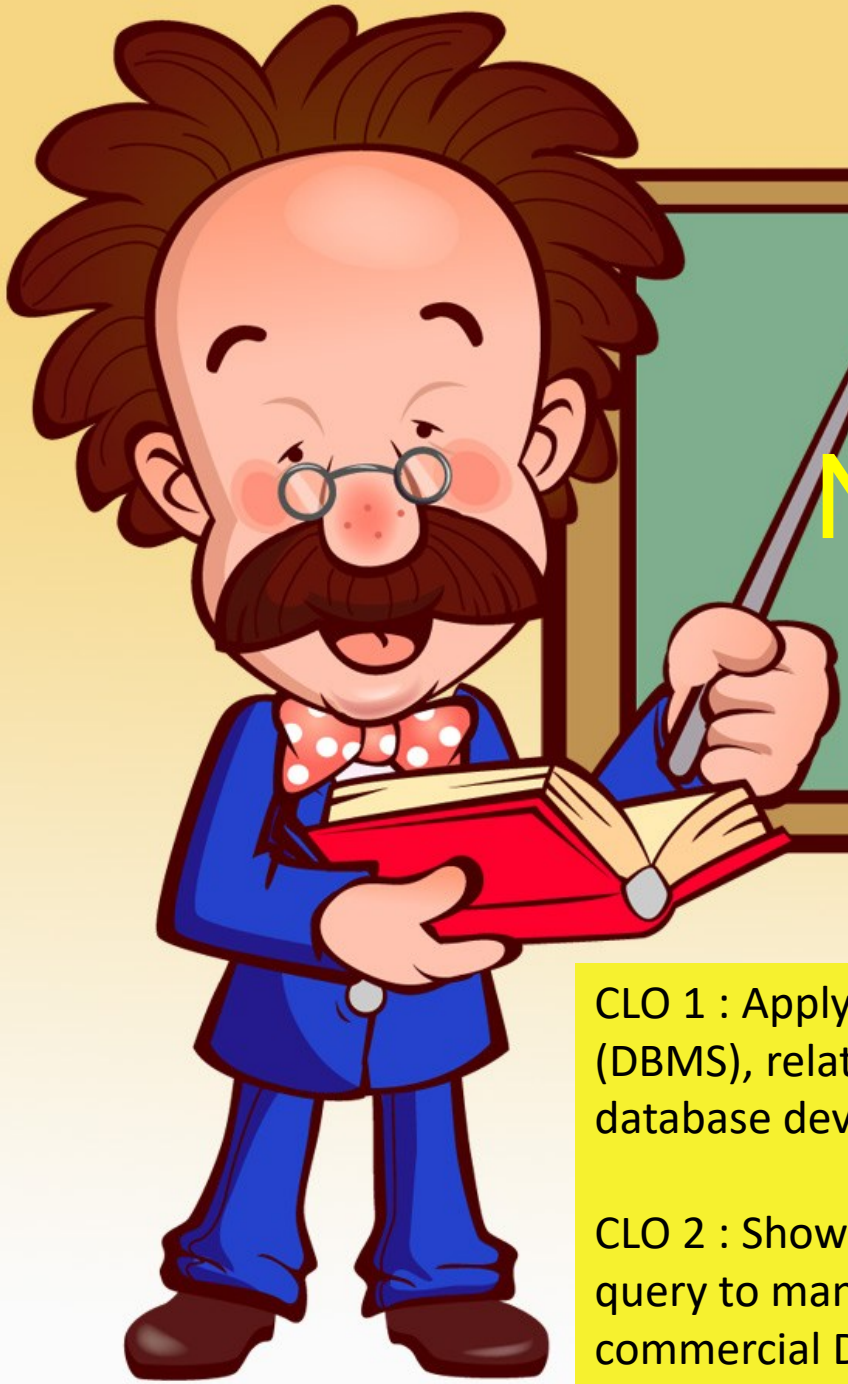


TOPIC 3
NORMALIZATION

CLO 1 : Apply fundamental of Database Management System (DBMS), relational data model and normalization concepts in database development process.

