

# What's covered here?

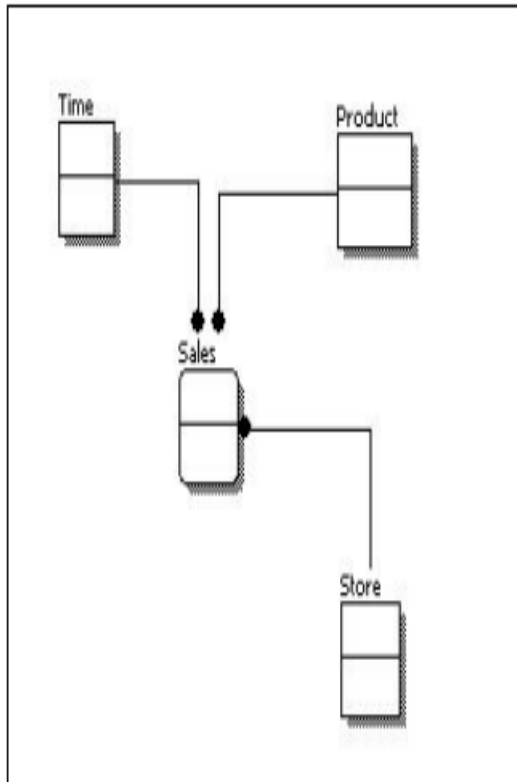
- Normalization

# Purpose of Normalization

- Produce a set of suitable entities that support the data requirements of an enterprise
- Minimize data redundancy
- Provide reporting flexibility
- Maintain data integrity

