Introduction to Databases

CT042-3-1-IDB (version1)

Database Normalization

Instructor: Bidur Devkota

Topic & Structure of The Lesson

- Functional dependency
- Un-Normalized form
- First normal form
- Second normal form
- Third normal form

Learning Outcomes

- **At the end of this topic, You should be able to**
  - **Describe functional dependency**
  - **Perform database normalization**

Key Terms You Must Be Able To Use

- **If you have mastered this topic, you should be able to use the following terms correctly in your assignments and exams:**
  - **UNF**
  - **1NF**
  - **2NF**
  - **3NF**
  - **Partial dependency**
  - **Transitive dependency**

Database Tables and Normalization

- Normalization
  - Works through a series of stages called normal forms:
    - First normal form (1NF)
    - Second normal form (2NF)
    - Third normal form (3NF)
    - Boyce-Codd Normal Form (BCNF / 3.5NF)
    - Fourth normal form (4NF)
    - Fifth Normal Form (5NF)

Database Tables and Normalization

- Normalization
  - Process for **evaluating and correcting table structures** to **minimize data redundancies**
    - **Reduces data anomalies.**
  - It works for:
    - No information redundancy
    - No anomalies

Database Tables and Normalization

Employee Number	First Name	Last Name	Date of Birth	Department Code	Department Name	Department Head
1001	Steave	Jakson	25-09-1985	SA001	Sales	Paul Colgan
1002	Kitty	Mathew	06-04-1998	ACC008	Accounts	Jerry Mathew
1003	Meena	Patel	11-05-1992	SA001	Sales	Paul Colgan
1004	Nancy	Samual	02-12-1996	ACC008	Accounts	Jerry Mathew
1005	Michael	Smith	28-03-1995	SA001	Sales	Paul Colgan
1006	James	Garcia	22-01-1994	SA002	Sales	David Smith
1007	Nancy	Samual	11-02-1996	ACC008	Accounts	Charles Williams

Duplicate Data Is Called Redundant Data.

SA001	Sales	Paul Colgan
ACC008	Accounts	Jerry Mathew

https://www.learncomputerscienceonline.com/database-normalization/

Database Tables and Normalization

- **Anomalies causes database to become inconsistent.**
- **Example Scenario**:
  - **Initially** Database is in a **consistent state.**
  - User runs **valid queries** which causes **inconsistency.**
- **Update anomaly:**
  - the action that causes the error is a row update,
- Similarly can be **Insertion or deletion anomaly** .

Database Tables and Normalization

- **Anomalies causes database to become inconsistent.**
- E.g. The address for *Employee ID 519* occurs multiple times in the table.
- User updates address of the employee in **only 1** row.
- **update anomaly** occurred**:**
  - inconsistent data in the table
    - the employee cannot have two addresses at once.

Employees' Skills

Employee ID	Employee Address	Skill
426	87 Sycamore Grove	Typing
426	87 Sycamore Grove	Shorthand
519	94 Chestnut Street	Public Speaking
519	96 Walnut Avenue	Carpentry

https://en.wikibooks.org/wiki/Relational_Database_Design/Normalization

Database Tables and Normalization

A · P · U
ASIA PACIFIC UNIVERSITY OF TECHNOLOGY & INNOVATION

1NF - First normal form

- Eliminate repeating groups
- Identify PK → result is functional dependency

2NF - Second normal form

- Eliminate partial dependencies

3NF - Third normal form

- Eliminate transitive dependencies

https://bit.ly/3hNavTp

Database Tables and Normalization (continued)

- Normalization (continued)
  - 2NF is better than 1NF; 3NF is better than 2NF
  - For most business database design purposes, 3NF is as high as we need to go in normalization process
  - Highest level of normalization is not always most desirable

The Need for Normalization

- Example: 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, report is generated that contains information displayed in Table 5.1

The Need for Normalization (continued)

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	$ 85.50	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	25.6	$ 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	45.0	$ 2,021.80
				Subtotal			**$ 7,172.47**
22	Rolling Tide	105	Alice K. Johnson	Database Designer	$105.00	65.7	$ 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	Starflight	107	Maria D. Alonzo	Programmer	$ 35.75	25.6	$ 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			**$48,942.09**

Note: * indicates project leader.

