



Introduction to Database

Topics

- Database
- DBMS
- Relational Database
- Keys
- ER Model & Normalization
- SQL



Database

- A database is an **ordered collection of data**
- The word “Database” can be quite vague. People tend to clump a combination of three things when referring to a database
 - The Data
 - The Database Management System
 - Database Application



The Database Management System (DBMS)

- A DBMS is a Software that is designed to model data and provide interaction with a user
 - Data is normally stored in one or more files
 - Data is usually stored in **tables**
 - Table contain **rows** and **columns** much like a **spreadsheet**
 - Data can be cross referenced among tables. This functionality is used in **“Relational Databases”**

DBMS

381 systems in ranking, December 2021

Rank			DBMS	Database Model	Score		
Dec 2021	Nov 2021	Dec 2020			Dec 2021	Nov 2021	Dec 2020
1.	1.	1.	Oracle +	Relational, Multi-model i	1281.74	+9.01	-43.86
2.	2.	2.	MySQL +	Relational, Multi-model i	1206.04	-5.48	-49.41
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model i	954.02	-0.27	-84.07
4.	4.	4.	PostgreSQL +	Relational, Multi-model i	608.21	+10.94	+60.64
5.	5.	5.	MongoDB +	Document, Multi-model i	484.67	-2.67	+26.95
6.	6.	↑ 7.	Redis +	Key-value, Multi-model i	173.54	+2.04	+19.91
7.	7.	↓ 6.	IBM Db2	Relational, Multi-model i	167.18	-0.34	+6.74
8.	8.	8.	Elasticsearch	Search engine, Multi-model i	157.72	-1.36	+5.23
9.	9.	9.	SQLite +	Relational	128.68	-1.12	+7.00
10.	↑ 11.	↑ 11.	Microsoft Access	Relational	125.99	+6.75	+9.25



Relational Database

- A relational database represents a collection of related (two-dimensional) tables
- Each of the tables are similar to an Excel spreadsheet
 - with **a fixed number of named columns** (the attributes or properties of the table)
 - **any number of rows** of data.

Simple Example

id	firstname	lastname	email
1	Molta	Sansard	molta.san@gmail.com
2	Jutamas	Saengvanich	jutamas.sae@gmail.com
3	Winai	Kulthayawat	winai.kul@gmail.com

Database Name: MyStore

Table Name: Users

Column/ Fields: id, firstname, lastname, email

Row/ Records: 3 Users

Database Design – Goal

- Maintain a balance Database Size and Performance



Database Design – Goal



How uploader's
details are stored

Table Name: Photo

ID	Title	Uploader Name	Uploader PhoneNum	Uploader Address
1	The tram	Chris Wong	252-608-5628	572 Lost Lake Crossing
2	Little kitten	Chris Wong	252-608-5628	572 Lost Lake Crossing
3	Chamomile	Chris Wong	252-608-5628	572 Lost Lake Crossing
4	New bus route	Wanda A. Allison	808-839-2217	2017 Green Impasse
5	Ferry pier 9	Chris Wong	252-608-5628	572 Lost Lake Crossing
6	The coast	Chris Wong	252-608-5628	572 Lost Lake Crossing
7	Rock climbing	Wanda A. Allison	808-839-2217	2017 Green Impasse
8	Fountain	Chris Wong	252-608-5628	572 Lost Lake Crossing
9	Sunset	Wanda A. Allison	808-839-2217	2017 Green Impasse

Database Design – Goal

Table Name: Photo

ID	Title	Uploader Name	Uploader PhoneNum	Uploader Address
1	The tram	Chris Wong	252-608-5628	572 Lost Lake Crossing
2	Little kitten	Chris Wong	252-608-5628	572 Lost Lake Crossing
3	Chamomile	Chris Wong	252-608-5628	572 Lost Lake Crossing
4	New bus route	Wanda A. Allison	808-839-2217	2017 Green Impasse
5	Ferry pier 9	Chris Wong	252-608-5628	572 Lost Lake Crossing
6	The coast	Chris Wong	252-608-5628	572 Lost Lake Crossing
7	Rock climbing	Wanda A. Allison	808-839-2217	2017 Green Impasse
8	Fountain	Chris Wong	252-608-5628	572 Lost Lake Crossing
9	Sunset	Wanda A. Allison	808-839-2217	2017 Green Impasse



