Database Design

- **Two Approaches (Techniques) to Database Design**
  - Data Modeling
    - A conceptual picture of your database
    - Drawn with a computer-based tool
  - Normalization
    - A method of cleaning up and organizing your data
    - Structures the data to fit the RELATIONAL MODEL
    - A set of rules you apply to the data to clean up the data

Database Design

- **Best Approach**
  - Combine the Two
    - We construct a data model
    - We normalize the data as we model
  - Iterative process
    - Model some, normalize some
    - Model some more, normalize some more
- **BUT, we must FIRST learn the rules of Normalization**

Normalization

- A structured, defined, detailed process
- Prepares data to make sure it complies with the rules of the RELATIONAL MODEL
- Often seen by students as a CONFUSING process
- Concepts are somewhat ACADEMIC and THEORETICAL

Objective of Normalization

- **To Arrange the data into a series of clearly defined, well-organized RELATIONS**
  - Each with a **primary key**
  - All attributes are **functionally dependent** on the primary key
  - Each with required **foreign keys** referencing other relations

Normalization

- **Normalization: follow the Rules of "ONE"**
  - One entity = one table
  - One occurrence of that entity = one row
  - One attribute = one column
  - All columns in the row describe one occurrence of that entity
  - All columns are functionally dependent on one primary key

Normalization

- **Normalization: Rules of “ONE” EXAMPLE**
  - Imagine a system used by a university’s Registrar’s office
  - We keep data about STUDENTS (entity)
  - For a student (one occurrence of the entity) we store FirstName, LastName, MiddleInitial, PhoneNumber, Address (attributes)
  - We identify a student by their StudentID (primary key)

Normalization

- **Entity = STUDENT**
- **Attributes = (StudentID, FirstName, LastName, MiddleInitial, PhoneNumber, Address)**

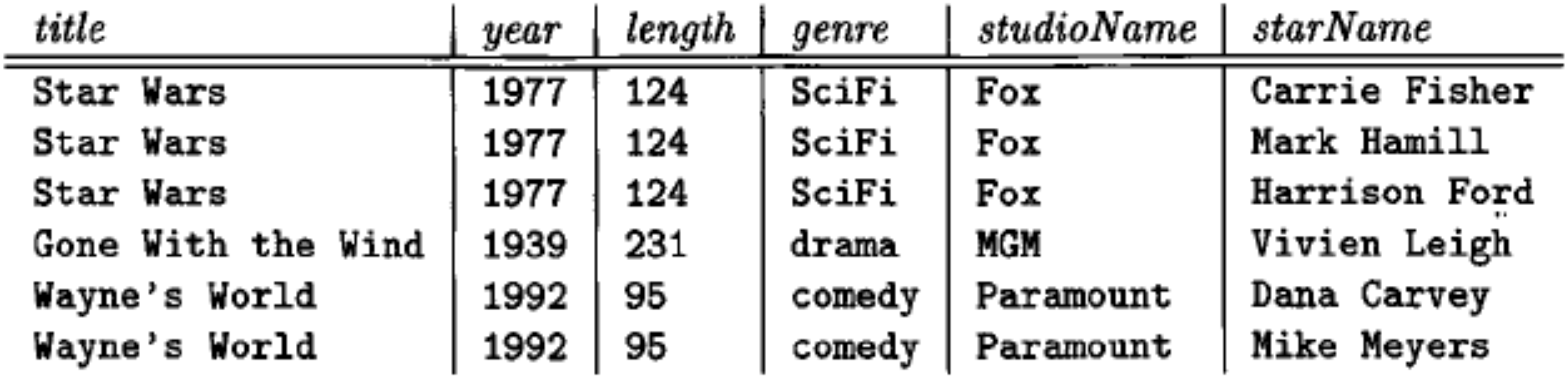
- **One entity, one primary key, one student contained in one row, all columns are functionally dependent on the one primary key**

Normalization

- **How do we achieve this?**
  - Key Analysis using Functional Dependencies
  - Identify and correct anomalies
  - Decomposition of a single table into multiple tables

Normalization

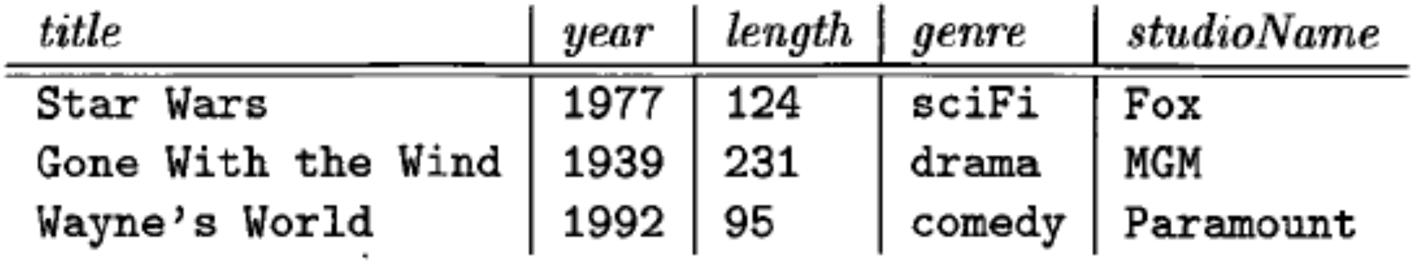
- **Anomalies**
  - Redundancy – storing a piece of data more than once
  - Update/Delete errors



title	*year*	*length*	*genre*	*studioName*	*starName*
Star Wars	1977	124	SciFi	Fox	Carrie Fisher
Star Wars	1977	124	SciFi	Fox	Mark Hamill
Star Wars	1977	124	SciFi	Fox	Harrison Ford
Gone With the Wind	1939	231	drama	MGM	Vivien Leigh
Wayne's World	1992	95	comedy	Paramount	Dana Carvey
Wayne's World	1992	95	comedy	Paramount	Mike Meyers