The Need for Normalization (continued)

FIGURE 5.1 Tabular representation of 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

The Need for Normalization (continued)

- Structure of data set in Figure 5.1 does not handle data very well
- The table structure appears to work; report generated with ease
- Unfortunately, report may yield different results depending on what data anomaly has occurred

The Normalization Process

- Each table represents a single subject
- No data item will be unnecessarily stored in more than one table → **No Duplication**
- All attributes in a table are dependent on the primary key

The Normalization Process (continued)

TABLE 5.2 Normal Forms

NORMAL FORM	CHARACTERISTIC	SECTION
First normal form (1NF)	Table format; no repeating groups and PK identified	5.3.1
Second normal form (2NF)	1NF and no partial dependencies	5.3.2
Third normal form (3NF)	2NF and no transitive dependencies	5.3.3
Boyce-Codd normal form (BCNF)	Every determinant is a candidate key (special case of 3NF)	5.6.1
Fourth normal form (4NF)	3NF and no independent multivalued dependencies	5.6.2

Normalization stages

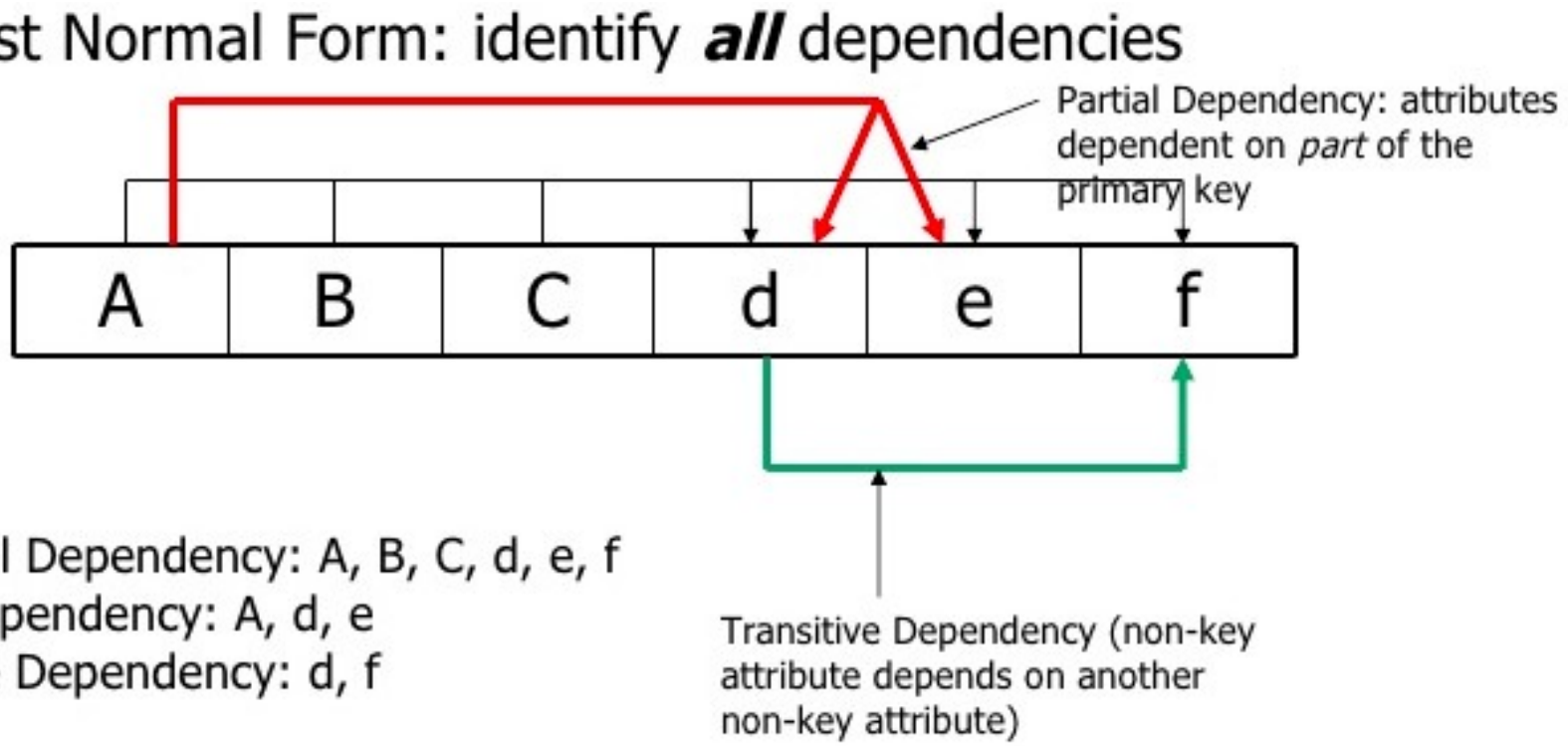
- 1NF - First normal form
  - Eliminate repeating groups
  - Identify PK → result is functional dependency
- 2NF - Second normal form
  - Eliminate partial dependencies
- 3NF - Third normal form
  - Eliminate transitive dependencies