**Access both photo and
uploader details when
querying a single record**

Database Design – Goal

Table Name: Photo

ID	Title	Uploader Name	Uploader PhoneNum	Uploader Address
1	The tram	Chris Wong	252-608-5628	572 Lost Lake Crossing
2	Little kitten	Chris Wong	252-608-5628	572 Lost Lake Crossing
3	Chamomile	Chris Wong	252-608-5628	572 Lost Lake Crossing
4	New bus route	Wanda A. Allison	808-839-2217	2017 Green Impasse
5	Ferry pier 9	Chris Wong	252-608-5628	572 Lost Lake Crossing
6	The coast	Chris Wong	252-608-5628	572 Lost Lake Crossing
7	Rock climbing	Wanda A. Allison	808-839-2217	2017 Green Impasse
8	Fountain	Chris Wong	252-608-5628	572 Lost Lake Crossing
9	Sunset	Wanda A. Allison	808-839-2217	2017 Green Impasse

However, this may create a lot of redundant data

Database Design – Goal

Table Name: Photo

ID	Title	Uploader ID
1	The tram	1
2	Little kitten	1
3	Chamomile	1
4	New bus route	2
5	Ferry pier 9	1
6	The coast	1
7	Rock climbing	2
8	Fountain	1
9	Sunset	2

Table Name: Uploader

ID	Name	PhoneNum	Address
1	Chris Wong	252-608-5628	572 Lost Lake Crossing
2	Wanda A. Allison	808-839-2217	2017 Green Impasse



ER Diagram

A Relational Database Overview

- A **database** is a means of storing information in such a way that **information** can be **retrieved** from it.
- A relational database is one that presents information in **tables** with **rows** and **columns**.
- A **table** is referred to as a relation in the sense that it is a collection of objects of the **same type**.
- Data in a **table** can be **related** according to common keys or concepts, and the ability to retrieve related data from a table is the basis for the term relational database.
- A Database Management System (**DBMS**) handles the way data is stored, maintained, and retrieved.



Structure of Relational Database

- A relational database is a **collection of tables**.
 - Each table has a **unique name**.
 - Each table consists of **multiple rows**.
 - Each row is a set of values that by-definition are related to each other in someway; these values conform to the attributes or **columns** of the table.
 - Each attribute of a table defines a set of **permitted values** for that attribute; this set of permitted set is the domain of that attribute.

Tables

- A table is a collection of related data held in a structured format within a database.
- A table consists of columns and rows.
 - Each row in a table represents a set of related data (tuple).
 - Each row does not allow duplicates.
 - Every row in the table has the same structure.
 - Order of rows does not matter.
 - Order of column (fields) does not matter.

Example

- Employee Table

The diagram illustrates the components of a database table. A table with three columns (EmployeeID, Name, Salary) and four rows is shown. Annotations include: a blue arrow pointing to the 'Name' column header labeled 'Column'; a yellow arrow pointing to the top-right corner of the table labeled 'Schema'; and a green arrow pointing to the second row labeled 'Row'.

