Normalization of Database Tables

In this chapter, you will learn:

- What normalization is and what role it plays in the database design process
- About the normal forms 1NF, 2NF, 3NF, BCNF, and 4NF
- How normal forms can be transformed from lower normal forms to higher normal forms
- How normalization and ER modeling are used concurrently to produce a good database design
- How some situations require denormalization to generate information efficiently

Database Tables and Normalization

- Normalization
  - Process for evaluating and correcting table structures to minimize data redundancies
    - Reduces data anomalies
  - Works through a series of stages called normal forms:
    - First normal form (1NF)
    - Second normal form (2NF)
    - Third normal form (3NF)

Database Tables and Normalization (continued)

- Normalization (continued)
  - 2NF is better than 1NF; 3NF is better than 2NF
  - For most business database design purposes, 3NF is as high as we need to go in normalization process
  - Highest level of normalization is not always most desirable

The Need for Normalization

- Example: Company that manages building projects
  - Charges its clients by billing hours spent on each contract
  - Hourly billing rate is dependent on employee’s position
  - Periodically, report is generated that contains information displayed in Table 5.1

The Need for Normalization (continued)

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	$ 85.50	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
				Subtotal			**$10,549.70**
18	Amber Wave	114	Annelise Jones	Applications Designer	$ 48.10	25.6	$ 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	$ 3,135.70
		112	Darlene M. Smithson	DSS Analyst	$ 45.95	45.0	$ 2,021.80
				Subtotal			**$ 7,172.47**
22	Rolling Tide	105	Alice K. Johnson	Database Designer	$105.00	65.7	$ 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
				Subtotal			**$13,660.10**
25	Starflight	107	Maria D. Alonzo	Programmer	$ 35.75	25.6	$ 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
				Subtotal			**$17,559.82**
				Total			**$48,942.09**

Note: * indicates project leader.

The Need for Normalization (continued)

FIGURE 5.1 Tabular representation of the report format

Table name: RPT_FORMAT   Database name: Ch05_ConstructCo

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

The Need for Normalization (continued)

- Structure of data set in Figure 5.1 does not handle data very well
- The table structure appears to work; report generated with ease
- Unfortunately, report may yield different results depending on what data anomaly has occurred

The Normalization Process

- Each table represents a single subject
- No data item will be unnecessarily stored in more than one table
- All attributes in a table are dependent on the primary key

The Normalization Process (continued)

TABLE 5.2 Normal Forms

NORMAL FORM	CHARACTERISTIC	SECTION
First normal form (1NF)	Table format; no repeating groups and PK identified	5.3.1
Second normal form (2NF)	1NF and no partial dependencies	5.3.2
Third normal form (3NF)	2NF and no transitive dependencies	5.3.3
Boyce-Codd normal form (BCNF)	Every determinant is a candidate key (special case of 3NF)	5.6.1
Fourth normal form (4NF)	3NF and no independent multivalued dependencies	5.6.2

Conversion to First Normal Form

- Repeating group
  - Derives its name from the fact that a group of multiple entries of same type can exist for any single key attribute occurrence
- Relational table must not contain repeating groups
- Normalizing table structure will reduce data redundancies
- Normalization is three-step procedure

Conversion to First Normal Form (continued)

- Step 1: Eliminate the Repeating Groups
  - Present data in tabular format, where each cell has single value and there are no repeating groups
  - Eliminate repeating groups, eliminate nulls by making sure that each repeating group attribute contains an appropriate data value

Conversion to First Normal Form (continued)

FIGURE 5.2 A table in first normal form

Table name: DATA_ORG_1NF | Database name: Ch05_ConstructCo

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	12.6
15	Evergreen	102	David H. Senior	Systems Analyst	$96.75	23.8
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	32.4
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. Ramoras	Systems Analyst	$96.75	48.4
22	Rolling Tide	113	Delbert K. Joenbrood *	Applications Designer	$48.10	23.6
22	Rolling Tide	111	Geoff B. Wabash	Clerical Support	$26.87	22.0
22	Rolling Tide	106	William Smithfield	Programmer	$35.75	12.8
25	Starflight	107	Maria D. Alonzo	Programmer	$35.75	24.6
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	23.6
25	Starflight	118	James J. Frommer	General Support	$18.36	30.5
25	Starflight	112	Darlene M. Smithson	DSS Analyst	$45.95	41.4

Conversion to First Normal Form (continued)

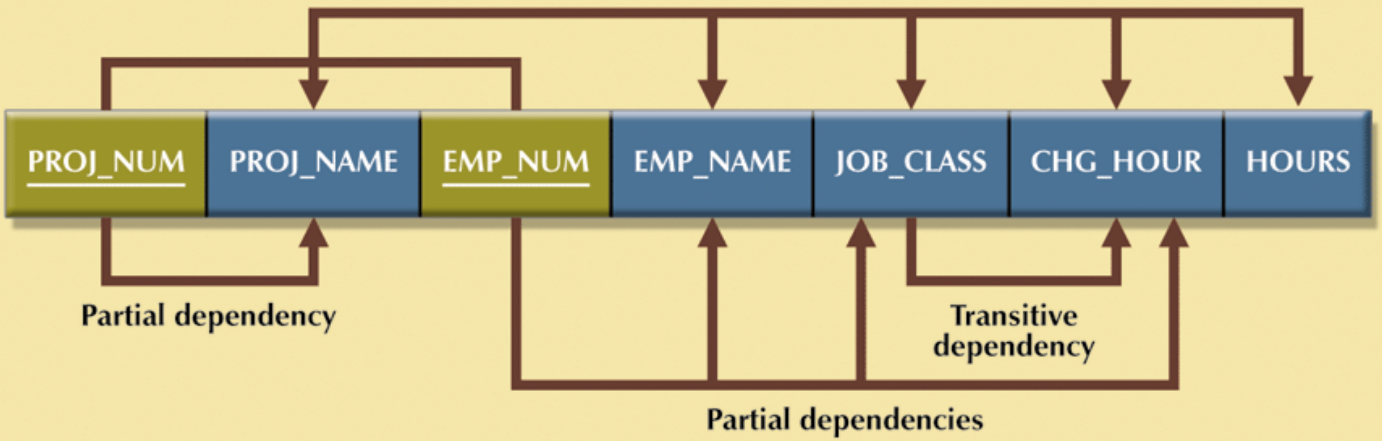
- Step 2: Identify the Primary Key
  - Primary key must uniquely identify attribute value
  - New key must be composed