CLO 2 : Show a well-structured database using the database query to manipulate a database with an appropriate commercial Database Management System.

What is Normalization?

- The process of removing redundant data from your tables in to improve storage efficiency, data integrity, and scalability.

- Normalization generally involves splitting existing tables into multiple ones, which must be re-joined or linked each time a query is issued.

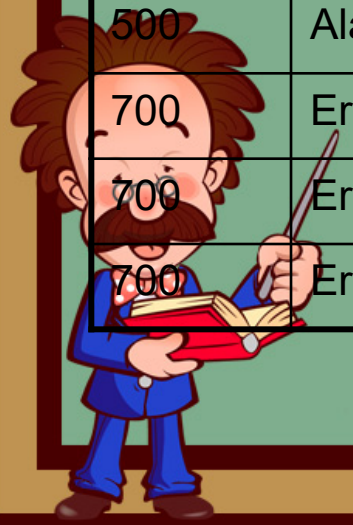
- Normalization is a process for assigning attributes to entities. It reduces data redundancies and helps eliminate the data anomalies.

The Importance of Normalization in Database

- Save typing of repetitive data
- Increase flexibility to query, sort, summarize, and group data (Simpler to manipulate data!)
- Avoid frequent restructuring of tables and other objects to accommodate new data
- Reduce disk space

A Typical Spreadsheet File

Emp No	Employee Name	Time Card No	Time Card Date	Dept No	Dept Name
10	Thomas Arquette	106	11/02/2002	20	Marketing
10	Thomas Arquette	106	11/02/2002	20	Marketing
10	Thomas Arquette	106	11/02/2002	20	Marketing
10	Thomas Arquette	115	11/09/2002	20	Marketing
99	Janice Smitty			10	Accounting
500	Alan Cook	107	11/02/2002	50	Shipping
500	Alan Cook	107	11/02/2002	50	Shipping
700	Ernest Gold	108	11/02/2002	50	Shipping
700	Ernest Gold	116	11/09/2002	50	Shipping
700	Ernest Gold	116	11/09/2002	50	Shipping



Employee, Department, and Time Card Data in Three Tables

Table: Employees

EmpNo	EmpFirstName	EmpLastName	DeptNo
10	Thomas	Arquette	20
500	Alan	Cook	50
700	Ernest	Gold	50
99	Janice	Smitty	10

Table: Departments

DeptNo	DeptName
10	Accounting
20	Marketing
50	Shipping

Table: Time Card Data

TimeCardNo	EmpNo	TimeCardDate
106	10	11/02/2002
107	500	11/02/2002
108	700	11/02/2002
115	10	11/09/2002
116	700	11/09/2002

Anomalies

- Problems that can occur in poorly planned, un-normalized databases where all the data is stored in one table (a flat-file database).

Types of Anomalies

- Insertion anomalies
  - An Insert Anomaly occurs when certain attributes cannot be inserted into the database without the presence of other attributes.
- Deletion anomalies
  - A Delete Anomaly exists when certain attributes are lost because of the deletion of other attributes.

- Update anomalies
  - An Update Anomaly exists when one or more instances of duplicated data is updated, but not all.

Functional Dependencies: Definitions

- ***Multivalued Attributes*** (or ***repeating groups***): non-key attributes or groups of non-key attributes the values of which are not uniquely identified by (directly or indirectly) (not functionally dependent on) the value of the Primary Key (or its part).

STUDENT

Stud_ID	Name	Course_ID	Units
101	Lennon	MSI 250	3.00
101	Lennon	MSI 415	3.00
125	Johnson	MSI 331	3.00

Partial Dependencies

- ***Partial Dependency*** – when an non-key attribute is determined by a part, but not the whole, of a **COMPOSITE** primary key.

