Ch 4

# Database Management System

Sumayyea Salahuddin (Lecturer)
Dept. of Computer Systems Eng.
UET Peshawar

# Overview

- Case Study: SoftCo

- Normalization
  - What is Normalization?
  - Why it is needed?
  - Normal Forms
    1. Conversion to first normal form
    2. Conversion to second normal form
    3. Conversion to third normal form

# Case Study: SoftCo

- Example: A company that manages building projects
  - Charges its clients by billing hours spent on each contract
  - Hourly billing rate is dependent on employee's position
  - Periodically, a report is generated that contains information displayed in Table 5.1

# SoftCo – 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 | $3,135.70 |
| | | 112 | Darlene M. Smithson | DSS Analyst | $45.95 | 44.0 | $2,021.80 |
| | | | | **Subtotal** | | | **$7,171.47** |
| 22 | Rolling Tide | 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 | Starlight | 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** |

* Indicates project leader

# SoftCo – Table Representation

FIGURE 5.1 A TABLE IN THE REPORT FORMAT

Table name: RPT_FORMAT — Database name: Ch05_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 |

# SoftCo – Table Issues

- PK(PROJ_NUM) contains nulls.
- Data inconsistencies: JOB_CLASS value "Elect. Engineer" might be entered as "Elect Eng" or "El. Eng." or "EE"
- Data redundancies which yield data anomalies:
  - Update Anomalies: Modifying the JOB_CLASS for employee number 105 requires (potentially) many alternations, one for each EMP_NUM = 105
  - Insertion Anomalies: Just to complete row definition, a new employee must be assigned to a project. If the employee is not yet assigned, a phantom project must be created to complete the employee data entry.
  - Deletion Anomalies: Suppose that only one employee is associated with a given project. If that employee leaves the company and the employee data are deleted, the project information will be deleted. To prevent the loss of the project information, a fictitious employee must be created just to save the project information.

# Normalization

- Normalization
  - Process for evaluating and correcting table structure to minimize data redundancies
    - **Helps eliminate data anomalies**
  - Works through a series of stages called normal forms
    - Normal Form (1NF)
    - Second Normal Form (2NF)
    - Third Normal Form (3NF)
  - 2NF is better than 1NF; 3NF is better than 2NF
  - For most business database design purposes, 3NF is highest we need to go in the normalization process
  - Highest level of normalization is not always most desirable

# Conversion to First Normal Form (1NF)

- **Repeating Group**
  - Derives its name from the fact that a group of multiple (related) entries can exist for any *single* key attribute occurrence
- Relational table must not contain repeating groups
- Normalizing the table structure will reduce these data redundancies
- Normalization is three-step procedure

# Step 1: Eliminate the Repeating Groups

- Present data in a tabular format, where each cells has a single value and there are no repeating groups
- Eliminate repeating groups by eliminating null, making sure that each repeating group attributes contain an appropriate data value

