Ch 4

Database Management Systems

Sumayyea Salahuddin (Lecturer)
Dept. of Computer Systems Eng.
UET Peshawar

Overview

- Normalization Review
- Boyce – Codd Normal Form [BCNF]
- Higher Normal Forms
- Case Study – Contacting Company

Limitations on System-Assigned Keys

- System-assigned primary key may not prevent confusing entries
- Data entries in Table 5.2 are inappropriate because they duplicate existing records
  - Yet there has been no violation of either entity integrity or referential integrity

TABLE 5.2 DUPLICATE ENTRIES IN THE JOB TABLE

JOB_CODE	JOB_DESCRIPTION	JOB_CHG_HOUR
511	Programmer	$35.75
512	Programmer	$35.75

The Boyce-Codd Form (BCNF)

- Every determinant in the table is a candidate key
  - Has same characteristics as primary key, but for some reason, not chosen to be primary key
- If a table contains only one candidate key, the 3NF and the BCNF are equivalent
- BCNF can be violated only if the table contains more than one candidate key
- Most designers consider the Boyce-Codd normal form (BCNF) as a special case of 3NF
- A table is in 3NF if it is in 2NF and there are no transitive dependencies

The Boyce-Codd Form (BCNF) cont…

- A table can be in 3NF and not be in BCNF
  - A transitive dependency exists when one nonprime attribute is dependent on another nonprime attribute
  - A non-key attribute is the determinant of a key attribute

