BACS3183
Advanced Database Management

Chapter 5

Relational Database Design

Learning Outcomes

At the end of this class, you should be able to

- **Explain normalization and its role in the database design process.**
- **Determine the functional dependency between two or more attributes that are a subset of a relation.**
- **Describe the properties of 1NF, 2NF, 3NF, BCNF.**
- **Transform lower normal forms to higher normal forms**

1. Relational Database Design

- **Relational database design requires that we find a "good" collection of relation schemas**
- **A bad design may lead to**
  - **Repetition of Information.**
    - **Leads to anomalies**
  - **Loss of information (lossy decomposition).**
- **Design Goals:**
  - **Avoid redundant data**
  - **Ensure that relationships among attributes are represented.**
  - **Facilitate the checking of updates for violation of integrity constraints**

Problems of Data Redundancy

- **Update anomalies include**
  - **Insertion**
  - **Deletion**
  - **Modification**

Problems of Data Redundancy

Modification Anomaly

The same information can be expressed on multiple rows; therefore modification to the table may result in logical inconsistencies

Employees' Skills

Employee ID	Employee Address	Skill
426	87 Sycamore Grove	Typing
426	87 Sycamore Grove	Shorthand
519	94 Chestnut Street	Public Speaking
519	96 Walnut Avenue	Carpentry

Problems of Data Redundancy

Insertion Anomaly

There are circumstances in which certain facts cannot be recorded at all

Faculty and Their Courses

Faculty ID	Faculty Name	Faculty Hire Date	Course Code
389	Dr. Giddens	10-Feb-1985	ENG-206
407	Dr. Saperstein	19-Apr-1999	CMP-101
407	Dr. Saperstein	19-Apr-1999	CMP-201
424	Dr. Newsome	29-Mar-2007	?

Until the new faculty member, Dr. Newsome, is assigned to teach at least one course, his details cannot be recorded

Problems of Data Redundancy

Deletion Anomaly

Under certain circumstances, deletion of data representing certain facts necessitates deletion of data representing completely different facts.

Faculty and Their Courses

Faculty ID	Faculty Name	Faculty Hire Date	Course Code
389	Dr. Giddens	10-Feb-1985	ENG-206
407	Dr. Saperstein	19-Apr-1999	CMP-101
407	Dr. Saperstein	19-Apr-1999	CMP-201

DELETE

All information about Dr. Giddens is lost if he is temporarily stopped to be assigned to any courses.

2. Functional Dependency

- Vital for the redesign of database schemas to eliminate redundancy
- Enable systematic improvement of database designs

Functional Dependency

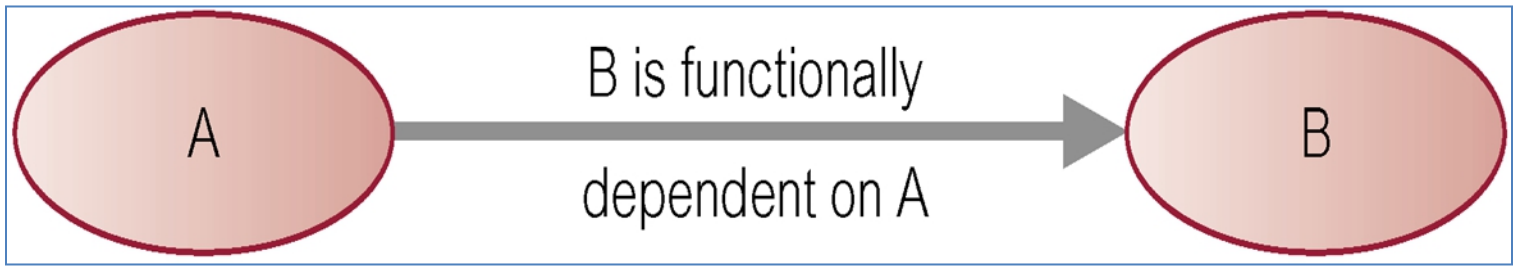
- **Functional dependency describes relationship between attributes.**
- **For example,**

  If A and B are attributes of relation R,

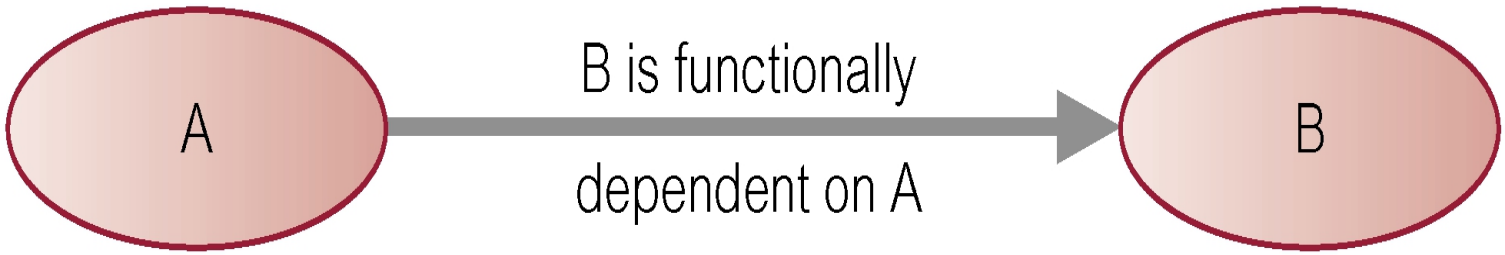
  B is functionally dependent on A

  (denoted A → B),

  if each value of A in R is associated with exactly one value of B in R.



