Database Management Systems

Associate Professor Dr. Raed Ibraheem Hamed

University of Human Development,
College of Science and Technology
Computer Science Department

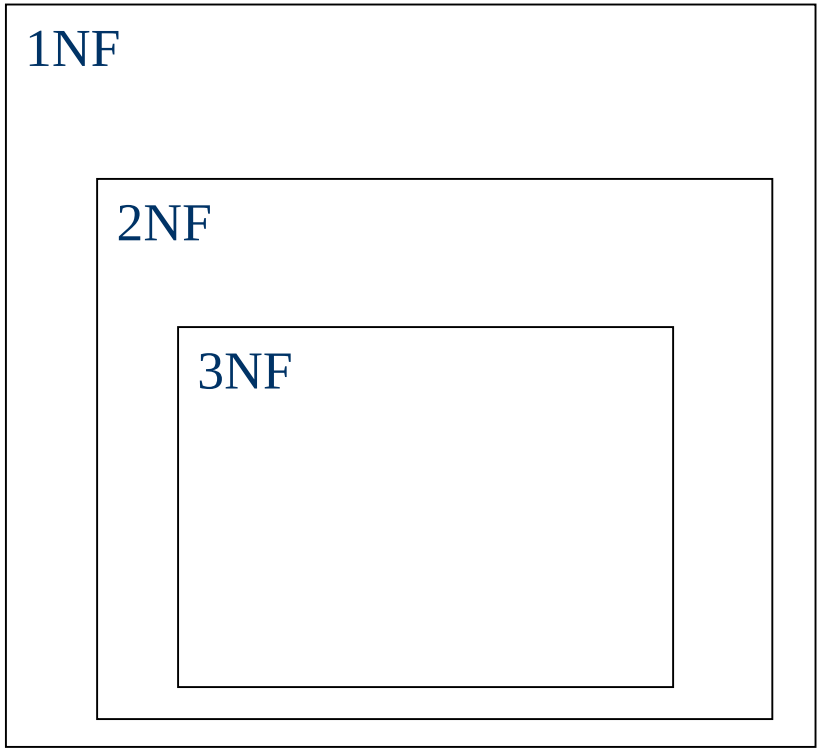
2015 – 2016

Database Tables and Normalization

- **Normalization** is a process that "improves" a database design by generating relations that are of higher normal forms. It reduces data redundancies and helps eliminate the data anomalies.
- **Normalization works through a series of stages called normal forms:**
  - First normal form (1NF)
  - Second normal form (2NF)
  - Third normal form (3NF)

Normalization

We consider a relation in 3NF to be fully normalized.