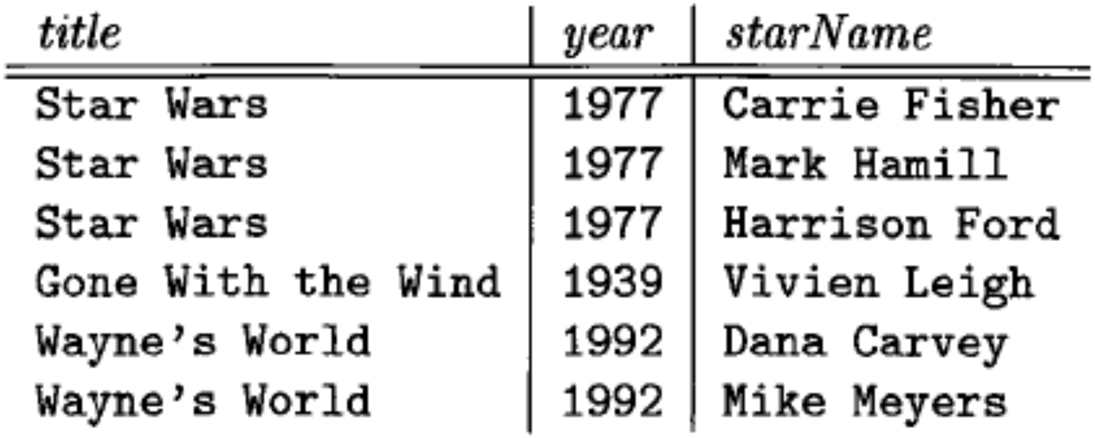
Normalization

- **Decomposing this relation into two removes the anomalies**



title	*year*	*length*	*genre*	*studioName*
Star Wars	1977	124	sciFi	Fox
Gone With the Wind	1939	231	drama	MGM
Wayne's World	1992	95	comedy	Paramount

(b) The relation Movies2.



title	*year*	*starName*
Star Wars	1977	Carrie Fisher
Star Wars	1977	Mark Hamill
Star Wars	1977	Harrison Ford
Gone With the Wind	1939	Vivien Leigh
Wayne's World	1992	Dana Carvey
Wayne's World	1992	Mike Meyers

(b) The relation Movies3.

Normalization

- **Normalization Step-by-Step**
  - First Normal Form
    - Remove any multi-valued cells and/or any rows requiring a specific sequence
  - Second Normal Form
    - For entities with concatenated keys, make sure that all attributes are dependent on the full key
  - Third Normal Form
    - Make sure that no attributes are dependent on any other non-key attributes
  - Fourth Normal Form
    - Eliminate multi-value dependencies

Normalization

- **First Normal Form**
  - Remove Repeating Groups of attributes
  - Put repeating attributes into a new relation
    - Create a key for the new relation
      - Probably a Concatenated key
      - The key of the Original relation + an identifier for each occurrence of the Repeating Group of attributes

Normalization

- **Second Normal Form**
  - Every non-key attribute is functionally dependent on the entire primary key
  - Only meaningful when there is a concatenated key
  - Create a new relation
    - Create a key for the new relation
  - Migrate only the dependent columns to the new relation

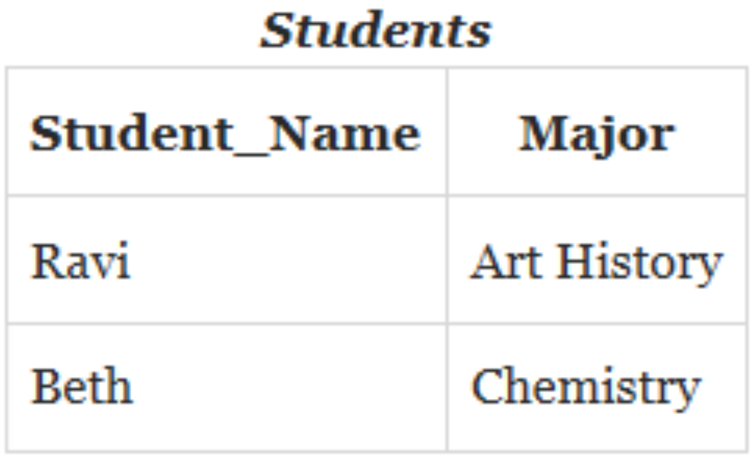
Normalization

- **Third Normal Form**
  - No non-key attributes are dependent on any other non-key attributes
  - Create a new relation
    - Create a key for the new relation
  - Migrate only the dependent columns to the new relation

- **Fourth Normal Form**
  - A multi-valued dependency is this:
    - Multiple independent attributes are dependent on the same determinant
    - Example:
      - One Customer may have multiple “contacts” of different “types”