# Data Organization: First Normal Form (1NF)

FIGURE 5.2 DATA ORGANIZATION: FIRST NORMAL FORM

**Table name: DATA_ORG_1NF** **Database name: Ch05_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 |

# Step 2: Identify the Primary Key

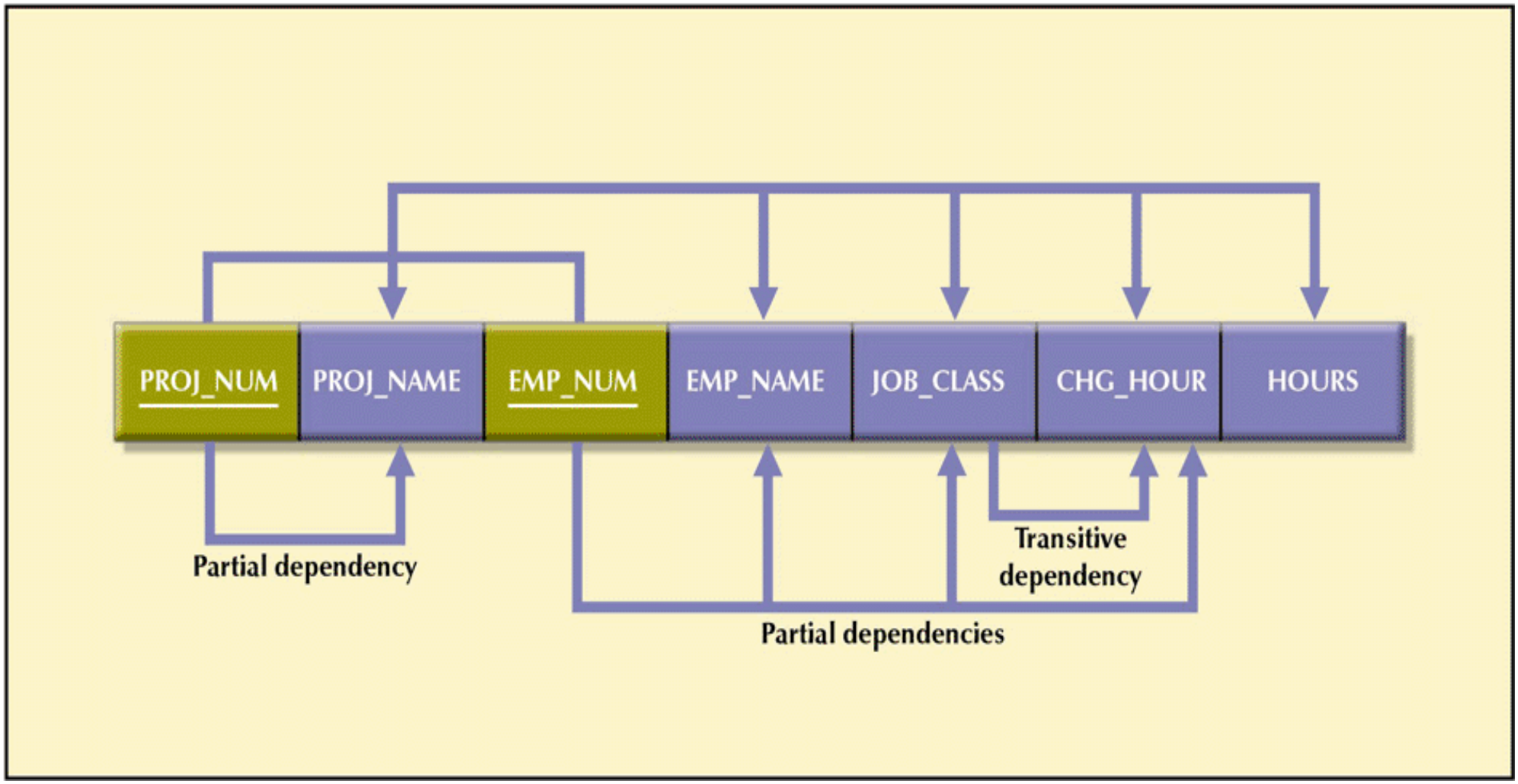
- Primary key must uniquely identify attribute value
- New key must be composed

# Step 3: Identify All Dependencies

- Dependencies can be depicted with the help of a diagram
- Dependency diagram:
  - Depicts all dependencies found within a given table structure
  - Helpful in getting bird's eye view of all relationships among table's attributes
  - Use makes it much less likely that an important dependency will be overlooked

# A Dependency Diagram: First Normal Form (1NF)

FIGURE 5.3 A DEPENDENCY DIAGRAM: FIRST NORMAL FORM (1NF)



# First Normal Form (1NF)

- Tabular format in which:
  - All key attributes are defined
  - There are no repeating groups in the table
  - All attributes are dependent on primary key
- All relational tables satisfy 1NF requirements
- Some tables contain partial dependencies
  - Dependencies based on only part of the primary key
  - Sometimes used for performance reasons, but should be used with caution
  - Still subject to data redundancies

# Task 1

- Table 4-3 contains sample data for **parts** and for **vendors** who supply those parts. In discussing these data with users, we find that part numbers (but not descriptions) uniquely identify parts and that vendor names uniquely identify vendors.
  1) Convert this table to a relation (named **PART SUPPLIER**) in **1NF.** Illustrate the relation with the sample data in the table.
  2) List the functional dependencies in PART SUPPLIER and identify a candidate key.
  3) For the relation PART SUPPLIER, identify each of the following: an insert anomaly, a delete anomaly, and a modification anomaly.

