

پوهنتون کابل

پوهنځی کمپیوتر ساینس

Introduction to Database and Data Models

Lectures 11-13

تهیه کننده : پوهنیار محمد شعیب "زرین خیل"
سال : 1389

***Introduction to Database
and Data Models
- Normalization***

11

By: M Shuaib Zarinkhail

2010

Example 1

Dependency on entire primary key

<u>EmpID</u>	<u>CourseTitle</u>	Name	DeptName	Salary	DateCompleted
--------------	--------------------	------	----------	--------	---------------

Dependency on only *part* of the key

EmpID, CourseTitle → DateCompleted
EmpID → Name, DeptName, Salary

Therefore, NOT in 2nd Normal Form!

Getting it into 2NF

Decomposed into two separate relations

<u>EmpID</u>	Name	DeptName	Salary
--------------	------	----------	--------

<u>EmpID</u>	<u>CourseTitle</u>	DateCompleted
--------------	--------------------	---------------

Both are full
functional
dependencies

Example 2

- ENGINEER(Emp#, Name, Dept, Svc#, SvcName)
- FDs:
 - Emp# → Name
 - Emp# → Dept
 - Svc# → SvcName

This relation is not in 2NF, Why?

Example 2 Cont...

- Possible Solution

ENGINEER(Emp#, Name, Dept)
SERVICE(Svc#, SvcName)

Does this solve the problem?

Example 2 Cont...

- Second Possible Solution

ENGINEER(Emp#, Name, Dept)

SERVICE(Svc#, SvcName)

ENGSERVICE(Emp#, Svc#)

Does this solve the problem?

Conversion to 2NF

- Start with 1NF format:
 - Write each key component on a separate line
 - Write original key on last line
 - Each component (theme) is a new table
 - Write dependent attributes after each key

PROJECT (PrjNum, PrjName)

EMPLOYEE (EmpNum, EmpName, JobClass, ChgHour)

ASSIGN (PrjNum, EmpNum, Hours)

Example 3

Title	Year	Length	Type	Studio	Star
Star Wars	1977	124	Color	Fox	C. Fisher
Star Wars	1977	124	Color	Fox	M. <u>Hamil</u>
Star Wars	1977	124	Color	Fox	H. Ford
Alien	1979	117	Color	Paramount	S. Weaver
Aliens	1986	137	Color	Paramount	S. Weaver
Alien3	1992	113	Color	Paramount	S. Weaver
Annie Hall	1977	93	Color	Warner Bros	W. Allen
Annie Hall	1977	93	Color	Warner Bros	D. <u>Keaton</u>
Chaplin	1992	124	B&W	MGM	R. Downey
Dr. Strangelove	1964	93	B&W	Paramount	R. Torn
Restoration	1995	117	Color	Miramax	R. Downey

Example 3 Cont...

- Is the example in 1NF?
- What is the primary key for this relation?
- What functional dependencies exist in this relation?

Example 3 Cont...

- Is the example in 2NF?
- What is the partial dependency for this relation?
- **Partial Dependency:** When a Non-key attribute depends on part of a primary key

Example 3 Cont...

- Is the example in 2NF?
- If not, how do you fix it?

Example 4

- CHILD(Child#, ChildName, Preschool#, PreschoolLocation)
- FDs
 - Child# → ChildName
 - Child# → Preschool#
 - Preschool# → PreschoolLocation

Why isn't this relation in 3NF?

Example 4 Cont...

Solution:

- CHILD(Child#, ChildName, *Preschool#*)
- PRESCHOOL(Preschoo#,
PreschoolLocation)

Example 5

VIN	Make	Model	Year	NID	Owner
1119	Toyota	Corolla	1988	3373	Jawid
2235	Ford	Windstar	1998	4992	Karim
4522	GM	GMC	1995	5268	Sahar
3351	Sobarou	Outback	2000	4439	Daud

**VEHICLE Relation:
VIN is the Primary Key!**

Example 5 Cont...

- This relation is in what normal form?
- What are the functional dependencies?
- Are there any transitive dependencies?
- Are there any partial dependencies?

- How do we upgrade it to 3NF?

Example 6

SALES(Cust_ID, Name, SalesPerson, Region)

SALES

<u>Cust_ID</u>	Name	Salesperson	Region
8023	Anderson	Smith	South
9167	Bancroft	Hicks	West
7924	Hobbs	Smith	South
6837	Tucker	Hernandez	East
7018	Arnold	Faulb	North
8596	Eckersley	Hicks	West

27

Example 6 Cont...