a relation in 3NF is also in 2NF

a relation in 2NF is also in 1NF

a relation in 1NF is need a process

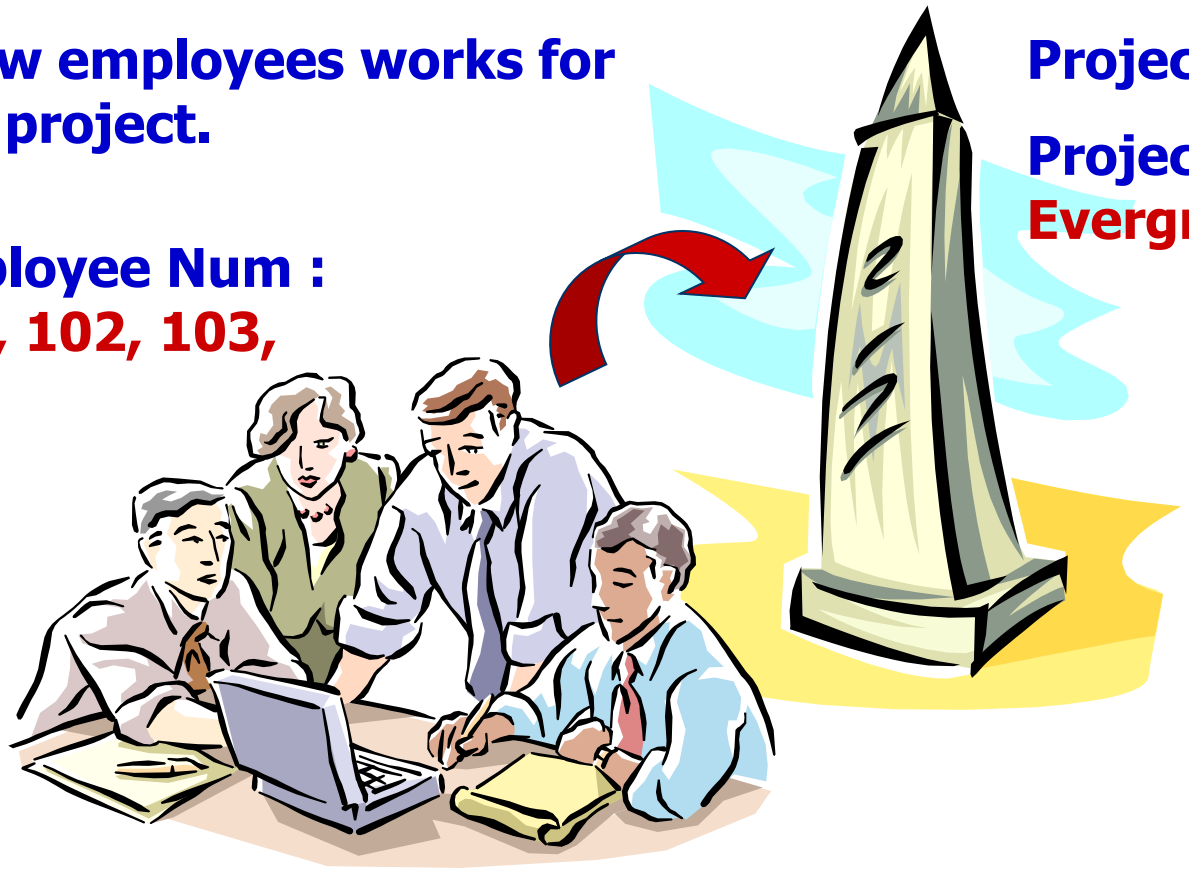
Database Tables and Normalization

- **The Need for Normalization**
  - **Case of a Construction Company**
    - **Building project -- Project number, Name, Employees assigned to the project.**
    - **Employee -- Employee number, Name, Job classification**
    - **The company charges its clients by billing the hours spent on each project. The hourly billing rate is dependent on the employee's position.**
    - **Periodically, a report is generated.**
    - **The table whose contents correspond to the reporting requirements is shown in Table 5.1.**

Scenario

A few employees works for one project.

Employee Num : 101, 102, 103, 105



Project Num : 15

Project Name : Evergreen

Sample Form

Project Num : 15

Project Name : Evergreen

Emp Num	Emp Name	Job Class	Chr Hours	Hrs Billed	Total
101					
102					
103					
105					

A Sample Report Layout

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	$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		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		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: * indicates project leader

Table Structure Matches the Report Format

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 5.1 A TABLE WHOSE STRUCTURE MATCHES THE REPORT FORMAT

Database Tables and Normalization

- **Problems with the Figure 5.1**
  - **The project number is intended to be a primary key, but it contains nulls.**
  - **The table displays data redundancies.**
  - **The table entries invite data inconsistencies.**
  - **The data redundancies yield the following anomalies:**
    - **Update anomalies.**
    - **Insertion anomalies.**
    - **Deletion anomalies.**

First Normal Form (1 NF)

- **1NF Definition**
  - The term first normal form (1NF) describes the tabular format in which:
    - All the key attributes are defined.
    - There are no repeating groups in the table.
    - All attributes are dependent on the primary key.