Functional Dependency



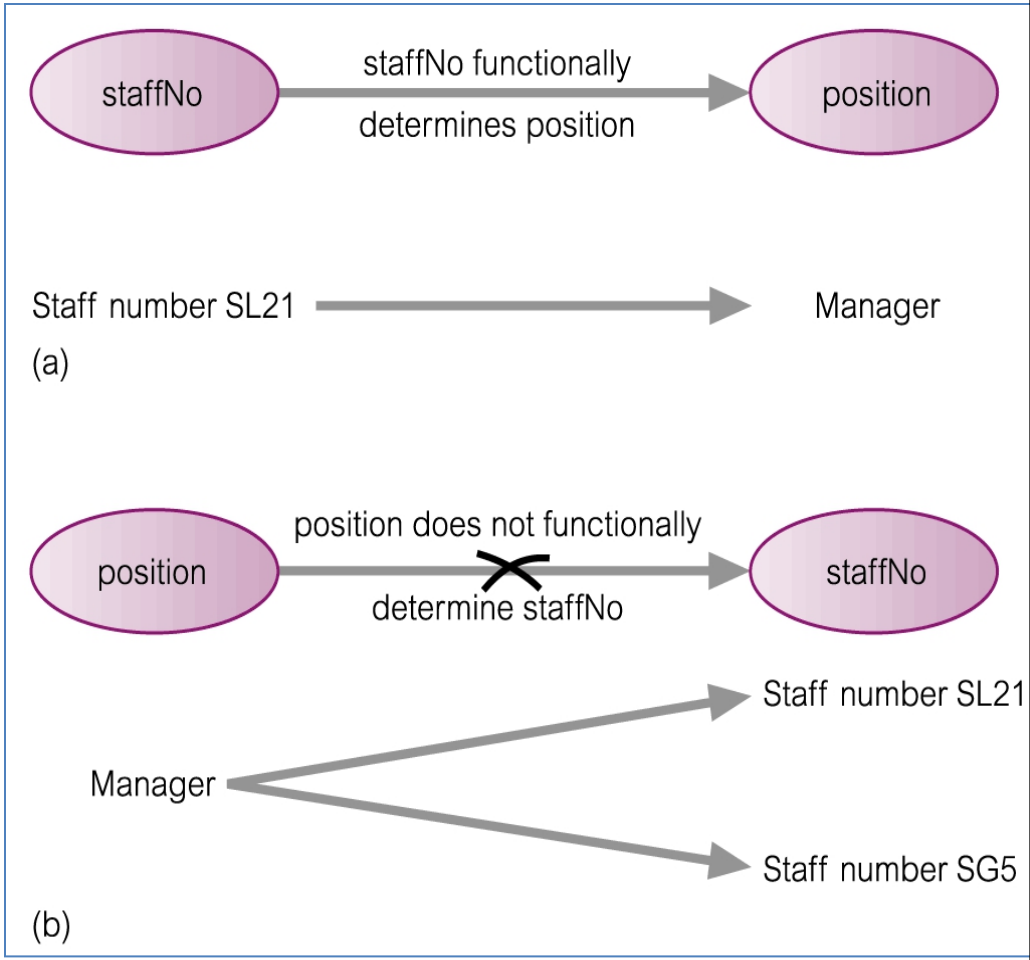
- **The *determinant* of a functional dependency refers to the attribute or group of attributes on the left-hand side of the arrow.**

Functional Dependency

***A → B* (B is functionally dependent on A)**
if each value of A in R is associated with
exactly one value of B in R.

Staff

staffNo	sName	position	salary	branchNo
SL21	John White	Manager	30000	B005
SG37	Ann Beech	Assistant	12000	B003
SG14	David Ford	Supervisor	18000	B003
SA9	Mary Howe	Assistant	9000	B007
SG5	Susan Brand	Manager	24000	B003
SL41	Julie Lee	Assistant	9000	B005



Example Functional Dependency

Staff

staffNo	sName	position	salary	branchNo
SL21	John White	Manager	30000	B005
SG37	Ann Beech	Assistant	12000	B003
SG14	David Ford	Supervisor	18000	B003
SA9	Mary Howe	Assistant	9000	B007
SG5	Susan Brand	Manager	24000	B003
SL41	Julie Lee	Assistant	9000	B005

Branch

branchNo	bAddress
B005	22 Deer Rd, London
B007	16 Argyll St, Aberdeen
B003	163 Main St, Glasgow

Staff Branch

staffNo	sName	position	salary	branchNo	bAddress
SL21	John White	Manager	30000	B005	22 Deer Rd, London
SG37	Ann Beech	Assistant	12000	B003	163 Main St, Glasgow
SG14	David Ford	Supervisor	18000	B003	163 Main St, Glasgow
SA9	Mary Howe	Assistant	9000	B007	16 Argyll St, Aberdeen
SG5	Susan Brand	Manager	24000	B003	163 Main St, Glasgow
SL41	Julie Lee	Assistant	9000	B005	22 Deer Rd, London

Example Functional Dependency that holds for all Time

- Consider the values shown in staffNo and sName attributes of the Staff relation
- The functional dependency is

staffNo → sName

sName → staffNo ??

Staff

staffNo	sName	position	salary	branchNo
SL21	John White	Manager	30000	B005
SG37	Ann Beech	Assistant	12000	B003
SG14	David Ford	Supervisor	18000	B003
SA9	Mary Howe	Assistant	9000	B007
SG5	Susan Brand	Manager	24000	B003
SL41	Julie Lee	Assistant	9000	B005

Full Functional Dependency

Full functional dependency indicates that if A and B are attributes of a relation, B is fully functionally dependent on A, if B is functionally dependent on A, but not on any proper subset of A.

- **Exists in the Staff relation : staffNo, sName → branchNo**
- **True - each value of (staffNo, sName) is associated with a single value of branchNo.**
- **However, branchNo is also functionally dependent on a subset of (staffNo, sName), namely staffNo. Example above is a *partial dependency*.**

staffNo	sName	position	salary	branchNo
SL21	John White	Manager	30000	B005
SG37	Ann Beech	Assistant	12000	B003
SG14	David Ford	Supervisor	18000	B003
SA9	Mary Howe	Assistant	9000	B007
SG5	Susan Brand	Manager	24000	B003
SL41	Julie Lee	Assistant	9000	B005

Transitive Dependency

- A transitive dependency in a relation can potentially cause update anomalies.

- Transitive dependency describes a condition where A, B, and C are attributes of a relation such that

  if A → B and B → C, then

  C is transitively dependent on A via B

  (provided that A is not functionally dependent on B or C).