A Table that is in 3NF but not in BCNF

FIGURE 5.7 A TABLE THAT IS IN 3NF BUT NOT IN BCNF

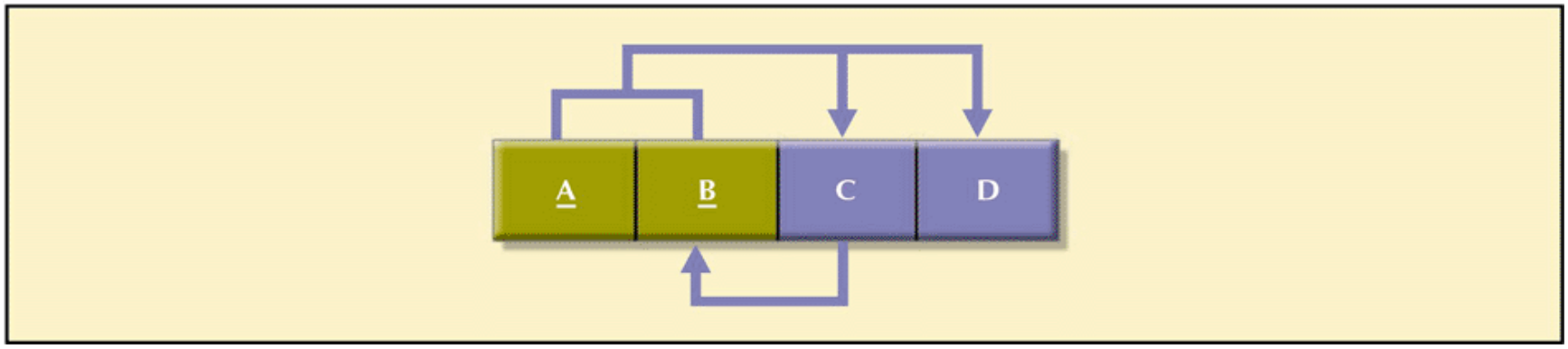
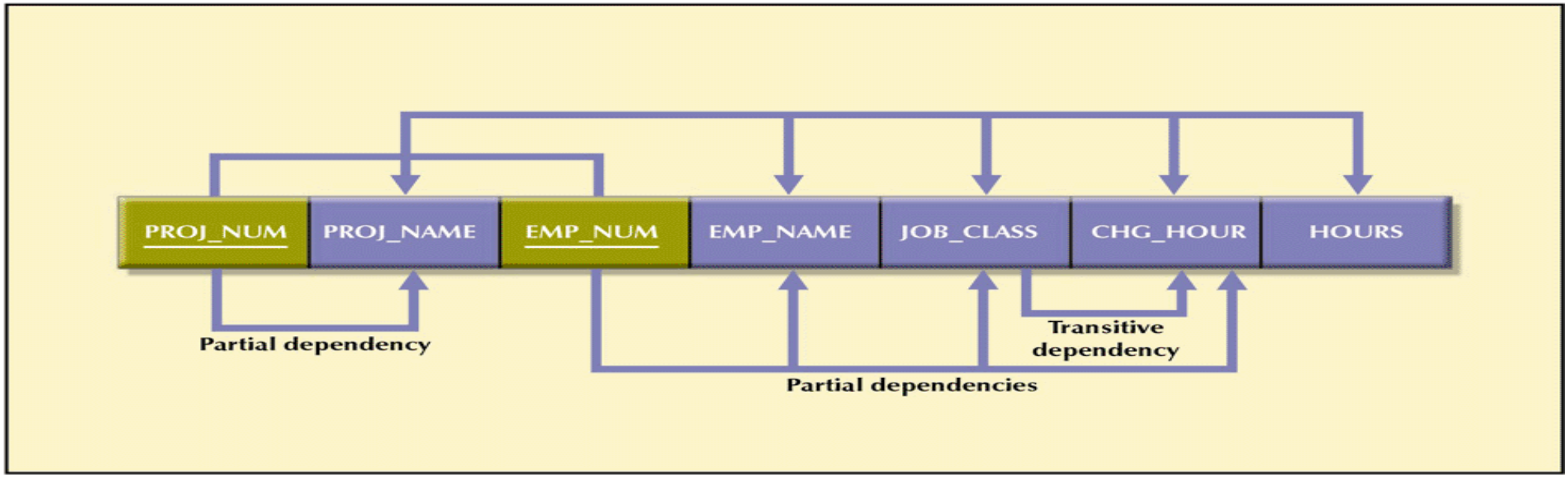
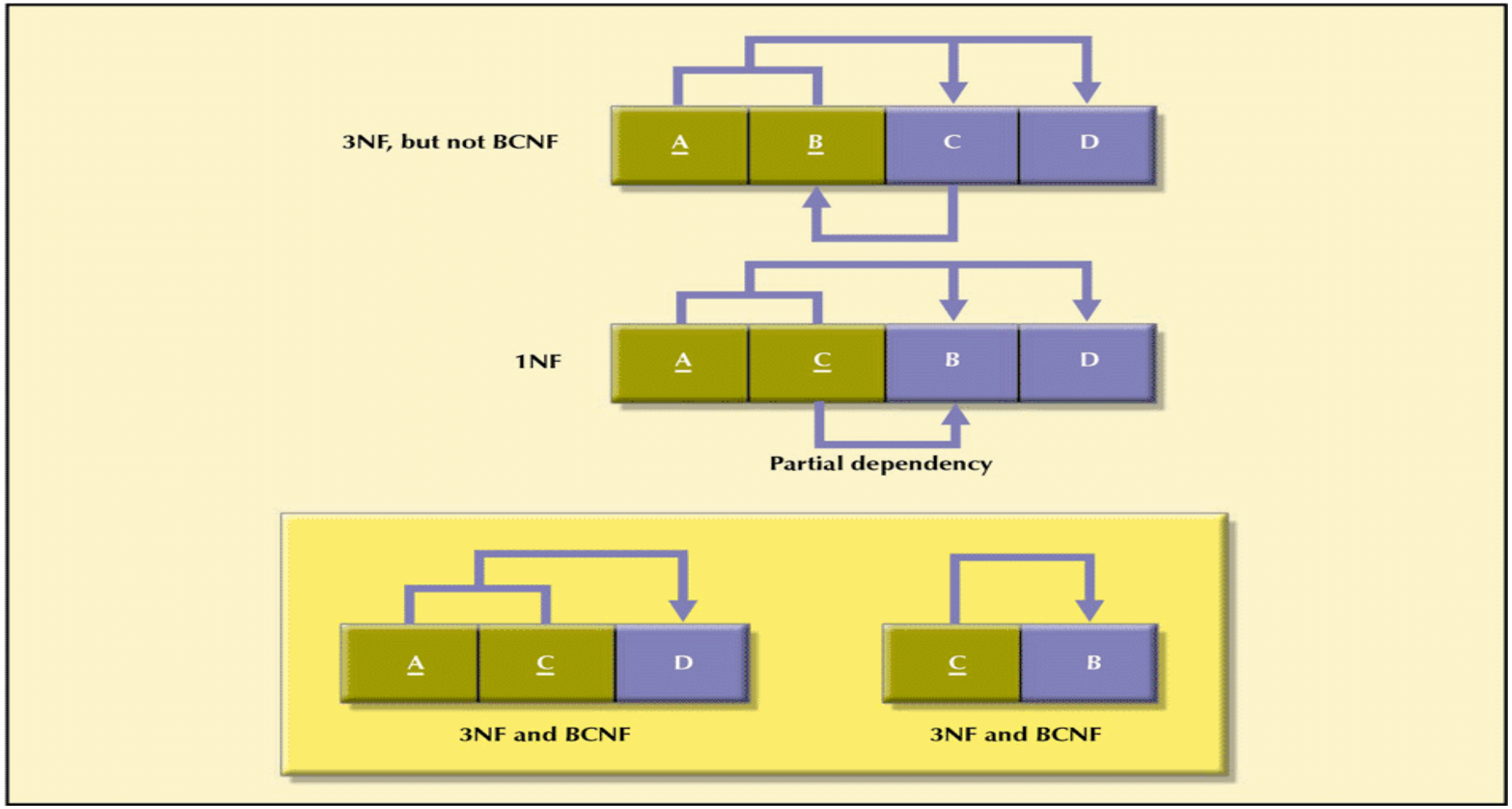


FIGURE 5.3 A DEPENDENCY DIAGRAM: FIRST NORMAL FORM (1NF)