Database Tables and Normalization

- **Conversion to First Normal Form**
  - **A relational table must not contain repeating groups.**
  - **Repeating groups can be eliminated by adding the appropriate entry at the primary key column(s).**

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

FIGURE 5.2 THE EVERGREEN DATA

Data Organization: First Normal Form

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 5.1 A TABLE WHOSE STRUCTURE MATCHES THE REPORT FORMAT

Before

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	12.5
15	Evergreen	102	David H. Senior	Systems Analyst	$96.75	23.9
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	32.1
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	48.9
22	Rolling Tide	113	Delbert K. Joenbrood *	Applications Designer	$48.10	23.6
22	Rolling Tide	111	Geoff B. Wabash	Clerical Support	$26.87	22.5
22	Rolling Tide	106	William Smithfield	Programmer	$35.75	12.1
25	Starflight	107	Maria D. Alonzo	Programmer	$35.75	24.7
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	23.9
25	Starflight	118	James J. Frommer	General Support	$18.36	30.2
25	Starflight	112	Darlene M. Smithson	DSS Analyst	$45.95	41.4

FIGURE 5.3 DATA ORGANIZATION: FIRST NORMAL FORM

After

Second Normal Form (2 NF)

- Dependency Diagram

  - The primary key components are underlined .
  - The arrows above entities indicate all desirable dependencies, i.e., dependencies that are based on PK.
  - The arrows below the dependency diagram indicate less desirable dependencies -- partial dependencies and transitive dependencies.

Dependency Diagram Concept

Example: Suppose we keep track of employee email addresses, and we only track one **email address** for each employee.

Suppose each employee is identified by their unique **employee number**. We say there is a **functional dependency** of email address on employee number:

employee number → email address

Functional Dependencies

EmpNum	**EmpEmail**	**EmpFname**	**EmpLname**
123	jdoe@abc.com	John	Doe
456	psmith@abc.com	Peter	Smith
555	alee1@abc.com	Alan	Lee
633	pdoe@abc.com	Peter	Doe
787	alee2@abc.com	Alan	Lee

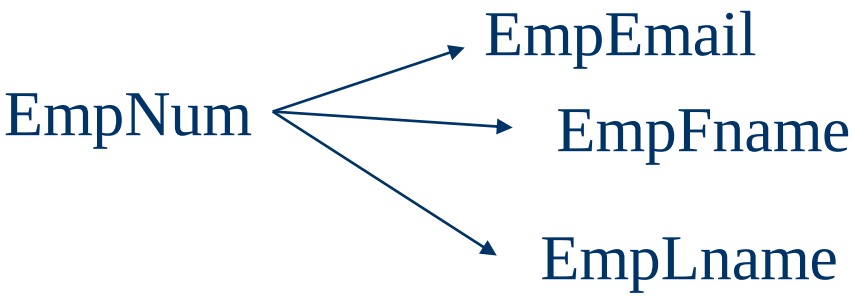
If EmpNum is the **PK** then the **FD**s:

EmpNum → EmpEmail
EmpNum → EmpFname
EmpNum → EmpLname

Functional Dependencies

EmpNum → EmpEmail
EmpNum → EmpFname
EmpNum → EmpLname

3 different ways you might see FDs depicted



Transitive Dependency Concept

Transitive dependency

Consider attributes A, B, and C, and where

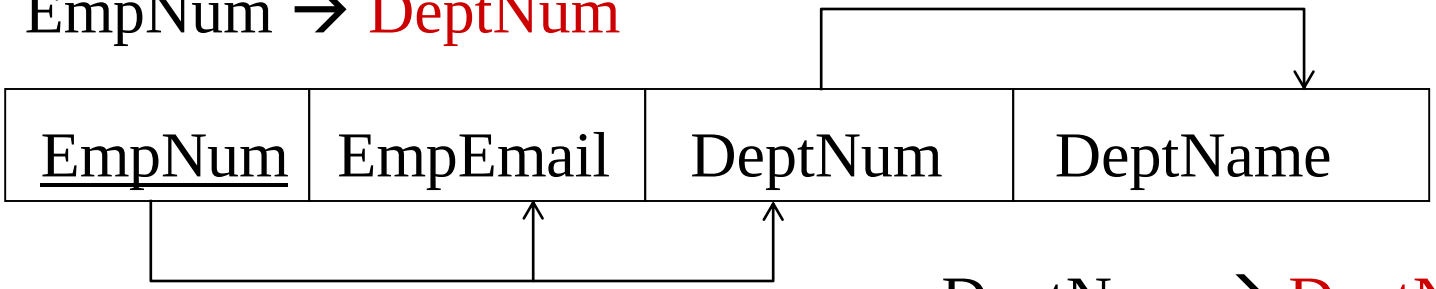
A → B and B → C.

Functional dependencies are transitive, which means that we also have the functional dependency A → C

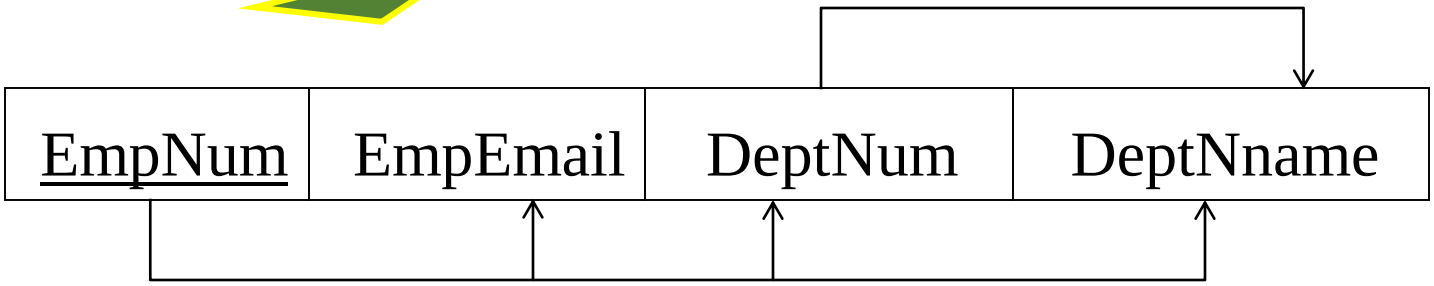
We say that **C is transitively dependent on A** through B.

Transitive dependency

EmpNum → DeptNum



DeptNum → DeptName

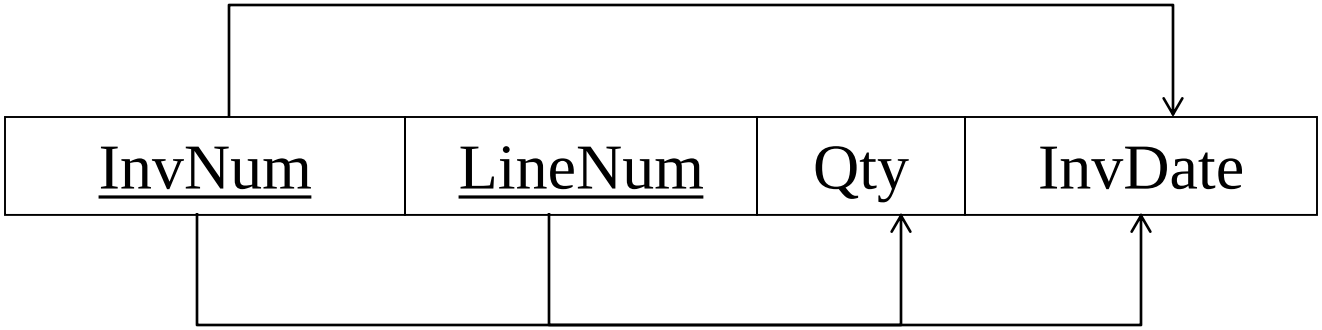


DeptName is *transitively dependent* on EmpNum via DeptNum

EmpNum → DeptName

Partial dependency Concept

A **partial dependency** exists when an attribute B is functionally dependent on an attribute A, and A is a component of a **multipart candidate key.**



