Concepts of Database Management
Seventh Edition

Chapter 5

Database Design 1: Normalization

Introduction

- **Normalization process**
  - Identifying potential problems, called **update anomalies**, in the design of a relational database
  - Methods for correcting these problems
- **Normal form**: table has desirable properties
  - First normal form (1NF)
  - Second normal form (2NF)
  - Third normal form (3NF)
  - Fourth normal form (4NF)

Introduction (continued)

- Normalization
  - Table in first normal form better than table not in first normal form
  - Table in second normal form better than table in first normal form, and so on
  - Goal: new collection of tables that is free of update anomalies

Functional Dependence

- Column B is **functionally dependent** on column A
  - Each value for A is associated with exactly one value of B

    $A \rightarrow B$
  - A **functionally determines** B

Functional Dependence (continued)

Rep

RepNum	LastName	FirstName	Street	City	State	Zip	Commission	PayClass	Rate
20	Kaiser	Valerie	624 Randall	Grove	FL	33321	$20,542.50	1	0.05
35	Hull	Richard	532 Jackson	Sheldon	FL	33553	$39,216.00	2	0.07
65	Perez	Juan	1626 Taylor	Fillmore	FL	33336	$23,487.00	1	0.05

FIGURE 5-2: Rep table with additional column, PayClass

Functional Dependence (continued)

Rep

RepNum	LastName	FirstName	Street	City	State	Zip	Commission	Rate
20	Kaiser	Valerie	624 Randall	Grove	FL	33321	$20,542.50	0.05
35	Hull	Richard	532 Jackson	Sheldon	FL	33553	$39,216.00	0.07
65	Perez	Juan	1626 Taylor	Fillmore	FL	33336	$23,487.00	0.05

FIGURE 5-3: Rep table

Rep

RepNum	LastName	FirstName	Street	City	State	Zip	Commission	Rate
20	Kaiser	Valerie	624 Randall	Grove	FL	33321	$20,542.50	0.05
35	Hull	Richard	532 Jackson	Sheldon	FL	33553	$39,216.00	0.07
65	Perez	Juan	1626 Taylor	Fillmore	FL	33336	$23,487.00	0.05
85	Kaiser	Wil	172 Bahia	Norton	FL	39281	$0.00	0.05

FIGURE 5-4: Rep table with second rep named Kaiser added

Keys

- Column A (or a collection of columns) is the **primary key** for a relation R
  - Property 1: *all* columns in R are functionally dependent on A
  - Property 2: no subcollection of columns in A also have Property 1
- **Candidate key**: column(s) on which all columns in table are functionally dependent
- **Alternate keys**: candidate keys not chosen as primary key

First Normal Form

- **Repeating group**: multiple entries for a single record
- **Unnormalized relation**: contains a repeating group
- Table (relation) in **first normal form (1NF)** does not contain repeating groups

First Normal Form (continued)

Orders (<u>OrderNum</u>, OrderDate, (PartNum, NumOrdered))

Orders

OrderNum	OrderDate	PartNum	NumOrdered
21608	10/20/2013	AT94	11
21610	10/20/2013	DR93 DW11	1 1
21613	10/21/2013	KL62	4
21614	10/21/2013	KT03	2
21617	10/23/2013	BV06 CD52	2 4
21619	10/23/2013	DR93	1
21623	10/23/2013	KV29	2

FIGURE 5-5: Sample unnormalized table

First Normal Form (continued)

Orders (<u>OrderNum</u>, OrderDate, <u>PartNum</u>, NumOrdered)

Orders

OrderNum	OrderDate	PartNum	NumOrdered
21608	10/20/2013	AT94	11
21610	10/20/2013	DR93	1
21610	10/20/2013	DW11	1
21613	10/20/2013	KL62	4
21614	10/20/2013	KT03	2
21617	10/20/2013	BV06	2
21617	10/20/2013	CD52	4
21619	10/20/2013	DR93	1
21623	10/20/2013	KV29	2

FIGURE 5-6: Result of normalization (conversion to first normal form)

