IST659 Data Admin Concepts and Database Management

Normalization

week 4

Acknowledgements and caveat

These course materials draw liberally, with permission, from the following sources:

- IST659 classes taught by Prof. Bei Yu
- IST659 classes taught by Prof. Susan Dischiave
- IST 659 classes taught by Prof. Yun Huang
- Instructor resources provided by our text book ”Modern Database Management”

Caveat (beware): At best, PowerPoint slides are only a pale imitation of the entirety of a class meeting. In IST659 in particular, the lectures will cover topics beyond what appears in these slides. Don’t rely on them as a substitute for attending class.

Relational Data Model

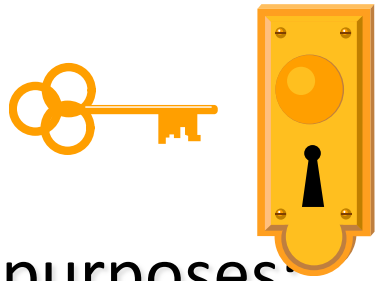
COMPONENTS OF RELATIONAL MODEL

- Data structure
  - Tables (relations), rows, columns
- Data integrity
  - Mechanisms for implementing business rules that maintain integrity of manipulated data
- Data manipulation
  - Powerful SQL operations for retrieving and modifying data

Relation

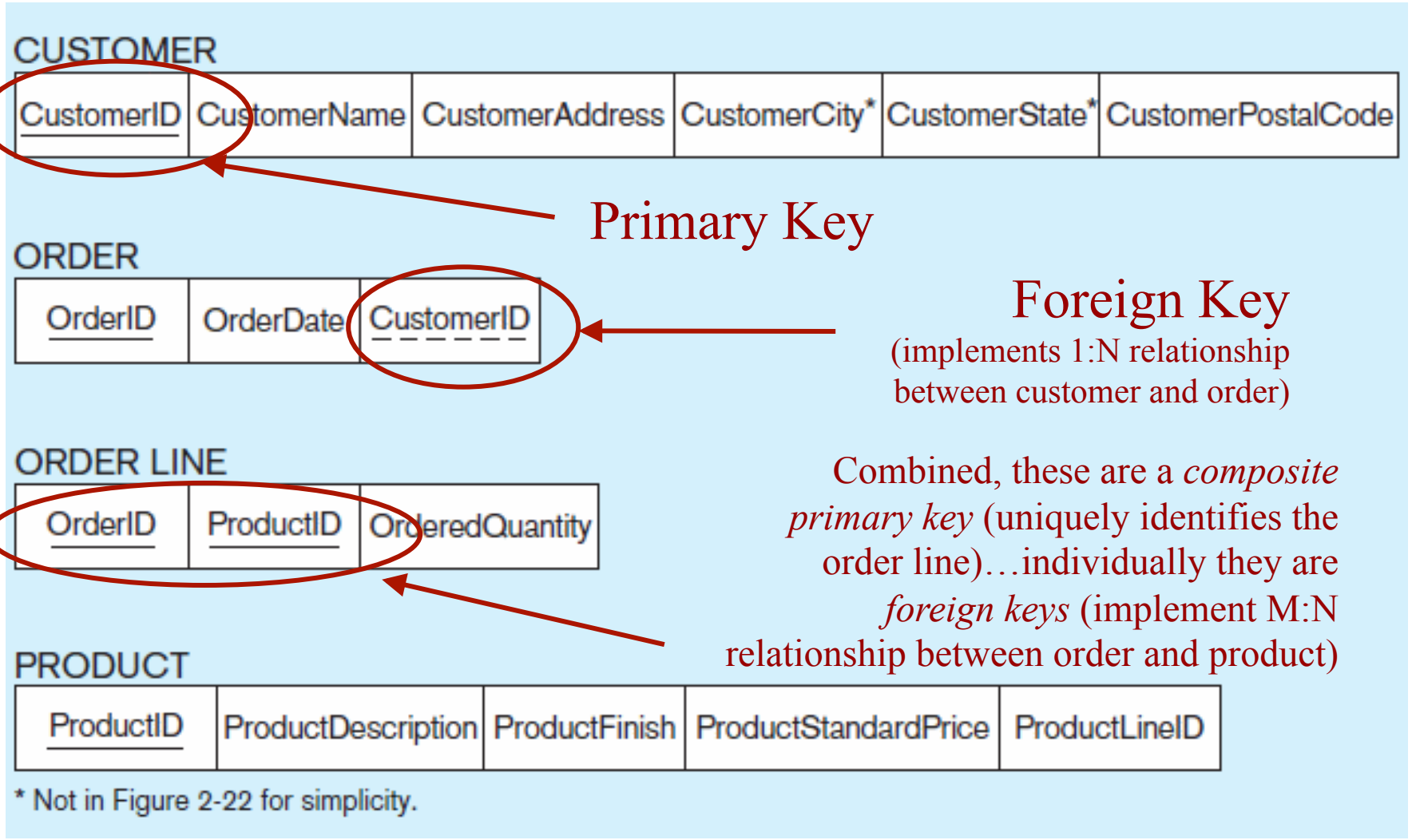
- A relation is a named, two-dimensional table of data.
- A table consists of rows (records) and columns (attribute or field).
- Requirements for a table to qualify as a relation:
  - It must have a unique name.
  - Every attribute value must be atomic (not multivalued, not composite).
  - Every row must be unique (can't have two rows with exactly the same values for all their fields).
  - Attributes (columns) in tables must have unique names.
  - The order of the columns must be irrelevant.
  - The order of the rows must be irrelevant.

KEY FIELDS



- Keys are special fields that serve two main purposes:
  - ***Primary keys*** are unique identifiers of the relation. Examples include employee numbers, social security numbers, etc. *This guarantees that all rows are unique.*
  - ***Foreign keys*** are identifiers that enable a dependent relation (on the many side of a relationship) to refer to its parent relation (on the one side of the relationship).

- Keys can be ***simple*** (a single field) or ***composite*** (more than one field).
- Keys usually are used as indexes to speed up the response to user queries (more on this in Chapter 5).


CUSTOMER
CustomerID
CustomerName
CustomerAddress
CustomerCity*
CustomerState*
CustomerPostalCode
Primary Key
ORDER
OrderID
OrderDate
CustomerID
Foreign Key
(implements 1:N relationship between customer and order)
ORDER LINE
OrderID
ProductID
OrderedQuantity
Combined, these are a composite primary key (uniquely identifies the order line)…individually they are foreign keys (implement M:N relationship between order and product)
PRODUCT
ProductID
ProductDescription
ProductFinish
ProductStandardPrice
ProductLineID
* Not in Figure 2-22 for simplicity.

Three integrity constraints

- To facilitate maintaining the accuracy and integrity of data in the database
  - Entity integrity
  - Domain integrity
  - Referential integrity

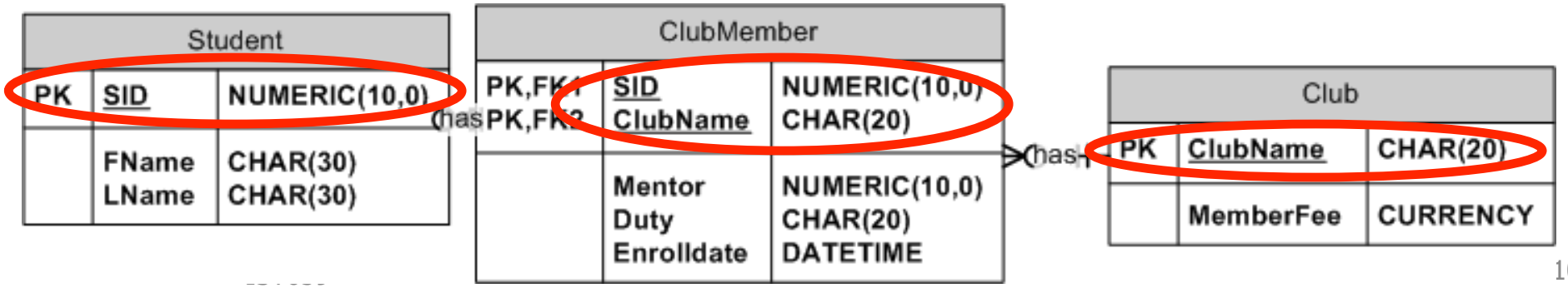
Integrity constraints (I)

- Entity integrity (primary key)
  - Every entity has a primary key
  - And thus its value in every row can not be NULL

SID	Fname	Lname	Birthdate
100001	Jim	Gray	07/17/1970
100002	Richard	Lancaster	07/17/1970
100003	Jenny	White	06/08/1975
100001	Jim	Gray	07/17/1970

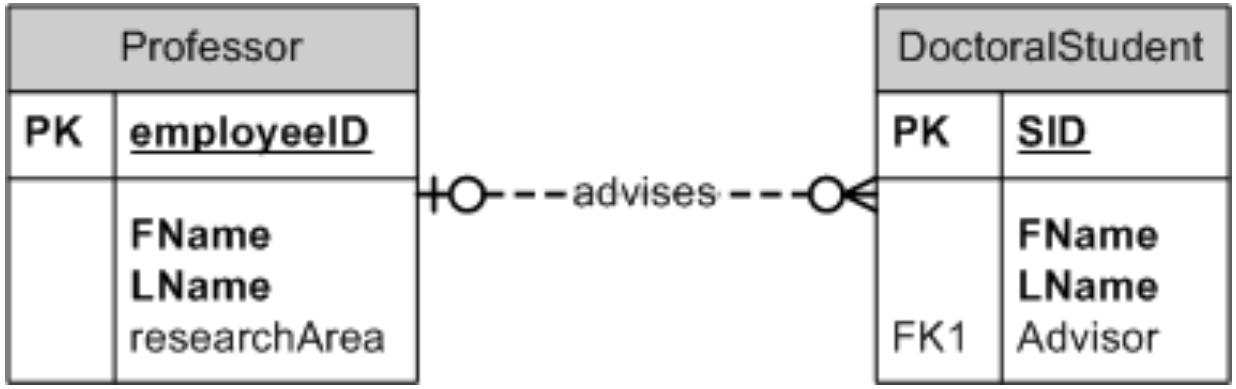
Integrity constraints (II)

- Domain integrity
  - Every column has a data type
  - The allowable value for an attribute has to conform to the data type.
    - For example, if it is a number you can't put in a text string)
  - A foreign key in one entity must have the same data type as its corresponding primary key in another entity does



Integrity constraints (III)

- Referential integrity (foreign key)
  - A set of attributes that is primary key in another table (aka parent table).
  - Its value in every row has to match a primary key value in the parent table table, or NULL



Referential integrity

- Delete Rules
  - **Restrict**–don't allow delete of "parent" side if related rows exist in "dependent" side
  - **Cascade**–automatically delete "dependent" side rows that correspond with the "parent" side row to be deleted
  - **Set-to-Null**–set the foreign key in the dependent side to null if deleting from the parent side → not allowed for weak entities

Referential Integrity Constraints and Delete Rules

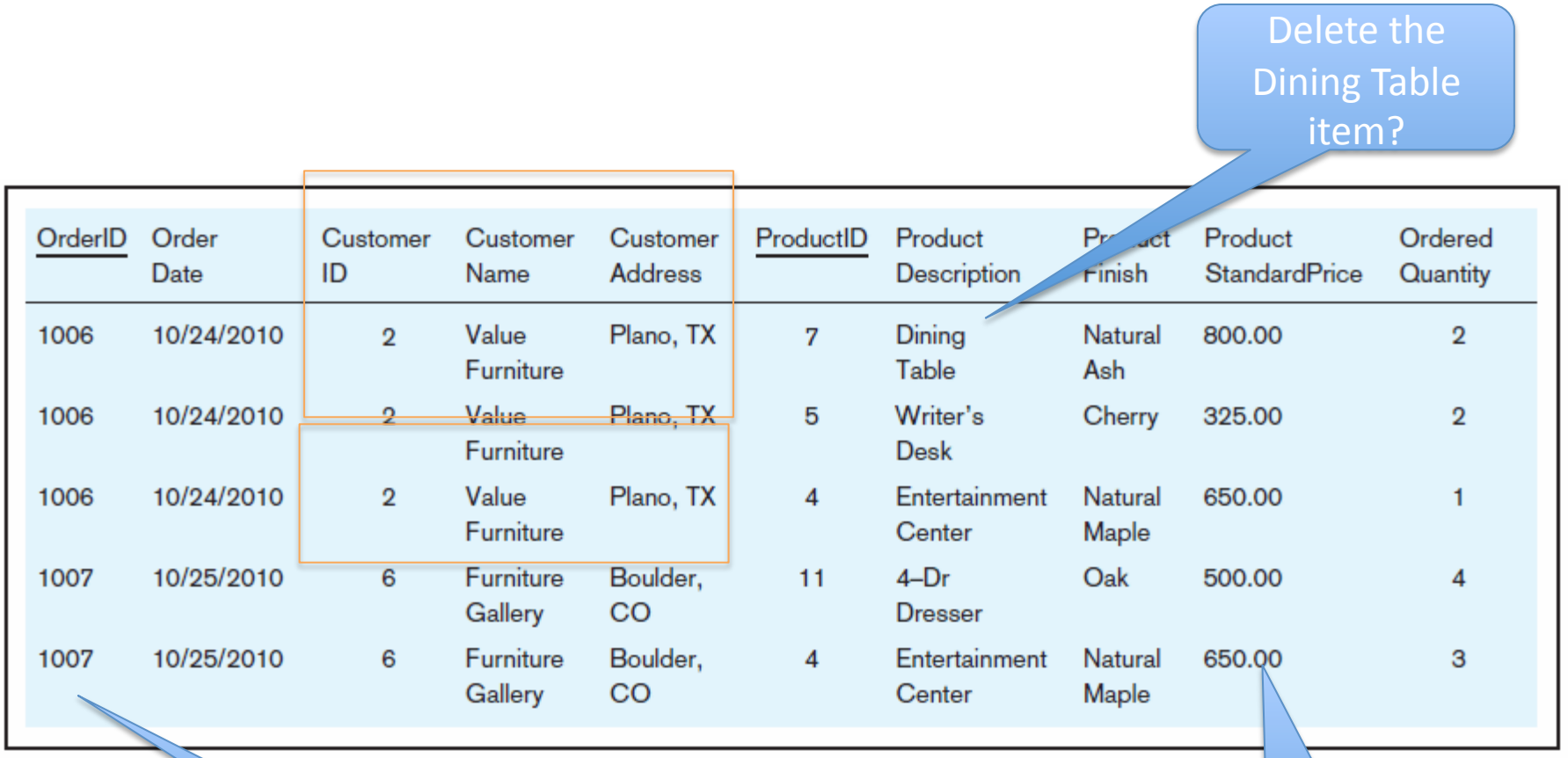
Example in MySQL

```
CREATE TABLE categories (
    id int unsigned not null primary key,
    name varchar(255) default null
);
CREATE TABLE products (
    id int unsigned not null primary key,
    name varchar(255) default null
);

CREATE TABLE categories_products (
    category_id int unsigned not null,
    product_id int unsigned not null,
    PRIMARY KEY (category_id, product_id),
    KEY pkey (product_id),
    FOREIGN KEY (category_id) REFERENCES categories (id)
       ON DELETE CASCADE
       ON UPDATE CASCADE,
    FOREIGN KEY (product_id) REFERENCES products (id)
       ON DELETE CASCADE
       ON UPDATE CASCADE,

);
```