Decomposition to BCNF

FIGURE 5.8 DECOMPOSITION TO BCNF



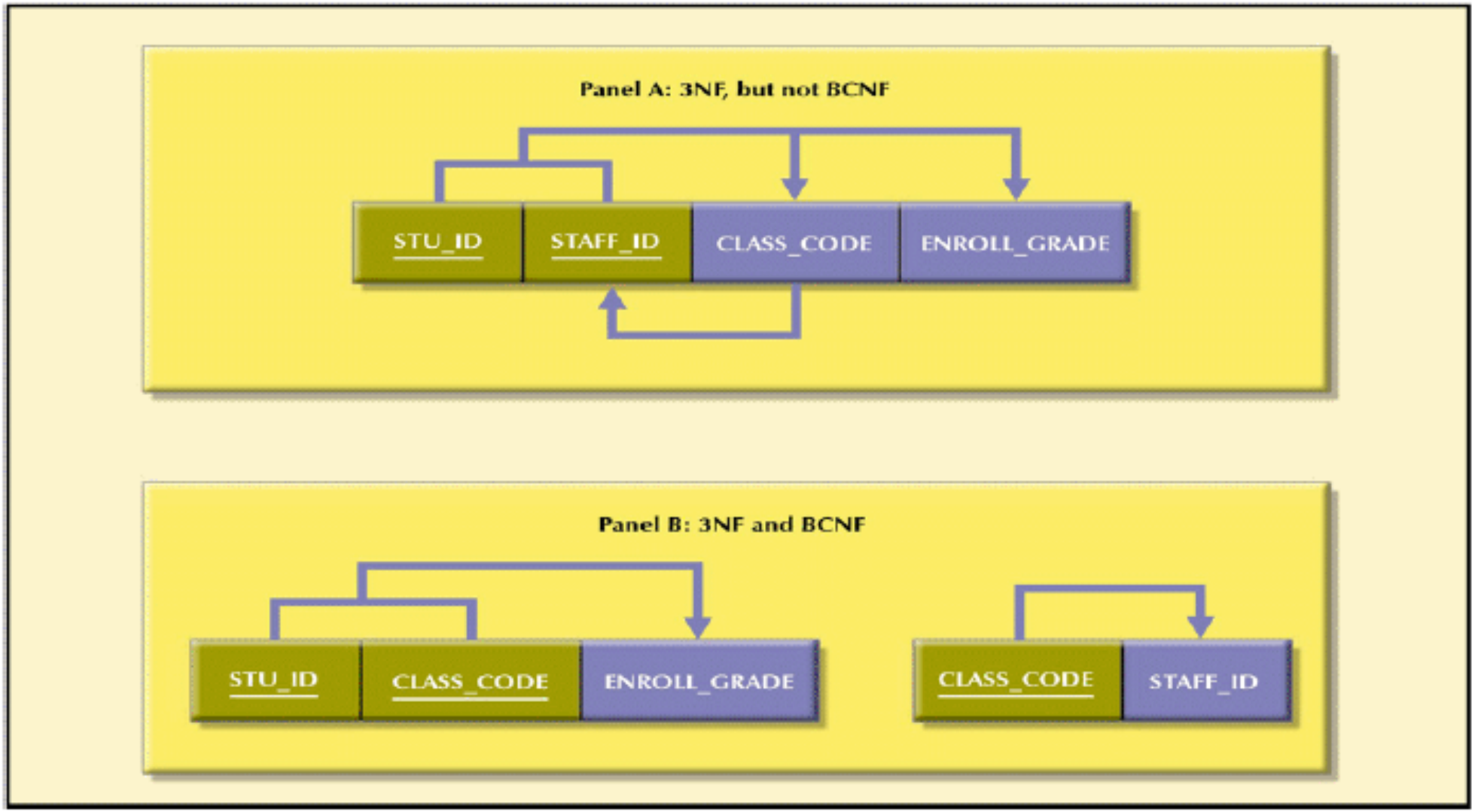
Sample Data for a BCNF Conversion

TABLE 5.3 SAMPLE DATA FOR A BCNF CONVERSION

STU_ID	STAFF_ID	CLASS_CODE	ENROLL_GRADE
125	25	21334	A
125	20	32456	C
135	20	28458	B
144	25	27563	C
144	20	32456	B

Another BCNF Decomposition

FIGURE 5.9 ANOTHER BCNF DECOMPOSITION



Higher-Level Normal Forms

- In some databases, multiple multi-valued attributes exist

FIGURE 5.15 TABLES WITH MULTIVALUED DEPENDENCIES

Database name: Ch05_Service

Table name: VOLUNTEER_V1

EMP_NUM	ORG_CODE	ASSIGN_NUM
10123	RC	1
10123	UW	3
10123		4

Table name: VOLUNTEER_V2

EMP_NUM	ORG_CODE	ASSIGN_NUM
10123	RC	
10123	UW	
10123		1
10123		3
10223		4

Table name: VOLUNTEER_V3

EMP_NUM	ORG_CODE	ASSIGN_NUM
10123	RC	1
10123	RC	3
10123	UW	4

Fourth Normal Form

- Table is in fourth normal form (4NF) if
  - It is in 3NF
  - Has no multiple sets of multi-valued dependency
- 4NF is largely academic if table conform to the following two rules
  - All attributes are dependent on primary key but independent of each other
  - No row contain two or more multi-valued facts about an entity

