NORMALIZATION

Relational Database Design _Part 2b

Transitive Dependency

2NF removes redundancy due to partial dependency.

But 2NF is not free from redundancy due to transitive dependency.

Transitive dependency is a FD which holds by virtue of transitivity. A transitive dependency can occur only in a Relation that has 3 or more attributes.

Let A,B and C are 3 distinct attributes of a relation R
Suppose all the three of the following conditions hold

$A \rightarrow B$

it is not $B \rightarrow A$

$B \rightarrow C$

Then the FD $A \rightarrow C$ is a transitive dependency

(Transitive dependency occurs when a non key attribute determines another non key attribute)

Third Normal Form

For a Relation R to be in third normal form (3NF)

- R should be in 2NF
- R should not contain any transitive dependency.(or every nonprime attribute of R is non transitively dependent on every key of R)
- No nonprime attribute should depend on another nonprime attribute

R is in 3NF iff for each of its functional dependencies $X \rightarrow A$, at least one of the following condition holds.

- X contains A (ie. A is a subset of X ,meaning $X \rightarrow A$ is a trivial functional dependency),or
- X is a super key, or
- Every element of A-X, the set difference between A and X , is a prime attribute (ie. each attribute ibn A-X is contained in some candidate key)

Third Normal Form

In simple words

- A relational schema R is 3NF if for every FD

  $X \rightarrow A$ associated with R

  - $A \subseteq X$ (ie. the FD is trivial) **or**
  - X is a super key of R **or**
  - A is part of some key (not just super key)

ie.at least one of the above holds

3NF Example

Relation *dept_advisor*:

- *dept_advisor* (*s_ID, i_ID, dept_name*)
  $F = \{s\_ID, dept\_name \rightarrow i\_ID,\ i\_ID \rightarrow dept\_name\}$
- Two candidate keys: *s_ID, dept_name,* and *i_ID, s_ID*
- *R* is in 3NF
  - $s\_ID, dept\_name \rightarrow i\_ID$
  - *s_ID,dept_name* is a superkey
  - $i\_ID \rightarrow dept\_name$
    - *dept_name* is contained in a candidate key

So this Relation is in 3NF.But there is some redundancy in the schema

Example

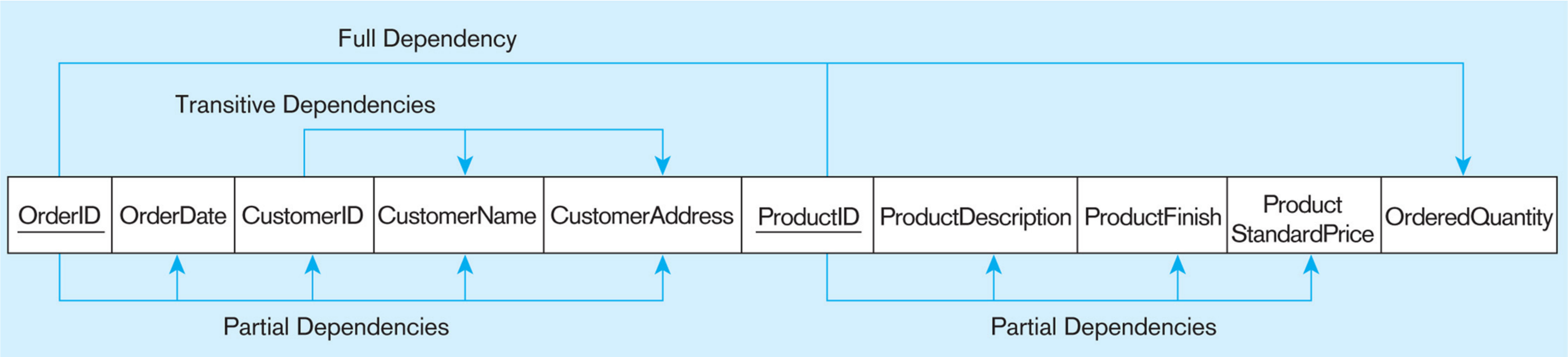
OrderID	Order Date	Customer ID	Customer Name	Customer Address	ProductID	Product Description	Product Finish	Product StandardPrice	Ordered Quantity
1006	10/24/2010	2	Value Furniture	Plano, TX	7	Dining Table	Natural Ash	800.00	2
					5	Writer's Desk	Cherry	325.00	2
					4	Entertainment Center	Natural Maple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

Note: This is NOT a relation.
WHY?

INF

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

This is relation in 1NF but not a well-structured one. WHY?