Data Redundancy may break Data Integrity Rules

- Data redundancy is bad.
  - Waste storage
  - Redundancies in a table may result in errors or inconsistencies (called anomaly) when a user attempts to update the data in the table.
- **Three types of anomalies**
  - **Insertion Anomaly**– occurs when certain attributes cannot be inserted into the database without the presence of other attributes.
  - **Deletion Anomaly**– when certain attributes are lost because of the deletion of other attributes
  - **Update Anomaly**–changing data in a row forces changes to other rows because of duplication, otherwise the data will be inconsistent.

An example of Data Redundancy



OrderID	Order Date	Customer ID	Customer Name	Customer Address	ProductID	Product Description	Product Finish	Product StandardPrice	Ordered Quantity
1006	10/24/2010	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
1006	10/24/2010	2	Value Furniture	Plano, TX	5	Writer's Desk	Cherry	325.00	2
1006	10/24/2010	2	Value Furniture	Plano, TX	4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2010	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
1007	10/25/2010	6	Furniture Gallery	Boulder, CO	4	Entertainment Center	Natural Maple	650.00	3

FIGURE 4-26 IN[illegible]ICE relation (1NF) (Pine Valley Furniture Company)

ANOMALIES in the Example

- Insertion–if new product is ordered for order 1007 of existing customer, customer data must be re-entered, causing duplication
- Deletion–if we delete the Dining Table from Order 1006, we lose information concerning this item's finish and price
- Update–changing the price of product ID 4 requires update in multiple records

Why do these anomalies exist?
Because there are multiple themes (entity types) in one relation. This results in duplication and an unnecessary dependency between the entities.

Normalization

DATA NORMALIZATION

- Primarily a tool to validate and improve a logical design so that it satisfies certain constraints that ***avoid unnecessary duplication of data***
- Data normalization is a systematic approach to removing data redundancy, a process of decomposing relations with anomalies to produce smaller, well-structured relations
- Goal is to avoid the three anomalies

General rule of thumb: A table should not pertain to more than one entity type.

Concepts

- Normalization can be achieved in stages, each of which corresponds to a normal form.
- Normal form: a state of a relation that requires that certain rules regarding relationships between attributes (or functional dependencies) are satisfied.

Functional Dependency & Keys

- Functional Dependency: The value of one attribute (the *determinant*) determines the value of another attribute

  - Partial Functional Dependency: a functional dependency in which one or more nonkey attributes are functionally dependent on part (but not all) of the primary key.

  - Transitive Dependency: a functionary dependency between the primary key and one or more nonkey attributes that are dependent on the primary key via another nonkey attribute.

Understanding Functional Dependence

- For attributes A and B, B is **functionally dependent** on A means each value in column A determines one and only one value in column B.
  - Written: A → B
  - A determines B
  - B is the determinant
  - **Ex:** SSN → Name (Name is functionally dependent on SSN)

SSN	Name
123-45-6789	George Foreman
123-46-9987	Georgeina Forman
123-02-0902	George Foreman
123-02-0993	George Foreman

Activity

Driver Permissions Report

Driver ID #	Driver Name	Driver Chg/Hr	Driver Territories	Vehicle Lic Plate	Vehicle Make	Vehicle Model	Vehicle Size	Vehicle Chg/Hr	Permission Exp. Date
101	Bill Melator	$100.00	West, North, Central	PPF673	Cadillac	Escalde	M	$ 100.00	12/31/04
101	Bill Melator	$100.00	West, North, Central	PXK3D7T	Chevy	Tahoe	L	$ 120.00	12/31/04
101	Bill Melator	$100.00	West, North, Central	445GH2	Lincon	Towncar	S	$ 80.00	1/30/13
101	Bill Melator	$100.00	West, North, Central	59DLLK	Lincon	Continental	S	$ 80.00	4/30/15
102	Willie Work	$75.00	South, East	PXK3D7T	Chevy	Tahoe	L	$ 120.00	1/15/14
102	Willie Work	$75.00	South, East	663ETMP	Chevy	Suburban	L	$ 120.00	4/1/15
103	Sal Ladd	$75.00	Central, East, North	667GM8	Audi	A8	M	$ 100.00	9/1/15
103	Sal Ladd	$75.00	Central, East, North	445GH2	Lincon	Towncar	S	$ 80.00	9/1/14
103	Sal Ladd	$75.00	Central, East, North	59DLLK	Lincon	Continental	S	$ 80.00	12/31/12
104	Carol Ling	$100.00	Central, West	667GM8	Audi	A8	M	$ 100.00	7/31/12
104	Carol Ling	$100.00	Central, West	PPF673	Cadillac	Escalde	M	$ 100.00	7/31/13
104	Carol Ling	$100.00	Central, West	59DLLK	Lincon	Continental	S	$ 80.00	10/1/14

Identify the:

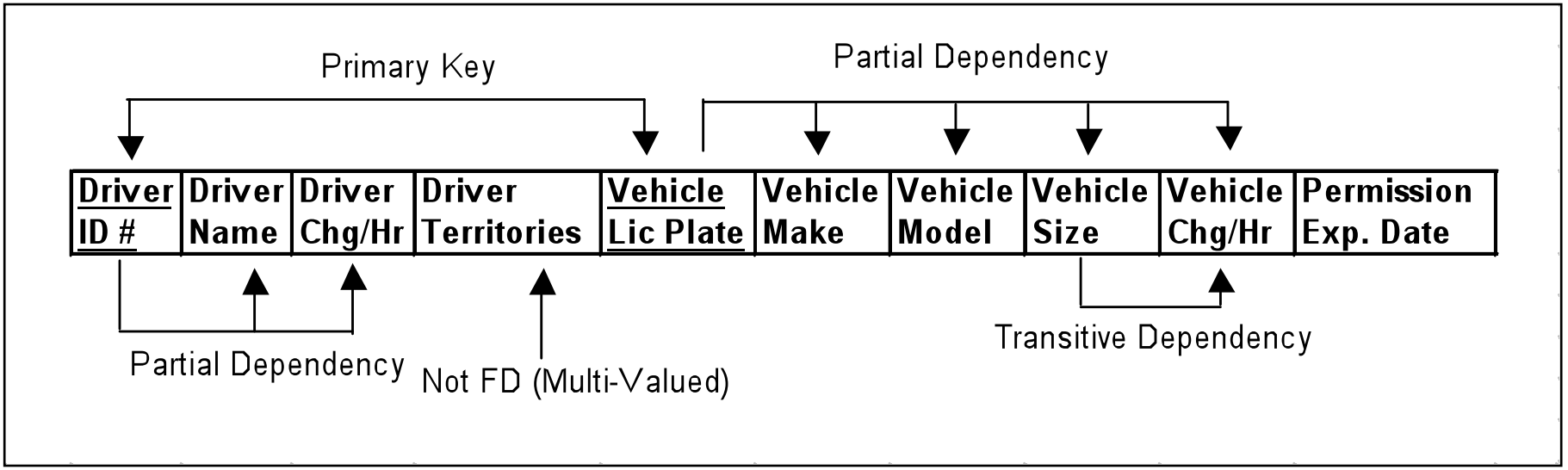
- Chosen Key?
- Key Attributes?
- Non-Key Attributes

Identify the:

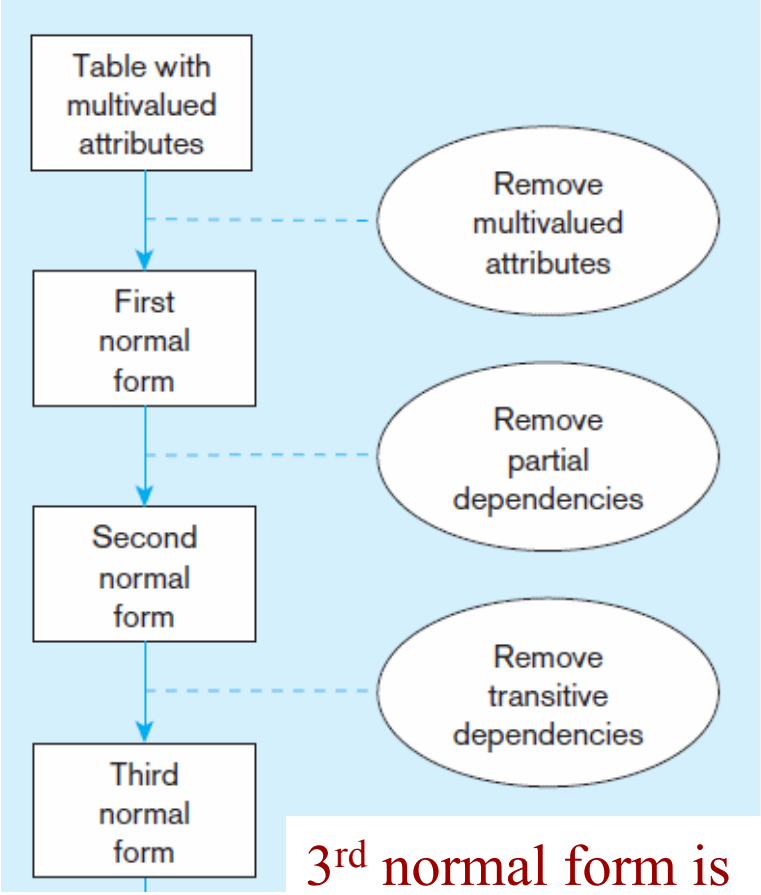
- Functional Dependencies (WRT the Key)
- Partial Functional Dependencies (WRT part of the Key)
- Transitive Functional Dependencies (WRT some non-prime attribute)

The Dependency Diagram

- The **Dependency** Diagram is a Very Useful Tool. It depicts the dependencies which exist among the attributes.



Data Normalization Process to 3rd Normal Form (3NF)



3rd normal form is generally considered sufficient

Step 1: check 1NF

Assume you are given an excel sheet as follows:

Figure 5-25 INVOICE date (Pine Valley Furniture Company)

Order_ID	Order_ Date	Customer_ ID	Customer_ Name	Customer_ Address	Product_ID	Product_ Description	Product_ Finish	Unit_ Price	Ordered_ Quantity
1006	10/24/2008	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
					5	Writer's Desk	Cherry	325.00	2
					4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
					4	Entertainment Center	Natural Maple	650.00	3

1st Normal Form (1NF)

- All relations are in 1NF
- A relation is a named, two-dimensional table of data which satisfies the following requirements:
  - There is no row or column order
  - Every column must have a unique name
  - Every row must be unique (primary key exists)
  - Every value is atomic
  - No repeating columns

Table with a row order, not in 1st normal form

OrderID	Order Date	Customer ID	Customer Name	Customer Address	ProductID	Product Description	Product Finish	Product StandardPrice	Ordered Quantity
1006	10/24/2010	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
					5	Writer's Desk	Cherry	325.00	2
					4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2010	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
					4	Entertainment Center	Natural Maple	650.00	3

FIGURE 4-25 INVOICE data (Pine Valley Furniture Company)

Note: This is NOT a relation.

