

Session 3-5 Normalization

Kyunghee Lee, PhD



WAYNE STATE
Mike Ilitch School of Business

Relation Anomalies to Cause Data Duplication

- **Insertion Anomaly:** adding new rows forces user to create duplicate data
- **Deletion Anomaly:** deleting rows may cause a loss of data that would be needed for other future rows
- **Update (Modification) Anomaly:** changing data in a row forces changes to other rows because of duplication



Normalization to Avoid Unnecessary Duplication

- The process of decomposing relations with anomalies to produce smaller, **well-structured** relations
 - A relation that contains **minimal data redundancy** and allows users to insert, delete, and update rows without causing data inconsistencies
- General rule of thumb: A table should not pertain to more than one entity type



Is This Table Well-structured?

<u>Emp_ID</u>	Name	Dept_Name	Salary	Course_Title	Date_Completed
100	Margaret Simpson	Marketing	48,000	SPSS	6/19/200X
100	Margaret Simpson	Marketing	48,000	Surveys	10/7/200X
140	Alan Beeton	Accounting	52,000	Tax Acc	12/8/200X
110	Chris Lucero	Info Systems	43,000	Visual Basic	1/12/200X
110	Chris Lucero	Info Systems	43,000	C++	4/22/200X
190	Lorenzo Davis	Finance	55,000		
150	Susan Martin	Marketing	42,000	SPSS	6/19/200X
150	Susan Martin	Marketing	42,000	Java	8/12/200X

Insertion–can't enter a new class without having some employee take that class

Deletion–if we remove employee 140, we lose information about the existence of a Tax Acc class

Modification–giving a salary increase to employee 100 forces us to update multiple records



WAYNE STATE

Mike Ilitch School of Business

Anomaly Example

StaffBranch

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

If StaffNo is primary key
what happens if we insert a
new branch with no staff
yet?

INSERT ANOMALY

What if the staffNO
SA9 left the organization?
What would happen to
Branch B007?

DELETE ANOMALY

What if
someone spelled
'Glasgow' as
'Glasgoaw' in
some rows?

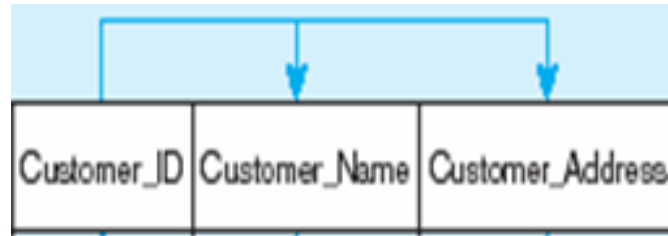
UPDATE ANOMALY



WAYNE STATE
Mike Ilitch School of Business

Functional Dependencies and Keys

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



Customer_ID → Customer_Name, Customer_Address

- **Candidate Key:** A unique identifier. One of the candidate keys will become the primary key
- **Full dependency:** Each non-key field is functionally dependent on every candidate key



WAYNE STATE

Mike Ilitch School of Business

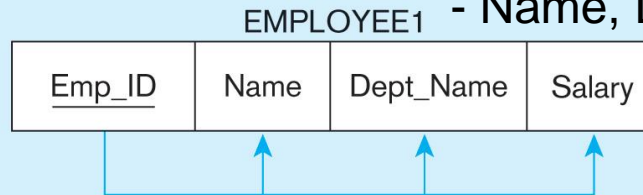
Functional Dependencies

Candidate key?

