Database Principles: Fundamentals of Design, Implementation, and Management

Tenth Edition

Chapter 9

Normalizing Database Designs

Objectives

- In this chapter, students 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
  - That normalization and ER modeling are used concurrently to produce a good database design
  - That 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
  - Series of stages called normal forms:
    - First normal form (1NF)
    - Second normal form (2NF)
    - Third normal form (3NF)

Database Tables and Normalization (cont'd.)

- Normalization (continued)
  - 2NF is better than 1NF; 3NF is better than 2NF
  - For most business database design purposes, 3NF is as high as needed in normalization
  - Highest level of normalization is not always most desirable
- Denormalization produces a lower normal form
  - Increased performance but greater data redundancy

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 such as displayed in Table 6.1

TABLE 9.1 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
				Subtotal			**$10,549.70**
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
				Subtotal			**$ 7,171.47**
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
				Subtotal			**$13,660.10**
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
				Subtotal			**$17,559.82**
				Total			**$48,941.09**

Note: A * indicates the project leader.

The Need for Normalization (cont'd.)

- Structure of data set in Figure 6.1 does not handle data very well
- Table structure appears to work; report is generated with ease
- Report may yield different results depending on what data anomaly has occurred
- Relational database environment is suited to help designer avoid data integrity problems

The Normalization Process

- Each table represents a single subject
- No data item will be unnecessarily stored in more than one table
- All nonprime attributes in a table are dependent on the primary key
- Each table is void of insertion, update, and deletion anomalies

TABLE 9.2 Normal Forms

NORMAL FORM	CHARACTERISTIC	SECTION
First normal form (1NF)	Table format, no repeating groups, and PK identified	9.3.1
Second normal form (2NF)	1NF and no partial dependencies	9.3.2
Third normal form (3NF)	2NF and no transitive dependencies	9.3.3
Boyce-Codd normal form (BCNF)	Every determinant is a candidate key (special case of 3NF)	9.6.1
Fourth normal form (4NF)	3NF and no independent multivalued dependencies	9.6.2

The Normalization Process (cont'd.)

- Objective of normalization is to ensure that all tables are in at least 3NF
- Higher forms are not likely to be encountered in business environment
- Normalization works one relation at a time
- Progressively breaks table into new set of relations based on identified dependencies

TABLE 9.3 Functional Dependence Concepts

CONCEPT	DEFINITION
Functional dependence	The attribute *B* is fully functionally dependent on the attribute *A* if each value of *A* determines one and only one value of *B*. Example: PROJ_NUM → PROJ_NAME (read as *PROJ_NUM functionally determines PROJ_NAME*) In this case, the attribute PROJ_NUM is known as the determinant attribute, and the attribute PROJ_NAME is known as the dependent attribute.
Functional dependence (generalized definition)	Attribute *A* determines attribute *B* (that is, *B* is functionally dependent on *A*) if all of the rows in the table that agree in value for attribute *A* also agree in value for attribute *B*.
Fully functional dependence (composite key)	If attribute *B* is functionally dependent on a composite key *A* but not on any subset of that composite key, the attribute *B* is fully functionally dependent on *A*.

The Normalization Process (cont'd.)

- Partial dependency
  - Exists when there is a functional dependence in which the determinant is only part of the primary key
- Transitive dependency
  - Exists when there are functional dependencies such that X → Y, Y → Z, and X is the primary key

Conversion to First Normal Form

- Repeating group
  - 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 (cont'd.)

- Step 1: Eliminate the Repeating Groups
  - Eliminate nulls: each repeating group attribute contains an appropriate data value
- Step 2: Identify the Primary Key
  - Must uniquely identify attribute value
  - New key must be composed
- Step 3: Identify All Dependencies
  - Dependencies are depicted with a diagram

