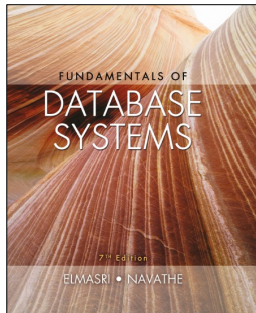


## Fundamentals of Database Systems

Seventh Edition



### Chapter 14

Basics of Functional Dependencies and Normalization for Relational Databases

1

## Database Tables and Normalization

- Table is basic building block in database design
- The table's structure is of great interest of us
- How do you recognize poor table structure, and produce a good table? NORMALIZATION
- Normalization is process for assigning attributes to entities
  - Reduces data redundancies
  - Helps eliminate data anomalies
  - Produces controlled redundancies to link tables
- Normalization works through a series of stages called normal forms
  - 1NF - First normal form
  - 2NF - Second normal form
  - 3NF - Third normal form

2

**TABLE 7.1 A sample report layout**

Proj. Num.	Project Name	Employee Number	Employee Name	Job Class	Chg/Hour	Hours Billed	Total Charge
15	Evergreen	103	Mzwandile E. Baloyi	Elect. Engineer	€67.55	23.8	€1 607.69
		101	John G. News	Database Designer	€82.95	19.4	€1 999.23
		105	Alice K. Johnson*	Database Designer	€82.95	35.7	€2 991.32
		106	William Smithfield	Programmer	€26.66	12.6	€335.92
		102	Kavyara H. Moonsamy	Systems Analyst	€76.43	23.8	€1 819.03
			<b>Subtotal</b>				<b>€8 333.19</b>
18	Amber Wave	114	Annelise Jones	Applications Designer	€38.00	25.6	€972.80
		118	James J. Frommer	General Support	€14.50	45.3	€656.85
		104	Noxolo K. Maseki*	Systems Analyst	€76.43	32.4	€2 476.33
		112	Darlene M. Smithson	DSS Analyst	€36.30	45.0	€1 633.50
			<b>Subtotal</b>				<b>€5 739.48</b>
22	Rolling Tide	105	Alice K. Johnson	Database Designer	€82.95	65.7	€5 449.82
		104	Noxolo K. Maseki	Systems Analyst	€76.43	48.4	€3 699.21
		113	Delbert K. Joenbrood*	Applications Designer	€38.00	23.6	€896.80
		111	Geoff B. Wabash	Clerical Support	€21.23	22.0	€467.06
		106	William Smithfield	Programmer	€28.24	12.8	€361.47
			<b>Subtotal</b>				<b>€10 874.36</b>
25	Starflight	107	Maria D. Alonzo	Programmer	€28.24	25.6	€722.94
		115	Travis B. Bawangi	Systems Analyst	€76.43	45.8	€3 500.49
		101	John G. News*	Database Designer	€82.95	56.3	€4 670.09
		114	Annelise Jones	Applications Designer	€38.00	33.1	€1 257.80
		108	Krishshanth B. Khan	Systems Analyst	€76.43	23.6	€1 803.75
			<b>Subtotal</b>				<b>€13 900.14</b>
			<b>Total</b>				<b>€38 942.09</b>

