key.

**Break into two relations:** STUDENT-2 and STUDENT-MAJOR

**STUDENT-2** (StudentNumber, StudentName)

**STUDENT-MAJOR** (StudentNumber, Major)

**FOR STUDENT-2:**

**Functional Dependencies:**

StudentNumber → StudentName

**STUDENT\_2 Candidate Keys:** StudentNumber

*Is every determinant a candidate key?*

**YES** — StudentNumber is a candidate key — Normalization complete!

## Chapter Two — The Relational Model

---

**FOR STUDENT-MAJOR:**

**Functional Dependencies:**

**(StudentNumber, Major) → StudentNumber**

**(StudentNumber, Major) → Major**

**STUDENT\_2 Candidate Keys: (StudentNumber, Major)**

**Is every determinant a candidate key?**

**YES — (StudentNumber, Major) is a candidate key — Normalization complete!**

**FINAL NORMALIZED REALTIONS:**

**STUDENT-2 (StudentNumber, StudentName)**

**STUDENT-MAJOR (StudentNumber, Major)**

**STUDENT-SIBLING (StudentNumber, SiblingName)**

- 2.42 *The text states that one can argue that “the only reason for having relations is to store instances of functional dependencies.” Explain what this means in your own words.*

In a properly normalized relation, each row of the relation consists of a primary key value, which is a determinant, and attribute values which are all functionally dependent on the primary key. Thus, properly normalized relations store instances of functional dependencies, and only instances of functional dependencies. So we can say that the purpose of relations is to store instances of functional dependencies.

**ANSWERS TO GARDEN GLORY PROJECT QUESTIONS**

Figure 2-29 shows data that Garden Glory collects about properties and services.

PropertyName	Type	Street	City	ZIP	ServiceDate	Description	Amount
Eastlake Building	Office	123 Eastlake	Seattle	98119	5/5/2008	Lawn Mow	\$ 42.50
Elm St Apts	Apartment	4 East Elm	Lynnwood	98223	5/8/2008	Lawn Mow	\$ 123.50
Jeferson Hill	Office	42 West 7th St	Bellevue	98040	5/8/2008	Garden Service	\$ 53.00
Eastlake Building	Office	123 Eastlake	Seattle	98119	5/10/2008	Lawn Mow	\$ 42.50
Eastlake Building	Office	123 Eastlake	Seattle	98119	5/12/2008	Lawn Mow	\$ 42.50
Elm St Apts	Apartment	4 East Elm	Lynnwood	98223	5/15/2008	Lawn Mow	\$ 123.50
Eastlake Building	Office	123 Eastlake	Seattle	98119	5/19/2008	Lawn Mow	\$ 42.50

- A. Using these data, state assumptions about functional dependencies among the columns of data. Justify your assumptions on the basis of these sample data and also on the basis of what you know about service businesses.

From the data it appears that there are many functional dependencies that could be defined. Some examples are:

**PropertyName → Type**

**(PropertyName, Street) → (Type, City, Zip)**

**(PropertyName, City) → (Type, Street, Zip)**

**(PropertyName, Zip) → (Type, Street, City)**

**(PropertyName, Description, ServiceDate) → Amount**

None of these seem to be more than just coincidence, however. It would seem, for example, that an ‘Elm St Apts’ could exist in more than one city — there are certainly enough cities with a street name Elm Street! There is simply not enough data to reply on it. Logically, it seems that we need one ID column—a surrogate key will be required here.

With regard to services, it would seem likely that a given service could be given to the same property, but on different dates. So, if we had a good determinant for property, then the last functional dependency would be true. So, the following seems workable:

**PropertyID → (PropertyName, Type, Street, City, Zip)**

**(PropertyID, Description, ServiceDate) → Amount**

- B. Given your assumptions, comment on the appropriateness of the following designs.

- PROPERTY (PropertyName, Type, Street, City, Zip, ServiceDate, Description, Amount)  
NOT GOOD: For example, PropertyName does not determine ServiceDate.
- PROPERTY (PropertyName, Type, Street, City, Zip, ServiceDate, Description, Amount)  
NOT GOOD: There may be more than one service on a given date.