A Set of Tables in 4NF

FIGURE 5.16 A SET OF TABLES IN 4NF

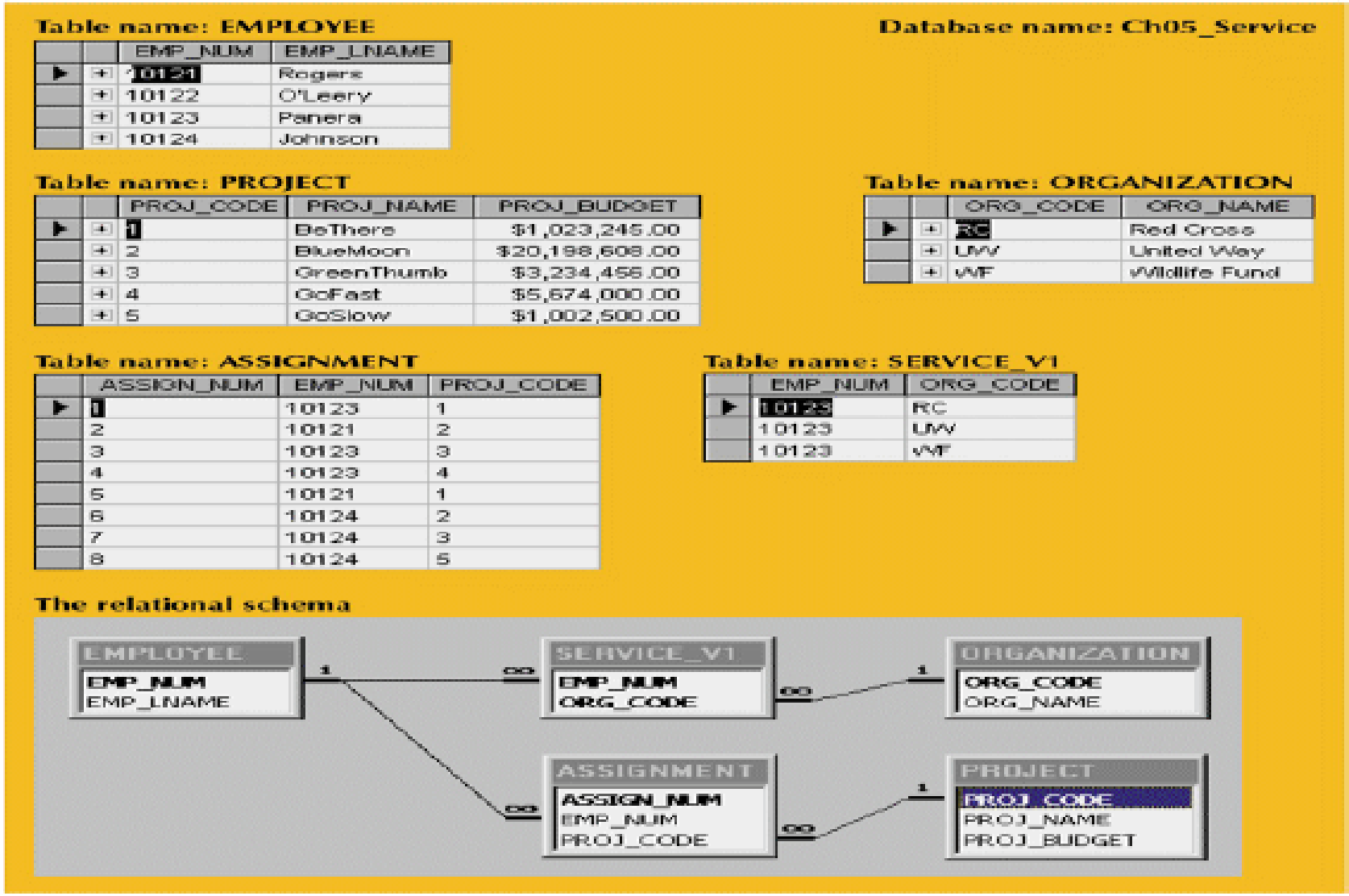


Table name: EMPLOYEE

Database name: Ch05_Service

EMP_NUM	EMP_LNAME
10121	Rogers
10122	O'Leery
10123	Panera
10124	Johnson

Table name: PROJECT

PROJ_CODE	PROJ_NAME	PROJ_BUDGET
1	BeThere	$1,023,245.00
2	BlueMoon	$20,198,608.00
3	GreenThumb	$3,234,456.00
4	GoFast	$5,674,000.00
5	GoSlow	$1,002,500.00

Table name: ORGANIZATION

ORG_CODE	ORG_NAME
RC	Red Cross
UW	United Way
WF	Wildlife Fund

Table name: ASSIGNMENT

ASSIGN_NUM	EMP_NUM	PROJ_CODE
1	10123	1
2	10121	2
3	10123	3
4	10123	4
5	10121	1
6	10124	2
7	10124	3
8	10124	5

Table name: SERVICE_V1

EMP_NUM	ORG_CODE
10123	RC
10123	UW
10123	WF

The relational schema

Normalization and Database Design

- Normalization should be part of design process
- Make sure that proposed entities meet required normal form before table structures are created
- Many real-world databases have been improperly designed or burdened with anomalies if improperly modified during course of time
- You may be asked to redesign and modify existing databases
- ER Diagram
  - Provides the big picture, or macro view, of an organization's data requirements and operations
  - Created through an iterative process
    - Identifying relevant entities, their attributes ad their relationship
    - Use results to identify additional entities and attributes

Normalization and Database Design (cont…)

- Normalization procedures
  - Focus in the characteristics of specific entities
  - A micro view of the entities within the ER diagram
- Difficult to separate normalization process from ER modeling
- Two techniques (i.e. ER & Normalizations) should be used concurrently

Case Study

CONTRACTING COMPANY

Business Rules of Contracting Company

- The company manages many projects
- Each project requires the services of many employees
- An employee may be assigned to several different projects
- Some employees are not assigned to a project and perform duties not specifically related to a project. Some employee are part of a labor pool, to be shared by all project teams. For instance, company's executive secretary not assigned to any one particular project
- Each employee has a single job classification. That job classification determines the hourly billing rate
- Many employees can have the same job classification. For instance, the company employs more than one electrical engineer

