

Normalization

- Widely used as a guide in designing relational databases
- Eliminate redundant data
- Eliminate modifications that make the data inconsistent
- Often includes a process of ‘decomposition’ into a set of smaller tables
 - Removes multi-valued attributes and repeating groups
- Good database design is very important
 - Although SQL usage is very important, it cannot fix a badly designed database

Advantages and Disadvantages of Normalizing

- Advantages
 - ER diagrams
 - Avoid update anomalies
 - Compact
 - Data access
 - Joins
- Disadvantages
 - Numerous tables
 - Maintenance
 - Resource usage

Eliminating Data Inconsistencies

- **Anomalies**
 - Can occur from poorly designed table structures is duplicate data, redundant data and difficulties in using the data
- **Ideal Fields**
 - The best way to detect fields that may lead to anomalies is ensuring that each field designed meets the following criteria:
 - Represents a characteristic of the table subject
 - Contains only a single value
 - Should not be de-constructed into smaller components
 - Does not contain a calculated or concatenated value
 - Unique within the entire database structure
 - Identical field types in relationship

Normal Forms

- Many successive levels of normalization
 - Provides stronger guarantees for data updates
- First three are most common
 - **First Normal Form (1NF)** -- contains no repeating groups within rows
 - **Second Normal Form (2NF)** -- normalized at the first level and every non-key (supporting) column value is dependent on the primary key
 - A non-key value must depend on every column in a *composite* primary key, as well
 - **Third Normal Form (3NF)** -- normalized at the first and second level, dependent solely on the primary key and no other non-key (supporting) column value

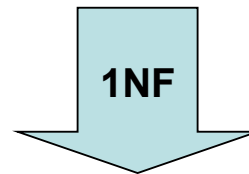
Normalization Process Example

- Furniture store inventory
 - You want to turn your inventory spreadsheet into a database, so you can...
 - do intelligent searches
 - Modifications
 - prepare for future growth of your business.
 - Very simple example to demonstrate normalization process
- Original inventory spread sheet

Inventory								
sID	sLoc	sPostal	pID1	pName1	pQty1	pID2	pName2	pQty2
1	Holtsville	00501	1	bed	15	2	chair	4
2	Waukesha	53146	1	bed	4	3	table	6
3	Waukesha	53146	2	chair	8	4	sofa	4
4	Ketchikan	99950	2	chair	24	4	sofa	10

First Normal Form

Inventory								
sID	sLoc	sPostal	pID1	pName1	pQty1	pID2	pName2	pQty2
1	Holtsville	00501	1	bed	15	2	chair	4
2	Waukesha	53146	1	bed	4	3	table	6
3	Waukesha	53146	2	chair	8	4	sofa	4
4	Ketchikan	99950	2	chair	24	4	sofa	10



Inventory_1NF						
PK	sID	sLoc	sPostal	pID*	pName	pQty
	1	Holtsville	00501	1	bed	15
	1	Holtsville	00501	2	chair	4
	2	Waukesha	53146	1	bed	4
	2	Waukesha	53146	3	table	6
	3	Waukesha	53146	2	chair	8
	3	Waukesha	53146	4	sofa	4
	4	Ketchikan	99950	2	chair	24
4	Ketchikan	99950	4	sofa	10	

* This column would not usually appear in a first normal form; however, it is here to demonstrate a connection to second normal form.

Second Normal Form

Inventory_1NF					
sID	sLoc	sPostal	pID	pName	pQty
1	Holtsville	00501	1	bed	15
1	Holtsville	00501	2	chair	4
2	Waukesha	53146	1	bed	4
2	Waukesha	53146	3	table	6
3	Waukesha	53146	2	chair	8
3	Waukesha	53146	4	sofa	4
4	Ketchikan	99950	2	chair	24
4	Ketchikan	99950	4	sofa	10

2NF

1NF

Supplier_2NF		
sID	sLoc	sPostal
1	Holtsville	00501
2	Waukesha	53146
3	Waukesha	53146
4	Ketchikan	99950

PK

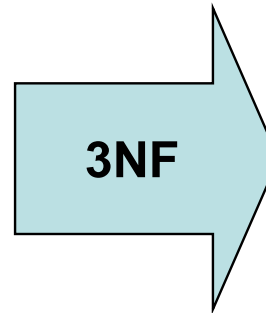
Part_1NF			
sID	pID	pName	pQty
1	1	bed	15
1	2	chair	4
2	1	bed	4
2	3	table	6
3	2	chair	8
3	4	sofa	4
4	2	chair	24
4	4	sofa	10