FIGURE 9.2 A table in first normal form

Table name: DATA_ORG_1NF

Database name: Ch09_ConstructCo

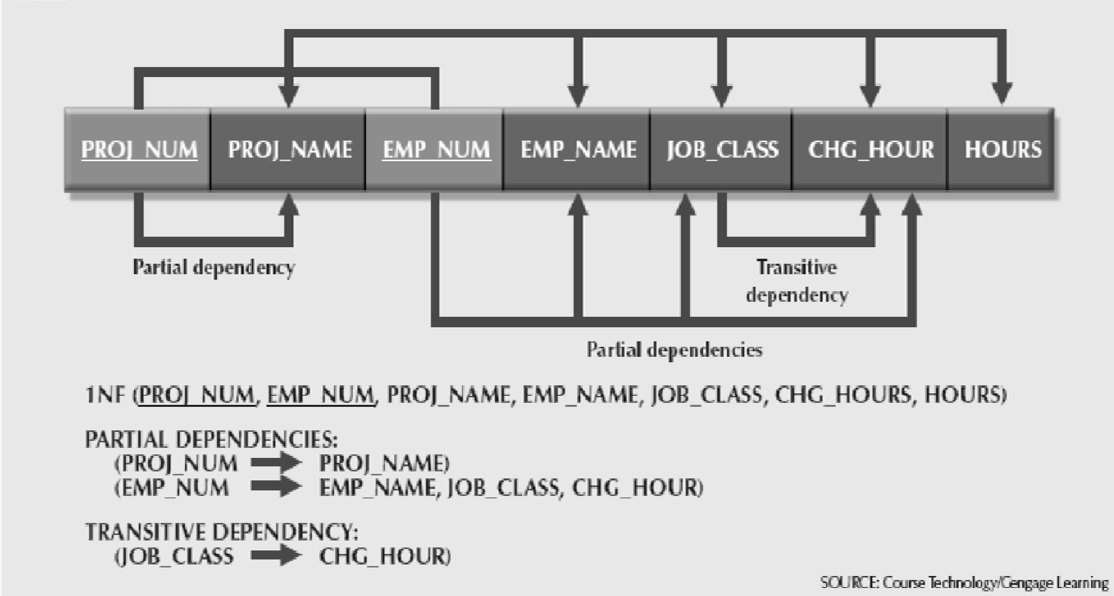
PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	June E. Arbough	Elect. Engineer	84.50	23.8
15	Evergreen	101	John G. News	Database Designer	105.00	19.4
15	Evergreen	105	Alice K. Johnson *	Database Designer	105.00	35.7
15	Evergreen	106	William Smithfield	Programmer	35.75	12.6
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	Starlight	107	Maria D. Alonzo	Programmer	35.75	24.6
25	Starlight	115	Travis B. Bawangi	Systems Analyst	96.75	45.8
25	Starlight	101	John G. News *	Database Designer	105.00	56.3
25	Starlight	114	Annelise Jones	Applications Designer	48.10	33.1
25	Starlight	108	Ralph B. Washington	Systems Analyst	96.75	23.6
25	Starlight	118	James J. Frommer	General Support	18.36	30.5
25	Starlight	112	Darlene M. Smithson	DSS Analyst	45.95	41.4

SOURCE: Course Technology/Cengage Learning

Conversion to First Normal Form (cont'd.)

- 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 you will overlook an important dependency