Candidate keys: {InvNum, LineNum} **InvDate** is *partially dependent* on {InvNum, LineNum} as **InvNum** is a determinant of **InvDate** and **InvNum** is part of a candidate key

Dependency Diagram Example

personid	projectid	personname	projectcode	projectname	personphone
1	1	john	PROJ1	DBproject	345
2	1	sam	PROJ1	DBproject	754
3	1	anil	PROJ1	DBproject	456
4	1	das	PROJ1	DBproject	123
1	2	john	PROJ2	webproject	345
2	2	sam	PROJ2	webproject	754
3	2	anil	PROJ2	webproject	456
4	2	das	PROJ2	webproject	123

Dependency Diagram Result

person

personid	personname	personphone
1	john	345
2	sam	754
3	anil	456
4	das	123

project

projectid	projectcode	projectname
1	PROJ1	DBproject
2	PROJ2	webproject

personproject

personid	projectid
1	1
2	1
3	1
4	1
1	2
2	2
3	2
4	2

Dependency Diagram

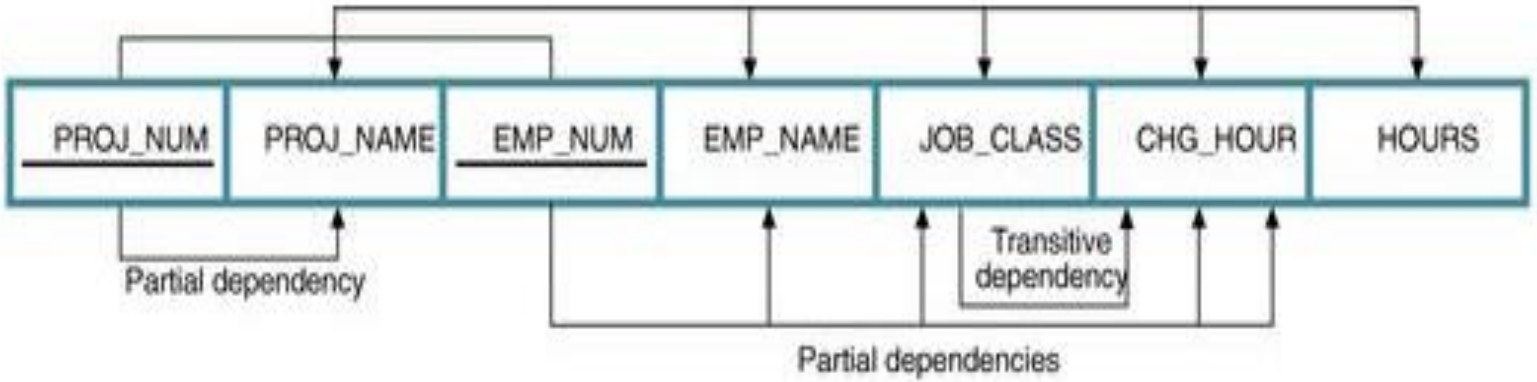


FIGURE 5.4 A DEPENDENCY DIAGRAM: FIRST NORMAL FORM (1NF)

Second Normal Form (2 NF)

All attributes dependent on full primary key

- **Conversion to Second Normal Form**
  - Starting with the 1NF format, the database can be converted into the 2NF format by
    - Writing each key component on a separate line, and then writing the original key on the last line and
    - Writing the dependent attributes after each new key.