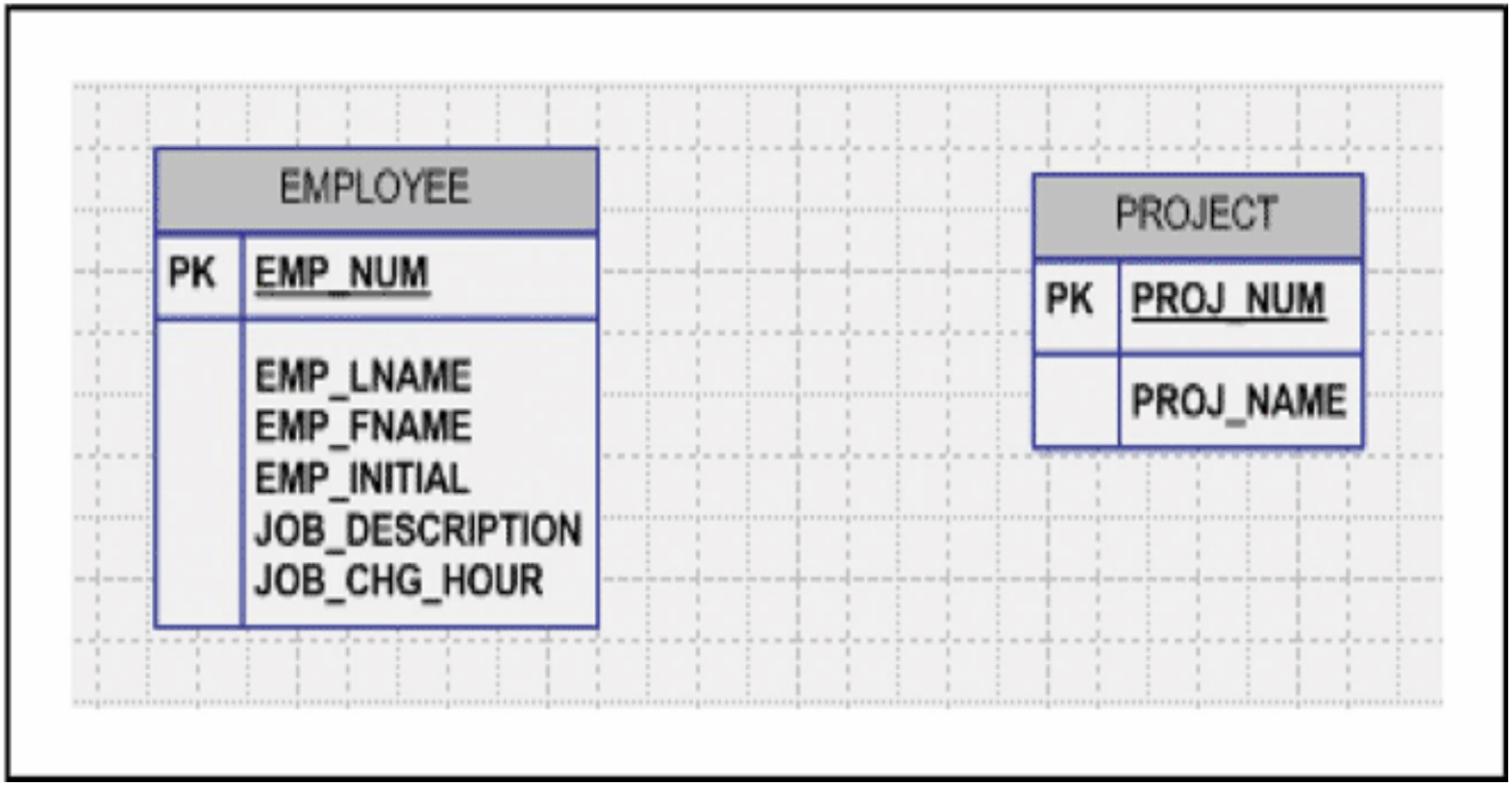
Initial Relational Schema

- PROJECT (PROJ_NUM, PROJ_NAME)
- EMPLOYEE (EMP_NUM, EMP_LNAME, EMP_FNAME, EMP_INITIAL, JOB_DESCRIPTION, JOP_CHG_HOUR)

- Is PROJECT in normal form? If yes, which one? If not, what dependencies exists?
- Is EMPLOYEE in normal form? If yes, which one? If not, what dependencies exists?