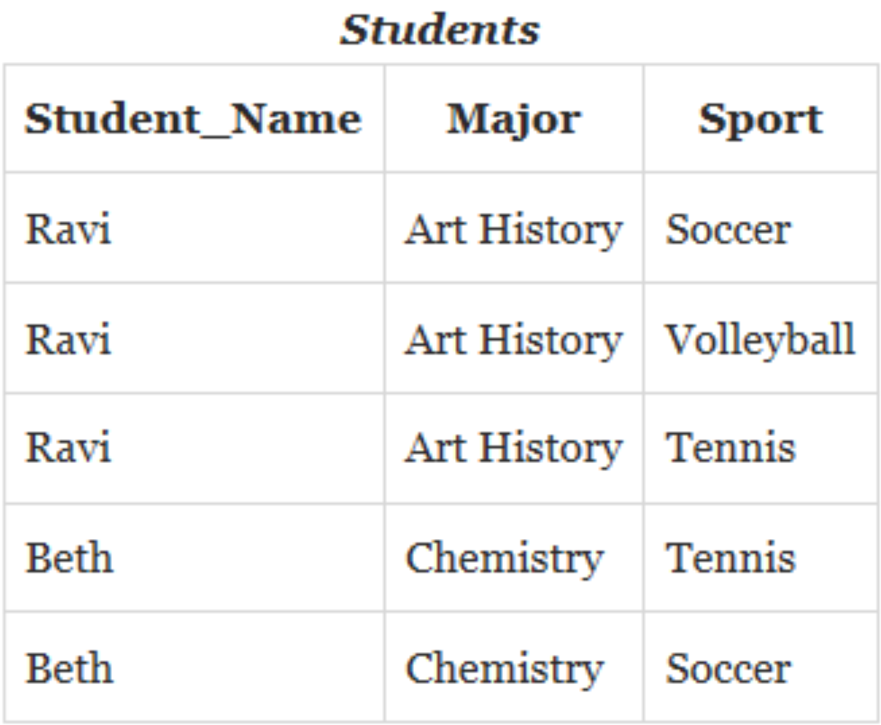
Normalization

- **Multivalued Dependency example**



Students

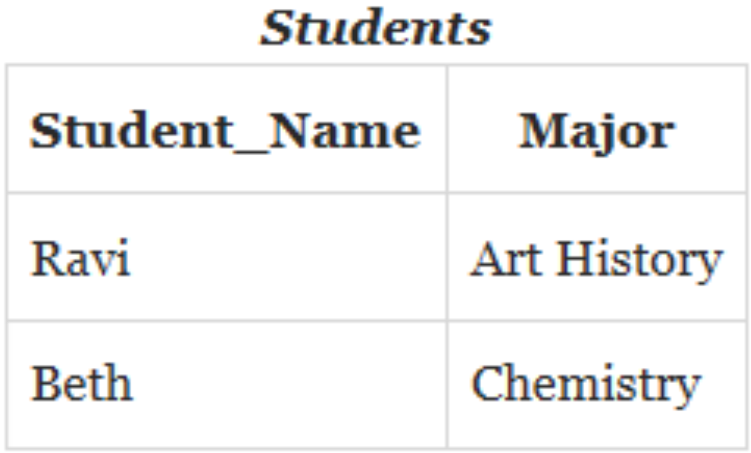
Student_Name	Major
Ravi	Art History
Beth	Chemistry



Students

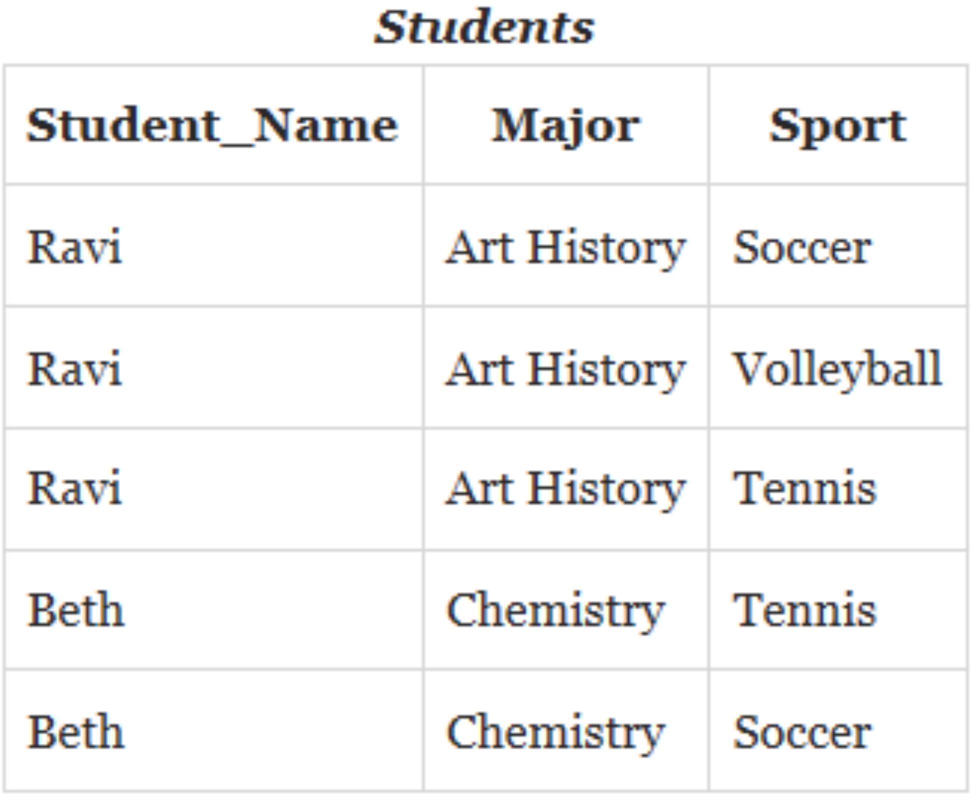
Student_Name	Major	Sport
Ravi	Art History	Soccer
Ravi	Art History	Volleyball
Ravi	Art History	Tennis
Beth	Chemistry	Tennis
Beth	Chemistry	Soccer