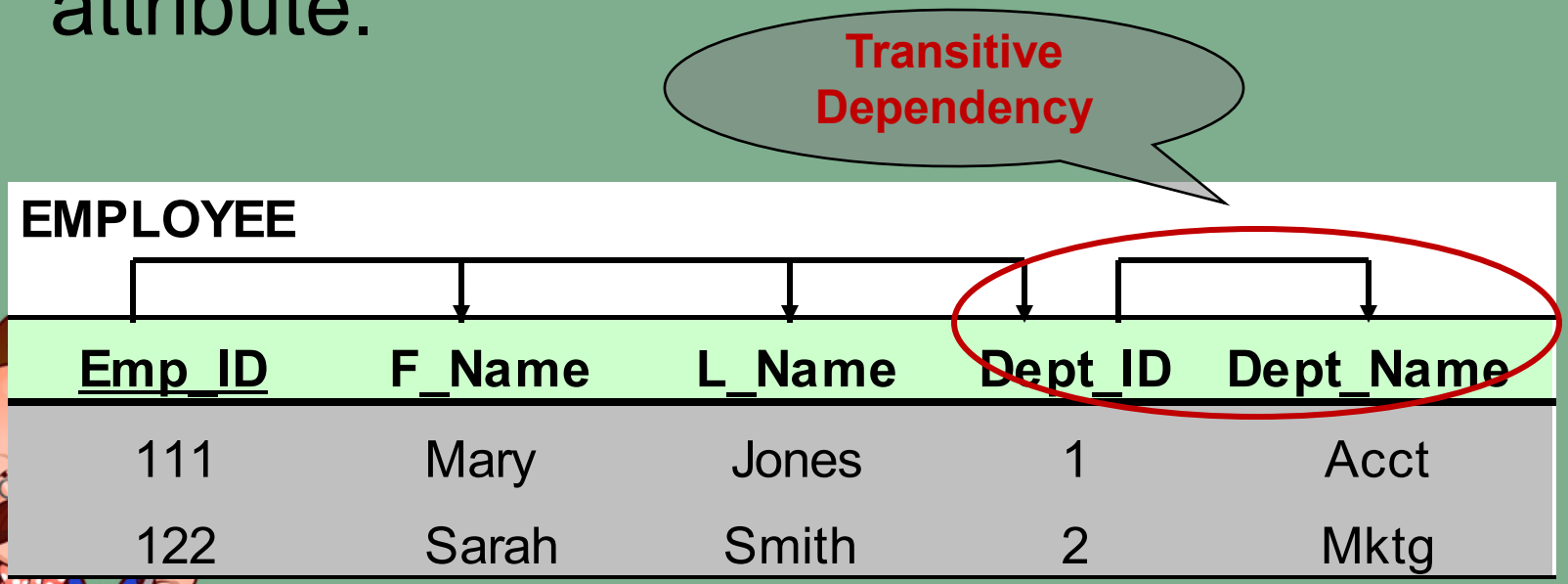
CUSTOMER

Partial Dependency

Cust_ID	Name	Order_ID
101	AT&T	1234
101	AT&T	156
125	Cisco	1250

Transitive Dependencies

- ***Transitive Dependency*** – when a non-key attribute determines another non-key attribute.



Normal Forms : Review

- Unnormalized – There are multivalued attributes or repeating groups
- 1 NF – No multivalued attributes or repeating groups.
- 2 NF – 1 NF plus no partial dependencies
- 3 NF – 2 NF plus no transitive dependencies

- All attributes depend on the key, the whole key and nothing but the key.
- 1NF - Keys and no repeating groups
- 2NF - No partial dependencies
- 3NF - All determinants are candidate keys / no transitive dependencies
- BCNF - is a higher version of the Third Normal Form. This form deal with certain type of anomaly that is not handled by 3NF. A 3NF table which does not have multiple overlapping candidate keys is said to be in BCNF

UNF

1NF

2NF

3NF

BCNF

Define primary key/
remove repeating group

Remove partial dependencies

Remove transitive dependencies

Summary !!!!