EmployeeID	Name	Salary
45180	Nancy Bush	30,000
57138	Andrew Jackson	25,000
58015	Janet Lowe	25,000
49222	Margaret Danes	20,000

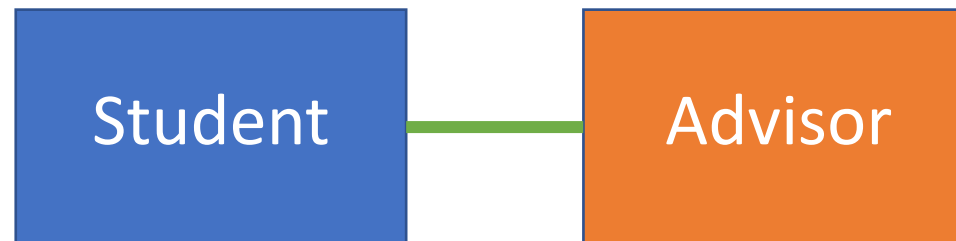
Attributes

- An attribute is a property of a given entity.
- The value of each attribute contains only a single value from that domain.
- Each attribute contains only atomic (indivisible) values.

ID	NAME	ADDRESS				
ID	NAME	ADDR	STREET	SUB_DIS	DISTRICT	PROVINCE

Keys

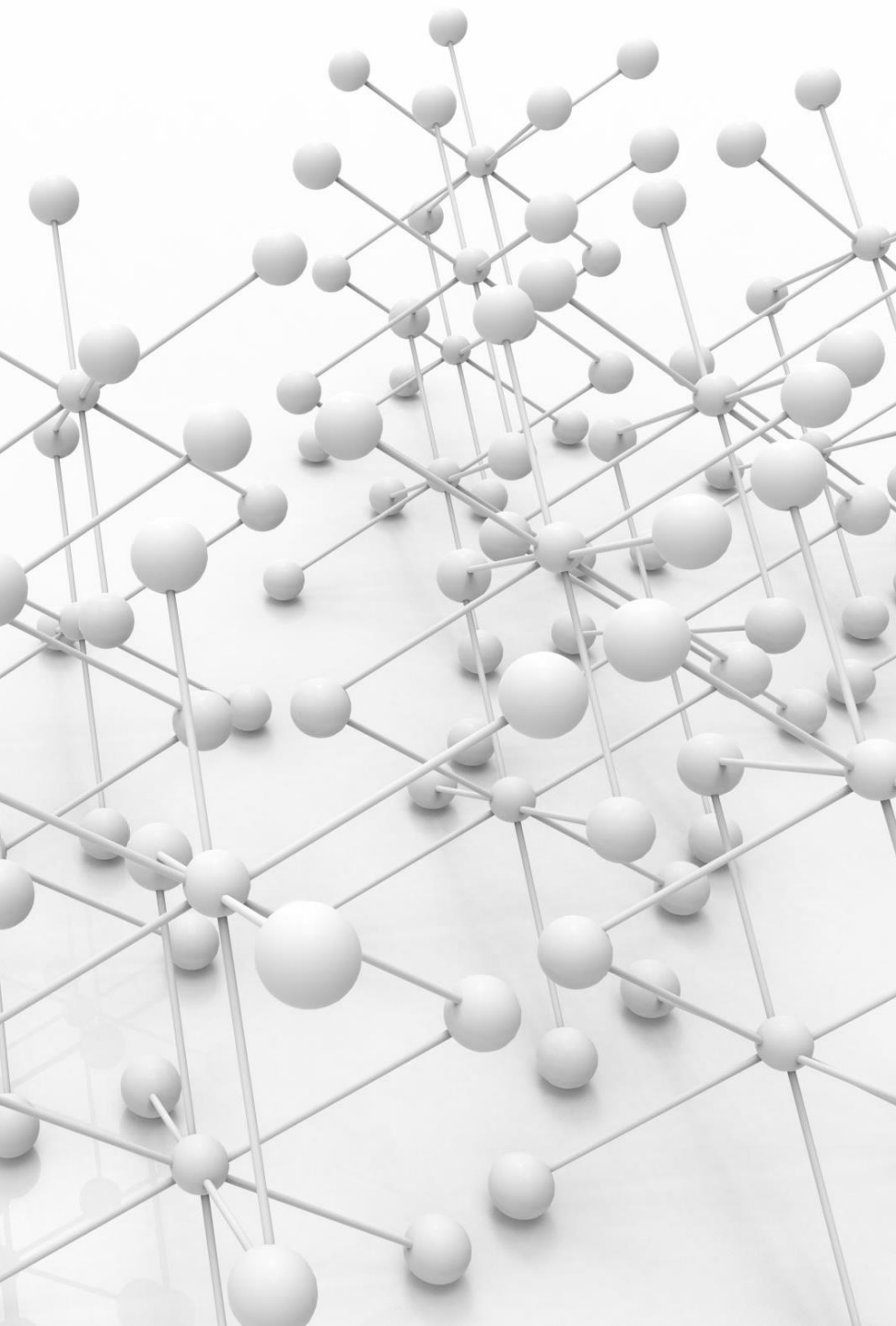
- **Primary key** : A candidate key that is most appropriate to become main key of the table
 - Uniqueness
 - Minimum
- **Composite key**: consist of two or more attributes that uniquely identify an entity occurrence.
- **Foreign key**: an attribute (or collection of attributes) in one table that uniquely identifies a row of another table.



