Normalization of Database Tables

관계형 데이터베이스의 설계에서 중복을 최소화하게 데이터를 구조화하는 프로세스
이상이 있는 관계를 재구성하여 작고 잘 조직된 관계를 생성

갱신이상**:** 보험기록에 에이젼트 번호와 이름이 중복해서 들어가 있을 경우 에이젼트가 결혼하여 **Last Name**이 바뀌면 모든 레코드를 수정해야 함**.**

삽입 이상**:** 교수와 강의라는 테이블에 교수 **ID,** 교수 이름**,** 교수 고용일자**,** 강의 코드가 있다고 했을 때 새로운 교수를 고용하면 그가 맡은 강의가 없을 수 있어 강의코드가 **NULL**이 됨**.**

삭제이상**:** 위의 교수와 강의 테이블에서 강의를 일시적으로 중단하고자 할 때 교수 자체가 사라지는 현상이 발생함**.**

1NF: Eliminate the repeating groups, Identify PK and all dependencies

2NF: Minimize partial dependencies

3NF: Minimize transitive dependencies

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Database Tables and Normalization

- Normalization
  - Process for evaluating and correcting table structures to minimize data redundancies
    - helps eliminate data anomalies
  - Works through a series of stages called normal forms:
    - First normal form (1NF)
    - Second normal form (2NF)
    - Third normal form (3NF)
    - ...

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Database Tables and Normalization

- 2NF is better than 1NF; 3NF is better than 2NF
- For most business database design purposes, 3NF is the highest we need to go in the normalization process
- Highest level of normalization is not always most desirable ***Accuracy VS Speed***

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

The Need for Normalization

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

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

The Need for Normalization (2)

- Structure of data set in the previous figure does not handle data very well
  - PROJ_NUM is apparently intended to be a primary key, or at least a part of a primary key, but it contains nulls
  - Data inconsistencies (e.g.: Database Designer, DB Designer
  - Data redundancies (e.g.: update, insertion, and deletion anomalies) ***Possible redundant inputs***

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Conversion to First Normal Form (1NF)

- Step 1: Eliminate the Repeating Groups
- Step 2: Identify the Primary Key
- Step 3: Identify all Dependencies

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

1NF Step 1: Eliminate Repeating Groups

- Present data in a tabular format, where each cell has a single value and there are no repeating groups
- Eliminate repeating groups by eliminating nulls, making sure that each repeating group attribute contains an appropriate data value