1st normal form (1NF)

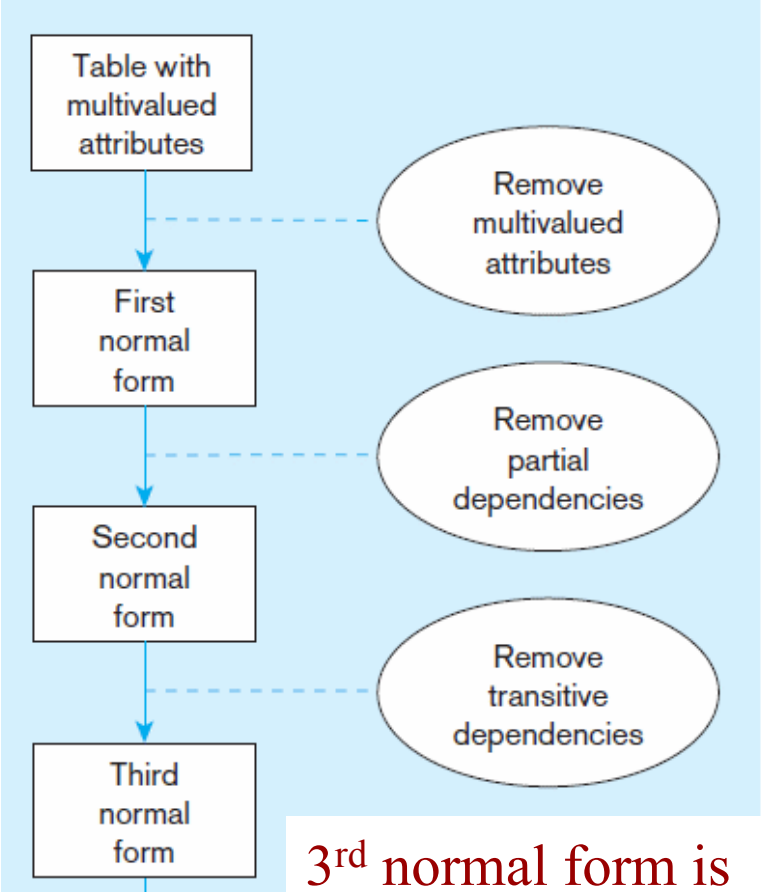
Table with no row order and unique rows

OrderID	Order Date	Customer ID	Customer Name	Customer Address	ProductID	Product Description	Product Finish	Product StandardPrice	Ordered Quantity
1006	10/24/2010	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
1006	10/24/2010	2	Value Furniture	Plano, TX	5	Writer's Desk	Cherry	325.00	2
1006	10/24/2010	2	Value Furniture	Plano, TX	4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2010	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
1007	10/25/2010	6	Furniture Gallery	Boulder, CO	4	Entertainment Center	Natural Maple	650.00	3

FIGURE 4-26 INVOICE relation (1NF) (Pine Valley Furniture Company)

Note: This is a relation, but not a well-structured one.

Data Normalization Process to 3rd Normal Form (3NF)



- Check if each entity is 1NF, if not convert it to 1NF
  - No row or column order
  - Primary key exists
  - Every value is atomic, no multivalued attributes
  - No repeating columns
  - All relations are in 1NF, e.g. MS Access Tables

3rd normal form is generally considered sufficient

Step 2: check functional dependencies to find redundant data groups

Figure 5-26 INVOICE relation (1NF) (Pine Valley Furniture Company)

Order_ID	Order_ Date	Customer_ ID	Customer_ Name	Customer_ Address	Product_ID	Product_ Description	Product_ Finish	Unit_ Price	Ordered_ Quantity
1006	10/24/2008	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	5	Writer's Desk	Cherry	325.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	4	Entertainment Center	Natural Maple	650.00	3

Step 2: check functional dependencies to find redundant data groups

Figure 5-26 INVOICE relation (1NF) (Pine Valley Furniture Company)

Order_ID	Order_ Date	Customer_ ID	Customer_ Name	Customer_ Address	Product_ID	Product_ Description	Product_ Finish	Unit_ Price	Ordered_ Quantity
1006	10/24/2008	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	5	Writer's Desk	Cherry	325.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	4	Entertainment Center	Natural Maple	650.00	3

What is the primary key of this 1NF table? (order_ID, product_ID) assuming any product would appear at most once in each order

Figure 5-26 INVOICE relation (1NF) (Pine Valley Furniture Company)

Order_ID	Order_ Date	Customer_ ID	Customer_ Name	Customer_ Address	Product_ID	Product_ Description	Product_ Finish	Unit_ Price	Ordered_ Quantity
1006	10/24/2008	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	5	Writer's Desk	Cherry	325.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	4	Entertainment Center	Natural Maple	650.00	3

Figure 5-26 INVOICE relation (1NF) (Pine Valley Furniture Company)

Order_ID	Order_ Date	Customer_ ID	Customer_ Name	Customer_ Address	Product_ID	Product_ Description	Product_ Finish	Unit_ Price	Ordered_ Quantity
1006	10/24/2008	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	5	Writer's Desk	Cherry	325.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	4	Entertainment Center	Natural Maple	650.00	3

Order_ID ➔ Order_Date, Customer_ID, Customer_Name, Customer_Address

Figure 5-26 INVOICE relation (1NF) (Pine Valley Furniture Company)

Order_ID	Order_ Date	Customer_ ID	Customer_ Name	Customer_ Address	Product_ID	Product_ Description	Product_ Finish	Unit_ Price	Ordered_ Quantity
1006	10/24/2008	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	5	Writer's Desk	Cherry	325.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	4	Entertainment Center	Natural Maple	650.00	3

Order_ID ➔ Order_Date, Customer_ID, Customer_Name, Customer_Address

Figure 5-26 INVOICE relation (1NF) (Pine Valley Furniture Company)

Order_ID	Order_ Date	Customer_ ID	Customer_ Name	Customer_ Address	Product_ID	Product_ Description	Product_ Finish	Unit_ Price	Ordered_ Quantity
1006	10/24/2008	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	5	Writer's Desk	Cherry	325.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	4	Entertainment Center	Natural Maple	650.00	3

Order_ID ➔ Order_Date, Customer_ID, Customer_Name, Customer_Address

Customer_ID ➔ Customer_Name, Customer_Address

Figure 5-26 INVOICE relation (1NF) (Pine Valley Furniture Company)

Order_ID	Order_ Date	Customer_ ID	Customer_ Name	Customer_ Address	Product_ID	Product_ Description	Product_ Finish	Unit_ Price	Ordered_ Quantity
1006	10/24/2008	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	5	Writer's Desk	Cherry	325.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	4	Entertainment Center	Natural Maple	650.00	3

Order_ID ➔ Order_Date, Customer_ID, Customer_Name, Customer_Address

Customer_ID ➔ Customer_Name, Customer_Address

Figure 5-26 INVOICE relation (1NF) (Pine Valley Furniture Company)

Order_ID	Order_ Date	Customer_ ID	Customer_ Name	Customer_ Address	Product_ID	Product_ Description	Product_ Finish	Unit_ Price	Ordered_ Quantity
1006	10/24/2008	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	5	Writer's Desk	Cherry	325.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	4	Entertainment Center	Natural Maple	650.00	3

Order_ID ➔ Order_Date, Customer_ID, Customer_Name, Customer_Address

Customer_ID ➔ Customer_Name, Customer_Address

Product_ID ➔ Product_Description, Product_Finish, Unit_Price

Figure 5-26 INVOICE relation (1NF) (Pine Valley Furniture Company)

Order_ID	Order_ Date	Customer_ ID	Customer_ Name	Customer_ Address	Product_ID	Product_ Description	Product_ Finish	Unit_ Price	Ordered_ Quantity
1006	10/24/2008	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	5	Writer's Desk	Cherry	325.00	2
1006	10/24/2008	2	Value Furniture	Plano, TX	4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	4	Entertainment Center	Natural Maple	650.00	3

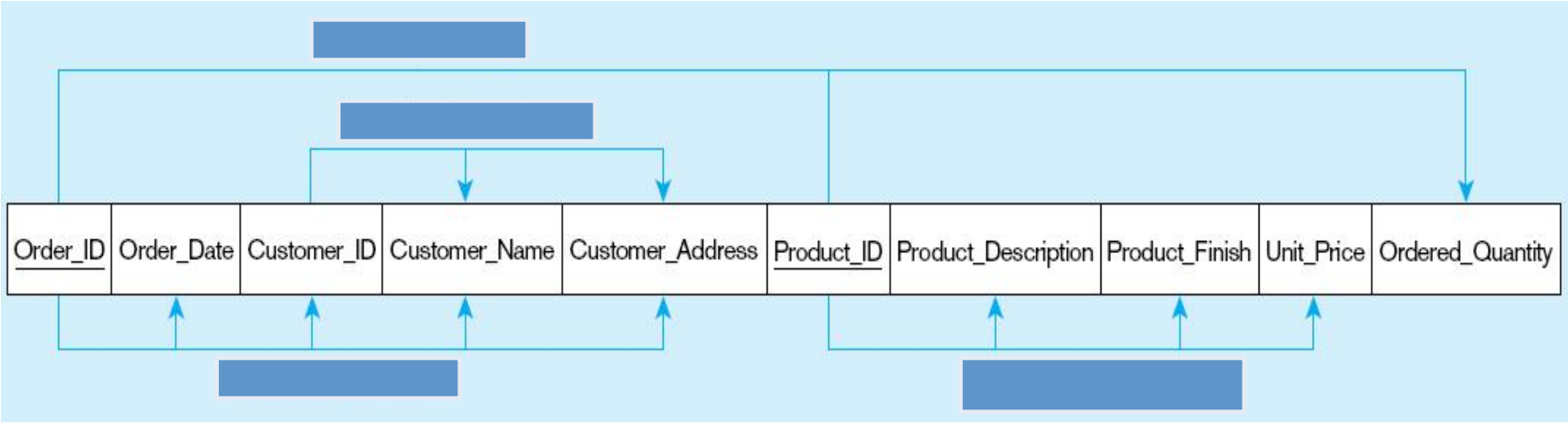
Order_ID ➔ Order_Date, Customer_ID, Customer_Name, Customer_Address

Customer_ID ➔ Customer_Name, Customer_Address

Product_ID ➔ Product_Description, Product_Finish, Unit_Price

Order_ID, Product_ID ➔ Order_Quantity

Figure 5-27 Functional dependency diagram for INVOICE



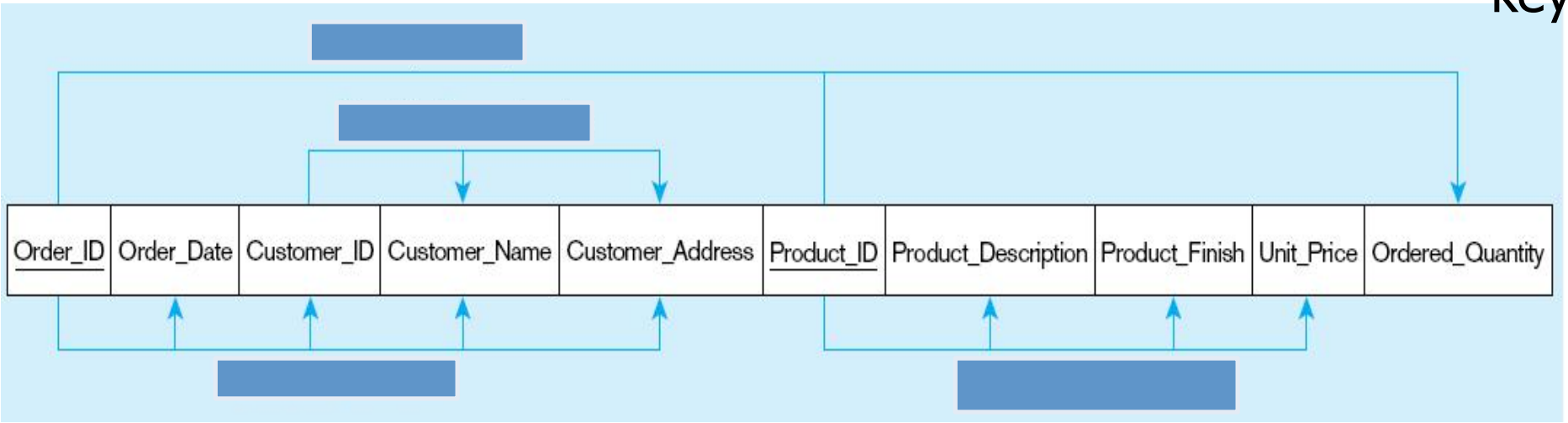
Order_ID ➔ Order_Date, Customer_ID, Customer_Name, Customer_Address

Customer_ID ➔ Customer_Name, Customer_Address

Product_ID ➔ Product_Description, Product_Finish, Unit_Price

Order_ID, Product_ID ➔ Order_Quantity

A partial dependency exists when a nonkey attribute is functionally dependent on part (but not all) of the primary key



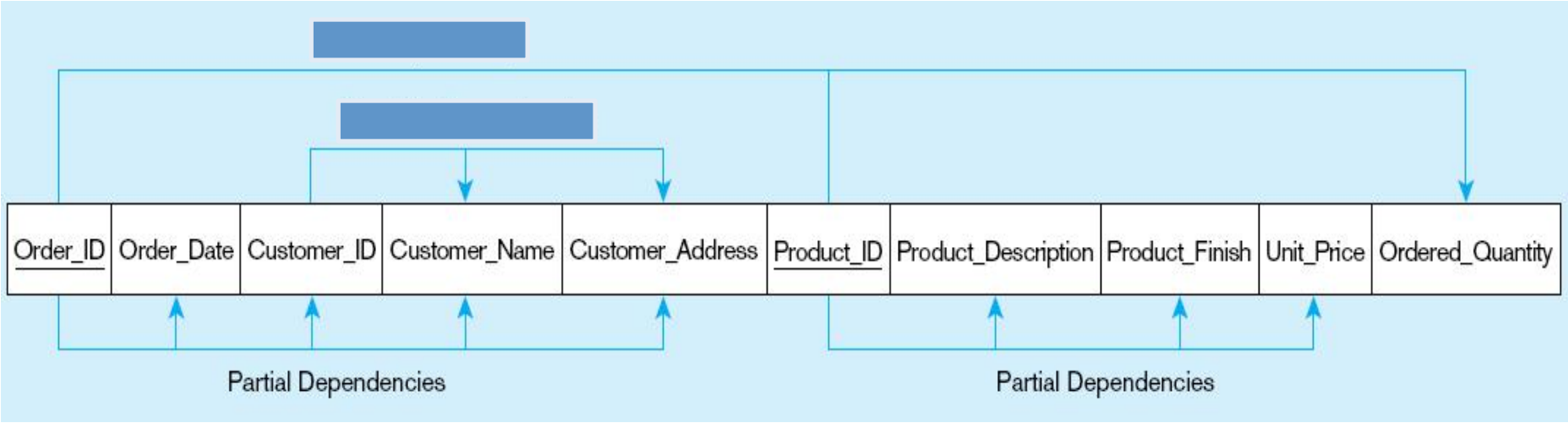
Order_ID ➔ Order_Date, Customer_ID, Customer_Name, Customer_Address