FIGURE 9.3 First normal form (1NF) dependency diagram



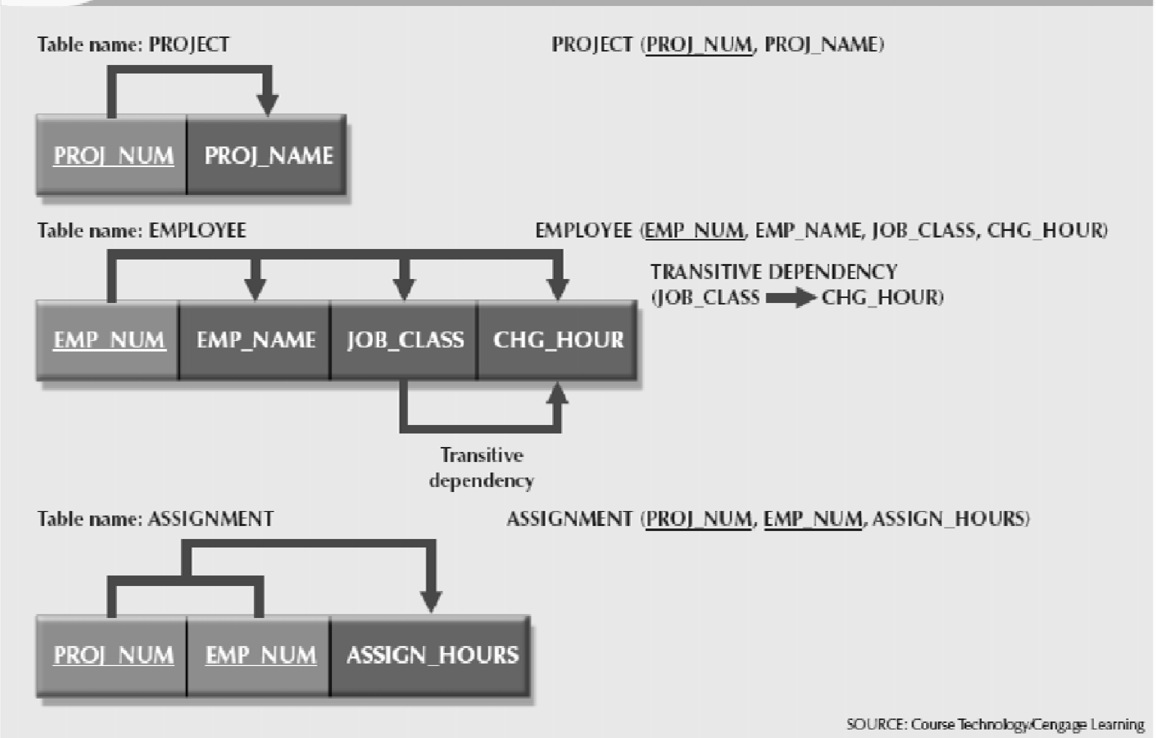
Conversion to First Normal Form (cont'd.)

- First normal form describes tabular format:
  - All key attributes are defined
  - 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 are based on part of the primary key
  - Should be used with caution

Conversion to Second Normal Form

- Step 1: Make New Tables to Eliminate Partial Dependencies
  - Write each key component on separate line, then write original (composite) key on last line
  - Each component will become key in new table
- Step 2: Reassign Corresponding Dependent Attributes
  - Determine attributes that are dependent on other attributes
  - At this point, most anomalies have been eliminated

FIGURE 9.4 Second normal form (2NF) conversion results



Conversion to Second Normal Form (cont'd.)

- 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

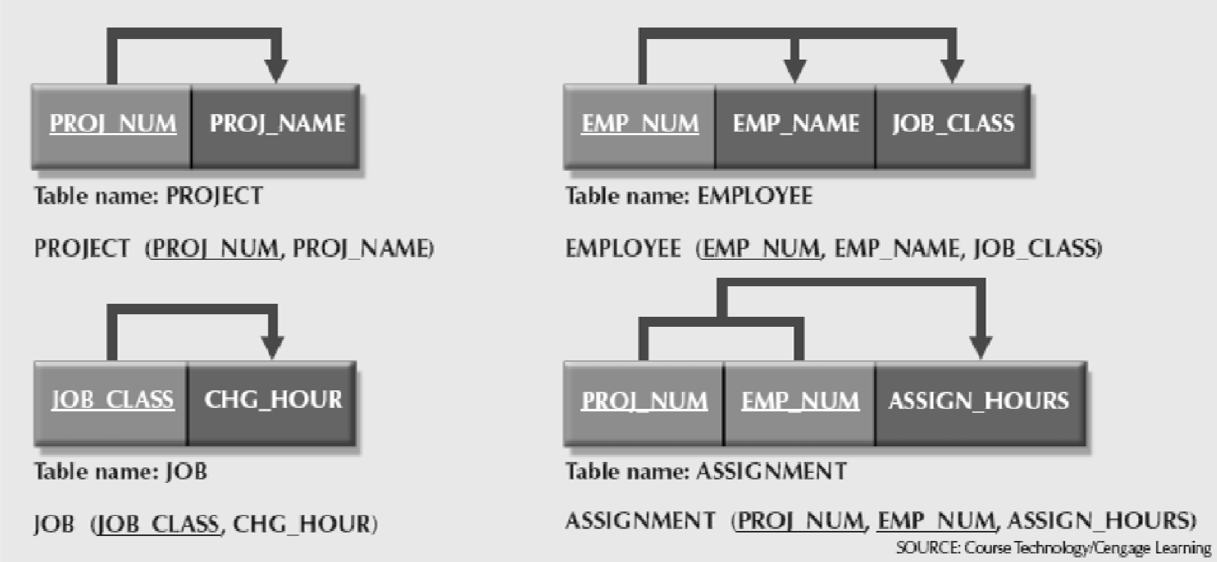
Conversion to Third Normal Form

- Step 1: Make New Tables to Eliminate Transitive Dependencies
  - 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 (cont'd.)