Note: \* indicates project leader

A Sample Report Layout of an IT firm that manages several projects

3

**FIGURE 7.1 Tabular representation of the report format**

Database name: Ch07\_ConstrucGo

Table name: RPT\_FORMAT

PROJ. NUM	PROJ. NAME	EMP. NUM	EMP. NAME	JOB. CLASS	CHG. HOUR	HOURS
15	Evergreen	103	Mzwandile E. Baloyi	Elect. Engineer	€67.55	23.80
		101	John G. News	Database Designer	€82.95	19.40
		105	Alice K. Johnson *	Database Designer	€82.95	35.70
		106	William Smithfield	Programmer	€26.66	12.60
		102	Kavyara H. Moonsamy	Systems Analyst	€76.43	23.80
18	Amber Wave	114	Annelise Jones	Applications Designer	€38.00	24.60
		118	James J. Frommer	General Support	€14.50	45.30
		104	Noxolo K. Maseki *	Systems Analyst	€76.43	32.40
		112	Darlene M. Smithson	DSS Analyst	€36.30	44.00
		105	Alice K. Johnson	Database Designer	€82.95	64.70
22	Rolling Tide	104	Noxolo K. Maseki	Systems Analyst	€76.43	48.40
		113	Delbert K. Joenbrood *	Applications Designer	€38.10	23.60
		111	Geoff B. Wabash	Clerical Support	€21.23	22.00
		106	William Smithfield	Programmer	€28.24	12.80
		107	Maria D. Alonzo	Programmer	€28.24	24.80
25	Starflight	115	Travis B. Bawangi	Systems Analyst	€76.43	45.80
		101	John G. News *	Database Designer	€82.95	56.30
		114	Annelise Jones	Applications Designer	€38.00	33.10
		108	Krishshanth B. Khan	Systems Analyst	€76.43	23.60
		118	James J. Frommer	General Support	€14.50	30.50
		112	Darlene M. Smithson	DSS Analyst	€36.30	41.40

Need for Normalization

A table's structure matches the report format

4

## Observations

The structure does not match the requirements in Relational Database Model. It does not handle data very well. Examples;

- The PROJ\_NUM is apparently intended to be a primary key, or at least a part of a primary key, but it contains nulls.
- The table entries invite the data inconsistencies. (the JOB\_CLASS value "Elect. Engineer" might be entered as "Elect. Eng." in some cases, "El. Eng." and "EE").
- The table displays data redundancies. These data redundancies yield the following anomalies.
  - Update Anomalies.** Modifying the JOB\_CLASS for employee 105 requires many alterations.
  - Insertion Anomalies.** Just to complete a row definition, a new employee must be assigned to a project. If the employee is not yet assigned, the imaginary project must be created in order to complete the employee data entry.
  - Deletion Anomalies.** If employee 103 quits, deletions must be made for every entry in which EMP\_NUM = 103. As such deletions are made, other vital data are lost, too.

Also, every time another employee assigned to a project, some data entries (such as PROJ\_NAME, EMP\_NAME, CHG\_HOUR) are unnecessarily repeated. Since, all copies of redundant data is not identical, these confusions are data integrity problems.

5

## Conversion to 1NF

- Repeating groups must be eliminated
  - Proper primary key developed
    - Uniquely identifies attribute values (rows)
    - Combination of PROJ\_NUM and EMP\_NUM
  - Dependencies can be identified
    - The arrows above entities indicate all desirable dependencies, that is, dependencies that are based on the primary key
    - The arrows below the dependency diagram indicate less desirable dependencies
      - Partial Dependency**
        - Dependencies based on part of composite primary key (i.e. You need to know only the PROJ\_NUM to determine PROJ\_NAME, that is the PROJ\_NAME is dependent on only part of the primary key)
      - Transitive Dependency**
        - Dependency of one nonkey attribute on another nonkey attribute (i.e. CHG\_HOUR is dependent on JOB\_CLASS. Because neither CHG\_HOUR nor JOB\_CLASS is a key attribute)