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

1NF Data Organization

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

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

1NF Step 2: Identify the Primary Key

- Primary key must uniquely identify attribute value

  PROJ_NUM & EMP_NUM

- New key must be composed

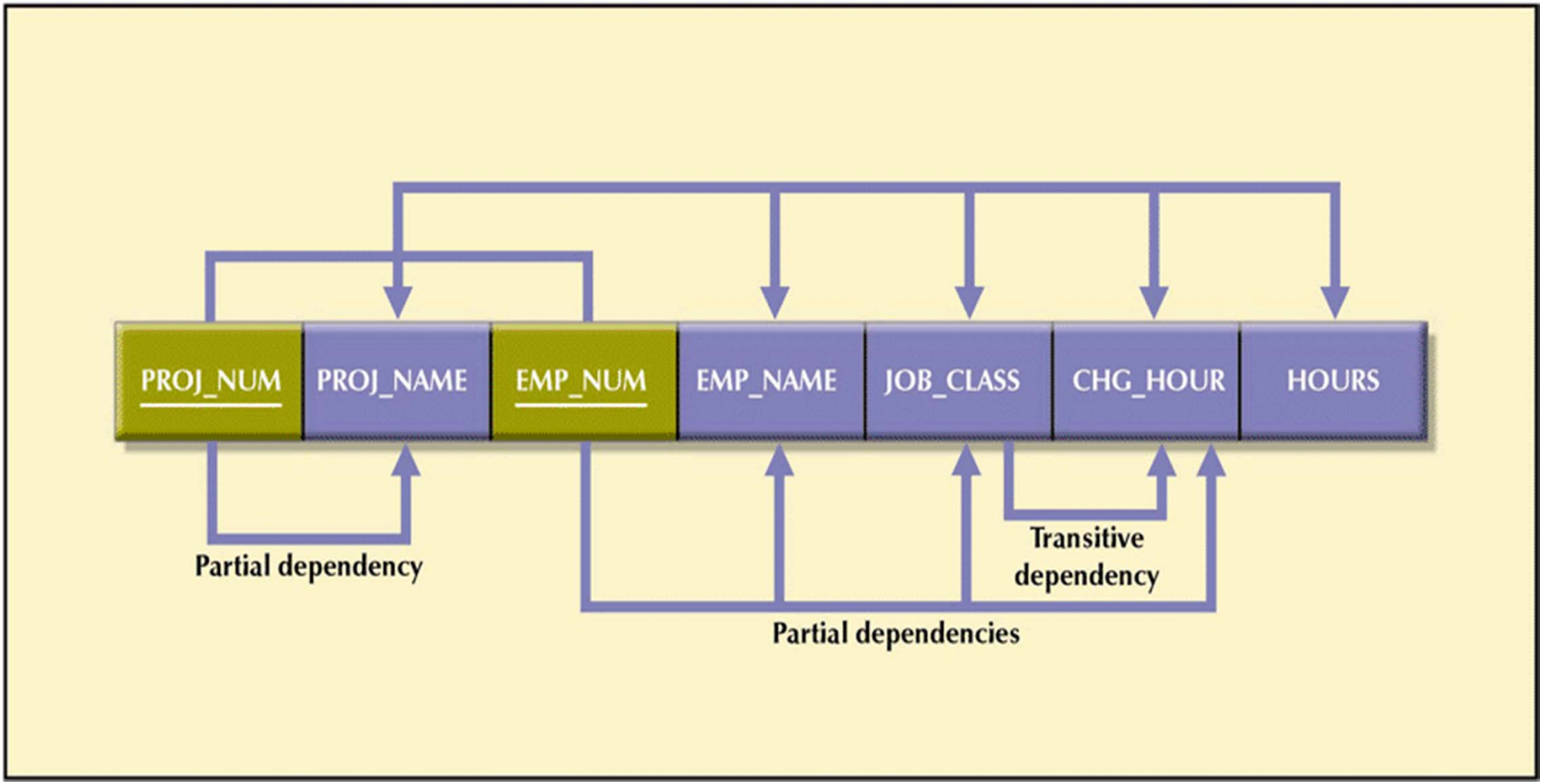
From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

1NF 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
  - Partial dependencies: based on only a part of a composite primary key
  - Transitive dependencies: dependency of one nonprime attribute on another nonprime attribute.
  - Partial and transitive dependencies are not desirable.

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

A Dependency Diagram: 1NF



From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

First Normal Form: Summary

- 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

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Conversion to Second Normal Form

- Relational database design can be improved by converting the database into second normal form (2NF) ***To minimize partial dependencies***

- Two steps:
  - 2NF Step 1: Identify All Key Components
  - 2NF Step 2: Identify the Dependent Attributes