The background of the slide features a dark gray surface with a grid of thin black lines. Scattered across this grid are numerous 3D cubes of varying sizes and orientations, some of which are connected by thin black lines, creating a complex, abstract geometric pattern.

Database Design Notations

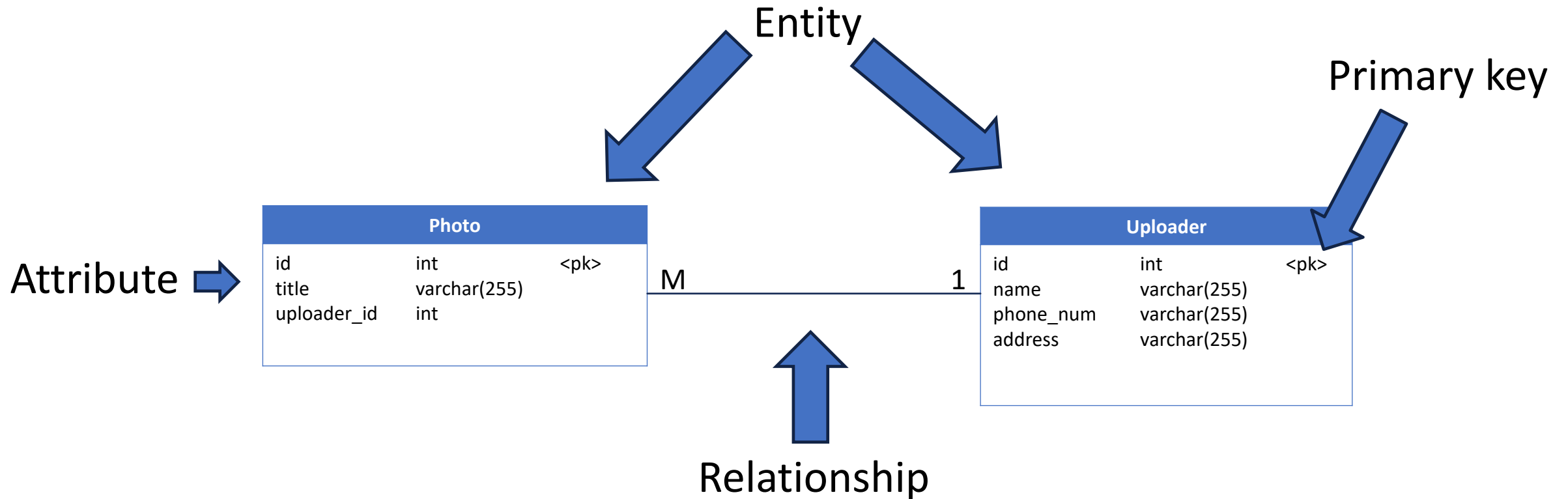
- Entity Relationship Diagram is:
 - Language for expressing database
 - Visual language
 - Facilitate communication
 - Presents data structure & interrelationships