Example Transitive Dependency

Staff Branch

staffNo	sName	position	salary	branchNo	bAddress
SL21	John White	Manager	30000	B005	22 Deer Rd, London
SG37	Ann Beech	Assistant	12000	B003	163 Main St, Glasgow
SG14	David Ford	Supervisor	18000	B003	163 Main St, Glasgow
SA9	Mary Howe	Assistant	9000	B007	16 Argyll St, Aberdeen
SG5	Susan Brand	Manager	24000	B003	163 Main St, Glasgow
SL41	Julie Lee	Assistant	9000	B005	22 Deer Rd, London

staffNo → sName, position, salary, branchNo, bAddress

branchNo → bAddress

Transitive dependency,

staffNo → branchNo,

branchNo → bAddress

Thus bAddress is transitively dependent on staffNo via branchNo

3. Normalization

- FDs are the basis of normalization -- a formal methodology for refining and creating good relational designs

One common schema refinement techniques - decomposition

Starts with a relational schema, and uses the FDs to guide the schema decomposition,

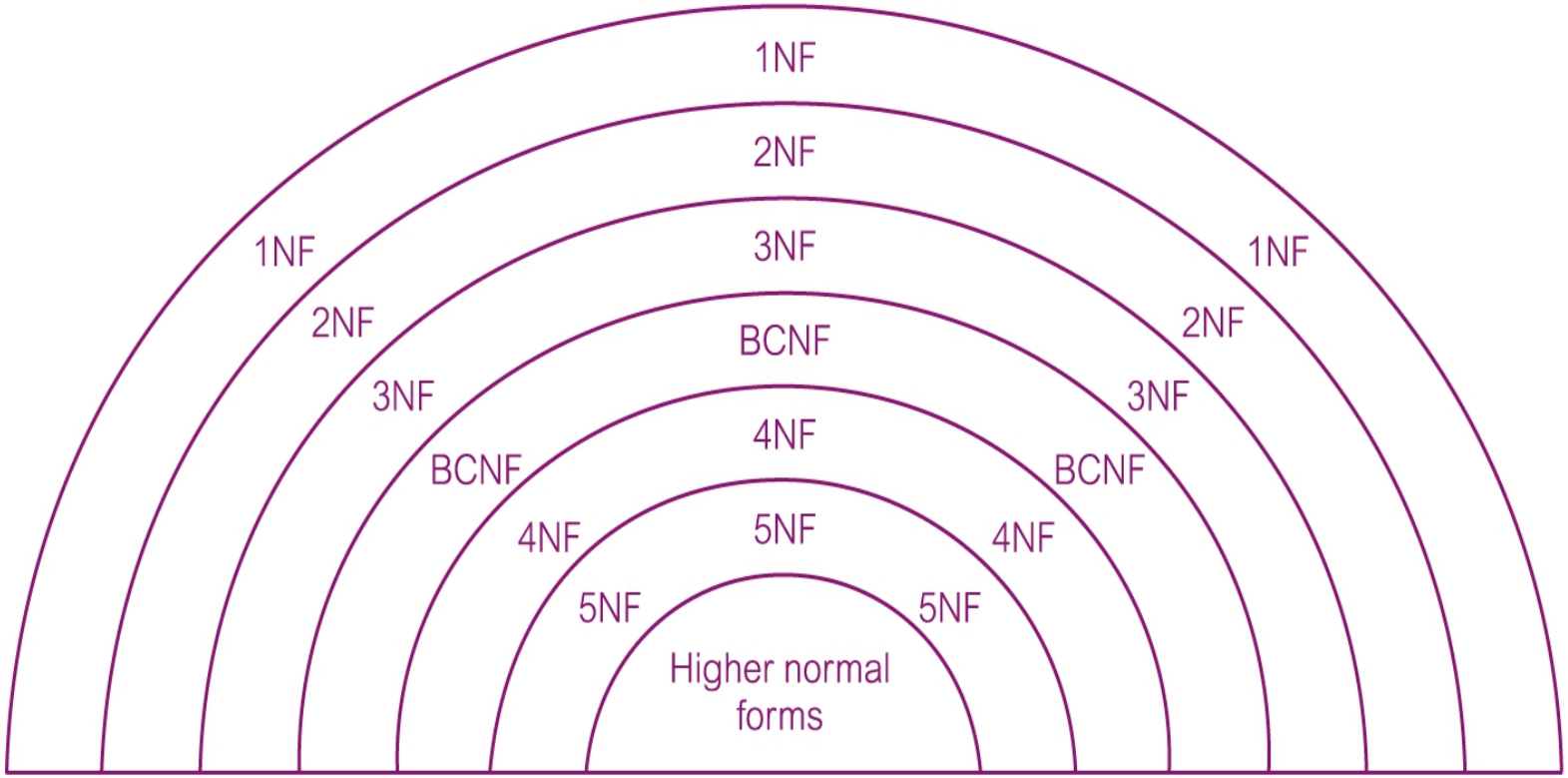
e.g., R (ABCD)

R1(AB) **R2 (BCD)**

Properties Of Decomposition

- **Two important properties:**
  - ***Lossless-join property* : the final relations must contain exactly the information contained in the original relation, without losing any nor adding any other ones.**
  - ***Dependency preservation property* : the final relations must be characterized by the same FDs as the original relation**

The Process of Normalization



Normalization is a formal methodology for refining and creating good relational designs

Data Source Example

Project Management Report

Project Code: PC010
Project Title: Pensions System

Project Manager: M Phillips **Project Budget:** £24,500

Employee No.	Employee Name	Department No.	Department Name	Hourly Rate
S10001	A Smith	L004	IT	£22.00
S10030	L Jones	L023	Pensions	£18.50
S21010	P Lewis	L004	IT	£21.00
S00232	R Smith	L003	Programming	£26.00

Total Staff on Project: 4 **Average Hourly Rate:** £21.88

Calculated Fields

Unnormalized Form (UNF)

- A table that contains one or more repeating groups.

- To create an unnormalized table
  - Transform the data from the information source (e.g. form) into table format with columns and rows.

Project Code: PC010
Project Title: Pensions System

Project Manager: M Phillips **Project Budget:** £24,500

Employee No.	Employee Name	Department No.	Department Name	Hourly Rate
S10001	A Smith	L004	IT	£22.00
S10030	L Jones	L023	Pensions	£18.50
S21010	P Lewis	L004	IT	£21.00
S00232	R Smith	L003	Programming	£26.00

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

Any Insertion, Deletion and Modification Anomalies?
Which normal form is the table in?

First Normal Form (1NF)

- **A relation in which the intersection of each row and column contains one and only one value.**

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

Not in 1NF

UNF to 1NF

- Nominate an attribute or group of attributes to act as the key for the unnormalized table.
- Identify the repeating group(s) in the unnormalized table which repeats for the key attribute(s).

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

UNF to 1NF

- Remove the repeating group by
  - Entering appropriate data into the empty columns of rows containing the repeating data ('flattening' the table).

  Or by

  - Placing the repeating data along with a copy of the original key attribute(s) into a separate relation.