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

2NF Step 1: Identify All Key Components

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

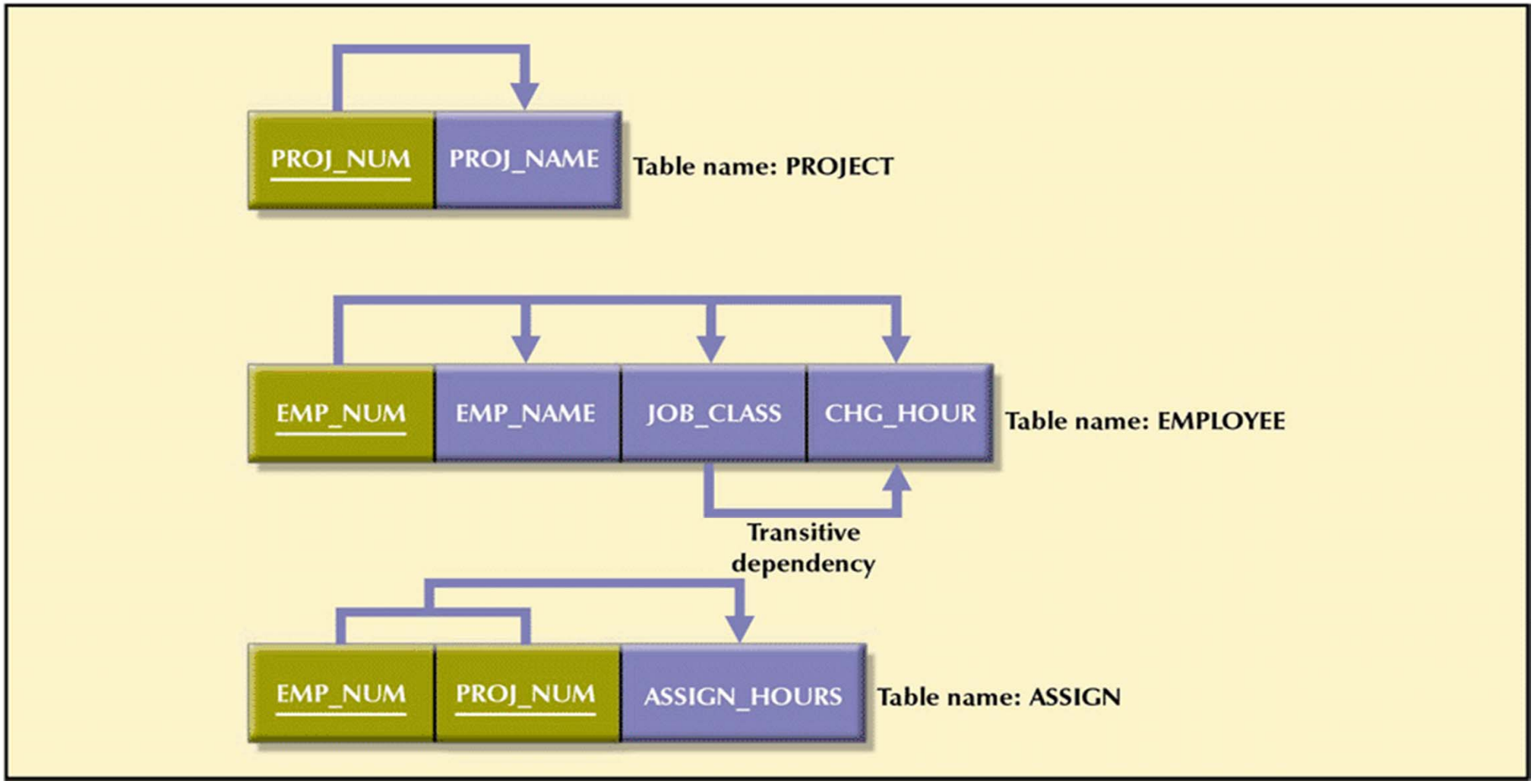
From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

2NF Step 2: Identify Dependent Attributes

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

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Second Normal Form (2NF) Conversion Results



From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Second Normal Form

- 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

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Conversion to Third Normal Form (3NF)

To minimize transitive dependencies

- Data anomalies created are easily eliminated by completing three steps
  - 3NF Step 1: Identify Each New Determinant
  - 3NF Step 2: Identify the Dependent Attributes
  - 3NF Step 3: Remove the Dependent Attributes from Transitive Dependencies