Second Normal Form

Orders

OrderNum	OrderDate	PartNum	Description	NumOrdered	QuotedPrice
21608	10/20/2013	AT94	Iron	11	$21.95
21610	10/20/2013	DR93	Gas Range	1	$495.00
21610	10/20/2013	DW11	Washer	1	$399.99
21613	10/21/2013	KL62	Dryer	4	$329.95
21614	10/21/2013	KT03	Dishwasher	2	$595.00
21617	10/23/2013	BV06	Home Gym	2	$794.95
21617	10/23/2013	CD52	Microwave Oven	4	$150.00
21619	10/23/2013	DR93	Gas Range	1	$495.00
21623	10/23/2013	KV29	Treadmill	2	$1290.00

FIGURE 5-7: Sample Orders table

Second Normal Form (continued)

Orders (<u>OrderNum</u>, OrderDate, <u>PartNum</u>,
Description, NumOrdered, QuotedPrice)

- Functional dependencies:

  OrderNum → OrderDate

  PartNum → Description

  OrderNum, PartNum → NumOrdered, QuotedPrice,
  OrderDate, Description

Second Normal Form (continued)

- Update anomalies
  - Update
  - Inconsistent data
  - Additions
  - Deletions
- **Nonkey column** (**nonkey attribute**): not part of primary key

Second Normal Form (continued)

- Table (relation) in **second normal form (2NF)**
  - Table is in first normal form
  - No **nonkey column** is dependent on only a portion of primary key
- **Dependency diagram**: arrows indicate all functional dependencies
  - Arrows above boxes: normal dependencies
  - Arrows below boxes: partial dependencies
- **Partial dependencies**: only on a portion of the primary key

Second Normal Form (continued)

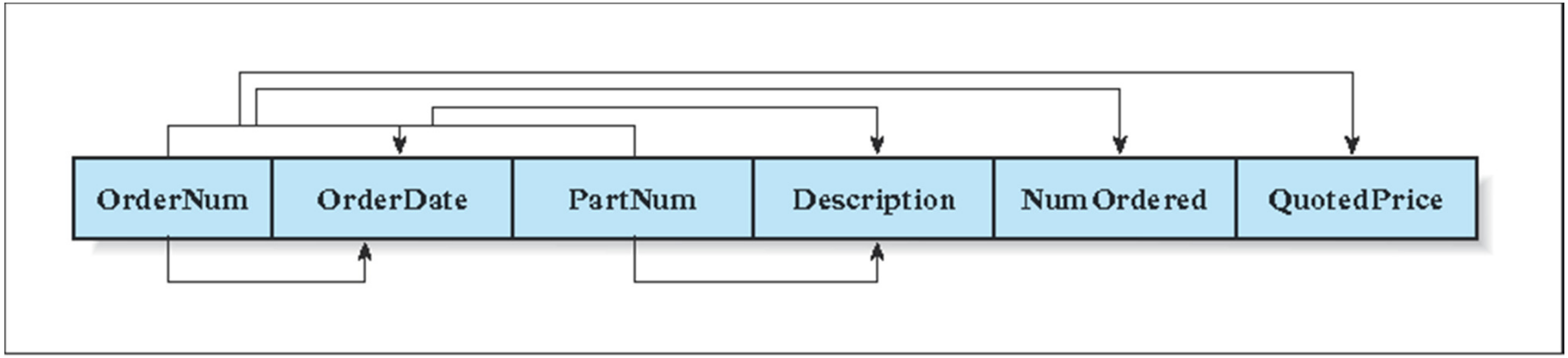


FIGURE 5-8: Dependences in the Orders table

Second Normal Form (continued)

Orders

OrderNum	OrderDate	PartNum	Description	NumOrdered	QuotedPrice
21608	10/20/2013	AT94	Iron	11	$21.95
21610	10/20/2013	DR93	Gas Range	1	$495.00
21610	10/20/2013	DW11	Washer	1	$399.99
21613	10/21/2013	KL62	Dryer	4	$329.95
21614	10/21/2013	KT03	Dishwasher	2	$595.00
21617	10/23/2013	BV06	Home Gym	2	$794.95
21617	10/23/2013	CD52	Microwave Oven	4	$150.00
21619	10/23/2013	DR93	Gas Range	1	$495.00
21623	10/23/2013	KV29	Treadmill	2	$1290.00