ER Model: Entity-Relationship Model

- A graphical representation of entities and their relationships to each other
- There are 3 basic elements
 - Entity : real-world object or thing with an independent existence and distinguishable from other objects.
 - Attribute : a property or characteristic of an entity
 - Relationship : How entities associated with each other

Database Design Notations



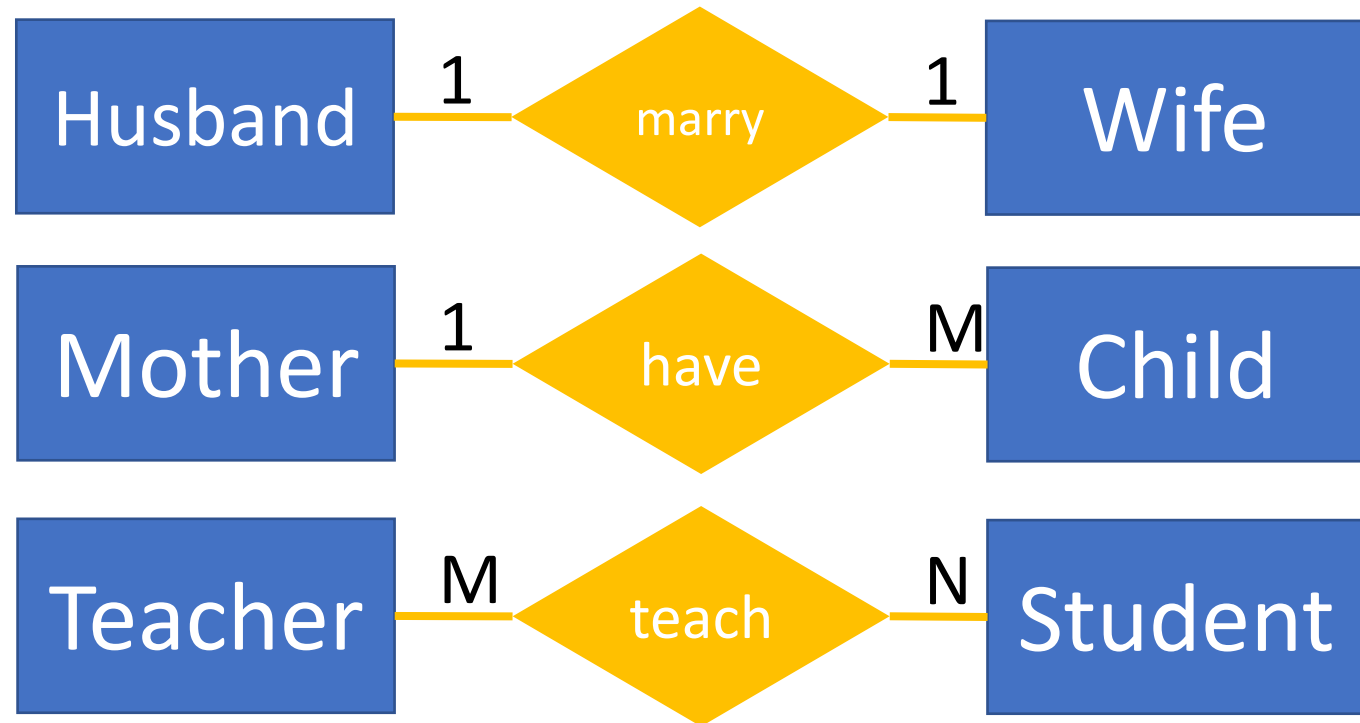