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

FIGURE 9.5 Third normal form (3NF) conversion results



SOURCE: Course Technology/Cengage Learning

Conversion to Third Normal Form (cont'd.)

- 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 should be cleaned up to eliminate initial partial and transitive dependencies
- Normalization cannot, by itself, be relied on to make good designs
- Valuable because it helps eliminate data redundancies

Improving the Design (cont'd.)

- 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

Improving the Design (cont'd.)

- Identify New Relationships
- Refine Primary Keys as Required for Data Granularity
- Maintain Historical Accuracy
- Evaluate Using Derived Attributes

FIGURE 9.6 The completed database

Table name: PROJECT

Table name: JOB   Database name: Ch09_ConstructCo

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: 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

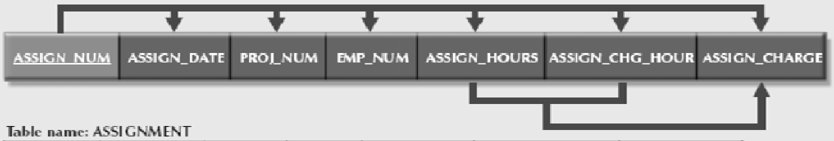


Table name: ASSIGNMENT

ASSIGN_NUM	ASSIGN_DATE	PROJ_NUM	EMP_NUM	ASSIGN_HOURS	ASSIGN_CHG_HOUR	ASSIGN_CHARGE
1001	04-Mar-12	15	103	2.6	84.50	219.70
1002	04-Mar-12	18	118	1.4	18.36	25.70
1003	05-Mar-12	15	101	3.6	105.00	378.00
1004	05-Mar-12	22	113	2.5	48.10	120.25
1005	05-Mar-12	15	103	1.9	84.50	160.55
1006	05-Mar-12	25	115	4.2	96.75	406.35
1007	05-Mar-12	22	105	5.2	105.00	546.00
1008	05-Mar-12	25	101	1.7	105.00	178.50
1009	05-Mar-12	15	105	2.0	105.00	210.00
1010	06-Mar-12	15	102	3.8	96.75	367.65
1011	06-Mar-12	22	104	2.6	96.75	251.55
1012	06-Mar-12	15	101	2.3	105.00	241.50
1013	06-Mar-12	25	114	1.8	48.10	86.58
1014	06-Mar-12	22	111	4.0	26.87	107.48
1015	06-Mar-12	25	114	3.4	48.10	163.54
1016	06-Mar-12	18	112	1.2	45.95	55.14
1017	06-Mar-12	18	118	2.0	18.36	36.72
1018	06-Mar-12	18	104	2.6	96.75	251.55
1019	06-Mar-12	15	103	3.0	84.50	253.50
1020	07-Mar-12	22	105	2.7	105.00	283.50
1021	08-Mar-12	25	108	4.2	96.75	406.35
1022	07-Mar-12	25	114	5.8	48.10	278.98
1023	07-Mar-12	22	106	2.4	35.75	85.80

SOURCE: Course Technology/Cengage Learning

FIGURE 9.6 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

SOURCE: Course Technology/Cengage Learning

Surrogate Key Considerations

- When primary key is considered to be unsuitable, designers use surrogate keys
- Data entries in Table 6.4 are inappropriate because they duplicate existing records
  - No violation of entity or referential integrity

TABLE 9.4 Duplicate Entries in the JOB Table

JOB_CODE	JOB_DESCRIPTION	JOB_CHG_HOUR
511	Programmer	$35.75
512	Programmer	$35.75

Higher-Level Normal Forms

- Tables in 3NF perform suitably in business transactional databases
- Higher-order normal forms are useful on occasion
- Two special cases of 3NF:
  - Boyce-Codd normal form (BCNF)
  - Fourth normal form (4NF)

The Boyce-Codd Normal Form

- 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 (cont'd.)

