

pID	name	team	teamPhone	pos1	pos2	pos3
1	Pessi	Argentina	54-11-1000-1000	str	for	∅
2	Ricardo	Portugal	351-2-7777-7777	rm	dm	∅
3	Neumann	Brazil	55-21-4040-2020	for	lb	rb
4	Baily	Wales	44-29-1876-1876	dm	st	∅
5	Marioso	Argentina	54-11-1000-1000	sw	dm	st
6	Pare	Brazil	55-21-4040-2020	go	∅	∅

pID	name	team	teamPhone	pos1	pos2	pos3
1	Pessi	Argentina	54-11-1000-1000	str	for	∅
2	Ricardo	Portugal	351-2-7777-7777	rm	dm	∅
3	Neumann	Brazil	55-21-4040-2020	for	lb	rb
4	Baily	Wales	44-29-1876-1876	dm	st	∅
5	Marioso	Argentina	54-11-1000-1000	sw	dm	st
6	Pare	Brazil	55-21-4040-2020	go	∅	∅
∅	∅	Iceland	354-5-109-20	∅	∅	∅

Insert anomaly

pID	name	team	teamPhone	pos1	pos2	pos3
1	Pessi	Argentina	54-11-1000-1000	str	for	∅
2	Ricardo	Portugal	351-2-7777-7777	rm	dm	∅
3	Neumann	Brazil	55-21-4040-2020	for	lb	rb
4	Baily	Wales	44-29-1876-1876	dm	st	∅
5	Marioso	Argentina	54-11-1000-1000	sw	dm	st
6	Pare	Brazil	55-21-4040-2020	go	∅	∅

Update anomaly

pID	name	team	teamPhone	pos1	pos2	pos3
1	Pessi	Argentina	54-11-1000-1000	str	for	∅
2	Ricardo	Portugal	351-2-7777-7777	rm	dm	∅
3	Neumann	Brazil	55-21-4040-2020	for	lb	rb
4	Baily	Wales	44-29-1876-1876	dm	st	∅
5	Marioso	Argentina	54-11-1000-1000	sw	dm	st
6	Pare	Brazil	55-21-4040-2020	go	∅	∅

Delete anomaly

Normal Forms

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

Cengage Learning © 2015

Functional Dependence Concepts

Concept	Definition
Functional dependence (SKU, Price)	The attribute B is fully functionally dependent on the attribute A if each value of A determines one and only one value of B.
Functional dependence (Generalized definition) (Department, DeptName)	Attribute A determines attribute B 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) (LicenseNumber, State, VIN)	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.

Table name: STUDENT

Database name: Ch03_TinyCollege

STU_NUM	STU_LNAME	STU_FNAME	STU_INIT	STU_DOB	STU_HRS	STU_CLASS	STU_GPA	STU_TRANSFER	DEPT_CODE	STU_PHONE	PROF_NUM
321452	Bowser	William	C	12-Feb-1985	42	So	2.84	No	BIOL	2134	205
324257	Smithson	Anne	K	15-Nov-1991	81	Jr	3.27	Yes	CIS	2256	222
324258	Brewer	Juliette		23-Aug-1979	36	So	2.26	Yes	ACCT	2256	228
324269	Oblonski	Walter	H	16-Sep-1986	66	Jr	3.09	No	CIS	2114	222
324273	Smith	John	D	30-Dec-1968	102	Sr	2.11	Yes	ENGL	2231	199
324274	Katinga	Raphael	P	21-Oct-1989	114	Sr	3.15	No	ACCT	2267	228
324291	Robertson	Gerald	T	08-Apr-1983	120	Sr	3.87	No	EDU	2267	311
324299	Smith	John	B	30-Nov-1996	15	Fr	2.92	No	ACCT	2315	230

- STU_NUM = Student number
- STU_LNAME = Student last name
- STU_FNAME = Student first name
- STU_INIT = Student middle initial
- STU_DOB = Student date of birth
- STU_HRS = Credit hours earned
- STU_CLASS = Student classification
- STU_GPA = Grade point average
- STU_TRANSFER = Student transferred from another institution
- DEPT_CODE = Department code
- STU_PHONE = 4-digit campus phone extension
- PROF_NUM = Number of the professor who is the student's advisor