PK

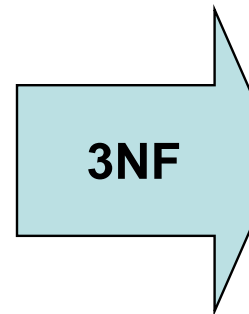
PK

Third Normal Form (1/2)

Part_1NF			
sID	pID	pName	pQty
1	1	bed	15
1	2	chair	4
2	1	bed	4
2	3	table	6
3	2	chair	8
3	4	sofa	4
4	2	chair	24
4	4	sofa	10

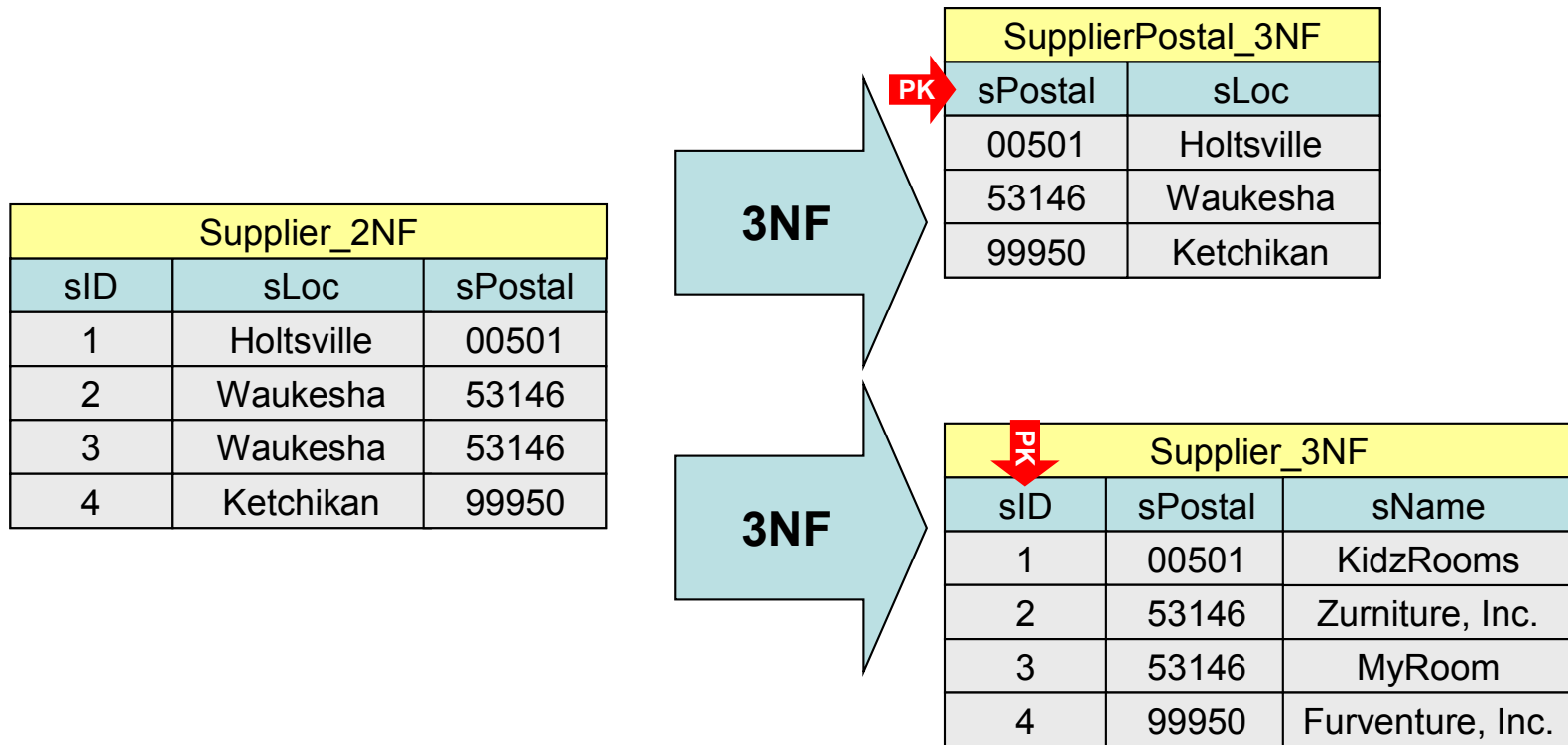


Part_3NF		
sID	pID	Qty
1	1	15
1	2	4
2	1	4
2	3	6
3	2	8
3	4	4
4	2	24
4	4	10



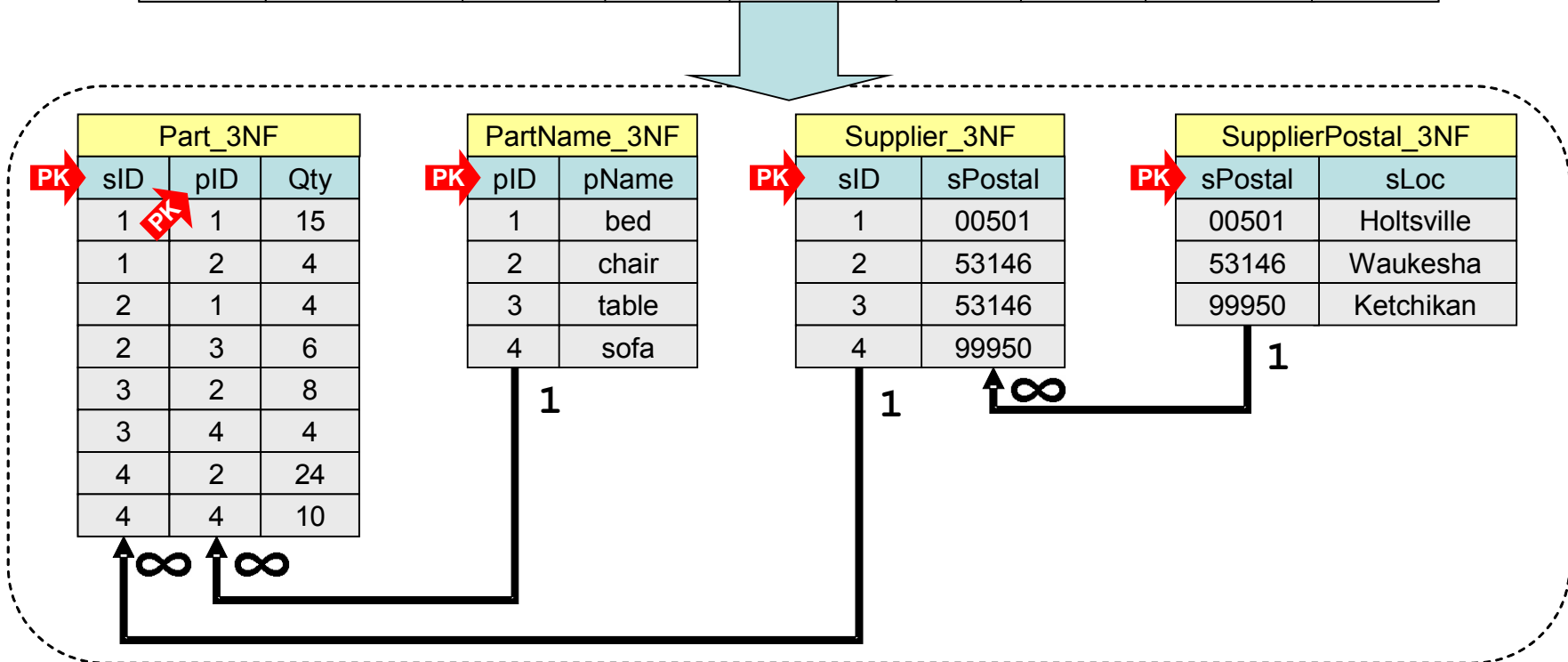
PartName_3NF	
pID	pName
1	bed
2	chair
3	table
4	sofa

Third Normal Form (2/2)



Normalized 'Furniture Store' Database

Inventory								
sID	sLoc	sPostal	pID1	pName1	pQty1	pID2	pName2	pQty2
1	Holtsville	00501	1	bed	15	2	chair	4
2	Waukesha	53146	1	bed	4	3	table	6
3	Waukesha	53146	2	chair	8	4	sofa	4
4	Ketchikan	99950	2	chair	24	4	sofa	10



Making Intelligent Choices

- Normalization is just common sense
- Put “pen to paper” and document the process
- Eliminate repeating groups
- Make sure each column contains only one value
- Use candidate keys
- Sensible column naming
- Create new tables to better organize (if needed)