https://bit.ly/3hNavTp

Conversion to First Normal Form

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

Conversion to First Normal Form



Note: Capital letters refer to primary key, lower case letters refer to attributes.

https://bit.ly/3hNavTp

Conversion to First Normal Form (continued)

- Step 1: **Eliminate the Repeating Groups**
  - Present data in t**abular format,** where **each cell has single value** and there are **no repeating groups**
  - Eliminate **repeating groups**, **eliminate nulls** by making sure that each repeating group attribute contains an appropriate data value

Conversion to First Normal Form (continued)

Step 1
Eliminate Repeating Groups

Step 1 is to eliminate repeating groups from our database.

To eliminate the repeating groups, eliminate the nulls by making sure that each repeating group attribute contains an appropriate data value. That change converts the table in Figure 1.1 (above figure) to 1NF in Figure 1.2 (below figure).

Figure 1.1

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	Starlight	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	23.8
		118	James J. Frommer	General Support	18.36	30.5
		112	Darlene M. Smithson	DSS Analyst	45.95	41.4

Figure 1.2

Table name: DATA_ORDG_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	Starlight	107	Maria D. Alonzo	Programmer	35.75	24.6
25	Starlight	115	Travis B. Bawangi	Systems Analyst	96.75	45.8
25	Starlight	101	John G. News *	Database Designer	105.00	56.3
25	Starlight	114	Annelise Jones	Applications Designer	48.10	33.1
25	Starlight	108	Ralph B. Washington	Systems Analyst	96.75	23.8
25	Starlight	118	James J. Frommer	General Support	18.36	30.5
25	Starlight	112	Darlene M. Smithson	DSS Analyst	45.95	41.4

Mudasir Qazi - mudasirqazi00@gmail.com

https://bit.ly/2ZS7ydY

Conversion to First Normal Form (continued)

FIGURE 5.2 A table in first normal form

Eliminated repeating groups populated every cell of the 'table'.

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

Conversion to First Normal Form (continued)

- Step 2: **Identify the Primary Key**
  - Primary key must uniquely identify attribute value
  - New key must be composed

Conversion to First Normal Form (continued)