Orders

OrderNum	OrderDate
21608	10/20/2013
21610	10/20/2013
21613	10/21/2013
21614	10/21/2013
21617	10/23/2013
21619	10/23/2013
21623	10/23/2013

Part

PartNum	Description
AT94	Iron
BV06	Home Gym
CD52	Microwave Oven
DL71	Cordless Drill
DR93	Gas Range
DW11	Washer
FD21	Stand Mixer
KL62	Dryer
KT03	Dishwasher
KV29	Treadmill

OrderLine

OrderNum	PartNum	NumOrdered	QuotedPrice
21608	AT94	11	$21.95
21610	DR93	1	$495.00
21610	DW11	1	$399.99
21613	KL62	4	$329.95
21614	KT03	2	$595.00
21617	BV06	2	$794.95
21617	CD52	4	$150.00
21619	DR93	1	$495.00
21623	KV29	2	$1290.00

FIGURE 5-9: Conversion to second normal form

Third Normal Form

- Customer (CustomerNum, CustomerName, Balance, CreditLimit, RepNum, LastName, FirstName)

- Functional dependencies:
  - CustomerNum → CustomerName, Balance, CreditLimit, RepNum, LastName, FirstName
  - RepNum → LastName, FirstName

Third Normal Form (continued)

Customer

CustomerNum	CustomerName	Balance	CreditLimit	RepNum	LastName	FirstName
148	Al's Appliance and Sport	$6,550.00	$7,500.00	20	Kaiser	Valerie
282	Brookings Direct	$431.50	$10,000.00	35	Hull	Richard
356	Ferguson's	$5,785.00	$7,500.00	65	Perez	Juan
408	The Everything Shop	$5,285.25	$5,000.00	35	Hull	Richard
462	Bargains Galore	$3,412.00	$10,000.00	65	Perez	Juan
524	Kline's	$12,762.00	$15,000.00	20	Kaiser	Valerie
608	Johnson's Department Store	$2,106.00	$10,000.00	65	Perez	Juan
687	Lee's Sport and Appliance	$2,851.00	$5,000.00	35	Hull	Richard
725	Deerfield's Four Seasons	$248.00	$7,500.00	35	Hull	Richard
842	All Season	$8,221.00	$7,500.00	20	Kaiser	Valerie

FIGURE 5-10: Sample Customer table

Third Normal Form (continued)

- 2NF tables may still contain problems
  - Redundancy and wasted space
  - Update anomalies
    - Update
    - Inconsistent data
    - Additions
    - Deletions
- **Determinant**: column(s) that determines another column

Third Normal Form (continued)

- Table (relation) in **third normal form (3NF)**
  - It is in second normal form
  - Its only determinants are candidate keys

Third Normal Form (continued)

