



## **DATABASE (2)**

### **NORMALIZATION**

**M. Rasti-Barzoki**

**Website: [rasti.iut.ac.ir](http://rasti.iut.ac.ir)**

**91-2**



### Normalization

**Application of a number of rules to the relational model which will simplify the relations.**

- Avoid common database problems**
- Make the data as tightly bound as possible**
- Store the minimum amount of data**



# Management Information Systems

## Database

### Normalization

- Un-normalized Example

First Name	Last Name	Address	CustomerID	Order Date	OrderID	Quantity	ProductID	Product Name	UnitPrice	Units In Stock
Nicholas	Colon	9020 N.W. 75 Street	C0004	4/15/1997	O0001	4	P0014	HD Floppy Disks	\$9.99	500
Nicholas	Colon	9020 N.W. 75 Street	C0004	4/15/1997	O0001	1	P0027	Norton Utilities	\$115.95	150
Neil	Goodman	4215 South 81 Street	C0003	4/18/1997	O0002	1	P0001	Pentium desktop/166	\$1,899.00	50
Neil	Goodman	4215 South 81 Street	C0003	4/18/1997	O0002	1	P0006	15" SVGA Monitor	\$499.00	50
Neil	Goodman	4215 South 81 Street	C0003	4/18/1997	O0002	1	P0020	Fax/Modem 56 Kbps	\$189.95	35
Neil	Goodman	4215 South 81 Street	C0003	4/18/1997	O0002	1	P0022	Laser Printer	\$1,395.00	100
Jeffrey	Muddell	9522 S.W. 142 Street	C0006	4/18/1997	O0003	1	P0005	Pentium laptop/133	\$2,599.00	15
Jeffrey	Muddell	9522 S.W. 142 Street	C0006	4/18/1997	O0003	1	P0020	Fax/Modem 56 Kbps	\$189.95	35
Jeffrey	Muddell	9522 S.W. 142 Street	C0006	4/18/1997	O0003	1	P0022	Laser Printer	\$1,395.00	100
Ashley	Geoghegan	7500 Center Lane	C0007	4/18/1997	O0004	1	P0003	Pentium Pro desktop,	\$2,099.00	125
Ashley	Geoghegan	7500 Center Lane	C0007	4/18/1997	O0004	1	P0010	2 Gb SCSI Hard Drive	\$799.00	25
Ashley	Geoghegan	7500 Center Lane	C0007	4/18/1997	O0004	2	P0022	Laser Printer	\$1,395.00	100
Benjamin	Lee	1000 Call Street	C0001	4/20/1997	O0005	2	P0003	Pentium Pro desktop,	\$2,099.00	125
Benjamin	Lee	1000 Call Street	C0001	4/20/1997	O0005	2	P0012	CD-ROM: 8X	\$249.00	40
Benjamin	Lee	1000 Call Street	C0001	4/20/1997	O0005	2	P0016	2 Gb Tape Backup	\$179.95	15
Benjamin	Lee	1000 Call Street	C0001	4/21/1997	O0006	1	P0007	17" SVGA Monitor	\$899.00	25
Benjamin	Lee	1000 Call Street	C0001	4/21/1997	O0006	10	P0014	HD Floppy Disks	\$9.99	500

– به نظر شما چه مشکلی دارد؟ اگر سیستم ثبت نام دروس به این شکل باشد چه اتفاقی می افتد؟



# Management Information Systems

## Database

### Normalization

- Normalized Example

#### Customers

First Name ▾	Last Name ▾	Address ▾	CustomerID ▾
Benjamin	Lee	1000 Call Street	C0001
Eleanor	Milgrom	7245 NW 8 Street	C0002
Neil	Goodman	4215 South 81 Street	C0003
Nicholas	Colon	9020 N.W. 75 Street	C0004
Michael	Ware	276 Brickell Avenue	C0005
Jeffrey	Muddell	9522 S.W. 142 Street	C0006
Ashley	Geoghegan	7500 Center Lane	C0007
Serena	Sherard	5000 Jefferson Lane	C0008
Luis	Couto	455 Bargello Avenue	C0009