- Most designers consider the BCNF as a 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 fail to meet BCNF
  - No partial dependencies, nor does it contain transitive dependencies
  - A nonkey attribute is the determinant of a key attribute

© 2013 Cengage Learning. All Rights Reserved. This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition. May not be scanned, copied, duplicated, or posted to a publicly accessible website, in whole or in part.

FIGURE 9.7 A table that is in 3NF but not in BCNF

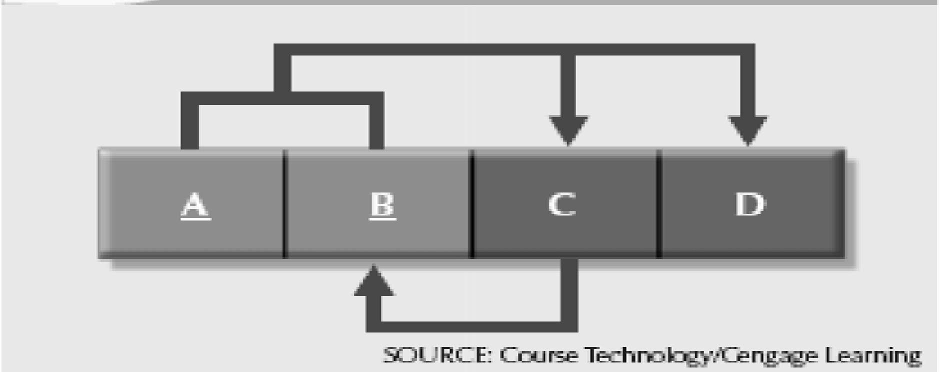
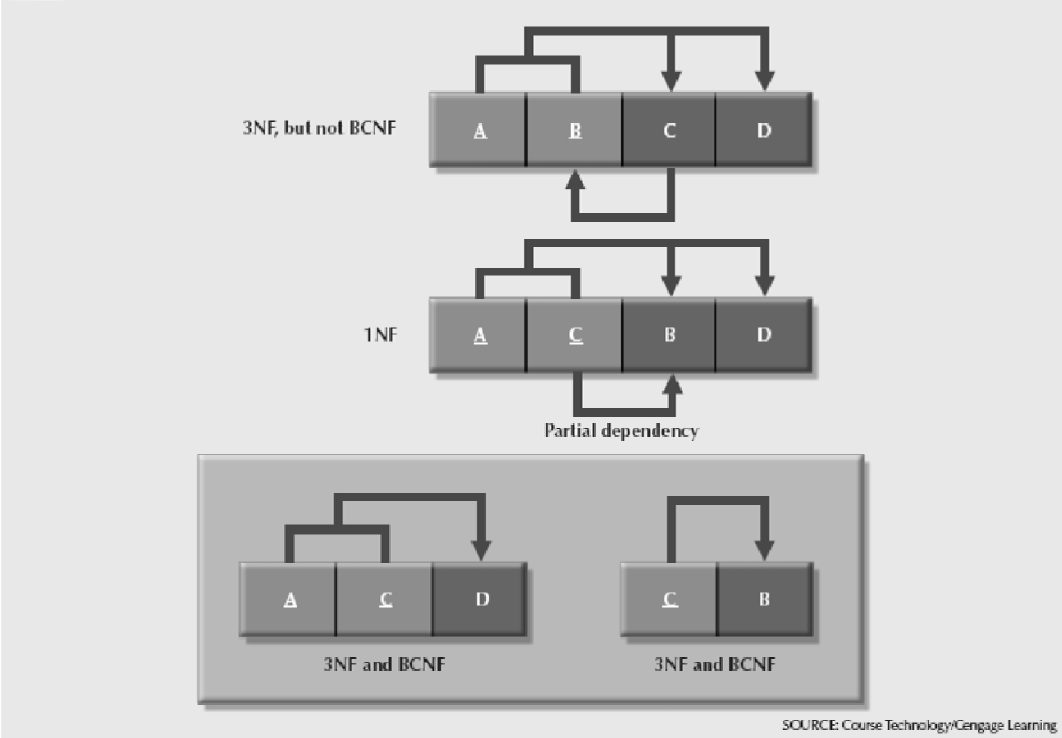


TABLE 9.5 Sample Data for a BCNF Conversion

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

FIGURE 9.8 Decomposition to BCNF



Fourth Normal Form (4NF)