Solution

- SALESPERSON(SalesPerson, Region)
- SALES(Cus_tID, Name, *SalesPerson*)

Example 7

SHIPMENT(Snum, Origin, Destination, Distance)

SHIPMENT

<u>Snum</u>	Origin	Destination	Distance
409	Seattle	Denver	1,537
618	Chicago	Dallas	1,058
723	Boston	Atlanta	1,214
824	Denver	Los Angeles	1,150
629	Minneapolis	St. Louis	587

Example 7 Cont...

Solution:

- SHIPTO(Snum, *Origin*, *Destination*)
- DISTANCES(Origin, Destination, Distance)

Boyce-Codd Normal Form (BCNF)

- A relation is in BCNF if all determinants are primary or candidate keys
- **Determinant:** the attribute (s) on the left side of a functional dependency
- **Candidate Key:** attribute (s) that could serve as a PK, but for some reason was not selected as PK

Example 8

- **ADVISOR(SID, Major, AName)**
 - A student can have one or more majors
 - A major can have several faculty members as advisors
 - A faculty member advises in only one major area
 - No two faculty members have the same name
 - A faculty member can advise zero or more students

Example 8 Cont...

- Candidate Key: (SID, AName)
- FD: AName \rightarrow Major
- This relation is at what level of normalization?
- How do we get this into BCNF?

***Introduction to Database
and Data Models
- Normalization***

12

By: M Shuaib Zarinkhail

2010

Example 09

	PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
▶	5	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

Figure 4.1

Ex. 09: Observations

- PROJ_NUM intended to be primary key
- Table entries invite data inconsistencies
- Table displays data anomalies
 - Update
 - Modifying JOB_CLASS
 - Insertion
 - New employee must be assigned project
 - Deletion
 - If employee deleted, other vital data lost

Ex. 09: Conversion to 1NF

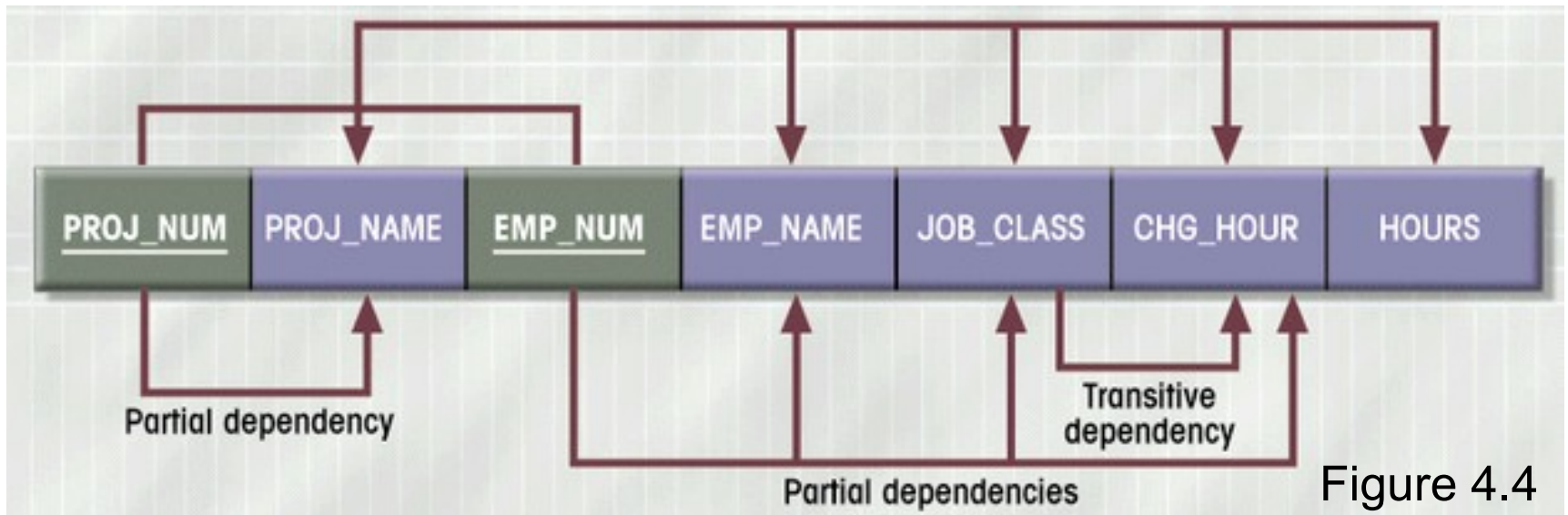
- Repeating groups must be eliminated
 - Proper primary key developed
 - Uniquely identify attribute values (rows)
 - Combination of PROJ_NUM and EMP_NUM

Ex. 09: Conversion to 1NF cont...