The Initial ERD for a Contracting Company

FIGURE 5.10 THE INITIAL ERD FOR A CONTRACTING COMPANY



Modified Relational Schema

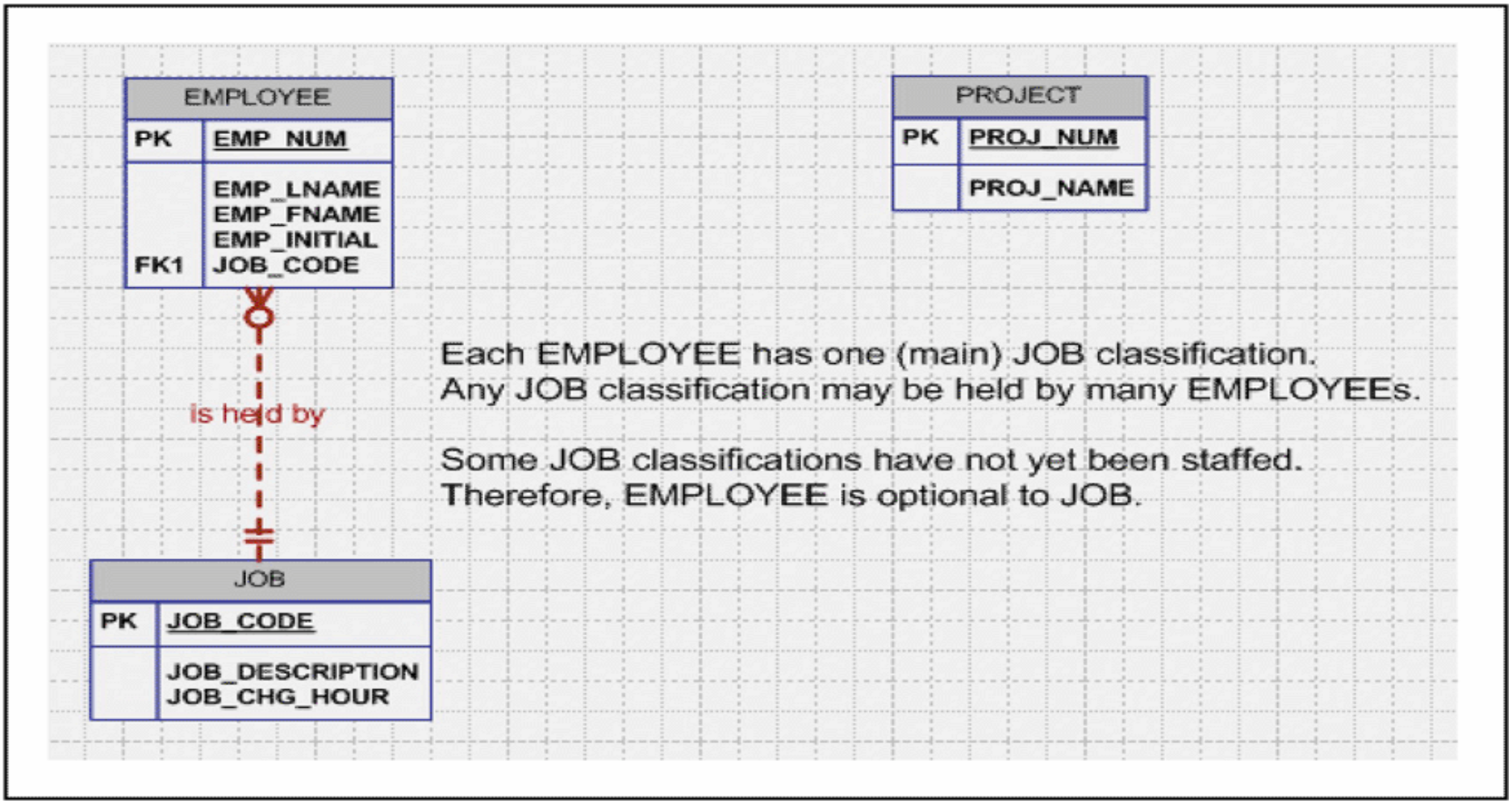
- PROJECT is in 3NF and needs no modifications.
- In EMPLOYEE schema, transitive dependency exists.
  - The JOB_DESCCRIPTION attribute defines job classifications such as Systems Analyst, Database Designer, and Programmer.
  - Job classifications determine the billing rate, JOB_CHG_HOUR.

Updated Schema:

- PROJECT (PROJ_NUM, PROJ_NAME)
- EMPLOYEE (EMP_NUM, EMP_LNAME, EMP_FNAME, EMP_INITIAL, JOB_CODE)
- JOB (JOB_CODE, JOB_DESCRIPTION, JOP_CHG_HOUR)

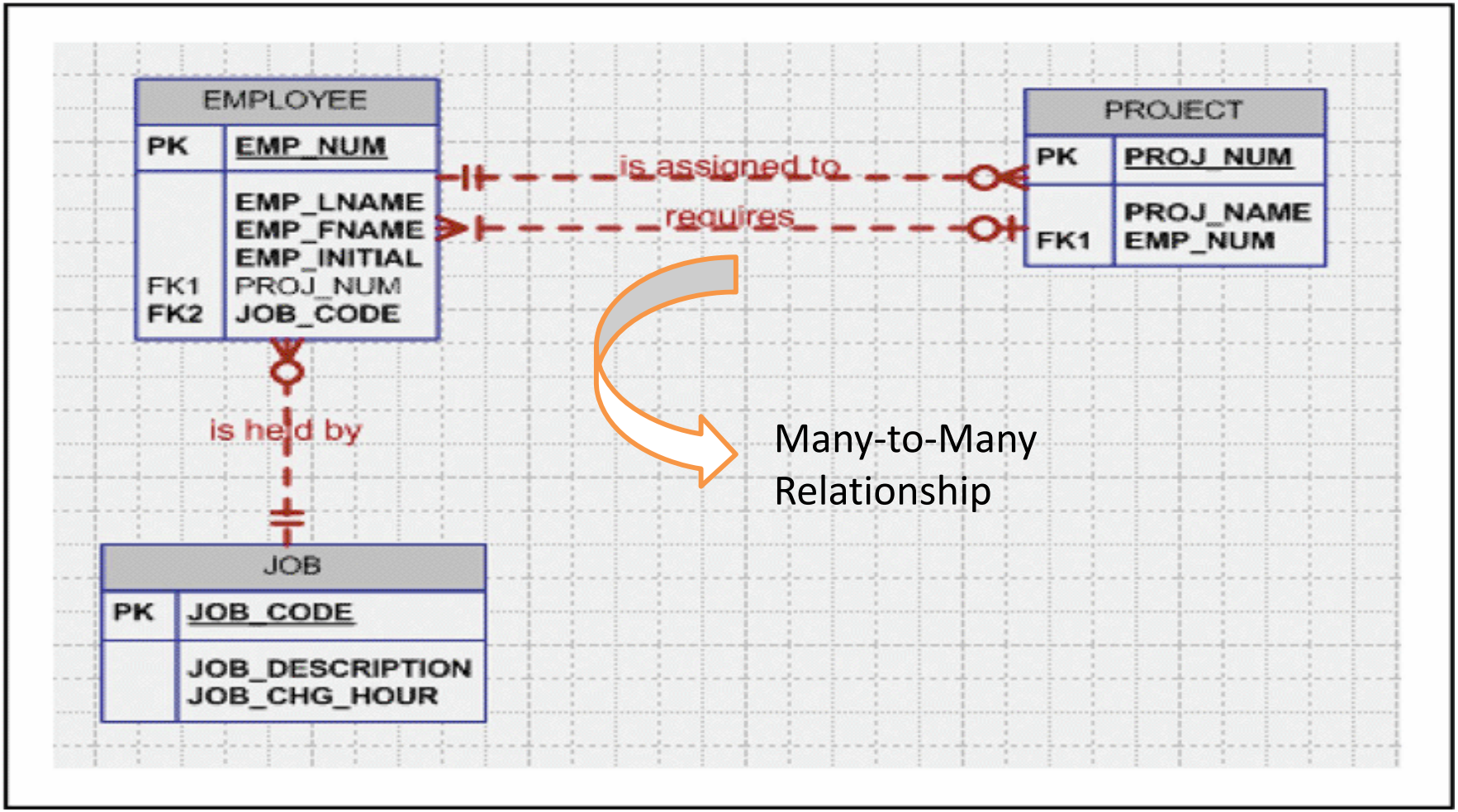
The Modified ERD for a Contracting Company