Normalization



Students

Student_Name	Major
Ravi	Art History
Beth	Chemistry



Students

Student_Name	Major	Sport
Ravi	Art History	Soccer
Ravi	Art History	Volleyball
Ravi	Art History	Tennis
Beth	Chemistry	Tennis
Beth	Chemistry	Soccer

- **Introduces anomalies**
- **Major and Sport both are FD on Student_Name, but are independent of each other.**

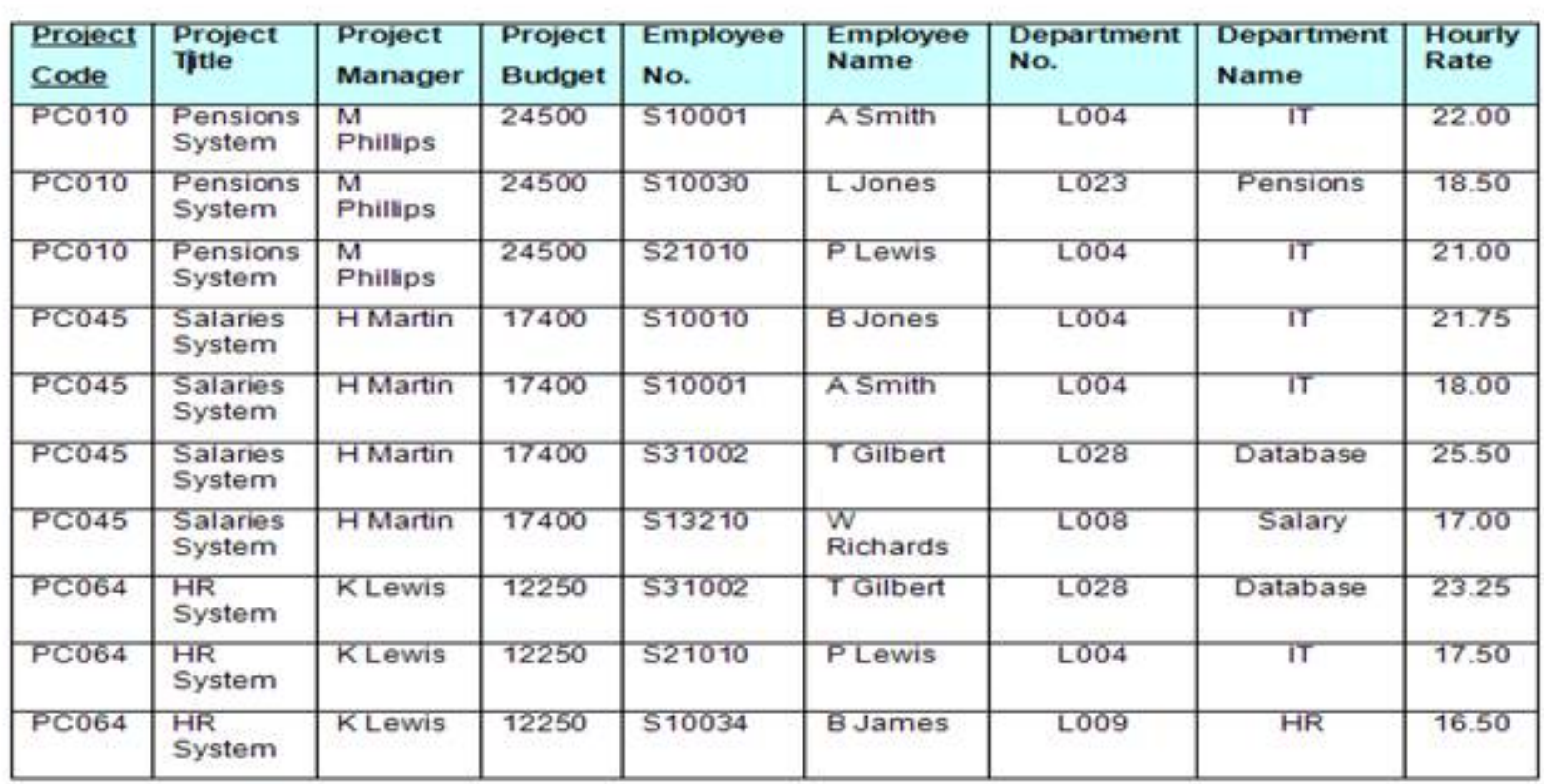
Normalization

Let’s work through some examples

- As you are designing a database, you must define your clients’ DATA REQUIREMENTS
- During interviews, Clients often refer you to their existing reports, spreadsheets, screens, etc.
- So, let’s think about FDs visible in this data
  - Think about Primary Keys
  - Determine which attributes are dependent on which primary keys