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

FIGURE 9.10 Tables with multivalued dependencies

Database name: Ch09_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
10123		4

Table name: VOLUNTEER_V3

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

SOURCE: Course Technology/Cengage Learning

FIGURE 9.11 A set of tables in 4NF

Database name: CH09_Service

Table name: PROJECT

PROJ_CODE	PROJ_NAME	PROJ_BUDGET
1	BeThere	1023245.00
2	BlueMoon	20198608.00
3	GreenThumb	3234456.00
4	GoFast	5674000.00
5	GoSlow	1002500.00

Table name: EMPLOYEE

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

Table name: ORGANIZATION

ORG_CODE	ORG_NAME
RC	Red Cross
UW	United Way
WF	Wildlife Fund

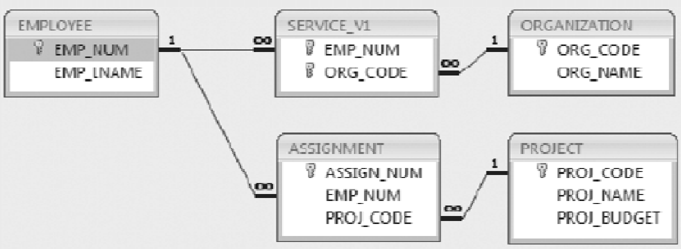
Table name: ASSIGNMENT

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



SOURCE: Course Technology/Cengage Learning

Normalization and Database Design

- Normalization should be part of the 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
- You may be asked to redesign and modify existing databases

Normalization and Database Design (cont'd.)

- ER diagram
  - Identify relevant entities, their attributes, and their relationships
  - Identify additional entities and attributes
- Normalization procedures
  - Focus on characteristics of specific entities
  - Micro view of entities within ER diagram
- Difficult to separate normalization process from ER modeling process