Conversion to First Normal Form (continued)

- Step 3: Identify All Dependencies
  - Dependencies can be depicted with help of a diagram
  - Dependency diagram:
    - Depicts all dependencies found within given table structure
    - Helpful in getting bird's-eye view of all relationships among table's attributes
    - Makes it less likely that will overlook an important dependency

Conversion to First Normal Form (continued)

FIGURE 5.3 First normal form (1NF) dependency diagram



1NF (PROJ_NUM, EMP_NUM, PROJ_NAME, EMP_NAME, JOB_CLASS, CHG_HOURS, HOURS)

PARTIAL DEPENDENCIES:
(PROJ_NUM → PROJ_NAME)
(EMP_NUM → EMP_NAME, JOB_CLASS, CHG_HOUR)

TRANSITIVE DEPENDENCY:
(JOB CLASS → CHG_HOUR)

Conversion to First Normal Form (continued)

- First normal form describes tabular format in which:
  - All key attributes are defined
  - There are no repeating groups in the table
  - All attributes are dependent on primary key
- All relational tables satisfy 1NF requirements
- Some tables contain partial dependencies
  - Dependencies based on only part of the primary key
  - Sometimes used for performance reasons, but should be used with caution
  - Still subject to data redundancies

Conversion to Second Normal Form

- Relational database design can be improved by converting the database into second normal form (2NF)
- Two steps

Conversion to Second Normal Form (continued)

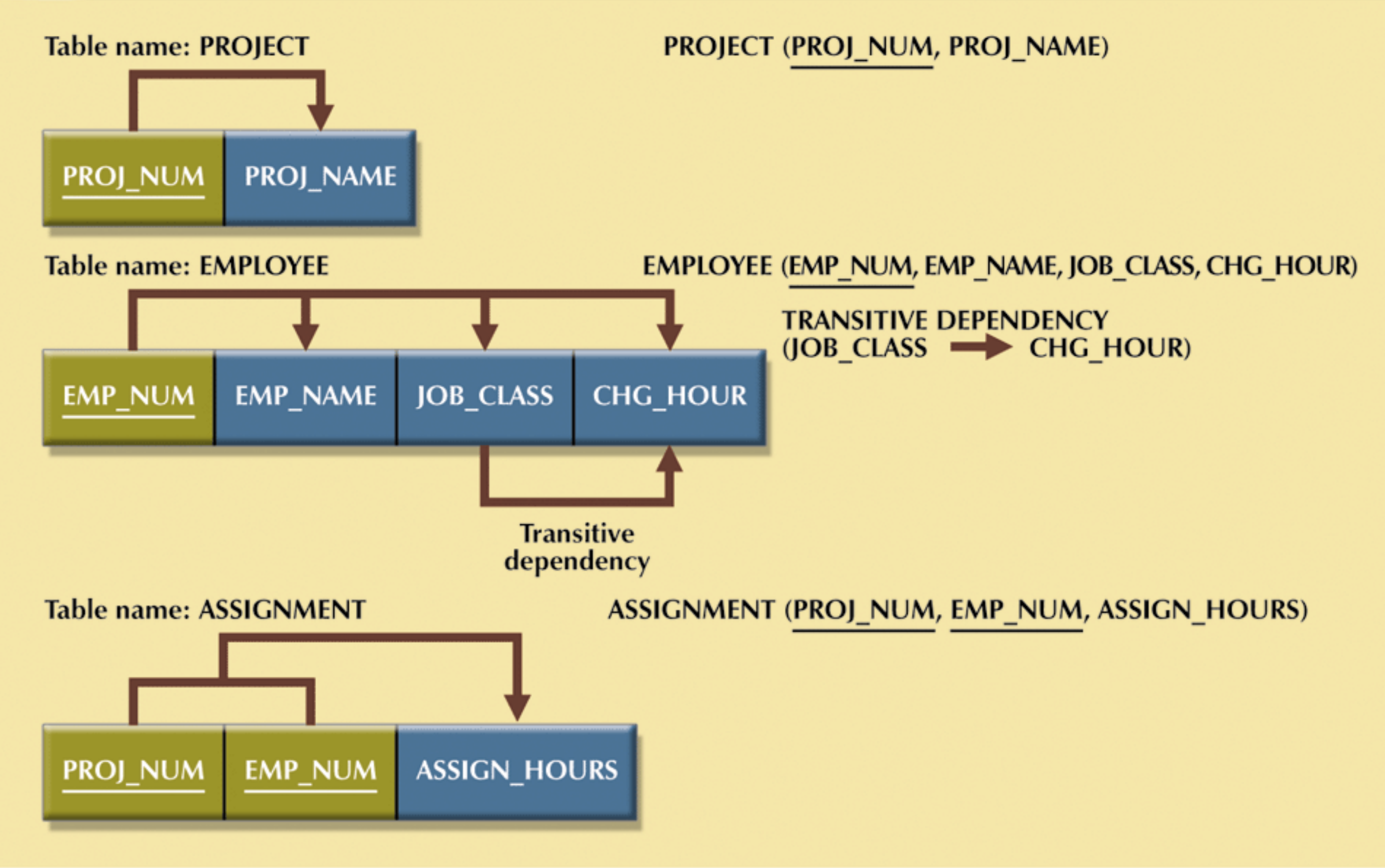
- Step 1: Write Each Key Component on a Separate Line
  - Write each key component on separate line, then write original (composite) key on last line
  - Each component will become key in new table

Conversion to Second Normal Form (continued)

- Step 2: Assign Corresponding Dependent Attributes
  - Determine those attributes that are dependent on other attributes
  - At this point, most anomalies have been eliminated

Conversion to Second Normal Form (continued)

FIGURE 5.4 Second normal form (2NF) conversion results



Conversion to Second Normal Form (continued)

- Table is in second normal form (2NF) when:
  - It is in 1NF and
  - It includes no partial dependencies:
    - No attribute is dependent on only portion of primary key

Conversion to Third Normal Form

- Data anomalies created are easily eliminated by completing three steps
- Step 1: Identify Each New Determinant
  - For every transitive dependency, write its determinant as PK for new table
    - Determinant
      - Any attribute whose value determines other values within a row

Conversion to Third Normal Form (continued)