STU_NUM → STU_GPA ✓

(STU_NUM, STU_LNAME) → STU_GPA ✗

Player (pID, pName, pAddr, pHSID, pHSName, pHSCity, pRank, pPri)

Tryout(pID, cName, pos, decision)

pRank	Priority
90-100	1
70-89	2
60-69	3

if s and t are tuples of R , then
 $s.pRank = t.pRank \rightarrow s.pPri = t.pPri$

Thus, this is a functional dependency for R :
 $pRank \rightarrow pPri$

Player (pID, pName, pAddr, pHSID, pHSName, pHSCity, pRank, pPri)

Tryout(pID, cName, pos, decision)

pRank	Priority
90-100	1
70-89	2
60-69	3

If students can only tryout for one position at any given college, then

Thus, this is a functional dependency for R :
 $pID, CName \rightarrow pos$

if we have a set of attributes A in Relation R that always determines all the other attributes B then we can say that A is a **key** for that Relation.

Activity: Find the FD's

- 1.(ISBN, Title, Author, PrintDate, Edition, Publisher)**
- 2.(NetID, LName, DeptCode, DeptName, CommitteeCode)**
- 3.(Department, CourseNumber, CourseName, CourseDescription)**
- 4.(ClientID, ClientName, ContractNum, ContractDate, ContractHours, ClientHourlyRate, ContractPrice)**
- 5.(FlightNum, Date, Airline, PlaneType, PlaneRange, NumSeats, SeatsSold, OriginAirport, DestAirport)**

Normalization Process

- 1. Create a detailed description of operations.**
- 2. Derive all the appropriate business rules from the description of operations.**
- 3. Model the data with the help of a good tool to produce an initial ERD blueprint.**
- 4. Use the normalization procedures to remove data redundancies. This process may produce additional entities.**
- 5. Revise the ERD created in step 3.**
- 6. Use the normalization procedures to audit the revised ERD. If additional data redundancies are discovered, repeat steps 4 and 5.**

Normal Forms

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

Cengage Learning © 2015

Table name: RPT_FORMAT

Database name: Ch06_ConstructCo

PROJ_NUM	PROJECT_NAME	EMP_NUMBER	EMP_NAME	JOB_CLASS	CHARGE_HOUR	HOURS_BILLED
15	Evergreen	103,101,105, 106, 102	June E. Arbough, John G. News, Alice K. Johnson *, William Smithfield, David H. Senior	Elec. Engineer, Database Designer, Database Designer, Programmer, Systems Analyst	85.5, 105., 105., 35.75, 98.75	23.8, 19.4, 35.7, 12.6, 23.8
18	Amber Wave	114, 118, 104, 112	Annelise Jones, James J. Frommer, Anne K. Ramoras *, Darlene M. Smithson	Applications Designer, General Support, Systems Analyst, DSS Analyst	48.1, 18.36, 96.75, 45.95	25.6, 45.3, 32.4, 45.
22	Rolling Tide	105, 104, 113, 111, 106	Alice K. Johnson, Anne K. Ramoras, Delbert K. Joenbrood *, Geoff B. Wabash, William Smithfield	DB Designer, Systems Analyst, Applications Designer, Clerical Support, Programmer	105., 96.75, 48.1, 26.87, 35.75	65.7, 48.4, 23.6, 22., 12.8
25	Starflight	107, 115, 101, 114, 108, 118, 112	Maria D. Alonzo, Travis B. Bawangi, John G. News *, Annelise Jones, Ralph B. Washington, James J. Frommer, Darlene M. Smithson	Programmer, Systems Analyst, Database Design, Applications Designer, Systems Analyst, General Support, DSS Analyst	35.75, 96.75, 105., 48.1, 96.75, 18.36, 45.95	25.6, 45.8, 56.3, 33.1, 23.6, 30.5, 41.4