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

3NF 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

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

3NF 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

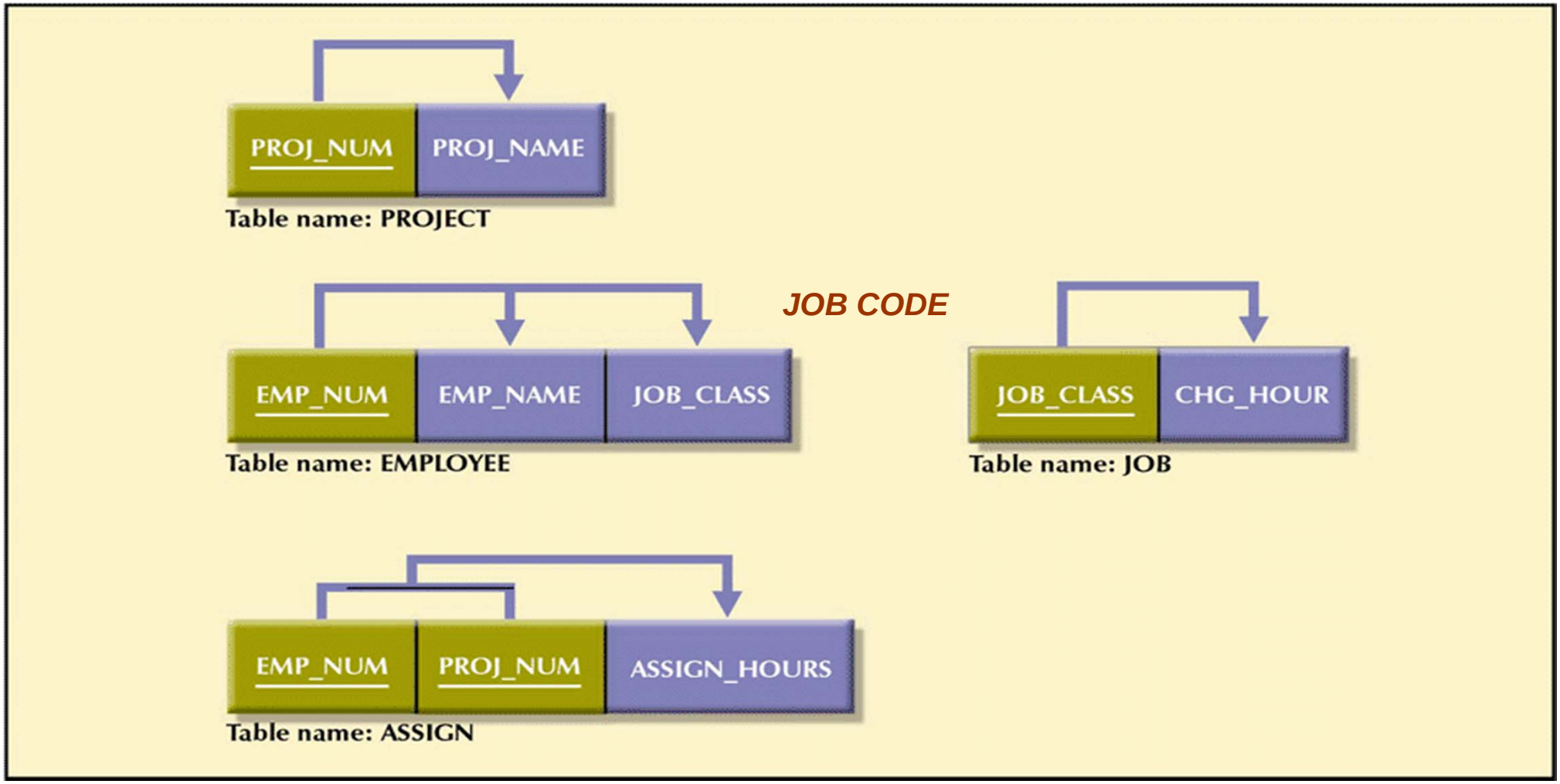
From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

3NF 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

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Third Normal Form (3NF) Conversion Results