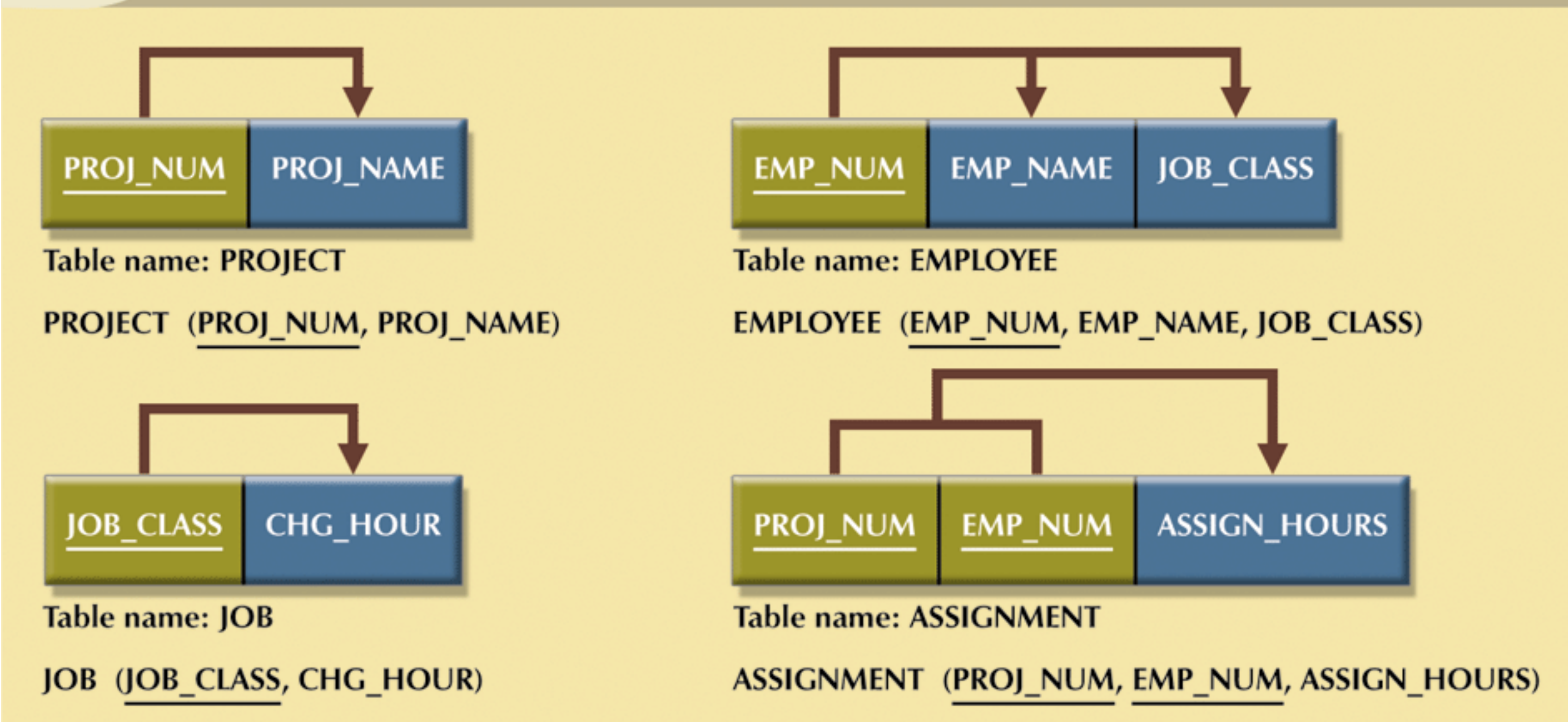
- Step 2: Identify the Dependent Attributes
  - Identify attributes dependent on each determinant identified in Step 1 and identify dependency
  - Name table to reflect its contents and function

Conversion to Third Normal Form (continued)

- Step 3: Remove the Dependent Attributes from Transitive Dependencies
  - Eliminate all dependent attributes in transitive relationship(s) from each of the tables that have such a transitive relationship
  - Draw new dependency diagram to show all tables defined in Steps 1–3
  - Check new tables as well as tables modified in Step 3 to make sure that each table has determinant and that no table contains inappropriate dependencies

Conversion to Third Normal Form (continued)

FIGURE 5.5 Third normal form (3NF) conversion results



Conversion to Third Normal Form (continued)

- A table is in third normal form (3NF) when both of the following are true:
  - It is in 2NF
  - It contains no transitive dependencies

Improving the Design

- Table structures are cleaned up to eliminate troublesome initial partial and transitive dependencies
- Normalization cannot, by itself, be relied on to make good designs
- It is valuable because its use helps eliminate data redundancies

Improving the Design (continued)

- Issues to address in order to produce a good normalized set of tables:
  - Evaluate PK Assignments
  - Evaluate Naming Conventions
  - Refine Attribute Atomicity
  - Identify New Attributes
  - Identify New Relationships
  - Refine Primary Keys as Required for Data Granularity
  - Maintain Historical Accuracy
  - Evaluate Using Derived Attributes

Improving the Design (continued)