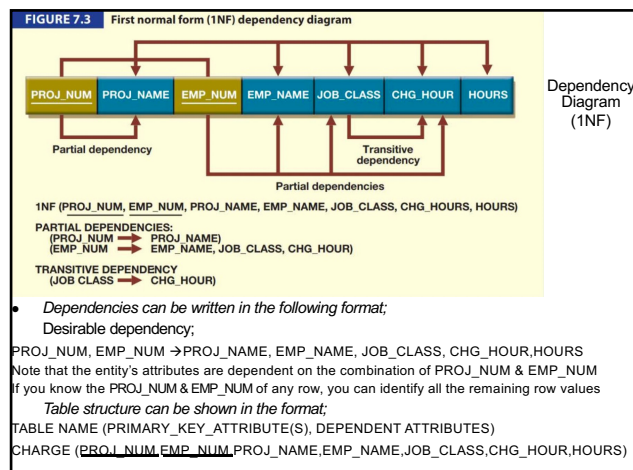
6

Database name: Ch07\_ConstructCo  
Table name: DATA\_ORG\_1NF

**Conversion to 1NF**

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	Mzwandile E. Baloyi	Elect. Engineer	€67.55	23.80
15	Evergreen	101	John G. News	Database Designer	€82.95	19.40
15	Evergreen	105	Alice K. Johnson *	Database Designer	€82.95	35.70
15	Evergreen	106	William Smithfield	Programmer	€26.66	12.60
15	Evergreen	102	Kavyara H. Moonsamy	Systems Analyst	€76.43	23.80
18	Amber Wave	114	Annelise Jones	Applications Designer	€38.00	24.60
18	Amber Wave	118	James J. Frommer	General Support	€14.50	45.30
18	Amber Wave	104	Noxolo K. Maseki *	Systems Analyst	€76.43	32.40
18	Amber Wave	112	Darlene M. Smithson	DSS Analyst	€36.30	44.00
22	Rolling Tide	105	Alice K. Johnson	Database Designer	€82.95	64.70
22	Rolling Tide	104	Noxolo K. Maseki	Systems Analyst	€76.43	48.40
22	Rolling Tide	113	Delbert K. Joenbrood *	Applications Designer	€38.00	23.60
22	Rolling Tide	111	Geoff B. Wabash	Clerical Support	€21.23	22.00
22	Rolling Tide	106	William Smithfield	Programmer	€28.24	12.80
25	Starflight	107	Maria D. Alonzo	Programmer	€28.24	24.60
25	Starflight	115	Travis B. Bawangi	Systems Analyst	€76.43	45.80
25	Starflight	101	John G. News *	Database Designer	€82.95	56.30
25	Starflight	114	Annelise Jones	Applications Designer	€38.00	33.10
25	Starflight	108	Krishshanth B. Khan	Systems Analyst	€76.43	23.60
25	Starflight	118	James J. Frommer	General Support	€14.50	30.50
25	Starflight	112	Darlene M. Smithson	DSS Analyst	€36.30	41.40

7



8

## 1NF Summarized

- The table is in first normal form (1NF) if:
- All key attributes defined
  - No repeating groups in table. (Each row, column intersection can contain one and only one value, not a set of values).
  - All attributes dependent on primary key

9

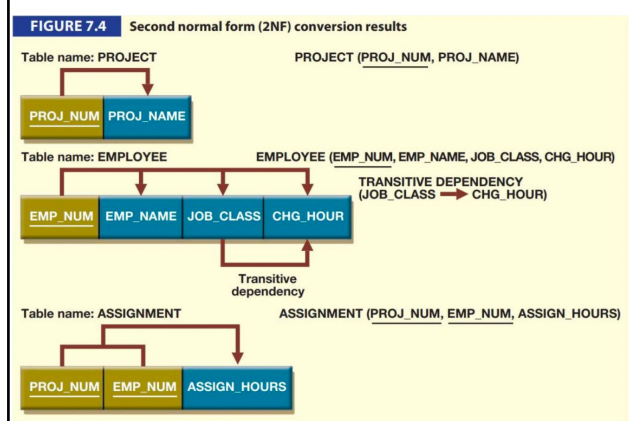
## Conversion to 2NF

- Start with 1NF format:
- Write each key component on separate line
- Write original key on last line
- Each component is 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)

10

## 2NF Conversion Results



11

## 2NF Summarized

The table is in second normal form (2NF) if:

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

NOTE: Because a partial dependency can exist only if a table's primary key is composed of several attributes, a table whose primary key consists of only a single attribute must automatically be in 2NF if it is in 1NF.

12

## Conversion to 3NF

- Create separate table(s) to eliminate transitive functional dependencies

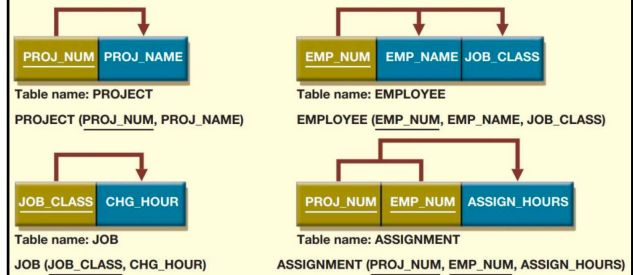
```
PROJECT (PROJ_NUM, PROJ_NAME)
ASSIGN (PROJ_NUM, EMP_NUM, HOURS)
EMPLOYEE (EMP_NUM, EMP_NAME, JOB_CLASS)
JOB (JOB_CLASS, CHG_HOUR)
```

\*JOB\_CLASS must be kept in the original 2NF table as a foreign key in order to establish a link between the original table and the newly created table.