#### Orders

Customer ID ▾	Order Date ▾	OrderID ▾
C0004	4/15/1997	O0001
C0003	4/18/1997	O0002
C0006	4/18/1997	O0003
C0007	4/18/1997	O0004
C0001	4/20/1997	O0005
C0001	4/21/1997	O0006
C0002	4/21/1997	O0007
C0002	4/22/1997	O0008
C0001	4/22/1997	O0009



# Management Information Systems

## Database

### Normalization

- Normalized Example

#### Order Details

Order ID	Quantity	Product ID
O0001	1	P0013
O0001	4	P0014
O0001	1	P0027
O0002	1	P0001
O0002	1	P0006
O0002	1	P0020
O0002	1	P0022
O0003	1	P0005
O0003	1	P0020
O0003	1	P0022

#### Products

ProductID	Product Name	UnitPrice	Units In Stock
P0001	Pentium desktop/166 with MMX	\$1,899.00	50
P0002	Pentium desktop/200 with MMX	\$1,999.00	25
P0003	Pentium Pro desktop/180	\$2,099.00	125
P0004	Pentium Pro desktop/200	\$2,299.00	25
P0005	Pentium laptop/133	\$2,599.00	15
P0006	15" SVGA Monitor	\$499.00	50
P0007	17" SVGA Monitor	\$899.00	25
P0008	20" Multisync Monitor	\$1,599.00	50
P0009	2.5 Gb IDE Hard Drive	\$399.00	15
P0010	2 Gb SCSI Hard Drive	\$799.00	25





### Normalization

- **Database Normalization Steps**

- 1NF - Remove repeating groups
- 2NF - Remove fields dependent only on part of the key field (Applies to concatenated keys)
- 3NF - Remove fields dependent only on other fields in that table

– برطرف کردن گروه های تکرار (نرمال نوع یک)

– برطرف کردن وابستگی های جزئی (نرمال نوع دو)

• هر صفت (فیلد) غیر کلید باید به تمام فیلدهای کلید وابسته باشد.

– برطرف کردن وابستگی های انتقالی (نرمال نوع سه)

• یک صفت (فیلد) غیر کلید نباید به یک فیلد غیر کلید وابسته باشد.



### Normalization

- توضیح و مثال (جزوه)



### Normalization

• مثال دو

#### – Case of a Construction Company

- **Building project -- Project number, Name, Employees assigned to the project.**
- **Employee -- Employee number, Name, Job classification**
- **The company charges its clients by billing the hours spent on each project. The hourly billing rate is dependent on the employee's position.**
- **Periodically, a report is generated.**





# Management Information Systems

## Database

### Normalization

A few employees works for one project.

Employee Num :  
**101, 102, 103,**

Project Num : **15**

Project Name :  
**Evergreen**





### Normalization

- **Sample Form**

**Project Num : 15**

**Project Name : Evergreen**



---

<b>Emp Num</b>	<b>Emp Name</b>	<b>Job Class</b>	<b>Chr Hours</b>	<b>Hrs Billed</b>	<b>Total</b>
<b>101</b>					
<b>102</b>					
<b>103</b>					
<b>105</b>					