FIGURE 5.6 The completed database

Table name: PROJECT

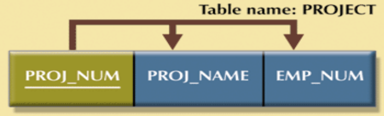


Table name: JOB

Database name: Ch05_ConstructCo

Table name: PROJECT

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

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: ASSIGNMENT

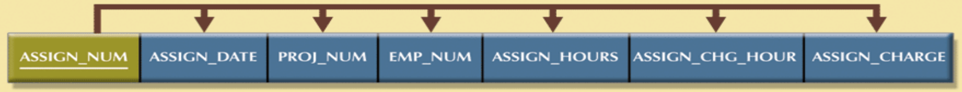


Table name: ASSIGNMENT

ASSIGN_NUM	ASSIGN_DATE	PROJ_NUM	EMP_NUM	ASSIGN_HOURS	ASSIGN_CHG_HOUR	ASSIGN_CHARGE
1001	04-Mar-06	15	103	2.6	$84.50	$219.70
1002	04-Mar-06	18	118	1.4	$18.36	$25.70
1003	05-Mar-06	15	101	3.6	$105.00	$378.00
1004	05-Mar-06	22	113	2.5	$48.10	$120.25
1005	05-Mar-06	15	103	1.9	$84.50	$160.55
1006	05-Mar-06	25	115	4.2	$96.75	$406.35
1007	05-Mar-06	22	105	5.2	$105.00	$546.00
1008	05-Mar-06	25	101	1.7	$105.00	$178.50
1009	05-Mar-06	15	105	2.0	$105.00	$210.00
1010	06-Mar-06	15	102	3.8	$96.75	$367.65
1011	06-Mar-06	22	104	2.6	$96.75	$251.55
1012	06-Mar-06	15	101	2.3	$105.00	$241.50
1013	06-Mar-06	25	114	1.8	$48.10	$86.58
1014	06-Mar-06	22	111	4.0	$26.87	$107.48
1015	06-Mar-06	25	114	3.4	$48.10	$163.54
1016	06-Mar-06	18	112	1.2	$45.95	$55.14
1017	06-Mar-06	18	118	2.0	$18.36	$36.72
1018	06-Mar-06	18	104	2.6	$96.75	$251.55
1019	06-Mar-06	15	103	3.0	$84.50	$253.50
1020	07-Mar-06	22	105	2.7	$105.00	$283.50
1021	08-Mar-06	25	108	4.2	$96.75	$406.35
1022	07-Mar-06	25	114	5.8	$48.10	$278.98
1023	07-Mar-06	22	106	2.4	$35.75	$85.80

Improving the Design (continued)

FIGURE 5.6 The completed database (continued)

Table name: EMPLOYEE

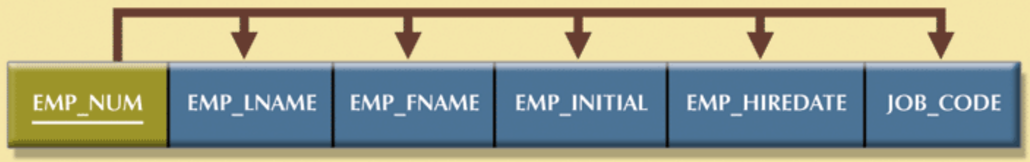


Table name: EMPLOYEE

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-06	510

Surrogate Key Considerations

- When primary key is considered to be unsuitable, designers use surrogate keys
- Data entries in Table 5.3 are inappropriate because they duplicate existing records
  - Yet there has been no violation of either entity integrity or referential integrity

Surrogate Key Considerations (continued)

TABLE 5.3 Duplicate Entries in the Job Table

JOB_CODE	JOB_DESCRIPTION	JOB_CHG_HOUR
511	Programmer	$35.75
512	Programmer	$35.75

The Boyce-Codd Normal Form (BCNF)

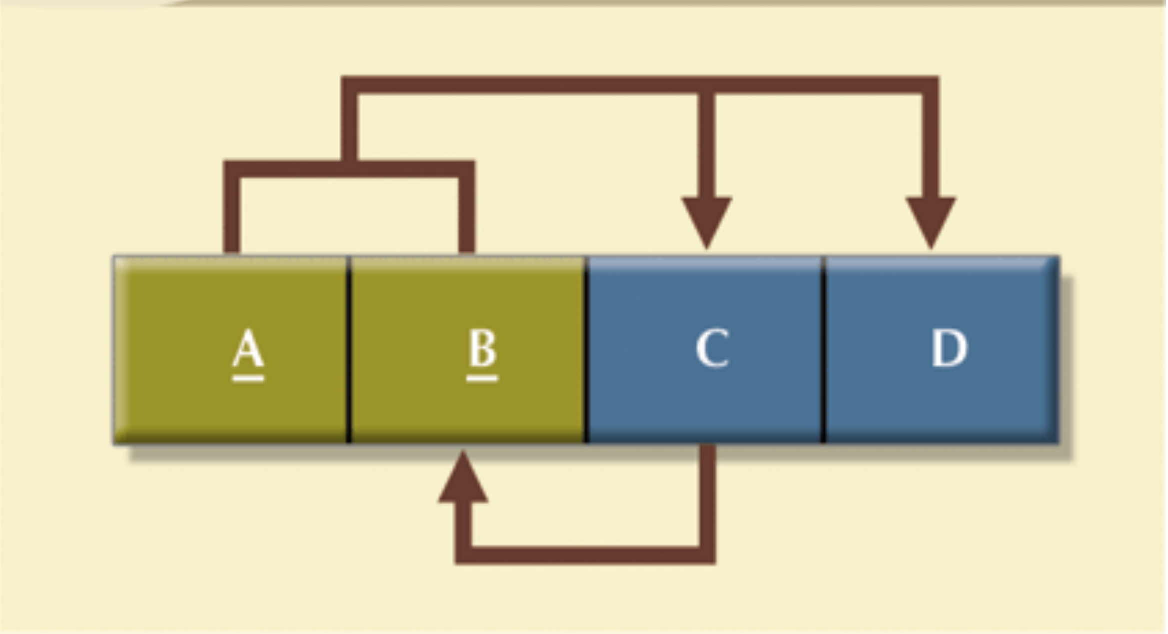
- Every determinant in table is a candidate key
  - Has same characteristics as primary key, but for some reason, not chosen to be primary key
- When table contains only one candidate key, the 3NF and the BCNF are equivalent
- BCNF can be violated only when table contains more than one candidate key

The Boyce-Codd Normal Form (BCNF) (continued)

- Most designers consider the BCNF as special case of 3NF
- Table is in 3NF when it is in 2NF and there are no transitive dependencies
- Table can be in 3NF and fails to meet BCNF
  - No partial dependencies, nor does it contain transitive dependencies
  - A nonkey attribute is the determinant of a key attribute

The Boyce-Codd Normal Form (BCNF) (continued)