Relationships - Participation

- Total participation
- Partial participation



Relationships - Cardinality

- One-to-One
- One-to-Many
- Many-to-Many



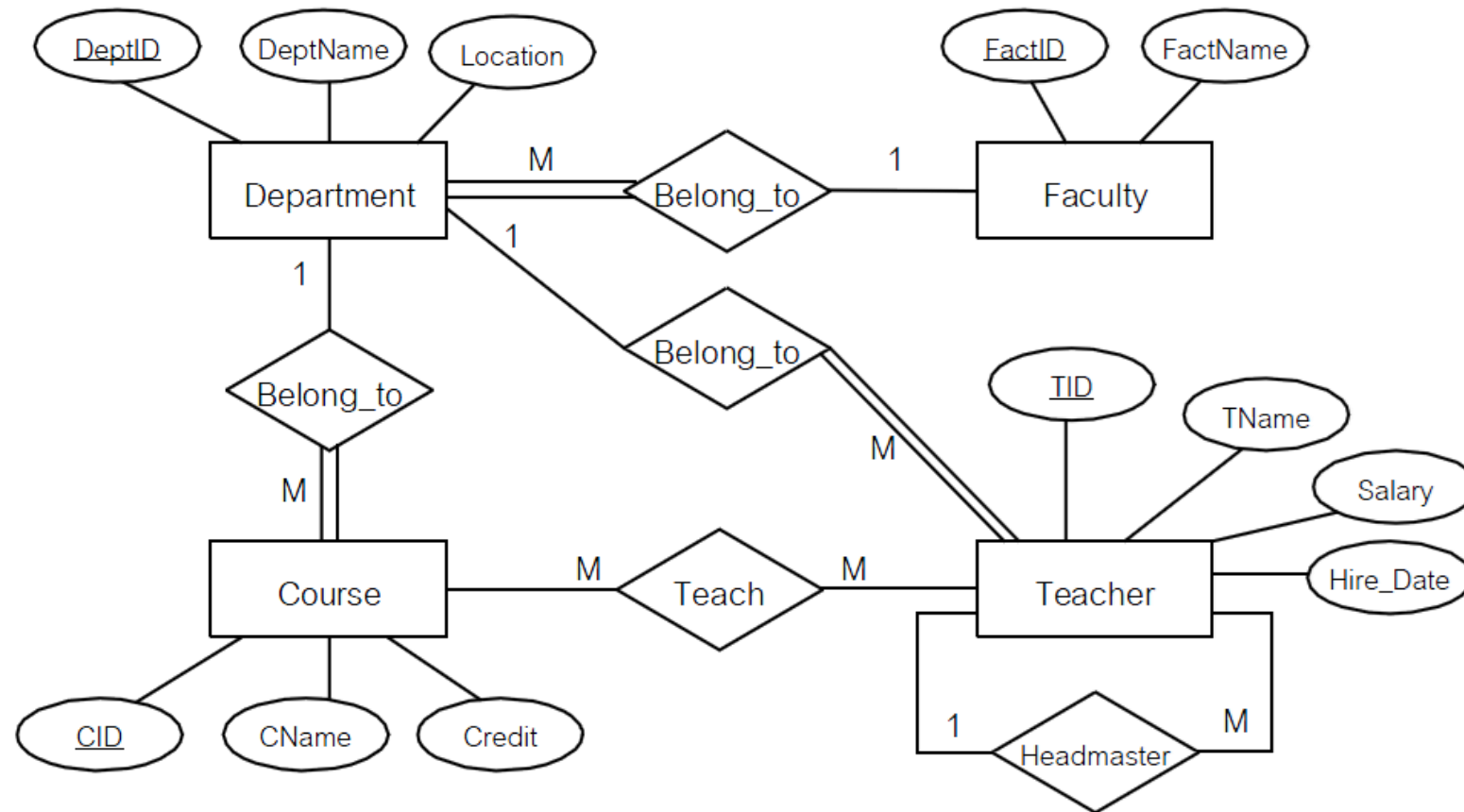
Problem 1:

- Draw an ER diagram for the following application for a store:
 - Each supplier has a unique name.
 - Each product has a unique product code.
 - A supplier can supply more than one products.
 - A product can be supplied by only one supplier.