**TABLE 4-3** Sample Data for Parts and Vendors

| Part No | Description | Vendor Name | Address | Unit Cost |
|---|---|---|---|---|
| 1234 | Logic chip | Fast Chips | Cupertino | 10.00 |
| | | Smart Chips | Phoenix | 8.00 |
| 5678 | Memory chip | Fast Chips | Cupertino | 3.00 |
| | | Quality Chips | Austin | 2.00 |
| | | Smart Chips | Phoenix | 5.00 |

# Conversion to Second Normal Form (2NF)

- Relational database design can be improved by converting the database into second normal form (2NF)

- Two steps

# Step 1: Identify All Key Components

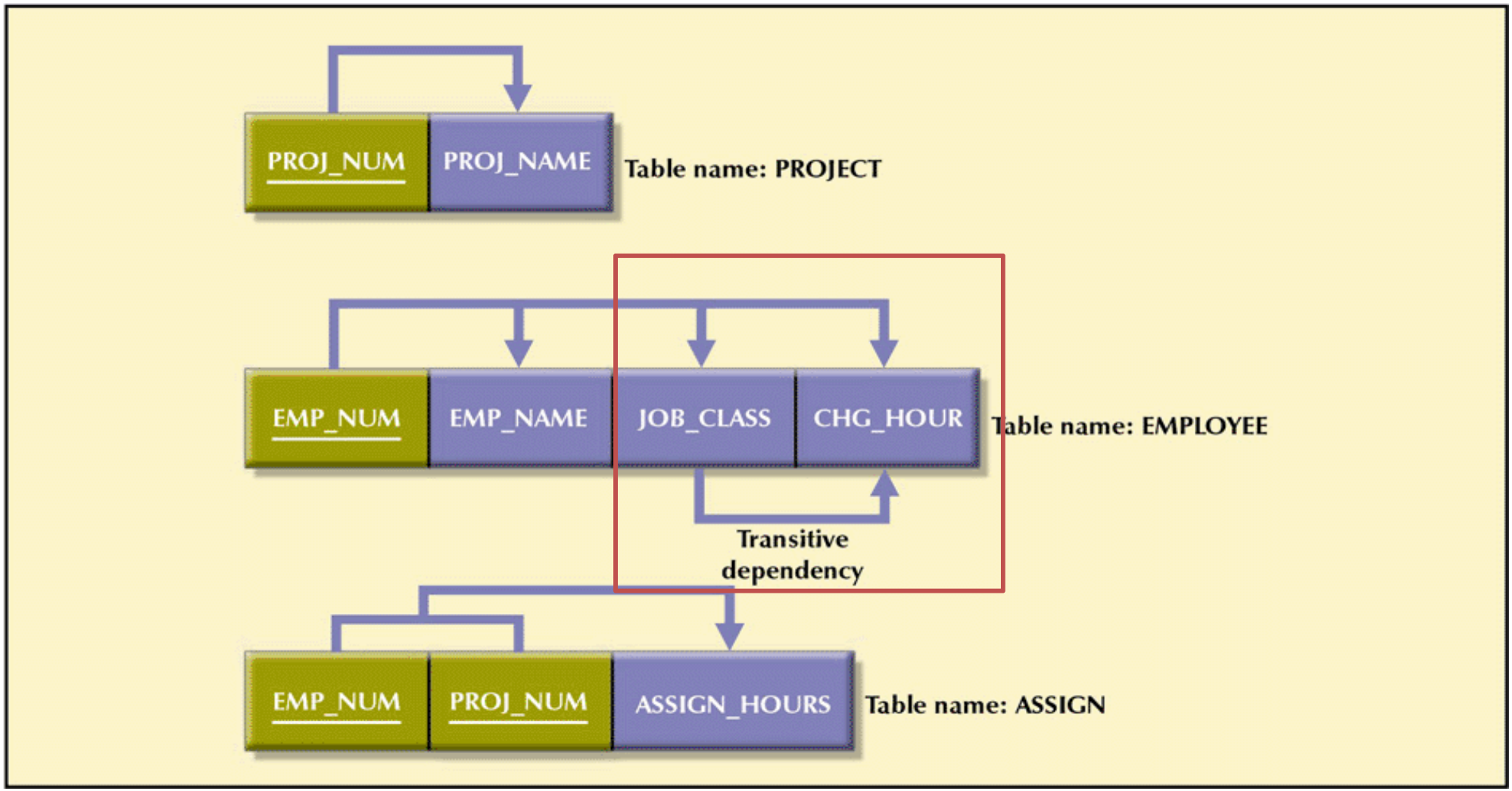
- Write each key component on separate line, and then write the original (composite) key on the last line
- Each component will become the key in a new table