Customer_ID ➔ Customer_Name, Customer_Address

Product_ID ➔ Product_Description, Product_Finish, Unit_Price

Order_ID, Product_ID ➔ Order_Quantity

A partial dependency exists when a nonkey attribute is functionally dependent on part (but not all) of the primary key



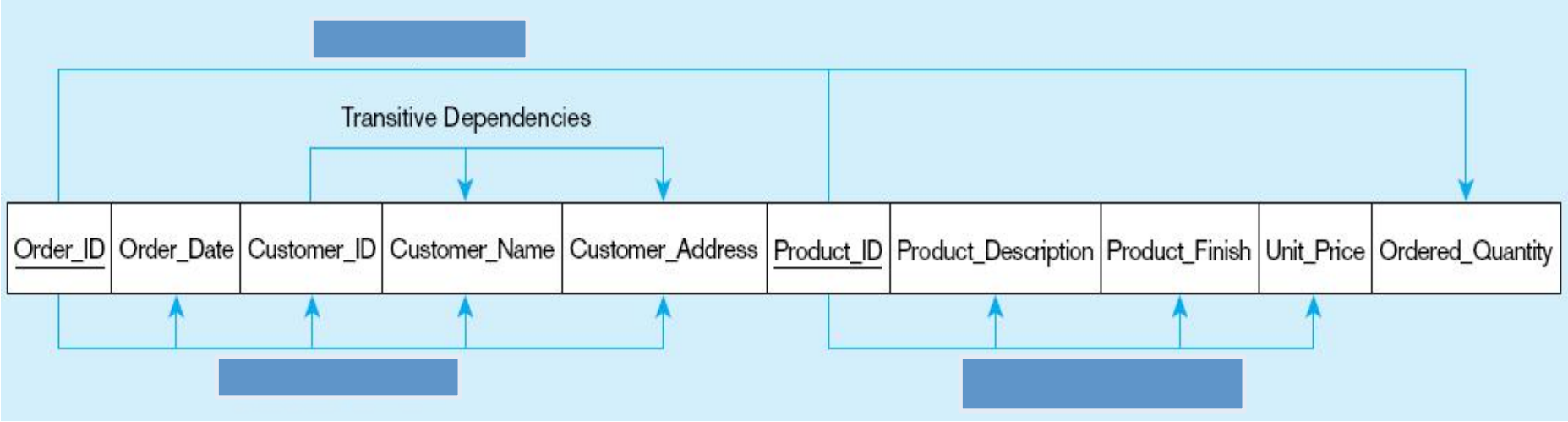
Order_ID ➔ Order_Date, Customer_ID, Customer_Name, Customer_Address

Customer_ID ➔ Customer_Name, Customer_Address

Product_ID ➔ Product_Description, Product_Finish, Unit_Price

Order_ID, Product_ID ➔ Order_Quantity (primary key)

A transitive dependency is a functional dependency between two (or more) nonkey attributes



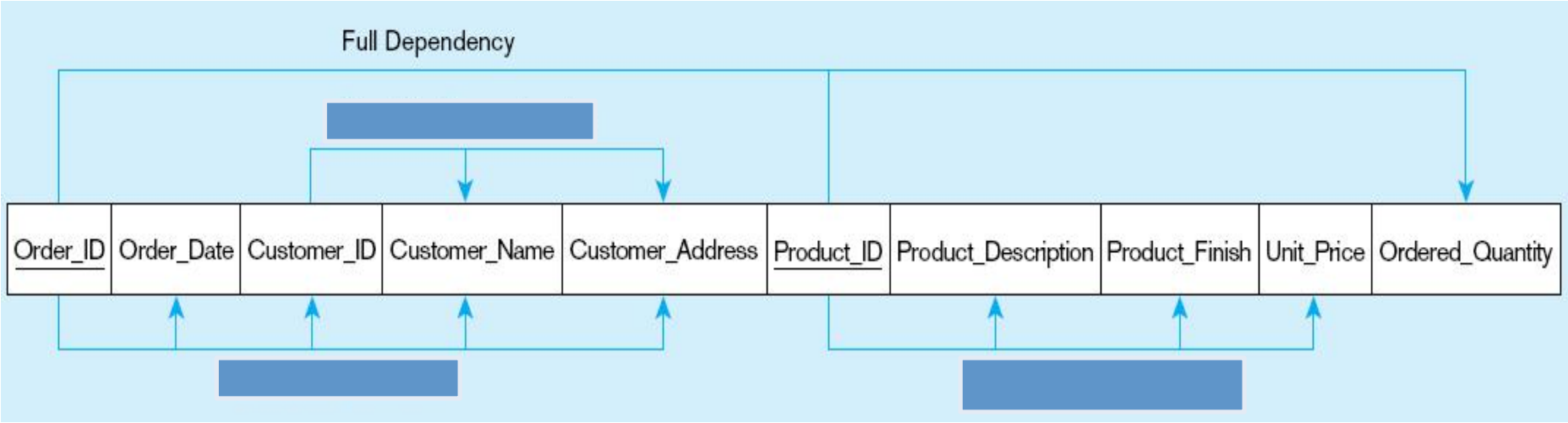
Order_ID ➔ Order_Date, Customer_ID, Customer_Name, Customer_Address

Customer_ID ➔ Customer_Name, Customer_Address

Product_ID ➔ Product_Description, Product_Finish, Unit_Price

Order_ID, Product_ID ➔ Order_Quantity

A full dependency exists when all nonkey attributes are functionally dependent on the entire primary key ONLY.



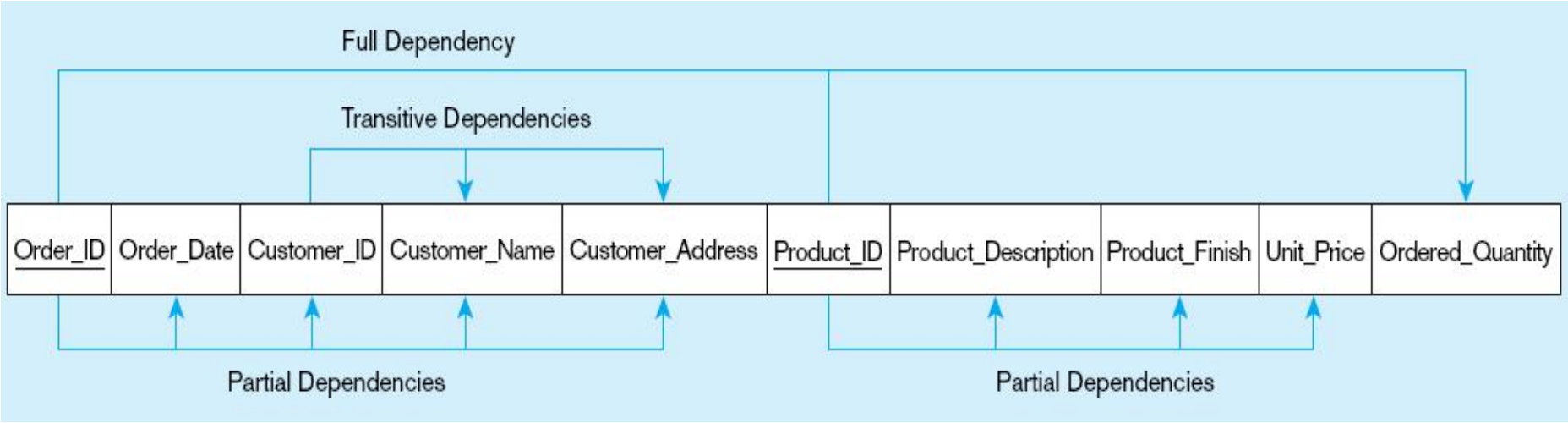
Order_ID ➔ Order_Date, Customer_ID, Customer_Name, Customer_Address

Customer_ID ➔ Customer_Name, Customer_Address

Product_ID ➔ Product_Description, Product_Finish, Unit_Price

Order_ID, Product_ID ➔ Order_Quantity

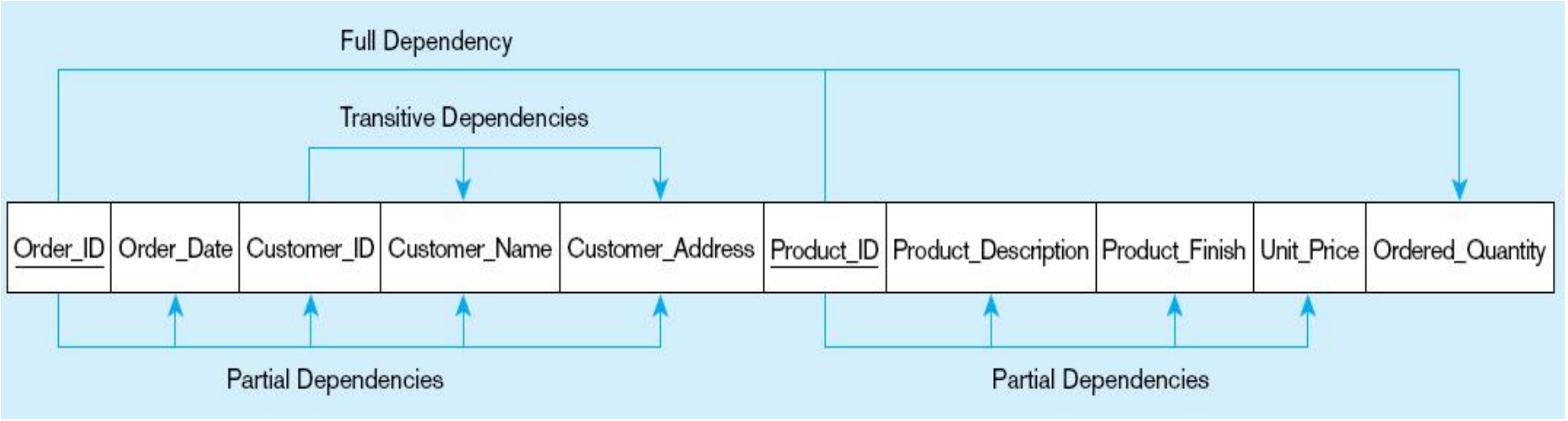
An Overview

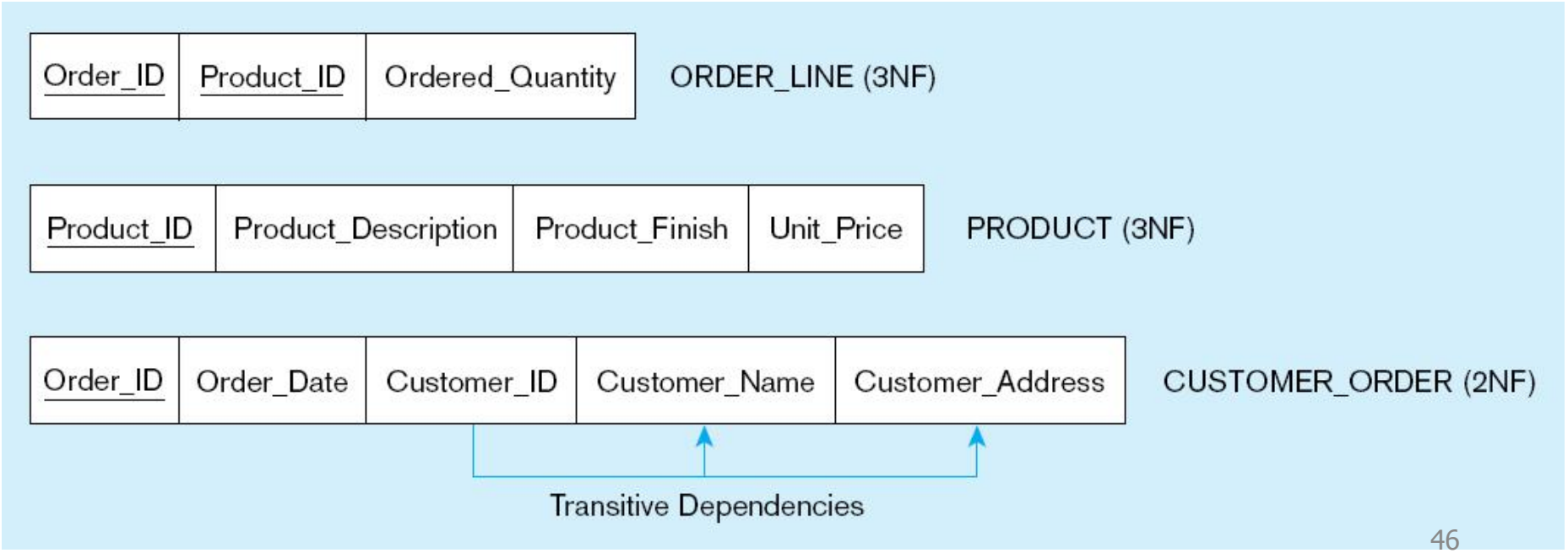

Full Dependency
Transitive Dependencies
Order_ID
Order_Date
Customer_ID
Customer_Name
Customer_Address
Product_ID
Product_Description
Product_Finish
Unit_Price
Ordered_Quantity
Partial Dependencies
Partial Dependencies

2nd NORMAL FORM (2NF)

- 1NF PLUS *every non-key attribute is fully functionally dependent on the ENTIRE primary key*
  - Every non-key attribute must be defined by the entire key, not by only part of the key
  - No partial functional dependencies
- Solution: create a new table for each partial dependency

Removing partial dependencies: 1NF-> 2NF





How do we associate M:N relationship?