‘Flattening the Table’

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

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

First Normal Form (1NF) Example

Normalized Set (1NF)

Employee'

Man#	Name	Birthdate
123	Michael Swart	Nov. 22

Children'

Man#	ChildName	BirthYear
123	Mini-me 1	2000
123	Mini-me 2	2002

JobHistory'

Man#	JobDate	Title
123	2000	Lackey
123	2002	Senior Lackey

SalaryHistory'

Man#	JobDate	SalaryDate	Salary
123	2000	2000	$1,000,000
123	2000	2001	$2,000,000
123	2002	2002	$3,000,000
123	2003	2003	$4,000,000

Is the above relations in 2NF and 3NF as well?

Second Normal Form (2NF)

- Based on the concept of full functional dependency.

- Full functional dependency indicates that

  if A and B are attributes of a relation,

  B is fully dependent on A if B is functionally dependent on A but not on any proper subset of A.

studID, studName → programmeCode

Second Normal Form (2NF)

- A relation that is
  - in 1NF and
  - every non primary-key attribute is fully functionally dependent on the primary key.

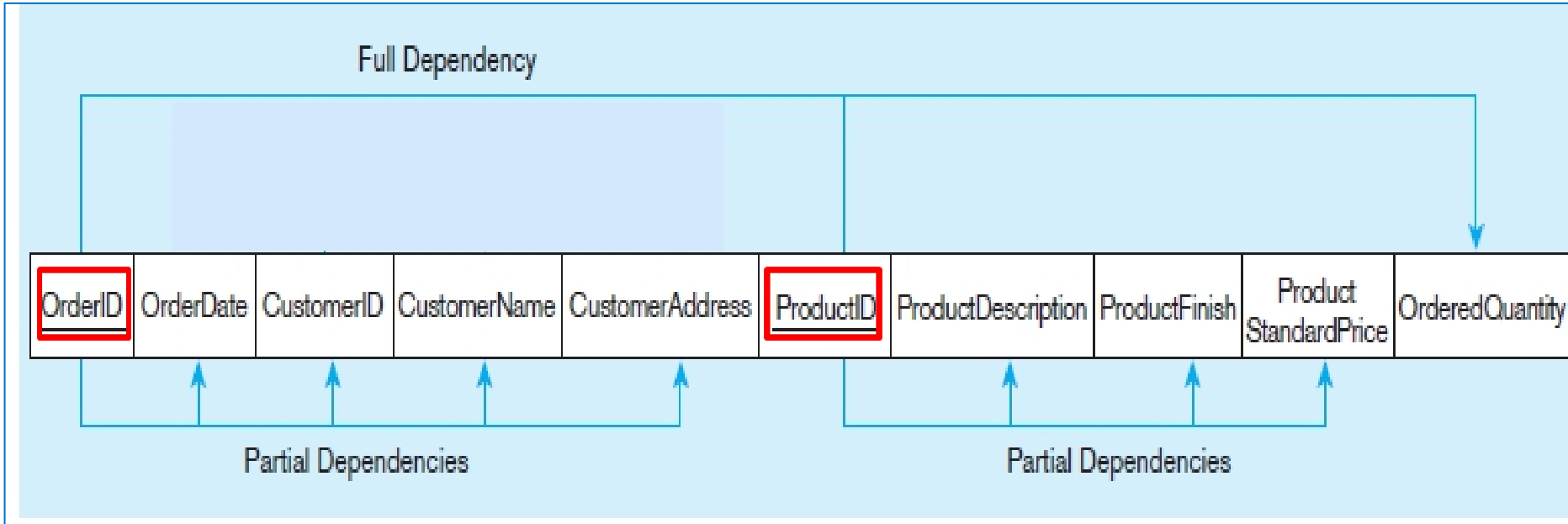
studID, studName → programmeCode

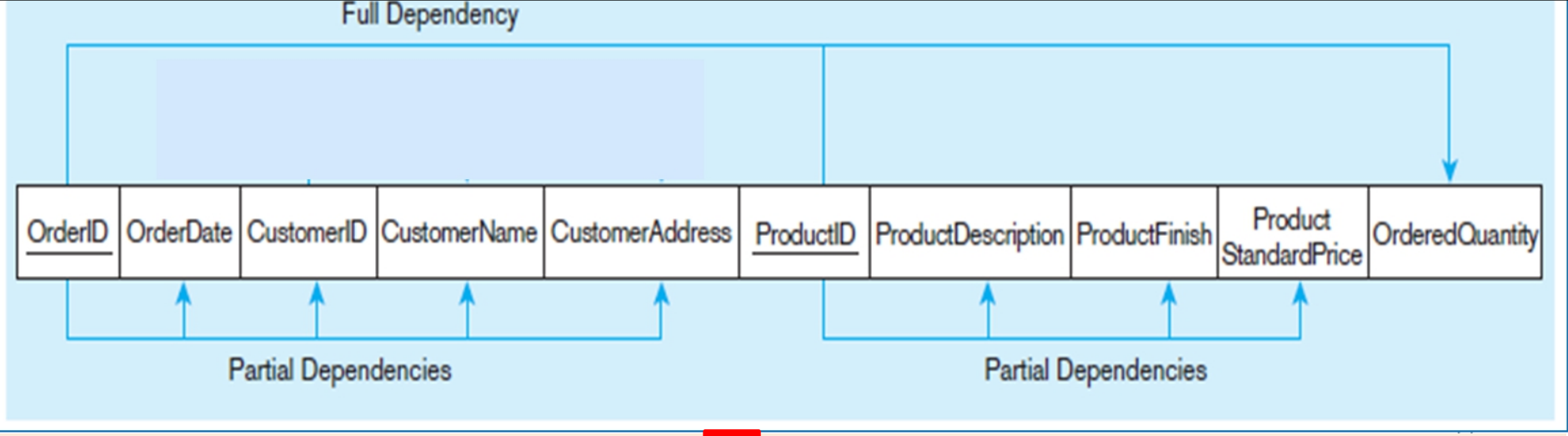
partial dependency exists

1NF to 2NF

- Identify the primary key for the 1NF relation.
- Identify the functional dependencies in the relation.
- If partial dependencies exist on the primary key remove them by placing them in
  a new relation along with a copy of their determinant.