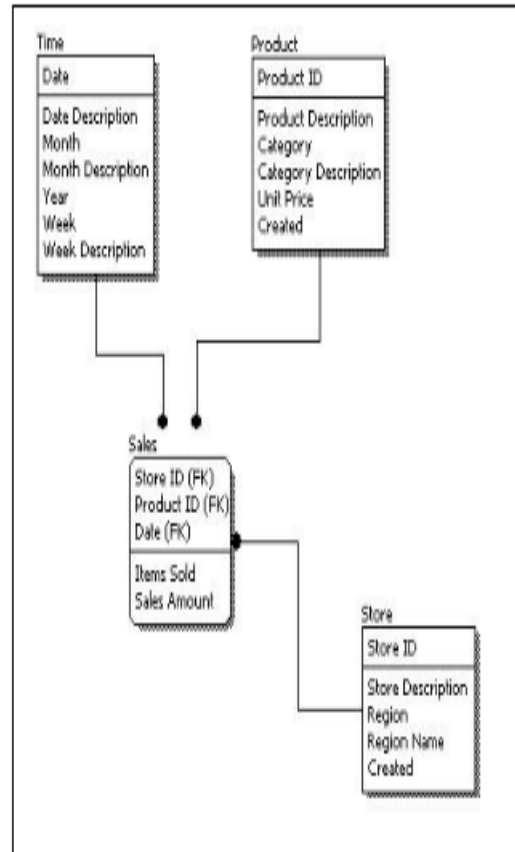
# Database Design Approach

## Conceptual Model Design

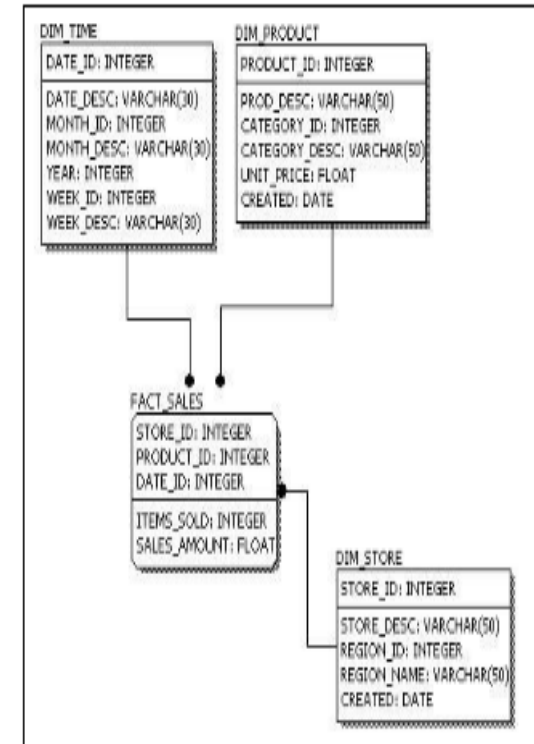


## Logical Model Design

**Do normalization here**



## Physical Model Design



# Functional Dependency

*Employee ID	Employee Name
101	John
102	Sara
103	Daniel
104	John
105	Chris

Functionally dependent on Employee ID

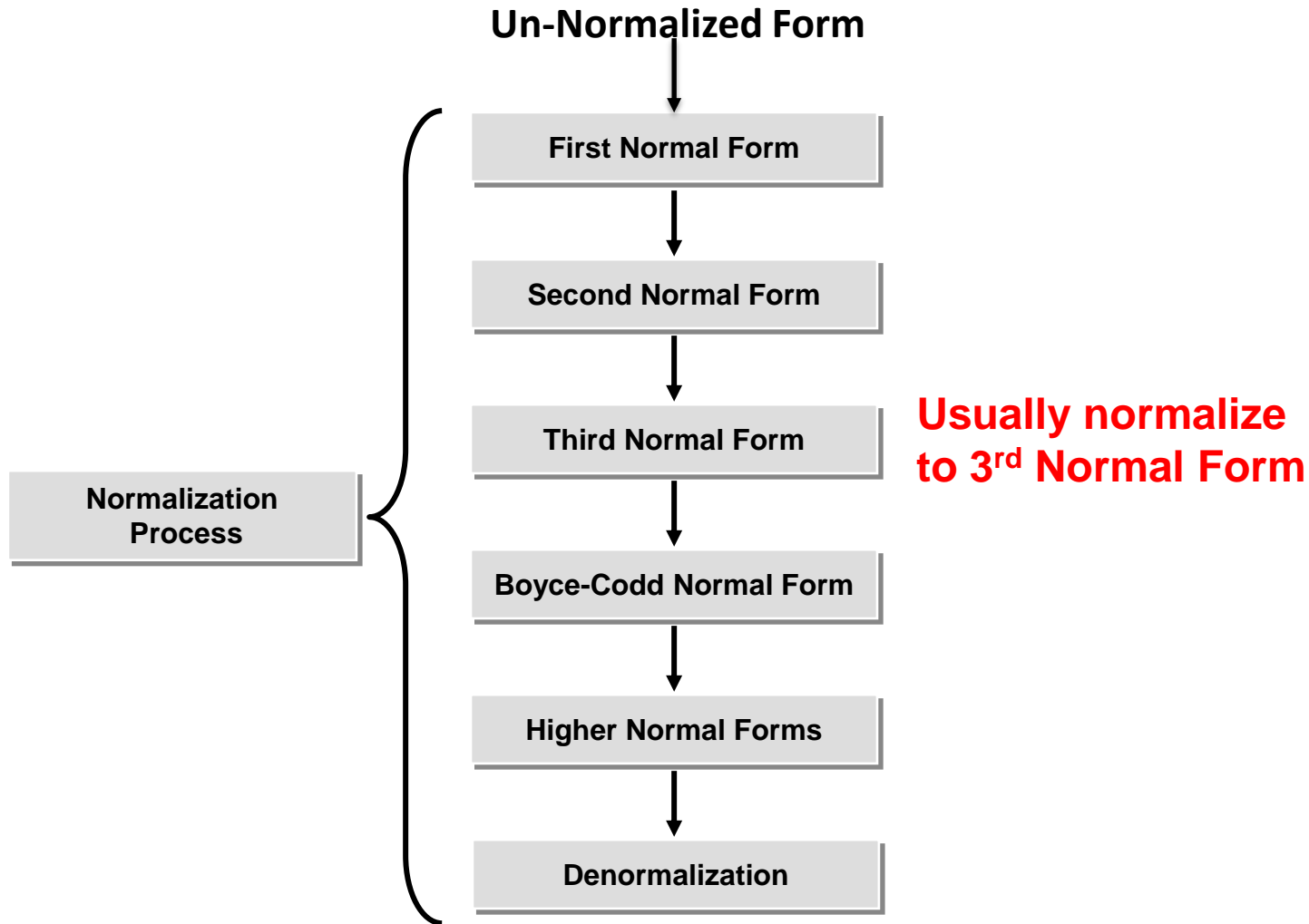
Employee ID  $\longrightarrow$  Employee Name

Employee Name  $\nrightarrow$  Employee ID

# Lossless-join and Dependency Preservation Properties

- Two important properties of decomposition
  - *Lossless-join property* enables us to find any instance of the original relation from corresponding instances in the smaller relations
  - *Dependency preservation property* enables us to enforce a constraint on the original relation by enforcing some constraint on each of the smaller relations

# The Normalization Process



# First Normal Form

- Remove multi-valued attributes
- Remove composite attributes
- Remove repeating groups
- Remove many-to-many relationship

**What are Multi-Valued Attributes?**



# Multi-Valued Attributes

* PersonID	Hobby
1	Reading, Cooking, Swimming
2	Playing Card, Gardening

Multi-Valued Attributes



**What are Composite Attributes?**

# Composite Attributes

*Person	Address
John Smith	1234 Darlene Way, San Jose, CA 95123
Peter Chang	2255 148th Ave, Bellevue 98008


