

4 Database Tables and Normalization

- ❖ **Normalization** is a process for assigning attributes to entities.
- ❖ 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)
 - ◆ Fourth normal form (4NF)
- ❖ Highest normalization is not always desirable (Joins vs Redundancy)

Is a Relation a File? A Table? What's a row?

Conceptual Model ERD	Relational Model	"User"	"Programmer"
Entity Set	Relation	Table	File
Entity instance	Tuple (row)	Row	Record
Attribute	Attribute (Field)	Column	Field, data element
Identifier	Primary Key	Key, ID Code	Record Key
Relationships	Foreign Key	Linking Value	Master Records

4 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.

4 TABLE 5.1 A Table Whose Format Matches the Report Format

Proj. Num.	Proj. Name	Emp. Num.	Emp. Name	Job Class.	Chg. Hour	Rate
10	Swingsets	100	John E. Brinkhoff	Elec. Engineer	\$94.00	22.8
		101	John G. Fries	Database Designer	\$100.00	19.8
		102	Alan K. Johnson	Electronics Designer	\$100.00	36.7
		103	William D. Mitchell	Programmer	\$30.75	12.6
		104	David H. Jones	Systems Analyst	\$96.75	22.8
Subtotal						\$43,246.76
11	Archie West	114	Arvides Jones	Applications Designer	\$96.75	24.6
		115	James J. Fryman	General Support	\$18.26	45.2
		116	Anne K. Ramirez	Systems Analyst	\$96.75	32.4
		117	Debra M. Johnson	DB Analyst	\$40.00	44.0
Subtotal						\$1,071.47
12	Paving Job	100	Alan K. Johnson	Database Designer	\$100.00	41.7
		104	Anne K. Ramirez	Systems Analyst	\$96.75	48.4
		111	Debra M. Johnson	Applications Designer	\$96.75	32.4
		111	David H. Jones	General Support	\$26.67	22.0
		106	William D. Mitchell	Programmer	\$30.75	12.6
Subtotal						\$1,248.16
13	Site Work	100	John G. Fries	Programmer	\$30.75	24.6
		112	John E. Brinkhoff	Systems Analyst	\$96.75	40.8
		101	John G. Fries	Database Designer	\$100.00	26.1
		114	Arvides Jones	Applications Designer	\$96.75	31.1
		106	Ralph B. Hurlingham	Systems Analyst	\$96.75	22.8
		116	James J. Fryman	General Support	\$18.26	39.2
Subtotal						\$1,289.21
Total						\$46,845.67

Figure 5.1: A Table Whose Structure Matches the Report Format

4 Problems with the Figure 5.1

- ❖ The project number is intended to be a primary key, but it contains blanks.
- ❖ The table displays data redundancies.
- ❖ The table entries invite data inconsistencies.
- ❖ The data redundancies yield the following anomalies:
 - ◆ Update anomalies.
 - don't have to change values in multiple places
 - ◆ Addition anomalies.
 - have all data you need to add instance
 - ◆ Deletion anomalies.
 - don't lose info by deleting instance

1NF = First Normal Form

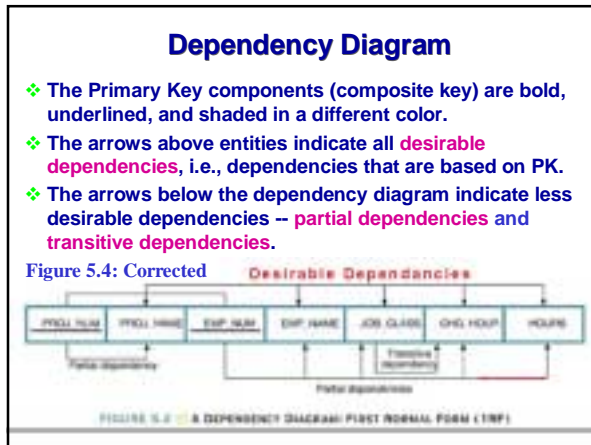
- ❖ Definition: Tabular format in which:
 - ◆ All the key attributes are defined.
 - ◆ There are no repeating groups in the table.
 - ◆ All attributes are dependent on primary key.
- ❖ Conversion to First Normal Form
 1. Fill in all the blanks in a table of data.
 2. Remove any duplicate records.
 - Exact record matches are redundant