SID	SName
1	Joe
2	Jane

CID	CName
659	database
565	datamining

SID	SName	CID
1	Joe	659,565
2	Jane	659

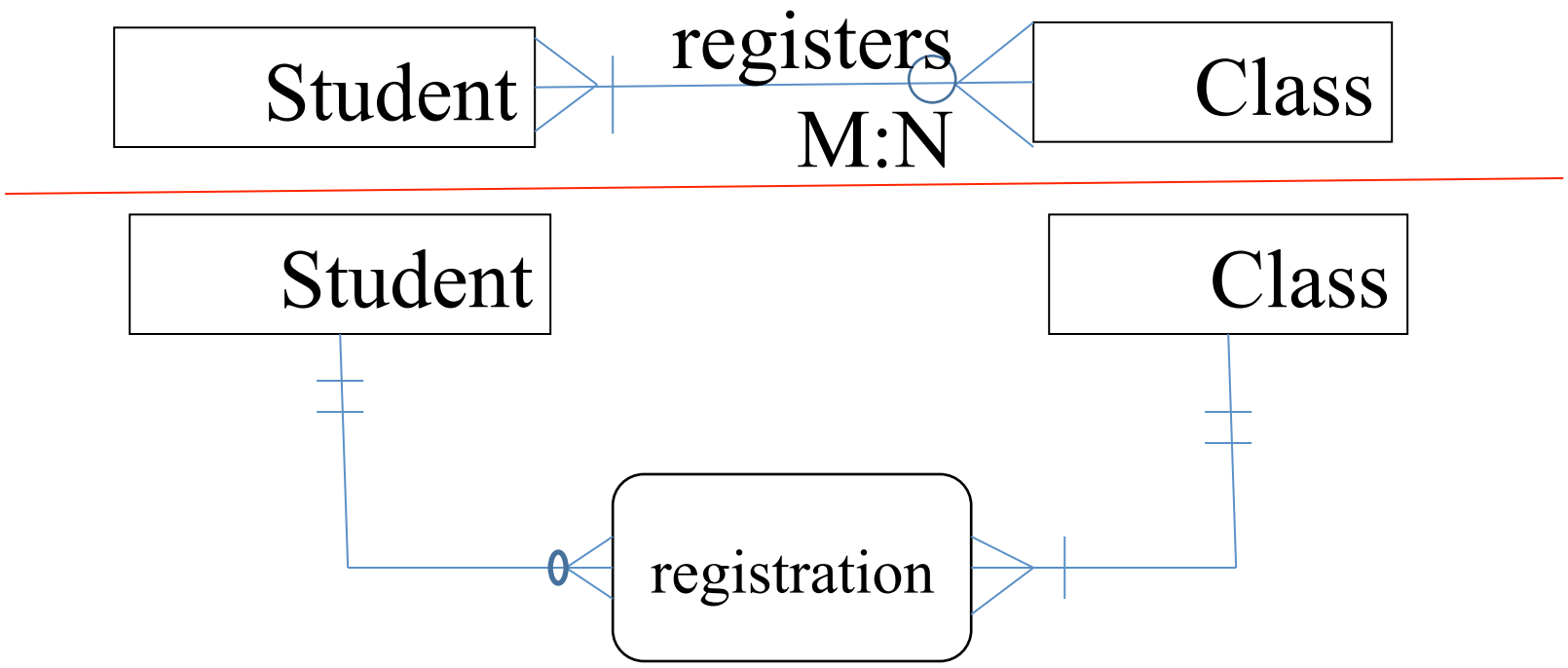
CID	CName	SID
659	database	1,2
565	datamining	1

Multi-valued attribute

Multi-valued attribute

What should we do now?

Associative entity



SID	SName
1	Joe
2	Jane

CID	CName
659	database
565	datamining

SID	SName
1	Joe
2	Jane

CID	CName
659	database
565	datamining

FK1

FK2

RID	SID	CID
1	1	659
2	1	565
3	2	659

SID	SName
1	Joe
2	Jane

CID	CName
659	database
565	datamining

SID	CID	SName
1	659	Joe
1	565	Joe
2	659	Jane

CID	CName
659	database
565	datamining

Will this work? PK (SID, CID)

SID	SName
1	Joe
2	Jane

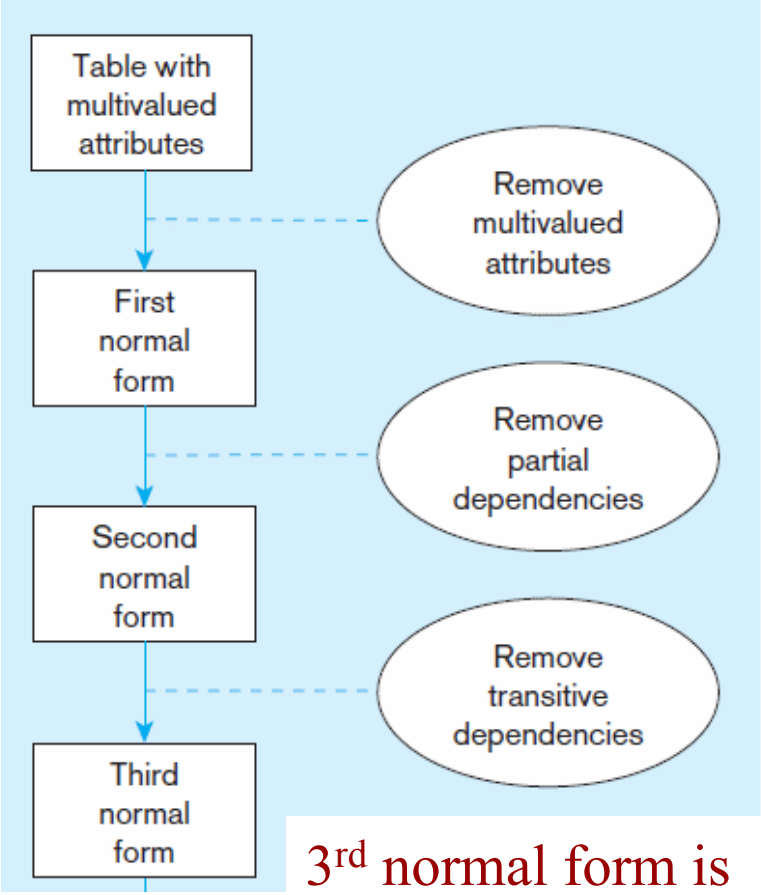
CID	CName
659	database
565	datamining

SID	CID	SName
1	659	Joe
1	565	Joe
2	659	Jane

CID	CName
659	database
565	datamining

Will this work? PK (SID, CID)
Partial dependency: SID->SName

Data Normalization Process to 3rd Normal Form (3NF)



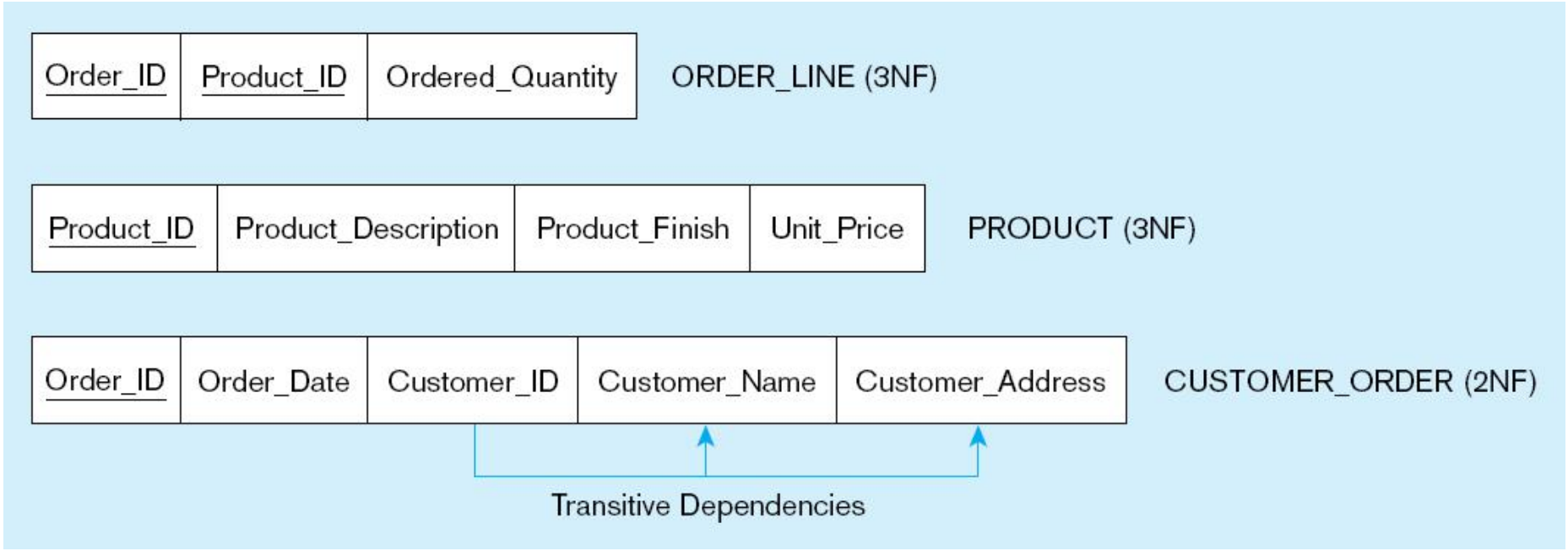
3rd normal form is generally considered sufficient

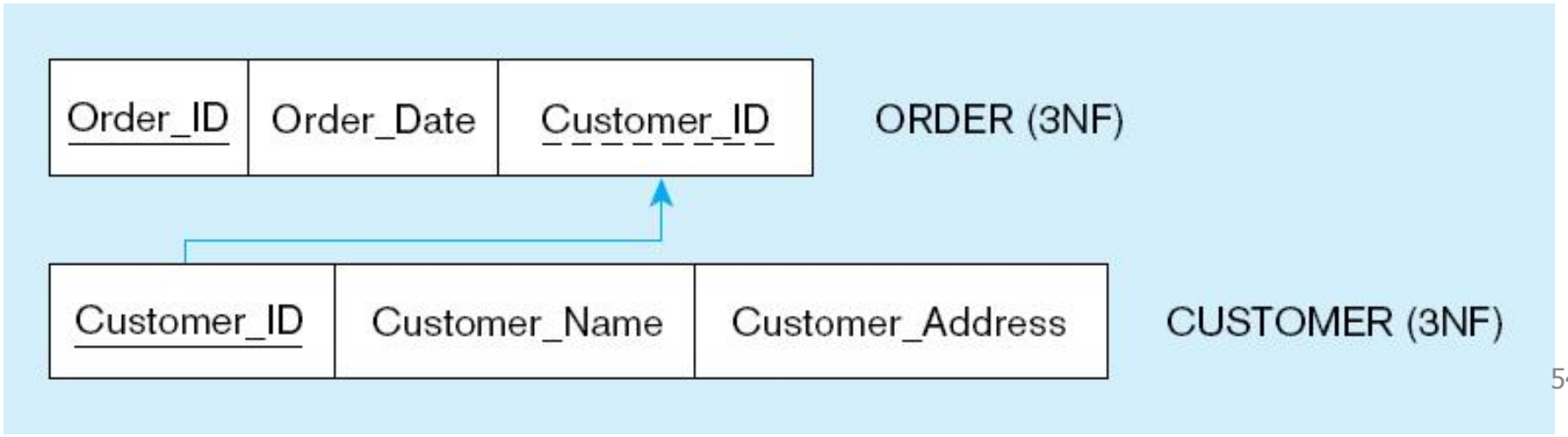
- 1NF
  - No row or column order
  - primary key exists
  - Every value is atomic, no multivalued attributes
  - No repeating columns
  - All relations are in 1NF, e.g. MS Access Tables
- Check the functional dependencies of each entity to find redundant data groups
- 2NF
  - Every non-key attribute must be defined by the entire key, not by only part of the key