FIGURE 6.1 TABULAR REPRESENTATION OF THE REPORT FORMAT

Table name: RPT_FORMAT

Database name: Ch06_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

FIGURE 6.2 A TABLE IN FIRST NORMAL FORM

Table name: DATA_ORG_1NF

Database name: Ch06_ConstructCo

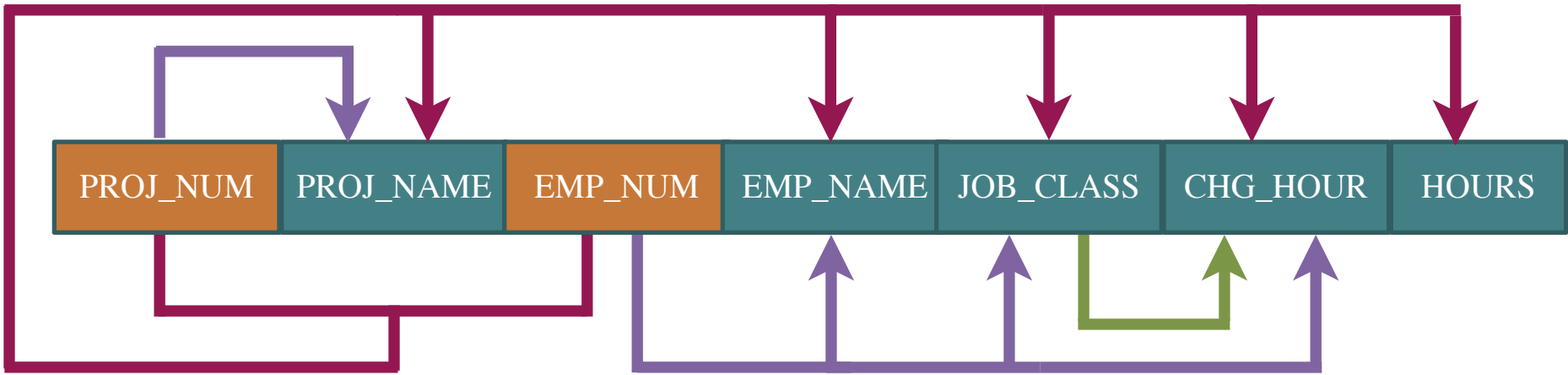
PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	June E. Arbough	Elect. Engineer	84.50	23.8
15	Evergreen	101	John G. News	Database Designer	105.00	19.4
15	Evergreen	105	Alice K. Johnson *	Database Designer	105.00	35.7
15	Evergreen	106	William Smithfield	Programmer	35.75	12.6
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

Surrogate Keys

- Used by designers when the primary key is considered to be unsuitable
- System-defined attribute
- Created and managed via the DBMS
- Have a numeric value which is automatically incremented for each new row

1st Normal Form

- Functional Dependency
- Partial Functional Dependency
- Transitive Functional Dependency



1NF: (PROJ_NUM, EMP_NUM, PROJ_NAME, EMP_NAME, JOB_CLASS, CHG_HOUR, HOURS)

PARTIAL FDs:

(PROJ_NUM → PROJ_NAME)

(EMP_NUM → EMP_NAME, JOB_CLASS, CHG_HOUR)

TRANSITIVE FD:

(JOB_CLASS → CHG_HOUR)

FIGURE 6.4 SECOND NORMAL FORM (2NF) CONVERSION RESULTS

Table name: PROJECT

PROJECT (PROJ_NUM, PROJ_NAME)

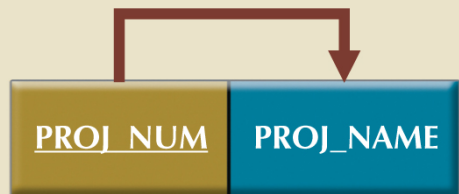
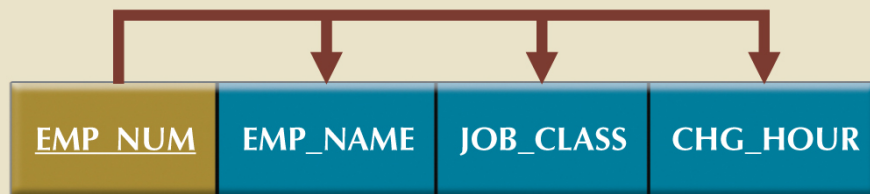