Convert ER Diagram to Relational Database

1. Entity → Table & Attribute → column
2. Weak Entity → Table with PK of main entity's PK
3. 1:1 Relationships → add PK of partial-side entity as FK in the total-side entity
4. 1:M Relationships → add PK of 1-side entity as FK in the M-side entity
5. M:N Relationships → Table with PK of both PKs from both entities
6. Multi-Valued Attribute → Table with PK of the main entity
7. Multi-Entity Relationships → Table with PK of all entities

Problem 2: Convert ER Diagram to Relational Database



Normalization

- relation schema \rightarrow "Normal Form"
- กำหนด Attribute ให้กับแต่ละตารางความสัมพันธ์ เพื่อให้ได้โครงสร้างของตารางที่ดี
 - ลดความซ้ำซ้อนของข้อมูล
 - ลดความผิดปกติของข้อมูล

ตัวอย่างการออกแบบฐานข้อมูลที่จะมีปัญหของความซ้ำซ้อนตามมา

EMP_ID	EMP_NAME	POSITION	SALARY	BRANCH_ID	ADDRESS
AM021	พีรพล โพธิ์ทัย	MANAGER	30000	B001	BANGKOK
AM035	ศิริรัตน์ สายคุณากร	ASSISTANT MANAGER	20000	B005	RATCHABURI
AM010	เพชรภรณ์ เปี่ยมศิริ	SECRETARY	20000	B004	CHIANGMAI
AM005	ธวัชชัย ด่านลาเคน	MANAGER	30000	B005	RATCHABURI

ตัวอย่างปัญหาความซ้ำซ้อนในข้อมูล

EMP_ID	EMP_NAME	POSITION	SALARY	BRANCH_ID	ADDRESS
AM021	พีรพล โพธิ์ทัย	MANAGER	30000	B001	BANGKOK
AM035	ศิริรัตน์ สายคุณากร	ASSISTANT MANAGER	20000	B005	RATCHABURI
AM010	เพชรภรณ์ เปี่ยมศิริ	SECRETARY	20000	B004	CHIANGMAI
AM005	ธวัชชัย ด่านลาเคน	MANAGER	30000	B005	RATCHABURI

ความผิดพลาดจากการเพิ่ม

- ถ้าต้องการเพิ่มพนักงานใหม่ ที่อยู่สาขา B004
 - จะต้องกรอก B004 และ ADDRESS คือ CHIANGMAI เพิ่มอีก

ตัวอย่างปัญหาความซ้ำซ้อนในข้อมูล

EMP_ID	EMP_NAME	POSITION	SALARY	BRANCH_ID	ADDRESS
AM021	พีรพล โพธิ์ทัย	MANAGER	30000	B001	BANGKOK
AM035	ศิริรัตน์ สายคุณากร	ASSISTANT MANAGER	20000	B005	RATCHABURI
AM010	เพชรภรณ์ เปี่ยมศิริ	SECRETARY	20000	B004	CHIANGMAI
AM005	ธวัชชัย ดำนลาเคน	MANAGER	30000	B005	RATCHABURI