FIGURE 5.11 THE MODIFIED ERD FOR A CONTRACTING COMPANY



The Incorrect Representation of a M:N Relationship

FIGURE 5.12 THE INCORRECT REPRESENTATION OF A M:N RELATIONSHIP



Final Relational Schema

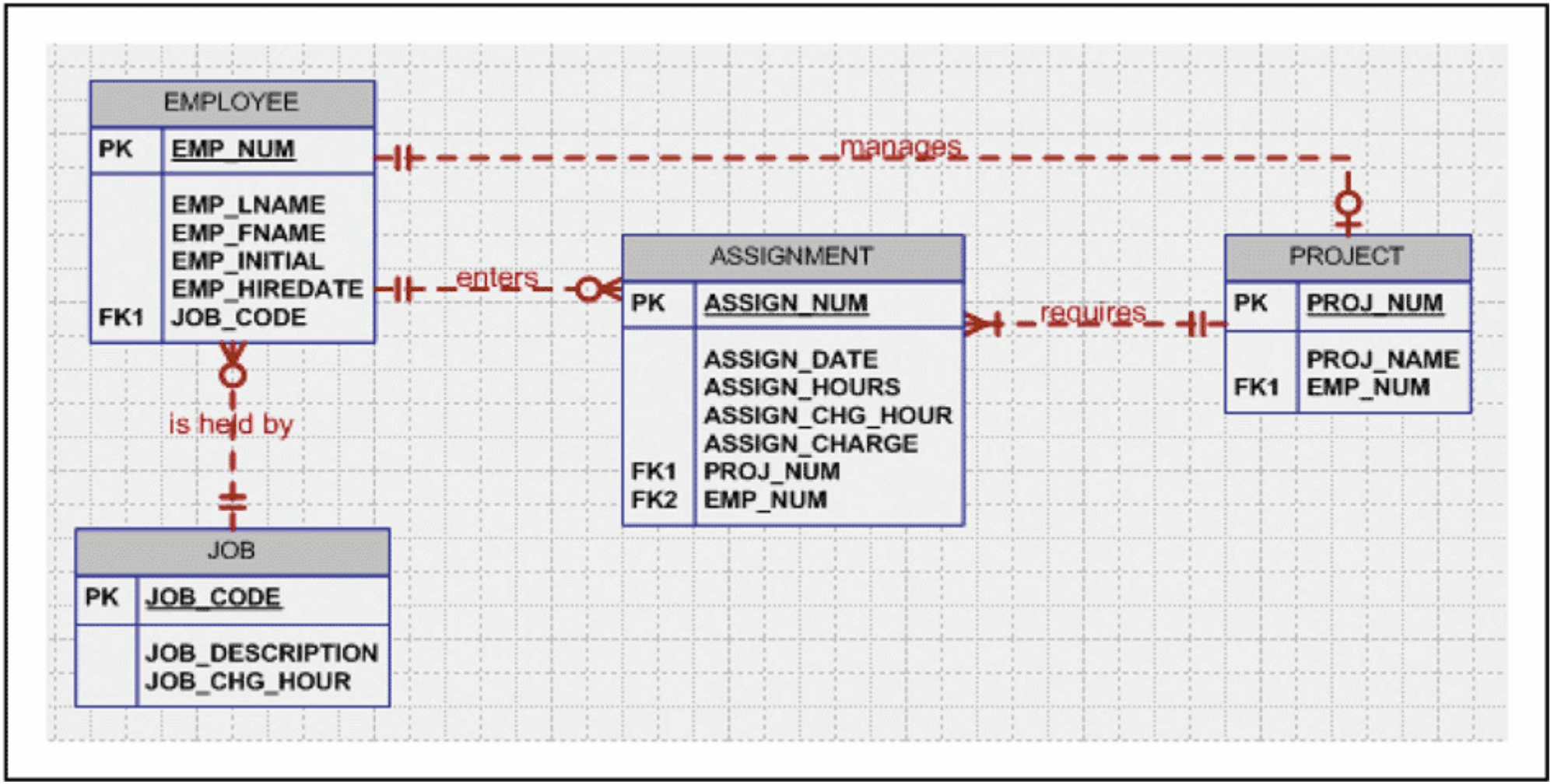
- To properly incorporate PROJECT and EMPLOYEE relationship, ASSIGNMENT associative entity must be defined.
- It keeps track of the assignment of employees to projects.