Order_ID ♏➔ Order_Date, Customer_ID, Customer_Name, Customer_Address

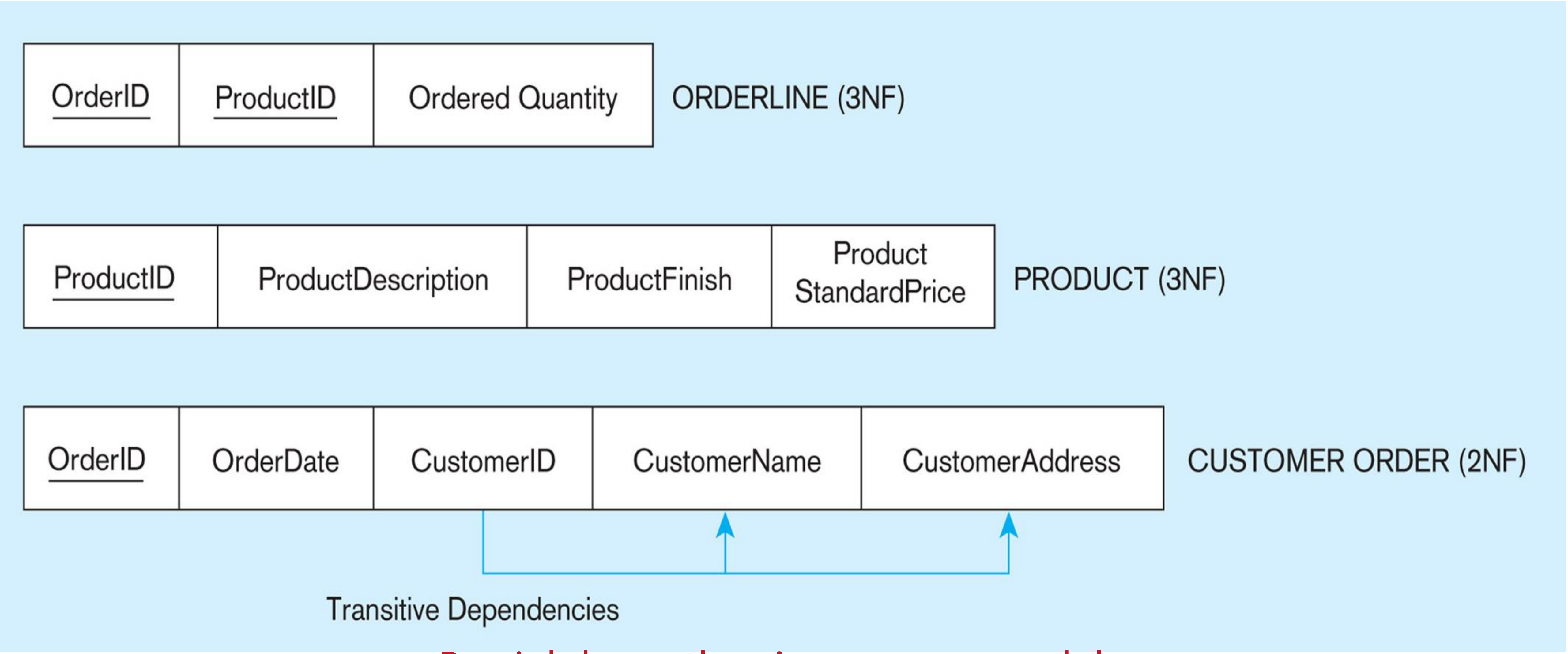
Customer_ID ♏➔ Customer_Name, Customer_Address