Table name: EMPLOYEE

EMPLOYEE (EMP_NUM, EMP_NAME, JOB_CLASS, CHG_HOUR)

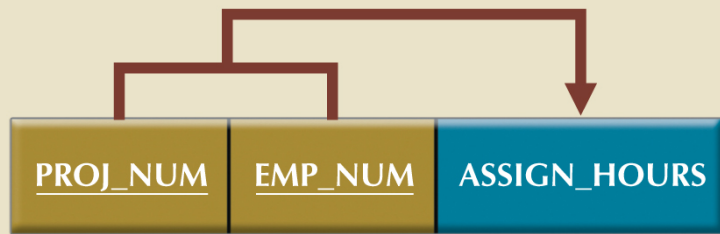


TRANSITIVE DEPENDENCY
(JOB_CLASS → CHG_HOUR)

Transitive
dependency

Table name: ASSIGNMENT

ASSIGNMENT (PROJ_NUM, EMP_NUM, ASSIGN_HOURS)



<i>ID</i>	<i>name</i>	<i>street</i>	<i>city</i>	<i>salary</i>
⋮				
57766	Kim	Main	Perryridge	75000
98776	Kim	North	Hampton	67000
⋮				

employee

<i>ID</i>	<i>name</i>
⋮	
57766	Kim
98776	Kim
⋮	

<i>name</i>	<i>street</i>	<i>city</i>	<i>salary</i>
⋮			
Kim	Main	Perryridge	75000
Kim	North	Hampton	67000
⋮			

natural join

<i>ID</i>	<i>name</i>	<i>street</i>	<i>city</i>	<i>salary</i>
⋮				
57766	Kim	Main	Perryridge	75000
57766	Kim	North	Hampton	67000
98776	Kim	Main	Perryridge	75000
98776	Kim	North	Hampton	67000
⋮				

FIGURE 6.5 THIRD NORMAL FORM (3NF) CONVERSION RESULTS

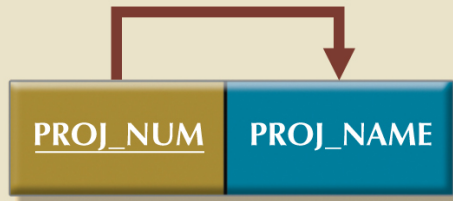


Table name: PROJECT

PROJECT (PROJ_NUM, PROJ_NAME)

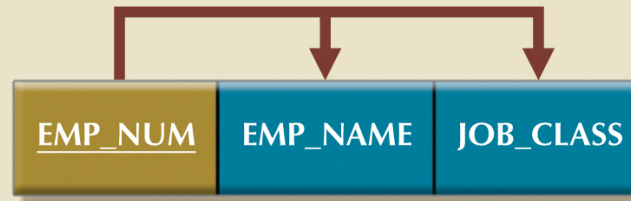


Table name: EMPLOYEE

EMPLOYEE (EMP_NUM, EMP_NAME, JOB_CLASS)

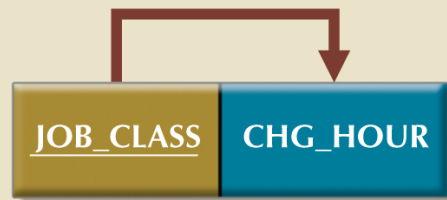


Table name: JOB

JOB (JOB_CLASS, CHG_HOUR)

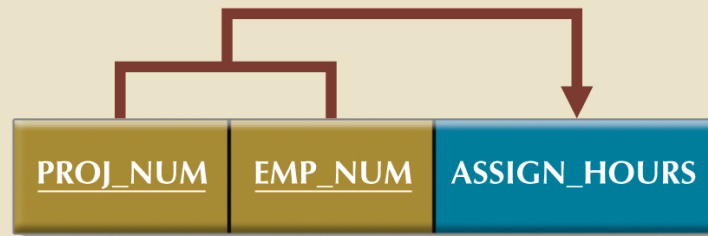


Table name: ASSIGNMENT

ASSIGNMENT (PROJ_NUM, EMP_NUM, ASSIGN_HOURS)