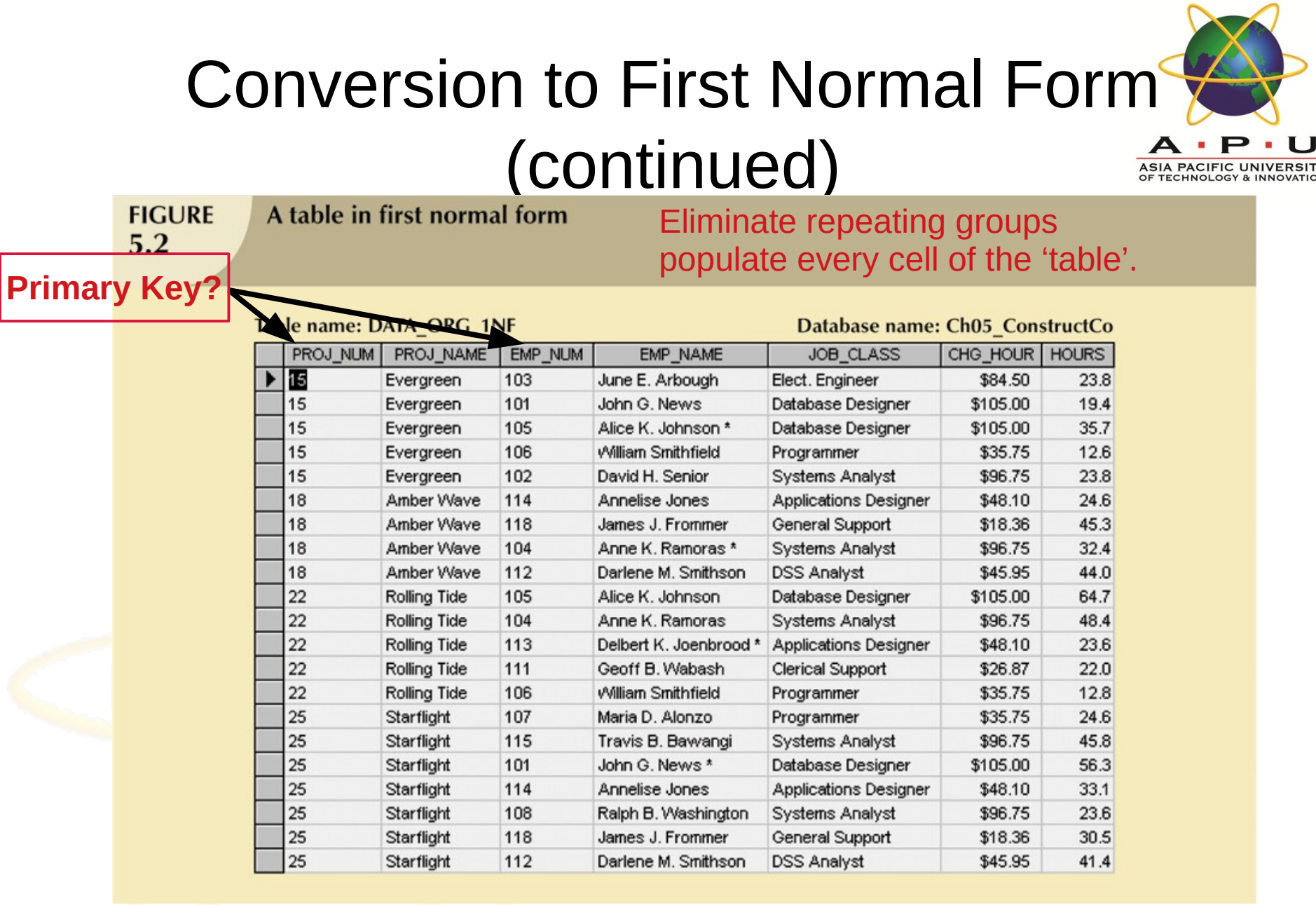


FIGURE 5.2 A table in first normal form

Eliminate repeating groups
populate every cell of the 'table'.

Primary Key?