# Step 2: Identify the Dependent Attributes

- Determine which attributes are dependent on which other attributes
- At this point, most anomalies have been eliminated

# Second Normal Form (2NF) Conversion Results

FIGURE 5.4 SECOND NORMAL FORM (2NF) CONVERSION RESULTS



# Second Normal Form (2NF)

- Table is in second normal form (2NF) if:
  - It is in 1NF and
  - It includes no partial dependencies
    - No attribute is dependent on only a portion of the primary key

# Task 2

- Develop a set of 2NF relations from dependency diagram of **PART Supplier** relation from Task 1.

# Conversion to Third Normal Form (3NF)

- Data anomalies created are easily eliminated by completing three steps

# Step 1: Identify Each New Determinant

- For every transitive dependency, write its determinant as a PK for a new table
  - Determinant
    - Any attribute whose value determines other values within a row

# Step 2: Identify the Dependent Attributes

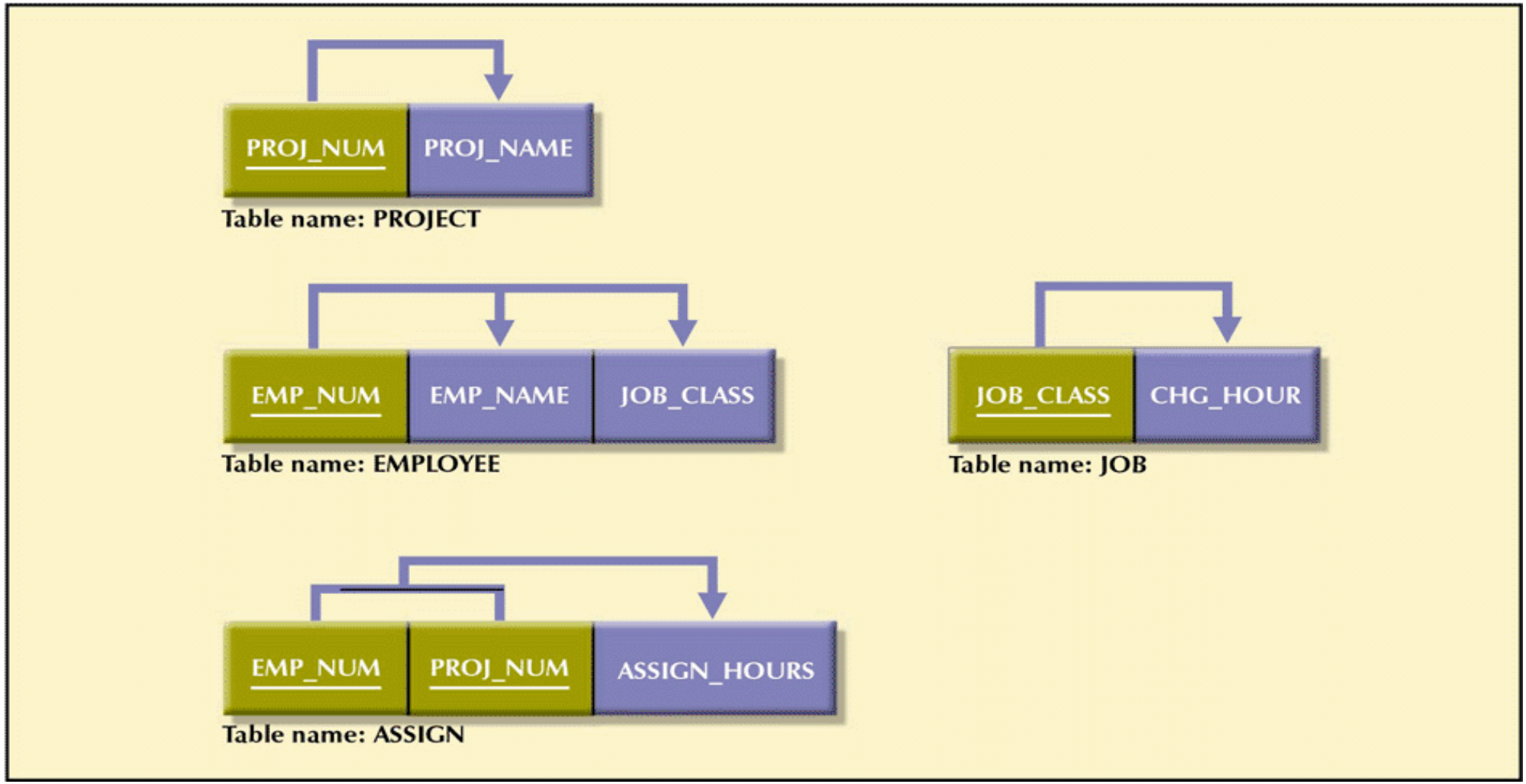
- Identify the attributes dependent on each determinant identified in Step 1 and identify the dependency
- Name the table to reflect its contents and function