Composite Attributes



# **What are Repeating Groups?**

# Repeating Groups

Logically repeating attributes



* PersonID	Hobby1	Hobby2	Hobby3	Hobby4
1	Reading	Cooking	Swimming	
2	Playing Card	Gardening		

# **What are Many-to-Many Relationships?**

# Many-to-Many Relationships

Each term students enroll in many courses

Students

*Student Name	*Contact	*Courses
Amity Rogers	661-555-9848	Math103
Lara Steele	661-555-8192	Math103
Phyllis Frank	661-555-0342	Phy111
Zena Lane	661-555-0934	Phy111
Brenda Michael	661-555-4881	Chem101
Amity Rogers	661-555-9848	Chem101
Sandra Dale	661-555-6270	Bio121
Phyllis Frank	661-555-0342	Bio121
Zena Lane	661-555-0934	Math103
Ora Stevenson	661-555-3503	Bio121
Rina Lowery	661-555-3682	Phy111
Brenda Michael	661-555-4881	Bio121

Courses

*Courses	*Student Name
Math103	Amity Rogers
	Lara Steele
	Kalia Wilcox
Phy111	Phyllis Frank
	Zena Lane
	Ora Stevenson
Chem101	Rina Lowery
	Brenda Michael
	Amity Rogers
Bio121	Sandra Dale
	Phyllis Frank

Enrollments

*Course ID	*Student ID
C01	101
C01	109
C01	112
C01	110
C01	105
C02	123
C02	105
C02	107
C02	129
C03	129
C03	121
C03	101
C03	109
C03	112
C04	110
C04	123
C04	117
C04	121

Students

*Student ID	Student Name	Contact
101	Amity Rogers	661-555-9848
109	Lara Steele	661-555-8192
112	Kalia Wilcox	661-555-8294
110	Sandra Dale	661-555-6270
123	Phyllis Frank	661-555-0342
105	Zena Lane	661-555-0934
117	Ora Stevenson	661-555-3503
129	Rina Lowery	661-555-3682
121	Brenda Michael	661-555-4881

Courses

*Course ID	Course Description
C01	Math103
C02	Phy111
C03	Chem101
C04	Bio121

With associative entity

## Un-normalized Form

### Employees

* Employee ID	Employee Name	Project 1 ID	Name	Project 2 ID	Name	Project 3 ID	Name
123	Dave Thompson	112-253-01	.....	112-478-03	.....		
124	Burt Rios	112-568-08	.....	113-235-19	.....	165-203-35	.....
125	Alice Parker	106-564-12	.....				



Primary key



# Employees

## Un-normalized Form

Logically repeating attributes

Composite

* Employee ID	Employee Name	Project 1 ID	Name	Project 2 ID	Name	Project 3 ID	Name
123	Dave Thompson	112-253-01	.....	112-478-03	.....		
124	Burt Rios	112-568-08	.....	113-235-19	.....	165-203-35	.....
125	Alice Parker	106-564-12	.....				

# Normalization

## Employees

<b>*EmployeeID</b>	<b>Employee First Name</b>	<b>Employee Last Name</b>
123	Dave	Thompson
124	Burt	Rios
125	Alice	Parker

## Projects

<b>*Project ID</b>	<b>Project Name</b>
112-253-01	ERP Web Frotend
112-478-03	Product Recommendation
112-568-08	CRM Upgrade
113-235-19	Inventory Data Pipeline
165-203-35	Website Retooing
106-564-12	Customer Deep Learning

# The First Normal Form

1NF

Employees

*EmployeeID	Employee First Name	Employee Last Name
123	Dave	Thompson
124	Burt	Rios
125	Alice	Parker

Projects

*Project ID	Project Name
112-253-01	ERP Web Frontend
112-478-03	Product Recommendation
112-568-08	CRM Upgrade
113-235-19	Inventory Data Pipeline
165-203-35	Website Retooling
106-564-12	Customer Deep Learning

Deployment

*Employee ID	* Project ID
123	112-253-01
123	112-478-03
124	112-568-08
124	113-235-19
124	165-203-35
125	106-564-12

# Second Normal Form

- Remove partial dependencies

**What is Partial Dependency?**

# Partial Dependency

Dependent only  
on Employee ID

## Assignment



<b>*Employee ID</b>	<b>Employee First Name</b>	<b>* Project ID</b>
101	John	112-253-01
102	Sara	112-478-03
103	Daniel	112-478-03
104	John	113-235-19
105	Chris	113-235-19

