

# CPSC 5021: Database Systems

Normalization of Database Tables

Lin Li

# Why Normalization?

**TABLE 6.1** A Sample Report Layout

PROJECT NUMBER	PROJECT NAME	EMPLOYEE NUMBER	EMPLOYEE NAME	JOB CLASS	CHARGE/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	Rolling Tide	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	Starflight	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: A \* indicates the project leader.

Project Number	Project Name	Employee Number	Employee Name	Job Class	Charge per Hour	Hours Billed
15	Evergreen	103	June E. Arbough	Elec. Engineer	\$84.50	23.8
15	Evergreen	101	Jone 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	12.6
15	Evergreen	102	David H. Senior	Systems Analyst	\$96.75	23.8
18	Amber Wave	114	Annelise Jones	Application Designer	\$48.10	24.6
18	Amber Wave	118	James J. Frommer	General Support	\$18.36	45.3
18	Amber Wave	104	Anne K. Romoras	Systems Analyst	\$96.75	32.4
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. Romoras	Systems Analyst	\$96.75	48.4
22	Rolling Tide	106	William Smithfield	Programmer	\$35.75	12.8

Problems?

- redundancies

Project Number	Project Name	Employee Number	Employee Name	Job Class	Charge per Hour	Hours Billed
15	Evergreen	103	June E. Arbough	Elec. Engineer	\$84.50	23.8
15	Evergreen	101	Jone 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	12.6
15	Evergreen	102	David H. Senior	Systems Analyst	\$96.75	23.8
18	Amber Wave	114	Annelise Jones	Application Designer	\$48.10	24.6
18	Amber Wave	118	James J. Frommer	General Support	\$18.36	45.3
18	Amber Wave	104	Anne K. Romoras	Systems Analyst	\$96.75	32.4
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. Romoras	Systems Analyst	\$96.75	48.4
22	Rolling Tide	106	William Smithfield	Programmer	\$35.75	12.8

Problems?

- Update anomalies

e.g. If we want to modify the Job classification & charge per hour for employee number 105, the update may cause anomalies.

Project Number	Project Name	Employee Number	Employee Name	Job Class	Charge per Hour	Hours Billed
15	Evergreen	103	June E. Arbough	Elec. Engineer	\$84.50	23.8
15	Evergreen	101	Jone 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	12.6
15	Evergreen	102	David H. Senior	Systems Analyst	\$96.75	23.8
18	Amber Wave	114	Annelise Jones	Application Designer	\$48.10	24.6
18	Amber Wave	118	James J. Frommer	General Support	\$18.36	45.3
18	Amber Wave	104	Anne K. Romoras	Systems Analyst	\$96.75	32.4
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. Romoras	Systems Analyst	\$96.75	48.4
26	Phantom Proj.	120	Anna Jones	Programmer	\$40.00	20.0

Problems?

- **Insertion anomalies**

e.g. To insert a new employee, but this employee is not assigned to a project yet. Has to make a phantom project so that this employee can be inserted (if the table does not allow NULL for Project Number and Project Name).

Project Number	Project Name	Employee Number	Employee Name	Job Class	Charge per Hour	Hours Billed
15	Evergreen	103	June E. Arbough	Elec. Engineer	\$84.50	23.8
15	Evergreen	101	Jone 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	12.6
15	Evergreen	102	David H. Senior	Systems Analyst	\$96.75	23.8
18	Amber Wave	114	Annelise Jones	Application Designer	\$48.10	24.6
18	Amber Wave	118	James J. Frommer	General Support	\$18.36	45.3
18	Amber Wave	104	Anne K. Romoras	Systems Analyst	\$96.75	32.4
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

Problems?

- **Deletion anomalies**

e.g. Suppose only one employee is related to the project 22. If the employee leaves the project, deletion of the employee record will make the project information be deleted as well (if values of the table cannot be NULL).

# Update Anomalies

## Employees' Skills



Different!

Employee ID	Employee Address	Skill
426	87 Sycamore Grove	Typing
426	87 Sycamore Grove	Shorthand
519	94 Chestnut Street	Public Speaking
519	96 Walnut Avenue	Carpentry

# Insertion Anomalies

## Faculty and Their Courses

Faculty ID	Faculty Name	Faculty Hire Date	Course Code
389	Dr. Giddens	10-Feb-1985	ENG-206
407	Dr. Saperstein	19-Apr-1999	CMP-101
407	Dr. Saperstein	19-Apr-1999	CMP-201

The record of  
new faculty Dr.  
Newsome  
**CANNOT** be  
inserted!  
(if the course  
code cannot be  
left NULL)

424	Dr. Newsome	29-Mar-2007
-----	-------------	-------------

?



# Deletion Anomalies

## Faculty and Their Courses

Faculty ID	Faculty Name	Faculty Hire Date	Course Code
389	Dr. Giddens	10-Feb-1985	ENG-206
407	Dr. Saperstein	19-Apr-1999	CMP-101
407	Dr. Saperstein	19-Apr-1999	CMP-201

The details of Dr. Giddens is  
LOST if he is temporarily not  
assigned to any course.

# Problems

- Redundancies
- Update anomalies
- Insertion anomalies
- Deletion anomalies



How to get rid of these problems?