Functional Dependency

More practice



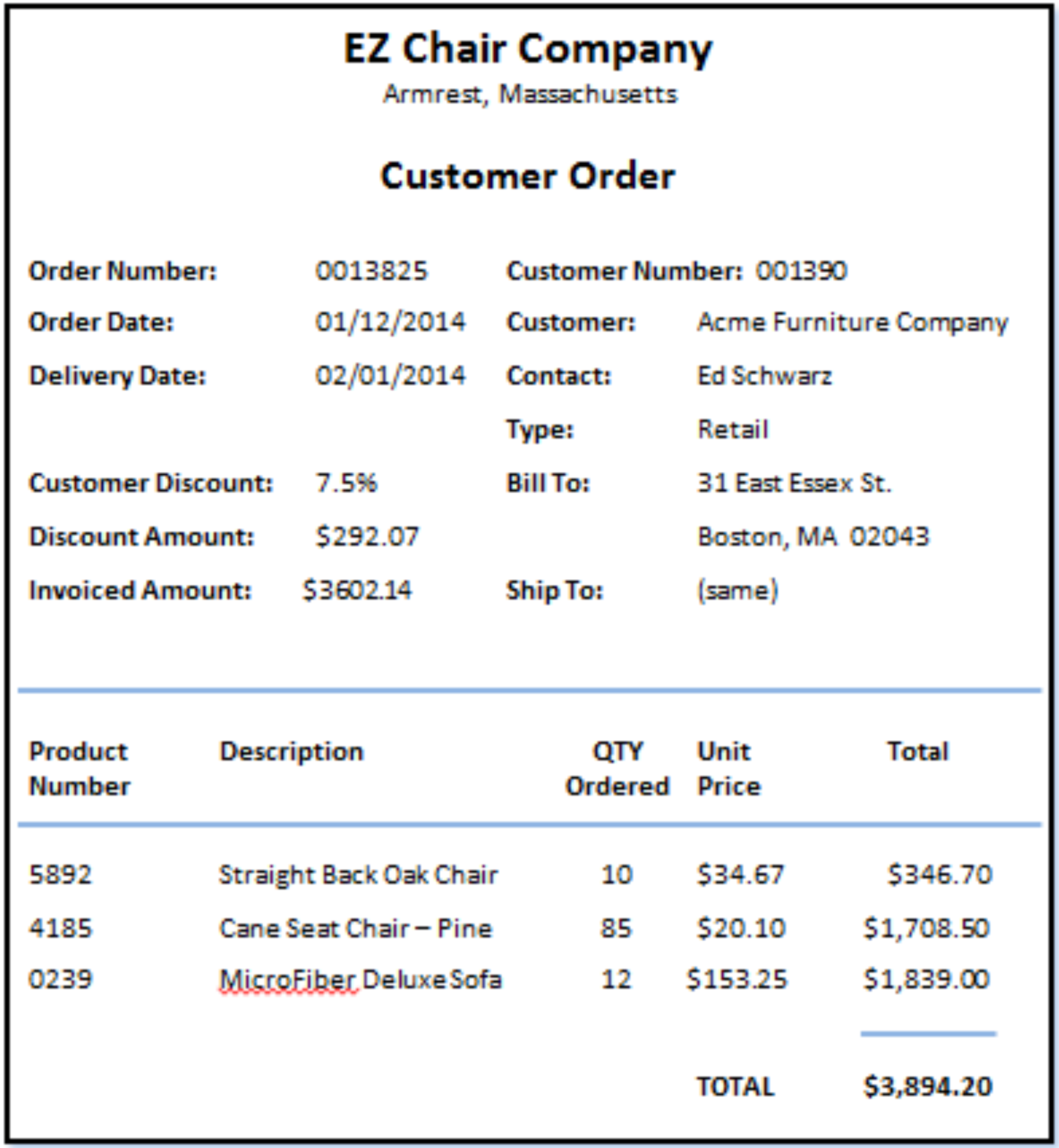
Project Code	Project Title	Project Manager	Project Budget	Employee No.	Employee Name	Department No.	Department Name	Hourly Rate
PC010	Pensions System	M Phillips	24500	S10001	A Smith	L004	IT	22.00
PC010	Pensions System	M Phillips	24500	S10030	L Jones	L023	Pensions	18.50
PC010	Pensions System	M Phillips	24500	S21010	P Lewis	L004	IT	21.00
PC045	Salaries System	H Martin	17400	S10010	B Jones	L004	IT	21.75
PC045	Salaries System	H Martin	17400	S10001	A Smith	L004	IT	18.00
PC045	Salaries System	H Martin	17400	S31002	T Gilbert	L028	Database	25.50
PC045	Salaries System	H Martin	17400	S13210	W Richards	L008	Salary	17.00
PC064	HR System	K Lewis	12250	S31002	T Gilbert	L028	Database	23.25
PC064	HR System	K Lewis	12250	S21010	P Lewis	L004	IT	17.50
PC064	HR System	K Lewis	12250	S10034	B James	L009	HR	16.50

Normalization

Another example:

- Imagine you are a consultant hired to design a database for a furniture manufacturer
- In defining requirements, you were given a sample customer invoice document
- Your task is to NORMALIZE this data

Normalization



EZ Chair Company

Armrest, Massachusetts

Customer Order

Order Number:	0013825	**Customer Number:**	001390
Order Date:	01/12/2014	**Customer:**	Acme Furniture Company
Delivery Date:	02/01/2014	**Contact:**	Ed Schwarz
		Type:	Retail
Customer Discount:	7.5%	**Bill To:**	31 East Essex St. Boston, MA 02043
Discount Amount:	$292.07		
Invoiced Amount:	$3602.14	**Ship To:**	(same)

Product Number	Description	QTY Ordered	Unit Price	Total
5892	Straight Back Oak Chair	10	$34.67	$346.70
4185	Cane Seat Chair – Pine	85	$20.10	$1,708.50
0239	MicroFiber Deluxe Sofa	12	$153.25	$1,839.00
			TOTAL	**$3,894.20**

Normalization

- **Steps:**
  - List all entities and attributes in UNNORMALIZED form
  - Then, put the data into First, Second, Third normal forms, following the rules:

Normalization

- **Normalization Step-by-Step**
  - First Normal Form
    - Break out any multi-valued/repeating values and/or any values requiring a specific sequence
  - Second Normal Form
    - Only for entities with concatenated keys: identify and break out any attributes dependent on a part of the concatenated key
  - Third Normal Form
    - Identify and break out any attributes that are dependent on any other non-key attributes