●13

## Conversion to 3NF

FIGURE 7.5 Third normal form (3NF) conversion results



●14

## 3NF Summarized

The table is in third normal form (3NF) if:

- In 2NF
- Contains no transitive dependencies

●15

## Normalization and Database Design

- Normalization should be part of the design process
- E-R Diagram provides a macro view of an organization's data requirements and operations
  - Created through an iterative process
  - Begins by identifying relevant entities, their attributes and their relationships
- Normalization provides micro view of entities
  - Focuses on characteristics of specific entities
  - May yield additional entities and attributes
- Difficult to separate normalization from E-R diagramming
- Business rules must be determined

●16

## Normalization and Database Design (Con't.)

- The operations of construction company can be summarized by the following business rules:
  - 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 employees are part of a labor pool, to be shared by all project teams. For example, the company's executive secretary would not be assigned to any one particular project.
  - Each employee has a (single) primary job classification. This job classification determines the hourly billing rate.
  - Many employees can have the same job classification. For example, the company employs more than one electrical engineer.

●17

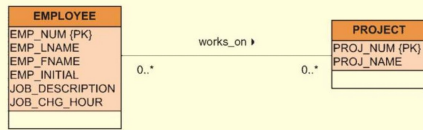
## Normalization and Database Design (Con't.)

- The operations of construction company can be summarized by the following business rules:
  - 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 employees are part of a labor pool, to be shared by all project teams. For example, the company's executive secretary would not be assigned to any one particular project.
  - Each employee has a (single) primary job classification. This job classification determines the hourly billing rate.
  - Many employees can have the same job classification. For example, the company employs more than one electrical engineer.
- Given this simple description of company's operations, two entities and their attributes are initially defined.
  - PROJECT (PROJ\_NUM, PROJ\_NAME)
  - EMPLOYEE (EMP\_NUM, EMP\_LNAME, EMP\_FNAME, EMP\_INITIAL, JOB\_DESCRIPTION, JOB\_CHG\_HOUR)

●18

## Initial ERD for an IT Firm

FIGURE 7.12 Initial contracting company ERD

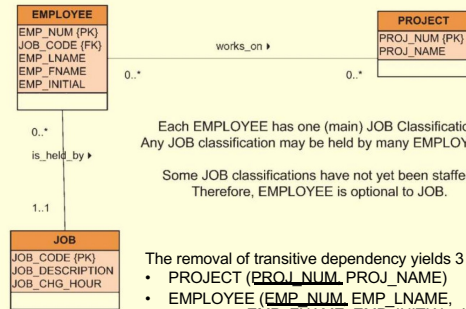


NOTE: Some employees are not assigned to a project. Therefore, PROJECT is optional to EMPLOYEE.

- PROJECT is in 3NF and needs no modification.
- EMPLOYEE needs additional examination.  
JOB\_DESCRIPTION → JOB\_CHG\_HOUR.  
Therefore, EMPLOYEE contains transitive dependency.

## Modified ERD for an IT Firm

FIGURE 7.13 Modified contracting company ERD



Each EMPLOYEE has one (main) JOB Classification. Any JOB classification may be held by many EMPLOYEEs.

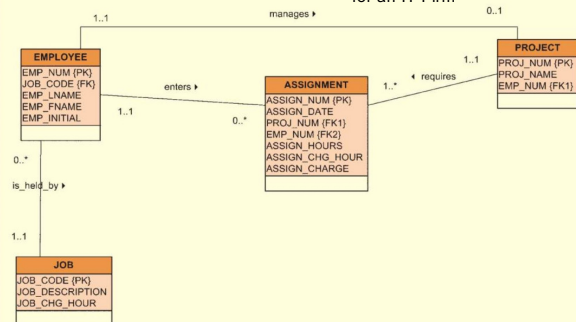
Some JOB classifications have not yet been staffed. Therefore, EMPLOYEE is optional to JOB.