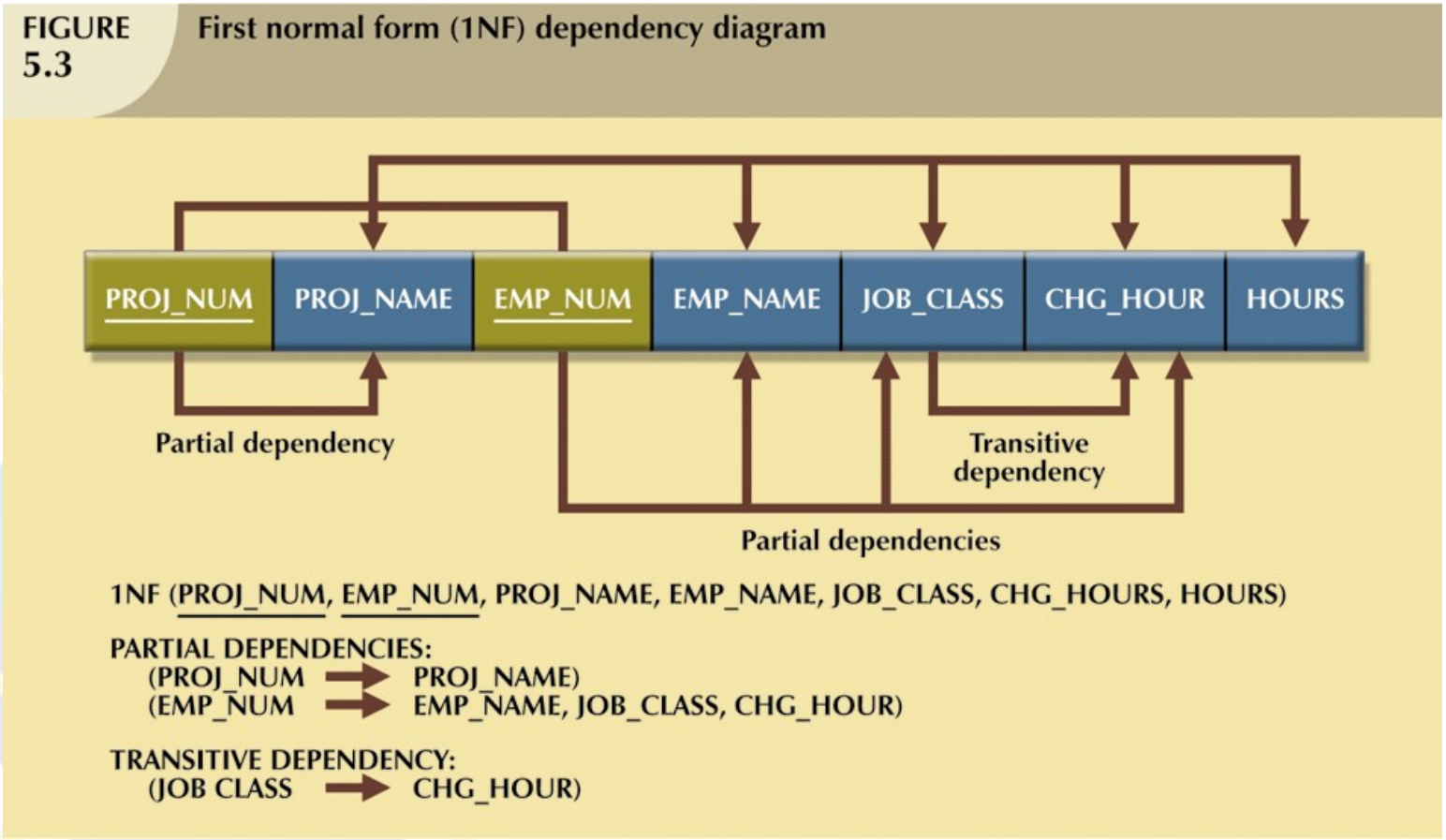
Table name: DATA_ORG_1NF Database name: Ch05_ConstructCo

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

Conversion to First Normal Form (continued)

- Step 3: **Identify All Dependencies**
  - Dependencies can be depicted with help of a diagram
  - Dependency diagram:
    - Depicts all dependencies found within given table structure
    - Helpful in getting bird's-eye view of all relationships among table's attributes
    - Makes it less likely that will overlook an important dependency

Conversion to First Normal Form (continued)



Conversion to First Normal Form (continued)