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.**
    - **Addition anomalies.**
    - **Deletion anomalies.**

- **Conversion to First Normal Form**
  - **A relational table must not contain repeating groups.**
  - **Repeating groups can be eliminated by adding the appropriate entry in at least 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

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

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

First Normal Form (1NF)

- **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.**

- **Dependency Diagram**
  - **The primary key components are bold, underlined, and shaded in a different color.**
  - **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.**

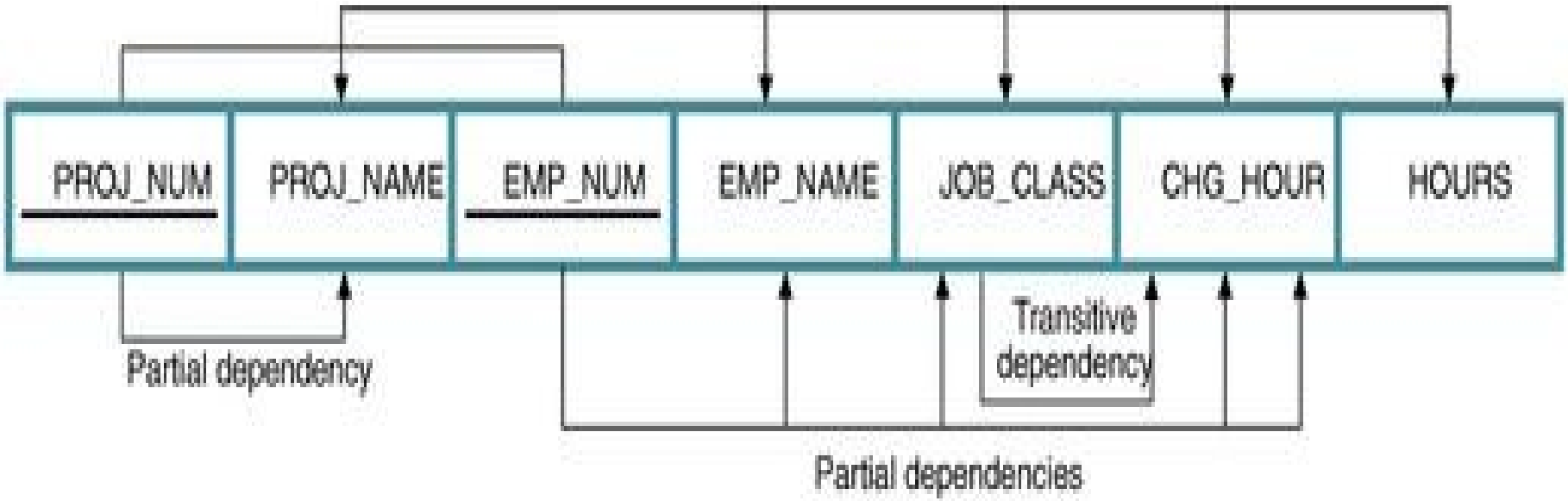


FIGURE 5.4 A DEPENDENCY DIAGRAM: FIRST NORMAL FORM (1NF)

Second Normal Form (2NF)