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



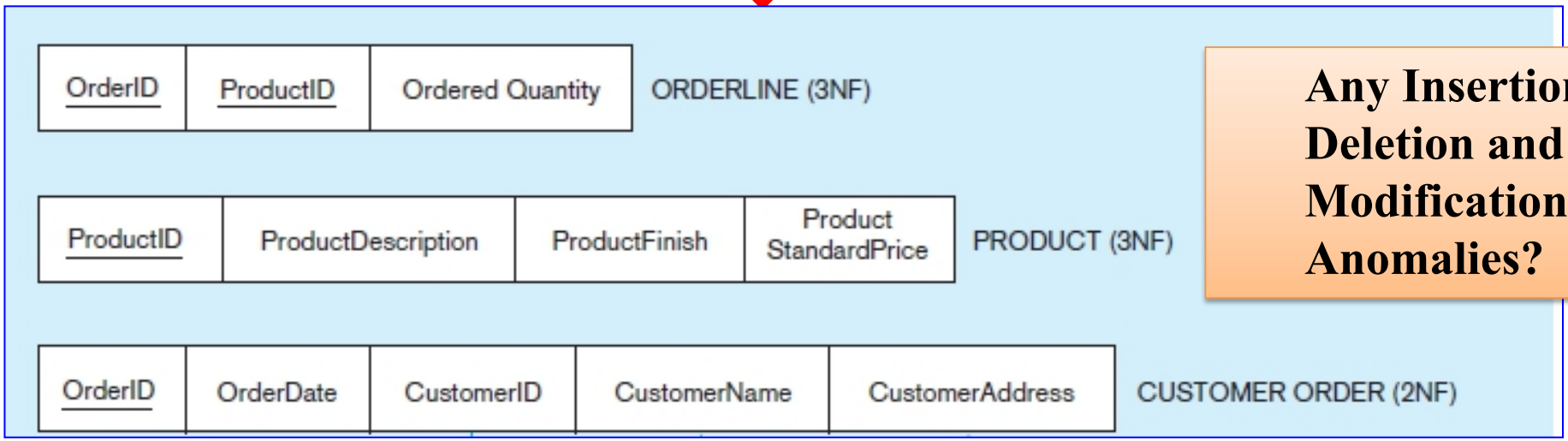


OrderID, ProductID → OrderQu tity

OrderID → OrderDate, Custome D, CustomerName, CustomerAddress

ProductID → ProductDescriptio roductFinish, ProductStandardPrice

Remove partial dependency



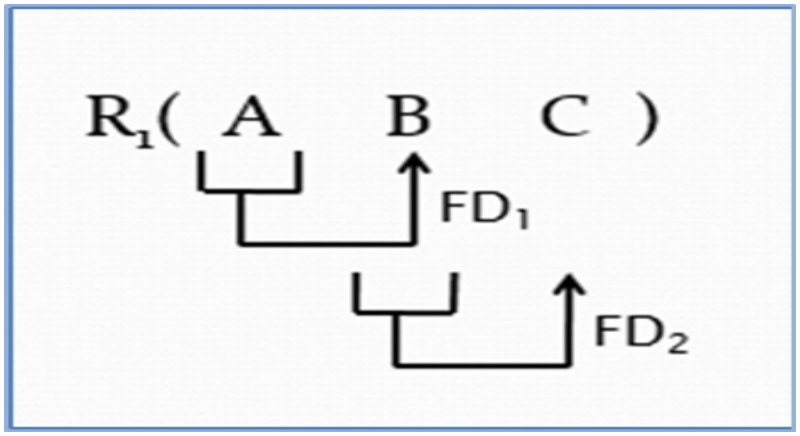
Any Insertion, Deletion and Modification Anomalies?

Third Normal Form (3NF)

- Based on the concept of transitive dependency.
- Transitive Dependency is a condition where
  - A, B and C are attributes of a relation such that

    if $A \rightarrow B$ and $B \rightarrow C$,

    then C is transitively dependent on A through B. (Provided that A is not functionally dependent on B or C).

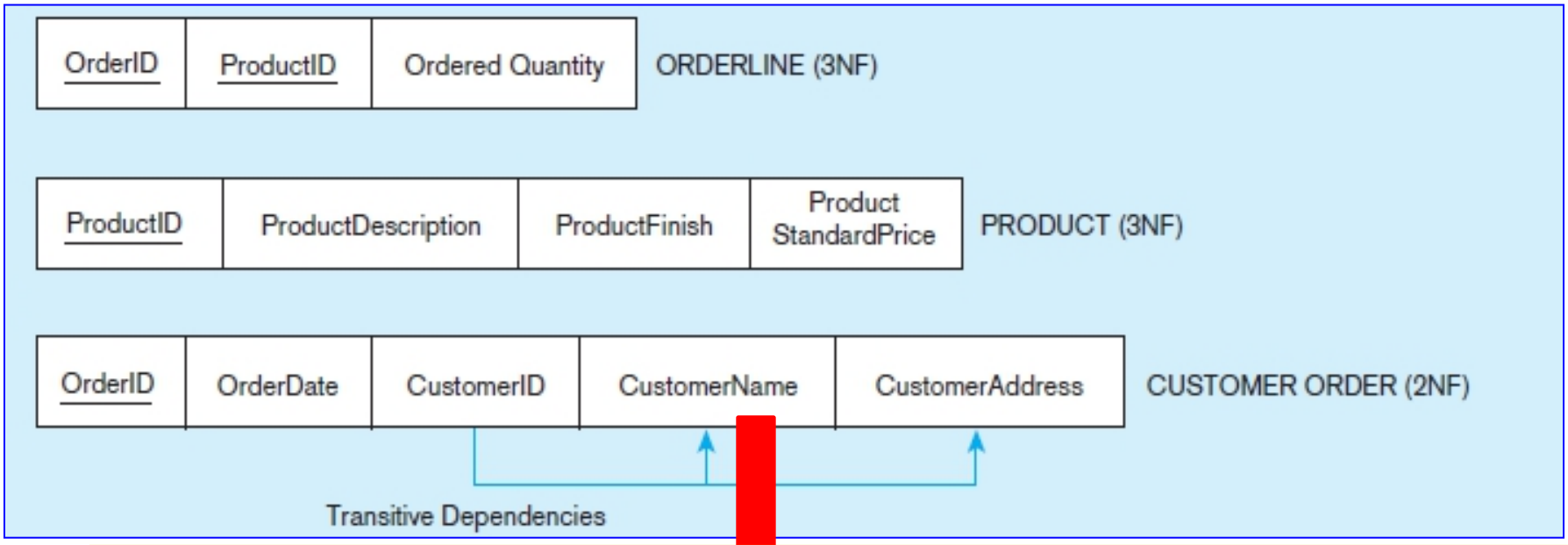


Third Normal Form (3NF)

- A relation that is
  - in 1NF and 2NF and
  - in which no non primary-key attribute is transitively dependent on the primary key.