3NF

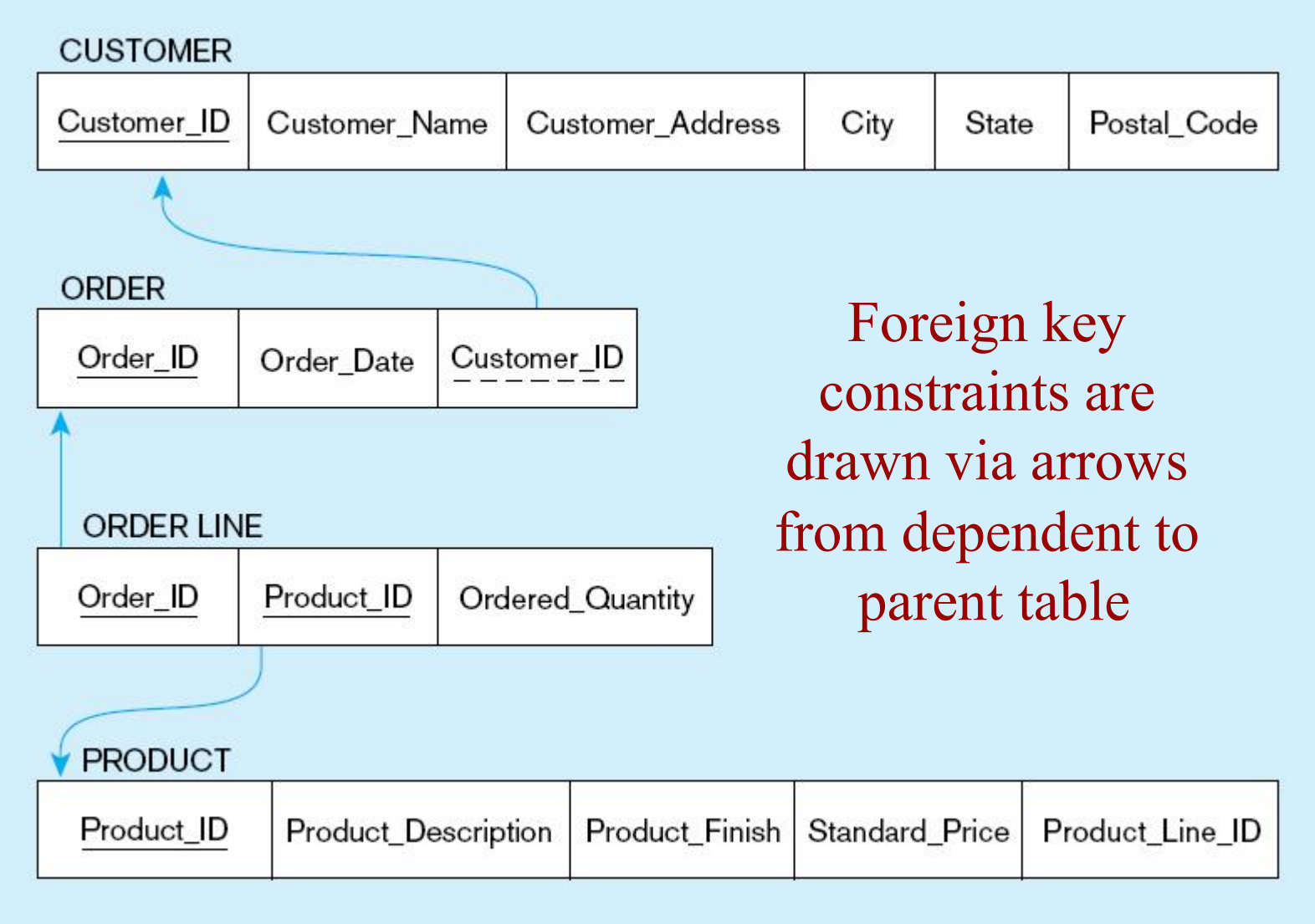
- 2NF PLUS ***no transitive dependencies*** (functional dependencies on non-primary-key attributes)
  - Each attribute must be dependent on the key, the whole key, and nothing but the key!
- Note: This is called transitive, because the primary key is a determinant for another attribute, which in turn is a determinant for a third
- Solution: Non-key determinant with transitive dependencies go into a new table; non-key determinant becomes primary key in the new table and stays as foreign key in the old table

Removing transitive dependencies: 2NF->3NF

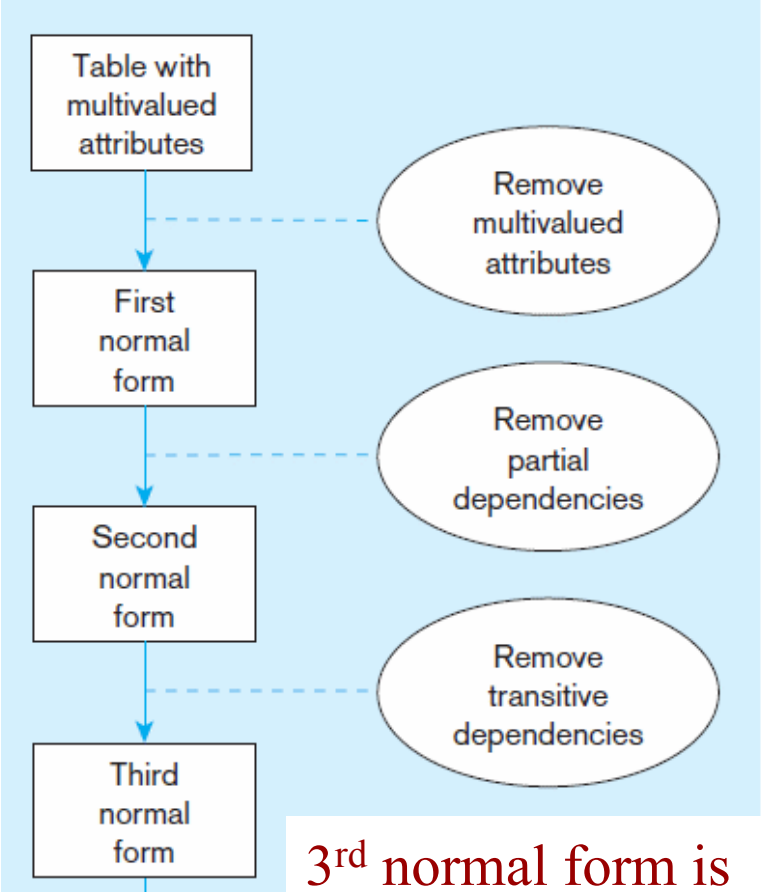




Connect the relations: foreign key constraints

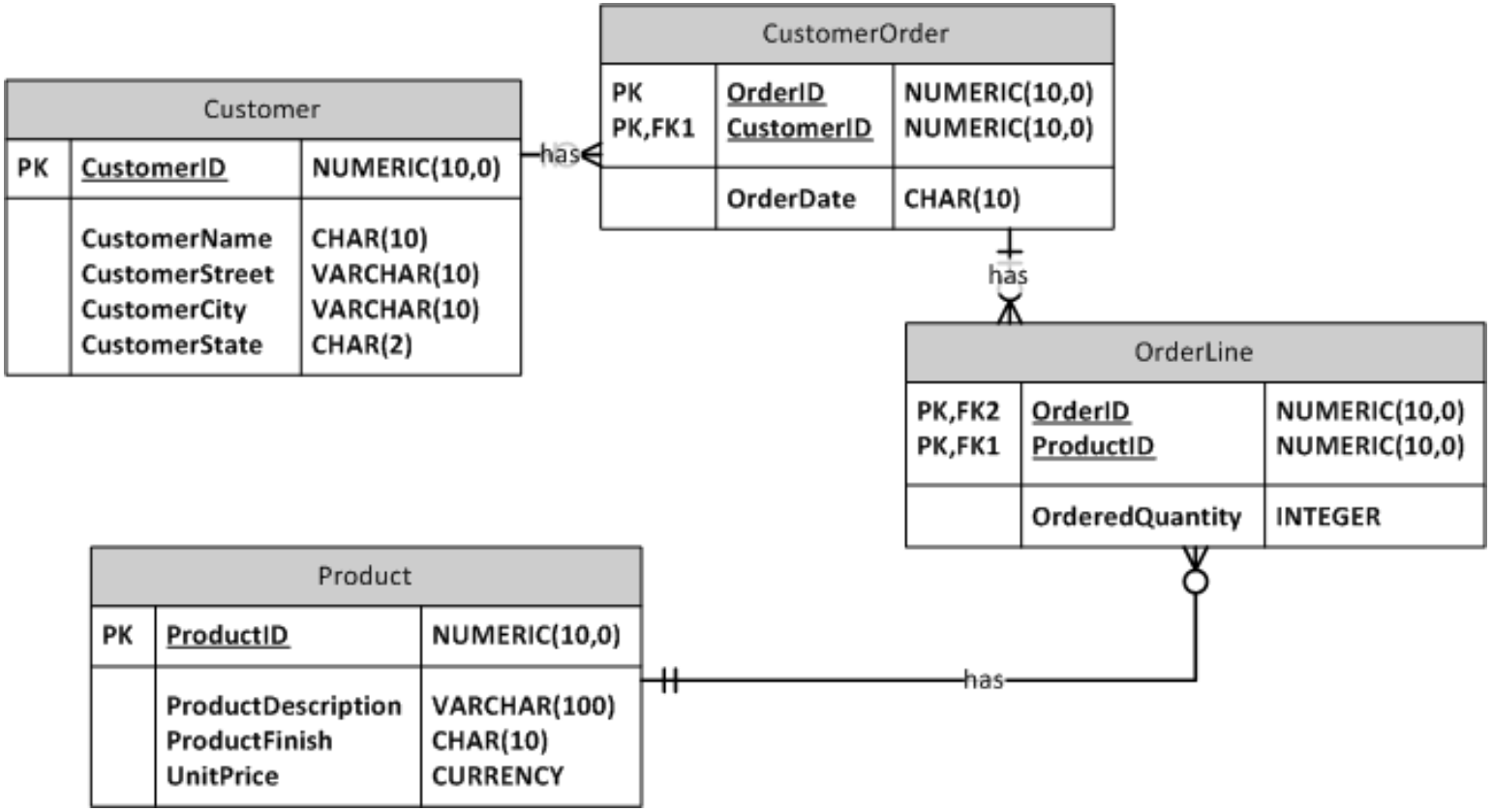


Data Normalization Process to 3rd Normal Form (3NF)



3rd normal form is generally considered sufficient

- 1NF
  - No row / column order
  - primary key exists
  - Every value is atomic, no multivalued attributes
  - No repeating columns
  - All relations are in 1NF, e.g. MS Access Tables
- Check the functional dependencies of each entity to find redundant data groups
- 2NF
  - Every non-key attribute must be defined by the entire key, not by only part of the key
- 3NF:
  - Each attribute must be dependent on the key, the whole key, and nothing but the key!



Every table in a 3NF data model should satisfy the requirement that "Each attribute must be dependent on the key, the whole key, and nothing but the key!"

Normalization is Important

- each table should satisfy the 3NF requirement that "all nonkey attributes should be dependent on the key, the whole key, and nothing but the key!"

Figure 5-25 INVOICE date (Pine Valley Furniture Company)

Order_ID	Order_ Date	Customer_ ID	Customer_ Name	Customer_ Address	Product_ID	Product_ Description	Product_ Finish	Unit_ Price	Ordered_ Quantity
1006	10/24/2008	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
					5	Writer's Desk	Cherry	325.00	2
					4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2008	6	Furniture Gallery	Boulder, CO	11	4–Dr Dresser	Oak	500.00	4
					4	Entertainment Center	Natural Maple	650.00	3

In Class Exercise: Student club database

Students participate in all kinds of club activities. A student can join many clubs and serve one duty in each club. A student may have a mentor in each club. A student can be a mentor for many other new students.

SID	Student Name	Mentor	Duty	Club	Member Fee	Enroll date
001	A. White	B. Gray	accountant	Fencing	20	08/31/2010
003	C. Black	D. Red	outreach	Chess	30	09/01/2010
002	B. Gray		president	Marathon	15	08/31/2008
00	E. Blue		president	Swim	40	08/30/2008
001	A. White	H. Purple	outreach	dance	50	09/25/2010

Example from ***Roger's Access Blog***

- Normalization_Example_1NF_3NF.docx on blackboard under week 4 lecture notes
- From ***http://rogersaccessblog.blogspot.com***

1NF?

Order1NF-wrong : Table

OrderNum	CustomerNum	OrderDate	Quantity	Item	ProductNum
112	444	2/1/2009	2	wrench	A7S
			3	O-ring	Z9Q
			1	saw	B7G
113	888	2/2/2009	2	clamp	L8U
			2	pail	G1W
114	333	2/3/2009	6	nail	T3B
			0		

Record: 2 of 6

1NF?

Order1NF-Wrong2 : Table

OrderNum	CustomerNum	OrderDate	Item
112	444	2/1/2009	2: wrench (A7S), 3: O-ring (Z9Q), 1: saw (B7G)
113	888	2/2/2009	2: clamp (L8U), 2: pail (G1W)
114	333	2/3/2009	6: nail (T3B)

Record: 1 of 3

1NF?

Order1NF-Wrong3 : Table

OrderNum	CustomerNum	OrderDate	Quantity	Item	Product
112	444	2/1/2009	2, 3, 1	wrench, O-ring, saw	A7S, Z9Q, B7G
113	888	2/2/2009	2, 2	clamp, pail	L8U, G1W
114	333	2/3/2009	6	nail	T3B

Record: 1 of 3

1NF?

Order1NF-Wrong4 : Table

OrderNum	CustomerNum	OrderDate	Qty1	Item1	ProdNum1	Qty2	Item2	ProdNum2	Qty3	Item3	ProdNum3
112	444	2/1/2009	2	wrench	A7S	3	O-ring	Z9Q	1	saw	B7G
113	888	2/2/2009	2	clamp	L8U	2	pail	G1W			
114	333	2/3/2009	6	nail	T3B						
			0			0			0		

Record: 1 of 3

1NF?