Product_ID ♏➔ Product_Description, Product_Finish, Unit_Price

Order_ID, Product_ID ♏➔ Order_Quantity

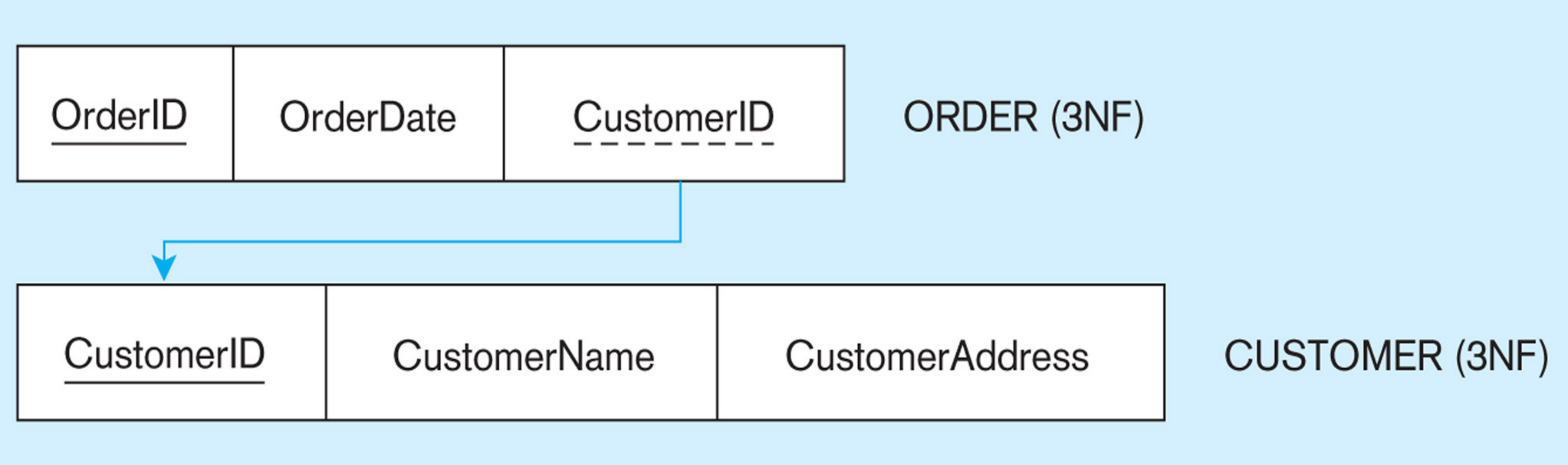
Therefore, NOT in 2NF Normal Form

2NF



Partial dependencies are removed, but there are still transitive dependencies

3NF


OrderID
OrderDate
CustomerID
ORDER (3NF)
CustomerID
CustomerName
CustomerAddress
CUSTOMER (3NF)

Boyce-Codd Normal Form

A relation schema *R* is in BCNF with respect to a set *F* of functional dependencies if for all functional dependencies in F^+ of the form

$$\alpha \rightarrow \beta$$

where $\alpha \subseteq R$ and $\beta \subseteq R$, at least one of the following holds:

- **$\alpha \rightarrow \beta$ is trivial (i.e., $\beta \subseteq \alpha$)**
- **α is a superkey for *R***

A Relation is in Boyce-Codd normal form (BCNF) if every determinant in the FD is a candidate key.

If a Relation contains only one candidate key, the 3NF and the BCNF are equivalent.

- BCNF is a special case of 3NF.
- BCNF eliminates all redundancy that can be discovered based on FD's

Boyce-Codd Normal Form

A relation schema *R* is in BCNF with respect to a set *F* of functional dependencies if for all functional dependencies in F^+ of the form

$$\alpha \rightarrow \beta$$

where $\alpha \subseteq R$ and $\beta \subseteq R$, at least one of the following holds:

- **$\alpha \rightarrow \beta$ is trivial (i.e., $\beta \subseteq \alpha$)**
- **α is a superkey for *R***

A Relation is in Boyce-Codd normal form (BCNF) if every determinant in the FD is a candidate key.

If a Relation contains only one candidate key, the 3NF and the BCNF are equivalent.

- BCNF is a special case of 3NF.
- BCNF eliminates all redundancy that can be discovered based on FD's

Boyce-Codd Normal Form (Cont.)

Example schema that is ***not*** in BCNF:

in_dep (*ID, name, salary, dept_name, building, budget*)

because :

- *dept_name*→ *building, budget* holds on *in_dep*
  BUT
- *dept_name* is not a superkey

When decompose *in_dept* into *instructor* and *department*

- *instructor* is in BCNF
- *department* is in BCNF

Decomposing a Schema into BCNF

Let R be a schema R that is not in BCNF. Let $\alpha \rightarrow \beta$ be the FD that causes a violation of BCNF.

We decompose R into:

- $(\alpha \cup \beta)$
- $(R - (\beta - \alpha))$

In our example of *in_dep*,

- R= (*ID, name, salary, dept_name, building, budget*)
- α = *dept_name*
- β = *building, budget*

and *in_dep* is replaced by

- $(\alpha \cup \beta)$ = (*dept_name, building, budget*)
- $(R - (\beta - \alpha))$ = (*ID, name, dept_name, salary*)

Comparison of BCNF and 3NF

Advantages of 3NF over BCNF.

It is always possible to obtain a 3NF design without sacrificing losslessness or dependency preservation.

Disadvantages of 3NF.

- We may have to use null values to represent some of the possible meaningful relationships among data items.
- There is the problem of repetition of information.