- **First normal form describes 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

Normalization stages

- 1NF - First normal form
  - Eliminate repeating groups
  - Identify PK → result is functional dependency
- 2NF - Second normal form
  - Eliminate partial dependencies
- 3NF - Third normal form
  - Eliminate transitive dependencies

https://bit.ly/3hNavTp

Conversion to Second Normal Form

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

Conversion to Second Normal Form (continued)

- Step 1: Write Each Key Component on a Separate Line
  - Write each key component on separate line, then write original (composite) key on last line
  - Each component will become key in new table

PROJECT (PROJ_NUM (pk), PROJ_NAME)

https://bit.ly/3hNavTp

EMPLOYEE (EMP_NUM (pk), EMP_NAME, JOB_CLASS, CHG_HOUR)

EMPLOYEE_PROJECT (PROJ_NUM (pk, fk), EMP_NUM (pk, fk), HOURS)

Attribute of hours is dependent on composite primary key

Conversion to Second Normal Form (continued)

- Step 2: Assign Corresponding Dependent Attributes
  - Determine those attributes that are dependent on other attributes
  - At this point, most anomalies have been eliminated

PROJECT (PROJ_NUM (pk), PROJ_NAME)

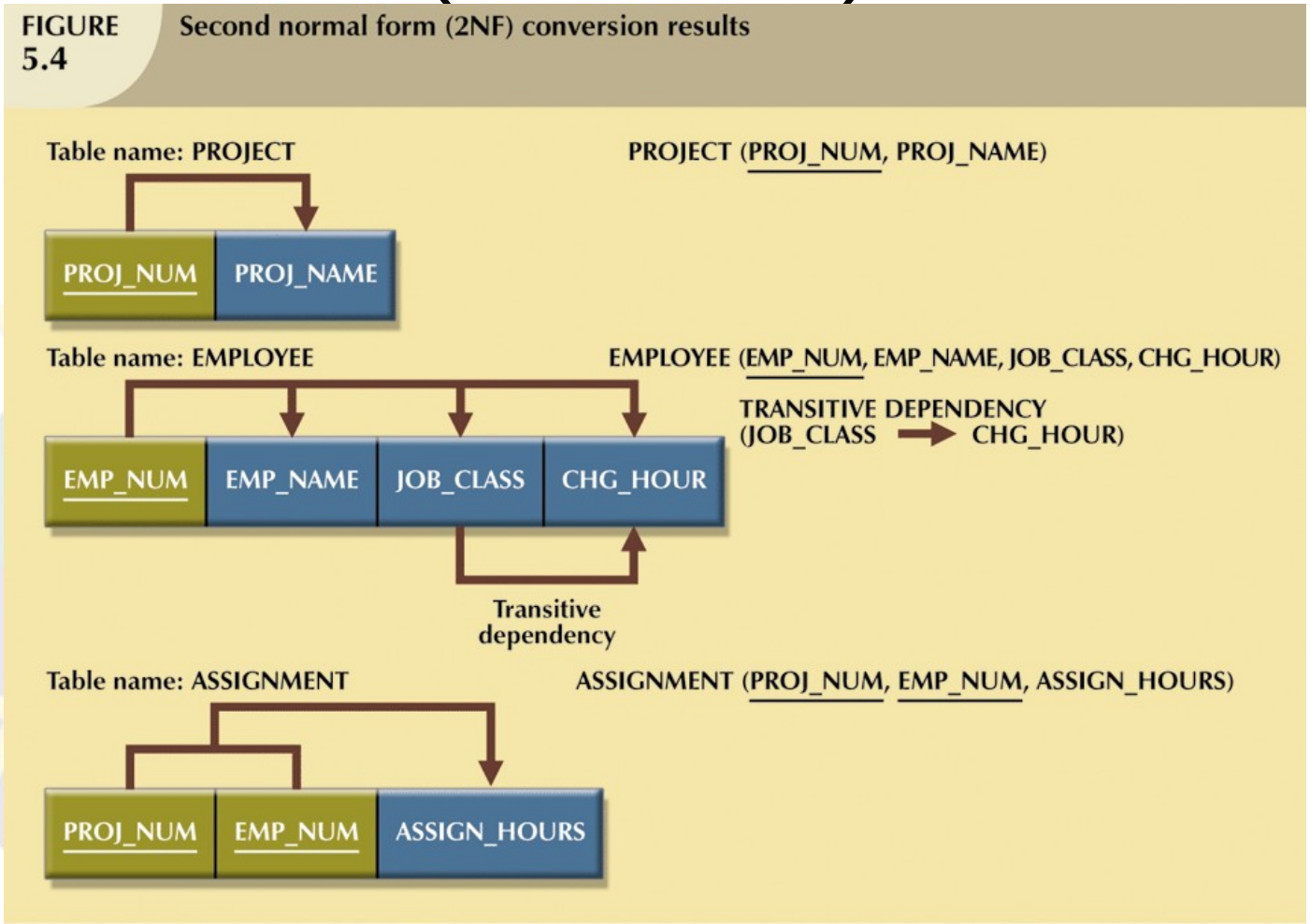
https://bit.ly/3hNavTp

EMPLOYEE (EMP_NUM (pk), EMP_NAME, JOB_CLASS, CHG_HOUR)