- The removal of transitive dependency yields 3 entities;
- PROJECT (PROJ\_NUM, PROJ\_NAME)
  - EMPLOYEE (EMP\_NUM, EMP\_LNAME, EMP\_FNAME, EMP\_INITIAL, JOB\_CODE)
  - JOB (JOB\_CODE, JOB\_DESCRIPTION, JOB\_CHG\_HOUR)

19

21

FIGURE 7.15 Final contracting company ERD  
Final ERD for an IT Firm



- PROJECT (PROJ\_NUM, PROJ\_NAME, EMP\_NUM)
- EMPLOYEE (EMP\_NUM, EMP\_LNAME, EMP\_FNAME, EMP\_INITIAL, EMP\_HIREDATE, JOB\_CODE)
- JOB (JOB\_CODE, JOB\_DESCRIPTION, JOB\_CHG\_HOUR)
- ASSIGN (ASSIGN\_NUM, ASSIGN\_DATE, ASSIGN\_HOURS, ASSIGN\_CHG\_HOUR, ASSIGN\_CHARGE, EMP\_NUM, PROJ\_NUM)

24

FIGURE 7.6 The completed database

Database name: Ch07\_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: EMPLOYEE

EMP_NUM	EMP_LNAME	EMP_FNAME	EMP_INITIAL	EMP_HIREDATE	JOB_CODE
101	News	John	G	08-Nov-10	502
102	Moonsamy	Kavyara	H	12-Jul-99	501
103	Baloyi	Mzwandile	E	01-Dec-07	503
104	Maseki	Noxolo	K	15-Nov-98	501
105	Johnson	Alice	K	01-Feb-04	502
106	Smithfield	William		22-Jun-15	500
107	Alonzo	Maria	D	10-Oct-04	500
108	Khan	Krishshanth	B	22-Aug-99	501
109	Smith	Larry	W	18-Jul-09	501
110	Olenko	Gerald	A	11-Dec-06	505
111	Wabash	Geoff	B	04-Apr-99	506
112	Smithson	Darlene	M	23-Oct-05	507
113	Joenbrood	Delbert	K	15-Nov-04	508
114	Jones	Annelise		20-Aug-01	508

26

Table name: ASSIGNMENT

ASSIGN_NUM	ASSIGN_DATE	PROJ_NUM	EMP_NUM	ASSIGN_HOURS	ASSIGN_CHG_HOUR	ASSIGN_CHARGE
1001	04-Mar-18	15	103	2.60	€67.55	€175.63
1002	04-Mar-18	18	118	1.40	€14.50	€20.30
1003	05-Mar-18	15	101	3.60	€82.95	€298.62
1004	05-Mar-18	22	113	2.50	€38.00	€95.00
1005	05-Mar-18	15	103	1.90	€67.55	€128.35
1006	05-Mar-18	25	115	4.20	€76.43	€321.01
1007	05-Mar-18	22	105	5.20	€82.95	€431.34
1008	05-Mar-18	25	101	1.70	€82.95	€141.02
1009	05-Mar-18	15	105	2.00	€82.95	€165.90
1010	06-Mar-18	15	102	3.80	€76.43	€290.43
1011	06-Mar-18	22	104	2.60	€76.43	€198.72
1012	06-Mar-18	15	101	2.30	€82.95	€190.79
1013	06-Mar-19	25	114	1.80	€38.00	€68.40
1014	06-Mar-19	22	111	4.00	€21.23	€84.92
1015	06-Mar-19	25	114	3.40	€38.00	€129.20
1016	06-Mar-19	18	112	1.20	€36.30	€43.56
1017	06-Mar-19	18	118	2.00	€14.50	€29.00
1018	06-Mar-19	18	104	2.60	€76.43	€198.72
1019	06-Mar-19	15	103	3.00	€67.55	€202.65
1020	07-Mar-19	22	105	2.70	€82.95	€223.97
1021	08-Mar-19	25	108	4.20	€76.43	€321.01
1022	07-Mar-19	25	114	5.80	€38.00	€220.40
1023	07-Mar-19	22	106	2.40	€28.24	€67.78

27

Table name: JOB

JOB_CODE	JOB_DESCRIPTION	JOB_CHG_HOUR
500	Programmer	€28.24
501	Systems Analyst	€76.43
502	Database Designer	€82.95
503	Electrical Engineer	€66.76
504	Mechanical Engineer	€53.64
505	Civil Engineer	€44.07
506	Clerical Support	€21.23