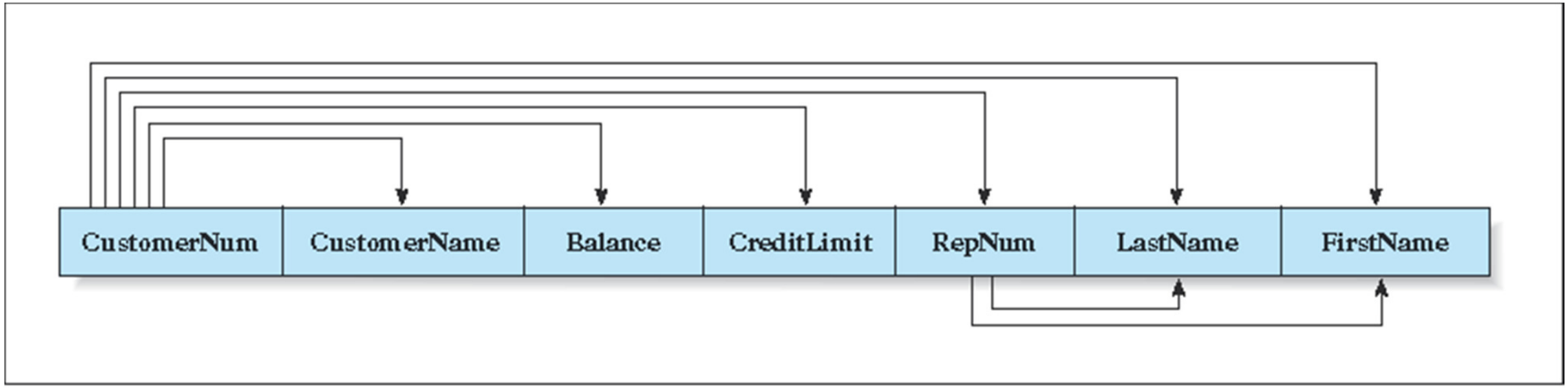


FIGURE 5-11: Dependencies in the Customer table

Third Normal Form (continued)

- Correction procedure
  - For each determinant that is not a candidate key, remove from table the columns that depend on this determinant
  - Create new table containing all columns from the original table that depend on this determinant
  - Make determinant the primary key of new table

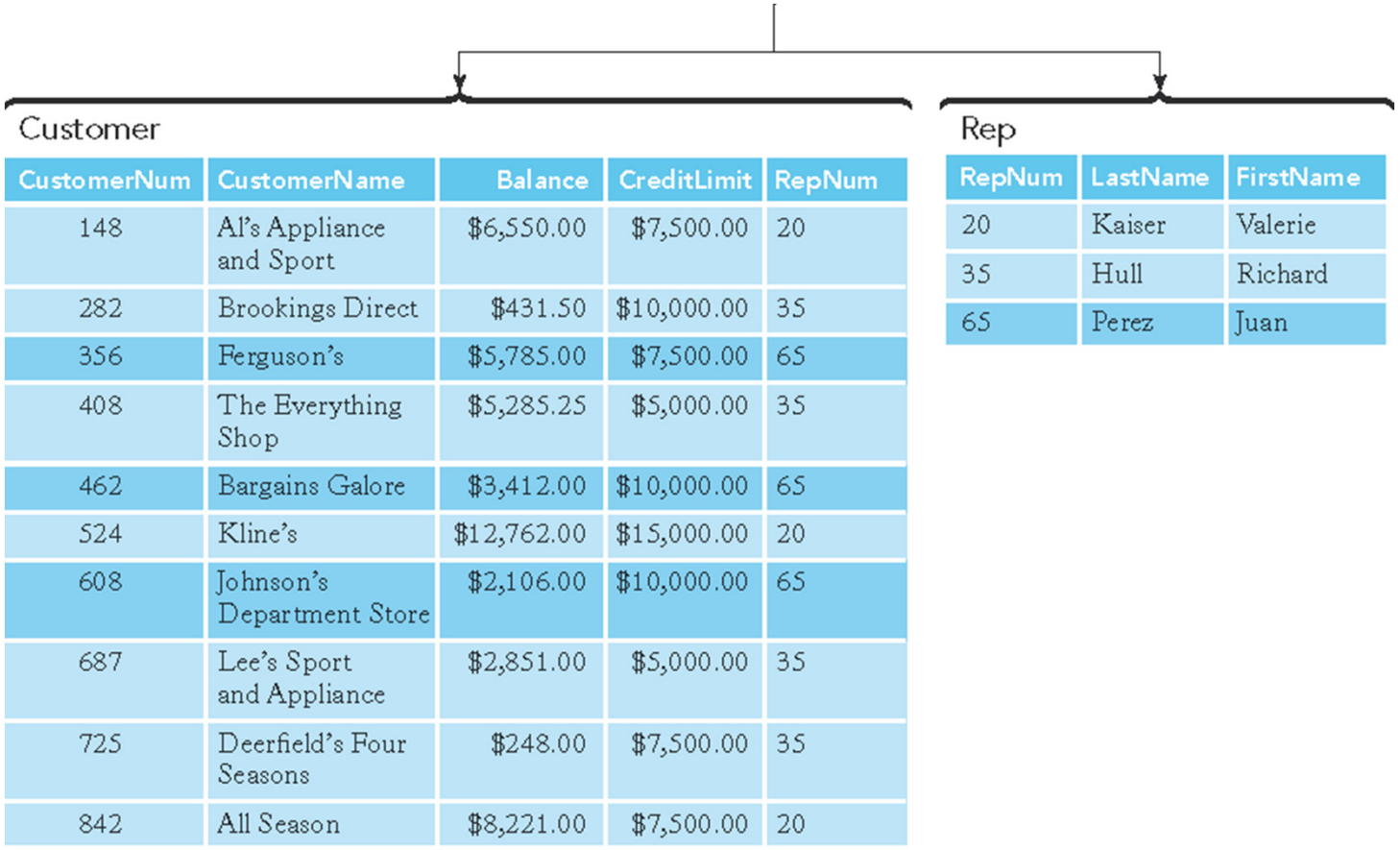
Third Normal Form (continued)