Un-Normalized

UNNORMALIZED

Customer Order

Order Number

Order Date

Delivery Date

Customer Discount

discount amount

invoiced amount

customer number

customer name

Contact

ContactType

bill to address

bill to city

bill to state

bill to zip

ship to address

ship to city

ship to state

ship to zip

Product Number

Description

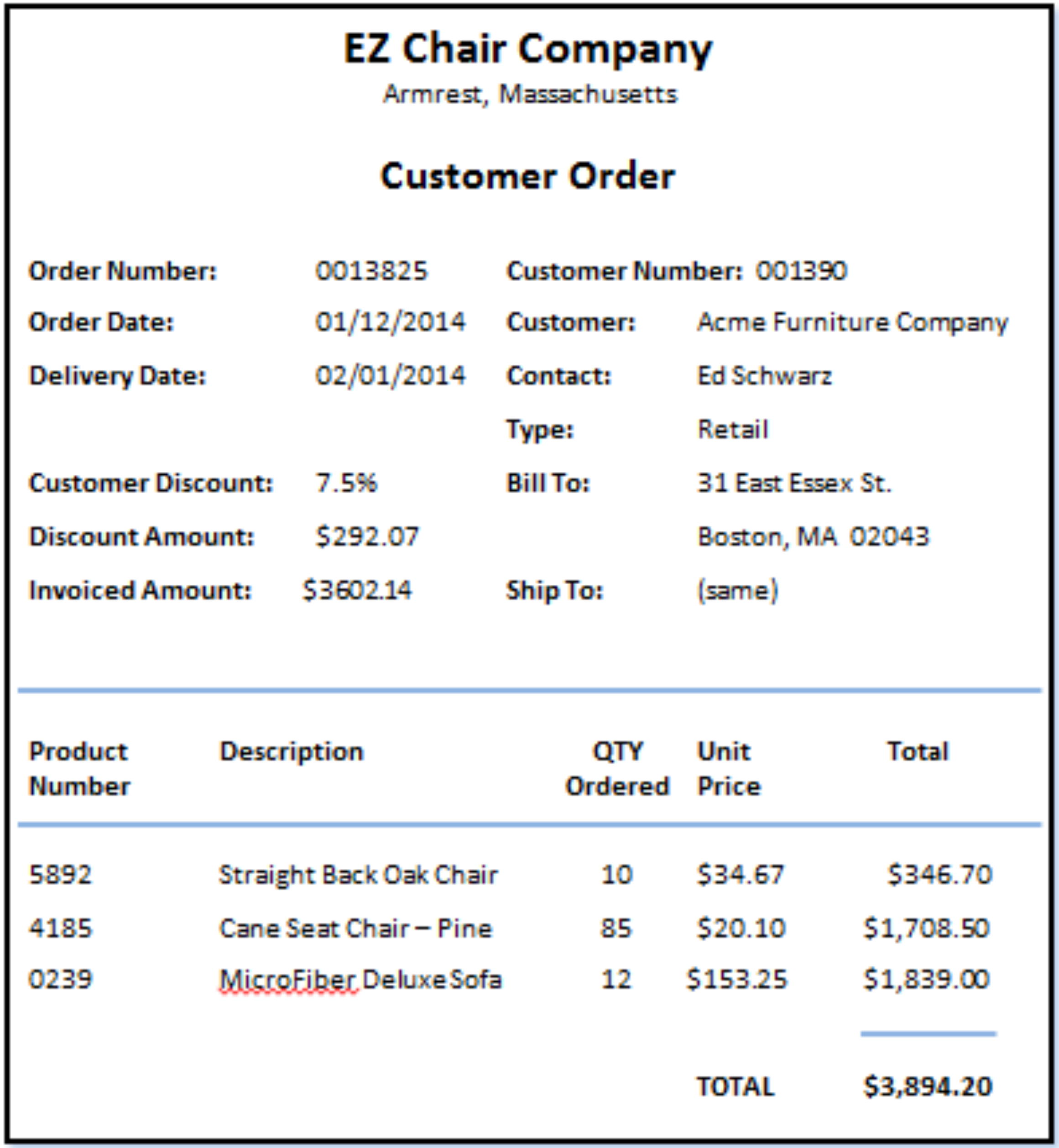
quantity ordered

unit price

order total

} Repeating/Multi-valued

Normalization



EZ Chair Company

Armrest, Massachusetts

Customer Order

Order Number:	0013825	Customer Number:	001390
Order Date:	01/12/2014	Customer:	Acme Furniture Company
Delivery Date:	02/01/2014	Contact:	Ed Schwarz
		Type:	Retail
Customer Discount:	7.5%	Bill To:	31 East Essex St.
Discount Amount:	$292.07		Boston, MA 02043
Invoiced Amount:	$3602.14	Ship To:	(same)

Product Number	Description	QTY Ordered	Unit Price	Total
5892	Straight Back Oak Chair	10	$34.67	$346.70
4185	Cane Seat Chair – Pine	85	$20.10	$1,708.50
0239	MicroFiber Deluxe Sofa	12	$153.25	$1,839.00
			TOTAL	$3,894.20