FIGURE 6.6 THE COMPLETED DATABASE

Table name: PROJECT

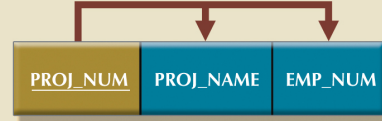


Table name: PROJECT

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

Table name: JOB

Database name: Ch06_ConstructCo



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

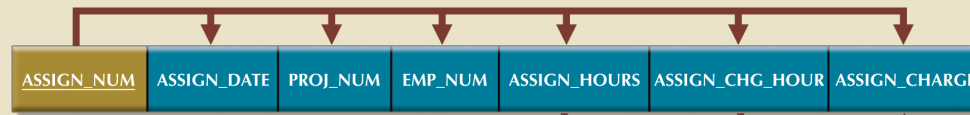


Table name: ASSIGNMENT

ASSIGN_NUM	ASSIGN_DATE	PROJ_NUM	EMP_NUM	ASSIGN_HOURS	ASSIGN_CHG_HOUR	ASSIGN_CHARGE
1001	04-Mar-16	15	103	2.6	84.50	219.70
1002	04-Mar-16	18	118	1.4	18.36	25.70
1003	05-Mar-16	15	101	3.6	105.00	378.00
1004	05-Mar-16	22	113	2.5	48.10	120.25
1005	05-Mar-16	15	103	1.9	84.50	160.55
1006	05-Mar-16	25	115	4.2	96.75	406.35
1007	05-Mar-16	22	105	5.2	105.00	546.00
1008	05-Mar-16	25	101	1.7	105.00	178.50
1009	05-Mar-16	15	105	2.0	105.00	210.00
1010	06-Mar-16	15	102	3.8	96.75	367.65
1011	06-Mar-16	22	104	2.6	96.75	251.55
1012	06-Mar-16	15	101	2.3	105.00	241.50
1013	06-Mar-16	25	114	1.8	48.10	86.58
1014	06-Mar-16	22	111	4.0	26.87	107.48
1015	06-Mar-16	25	114	3.4	48.10	163.54
1016	06-Mar-16	18	112	1.2	45.95	55.14
1017	06-Mar-16	18	118	2.0	18.36	36.72
1018	06-Mar-16	18	104	2.6	96.75	251.55
1019	06-Mar-16	15	103	3.0	84.50	253.50
1020	07-Mar-16	22	105	2.7	105.00	283.50
1021	08-Mar-16	25	108	4.2	96.75	406.35
1022	07-Mar-16	25	114	5.8	48.10	278.98
1023	07-Mar-16	22	106	2.4	35.75	85.80

FIGURE 6.6 THE COMPLETED DATABASE (CONTINUED)

Table name: EMPLOYEE

Database name: Ch06_ConstructCo

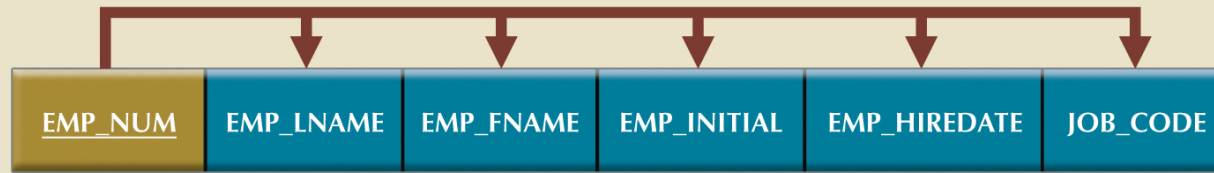


Table name: EMPLOYEE

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	Joebrood	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

Data depends on the key [1NF], the whole key [2NF], and nothing but the key [3NF].