Repeating groups must be eliminated

- Dependencies can be identified
 - Desirable dependencies based on PK
 - Less desirable dependencies
 - Partial
 - based on part of composite PK
 - Transitive
 - one non-key attribute depends on another non-key attribute

Dependency Diagram (1NF)



Data Organization: 1NF

	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

Figure 4.3

1NF Summarized

- All key attributes defined
- No repeating groups in table
- All attributes dependent on primary key

Conversion to 2NF

- Start with 1NF format:
- Write each key component on separate line
- Write original key on last line
- Each component creates a new table
- Write dependent attributes after each key

```
PROJECT (PROJ_NUM, PROJ_NAME)
EMPLOYEE (EMP_NUM, EMP_NAME, JOB_CLASS, CHG_HOUR)
ASSIGN (PROJ_NUM, EMP_NUM, HOURS)
```


2NF Conversion Results

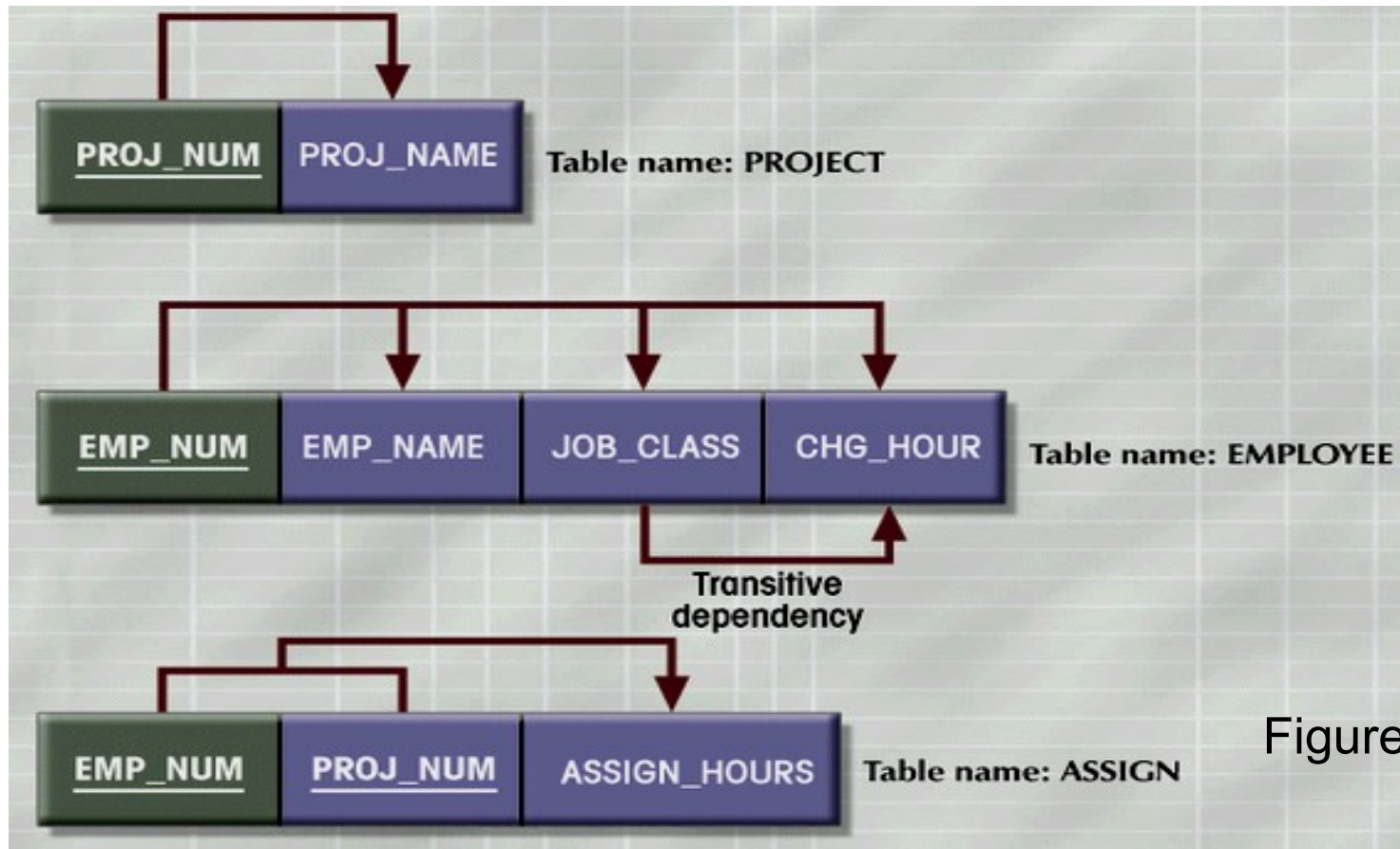


Figure 4.5

2NF Summarized

- In 1NF
- Includes no partial dependencies
 - No attribute dependent on a portion of PK
- Still possible to exhibit transitive dependency
 - Attributes may be functionally dependent on non-key attributes