# Management Information Systems

## Database

### Normalization

**TABLE 5.1** A SAMPLE REPORT LAYOUT

PROJ. NUM.	PROJECT NAME	EMPLOYEE NUMBER	EMPLOYEE NAME	JOB CLASS.	CHG/HOUR	HOURS BILLED	TOTAL CHARGE
15	Evergreen	103	June E. Arbough	Elec. Engineer	\$84.50	23.8	\$2,011.10
		101	John G. News	Database Designer	\$105.00	19.4	\$2,037.00
		105	Alice K. Johnson *	Database Designer	\$105.00	35.7	\$3,748.50
		106	William Smithfield	Programmer	\$35.75	12.6	\$450.45
		102	David H. Senior	Systems Analyst	\$96.75	23.8	\$2,302.65
						<b>Subtotal</b>	<b>\$10,549.70</b>
18	Amber Wave	114	Annelise Jones	Applications Designer	\$48.10	24.6	\$1,183.26
		118	James J. Frommer	General Support	\$18.36	45.3	\$831.71
		104	Anne K. Ramoras *	Systems Analyst	\$96.75	32.4	\$3,134.70
		112	Darlene M. Smithson	DSS Analyst	\$45.95	44.0	\$2,021.80
						<b>Subtotal</b>	<b>\$7,171.47</b>
22		105	Alice K. Johnson	Database Designer	\$105.00	64.7	\$6,793.50
		104	Anne K. Ramoras	Systems Analyst	\$96.75	48.4	\$4,682.70
		113	Delbert K. Joenbrood*	Applications Designer	\$48.10	23.6	\$1,135.16
		111	Geoff B. Wabash	Clerical Support	\$26.87	22.0	\$591.14
		106	William Smithfield	Programmer	\$35.75	12.8	\$457.60
						<b>Subtotal</b>	<b>\$13,660.10</b>
25		107	Maria D. Alonzo	Programmer	\$35.75	24.6	\$879.45
		115	Travis B. Bawangi	Systems Analyst	\$96.75	45.8	\$4,431.15
		101	John G. News *	Database Designer	\$105.00	56.3	\$5,911.50
		114	Annelise Jones	Applications Designer	\$48.10	33.1	\$1,592.11
		108	Ralph B. Washington	Systems Analyst	\$96.75	23.6	\$2,283.30
		118	James J. Frommer	General Support	\$18.36	30.5	\$559.98
		112	Darlene M. Smithson	DSS Analyst	\$45.95	41.4	\$1,902.33
						<b>Subtotal</b>	<b>\$17,559.82</b>
						<b>Total</b>	<b>48,941.09</b>

Note: \* indicates project leader





# Management Information Systems

## Database

### Normalization

- **Table Structure Matches the Report Format**

	PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
▶	15	Evergreen	103	June E. Arbough	Elect. Engineer	\$84.50	23.8
			101	John G. News	Database Designer	\$105.00	19.4
			105	Alice K. Johnson *	Database Designer	\$105.00	35.7
			106	William Smithfield	Programmer	\$35.75	12.6
			102	David H. Senior	Systems Analyst	\$96.75	23.8
	18	Amber Wave	114	Annelise Jones	Applications Designer	\$48.10	24.6
			118	James J. Frommer	General Support	\$18.36	45.3
			104	Anne K. Ramoras *	Systems Analyst	\$96.75	32.4
			112	Darlene M. Smithson	DSS Analyst	\$45.95	44.0
	22	Rolling Tide	105	Alice K. Johnson	Database Designer	\$105.00	64.7
			104	Anne K. Ramoras	Systems Analyst	\$96.75	48.4
			113	Delbert K. Joenbrood *	Applications Designer	\$48.10	23.6
			111	Geoff B. Wabash	Clerical Support	\$26.87	22.0
			106	William Smithfield	Programmer	\$35.75	12.8
	25	Starflight	107	Maria D. Alonzo	Programmer	\$35.75	24.6
			115	Travis B. Bawangi	Systems Analyst	\$96.75	45.8
			101	John G. News *	Database Designer	\$105.00	56.3
			114	Annelise Jones	Applications Designer	\$48.10	33.1
			108	Ralph B. Washington	Systems Analyst	\$96.75	23.6
			118	James J. Frommer	General Support	\$18.36	30.5
			112	Darlene M. Smithson	DSS Analyst	\$45.95	41.4