Robert Laurie - 7

Figure 5.3: Data Organization: First Normal Form

EMP_NUM	EMP_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
101	John E. Adams	101	John E. Adams	Genl. Support	\$18.50	35.7
102	John E. Adams	102	John E. Adams	Genl. Support	\$18.50	35.7
103	John E. Adams	103	John E. Adams	Genl. Support	\$18.50	35.7
104	John E. Adams	104	John E. Adams	Genl. Support	\$18.50	35.7
105	John E. Adams	105	John E. Adams	Genl. Support	\$18.50	35.7
106	John E. Adams	106	John E. Adams	Genl. Support	\$18.50	35.7
107	John E. Adams	107	John E. Adams	Genl. Support	\$18.50	35.7
108	John E. Adams	108	John E. Adams	Genl. Support	\$18.50	35.7
109	John E. Adams	109	John E. Adams	Genl. Support	\$18.50	35.7
110	John E. Adams	110	John E. Adams	Genl. Support	\$18.50	35.7
111	John E. Adams	111	John E. Adams	Genl. Support	\$18.50	35.7
112	John E. Adams	112	John E. Adams	Genl. Support	\$18.50	35.7
113	John E. Adams	113	John E. Adams	Genl. Support	\$18.50	35.7
114	John E. Adams	114	John E. Adams	Genl. Support	\$18.50	35.7
115	John E. Adams	115	John E. Adams	Genl. Support	\$18.50	35.7
116	John E. Adams	116	John E. Adams	Genl. Support	\$18.50	35.7
117	John E. Adams	117	John E. Adams	Genl. Support	\$18.50	35.7
118	John E. Adams	118	John E. Adams	Genl. Support	\$18.50	35.7
119	John E. Adams	119	John E. Adams	Genl. Support	\$18.50	35.7
120	John E. Adams	120	John E. Adams	Genl. Support	\$18.50	35.7

FIGURE 5.3 DATA ORGANIZATION: FIRST NORMAL FORM



2NF = Second Normal Form

- ❖ Definition: Tabular format in which:
 - ◆ It is in 1NF and
 - ◆ No Partial Dependencies exist
 - No attribute is dependent on only a portion of the primary key.
 - ◆ Transitive Dependencies are permitted
 - One or more attributes may be functionally dependent on nonkey attributes.
- ❖ Conversion to Second Normal Form
 1. Be in 1NF
 2. Remove partial dependencies.

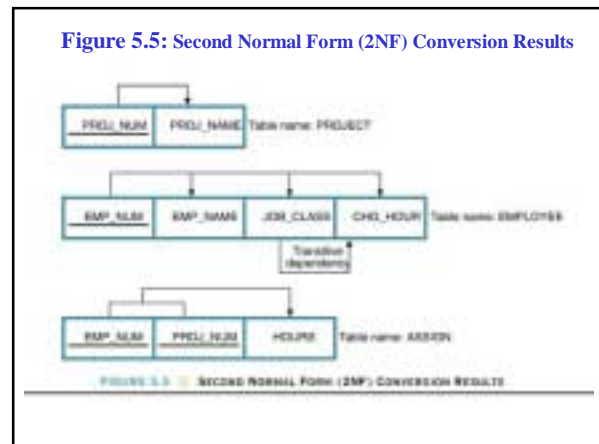
Robert Laurie - 10

Removing partial dependencies

- ❖ Goal: Every nonkey attribute is fully (not partially) functionally dependent on (determined by) the primary key.
- ❖ Method:
 1. Write key components on separate lines
 2. Write original key on the last line
 3. Write 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)

Robert Laurie - 11



4 3NF = Third Normal Form

- ❖ Definition: Tabular format in which:
 - ◆ It is in 2NF and
 - ◆ No Transitive Dependencies exist
 - One or more attributes may not be functionally dependent on nonkey attributes.
- ❖ Conversion to Third Normal Form
 1. Be in 2NF
 2. Remove Transitive Dependencies.