- Emp_ID

Non-key field

- Name, Dept_Name, Salary

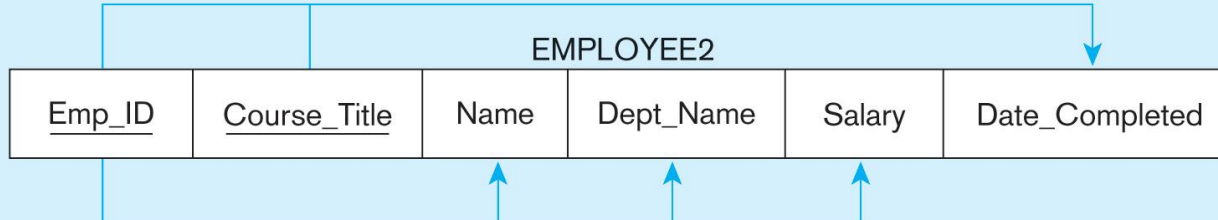


Candidate key?

- Emp_ID, Course_title

Non-key field

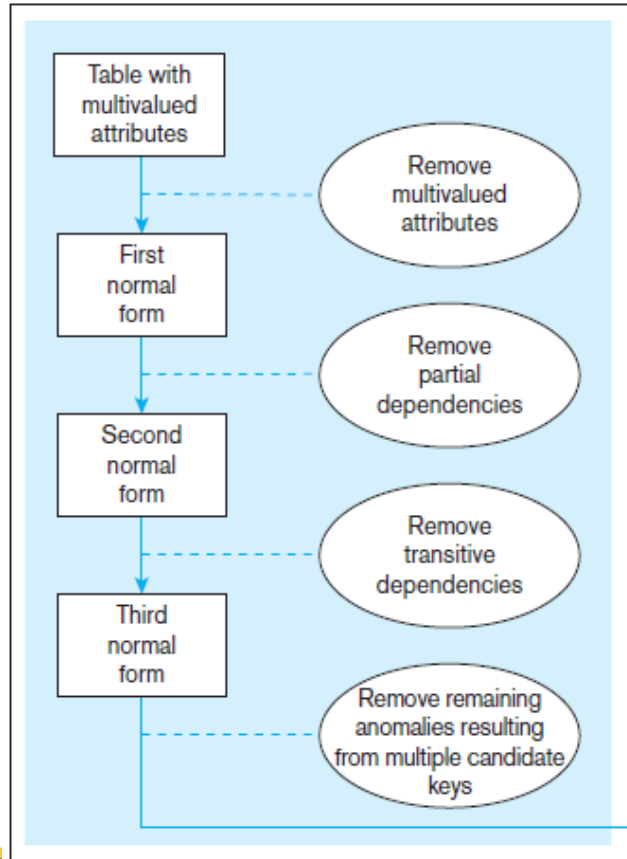
- Name, Dept_Name, Salary, Date_completed



WAYNE STATE

Mike Ilitch School of Business

The Process of Normalization



- Executed as series of steps to produce a specific normal form
- Relations become progressively more restricted (stronger) in format and also less vulnerable to update anomalies



WAYNE STATE

Mike Ilitch School of Business

Table with multivalued attributes, not in 1st normal form – Un-normalized form (UNF)

- A table that contains one or more repeating groups – **Multi-valued attributes**

<u>Order_ID</u>	Order_ Date	Customer_ ID	Customer_ Name	Customer_ Address	<u>Product_ID</u>	Product_ Description	Product_ Finish	Unit_ Price	Ordered_ Quantity
1006	10/24/2006	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	10/25/2006	6	Furniture Gallery	Boulder, CO	11	4-Dr Dresser	Oak	500.00	4
					4	Entertainment Center	Natural Maple	650.00	3



WAYNE STATE

Mike Ilitch School of Business

First Normal Form (1NF)

- No multivalued attributes (no repeating groups)
- Every attribute value is atomic
- Primary key identified

=> *All relations* are in 1st Normal Form (1NF)



Table with no Multivalued Attributes and Unique Rows (in 1NF)

PK (Composite)