FIGURE 5.1 A TABLE WHOSE STRUCTURE MATCHES THE REPORT FORMAT



### Normalization

- **Problems with the Figure 5.1**
  - **The project number is intended to be a primary key, but it contains nulls.**
  - **The table displays data redundancies.**
  - **The table entries invite data inconsistencies.**
  - **The data redundancies yield the following anomalies:**
    - **Update anomalies.**
    - **Addition anomalies.**
    - **Deletion anomalies.**



# Management Information Systems

## Database

### Normalization

- **Conversion to First Normal Form**
  - **A relational table must not contain repeating groups.**
  - **Repeating groups can be eliminated by adding the appropriate entry in at least the primary key column(s).**

	PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
▶	15	Evergreen	103	June E. Arbough	Elect. Engineer	\$84.50	23.8
			101	John G. News	Database Designer	\$105.00	19.4
			105	Alice K. Johnson *	Database Designer	\$105.00	35.7
			106	William Smithfield	Programmer	\$35.75	12.6
			102	David H. Senior	Systems Analyst	\$96.75	23.8

FIGURE 5.2 ■ THE EVERGREEN DATA





# Management Information Systems

## Database

### Normalization

- Data Organization: First Normal Form

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	03	June E. Arbough	Elect. Engineer	\$84.50	
		01	John G. News	Database Designer	\$105.00	
		05	Alice K. Johnson *	Database Designer	\$105.00	
		06	William Smithfield	Programmer	\$35.75	
		02	David H. Senior	Systems Analyst	\$96.75	
18	Amber Wave	14	Annelise Jones	Applications Designer	\$48.10	
		18	James J. Frommer	General Support	\$18.36	
		04	Anne K. Ramoras *	Systems Analyst	\$96.75	
		12	Darlene M. Smithson	DSS Analyst	\$45.95	
22	Rolling Tide	05	Alice K. Johnson	Database Designer	\$105.00	
		04	Anne K. Ramoras	Systems Analyst	\$96.75	
		13	Delbert K. Joenbrood *	Applications Designer	\$48.10	
		11	Geoff B. Wabash	Clerical Support	\$26.87	
		06	William Smithfield	Programmer	\$35.75	
25	Starflight	07	Maria D. Alonzo	Programmer	\$35.75	
		15	Travis B. Bawangi	Systems Analyst	\$96.75	
		01	John G. News *	Database Designer	\$105.00	
		14	Annelise Jones	Applications Designer	\$48.10	33.1
		08	Ralph B. Washington	Systems Analyst	\$96.75	23.6
		18	James J. Frommer	General Support	\$18.36	30.5
		12	Darlene M. Smithson	DSS Analyst	\$45.95	41.4

FIGURE 5.1 A 1

Before

HES THE REPORT FORMAT

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	June E. Arbough	Elect. Engineer	\$84.50	23.8
15	Evergreen	101	John G. News	Database Designer	\$105.00	19.4
15	Evergreen	105	Alice K. Johnson *	Database Designer	\$105.00	35.7
15	Evergreen	106	William Smithfield	Programmer	\$35.75	12.5
15	Evergreen	102	David H. Senior	Systems Analyst	\$96.75	23.9
18	Amber Wave	114	Annelise Jones	Applications Designer	\$48.10	24.6
18	Amber Wave	118	James J. Frommer	General Support	\$18.36	45.3
18	Amber Wave	104	Anne K. Ramoras *	Systems Analyst	\$96.75	32.1
18	Amber Wave	112	Darlene M. Smithson	DSS Analyst	\$45.95	44.0
22	Rolling Tide	105	Alice K. Johnson	Database Designer	\$105.00	64.7
22	Rolling Tide	104	Anne K. Ramoras	Systems Analyst	\$96.75	48.9
22	Rolling Tide	113	Delbert K. Joenbrood *	Applications Designer	\$48.10	23.6
22	Rolling Tide	111	Geoff B. Wabash	Clerical Support	\$26.87	22.5
22	Rolling Tide	106	William Smithfield	Programmer	\$35.75	12.1
25	Starflight	107	Maria D. Alonzo	Programmer	\$35.75	24.7
25	Starflight	115	Travis B. Bawangi	Systems Analyst	\$96.75	45.8
25	Starflight	101	John G. News *	Database Designer	\$105.00	56.3
25	Starflight	114	Annelise Jones	Applications Designer	\$48.10	33.1
25	Starflight	108	Ralph B. Washington	Systems Analyst	\$96.75	23.9
25	Starflight	118	James J. Frommer	General Support	\$18.36	30.2
25	Starflight	112	Darlene M. Smithson	DSS Analyst	\$45.95	41.4