1st NF

UNNORMALIZED

Customer Order

Order Number

Order Date

Delivery Date

Customer Discount

discount amount

invoiced amount

customer number

customer name

Contact

ContactType

bill to address

bill to city

bill to state

bill to zip

ship to address

ship to city

ship to state

ship to zip

Product Number

Description

quantity ordered

unit price

order total

→

FIRST NORMAL FORM

Customer Order

Order Number

Order Date

Delivery Date

Customer Discount

discount amount

invoiced amount

customer number

customer name

Contact

ContactType

bill to address

bill to city

bill to state

bill to zip

ship to address

ship to city

ship to state

ship to zip

order total

OrderProduct

Order number

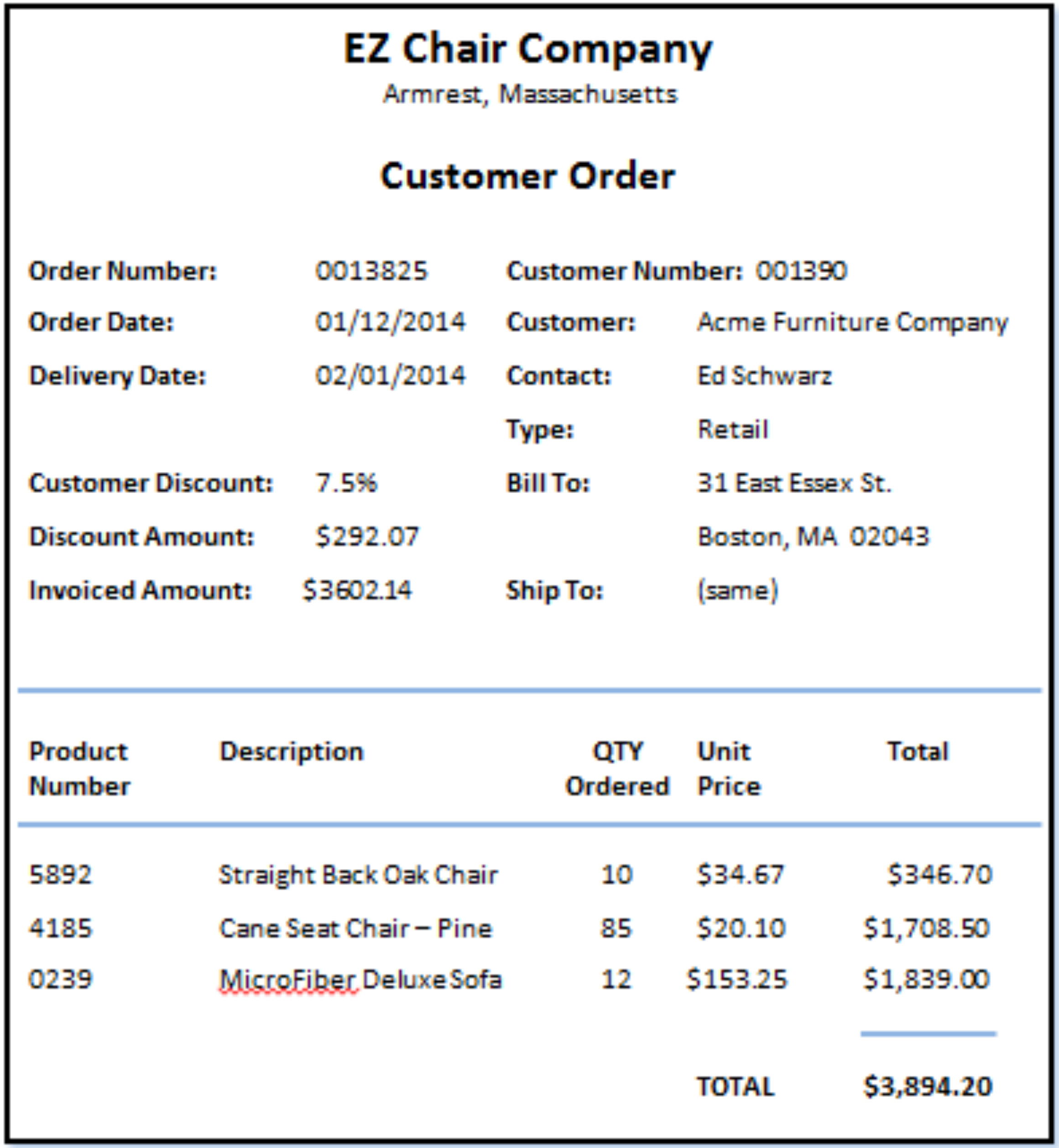
Product Number

Product Description

Quantity

unit price

Normalization



EZ Chair Company

Armrest, Massachusetts

Customer Order

Order Number:	0013825	Customer Number:	001390
Order Date:	01/12/2014	Customer:	Acme Furniture Company
Delivery Date:	02/01/2014	Contact:	Ed Schwarz
		Type:	Retail
Customer Discount:	7.5%	Bill To:	31 East Essex St.
Discount Amount:	$292.07		Boston, MA 02043
Invoiced Amount:	$3602.14	Ship To:	(same)

Product Number	Description	QTY Ordered	Unit Price	Total
5892	Straight Back Oak Chair	10	$34.67	$346.70
4185	Cane Seat Chair – Pine	85	$20.10	$1,708.50
0239	MicroFiber Deluxe Sofa	12	$153.25	$1,839.00
			TOTAL	$3,894.20

2nd NF

FIRST NORMAL FORM

Customer Order
Order Number
Order Date
Delivery Date
Customer Discount
discount amount
invoiced amount
customer number
customer name
Contact
ContactType
bill to address
bill to city
bill to state
bill to zip
ship to address
ship to city
ship to state
ship to zip
order total