EMPLOYEE_PROJECT (PROJ_NUM (pk, fk), EMP_NUM (pk, fk), HOURS)

Attribute of hours is dependent on composite primary key

Conversion to Second Normal Form (continued)



Conversion to Second Normal Form (continued)

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

Normalization stages

- 1NF - First normal form
  - Eliminate repeating groups
  - Identify PK → result is functional dependency
- 2NF - Second normal form
  - Eliminate partial dependencies
- 3NF - Third normal form
  - Eliminate transitive dependencies

https://bit.ly/3hNavTp

Conversion to Third Normal Form

- Create separate tables to eliminate transitive functional dependencies
- Identify any additional attributes needed in new table

JOB (JOB_CLASS (pk), CHG_HOUR)

PROJECT (PROJ_NUM (pk), PROJ_NAME)

EMPLOYEE (EMP_NUM (pk), EMP_NAME, JOB_CLASS (fk))

EMPLOYEE_PROJECT (PROJ_NUM (pk, fk), EMP_NUM (pk, fk), HOURS)

Steps are discussed in the up coming slides..

Conversion to Third Normal Form

- Data anomalies created are easily eliminated by completing three steps
- Step 1: Identify Each New Determinant
  - For every transitive dependency, write its determinant as PK for new table
    - Determinant
      - Any attribute whose value determines other values within a row

JOB (JOB_CLASS (pk),

Conversion to Third Normal Form (continued)