Conversion to 3NF

- Create separate table(s) to eliminate transitive dependencies

PROJECT (PROJ_NUM, PROJ_NAME)
ASSIGN (PROJ_NUM, EMP_NUM, HOURS)
EMPLOYEE (EMP_NUM, EMP_NAME, JOB_CLASS)
JOB (JOB_CLASS, CHG_HOUR)

3NF Summarized

- In 2NF
- Contains no transitive dependencies

Additional DB Enhancements

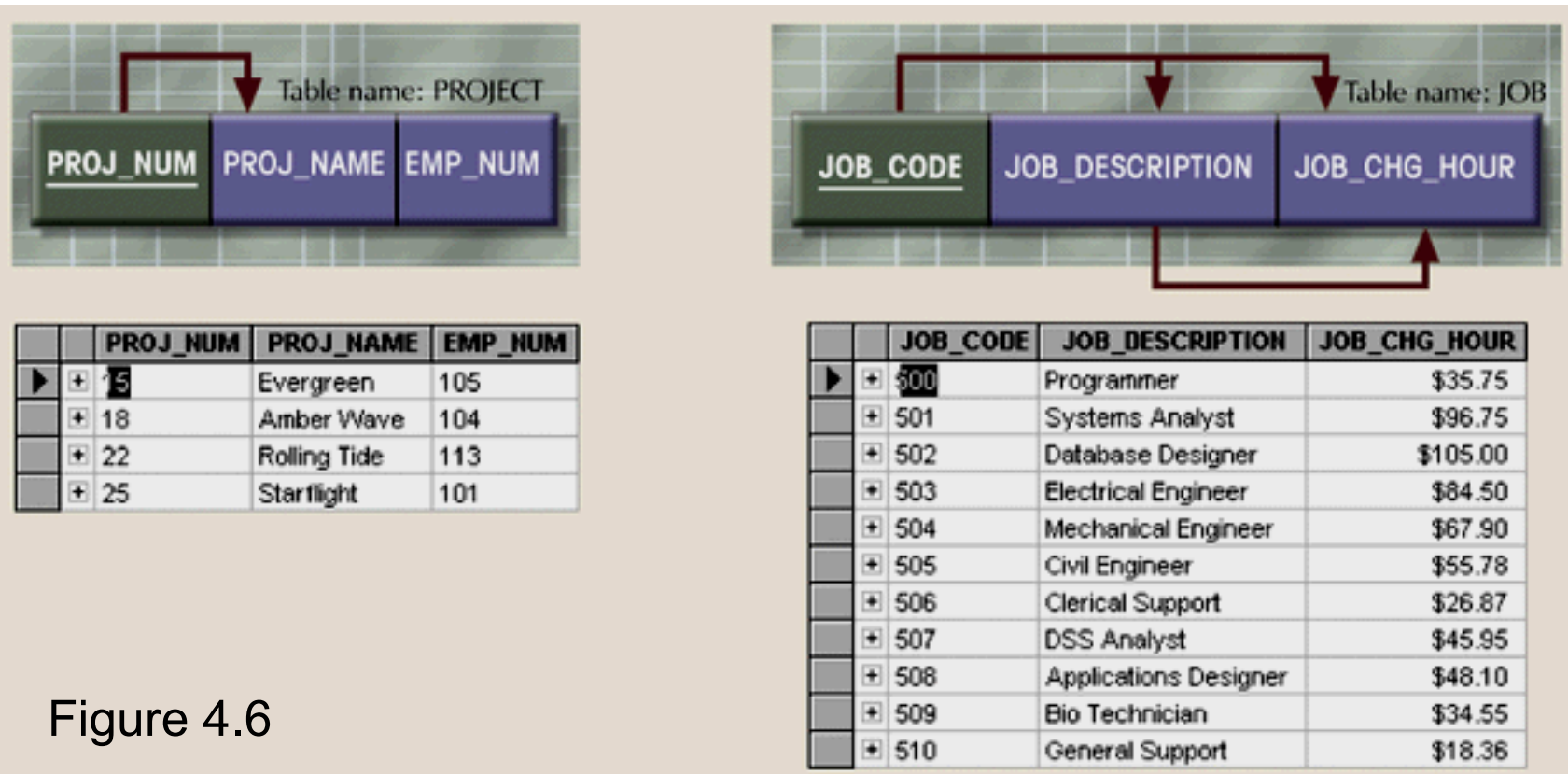
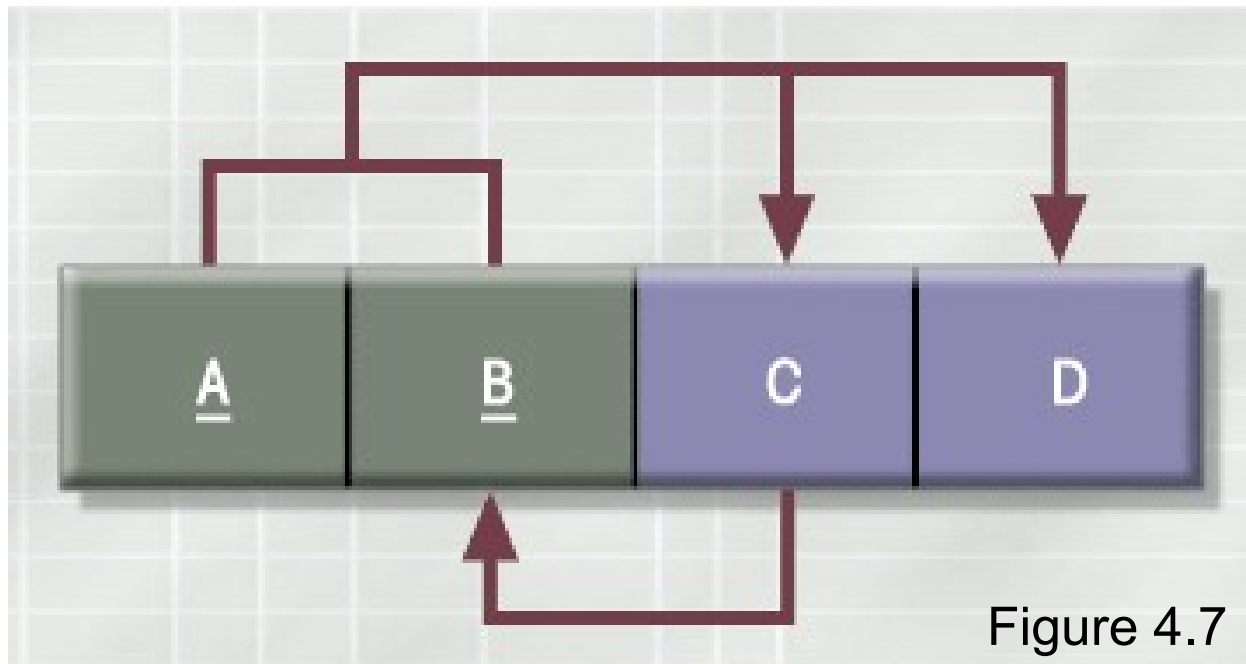