Customer

CustomerNum	CustomerName	Balance	CreditLimit	RepNum	LastName	FirstName
148	Al's Appliance and Sport	$6,550.00	$7,500.00	20	Kaiser	Valerie
282	Brookings Direct	$431.50	$10,000.00	35	Hull	Richard
356	Ferguson's	$5,785.00	$7,500.00	65	Perez	Juan
408	The Everything Shop	$5,285.25	$5,000.00	35	Hull	Richard
462	Bargains Galore	$3,412.00	$10,000.00	65	Perez	Juan
524	Kline's	$12,762.00	$15,000.00	20	Kaiser	Valerie
608	Johnson's Department Store	$2,106.00	$10,000.00	65	Perez	Juan
687	Lee's Sport and Appliance	$2,851.00	$5,000.00	35	Hull	Richard
725	Deerfield's Four Seasons	$248.00	$7,500.00	35	Hull	Richard
842	All Season	$8,221.00	$7,500.00	20	Kaiser	Valerie

FIGURE 5-12: Conversion to third normal form

Third Normal Form (continued)



Customer

CustomerNum	CustomerName	Balance	CreditLimit	RepNum
148	Al's Appliance and Sport	$6,550.00	$7,500.00	20
282	Brookings Direct	$431.50	$10,000.00	35
356	Ferguson's	$5,785.00	$7,500.00	65
408	The Everything Shop	$5,285.25	$5,000.00	35
462	Bargains Galore	$3,412.00	$10,000.00	65
524	Kline's	$12,762.00	$15,000.00	20
608	Johnson's Department Store	$2,106.00	$10,000.00	65
687	Lee's Sport and Appliance	$2,851.00	$5,000.00	35
725	Deerfield's Four Seasons	$248.00	$7,500.00	35
842	All Season	$8,221.00	$7,500.00	20

Rep

RepNum	LastName	FirstName
20	Kaiser	Valerie
35	Hull	Richard
65	Perez	Juan

FIGURE 5-12: Conversion to third normal form (continued)

Incorrect Decompositions

- Decomposition must be done using method described for 3NF
- Incorrect decompositions can lead to tables with the same problems as original table