FIGURE 5.7 A table that is in 3NF but not in BCNF

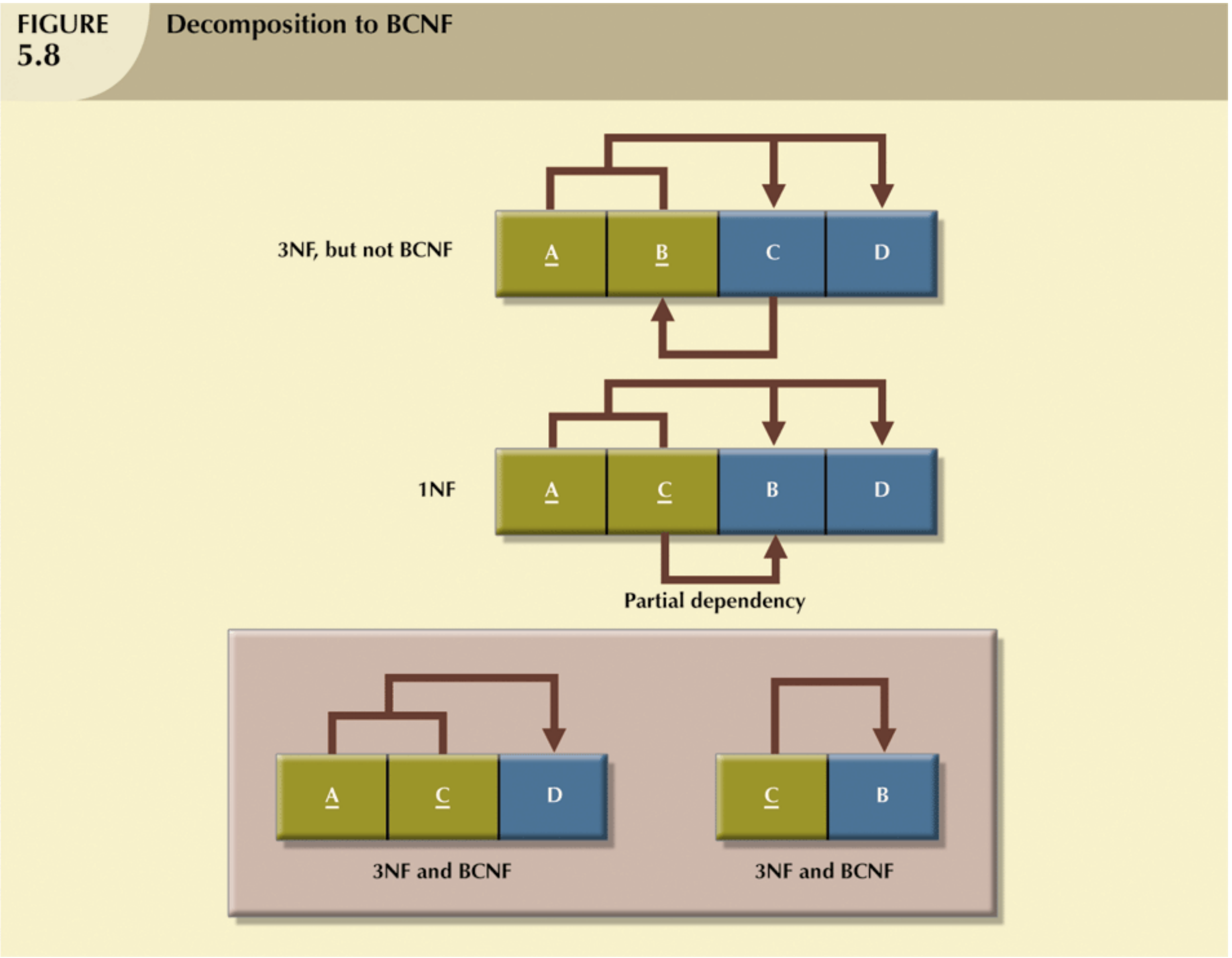


The Boyce-Codd Normal Form (BCNF) (continued)

TABLE 5.4 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

The Boyce-Codd Normal Form (BCNF) (continued)



Fourth Normal Form (4NF)

- Table is in fourth normal form (4NF) when both of the following are true:
  - It is in 3NF
  - Has no multiple sets of multivalued dependencies
- 4NF is largely academic if tables conform to following two rules:
  - All attributes must be dependent on primary key, but independent of each other
  - No row contains two or more multivalued facts about an entity

Fourth Normal Form (4NF) (continued)

FIGURE 5.10 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 (4NF) (continued)

FIGURE 5.11 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

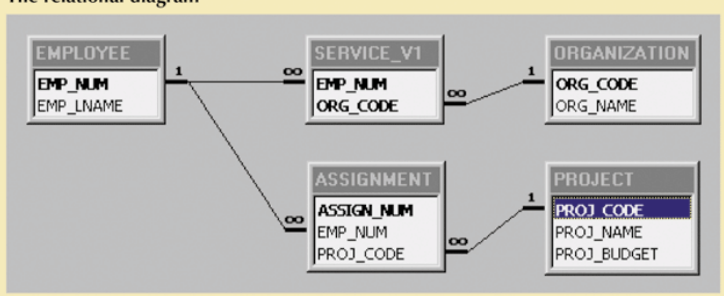
Table name: ASSIGNMENT

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 diagram



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

Normalization and Database Design (continued)

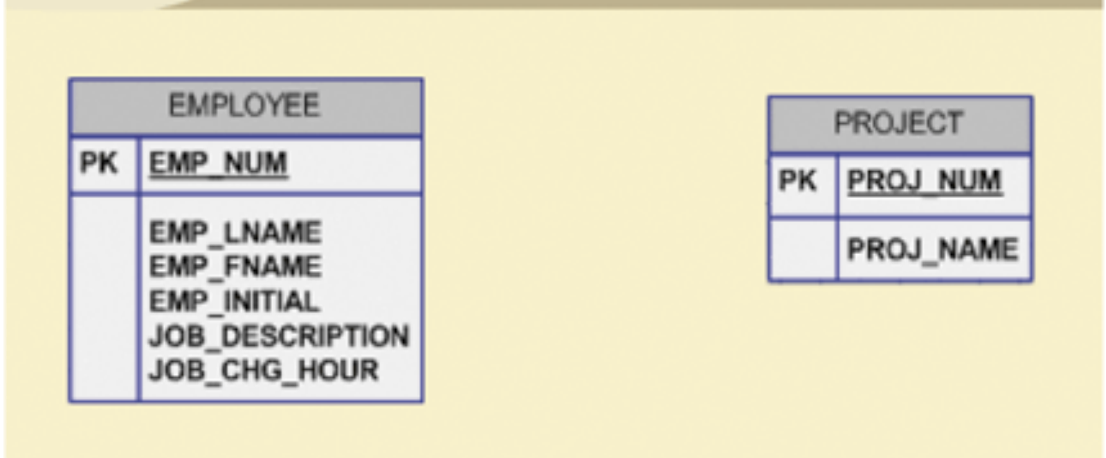
- ER diagram
  - Provides big picture, or macro view, of an organization’s data requirements and operations
  - Created through an iterative process
    - Identifying relevant entities, their attributes and their relationship
    - Use results to identify additional entities and attributes