FIGURE 9.12 Initial contracting company ERD

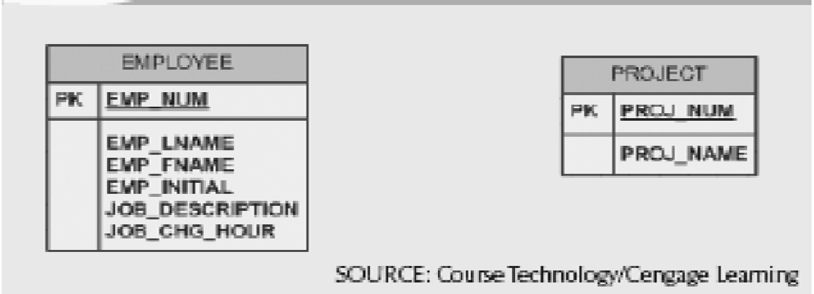


FIGURE 9.13 Modified contracting company ERD

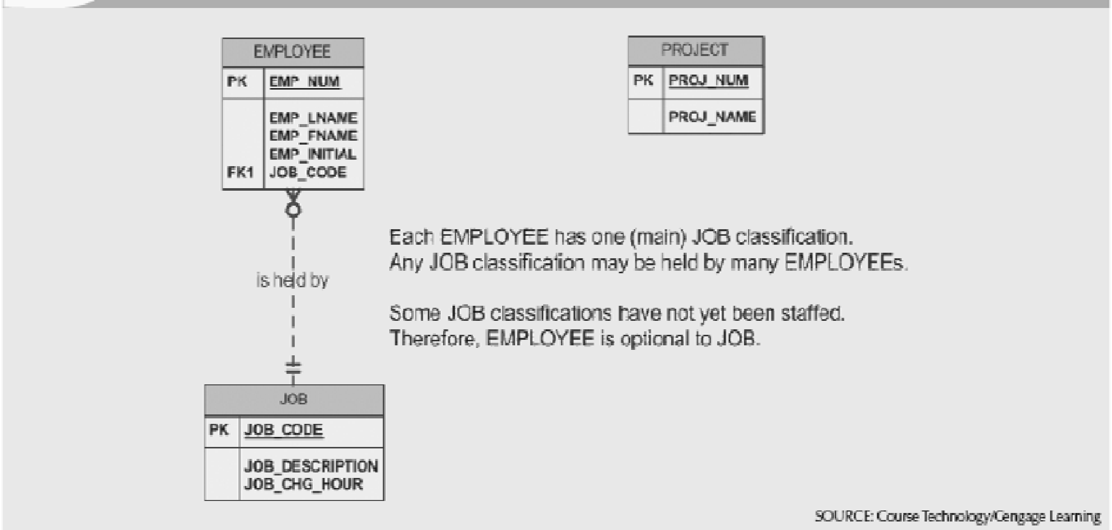
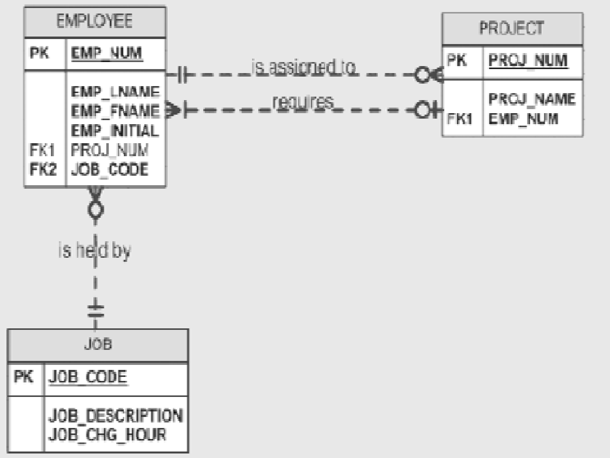
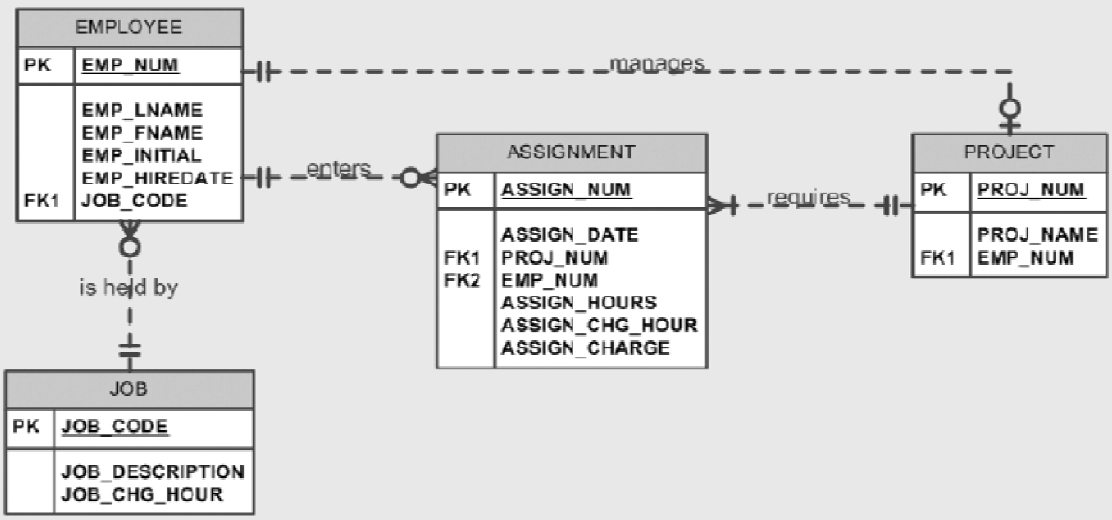


FIGURE 9.14 Incorrect M:N relationship representation



SOURCE: Course Technology/Cengage Learning

FIGURE 9.15 Final contracting company ERD



SOURCE: Course Technology/Cengage Learning

FIGURE 9.16 The implemented database

Table name: EMPLOYEE

Database name: Ch09_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-12	15	103	2.6	84.50	219.70
1002	04-Mar-12	18	118	1.4	18.36	25.70
1003	05-Mar-12	15	101	3.6	105.00	378.00
1004	05-Mar-12	22	113	2.5	48.10	120.25
1005	05-Mar-12	15	103	1.9	84.50	160.55
1006	05-Mar-12	25	115	4.2	96.75	406.35
1007	05-Mar-12	22	105	5.2	105.00	546.00
1008	05-Mar-12	25	101	1.7	105.00	178.50
1009	05-Mar-12	15	105	2.0	105.00	210.00
1010	06-Mar-12	15	102	3.8	96.75	367.65
1011	06-Mar-12	22	104	2.6	96.75	251.55
1012	06-Mar-12	15	101	2.3	105.00	241.50
1013	06-Mar-12	25	114	1.8	48.10	86.58
1014	06-Mar-12	22	111	4.0	26.87	107.48
1015	06-Mar-12	25	114	3.4	48.10	163.54
1016	06-Mar-12	18	112	1.2	45.95	55.14
1017	06-Mar-12	18	118	2.0	18.36	36.72
1018	06-Mar-12	18	104	2.6	96.75	251.55
1019	06-Mar-12	15	103	3.0	84.50	253.50
1020	07-Mar-12	22	105	2.7	105.00	283.50
1021	08-Mar-12	25	108	4.2	96.75	406.35
1022	07-Mar-12	25	114	5.8	48.10	278.98
1023	07-Mar-12	22	106	2.4	35.75	85.80