Final Schema:

- PROJECT (PROJ_NUM, PROJ_NAME, EMP_NUM)
- EMPLOYEE (EMP_NUM, EMP_LNAME, EMP_FNAME, EMP_INITIAL, JOB_CODE)
- JOB (JOB_CODE, JOB_DESCRIPTION, JOP_CHG_HOUR)
- ASSIGNMENT (ASSIGN_NUM, ASSIGN_DATE, PROJ_NUM, EMP_NUM, ASSIGN_HOURS, ASSIGN_CHG_HOURS, ASSIGN_CHARGE)

The Final (Implementable) ERD for a Contracting Company

FIGURE 5.13 THE FINAL (IMPLEMENTABLE) ERD FOR A CONTRACTING COMPANY



The Implemented Database for the Contracting Company

Table name: EMPLOYEE — **Database name: Ch05_ConstructCo**

EMP_NUM	EMP_LNAME	EMP_FNAME	EMP_INITIAL	EMP_HIREDATE	JOB_CODE
101	News	John	G	08-Nov-00	502
102	Senior	David	H	12-Jul-89	501
103	Arbough	June	E	01-Dec-97	503
104	Ramoras	Anne	K	15-Nov-88	501
105	Johnson	Alice	K	01-Feb-94	502
106	Smithfield	William		22-Jun-05	500
107	Alonzo	Maria	D	10-Oct-94	500
108	Washington	Ralph	B	22-Aug-89	501
109	Smith	Larry	W	18-Jul-99	501
110	Olenko	Gerald	A	11-Dec-96	505
111	Wabash	Geoff	B	04-Apr-89	506
112	Smithson	Darlene	M	23-Oct-95	507
113	Joenbrood	Delbert	K	15-Nov-94	508
114	Jones	Annelise		20-Aug-91	508
115	Bawangi	Travis	B	25-Jan-90	501
116	Pratt	Gerald	L	05-Mar-95	510
117	Williamson	Angie	H	19-Jun-94	509
118	Frommer	James	J	04-Jan-96	510

Table name: JOB

JOB_CODE	JOB_DESCRIPTION	JOB_CHG_HOUR
500	Programmer	35.75
501	Systems Analyst	96.75
502	Database Designer	105.00
503	Electrical Engineer	84.50
504	Mechanical Engineer	67.90
505	Civil Engineer	55.78
506	Clerical Support	26.87
507	DSS Analyst	45.95
508	Applications Designer	48.10
509	Bio Technician	34.55
510	General Support	18.36

Table name: PROJECT

PROJ_NUM	PROJ_NAME	EMP_NUM
15	Evergreen	105
18	Amber Wave	104
22	Rolling Tide	113
25	Starlight	101

Table name: ASSIGNMENT

ASSIGN_NUM	ASSIGN_DATE	PROJ_NUM	EMP_NUM	ASSIGN_HOURS	ASSIGN_CHG_HOUR	ASSIGN_CHARGE
1001	04-Mar-08	15	103	2.6	84.50	219.70
1002	04-Mar-08	18	118	1.4	18.36	25.70
1003	05-Mar-08	15	101	3.6	105.00	378.00
1004	05-Mar-08	22	113	2.5	48.10	120.25
1005	05-Mar-08	15	103	1.9	84.50	160.55
1006	05-Mar-08	25	115	4.2	96.75	406.35
1007	05-Mar-08	22	105	5.2	105.00	546.00
1008	05-Mar-08	25	101	1.7	105.00	178.50
1009	05-Mar-08	15	105	2.0	105.00	210.00
1010	06-Mar-08	15	102	3.8	96.75	367.65
1011	06-Mar-08	22	104	2.6	96.75	251.55
1012	06-Mar-08	15	101	2.3	105.00	241.50
1013	06-Mar-08	25	114	1.8	48.10	86.58
1014	06-Mar-08	22	111	4.0	26.87	107.48
1015	06-Mar-08	25	114	3.4	48.10	163.54
1016	06-Mar-08	18	112	1.2	45.95	55.14
1017	06-Mar-08	18	118	2.0	18.36	36.72
1018	06-Mar-08	18	104	2.6	96.75	251.55
1019	06-Mar-08	15	103	3.0	84.50	253.50
1020	07-Mar-08	22	105	2.7	105.00	283.50
1021	08-Mar-08	25	108	4.2	96.75	406.35
1022	07-Mar-08	25	114	5.8	48.10	278.98
1023	07-Mar-08	22	108	2.4	35.75	85.80

Summary

- Normalization is a table design technique aimed at minimizing data redundancies
- First three normal forms (1NF, 2NF, and 3NF) are most commonly encountered
- Normalization is an important part – but only a part – of the design process
- Continuing the iterative ER process until all entities and their attributes are defined and all equivalent tables are in 3NF
- A table in 3NF may contain multi-valued dependencies that produce either numerous null values or redundant data
- It may be necessary to convert a 3NF table to the fourth normal (4NF) by
  - Splitting such a table to remove multi-valued dependencies