FIGURE 5.3

After

RST NORMAL FORM



### Normalization

- **1NF Definition**
  - The term first normal form (**1NF**) describes the tabular format in which:
    - All the key attributes are defined.
    - There are no repeating groups in the table.
    - All attributes are dependent on the primary key.



### Normalization

- **Conversion to Second Normal Form**

- **Starting with the 1NF format, the database can be converted into the 2NF format by**

- **Writing each key component on a separate line, and then writing the original key on the last line and**
- **Writing the dependent attributes after each new key.**

PROJECT (PROJ\_NUM, PROJ\_NAME)

EMPLOYEE (EMP\_NUM, EMP\_NAME, JOB\_CLASS, CHG\_HOUR)

ASSIGN (PROJ\_NUM, EMP\_NUM, HOURS)



### Normalization

- **Second Normal Form (2 NF)**

- **A table is in 2NF if:**

- **It is in 1NF and**

- It includes no partial dependencies; that is, no attribute is dependent on only a portion of the primary key.

**(It is still possible for a table in 2NF to exhibit transitive dependency; that is, one or more attributes may be functionally dependent on non-key attributes.)**





# Management Information Systems

## Database

### Normalization

- **Dependency Diagram**

- The primary key components are bold, underlined, and shaded in a different color.
- The arrows above entities indicate all desirable dependencies, i.e., dependencies that are based on PK.
- The arrows below the dependency diagram indicate less desirable dependencies -- **partial dependencies** and **transitive dependencies**.

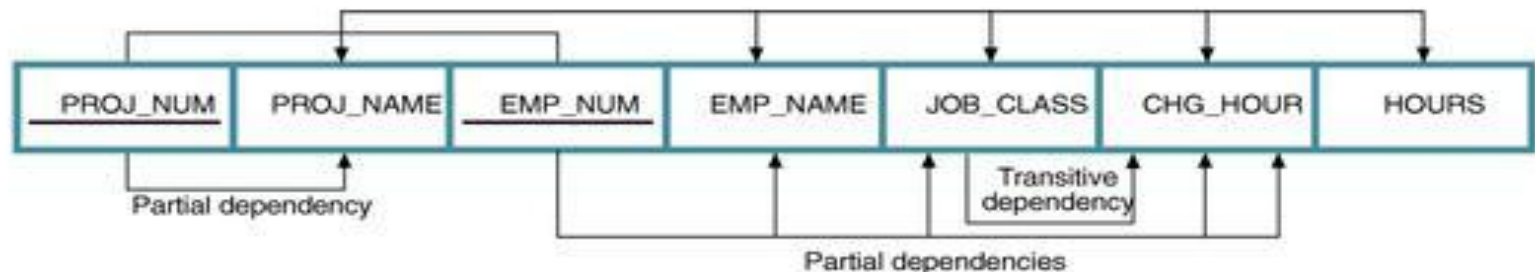


FIGURE 5.4 A DEPENDENCY DIAGRAM: FIRST NORMAL FORM (1NF)