# Step 3: Remove the Dependent Attributes from Transitive Dependencies

- Eliminate all dependent attributes in transitive relationship(s) from each table that has such a transitive relationship
- Draw a new dependency diagram to show all tables defined in Steps 1-3
- Check new tables and modified tables from Step 3 to make sure that each has a determinant and does not contain inappropriate dependencies

# Third Normal Form (3NF) Conversion Results

FIGURE 5.5 THIRD NORMAL FORM (3NF) CONVERSION RESULTS



# Third Normal Form (3NF)

- A table is in third normal form (3NF) if:
  - It is in 2NF and
  - It contains no transitive dependencies

# Task 3

- Given the relations in task 2, convert all the relations in 3NF.

# Improving the Design

- Table structures are cleaned up to eliminate the troublesome initial partial and transitive dependencies
- Normalization cannot, by itself, be relied on to make good designs
- It is valuable because its use helps eliminate data redundancies

# Improving the Design (cont…)

- The following changes were made:
  - PK assignment
  - Naming conventions
  - Attribute atomicity
  - Adding attributes
  - Adding relationships
  - Refining PKs
  - Maintaining historical accuracy
  - Using derived attributes

# The Completed Database

FIGURE 5.6 THE COMPLETED DATABASE


**Database name: Ch05_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: 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 |

# The Completed Database (cont…)

FIGURE 5.6 THE COMPLETED DATABASE (CONTINUED)

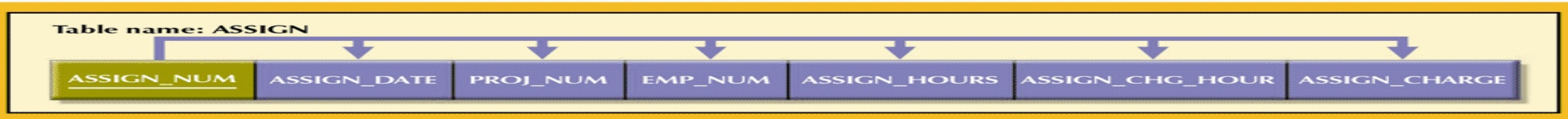



**Table name: ASSIGN** **Database name: Ch05_ConstructCo**

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


**Table name: EMPLOYEE**

| EMP_NUM | EMP_LNAME | EMP_FNAME | EMP_INITIAL | EMP_HIREDATE | JOB_CODE |
|---|---|---|---|---|---|
| 101 | News | John | G | 08-Nov-98 | 502 |
| 102 | Senior | David | H | 12-Jul-87 | 501 |
| 103 | Arbough | June | E | 01-Dec-94 | 503 |
| 104 | Ramoras | Anne | K | 15-Nov-85 | 501 |
| 105 | Johnson | Alice | K | 01-Feb-91 | 502 |
| 106 | Smithfield | William | | 22-Jun-03 | 500 |
| 107 | Alonzo | Maria | D | 10-Oct-91 | 500 |
| 108 | Washington | Ralph | B | 22-Aug-89 | 501 |
| 109 | Smith | Larry | W | 18-Jul-95 | 501 |
| 110 | Olenko | Gerald | A | 11-Dec-93 | 505 |
| 111 | Wabash | Geoff | B | 04-Apr-89 | 506 |
| 112 | Smithson | Darlene | M | 23-Oct-92 | 507 |
| 113 | Joenbrood | 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-04 | 510 |

# Summary

- Discussed concept of normalization
- Discussed its three normal forms
- Covered case study and tasks