## Chapter Two — The Relational Model

---

3. PROPERTY (PropertyName, Type, Street, City, Zip, ServiceDate, Description, Amount)  
NOT GOOD: For example, (PropertyName, ServiceDate) does not determine Description since there may be more than one service at a property on a given date.
4. PROPERTY (PropertyID, PropertyName, Type, Street, City, Zip, ServiceDate, Description, Amount)  
NOT GOOD: For example, PropertyID does not determine ServiceDate.
5. PROPERTY (PropertyID, PropertyName, Type, Street, City, Zip, ServiceDate, Description, Amount)  
NOT GOOD: For example, (PropertyID, ServiceDate) does not determine Description since there may be more than one service at a property on a given day.
6. PROPERTY (PropertyID, PropertyName, Type, Street, City, Zip, *ServiceDate*)  
and  
SERVICE (ServiceDate, Description, Amount)  
BETTER: ServiceDate is properly set up as a foreign key in PROPERTY. HOWEVER, will limit the system to only one service per property — the foreign key is in the wrong table!
7. PROPERTY (PropertyID, PropertyName, Type, Street, City, Zip, *ServiceDate*)  
and  
SERVICE (ServiceID, ServiceDate, Description, Amount)  
NOT GOOD: ServiceDate is supposedly a foreign key in PROPERTY, but isn't even a primary key in SERVICE. This simply doesn't work!
8. PROPERTY (PropertyID, PropertyName, Type, Street, City, Zip, *ServiceID*)  
and  
SERVICE (ServiceID, ServiceDate, Description, Amount, *PropertyID*)  
NOT GOOD: ServiceID is properly used as a foreign key in PROPERTY, but we also have PropertyID as a foreign key in SERVICE. This simply doesn't work — there cannot be two foreign keys like this! The question becomes, which one should we keep?
9. PROPERTY (PropertyID, PropertyName, Type, Street, City, Zip)  
and  
SERVICE (ServiceID, ServiceDate, Description, Amount, *PropertyID*)  
GOOD! Finally the relationship is set up correctly. Now we can have many services (even on the same date) for one property.

C. Suppose Garden Glory decides to add the following table.

**SERVICE\_FEE (PropertyID, ServiceID, ServiceDate, Amount)**

Add this table to what you consider to be the best design in your answer to question B. Modify the tables from question B as necessary to minimize the amount of duplicate data. Will this design work for the data in Figure 2-28? If not, modify the design so that this data will work. State the assumption implied by this design.

Here's the best design from Question B:

**PROPERTY(PropertyID, PropertyName, Type, Street, City, Zip)**

**SERVICE(ServiceID, ServiceDate, Description, Amount, PropertyID)**

Adding

**SERVICE-FEE (PropertyID, ServiceID, Description, Amount)**

Means that we need to take Description and Amount out of SERVICE. This leaves

**PROPERTY(PropertyID, PropertyName, Type, Street, City, Zip)**

**SERVICE (ServiceID, ServiceDate, PropertyID)**

**SERVICE-FEE (PropertyID, ServiceID, Description, Amount)**

These relations seem to imply that there is a standard description and fee for a service on a property, and that there is a need to record when that standard service was applied to the property. Thus, it would make sense to consider the relations in the order:

**PROPERTY(PropertyID, PropertyName, Type, Street, City, Zip)**

**SERVICE-FEE (PropertyID, ServiceID, Description, Amount)**

**SERVICE (ServiceID, ServiceDate, PropertyID)**

Note that PropertyID of SERVICE-FEE is a foreign key into PROPERTY. Now, for this design to make sense, we need to allow for multiple services on a property by making (ServiceID, ServiceDate, PropertyID) the key in SERVICE. This also means that (ServiceID, PropertyID) is a foreign key into SERVICE-FEE.

It also makes sense to rearrange the field in SERVICE so that we have:

**PROPERTY(PropertyID, PropertyName, Type, Street, City, Zip)**

**SERVICE-FEE (PropertyID, ServiceID, Description, Amount)**

**SERVICE (PropertyID, ServiceID, ServiceDate)**

This means that a service can be applied to a property on different, but multiple dates, that a property can have multiple, but different, services on the same date, and that a service can be applied to multiple, but different, properties on the same date. A property may not have the same service on the same date, however. All of this seems reasonable. Now we need to check with the users.



▶ **ANSWERS TO JAMES RIVER JEWELRY PROJECT QUESTIONS**

Figure 2-30 shows data that James River Jewelry collects for its frequent buyer program.