PROJECT (PROJ_NUM, PROJ_NAME)

EMPLOYEE (EMP_NUM, EMP_NAME, JOB_CLASS, CHG_HOUR)

ASSIGN (PROJ_NUM, EMP_NUM, HOURS)

Dependency Diagram

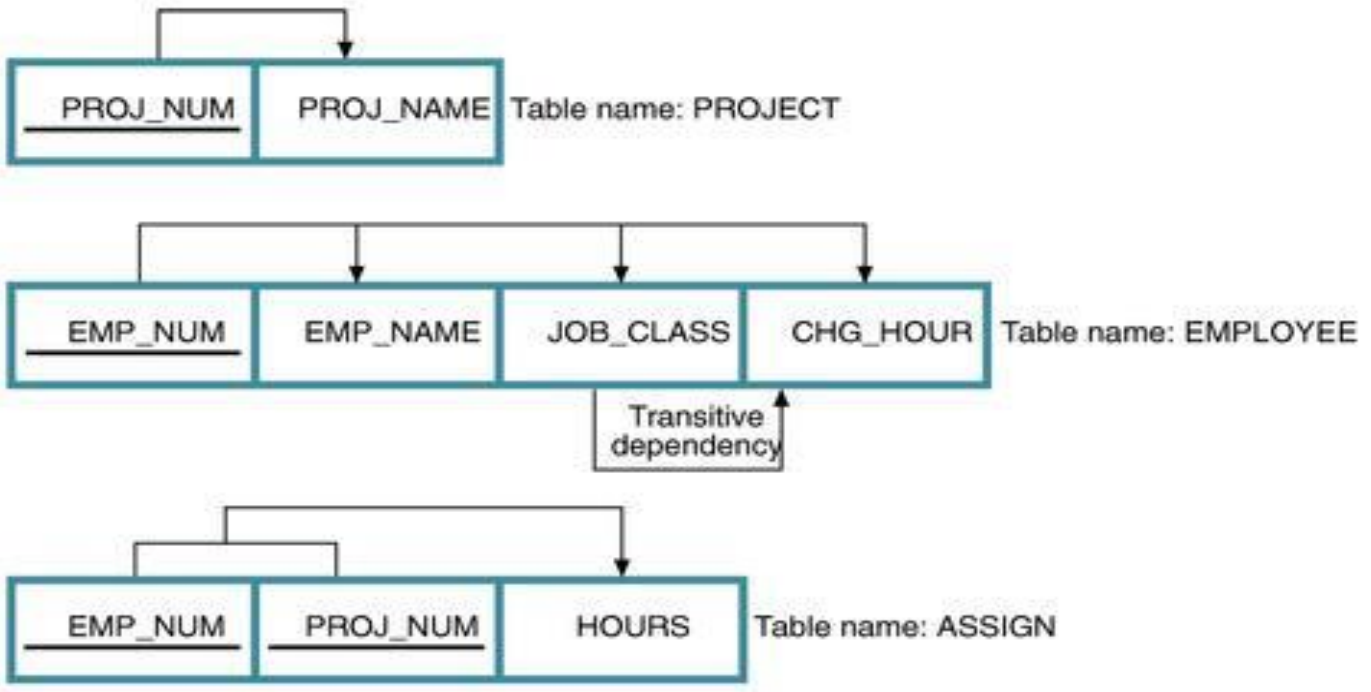


FIGURE 5.5 SECOND NORMAL FORM (2NF) CONVERSION RESULTS

Second Normal Form (2 NF)

A table is in 2NF if:

- It is in 1NF and
- It includes no partial dependencies; that is, no attribute is dependent on only a portion of the primary key.

  (It is still possible for a table in 2NF to exhibit **transitive dependency**; that is, one or more attributes may be functionally dependent on nonkey attributes.)

3NF- STOCKS example for 2nd normal form

The functional dependencies we can see are:

FD1: Symbol → Company , FD2: Company → place

so therefore: Symbol → place

This is a transitive dependency.