Figure 4.6

Boyce-Codd Normal Form (BCNF)

- Every determinant in the table is a candidate key
- Determinant is attribute whose value determines other values in a row
- 3NF table with one candidate key is already in BCNF

3NF Table Not in BCNF



Decomposition of Table Structure to Meet BCNF

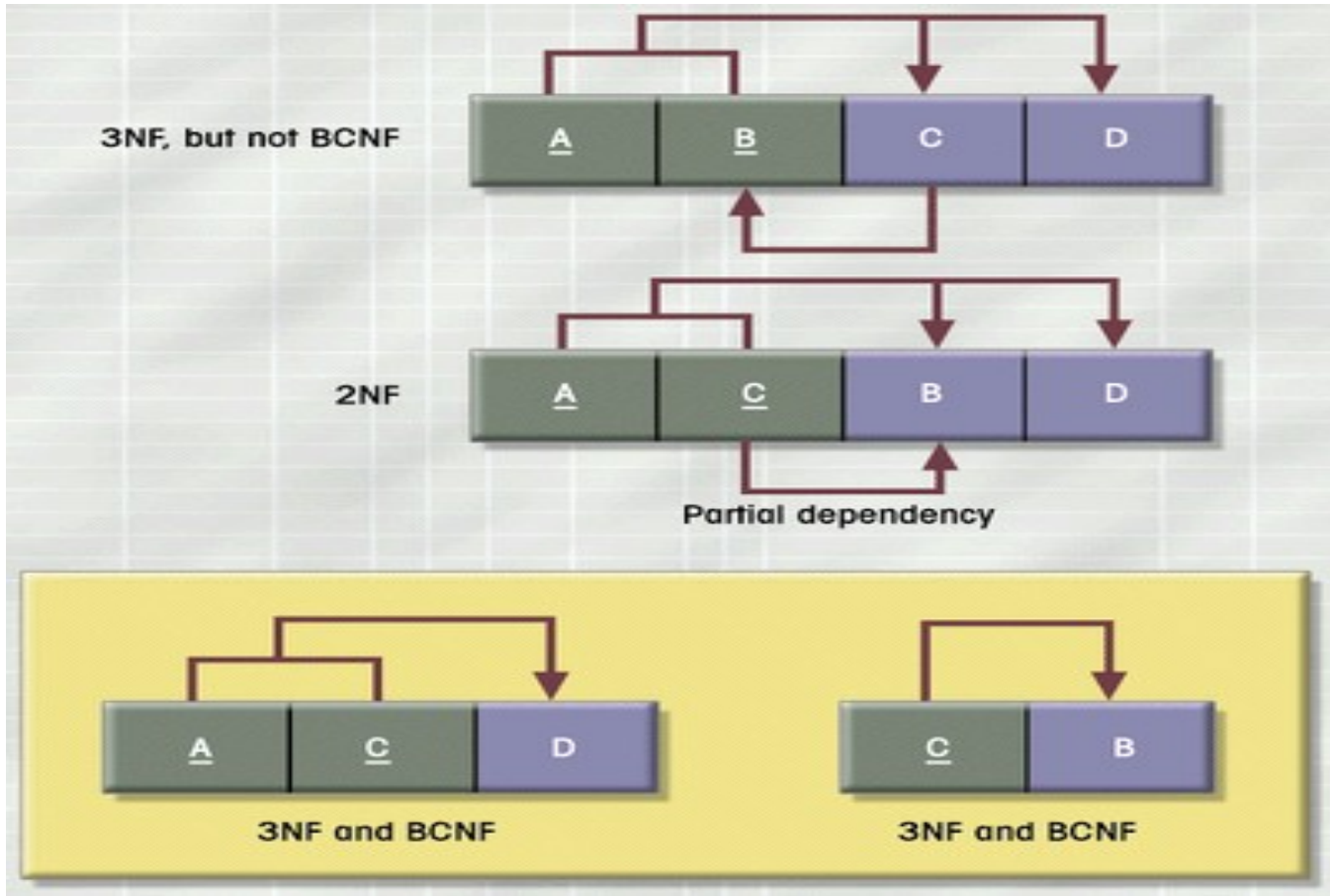


Figure 4.8

Decomposition into BCNF

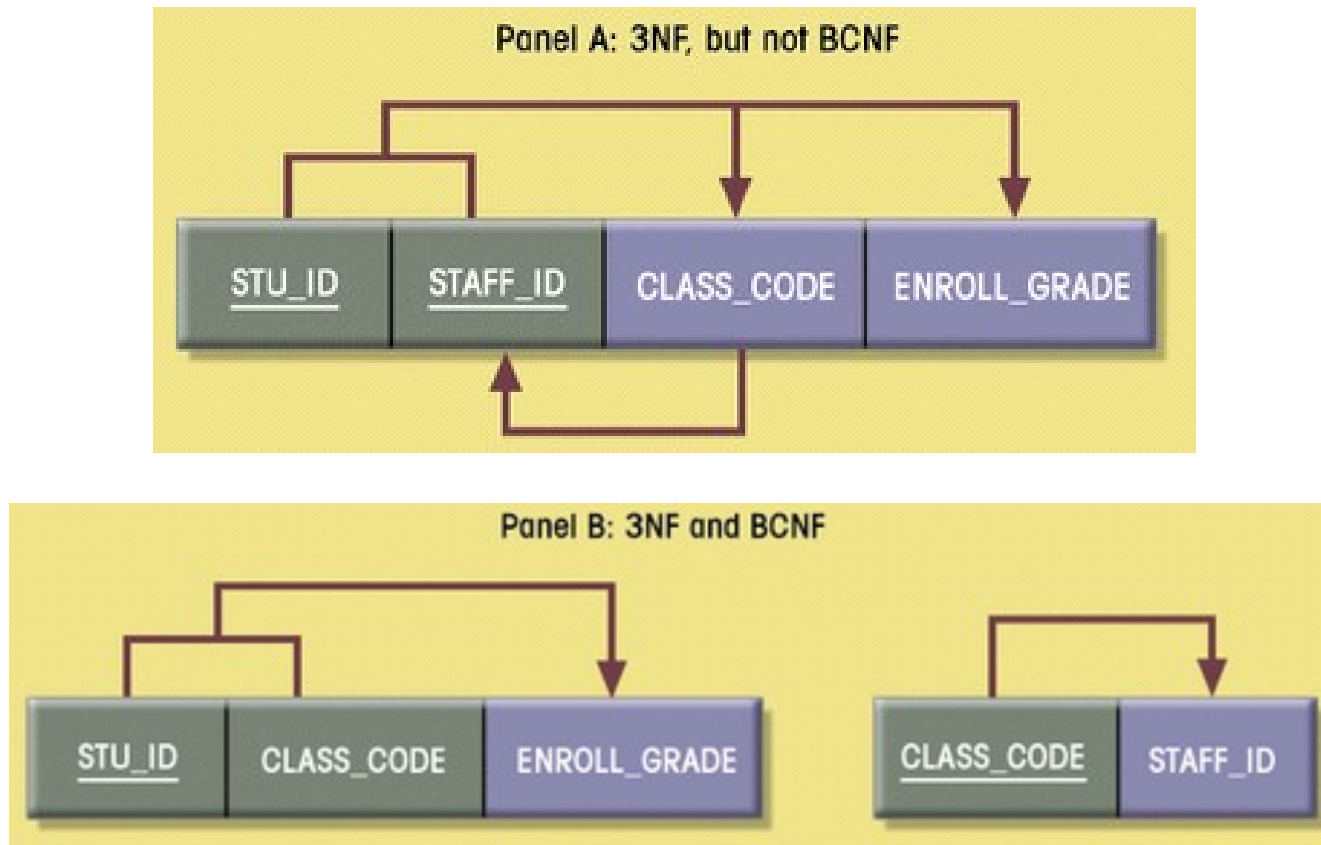


Figure 4.9

***Introduction to Database
and Data Models
- Normalization***

13

By: M Shuaib Zarinkhail

2010

Ex. 10 - 1NF

CourseID	CourseName	StudName	SSN	DOB	Prof	Bldg	Floor	RmNo	Grade
10058	CSIS 2520	Sam Hall	254638589	5/19/59	Harley	BB	4	405	C
10058	CSIS 2520	Ted Ringly	752135423	11/7/72	Harley	BB	4	405	B
25600	ART 1101	Jane King	232522654	1/5/81	Bartett	PI	2	201	A
38569	MATH 4454	John Smith	458963232	8/1/69	Harton	CL	1	106	D
10654	CSIS 3600	Jane King	232522654	1/5/81	Snipes	JM	2	217	A

Ex. 10 - 2NF

Grade

CourseID	SSN	Grade
10058	254638589	C
10058	752135423	B
25600	232522654	A
38569	458963232	D
10654	232522654	A

Student

SSN	StudName	DOB
254638589	Sam Hall	5/19/59
752135423	Ted Ringly	11/7/72
232522654	Jane King	1/5/81
458963232	John Smith	8/1/69

Course

CourseID	CourseName	Prof	Bldg	Floor	RmNo
10058	CSIS 2520	Harley	BB	4	405
25600	ART 1101	Bartett	PI	2	201
38569	MATH 4454	Harton	CL	1	106
10654	CSIS 3600	Snipes	JM	2	217

Ex. 10 - 3NF