Name	Phone	Email	InvoiceNumber	Date	PreTaxAmount
Elizabeth Stanley	555-236-7789	ES@somewhere.com	1000	5/5/2008	\$ 155.00
Fred Price	555-236-0091	FP@somewhere.com	1010	5/7/2008	\$ 203.00
Linda Becky	555-236-0392	LB@somewhere.com	1020	5/11/2008	\$ 75.00
Pamela Birch	555-236-4493	PB@somewhere.com	1030	5/15/2008	\$ 67.00
Richardo Romez	555-236-3334	RR@somewhere.com	1040	5/15/2008	\$ 330.00
Elizabeth Stanley	555-236-7789	ES@somewhere.com	1050	5/16/2008	\$ 25.00
Linda Becky	555-236-0392	LB@somewhere.com	1060	5/16/2008	\$ 45.00
Elizabeth Stanley	555-236-7789	ES@somewhere.com	1070	5/18/2008	\$ 445.00
Smantha Jackson	555-236-1095	SJ@somewhere.com	1080	5/19/2008	\$ 72.00

- A. *Using these data, state assumptions about functional dependencies among the columns of data. Justify your assumptions on the basis of these sample data and also on the basis of what you know about retail sales.*

From the data it would appear:

**Name → (Phone, Email)**

**Phone → (Name, Email)**

**Email → (Name, Phone)**

However, these are based on a very limited dataset and cannot be trusted. For example, name is not a good determinant in a retail application—there may be many customers with the same name. It’s also possible that some customers could have the same phone, even though they do not in this example.

Another functional dependency is:

**InvoiceNumber → (Name, Phone, Email, Date, PreTaxAmount)**

- B. *Given your assumptions, comment on the appropriateness of the following designs.*

- CUSTOMER (Name, Phone, Email, InvoiceNumber, Date, PreTaxAmount)**  
 NOT GOOD: For example, Name does not determine InvoiceNumber since one customer may have made more than one purchase.
- CUSTOMER (Name, Phone, Email, InvoiceNumber, Date, PreTaxAmount)**  
 TRUE, but not normalized. For example, Email → Phone. And why is InvoiceNumber the key for data about a customer?
- CUSTOMER (Name, Phone, Email, InvoiceNumber, Date, PreTaxAmount)**  
 GOOD FOR CUSTOMERS, BUT NOT INVOICES: Given a unique Email address, Email works as a key for customer data. Unfortunately, Email does not determine InvoiceNumber and therefore is not a sufficient key.

## Chapter Two — The Relational Model

---

4. CUSTOMER (CustomerID, Name, Phone, Email, InvoiceNumber, Date, PreTax Amount)  
GOOD FOR CUSTOMERS, BUT NOT INVOICES: A unique ID column is a good idea, and works as a key for customer data. Unfortunately, CustomerID does not determine InvoiceNumber and therefore is not a sufficient key.
5. CUSTOMER (Name, Phone, Email)  
and  
PURCHASE (InvoiceNumber, Date, PreTax Amount)  
GETTING BETTER, BUT INCOMPLETE. We cannot be sure Name is unique, and the relationship between CUSTOMER and PURCHASE is not defined.
6. CUSTOMER (Name, Phone, Email)  
and  
PURCHASE (InvoiceNumber, Date, PreTax Amount, *Email*)  
GOOD. The design breaks up the themes and has a proper foreign key. However, the use of Email as a primary key may be a problem if two customers share an Email address.
7. CUSTOMER (Name, Email)  
and  
PURCHASE (InvoiceNumber, Phone, Date, PreTax Amount, *Email*)  
A GOOD DESIGN WAS JUST MADE BAD AGAIN. The design breaks up the themes and has a proper foreign key. However, why was phone moved? It should have been left in CUSTOMER where it will only be entered once and where:

**Email → Phone**

- C. *Modify what you consider to be the best design in question B to include a column called AwardPurchaseAmount. The purpose of this column is to keep a balance of the customers' purchases for award purposes. Assume that returns will be recorded with invoices having a negative PreTaxAmount.*

The best design in Question B was number 6, so we'll put AwardPurchaseAmount in CUSTOMER. The result is:

**CUSTOMER (Name, Phone, Email, AwardPurchaseAmount)**

**PURCHASE (InvoiceNumber, Date, PreTax Amount, *Email*)**

However, with this design, there's no history of prior AwardPurchaseAmounts.

## Chapter Two — The Relational Model

---

- D. Add a new AWARD table to your answer to question C. Assume that the new table will hold data concerning the date and amount of an award that is given after a customer has purchased 10 items. Ensure that your new table has appropriate primary and foreign keys.