<u>Order_ID</u>	Order_ Date	Customer_ ID	Customer_ Name	Customer_ Address	<u>Product_ID</u>	Product_ Description	Product_ Finish	Unit_ Price	Ordered_ Quantity
1006	10/24/2006	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
1006	10/24/2006	2	Value Furniture	Plano, TX	5	Writer's Desk	Cherry	325.00	2
1006	10/24/2006	2	Value Furniture	Plano, TX	4	Entertainment Center	Natural Maple	650.00	1
1007	10/25/2006	6	Furniture Gallery	Boulder, CO	11	4-Dr Dresser	Oak	500.00	4
1007	10/25/2006	6	Furniture Gallery	Boulder, CO	4	Entertainment Center	Natural Maple	650.00	3



WAYNE STATE

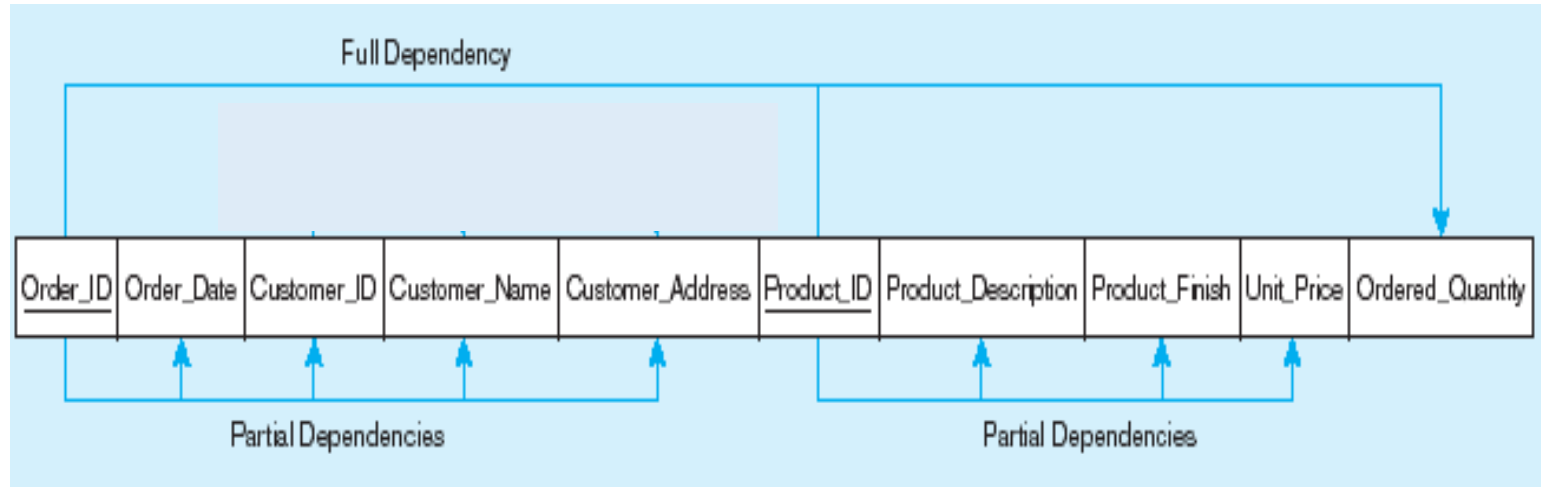
Mike Ilitch School of Business

Second Normal Form (2NF)

- 1NF + **no partial dependency**
 - Partial functional dependency: non-key attributes are functionally dependent on **part** (but not all) of the primary key
- Every non-key attribute must be defined by the entire key, not by only part of the primary key