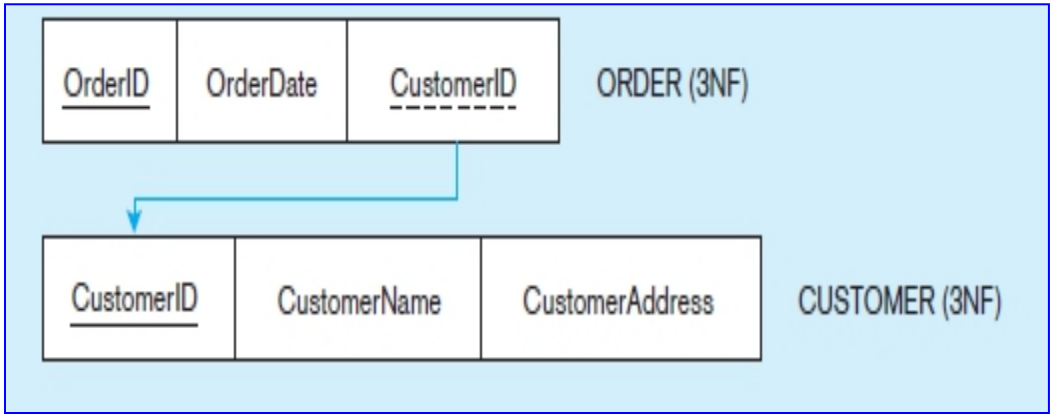
2NF to 3NF

- Identify the primary key in the 2NF relation.
- Identify functional dependencies in the relation.
- If transitive dependencies exist on the primary key remove them by placing them in a new relation along with a copy of their determinant.



OrderID → OrderDate, CustomerID
CustomerID → CustomerName, CustomerAddress

Remove transitive dependency



General Definitions of 2NF and 3NF

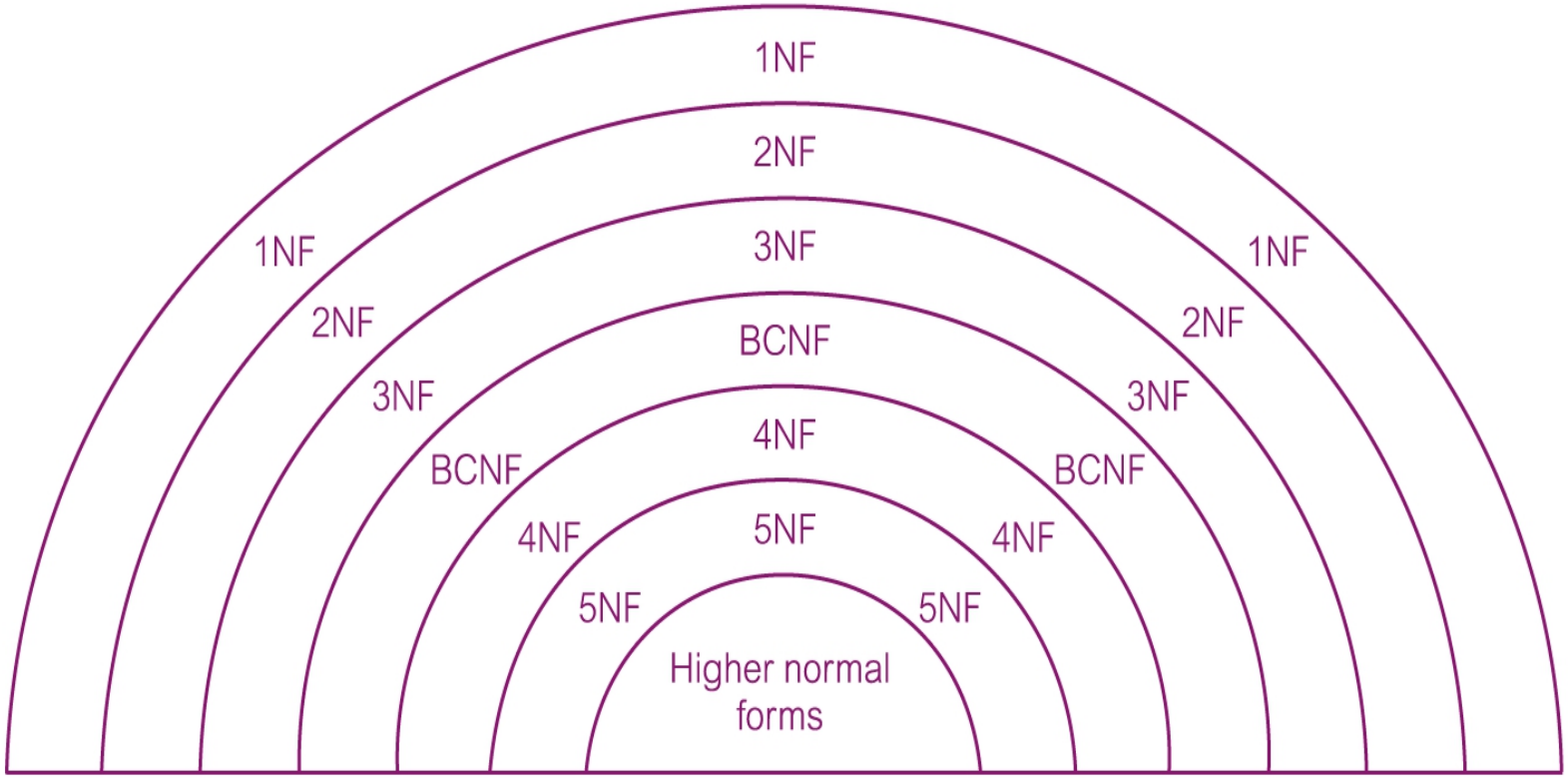
- **Second normal form (2NF)**
  - A relation that is in 1NF and every non primary-key attribute is fully functionally dependent on *any candidate key*.

- **Third normal form (3NF)**
  - A relation that is in 1NF and 2NF and in which no non primary-key attribute is transitively dependent on *any candidate key*.