Table name: EMPLOYEE

EMP_NUM	EMP_LNAME	EMP_FNAME	EMP_INITIAL	EMP_HIREDATE	JOB_CODE
101	News	John	G	08-Nov-10	502
102	Moonsamy	Kavyara	H	12-Jul-99	501
103	Baloyi	Mzwandile	E	01-Dec-07	503
104	Maseki	Noxolo	K	15-Nov-98	501
105	Johnson	Alice	K	01-Feb-04	502
106	Smithfield	William		22-Jun-15	500
107	Alonzo	Maria	D	10-Oct-04	500
108	Khan	Krishshanth	B	22-Aug-99	501
109	Smith	Larry	W	18-Jul-09	501
110	Olenko	Gerald	A	11-Dec-06	505
111	Wabash	Geoff	B	04-Apr-99	506
112	Smithson	Darlene	M	23-Oct-05	507
113	Joenbrood	Delbert	K	15-Nov-04	508
114	Jones	Annelise		20-Aug-01	508
115	Bawangi	Travis	B	25-Jan-00	501
116	Pratt	Gerald	L	05-Mar-05	510
117	Williamson	Angie	H	19-Jun-04	509
118	Frommer	James	J	04-Jan-16	510

28



## Denormalization

- Normalization is one of many database design goals
- Good database design also considers processing requirements
- As tables are decomposed to fit to normalization requirements, the number of database tables grows
- Joining the larger number of tables takes additional disk input/output (I/O) operations and processing, thereby reducing system speed
 

Advantage of higher processing speed  
AGAINST  
Disadvantage of data anomalies
- On the other hand, some anomalies are only theoretical interest
  - Should people worry that a ZIP\_CODE determines CITY in a CUSTOMER table whose primary key is the customer number?
  - ZIP (ZIP\_CODE, CITY) Is it really practical in real-world? NO!
- Normalization purity is difficult to preserve due to conflict in:
  - Design efficiency
  - Information requirements
  - Processing speed
- Use denormalization carefully!

29

## Summary

- Normalization is a technique used to design tables in which data redundancies are minimized.
- Higher normal forms are better than lower normal forms, because higher normal forms yield relatively fewer data redundancies.
- A table is in 1NF when all the key attributes are defined and when all remaining attributes are dependent on the primary key. A Table in 1NF can still contain both partial and transitive dependencies.
- A partial dependency is one in which an attribute is functionally dependent on only a part of a multiattribute (composite) primary key.
- A transitive dependency is one in which one attribute is functionally dependent on another nonkey attribute.

30

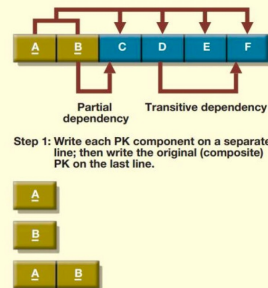
## Summary (Con't.)

- A table is in 2NF when it is in 1NF and contains no partial dependencies. A 1NF table is automatically in 2NF if its primary key is based on only a single attribute. A table in 2NF may still contain transitive dependencies.
- A table is in 3NF if it is in 2NF and contains no transitive dependencies.
- Normalization is part of design process. As entities and attributes are defined during the E-R modeling process, put each entity (set) through normalization checks and form new entities into the E-R diagram and continue the iterative E-R process until all entities and their attributes are defined and all equivalent tables are in 3NF.

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

31

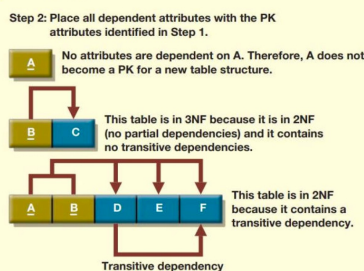
FIGURE 7.17 The initial 1NF structure



The Initial 1NF Structure

32

FIGURE 7.18 Identifying possible PK attributes



Identifying Possible PK Attributes

33

FIGURE 7.19 Table structures based on the selected PKs

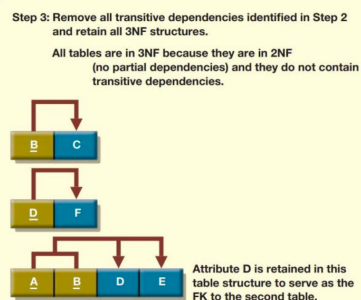


Table Structures Based on the Selected PKs

34