**Normalization:** process for correcting table structures to minimize data redundancies and reduce data anomalies

# Normalization

- First normal form (1NF)
- Second normal form (2NF)
- Third normal form (3NF)



Objective of normalization: Ensure that all tables are in at least 3NF

2NF is better than 1NF; 3NF is better than 2NF.

Higher normal forms usually increase joins

# 1NF

- A table is in first normal form if:
  - ✓ There are no repeating groups
  - ✓ All data values are atomic
  - ✓ Each field has a unique name
  - ✓ It has a primary key

# 1NF Example

Repeating Groups

Not 1NF

Student No.	Name	Address 1	Town	Post code	Gender	Course Code	Course Name	Teacher	Date Completed	Grade
TR4678D	Joseph Harvey	214 Aston Road	Birmingham	B57YH	M	72846	German	12 - H Strausse	12-06-2011	B
TR4678D	Joseph Harvey	214 Aston Road	Birmingham	B57YH	M	20194	ICT	37 - D Barton	06-07-2011	B
YU7423O	Maria Woods	47 West Bridge St.	Birmingham	B47UH	F	20194	ICT	37 - D Barton	05-05-2011	C
YU7423O	Maria Woods	47 West Bridge St.	Birmingham	B47UH	F	56490	Geography	23 - A Revell	17-06-2011	D
YU7423O	Maria Woods	47 West Bridge St.	Birmingham	B47UH	F	72846	German	12 - H Strausse	13-06-2010	C

Student No.	Name	Address 1	Town	Post code	Gender
TR4678D	Joseph Harvey	214 Aston Road	Birmingham	B57YH	M
YU7423O	Maria Woods	47 West Bridge St.	Birmingham	B47UH	F

Student No.	Course Code	Course Name	Teacher	Date Completed	Grade
TR4678D	72846	German	12 - H Strausse	12-06-2011	B
TR4678D	20194	ICT	37 - D Barton	06-07-2011	B
YU7423O	20194	ICT	37 - D Barton	05-05-2011	C
YU7423O	56490	Geography	23 – A Revell	17-06-2011	D
YU7423O	72846	German	12 - H Strausse	13-06-2010	C

**Eliminate the repeating groups**

When we remove fields from a table to form another table, we must always take a copy of the primary key from the table we removed them from.

**Fields are not atomic**

**Not 1NF**

All fields are atomic

“Non-atomic” fields are those which should be split into more than 1 field because they are hiding detail. e.g. Name hides the detail of first name and last name.

Student No.	First Name	Last Name	Address 1	Town	Post code	Gender
TR4678D	Joseph	Harvey	214 Aston Road	Birmingham	B57YH	M
YU7423O	Maria	Woods	47 West Bridge St.	Birmingham	B47UH	F

PK

Student No.	Course Code	Course Name	Teacher ID	Teacher Initial	Teacher Lastname	Date Completed	Grade
TR4678D	72846	German	12	H	Strausse	12-06-2011	B
TR4678D	20194	ICT	37	D	Barton	06-07-2011	B
YU7423O	20194	ICT	37	D	Barton	05-05-2011	C
YU7423O	56490	Geography	23	A	Revell	17-06-2011	D
YU7423O	72846	German	12	H	Strausse	13-06-2010	C

already

1NF

# 1NF Exercise

How to normalize this table to 1NF?

Order No.	Order Date	Customer ID	Customer Name	Customer Address	Post Code	Pizza Code	Pizza Name	Quantity	Price	Total Price
1250	12/11/2012	AAC45	Barry Baker	42 New St. Birmingham	B27YH	P	Pepperoni	1	\$5	\$22.5
1250	12/11/2012	AAC45	Barry Baker	42 New St. Birmingham	B27YH	MF	Meat Feast	1	\$6.5	\$22.5
1250	12/11/2012	AAC45	Barry Baker	42 New St. Birmingham	B27YH	V	Vegetarian	2	\$5.5	\$22.5
1251	13/11/2012	ABD67	Joanne Harris	13 Corporation St. Birmingham	BS10 4QT	HM	Ham& Mushroo m	3	\$5	\$34
1251	13/11/2012	ABD67	Joanne Harris	13 Corporation St. Birmingham	BS10 4QT	MF	Meat Feast	2	\$6.5	\$34
1251	13/11/2012	ABD67	Joanne Harris	13 Corporation St. Birmingham	BS10 4QT	THO	The Hot One	1	\$6	\$34



Order No.	Order Date	Customer ID	Customer First Name	Customer Last Name	Customer Address 1	Town	Post Code	Total Price
1250	12/11/2012	AAC45	Barry	Baker	42 New St.	Birmingham	B27YH	\$22.5
1251	13/11/2012	ABD67	Joanne	Harris	13 Corporation St.	Birmingham	BS10 4QT	\$34

in 1NF

Order No.	Pizza Code	Pizza Name	Quantity	Price
1250	P	Pepperoni	1	\$5
1250	MF	Meat Feast	1	\$6.5
1250	V	Vegetarian	2	\$5.5
1251	HM	Ham& Mushroom	3	\$5
1251	MF	Meat Feast	2	\$6.5
1251	THO	The Hot One	1	\$6