Superkey – an attribute or set of attributes that uniquely identifies a tuple within a relation.

Candidate key – a minimal superkey

The Process of Normalization



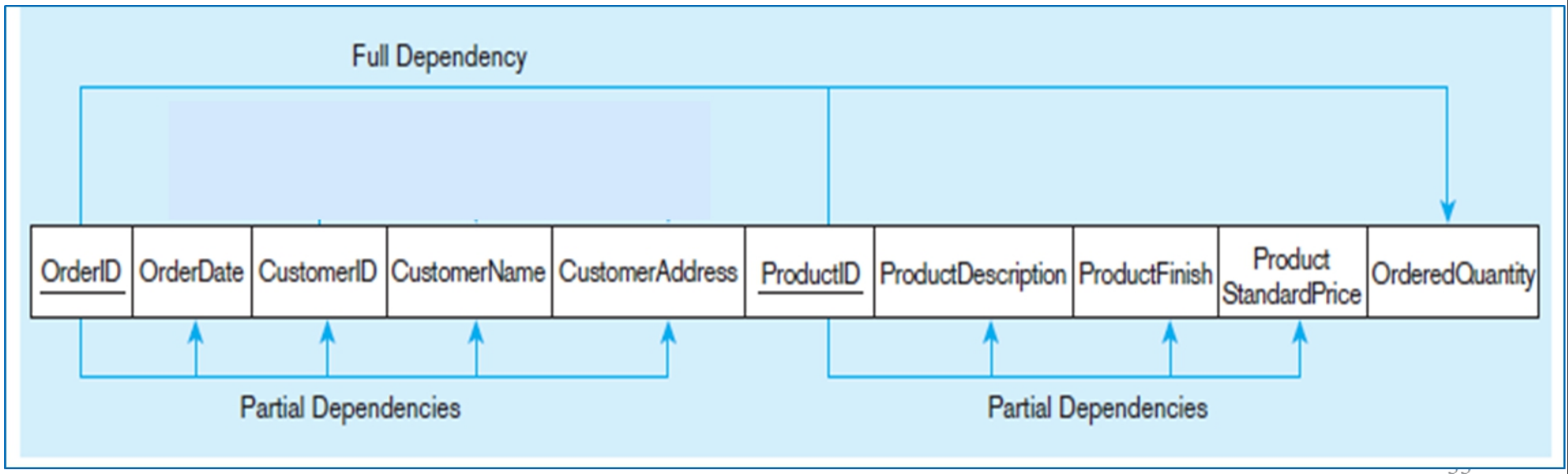
Normalization is a formal methodology for refining and creating good relational designs

UNF to 1NF

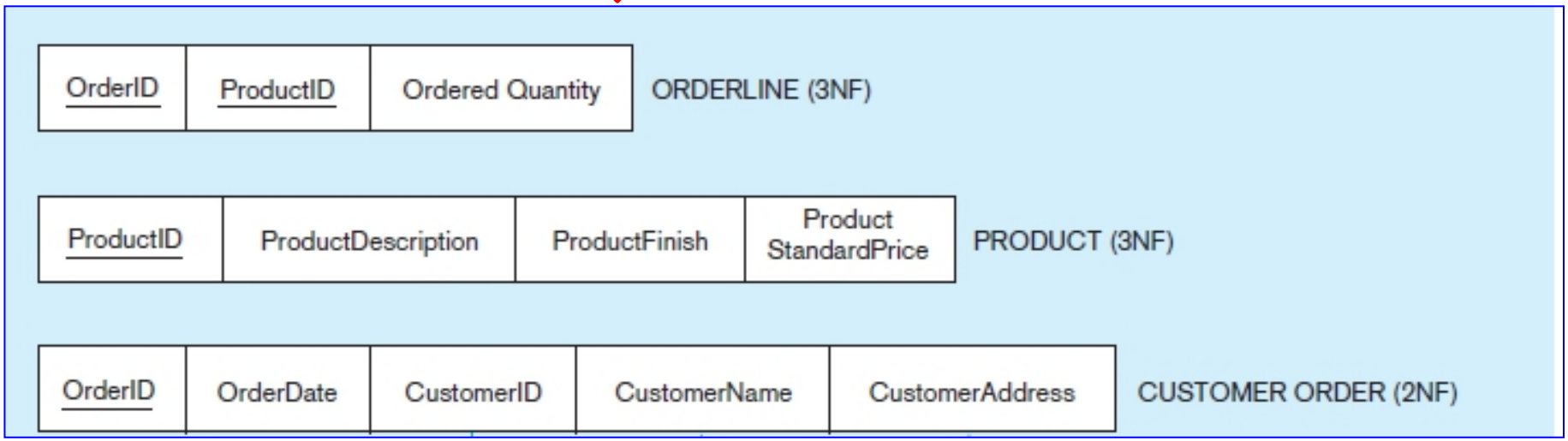
OrderID	Order Date	Customer ID	Customer Name	Customer Address	ProductID	Product Description	Product Finish	Product StandardPrice	Ordered Quantity
1006	10/24/2010	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
					5	Writer's Desk	Cherry	325.00	2
					4	Entertainment Center	Natural Maple	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