Robert Laurie - 13

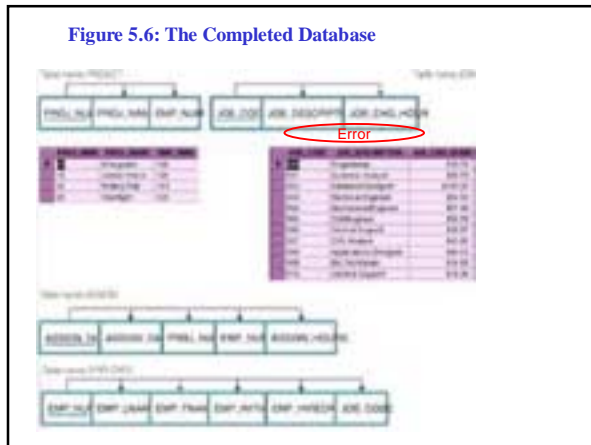
4 Remove Transitive Dependencies

- ❖ Goal: No non-key attribute determines another non-key attribute.
- ❖ Method:
 1. Create an additional table to remove the 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)
    
```

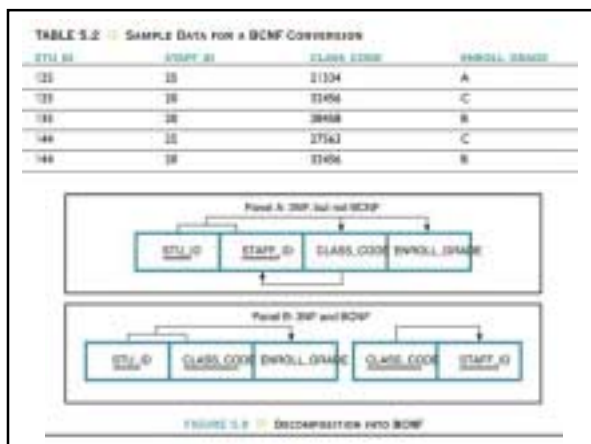
Robert Laurie - 14



4 Boyce-Codd Normal Form (BCNF)

- ❖ A table is in **Boyce-Codd normal form (BCNF)** if every determinant in the table is a candidate key.
 - ◆ A determinant is any attribute whose value determines other values with a row.
- ❖ If a table contains only one candidate key, the 3NF and the BCNF are equivalent.
- ❖ BCNF is a special case of 3NF.

Robert Laurie - 16



4 4NF = Fourth Normal Form

- ❖ Definition: Tabular format in which:
 - ◆ It is in 3NF and
 - ◆ No multiple sets of multivalued dependencies
- ❖ Conversion to Second Normal Form
 1. Be in 3NF
 2. Remove multiple sets of multivalued dependencies.
- ❖ Generally 3NF is enough normalization for business applications.
 - ◆ 4NF not done.

Robert Laurie - 18

Denormalization

- ❖ Normalization is only one of many database design goals.
- ❖ Normalized (decomposed) tables require additional processing, reducing system speed.
- ❖ Normalization purity is often difficult to sustain in the modern database environment. The conflict between design efficiency, information requirements, and processing speed are often resolved through compromises that include denormalization.

Robert Laurie - 19

Integrity constraints

- ❖ **Domain constraints**
 - ◆ Domain name, meaning, data type, size, allowable values/range
 - ◆ Part of "data dictionary"
- ❖ **Entity integrity**
 - ◆ Primary key: UNIQUE & valid, non-null
- ❖ **Referential integrity**
 - ◆ consistency in rows between relations
 - ◆ foreign key = primary key (or null)

Robert Laurie - 20

Database Design and Normalization Construction Company Example

- ❖ 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.
- ❖ Each employee has a (single) primary job classification. This job classification determines the hourly billing rate.