## Employees

**Before 2NF**

<b>*Employee ID</b>	<b>Employee First Name</b>	<b>* Project ID</b>	<b>Start Date</b>	<b>Project Manager Last Name</b>	<b>Duration</b>
101	John	112-253-01	2007-12-12	Albertson	18
102	Sara	112-478-03	2007-12-28	Albertson	20
103	Daniel	112-478-03	2007-12-28	Albertson	20
104	John	113-235-19	2008-02-12	Conway	22
105	Chris	113-235-19	2008-02-12	Conway	22

Dependent only  
on Employee ID

Dependent only on  
Project ID

### Employees

*Employee ID	Employee First Name	* Project ID	Start Date	Project Manager Last Name	Duration
101	John	112-253-01	2007-12-12	Albertson	18
102	Sara	112-478-03	2007-12-28	Albertson	20
103	Daniel	112-478-03	2007-12-28	Albertson	20
104	John	113-235-19	2008-02-12	Conway	22
105	Chris	113-235-19	2008-02-12	Conway	22

**Before 2NF**



Normalization

## Projects

* Project ID	Project Start Date	Project Manager Last Name	Project Duration (in months)
112-253-01	2007-12-12	Albertson	18
112-478-03	2007-12-28	Albertson	20
113-235-19	2008-02-12	Conway	22

## Employees

*Employee ID	Employee First Name
101	John
102	Sara
103	Daniel
104	John
105	Chris

# The Second Normal Form

**Projects**

* Project ID	Project Start Date	Project Manager Last Name	Project Duration (in months)
112-253-01	2007-12-12	Albertson	18
112-478-03	2007-12-28	Albertson	20
113-235-19	2008-02-12	Conway	22

**Employees**

*Employee ID	Employee First Name
101	John
102	Sara
103	Daniel
104	John
105	Chris

**Deployment**

*Employee ID	* Project ID
101	112-253-01
102	112-478-03
103	112-478-03
104	113-235-19
105	113-235-19

**2NF**


# Third Normal Form

- Remove transitive dependencies

**What is Transitive Dependency?**

# Transitive Dependency

## Students



* StudentID	StudentLName	StudentFName	Major	ProgramDirectorLN	ProgramDirectorFN
1	Smith	John	Biology	Black	Peter
2	Cuban	Mary	Pharmacy	Simpson	Connie
3	Whitmer	Marcus	Biology	Black	Peter

Diagram illustrating transitive dependencies in the 'Students' table:

- Primary key**: StudentID
- Non-key**: StudentLName, StudentFName, Major, ProgramDirectorLN, ProgramDirectorFN

Transitive dependencies are shown by curved arrows:

- A black arrow from StudentID to Major.
- A red arrow from Major to ProgramDirectorFN.

**Before 3NF**

## Employees

*Employee ID	Employee First Name	Employee Last Name	Manager ID	Manager First Name	Manager Last Name
101	John	Allen	201	Paul	Wright
102	Sara	Duran	203	Marsha	Albertson
103	Daniel	Friedman	234	Paul	Wright
104	John	Baker	556	Cindy	Conway
105	Chris	Spencer	556	Cindy	Conway

Primary key

Non-key

Non-key

Non-key

Non-key

Non-key

# Transitive Dependency

Before 3NF

Employees

*Employee ID	Employee First Name	Employee Last Name	Manager ID	Manager First Name	Manager Last Name
101	John	Allen	201	Paul	Wright
102	Sara	Duran	203	Marsha	Albertson
103	Daniel	Friedman	234	Paul	Wright
104	John	Baker	556	Cindy	Conway
105	Chris	Spencer	556	Cindy	Conway

Primary key

Non-key

Non-key

Non-key

Non-key

Non-key

Manager first and last name are dependent on Manager ID, a non-key field and not on the Employee ID, the key attribute

# Employees

Before 3NF

*Employee ID	Employee First Name	Employee Last Name	Manager ID	Manager First Name	Manager Last Name	Manager Mid Initial
101	John	Allen	201	Paul	Wright	R
102	Sara	Duran	203	Marsha	Albertson	A
103	Daniel	Friedman	234	Paul	Wright	K
104	John	Baker	556	Cindy	Conway	B
105	Chris	Spencer	556	Cindy	Conway	B

Deleting the details for Daniel Friedman will  
remove details of Paul K. Wright as well



# The Third Normal Form

**3NF**

Removing transitive dependency  
prevents loss of data

*Employee ID	Employee First Name	Employee Last Name	Manager ID
101	John	Allen	201
102	Sara	Duran	203
103	Daniel	Friedman	234
104	John	Baker	556
105	Chris	Spencer	556

* Manager ID	Manager First Name	Manager Last Name	Manager Mid Initial
201	Paul	Wright	R
203	Marsha	Albertson	A
234	Paul	Wright	K
556	Cindy	Conway	B