Order1NF-Wrong5 : Table

OrderNum	CustomerNum	OrderDate	Qty	wrench (A7S)	O-ring (Z9Q)	saw (B7G)	clamp (L8U)	pail (G1W)	nail (T3B)
112	444	2/1/2009	2	✓					
113	888	2/2/2009	2				✓	✓	
114	333	2/3/2009	6						✓
112	444	2/1/2009	3		✓				
112	444	2/1/2009	1			✓			
113	888	2/2/2009	2					✓	
			0						

Record: 1 of 6

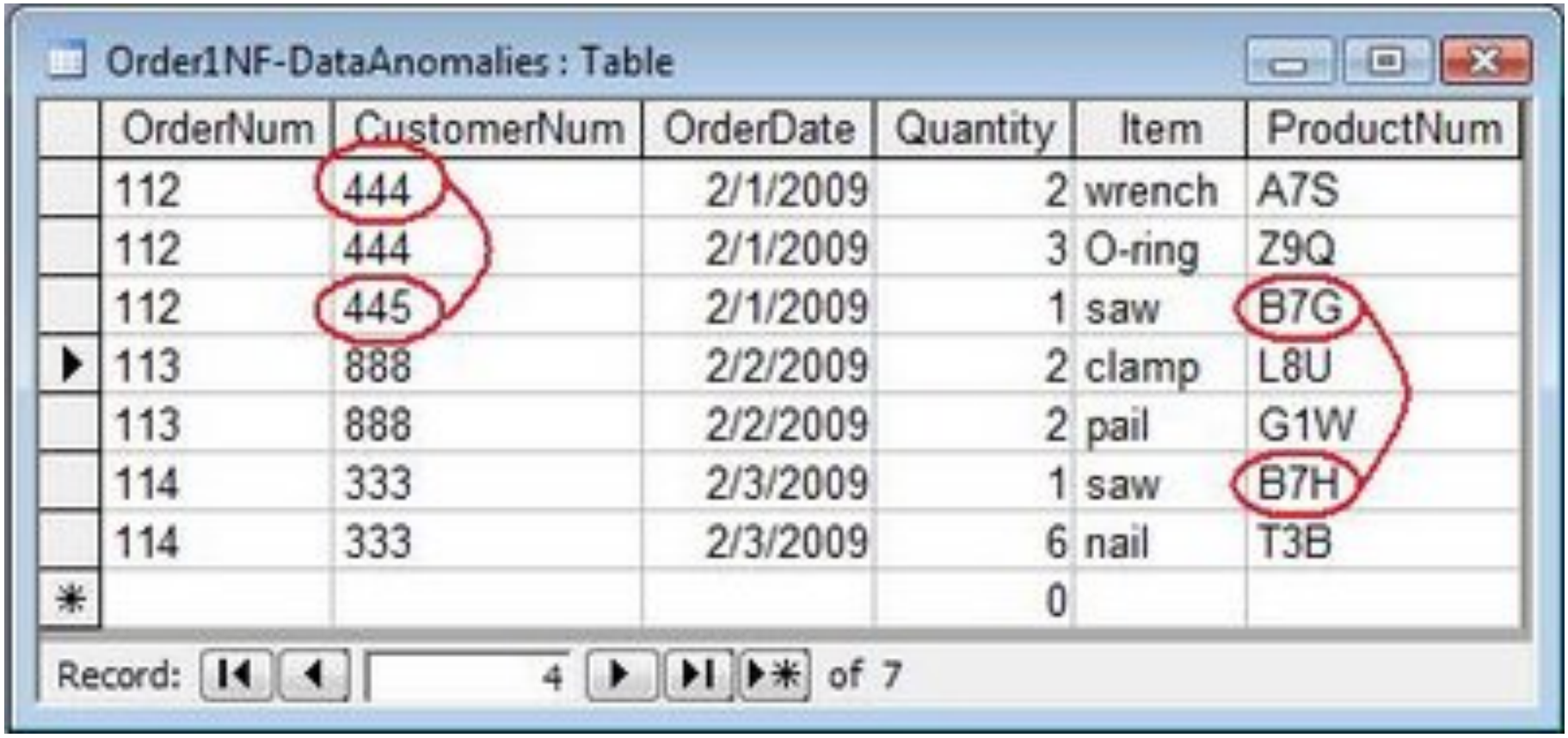
1NF?

Order1NF-Correct : Table

OrderNum	CustomerNum	OrderDate	Quantity	Item	ProductNum
112	444	2/1/2009	2	wrench	A7S
112	444	2/1/2009	3	O-ring	Z9Q
112	444	2/1/2009	1	saw	B7G
113	888	2/2/2009	2	clamp	L8U
113	888	2/2/2009	2	pail	G1W
114	333	2/3/2009	1	saw	B7G
114	333	2/3/2009	6	nail	T3B
			0		

Record: 1 of 7

Data anomalies

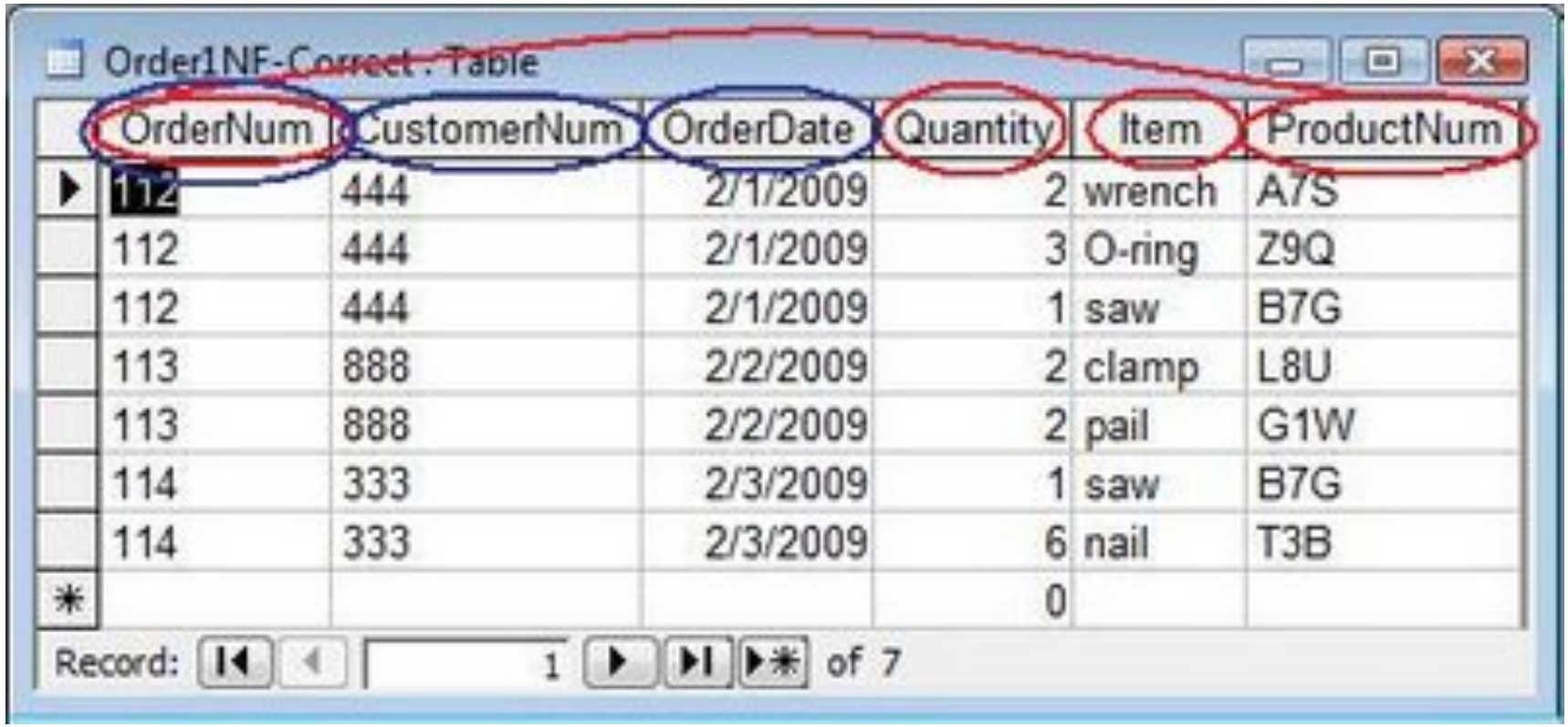


Order1NF-DataAnomalies : Table

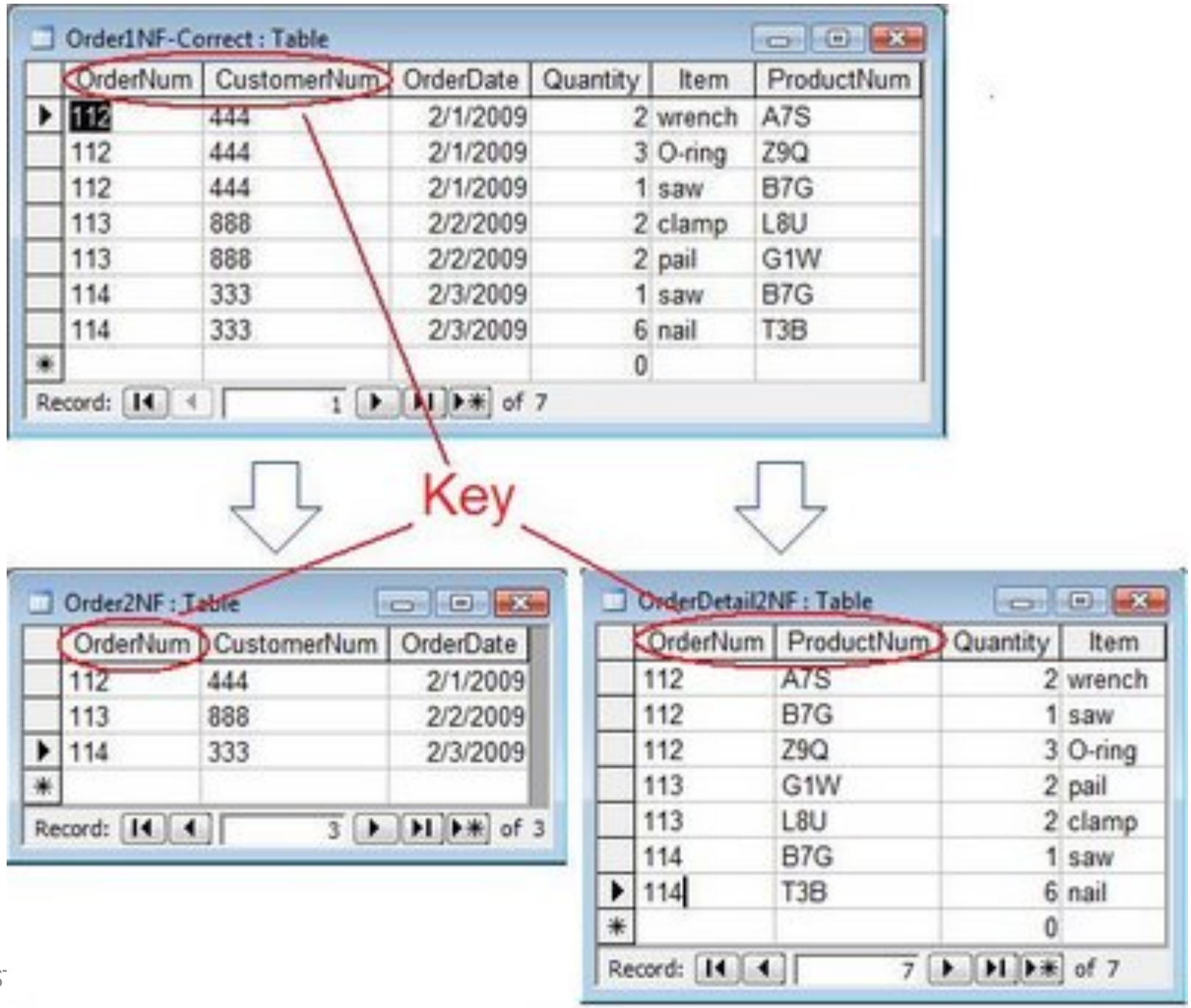
OrderNum	CustomerNum	OrderDate	Quantity	Item	ProductNum
112	444	2/1/2009	2	wrench	A7S
112	444	2/1/2009	3	O-ring	Z9Q
112	445	2/1/2009	1	saw	B7G
113	888	2/2/2009	2	clamp	L8U
113	888	2/2/2009	2	pail	G1W
114	333	2/3/2009	1	saw	B7H
114	333	2/3/2009	6	nail	T3B
			0		