Normalization and Database Design (continued)

- Normalization procedures
  - Focus on characteristics of specific entities
  - Represents micro view of entities within ER diagram
- Difficult to separate normalization process from ER modeling process
- Two techniques should be used concurrently

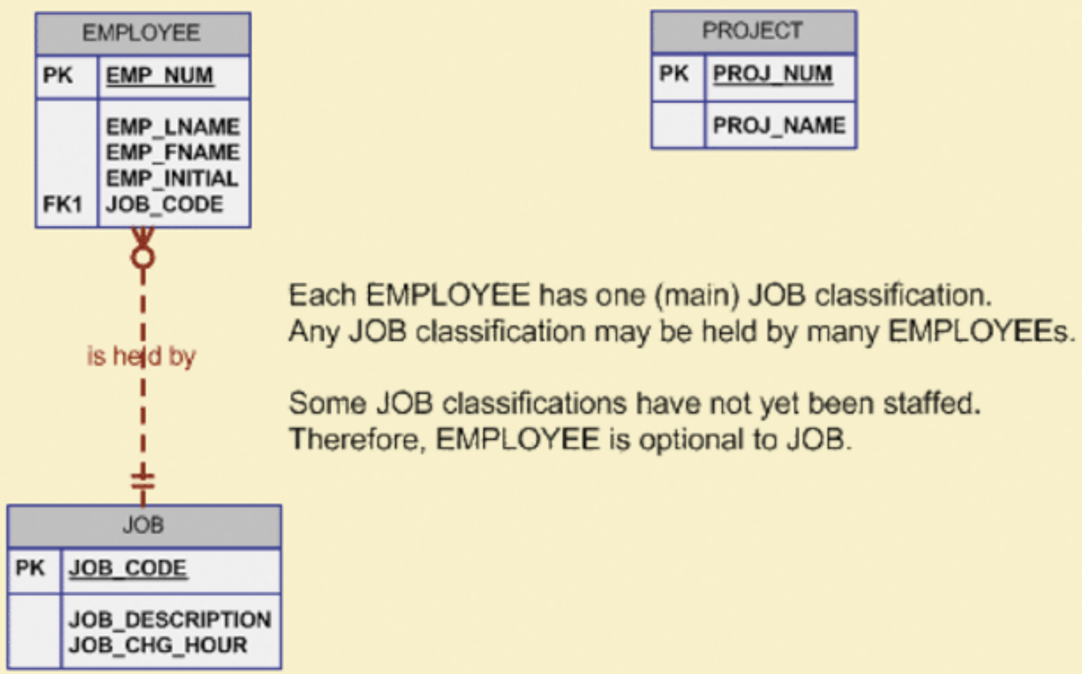
Normalization and Database Design (continued)

FIGURE 5.12 Initial contracting company ERD



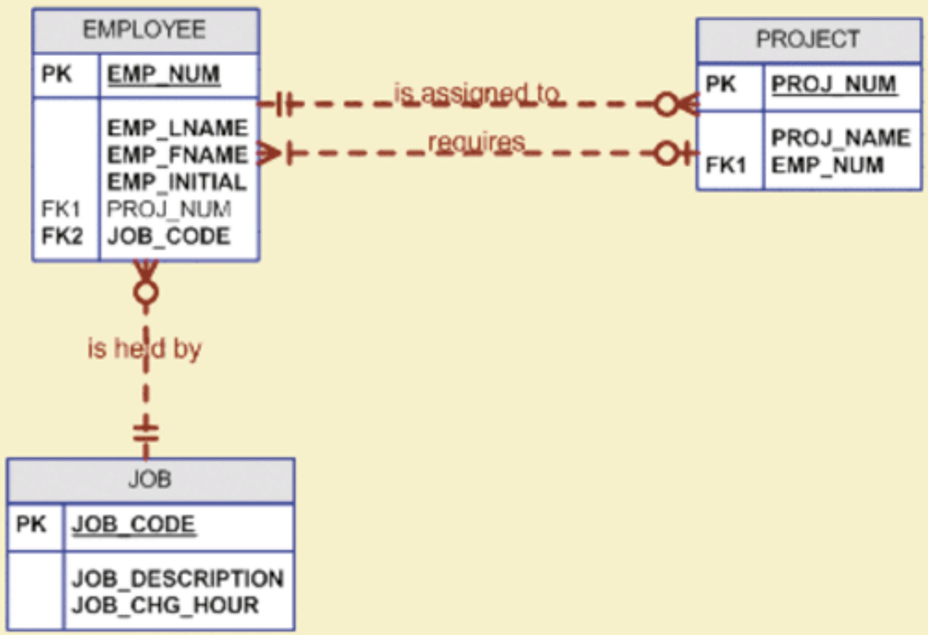
Normalization and Database Design (continued)