- **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

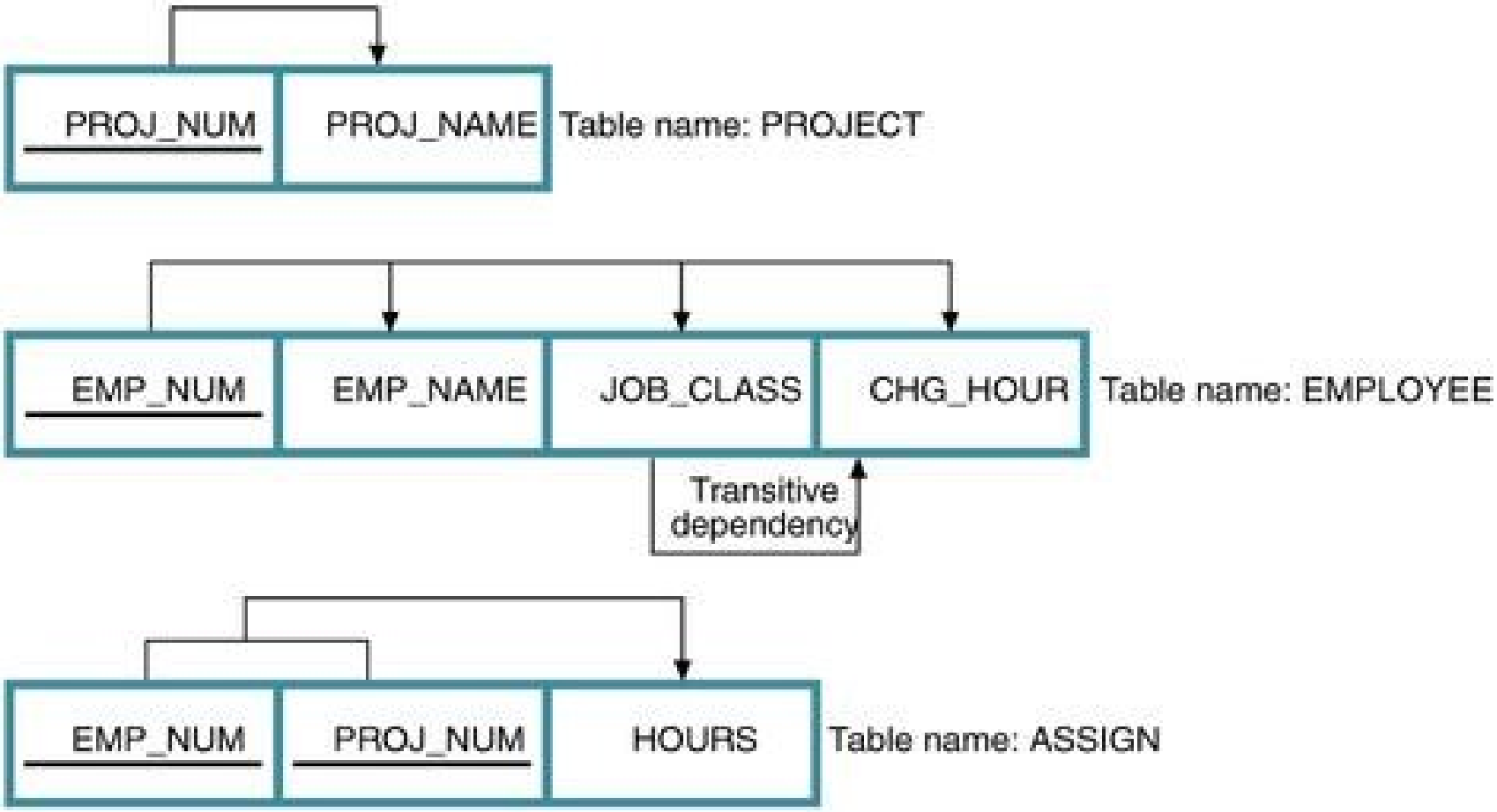


FIGURE 5.5 SECOND NORMAL FORM (2NF) CONVERSION RESULTS

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 non-key attributes.)

Third Normal Form (3NF)

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

PROJECT (PROJ_NUM, PROJ_NAME)

ASSIGN (PROJ_NUM, EMP_NUM, HOURS)

EMPLOYEE (EMP_NUM, EMP_NAME, JOB_CLASS)

JOB (JOB_CLASS, CHG_HOUR)