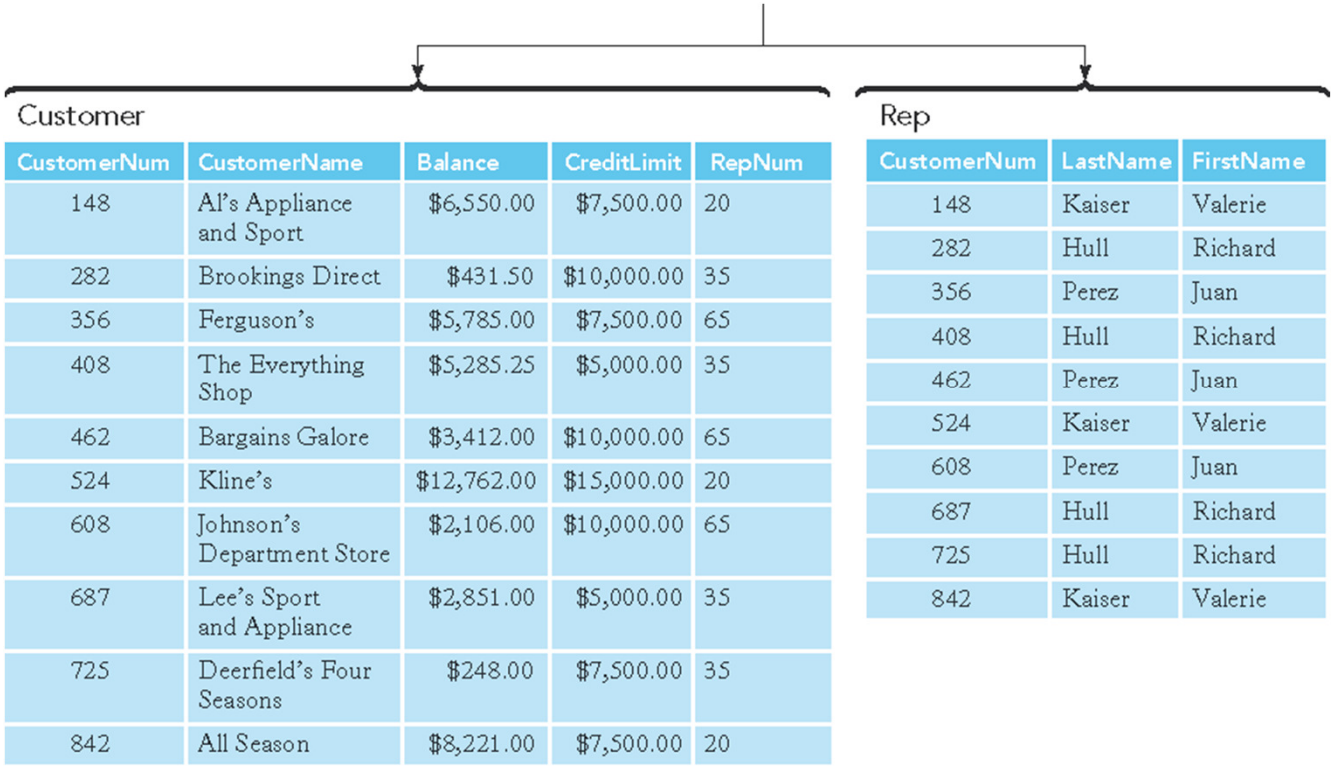
Incorrect Decompositions (continued)

Customer

CustomerNum	CustomerName	Balance	CreditLimit	RepNum	LastName	FirstName
148	Al's Appliance and Sport	$6,550.00	$7,500.00	20	Kaiser	Valerie
282	Brookings Direct	$431.50	$10,000.00	35	Hull	Richard
356	Ferguson's	$5,785.00	$7,500.00	65	Perez	Juan
408	The Everything Shop	$5,285.25	$5,000.00	35	Hull	Richard
462	Bargains Galore	$3,412.00	$10,000.00	65	Perez	Juan
524	Kline's	$12,762.00	$15,000.00	20	Kaiser	Valerie
608	Johnson's Department Store	$2,106.00	$10,000.00	65	Perez	Juan
687	Lee's Sport and Appliance	$2,851.00	$5,000.00	35	Hull	Richard
725	Deerfield's Four Seasons	$248.00	$7,500.00	35	Hull	Richard
842	All Season	$8,221.00	$7,500.00	20	Kaiser	Valerie

FIGURE 5-13: Incorrect decomposition of the Customer table

Incorrect Decompositions (continued)



Customer

CustomerNum	CustomerName	Balance	CreditLimit	RepNum
148	Al's Appliance and Sport	$6,550.00	$7,500.00	20
282	Brookings Direct	$431.50	$10,000.00	35
356	Ferguson's	$5,785.00	$7,500.00	65
408	The Everything Shop	$5,285.25	$5,000.00	35
462	Bargains Galore	$3,412.00	$10,000.00	65
524	Kline's	$12,762.00	$15,000.00	20
608	Johnson's Department Store	$2,106.00	$10,000.00	65
687	Lee's Sport and Appliance	$2,851.00	$5,000.00	35
725	Deerfield's Four Seasons	$248.00	$7,500.00	35
842	All Season	$8,221.00	$7,500.00	20

Rep

CustomerNum	LastName	FirstName
148	Kaiser	Valerie
282	Hull	Richard
356	Perez	Juan
408	Hull	Richard
462	Perez	Juan
524	Kaiser	Valerie
608	Perez	Juan
687	Hull	Richard
725	Hull	Richard
842	Kaiser	Valerie