Course

CourseID	CourseName	Prof	Bldg	RmNo
10058	CSIS 2520	Harley	BB	405
25600	ART 1101	Bartett	PI	201
38569	MATH 4454	Harton	CL	106
10654	CSIS 3600	Snipes	JM	217

Grade

CourseID	SSN	Grade
10058	254638589	C
10058	752135423	B
25600	232522654	A
38569	458963232	D
10654	232522654	A

Floor

RmNo	Floor
405	4
201	2
106	1
217	2

Student

SSN	StudName	DOB
254638589	Sam Hall	5/19/59
752135423	Ted Ringly	11/7/72
458963232	John Smith	8/1/69
232522654	Jane King	1/5/81

Ex. 10 - Is it in BCNF?

Yes ...

- Two possibilities here:
 1. CourseName is a candidate key (only one section per course) and determinant
 3. More than one section per course, then CourseName wouldn't be a determinant or candidate key

Example 11

- A cleaning company keeps records of its client offices
 - the employees that clean each office
 - and the amounts they are charged
- A record is kept each time an office is cleaned
- We will assume that no companies have the same name

Ex. 11

<u>Client Name</u>	Date	Address	EmpNo	Name	Service	Amount Due
Express Data	10/5/02	1025 Cobb Place	103	Joey Marx	Deluxe	\$255.99
Harry's Co.	8/9/02	432 Dallas St.	89	Mary Whittley	Basic	\$99.99
L's Sandwich	9/15/02	9876 Market Lane	103	Joey Marx	Basic	\$99.99
Harry's Co.	11/15/02	432 Dallas St.	89	Mary Whittley	Standard	\$145.99

Ex. 11 - 2NF

Client Name	Address
Express Data	1025 Cobb Place
Harry's Co.	432 Dallas St.