### Normalization

- **Dependency Diagram**

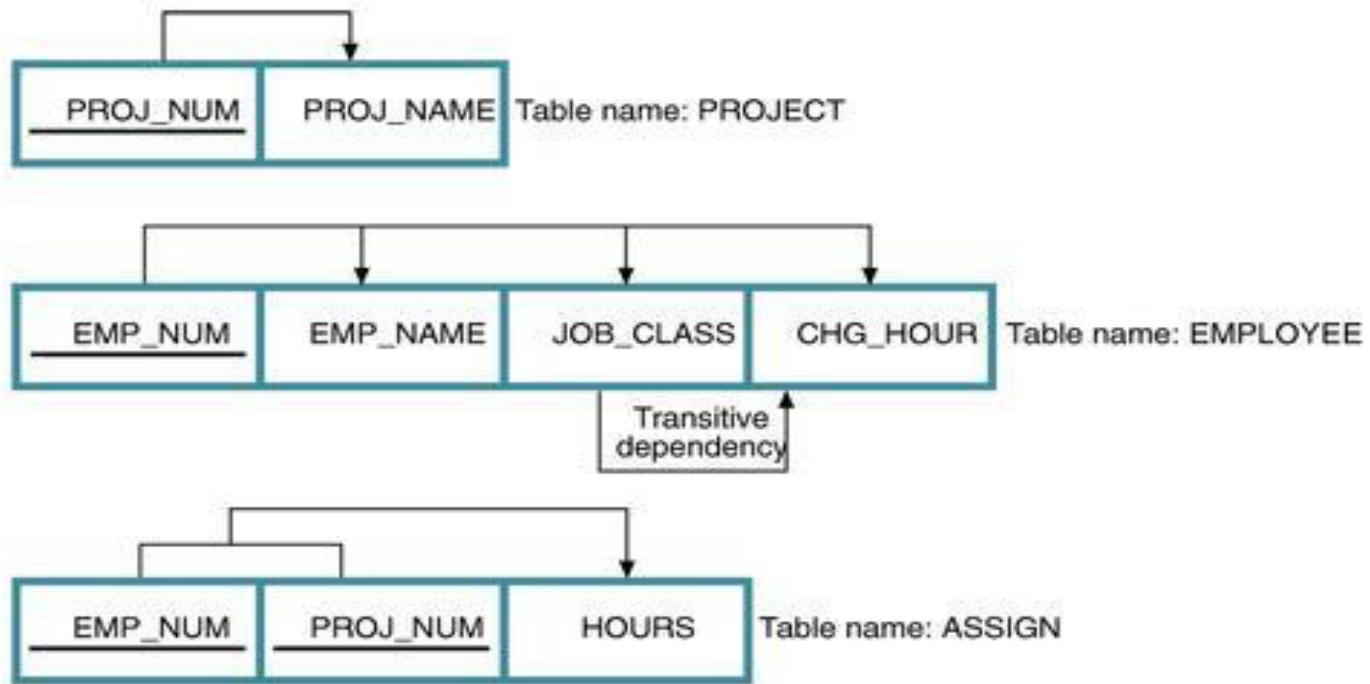


FIGURE 5.5 ■ SECOND NORMAL FORM (2NF) CONVERSION RESULTS





### Normalization

- **Third Normal Form (3 NF)**
- **Conversion to Third Normal Form**
  - **Create a separate table with attributes in a transitive functional dependence relationship.**

PROJECT (PROJ\_NUM, PROJ\_NAME)

ASSIGN (PROJ\_NUM, EMP\_NUM, HOURS)

EMPLOYEE (EMP\_NUM, EMP\_NAME, JOB\_CLASS)

JOB (JOB\_CLASS, CHG\_HOUR)



### Normalization

- **3NF Definition**

- **A table is in 3NF if:**

- **It is in 2NF and**

- **It contains no transitive dependencies.**



### نکات تکمیلی

- پاک سازی لیست نهاد / صفت

- صفات مترادف (synonyms)
- لغات یکسان برای صفات متفاوت
- اطلاعات اضافی
- اطلاعات دوجانبه



پایان