FIGURE 5-13: Incorrect decomposition of the Customer table (continued)

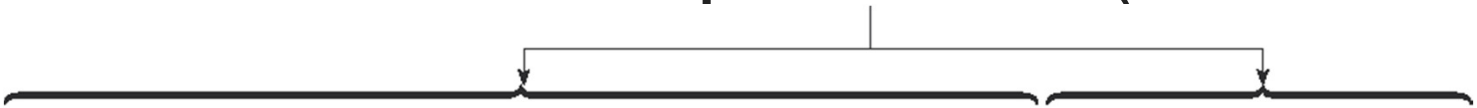
Incorrect Decompositions (continued)

Customer

CustomerNum	CustomerName	Balance	CreditLimit	RepNum	LastName	FirstName
148	Al's Appliance and Sport	$6,550.00	$7,500.00	20	Kaiser	Valerie
282	Brookings Direct	$431.50	$10,000.00	35	Hull	Richard
356	Ferguson's	$5,785.00	$7,500.00	65	Perez	Juan
408	The Everything Shop	$5,285.25	$5,000.00	35	Hull	Richard
462	Bargains Galore	$3,412.00	$10,000.00	65	Perez	Juan
524	Kline's	$12,762.00	$15,000.00	20	Kaiser	Valerie
608	Johnson's Department Store	$2,106.00	$10,000.00	65	Perez	Juan
687	Lee's Sport and Appliance	$2,851.00	$5,000.00	35	Hull	Richard
725	Deerfield's Four Seasons	$248.00	$7,500.00	35	Hull	Richard
842	All Season	$8,221.00	$7,500.00	20	Kaiser	Valerie

FIGURE 5-14: Second incorrect decomposition of the Customer table

Incorrect Decompositions (continued)



Customer

CustomerNum	CustomerName	Balance	CreditLimit	LastName	FirstName
148	Al's Appliance and Sport	$6,550.00	$7,500.00	Kaiser	Valerie
282	Brookings Direct	$431.50	$10,000.00	Hull	Richard
356	Ferguson's	$5,785.00	$7,500.00	Perez	Juan
408	The Everything Shop	$5,285.25	$5,000.00	Hull	Richard
462	Bargains Galore	$3,412.00	$10,000.00	Perez	Juan
524	Kline's	$12,762.00	$15,000.00	Kaiser	Valerie
608	Johnson's Department Store	$2,106.00	$10,000.00	Perez	Juan
687	Lee's Sport and Appliance	$2,851.00	$5,000.00	Hull	Richard
725	Deerfield's Four Seasons	$248.00	$7,500.00	Hull	Richard
842	All Season	$8,221.00	$7,500.00	Kaiser	Valerie

Rep

RepNum	LastName	FirstName
20	Kaiser	Valerie
35	Hull	Richard
65	Perez	Juan

FIGURE 5-14: Second incorrect decomposition of the Customer table (continued)

Multivalued Dependencies and Fourth Normal Form

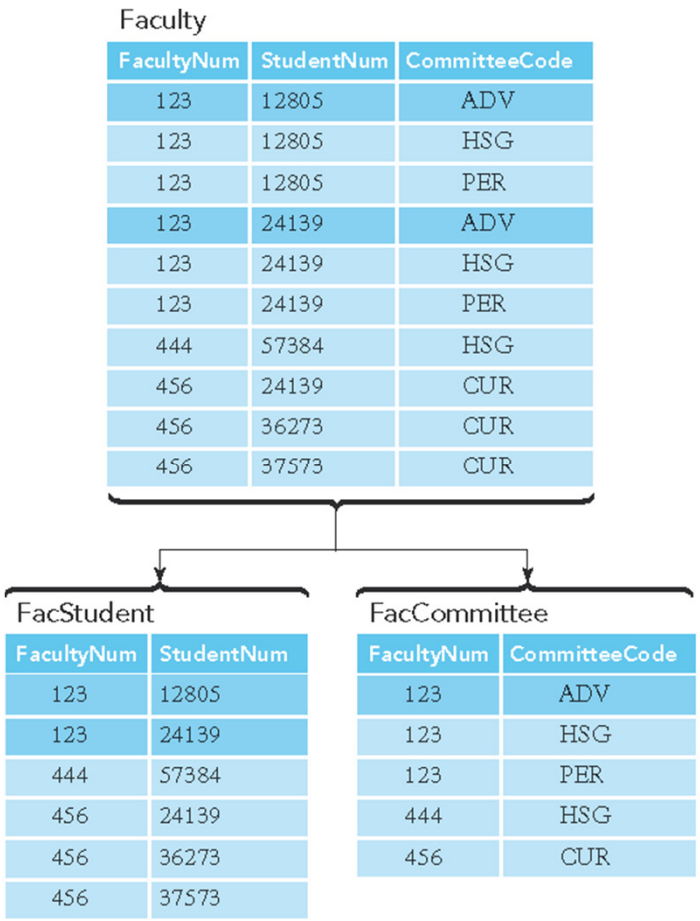
- 3NF tables may still contain problems
  - Updates
  - Additions
  - Deletions
- **Multivalued dependence** of column B on column A
  - "B is **multidependent** on A"
  - "A **multidetermines** B"
  - Each value for A is associated with a specific collection of values for B, and this collection is independent of any values for C