SOURCE: Course Technology/Cengage Learning

Denormalization

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

Denormalization (cont'd.)

- Joining the larger number of tables reduces system speed
- Conflicts are often resolved through compromises that may include denormalization
- Defects of unnormalized tables:
  - Data updates are less efficient because tables are larger
  - Indexing is more cumbersome
  - No simple strategies for creating virtual tables known as views

Data-Modeling Checklist

- Data modeling translates specific real-world environment into data model
  - Represents real-world data, users, processes, interactions
- Data-modeling checklist helps ensure that data-modeling tasks are successfully performed
- Based on concepts and tools learned in Part II

TABLE 9.7 Data-Modeling Checklist

BUSINESS RULES

- Properly document and verify all business rules with the end users.
- Ensure that all business rules are written precisely, clearly, and simply. The business rules must help identify entities, attributes, relationships, and constraints.
- Identify the source of all business rules, and ensure that each business rule is justified, dated, and signed off by an approving authority.

DATA MODELING

Naming conventions: All names should be limited in length (database-dependent size).

- Entity names:
  - Should be nouns that are familiar to business and should be short and meaningful
  - Should document abbreviations, synonyms, and aliases for each entity
  - Should be unique within the model
  - For composite entities, may include a combination of abbreviated names of the entities linked through the composite entity
- Attribute names:
  - Should be unique within the entity
  - Should use the entity abbreviation as a prefix
  - Should be descriptive of the characteristic
  - Should use suffixes such as _ID, _NUM, or _CODE for the PK attribute
  - Should not be a reserved word
  - Should not contain spaces or special characters such as @, !, or &
- Relationship names:
  - Should be active or passive verbs that clearly indicate the nature of the relationship

Entities:

- Each entity should represent a single subject.
- Each entity should represent a set of distinguishable entity instances.
- All entities should be in 3NF or higher. Any entities below 3NF should be justified.
- The granularity of the entity instance should be clearly defined.
- The PK should be clearly defined and support the selected data granularity.

Attributes:

- Should be simple and single-valued (atomic data)
- Should document default values, constraints, synonyms, and aliases
- Derived attributes should be clearly identified and include source(s)
- Should not be redundant unless this is required for transaction accuracy, performance, or maintaining a history
- Nonkey attributes must be fully dependent on the PK attribute

Relationships:

- Should clearly identify relationship participants
- Should clearly define participation, connectivity, and document cardinality

ER model:

- Should be validated against expected processes: inserts, updates, and deletes
- Should evaluate where, when, and how to maintain a history
- Should not contain redundant relationships except as required (see attributes)
- Should minimize data redundancy to ensure single-place updates
- Should conform to the minimal data rule: All that is needed is there, and all that is there is needed.

Summary

- Normalization minimizes data redundancies
- First three normal forms (1NF, 2NF, and 3NF) are most commonly encountered
- Table is in 1NF when:
  - All key attributes are defined
  - All remaining attributes are dependent on primary key

Summary (cont'd.)

- 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 the design process

Summary (cont'd.)

- Table in 3NF may contain multivalued dependencies
  - Numerous null values or redundant data
- Convert 3NF table to 4NF by:
  - Splitting table to remove multivalued dependencies
- Tables are sometimes denormalized to yield less I/O, which increases processing speed

© 2013 Cengage Learning. All Rights Reserved. This edition is intended for use outside of the U.S. only, with content that may be different from the U.S. Edition.
May not be scanned, copied, duplicated, or posted to a publicly accessible website, in whole or in part.