<u>Client Name</u>	Date	EmpNo	Name	Service	Amount Due
Express Data	10/5/02	103	Joey Marx	Deluxe	\$255.99
Harry's Co.	8/9/02	89	Mary Whittley	Basic	\$99.99
L's Sandwich	9/15/02	103	Joey Marx	Basic	\$99.99
Harry's Co.	11/15/02	89	Mary Whittley	Standard	\$145.99

Ex. 11 - 3NF

Client Name	Address
Express Data	1025 Cobb Place
Harry's Co.	432 Dallas St.

EmpNo.	Name
103	Joey Marx
89	Mary Whittlely

Service	Amount Due
Deluxe	\$255.99
Basic	\$99.99

<u>Client Name</u>	Date	EmpNo.	Service
Express Data	10/5/02	103	Deluxe
Harry's Co.	8/9/02	89	Basic
L's Sandwich	9/15/02	103	Basic
Harry's Co.	11/15/02	89	Standard

Ex. 11 - BCNF

- Already in BCNF (All determinants are primary or candidate keys)

Example 12

GROCERY STORE INVENTORY

- Problem Statement: You need to normalize an inventory database for a small town grocery store
- Assume that a given product will only be stored in one place (i.e. aisle) in the store

Ex. 12 - 1NF

Brand	Product	Size	Aisle
Hunt's	Canned tomatoes	#2 can	3
Contadina	Canned tomatoes	#3 can	3
Hunt's	Canned tomatoes	#3 can	3
Hunt's	Ketchup	12 oz.	5

Ex. 12 - 2NF

Brand	Product	Size
Hunt's	Canned tomatoes	#2 can
Contadina	Canned tomatoes	#3 can
Hunt's	Canned tomatoes	#3 can
Hunt's	Ketchup	12 oz.

Product	Aisle
Canned tomatoes	3
Ketchup	5

Ex. 12 - 3NF

- Already in 3NF because no transitive dependencies exist

Ex. 12 - BCNF

- Already in BCNF because all determinants are primary or candidate keys

Example 13

Prescription Exercise

- Problem Description: A pharmacy keeps track of their prescriptions using a database. When a patient comes in, the doctor's prescription is entered into the database, and any allowed generic drugs substitution are also entered.

Example 13

Assume the following:

- Rx# does not change when it is refilled
- A TrademarkDrug has only one GenericDrug substitute
- Date is modified every time a prescription is issued or refilled

Ex. 13 - 1NF

PatientID	Name	Address	RX_#	Trademark Drug	Generic Drug	Date
12345	Pheobe Townsend	123 Main St	123456	Zoloft	Sertraline	2/15/02
12345	Townsend	123 Main St	234567	Zoloft	Sertraline	3/15/02
56789	Thomas	5 West Ave	345678	Daypro	Oxyprozin	4/5/02
89012	Lisa Martin	7 Monroe Ave	456789	Paxil	Poroxetine	10/5/02
99910	Marcus Swift	100 Genesee Way	567890	Zoloft	Sertraline	8/1/01

Ex. 13 - 2NF

PatientID	Name	Address
12345	Pheobe Townsend	123 Main St
56789	John Thomas	5 West Ave
89012	Lisa Martin	7 Monroe Ave
99910	Marcus Swift	100 Genesee Way

RX_#	Date
123456	2/15/02
234567	3/15/02
345678	4/5/02
456789	10/5/02
567890	8/1/01

RX_#	Trademark Drug	Generic Drug	Patient ID.
123456	Zoloft	Sertraline	12345
234567	Zoloft	Sertraline	12345
345678	Daypro	Oxyprozin	56789
456789	Paxil	Poroxetine	89012
567890	Zoloft	Sertraline	99910

Ex. 13 - 3NF

PatientID	Name	Address
12345	Pheobe Townsend	123 Main St
56789	John Thomas	5 West Ave
89012	Lisa Martin	7 Monroe Ave
99910	Marcus Swift	100 Genesee Way

RX_#	Date
123456	2/15/02
234567	3/15/02
345678	4/5/02
456789	10/5/02
567890	8/1/01

RX_#	Trademark Drug	Patient ID
123456	Zoloft	12345
234567	Zoloft	12345
345678	Daypro	56789
456789	Paxil	89012
567890	Zoloft	99910

Trademark Drug	Generic Drug
Zoloft	Sertraline
Daypro	Oxyprozin
Paxil	Poroxetine

Ex. 13 - BCNF

- Already in BCNF, because all determinants are primary or candidate keys