- ❖ Many employees can have the same job classification.

Robert Laurie - 21

Normalization and Database Design

- ❖ **Two Initial Entities:**
 PROJECT (PROJ_NUM, PROJ_NAME)
 EMPLOYEE (EMP_NUM, EMP_LNAME, EMP_FNAME, EMP_INITIAL, JOB_DESCRIPTION, JOB_CHG_HOUR)
- ❖ **What Normal form are these in?**

Robert Laurie - 22

Normalization and Database Design

- ❖ **PROJECT is in 3NF**
 PROJECT (PROJ_NUM, PROJ_NAME)
- ❖ **EMPLOYEE is in 2NF**
 - ◆ Remove transitive dependency to get following two entities
 - EMPLOYEE (EMP_NUM, EMP_LNAME, EMP_FNAME, EMP_INITIAL, JOB_CODE)
 - JOB (JOB_CODE, JOB_DESCRIPTION, JOB_CHG_HOUR)

Robert Laurie - 23

The Modified ERD For A Contracting Company

Robert Laurie - 24

Figure 5.11

Slide Set 4: Normalization

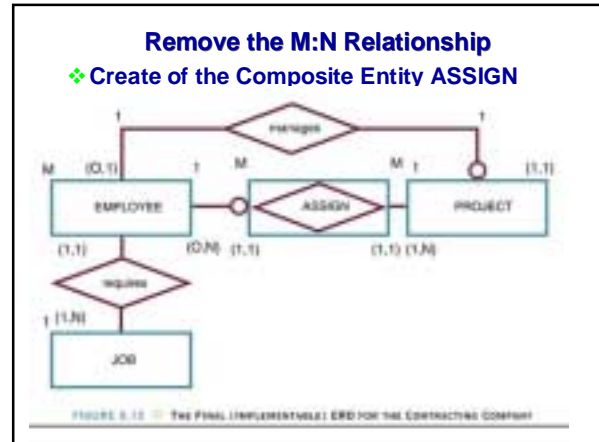
Normalization and Database Design

4

- ❖ Attribute **ASSIGN_HOUR** is assigned to the composite entity **ASSIGN**.
- ❖ “Manages” relationship is created between **EMPLOYEE** and **PROJECT**.

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, PROJ_NUM, EMP_NUM, ASSIGN_HOURS)

Robert Laurie - 25



The Relational Schema For The Contracting Company

4

FIGURE 5.13 The Relational Schema for the Contracting Company

Figure 5.13

Robert Laurie - 27

In Class Problem:

4

- ❖ Review tabular presentation of information in the following slide
- ❖ List & discuss the kinds of anomalies present in each table (see textbook discussion & prior example)
- ❖ Create 1NF relation; write in shorthand notation; identify candidate keys
- ❖ Analyze functional dependencies
- ❖ Write final form of 2NF relation(s)
- ❖ What is required for 3NF?

Robert Laurie - 28

PATIENT NO.	NAME	ADDRESS	DATE OF VISIT	DOCTOR	SYMPTOM
639147-0	Michael Smith	329 Fourth St.	3-24-94	Ryan	Fever
			7-6-94	Nelson	Sore throat
			8-16-94	Ryan	Cold
89043-0	Fred Jones	932 Third St.	2-20-94	Wood	Flu
			6-15-94	Nelson	Sore throat
			7-02-94	Ryan	Rash
90765-0	Sarah Smith	329 Fourth St.	3-24-94	Ryan	Fever
			10-28-94	Ryan	Bronchitis