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

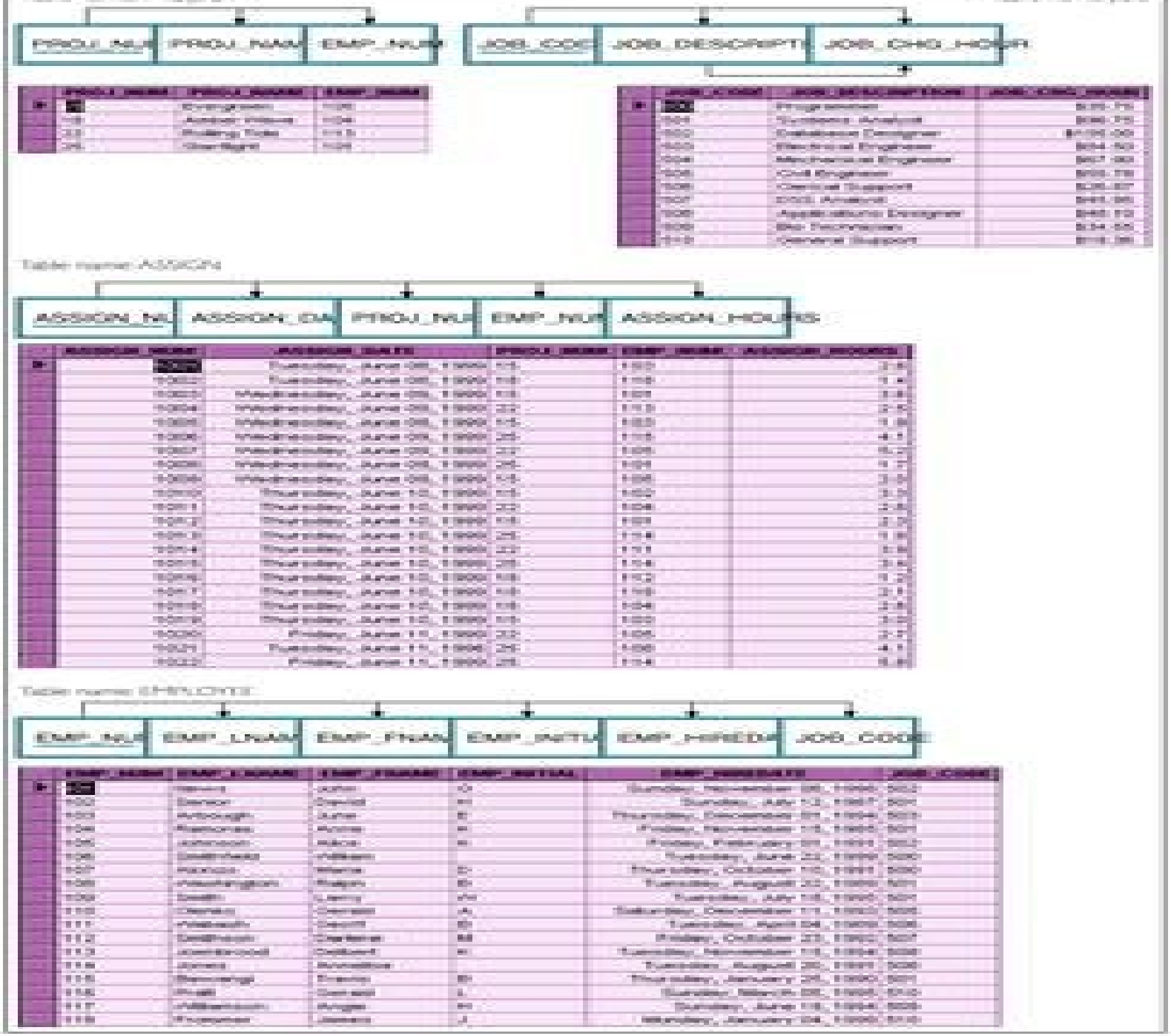


FIGURE 5.6 THE COMPLETED DATABASE

Define BCNF

- **Boyce Codd Normal Form (BCNF)** is considered a special condition of third Normal form.
- **A table is in BCNF if every determinant is a candidate key.**
- A table can be in 3NF but not in BCNF. This occurs when a non key attribute is a determinant of a key attribute.
- The dependency diagram may look like the one below