FIGURE 5.13 Modified contracting company ERD

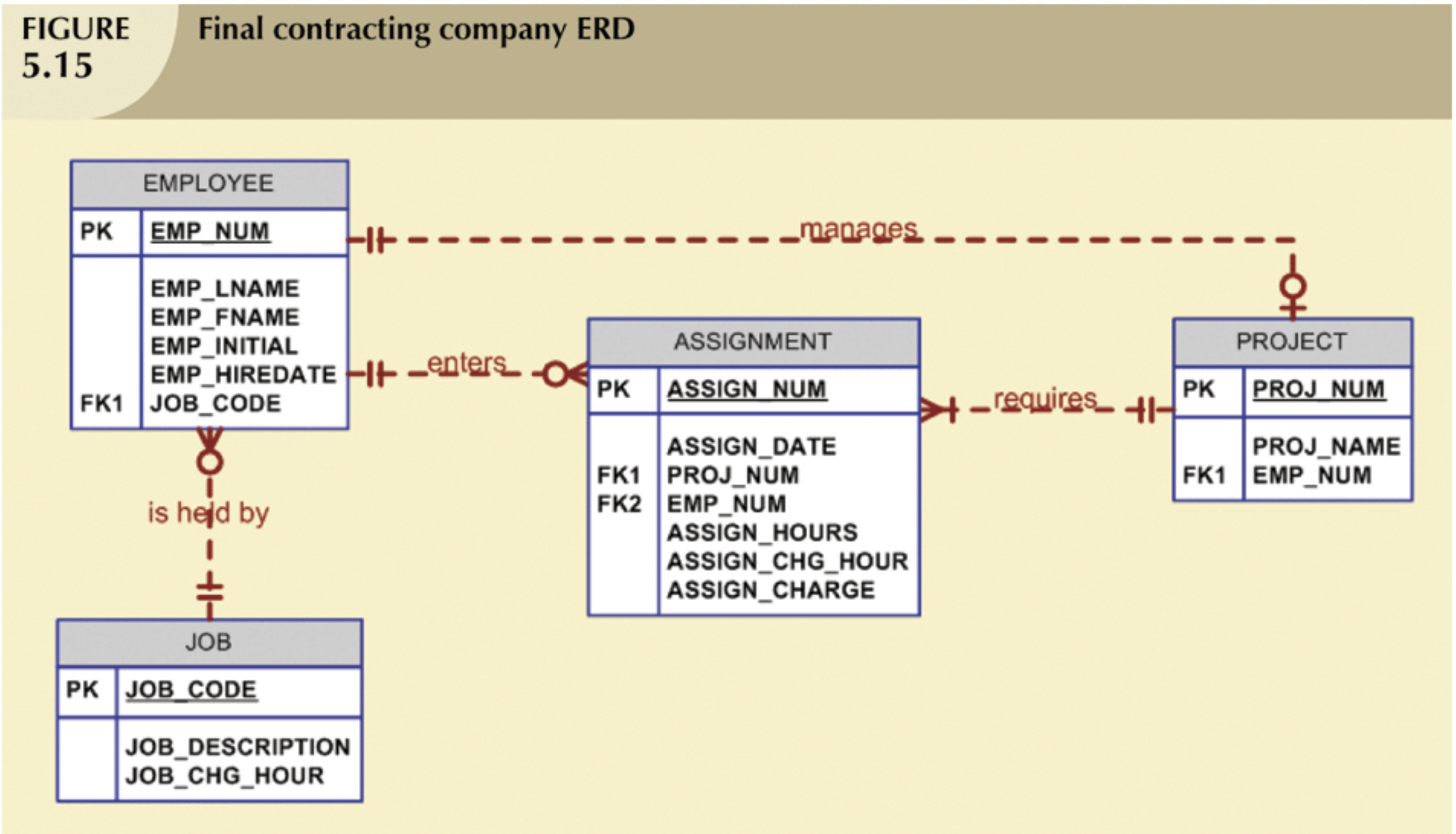


Normalization and Database Design (continued)

FIGURE 5.14 **Incorrect M:N relationship representation**



Normalization and Database Design (continued)



Normalization and Database Design (continued)

FIGURE 5.16 The implemented database

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-06	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	Starflight	101

Table name: ASSIGNMENT

ASSIGN_NUM	ASSIGN_DATE	PROJ_NUM	EMP_NUM	ASSIGN_HOURS	ASSIGN_CHG_HOUR	ASSIGN_CHARGE
1001	04-Mar-06	15	103	2.6	$84.50	$219.70
1002	04-Mar-06	18	118	1.4	$18.36	$25.70
1003	05-Mar-06	15	101	3.6	$105.00	$378.00
1004	05-Mar-06	22	113	2.5	$48.10	$120.25
1005	05-Mar-06	15	103	1.9	$84.50	$160.55
1006	05-Mar-06	25	115	4.2	$96.75	$406.35
1007	05-Mar-06	22	105	5.2	$105.00	$546.00
1008	05-Mar-06	25	101	1.7	$105.00	$178.50
1009	05-Mar-06	15	105	2.0	$105.00	$210.00
1010	06-Mar-06	15	102	3.8	$96.75	$367.65
1011	06-Mar-06	22	104	2.6	$96.75	$251.55
1012	06-Mar-06	15	101	2.3	$105.00	$241.50
1013	06-Mar-06	25	114	1.8	$48.10	$86.58
1014	06-Mar-06	22	111	4.0	$26.87	$107.48
1015	06-Mar-06	25	114	3.4	$48.10	$163.54
1016	06-Mar-06	18	112	1.2	$45.95	$55.14
1017	06-Mar-06	18	118	2.0	$18.36	$36.72
1018	06-Mar-06	18	104	2.6	$96.75	$251.55
1019	06-Mar-06	15	103	3.0	$84.50	$253.50
1020	07-Mar-06	22	105	2.7	$105.00	$283.50
1021	08-Mar-06	25	108	4.2	$96.75	$406.35
1022	07-Mar-06	25	114	5.8	$48.10	$278.98
1023	07-Mar-06	22	106	2.4	$35.75	$85.80

Denormalization

- Creation of normalized relations is important database design goal
- Processing requirements should also be a goal
- If tables decomposed to conform to normalization requirements:
  - Number of database tables expands

Denormalization (continued)

- Joining the larger number of tables takes additional input/output (I/O) operations and processing logic, thereby reducing system speed
- Conflicts between design efficiency, information requirements, and processing speed are often resolved through compromises that may include denormalization

Denormalization (continued)

- Unnormalized tables in production database tend to suffer from these defects:
  - Data updates are less efficient because programs that read and update tables must deal with larger tables
  - Indexing is more cumbersome
  - Unnormalized tables yield no simple strategies for creating virtual tables known as views

Denormalization (continued)

- Use denormalization cautiously
- Understand why—under some circumstances—unnormalized tables are better choice