— Unknown (to me, at least)

BCNF - Boyce-Codd Normal Form

(a.k.a. 3.5NF)

FIGURE 6.8 A TABLE THAT IS IN 3NF BUT NOT IN BCNF

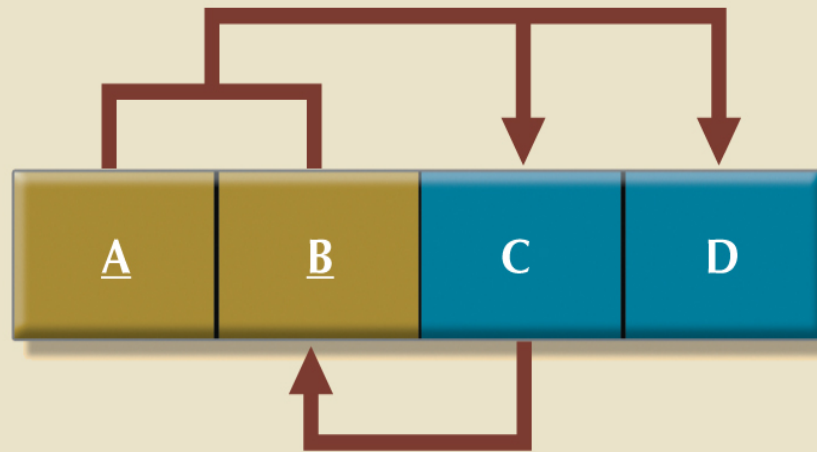
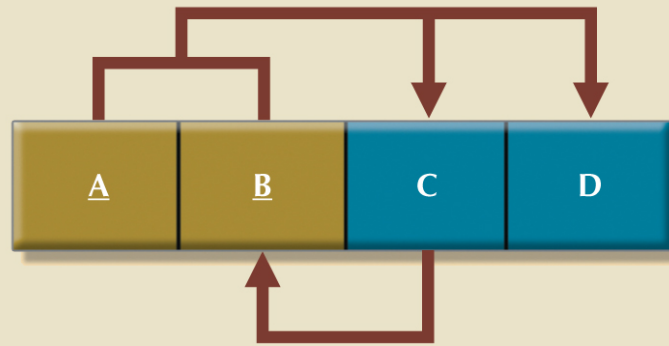
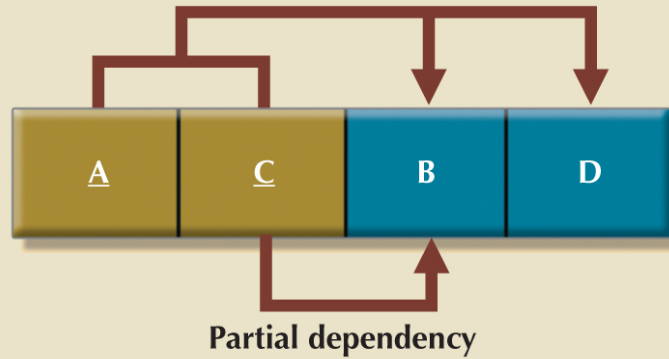


FIGURE 6.9 DECOMPOSITION TO BCNF

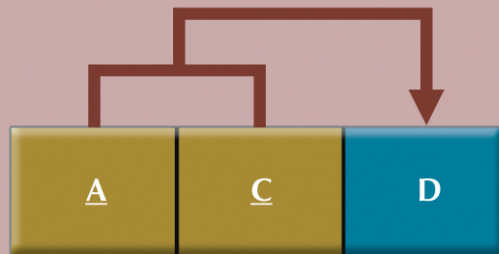
3NF, but not BCNF



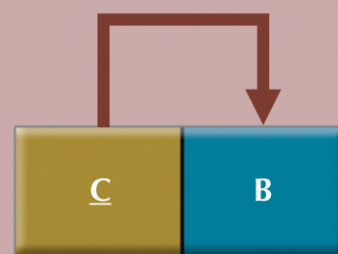
1NF



Partial dependency



3NF and BCNF



3NF and BCNF

Data-Modeling Checklist - part 1

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.

Cengage Learning © 2015

Data-Modeling Checklist - part 2

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

Data-Modeling Checklist - part 3

DATA MODELING

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

Data-Modeling Checklist - part 4

DATA MODELING

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 deletions
- 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.

Cengage Learning © 2015