The new table is:

**AWARD** (AwardID, AwardDate, AwardAmount, AwardPurchaseAmount, Email)

The other tables need to be adjusted, and the final design will be:

**CUSTOMER** (Name, Phone, Email)

**PURCHASE** (InvoiceNumber, Date, PreTax Amount, Email, AwardID)

**AWARD** (AwardID, AwardDate, AwardAmount, AwardPurchaseAmount, Email)

Placing AwardID into PURCHASE as a foreign key allows for each purchase to be allocated to a particular award. Business rules need to be in place to ensure that the count of the number of PURCHASEs having a positive PreTaxAmount minus the count of the number having a negative PreTaxAmount never exceeds 10 for any given AWARD row.

**▶ ANSWERS TO THE QUEEN ANNE CURIOSITY SHOP PROJECT QUESTIONS**

Figure 2-31 shows typical sales data for the Queen Anne Curiosity Shop, and Figure 2-32 shows typical purchase data.

Figure 2-31

LastName	FirstName	Phone	PurchaseDate	Item	Price	Tax	Total
Shire	Robert	206-524-2433	14-Dec-08	Antique Desk	3,000.00	249.00	3,249.00
Shire	Robert	206-524-2433	14-Dec-08	Antique Desk Chair	500.00	41.50	541.50
Goodyear	Katherine	206-524-3544	15-Dec-08	Dining Table Linens	1,000.00	83.00	1,083.00
Bancroft	Chris	425-635-9788	15-Dec-08	Candles	50.00	4.15	54.15
Griffith	John	206-524-4655	23-Dec-08	Candles	45.00	3.74	48.74
Shire	Robert	206-524-2433	5-Jan-09	Desk Lamp	250.00	20.75	270.75
Tierney	Doris	425-635-8677	10-Jan-09	Dining Table Linens	750.00	62.25	812.25
Anderson	Donna	360-538-7566	12-Jan-09	Book Shelf	250.00	20.75	270.75
Goodyear	Katherine	206-524-3544	15-Jan-09	Antique Chair	1,250.00	103.75	1,353.75
Goodyear	Katherine	206-524-3544	15-Jan-09	Antique Chair	1,750.00	145.25	1,895.25
Tierney	Doris	425-635-8677	25-Jan-09	Antique Candle Holders	350.00	29.05	379.05

Figure 2-32

Item	PurchasePrice	PurchaseDate	Vendor	Phone
Antique Desk	1,800.00	7-Nov-08	European Specialties	206-325-7866
Antique Desk	1,750.00	7-Nov-08	European Specialties	206-325-7866
Antique Candle Holders	210.00	7-Nov-08	European Specialties	206-325-7866
Antique Candle Holders	200.00	7-Nov-08	European Specialties	206-325-7866
Dining Table Linens	600.00	14-Nov-08	Linens and Things	206-325-6755
Candles	30.00	14-Nov-08	Linens and Things	206-325-6755
Desk Lamp	150.00	14-Nov-08	Lamps and Lighting	206-325-8977
Floor Lamp	300.00	14-Nov-08	Lamps and Lighting	206-325-8977
Dining Table Linens	450.00	21-Nov-08	Linens and Things	206-325-6755
Candles	27.00	21-Nov-08	Linens and Things	206-325-6755
Book Shelf	150.00	21-Nov-08	Harrison, Denise	425-746-4322
Antique Desk	1,000.00	28-Nov-08	Lee, Andrew	425-746-5433
Antique Desk Chair	300.00	28-Nov-08	Lee, Andrew	425-746-5433
Antique Chair	750.00	28-Nov-08	New York Brokerage	206-325-9088
Antique Chair	1,050.00	28-Nov-08	New York Brokerage	206-325-9088

## Chapter Two — The Relational Model

---

- A. *Using these data, state assumptions about functional dependencies among the columns of data. Justify your assumptions on the basis of these sample data and also on the basis of what you know about retail sales.*

From the sample sales data it would appear:

**LastName → (FirstName, Phone)**

**FirstName → (LastName, Phone)**

**Phone → (LastName, FirstName)**

**Price → (Tax, Total)**

**(LastName, Date, Item) → (Price, Tax, Total)**

**(FirstName, Date, Item) → (Price, Tax, Total)**

**(PhoneName, Date, Item) → (Price, Tax, Total)**

However, these are based on a very limited dataset and cannot be trusted. For example, name is not a good determinant in a retail application—there may be many customers with the same name. It's also possible that some customers could have the same phone, even though they do not in this example. The one trustable functional dependency here is:

**Price → (Tax, Total)**

From the sample purchase data it would appear:

**(Item, PurchasePrice) → (PurchaseDate, Vendor, Phone)**

**(Item, PurchasePrice, PurchaseDate) → (Vendor, Phone)**

**(PurchasePrice, PurchaseDate) → (Item, Vendor, Phone)**

**Vendor → Phone**

**Phone → Vendor**

However, these are based on a very limited dataset and cannot be trusted. For example, Item is not a good determinant in a retail application—there may be many Items with the same designator. It's also possible that multiple purchases on the same purchase date will be for the purchase price. The most trustworthy functional dependencies here are:

**Vendor → Phone**

**Phone → Vendor**

But, the first of these may fail if the vendor has multiple phone numbers.

## Chapter Two — The Relational Model

---

B. Given your assumptions, comment on the appropriateness of the following designs.

1. CUSTOMER (LastName, FirstName, Phone, Email, PurchaseDate, Item, Price, Tax, Total)

NOT GOOD. There may be many customers with the same last name.

2. CUSTOMER (LastName, FirstName, Phone, Email, PurchaseDate, Item, Price, Tax, Total)

NOT GOOD. There may be many customers with the same last name and first name.

3. CUSTOMER (LastName, FirstName, Phone, Email, PurchaseDate, Item, Price, Tax, Total)

NOT GOOD. Phone will be fairly unique, and combined with FirstName may overcome the problem of two people with the same phone number being in the customer list (but not necessarily — what if three students are sharing an apartment, and two of them are Bill Smith and Bill Jones?). However, this will not determine the purchase information of purchase date, etc.

4. CUSTOMER (LastName, FirstName, Phone, Email, PurchaseDate, Item, Price, Tax, Total)

NOT GOOD. There may be many customers with the same last name and first name, and some of them may make a purchase on the same date. The only good thing about this is that it is starting to address the purchase information. If (LastName, FirstName) was unique, then customers would be limited to one purchase per day.

5. CUSTOMER (LastName, FirstName, Phone, Email, PurchaseDate, Item, Price, Tax, Total)

NOT GOOD. We're still trying to make the unworkable actually work. Same objections as above in 4, except that now customers would be limited to one of a particular item per day. For example, a customer could not purchase two antique chairs on the same day!

6. CUSTOMER (LastName, FirstName, Phone, Email)

and

SALE (PurchaseDate, Item, Price, Tax, Total)

NOT GOOD. Finally the customer and purchase data is effectively broken up, but there is no foreign key to link the two tables. In CUSTOMER, using (LastName, FirstName) is still a problem. In SALE, with PurchaseDate as the primary key there can only be one purchase per day!

7. CUSTOMER (LastName, FirstName, Phone, Email, PurchaseDate)

and

SALE (PurchaseDate, Item, Price, Tax, Total)

NOT GOOD. We've got a foreign key of PurchaseDate in SALE. But everything else that was wrong in design 6 above is still a problem. And by using PurchaseDate as the foreign key, we limit the customer to only one purchase!

8. CUSTOMER (LastName, FirstName, Phone, Email)

and

SALE (PurchaseDate, Item, LastName, FirstName, Price, Tax, Total)

BETTER, BUT STILL NOT GOOD. We've got the foreign key in the proper place, but the customer is limited to buying only one of a particular item per day. Everything else that was wrong in design 6 above is still a problem. And by using PurchaseDate as the foreign key, we limit the customer to only one purchase! For example, a customer could not purchase two antique chairs on the same day!

However, this is the best design of the bunch.

- C. *Modify what you consider to be the best design in question B to include surrogate ID columns called CustomerID and SaleID. How does this improve the design?*

The best design in Question B was number 8, so we'll put in the ID columns. These columns will become the new primary keys, and we'll need to adjust the foreign key in SALE. The result is:

**CUSTOMER (CustomerID, LastName, FirstName, Phone, Email)**

**SALE (SaleID, CustomerID, PurchaseDate, Item, Price, Tax, Total)**

We now have a clean design, and CUSTOMER is in BCNF. SALE is not in BCNF since

**Price → (Tax, Total)**

We could further normalize SALE, but we will intentionally leave it this way. This is called denormalization, and is discussed in Chapter Five. The point is that creating the extra table (PRICE\_TAX\_TOTAL) is more trouble than its worth—can you imagine what the data for that table would look like?