Record: 4 of 7

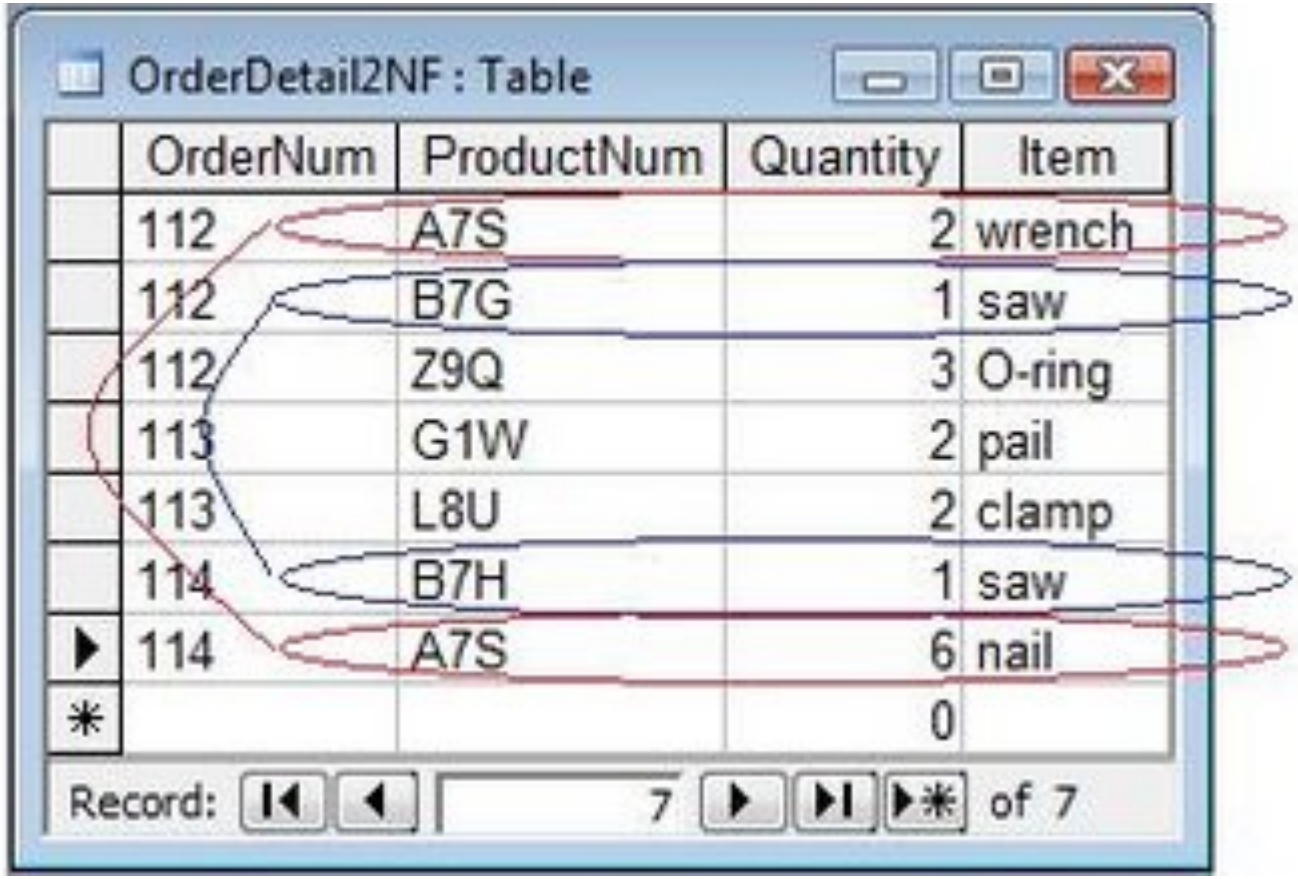
Functional dependency



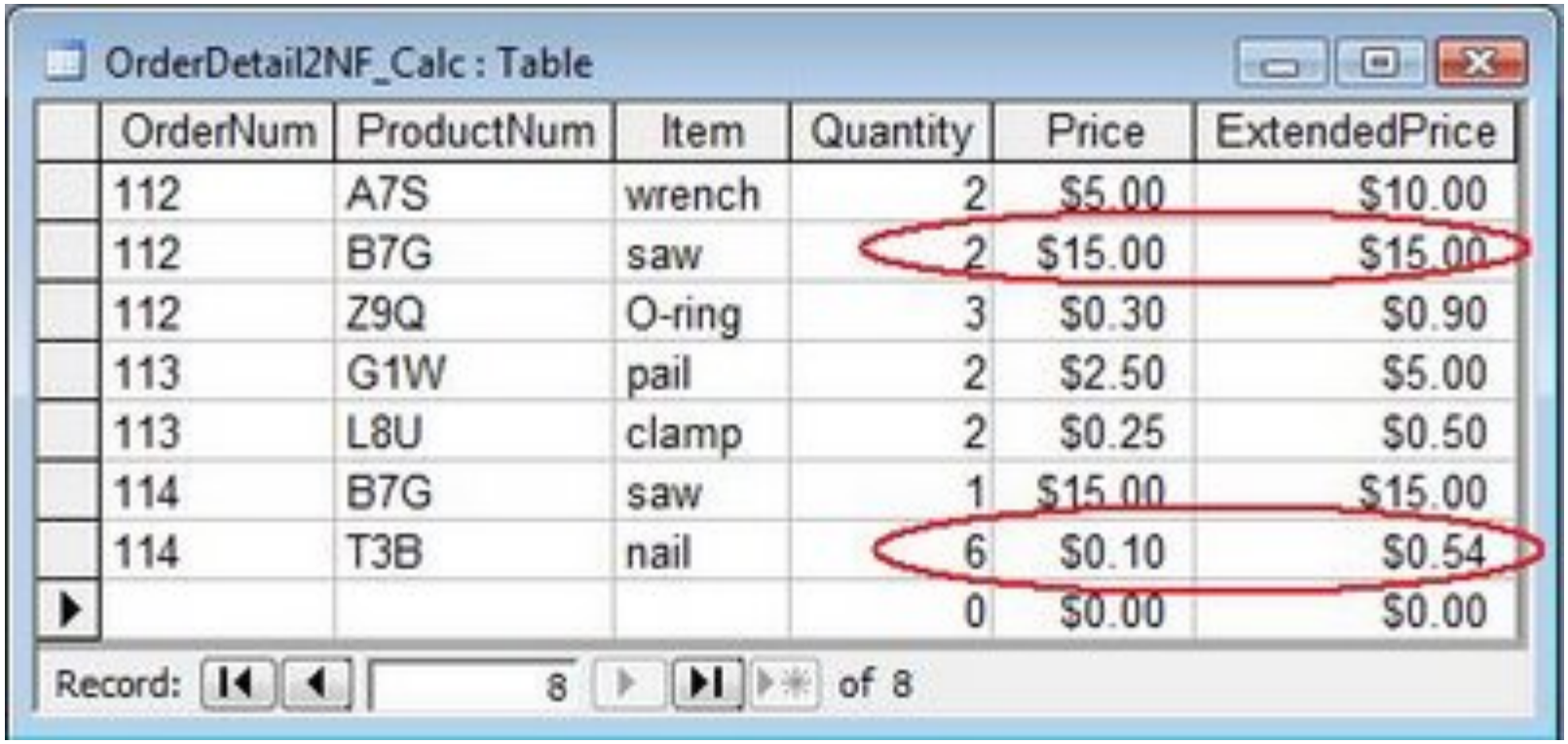
2NF


Order1NF-Correct : Table
OrderNum
CustomerNum
OrderDate
Quantity
Item
ProductNum
112 444 2/1/2009 2 wrench A7S
112 444 2/1/2009 3 O-ring Z9Q
112 444 2/1/2009 1 saw B7G
113 888 2/2/2009 2 clamp L8U
113 888 2/2/2009 2 pail G1W
114 333 2/3/2009 1 saw B7G
114 333 2/3/2009 6 nail T3B
0
Record: 1 of 7
Key
Order2NF : Table
OrderNum
CustomerNum
OrderDate
112 444 2/1/2009
113 888 2/2/2009
114 333 2/3/2009
Record: 3 of 3
OrderDetail2NF : Table
OrderNum
ProductNum
Quantity
Item
112 A7S 2 wrench
112 B7G 1 saw
112 Z9Q 3 O-ring
113 G1W 2 pail
113 L8U 2 clamp
114 B7G 1 saw
114 T3B 6 nail
0
Record: 7 of 7

Any issue?


OrderDetail2NF : Table
OrderNum
ProductNum
Quantity
Item
112
A7S
2
wrench
112
B7G
1
saw
112
Z9Q
3
O-ring
113
G1W
2
pail
113
L8U
2
clamp
114
B7H
1
saw
114
A7S
6
nail
0
Record:
7
of 7

Any issue?

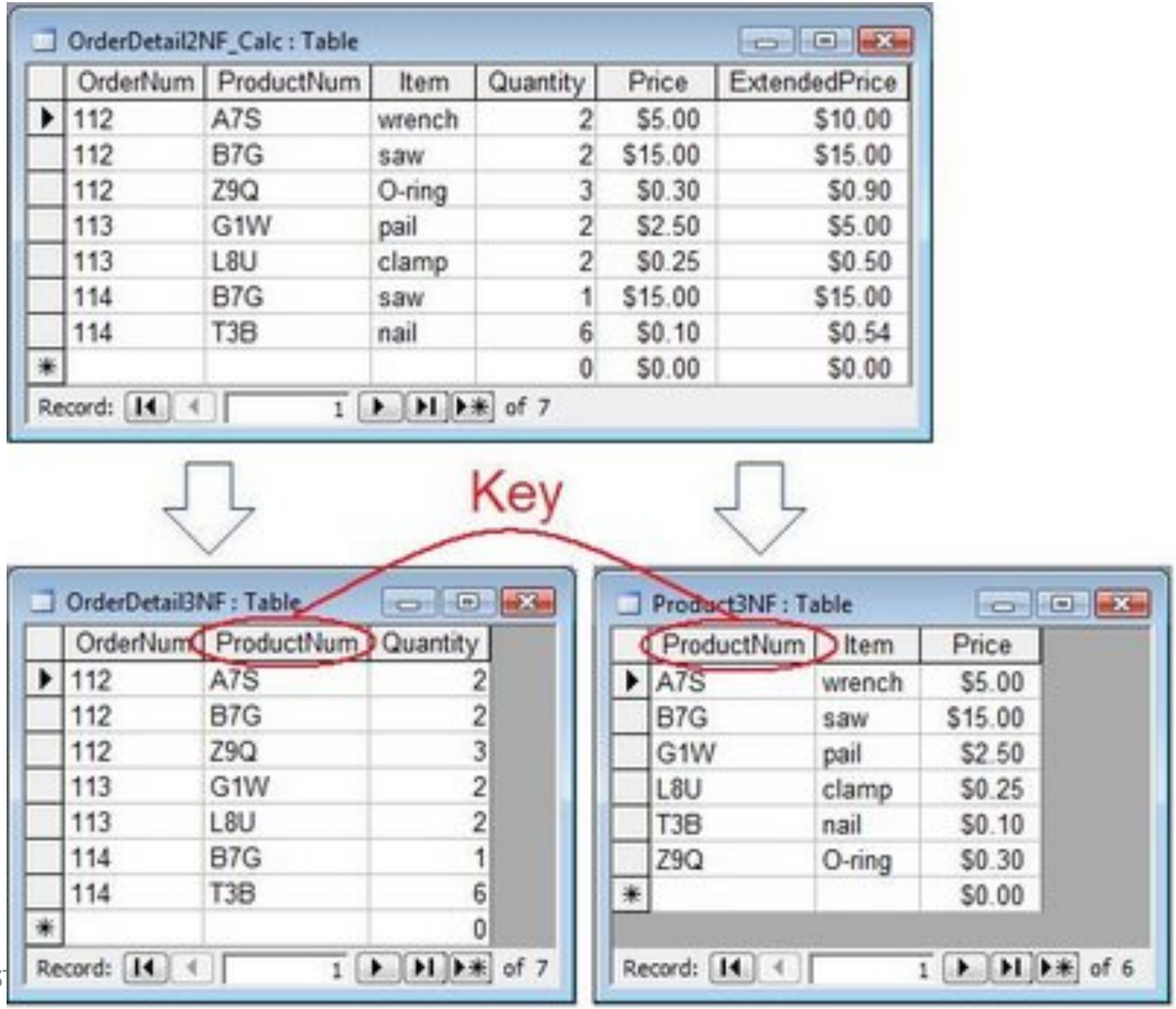


OrderDetail2NF_Calc : Table

OrderNum	ProductNum	Item	Quantity	Price	ExtendedPrice
112	A7S	wrench	2	$5.00	$10.00
112	B7G	saw	2	$15.00	$15.00
112	Z9Q	O-ring	3	$0.30	$0.90
113	G1W	pail	2	$2.50	$5.00
113	L8U	clamp	2	$0.25	$0.50
114	B7G	saw	1	$15.00	$15.00
114	T3B	nail	6	$0.10	$0.54
			0	$0.00	$0.00

Record: 8 of 8

3NF


OrderDetail2NF_Calc : Table
OrderNum | ProductNum | Item | Quantity | Price | ExtendedPrice
112 | A7S | wrench | 2 | $5.00 | $10.00
112 | B7G | saw | 2 | $15.00 | $15.00
112 | Z9Q | O-ring | 3 | $0.30 | $0.90
113 | G1W | pail | 2 | $2.50 | $5.00
113 | L8U | clamp | 2 | $0.25 | $0.50
114 | B7G | saw | 1 | $15.00 | $15.00
114 | T3B | nail | 6 | $0.10 | $0.54
0 | $0.00 | $0.00
Record: 1 of 7
Key
OrderDetail3NF : Table
OrderNum | ProductNum | Quantity
112 | A7S | 2
112 | B7G | 2
112 | Z9Q | 3
113 | G1W | 2
113 | L8U | 2
114 | B7G | 1
114 | T3B | 6
0
Record: 1 of 7
Product3NF : Table
ProductNum | Item | Price
A7S | wrench | $5.00
B7G | saw | $15.00
G1W | pail | $2.50
L8U | clamp | $0.25
T3B | nail | $0.10
Z9Q | O-ring | $0.30
$0.00
Record: 1 of 6