The solution again is to split this relation up into two new relations:

STOCK_SYMBOLS(Company, Symbol)

COMPANY_PLACE (Company, place)

This gives us the following sample data and FD for the new relations

FD1: Symbol → Company **FD2**: Company → place

Company	Symbol	place
Microsoft	MSFT	newyork
Oracle	ORCL	chicago

COMPANY_PLACE

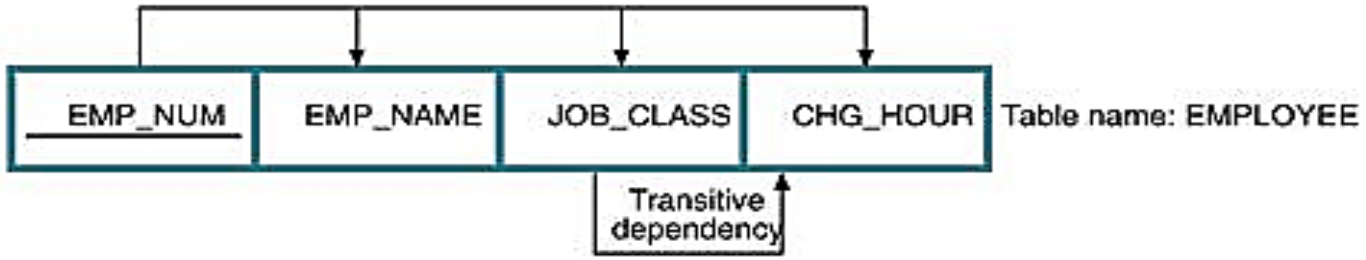
Company	place
Microsoft	newyork
Oracle	chicago

STOCK_SYMBOLS

Company	Symbol
Microsoft	MSFT
Oracle	ORCL

Third Normal Form (3 NF)

- **3NF Definition**
  - **A table is in 3NF if:**
    - **It is in 2NF and**
    - **It contains no transitive dependencies.**



Third Normal Form (3 NF)

All columns can be determined only by the key in the table and no other column

- **Conversion to Third Normal Form**
  - Create a separate table with attributes in a transitive functional dependence relationship.

PROJECT (PROJ_NUM, PROJ_NAME)

EMPLOYEE (EMP_NUM, EMP_NAME, JOB_CLASS)

ASSIGN (PROJ_NUM, EMP_NUM, HOURS)

JOB (JOB_CLASS, CHG_HOUR)

Normalization

- **Normalization will help us identify correct and appropriate TABLES.**
- **Until Now we have 4 tables**

PROJECT (PROJ_NUM, PROJ_NAME)

ASSIGN (PROJ_NUM, EMP_NUM, HOURS)

EMPLOYEE (EMP_NUM, EMP_NAME, JOB_CLASS)

JOB (JOB_CLASS, CHG_HOUR)

1st Normal Form Example

Un-normalized Students table:

Student#	AdvID	AdvName	AdvRoom	Class1	Class2
123	123A	James	555	102-8	104-9
124	123B	Smith	467	209-0	102-8

Normalized Students table:

Student#	AdvID	AdvName	AdvRoom	Class#
123	123A	James	555	102-8
123	123A	James	555	104-9
124	123B	Smith	467	209-0
124	123B	Smith	467	102-8

2nd Normal Form Example

Students table

Student#	AdvID	AdvName	AdvRoom
123	123A	James	555
124	123B	Smith	467

Registration table

Student#	Class#
123	102-8
123	104-9
124	209-0
124	102-8

3rd Normal Form Example

Students table:

Student#	AdvID	AdvName	AdvRoom
123	123A	James	555
124	123B	Smith	467

Student table:

Student#	AdvID
123	123A
124	123B

Advisor table:

AdvID	AdvName	AdvRoom
123A	James	555
123B	Smith	467

3rd Normal Form Example Cont.

Students table:

Student#	AdvID
123	123A
124	123B

Registration table:

Student#	Class#
123	102-8
123	104-9
124	209-0
124	102-8

Advisor table:

AdvID	AdvName	AdvRoom
123A	James	555
123B	Smith	467