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

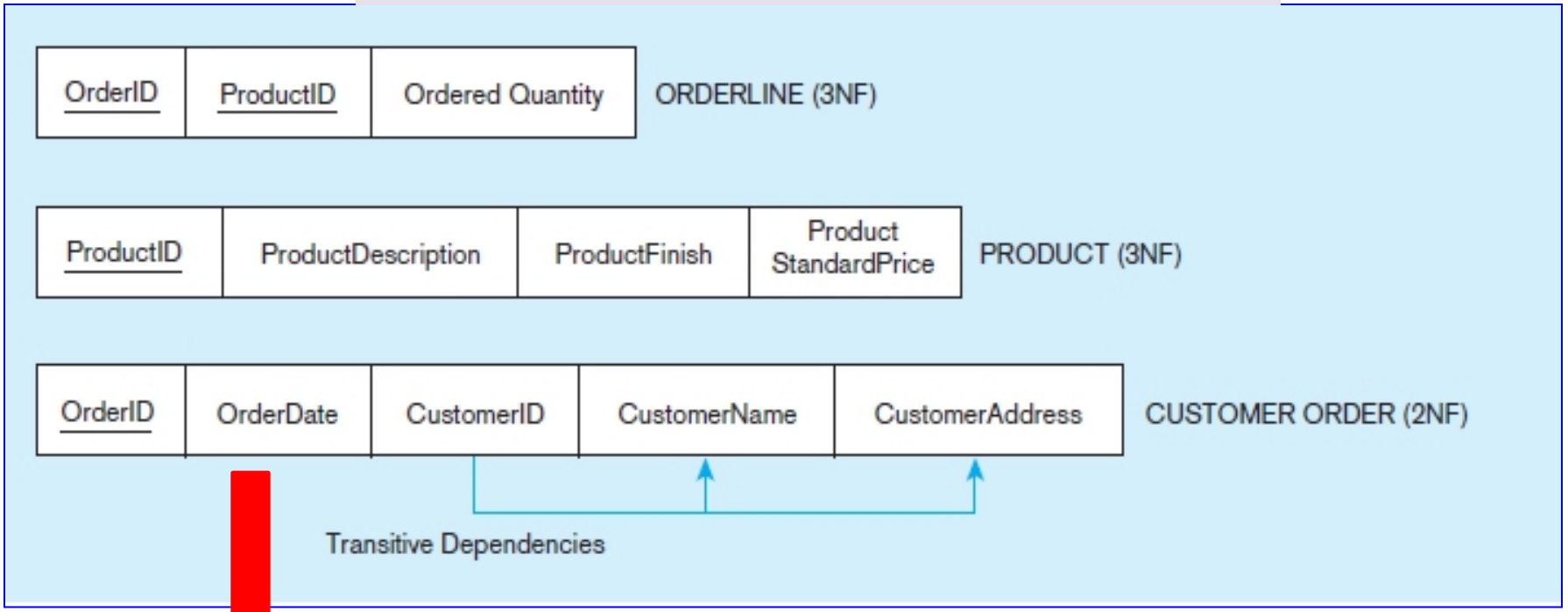
1NF to 2NF



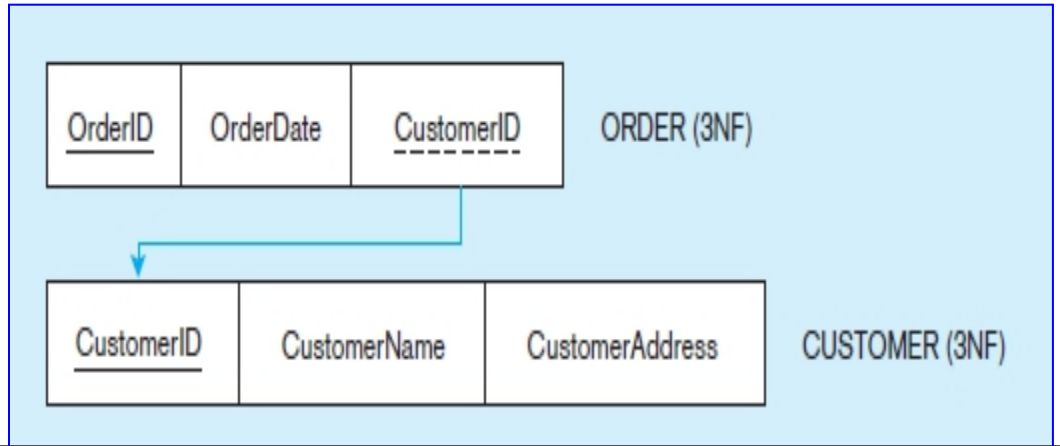
Remove partial dependency



2NF to 3NF



Remove transitive dependency



Boyce–Codd Normal Form (BCNF)

– A relation is in BCNF if and only if **every determinant is a candidate key.**

STUDENT_ADVISOR

	StudentID	Subject	AdvisorName
1	100	Math	Cauchy
2	200	Psychology	Jung
3	300	Math	Riemann
4	400	Math	Cauchy
5	500	Psychology	Perls
6	600	English	Austin
7	700	Psychology	Perls
8	700	Math	Riemann
9	800	Math	Cauchy
10	800	Psychology	Jung

- ✓ A student can take more than 1 majors
- ✓ A major can have several faculty members as advisors
- ✓ A faculty member advises in only one major area.

StudentID, Subject → AdvisorName

AdvisorName → Subject

AdvisorName is not a candidate key

Boyce–Codd Normal Form (BCNF)

STUDENT_ADVISOR

	StudentID	Subject	AdvisorName
1	100	Math	Cauchy
2	200	Psychology	Jung
3	300	Math	Riemann
4	400	Math	Cauchy
5	500	Psychology	Perls
6	600	English	Austin
7	700	Psychology	Perls
8	700	Math	Riemann
9	800	Math	Cauchy
10	800	Psychology	Jung

- ✓ **A student can take more than 1 majors**
- ✓ **A major can have several faculty members as advisors**
- ✓ **A faculty member advises in only one major area.**

Is this table in 3NF?

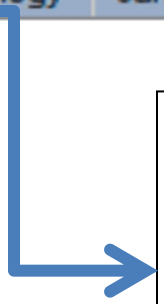
Any Insertion, Deletion and Modification Anomalies?

STUDENT_ADVISOR

	StudentID	Subject	AdvisorName
1	100	Math	Cauchy
2	200	Psychology	Jung
3	300	Math	Riemann
4	400	Math	Cauchy
5	500	Psychology	Perls
6	600	English	Austin
7	700	Psychology	Perls
8	700	Math	Riemann
9	800	Math	Cauchy
10	800	Psychology	Jung

StudentID, Subject → AdvisorName

AdvisorName → Subject



STUDENT_ADVISOR

	StudentID	AdvisorName
1	100	Cauchy
2	200	Jung
3	300	Riemann
4	400	Cauchy
5	500	Perls
6	600	Austin
7	700	Perls
8	700	Riemann
9	800	Cauchy
10	800	Jung

ADVISOR_SUBJECT

	AdvisorName	Subject
1	Austin	English
2	Cauchy	Math
3	Jung	Psychology
4	Perls	Psychology
5	Riemann	Math

References

- *Database Systems: A Practical Approach to Design, Implementation and Management.* Connolly, T. M. and Begg, C. E. .
- *Modern Database Management.* Hoffer, J.A., Prescott, M., and McFadden,F.
- *Database System Concepts* Silberschatz, A., Korth, H,. and Sudarshan, S.
- *Database Processing Fundamentals, Design & Implementation.* Kroenke D.