    $A \rightarrow\rightarrow B$

Multivalued Dependencies and Fourth Normal Form (continued)

- Table (relation) in **fourth normal form (4NF)**
  - It is in third normal form
  - No multivalued dependencies
- Converting table to fourth normal form
  - Split third normal form table into separate tables, each containing the column that multidetermines the others

Multivalued Dependencies and Fourth Normal Form (continued)



Faculty

FacultyNum	StudentNum	CommitteeCode
123	12805	ADV
123	12805	HSG
123	12805	PER
123	24139	ADV
123	24139	HSG
123	24139	PER
444	57384	HSG
456	24139	CUR
456	36273	CUR
456	37573	CUR

FacStudent

FacultyNum	StudentNum
123	12805
123	24139
444	57384
456	24139
456	36273
456	37573

FacCommittee

FacultyNum	CommitteeCode
123	ADV
123	HSG
123	PER
444	HSG
456	CUR

FIGURE 5-16: Conversion to fourth normal form

Multivalued Dependencies and Fourth Normal Form (continued)

Normal Form	Meaning/Required Conditions	Notes
First normal form	No repeating groups exist	
Second normal form	First normal form and no non-key column is dependent on only a portion of the primary key	Automatically second normal form if the primary key contains only a single column
Third normal form	Second normal form and the only determinants are candidate keys	Actually Boyce-Codd normal form (BCNF)
Fourth normal form	Third normal form and no multivalued dependencies exist	

FIGURE 5-17: Normal forms

Avoiding the Problem with Multivalued Dependencies

- Slightly more sophisticated method for converting unnormalized table to first normal form
  - Place each repeating group in separate table
  - Each table will contain all columns of a repeating group, and primary key of the original table
  - Primary key to each new table will be the concatenation of the primary keys of the original table and the repeating group

Application to Database Design

- Carefully convert tables to third normal form
- Review assumptions and dependencies periodically to see if changes to design are needed
- Splitting relations to achieve third normal form tables creates need for an interrelation constraint
  - **Interrelation constraint**: condition that involves two or more relations

Summary

- Column (attribute) B is functionally dependent on another column A (or collection of columns) when each value for A in the database is associated with exactly one value of B
- Column(s) A is the primary key if all other columns are functionally dependent on A and no subcollection of columns in A also have this property
- When there is more than one choice for primary key, one possibility is chosen to be *the* primary key; others called candidate keys

Summary (continued)

- Table (relation) in first normal form (1NF) does not contain repeating groups
- Nonkey column (or nonkey attribute) is not a part of the primary key
- Table (relation) is in the second normal form (2NF) when it is in 1NF and no nonkey column is dependent on only a portion of the primary key
- Determinant is a column that functionally determines another column

Summary (continued)

- Table (relation) is in third normal form (3NF) when it is in 2NF and its only determinants are candidate keys
- Collection of tables (relations) that is not in third normal form has inherent problems called update anomalies
- Table (relation) is in fourth normal form (4NF) when it is in 3NF and there are no multivalued dependencies