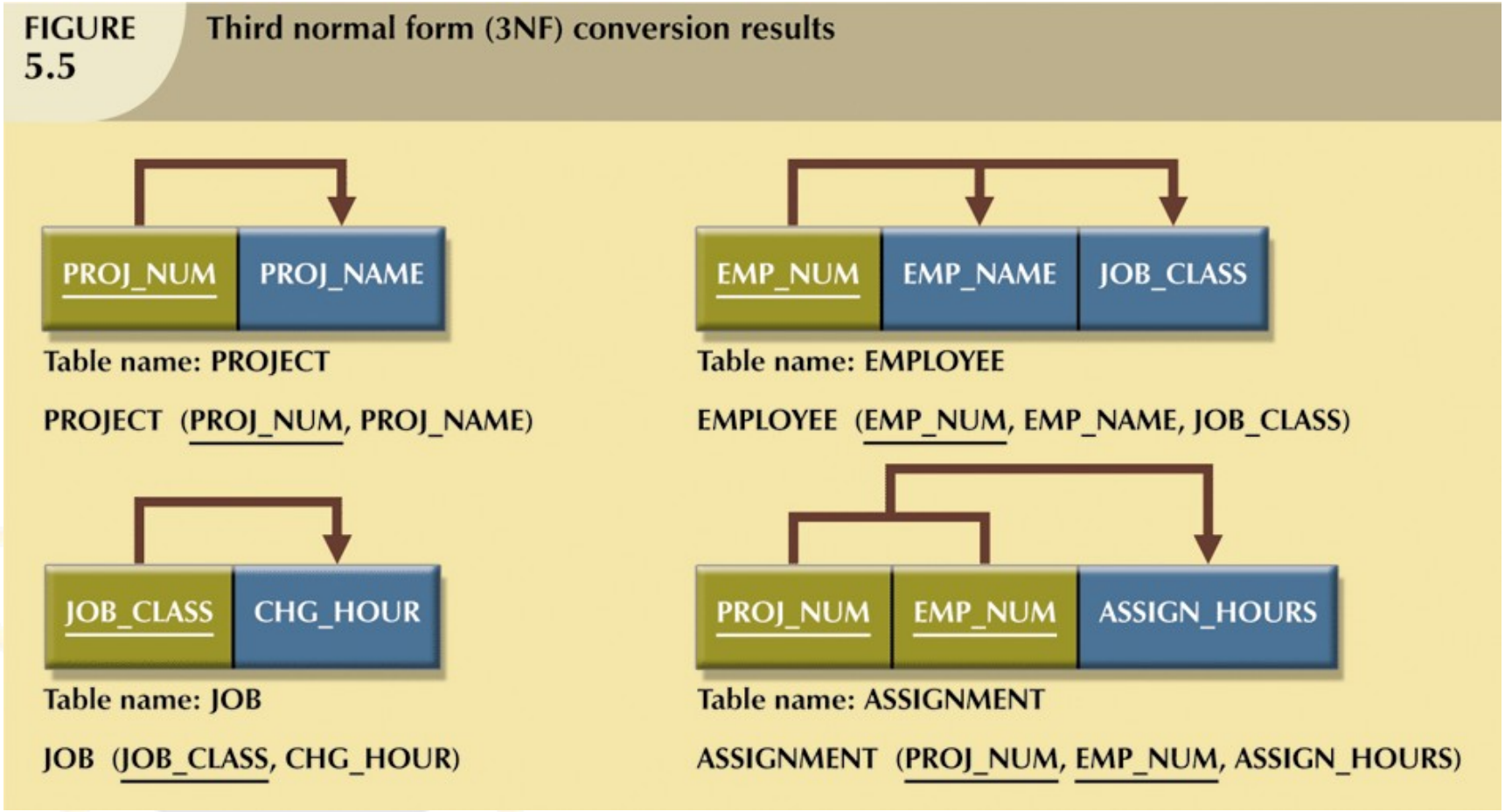
- Step 2: Identify the Dependent Attributes
  - Identify attributes dependent on each determinant identified in Step 1 and identify dependency
  - Name table to reflect its contents and function

JOB (JOB_CLASS (pk), CHG_HOUR)

Conversion to Third Normal Form (continued)

- Step 3: Remove the Dependent Attributes from Transitive Dependencies
  - Eliminate all dependent attributes in transitive relationship(s) from each of the tables that have such a transitive relationship
  - Draw new dependency diagram to show all tables defined in Steps 1–3
  - Check new tables as well as tables modified in Step 3 to make sure that each table has determinant and that no table contains inappropriate dependencies

Conversion to Third Normal Form (continued)



Conversion to Third Normal Form (continued)

- A table is in third normal form (3NF) when both of the following are true:
  - It is in **2NF**
  - It contains **no transitive dependencies**

Quick Review Question

- What is partial dependency?
- What is transitive dependency?

Summary of Main Teaching Points

- Normal forms can be transformed from lower normal forms to higher normal forms
- Normalization and ER modeling are used concurrently to produce a good database design

Question and Answer Session

A·P·U
ASIA PACIFIC UNIVERSITY
OF TECHNOLOGY & INNOVATION

Q & A

What we will cover next

- Structured Query Language (SQL)