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

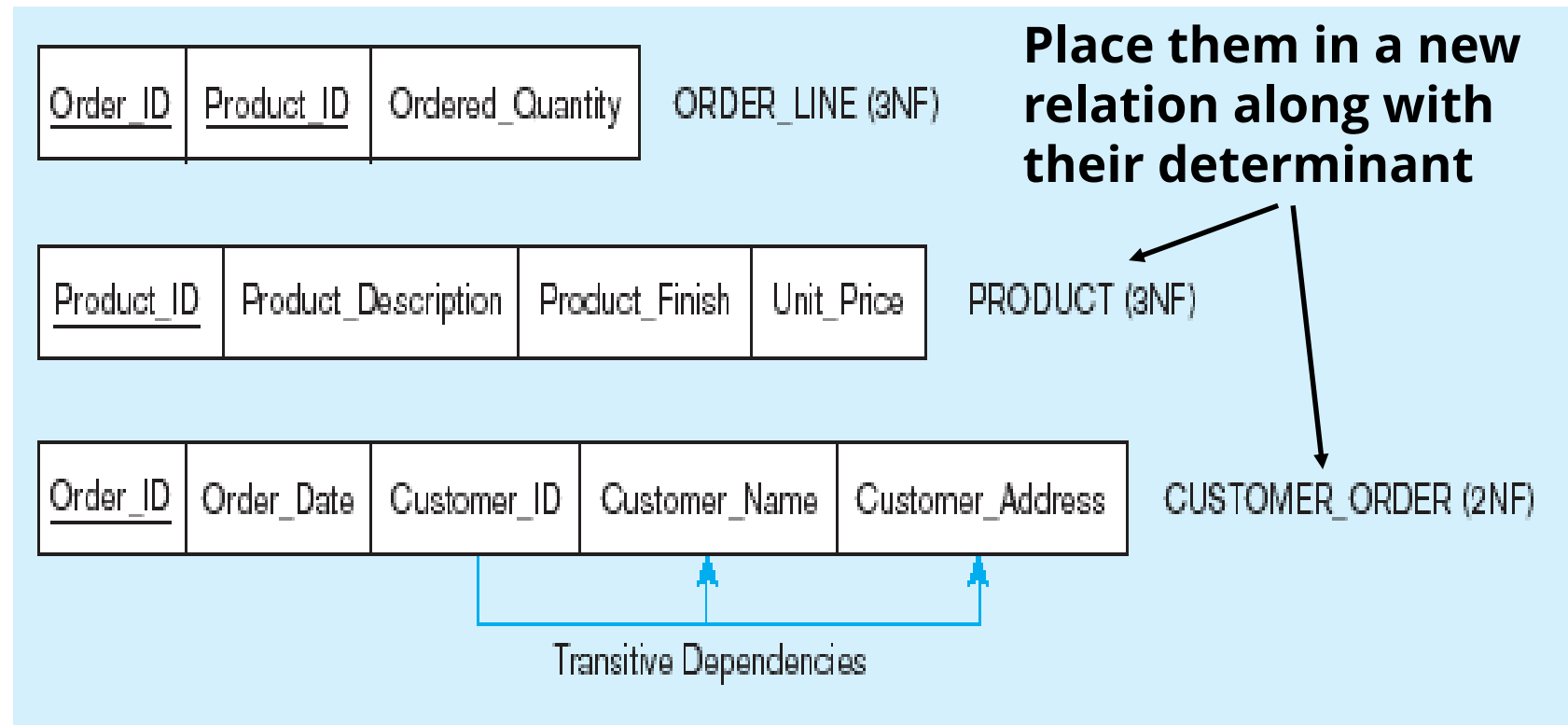


WAYNE STATE

Mike Ilitch School of Business

Therefore, NOT in 2NF

Removing Partial Dependencies

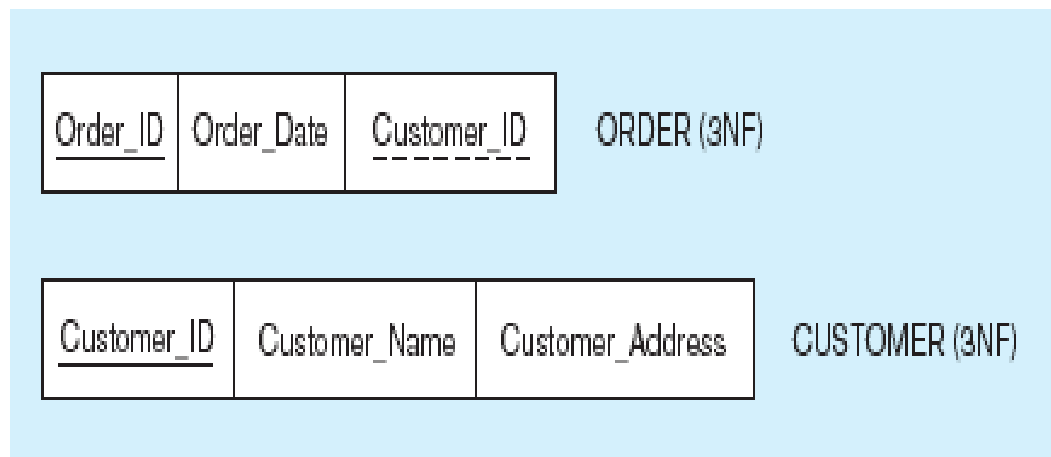
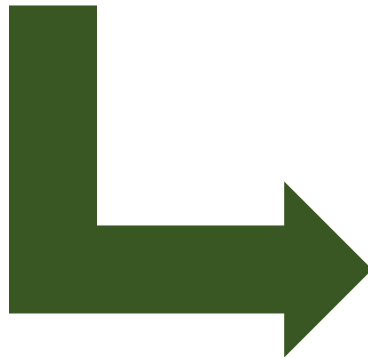
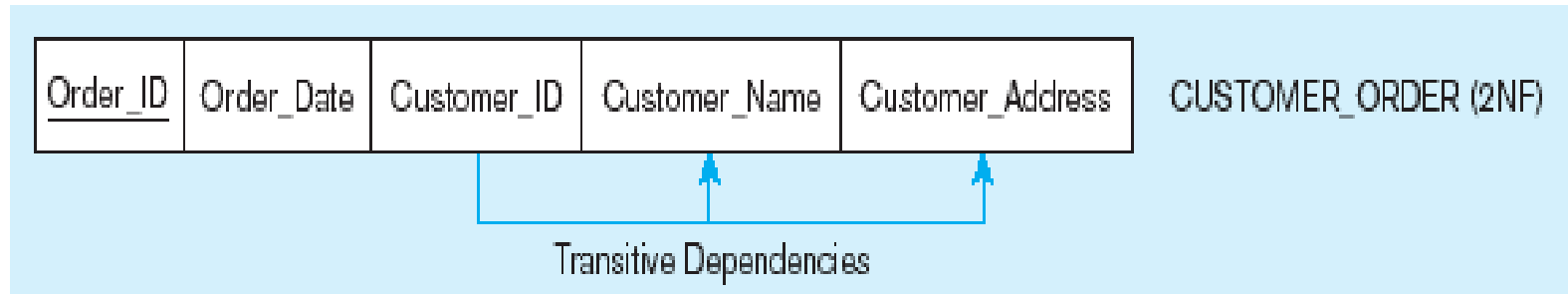


Third Normal Form (3NF)

- 2NF + **no transitive dependencies**
 - Transitive dependency: functional dependencies on non-primary-key attributes
- *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



Wrap-Up

Codd's Law

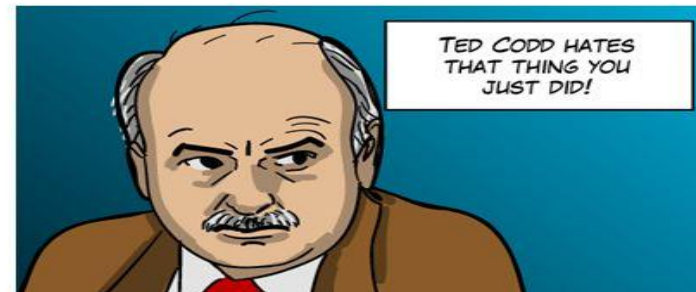
A non-key field must provide a fact about

NF1 - the *key*

NF2 - the *whole* key

NF3 - and *nothing but* the key

so help me Codd.



WAYNE STATE

Mike Ilitch School of Business