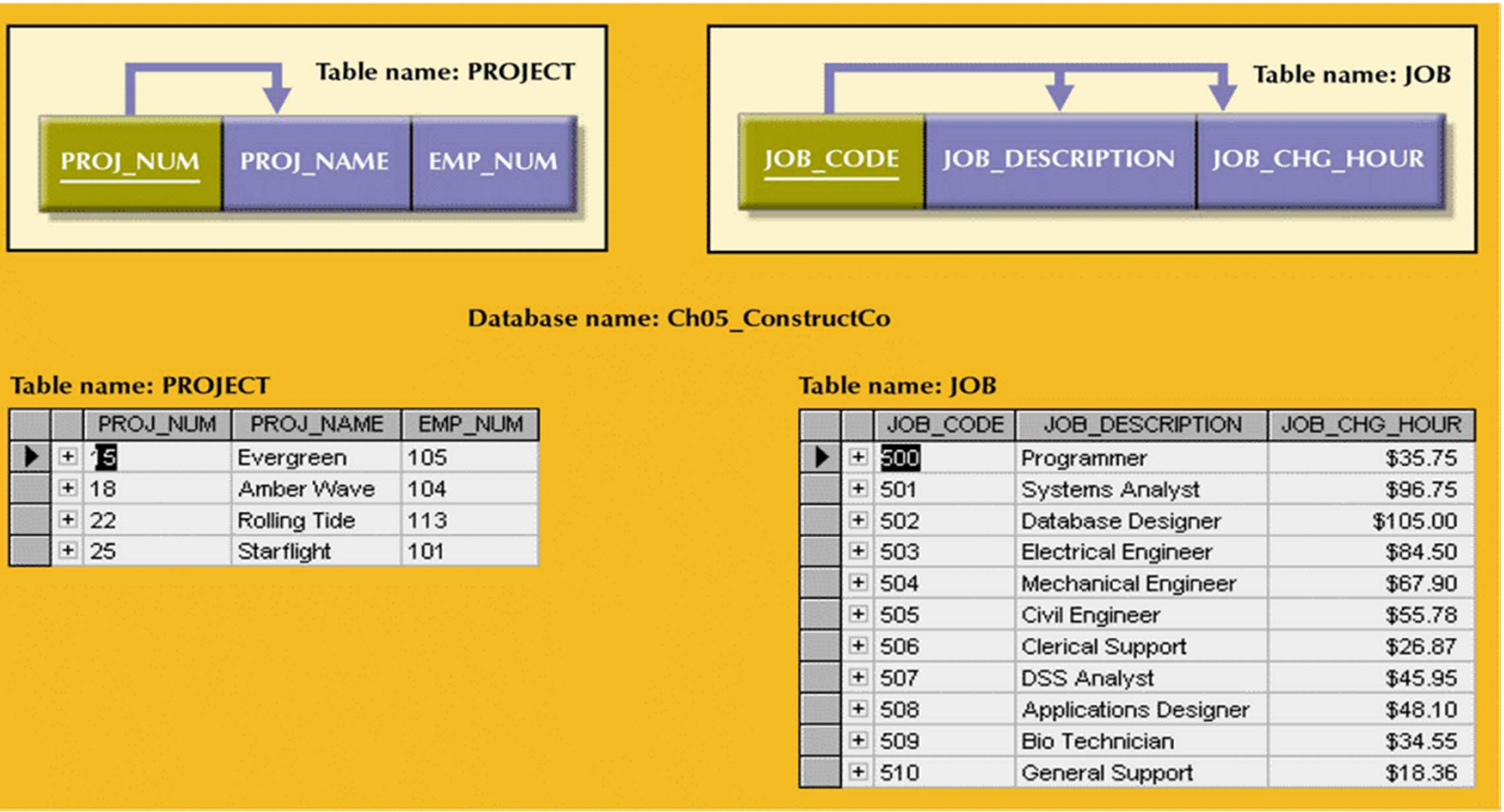
From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Third Normal Form