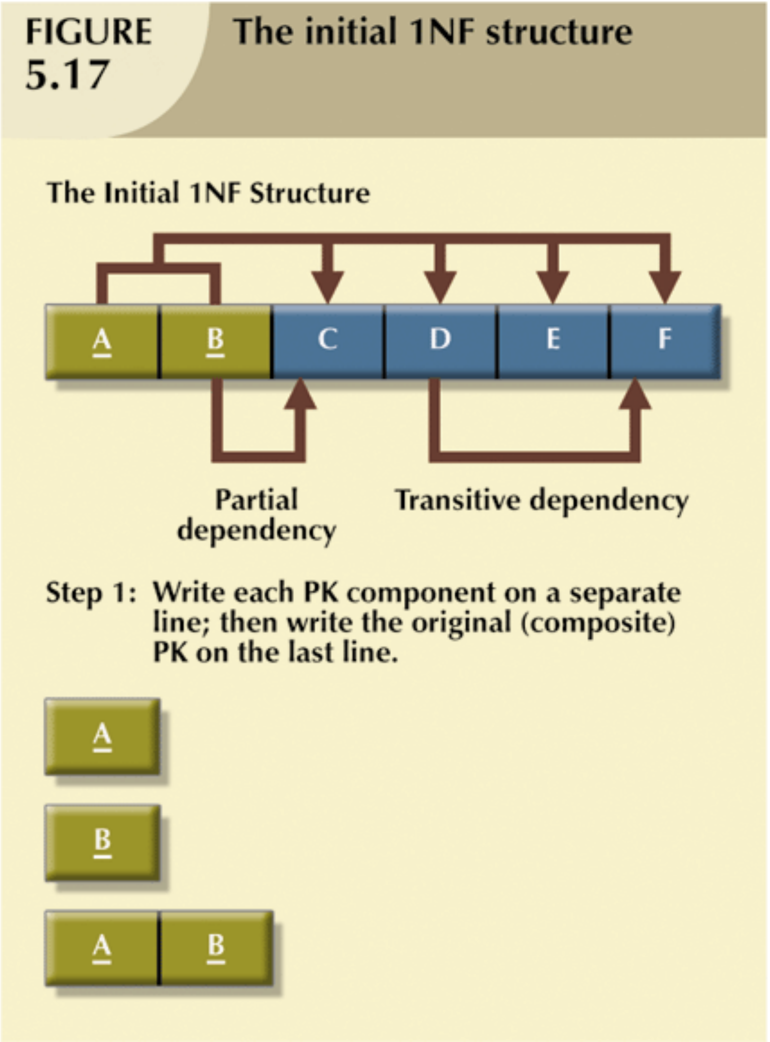
Summary

- Normalization is technique used to design tables in which data redundancies are minimized
- First three normal forms (1NF, 2NF, and 3NF) are most commonly encountered
- Table is in 1NF when all key attributes are defined and when all remaining attributes are dependent on primary key

Summary (continued)

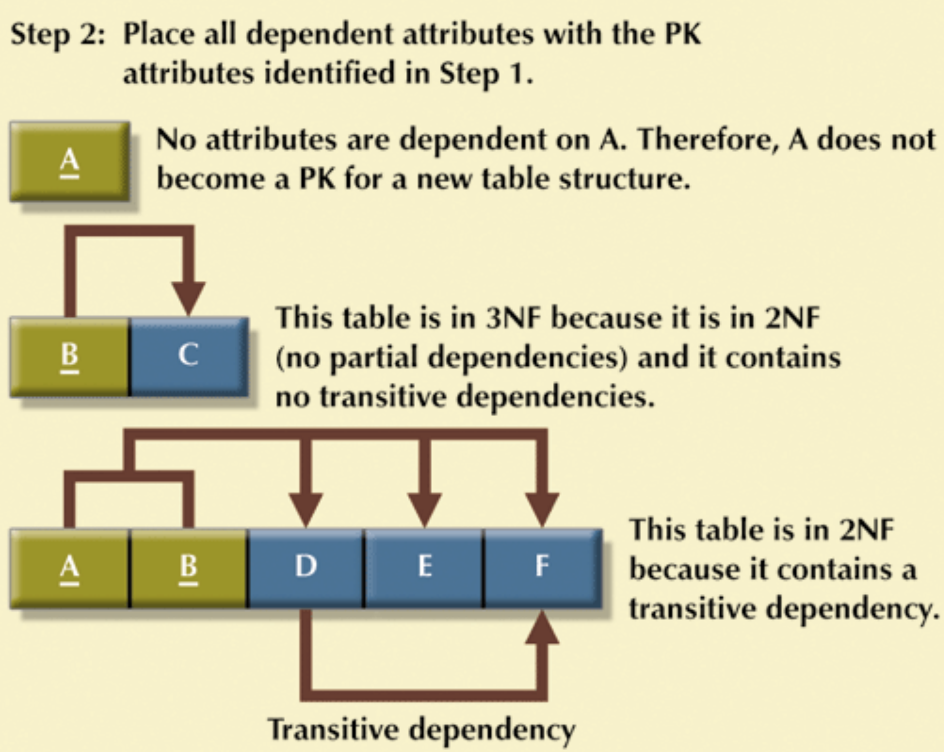
- Table is in 2NF when it is in 1NF and contains no partial dependencies
- Table is in 3NF when it is in 2NF and contains no transitive dependencies
- Table that is not in 3NF may be split into new tables until all of the tables meet 3NF requirements
- Normalization is important part—but only part—of design process

Summary (continued)



Summary (continued)

FIGURE 5.18 Identifying possible PK attributes

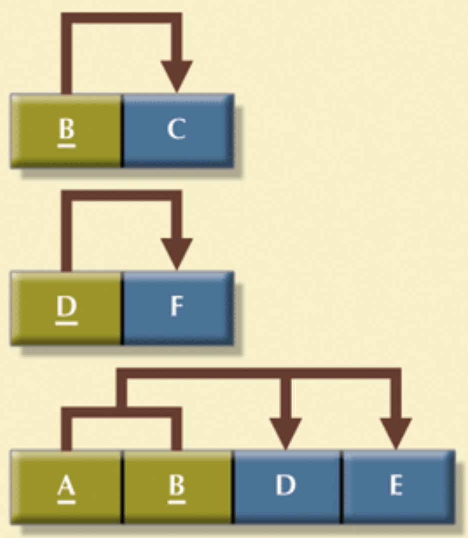


Summary (continued)

FIGURE 5.19 Table structures based on the selected PKs

Step 3: Remove all transitive dependencies identified in Step 2 and retain all 3NF structures.

All tables are in 3NF because they are in 2NF (no partial dependencies) and they do not contain transitive dependencies.



Attribute D is retained in this table structure to serve as the FK to the second table.

Summary (continued)

- Table in 3NF may contain multivalued dependencies that produce either numerous null values or redundant data
- It may be necessary to convert 3NF table to fourth normal form (4NF) by
  - Splitting table to remove multivalued dependencies
- Tables are sometimes denormalized to yield less I/O which increases processing speed