The primary key problems with both tables are resolved, and now a customer can purchase as many of an item on the same date as he or she wants to!

- D. *Modify the design in question C by breaking SALE into two relations named SALE and SALE\_ITEM. Modify columns and add additional columns as you think necessary. How does this improve the design?*

The main problem with the design in Question C is that only one item can be included in each sale. Moving items into a SALE\_ITEM table linked SALE will allow multiple items to be purchased as part of one sale. We'll need to include SaleID as part of a composite primary key so that the sale items are grouped according to their corresponding SALE. SaleID will also be the foreign key linking to SALE. Item and Price now belong in SALE\_ITEM, and we'll need to add a PreTaxTotal to SALE—tax will now only be calculated on the pretax total value of the sale. The result is:

**CUSTOMER** (CustomerID, LastName, FirstName, Phone, Email)

**SALE** (SaleID, CustomerID, PurchaseDate, PreTaxTotal, Tax, Total)

**SALE\_ITEM** (SaleID, SaleItemID, Item, Price)

We now have an improved clean design, and the CUSTOMER and SALE\_ITEM tables are in BCNF. SALE is still denormalized as discussed in Question C.

We have good primary and foreign keys, and now a customer can purchase as many of an item on the same date as he or she wants to and all items can be part of just one sale!

- E. *Given your assumptions, comment on the appropriateness of the following designs.*

1. PURCHASE (Item, PurchasePrice, PurchaseDate, Vendor, Phone)

NOT GOOD. There may be many purchases of the same item ("Candles").

2. PURCHASE (Item, PurchasePrice, PurchaseDate, Vendor, Phone)

NOT GOOD. There may be many purchases of the same item that cost the same amount of money ("Candles, \$30.00").

3. PURCHASE (Item, PurchasePrice, PurchaseDate, Vendor, Phone)

NOT GOOD. This limits purchases of a particular item to one per day.

4. PURCHASE (Item, PurchasePrice, PurchaseDate, Vendor, Phone)

NOT GOOD. This limits purchases of a particular item to one per vendor.

5. PURCHASE (Item, PurchasePrice, PurchaseDate)

and

VENDOR (Vendor, Phone)

NOT GOOD. It does, however separate the two themes of PURCHASE and VENDOR. However, there is no foreign key to link the tables. And this design still limits purchases of a particular item to one per day.



## Chapter Two — The Relational Model

---

6. PURCHASE (Item, PurchasePrice, PurchaseDate, Vendor)

and

VENDOR (Vendor, Phone)

BETTER, BUT STILL NOT GOOD. It separates the two themes of PURCHASE and VENDOR, which are now properly linked by a foreign key. However, this design still limits purchases of a particular item to one per day.

That said, this is the best design of the bunch.

F. *Modify what you consider to be the best design in question E to include surrogate ID columns called PurchaseID and VendorID. How does this improve the design?*

The best design in Question E was number 6, so we'll put in the ID columns. These columns will become the new primary keys, and we'll need to adjust the foreign key in PURCHASE. The result is:

**PURCHASE (PurchaseID, Item, PurchasePrice, PurchaseDate, VendorID)**

**VENDOR (VendorID, Vendor, Phone)**

We now have a clean design, and PURCHASE is in BCNF. VENDOR is not in BCNF since

**Vendor → (VendorID, Phone)**

**Phone → (VendorID, Vendor)**

We could further normalize VENDOR, but we will intentionally leave it this way. As discussed in Question D, this is called denormalization and is discussed in Chapter Five. The point is that creating the extra table (VENDOR\_PHONE) is more trouble than its worth.

The primary key problems with both tables are resolved, and now The Queen Anne Curiosity Shop can purchase as many of an item on the same date as needed.

G. *The relations in your design from question D and question F are not connected. Modify the database design so that sales data and purchase data are related.*

The connection between the two parts of the database design is the item being first purchased and then sold. Thus we can create an integrated design by replacing Item in SALE\_ITEM with PurchaseID as a foreign key. We will rename Price in SALE\_ITEM as SalePrice. Our final design will be:

**CUSTOMER (CustomerID, LastName, FirstName, Phone, Email)**

**SALE (SaleID, CustomerID, PurchaseDate, PreTaxTotal, Tax, Total)**

**SALE\_ITEM (SaleID, SaleItemID, PurchaseID, SalePrice)**

**PURCHASE (PurchaseID, Item, PurchasePrice, PurchaseDate, VendorID)**

**VENDOR (VendorID, Vendor, Phone)**