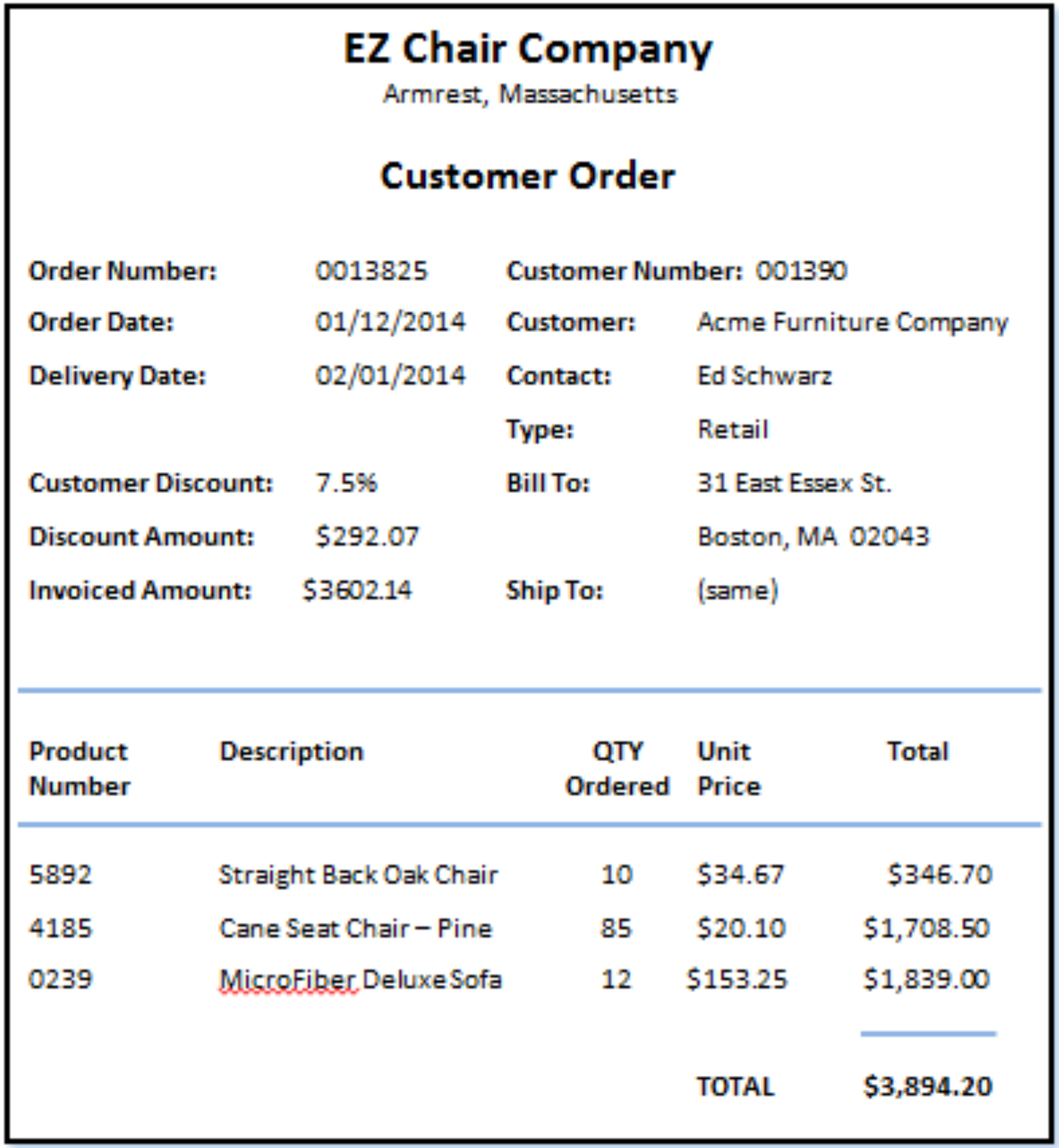
OrderProduct
Order number
Product Number
Product Description
Quantity
unit price

Customer Order
Order Number
Order Date
Delivery Date
Customer Discount
discount amount
invoiced amount
customer number
customer name
Contact
ContactType
bill to address
bill to city
bill to state
bill to zip
ship to address
ship to city
ship to state
ship to zip
order total

OrderProduct
Order number
Product Number
Quantity
unit price
total

Product
Product Number
Product Description

Normalization



EZ Chair Company

Armrest, Massachusetts

Customer Order

Order Number:	0013825	**Customer Number:**	001390
Order Date:	01/12/2014	**Customer:**	Acme Furniture Company
Delivery Date:	02/01/2014	**Contact:**	Ed Schwarz
		Type:	Retail
Customer Discount:	7.5%	**Bill To:**	31 East Essex St.
Discount Amount:	$292.07		Boston, MA 02043
Invoiced Amount:	$3602.14	**Ship To:**	(same)

Product Number	Description	QTY Ordered	Unit Price	Total
5892	Straight Back Oak Chair	10	$34.67	$346.70
4185	Cane Seat Chair – Pine	85	$20.10	$1,708.50
0239	MicroFiber Deluxe Sofa	12	$153.25	$1,839.00
			TOTAL	**$3,894.20**

SECOND NORMAL FORM

Customer Order

Order Number

Order Date
Delivery Date
Customer Discount
discount amount
invoiced amount
customer number
customer name
Contact
ContactType
bill to address
bill to city
bill to state
bill to zip
ship to address
ship to city
ship to state
ship to zip
order total

OrderProduct

Order number

Product Number

Quantity
unit price
total

Product

Product Number

Product Description

THIRD NORMAL FORM

3rd NF

Order

Order Number

Order Date
Delivery Date
discount amount
invoiced amount
customer number
order total

Customer

customer number

customer name
Contact
ContactType
bill to address
bill to city
bill to state
bill to zip
ship to address
ship to city
ship to state
ship to zip

OrderProduct

Order number

Product Number

Quantity
unit price
total

Product

Product Number

Product Description