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

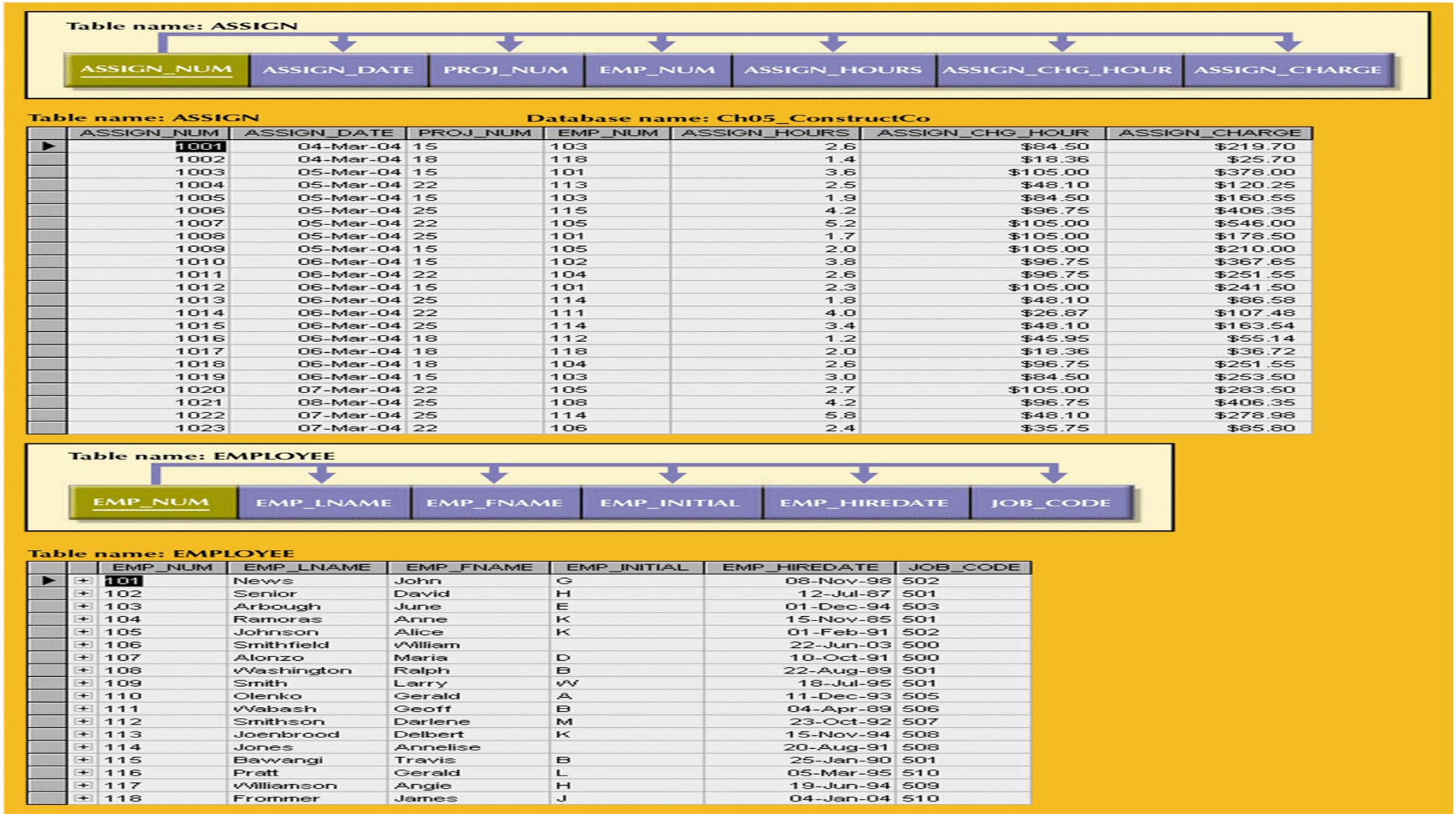
From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

The Completed Database



From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

The Completed Database (continued)



From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

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

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Improving the Design (2)

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

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Normalization and Database Design

- Normalization should be part of design process
- Make sure that proposed entities meet required normal form before table structures are created
- Many real-world databases have been improperly designed or burdened with anomalies if improperly modified during course of time
- You may be asked to redesign and modify existing databases

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Normalization and Database Design

- ER diagram
  - Provides the big picture, or macro view, of an organization's data requirements and operations
  - Created through an iterative process
    - Identifying relevant entities, their attributes and their relationship
    - Use results to identify additional entities and attributes

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Normalization and Database Design

- Normalization procedures
  - Focus on the characteristics of specific entities
  - A micro view of the entities within the ER diagram
- Difficult to separate normalization process from ER modeling process
- Two techniques should be used concurrently

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Denormalization

- Creation of normalized relations is important database design goal
- Processing requirements should also be a goal
- If tables decomposed to conform to normalization requirements
  - Number of database tables expands

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Denormalization (2)

- Joining larger number of tables takes additional disk input/output (I/O) operations and processing logic
  - Reduces system speed
- Conflicts among design efficiency, information requirements, and processing speed are often resolved through compromises that may include denormalization

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

Normalization Exercises

From Rob and Coronel (2004), Database Systems: Design, Implementation, and Management

1. To keep track of office furniture, computers, printers, and so on, the FOUNDIT Company uses the table structure shown in the table below.

Attribute Name	Sample Record	Sample Record
ITEM_ID	231-678	254-449
ITEM_LABEL	HP Deskjet 670	HP Scanner
ROOM_NUMBER	325	123
BLDG_CODE	NTC	TF
BLDG_NAME	Nottooclear	Toofar
BLDG_MANAGER	I. Right	M. Next

Given this information, draw the dependency diagram. Make sure you label the transitive and/or partial dependencies.

2. Starting with the dependency diagram drawn for Question 1, create a set of dependency diagrams that meet 3NF requirements. Rename attributes and create new entities and attributes as necessary.