ความผิดพลาดจากการเพิ่ม

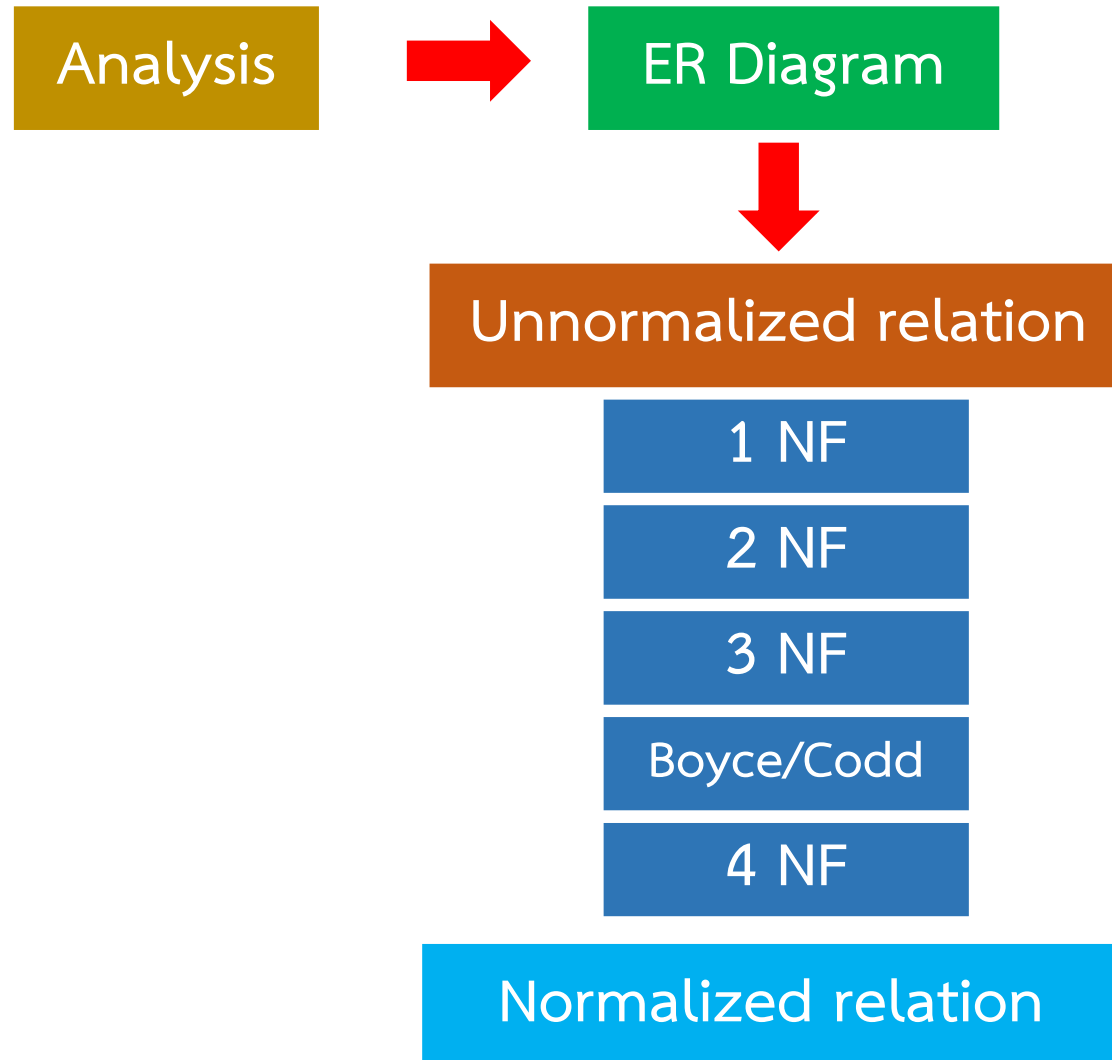
- ถ้าต้องการเพิ่มสาขา จะมีปัญหาคือ ตารางนี้มีทั้งข้อมูลพนักงานและข้อมูลสาขาอยู่รวมกัน
 - หากจะเพิ่มเฉพาะ รหัสสาขา และ ที่อยู่ ก็ไม่ได้เพราะ รหัสพนักงาน จะมีค่าว่างไม่ได้เพราะเป็น Primary Key ของตารางดังนั้นจะบันทึกได้ก็ต่อเมื่อมีพนักงานแล้ว

ตัวอย่างการออกแบบฐานข้อมูลที่ดี

<u>EMP_ID</u>	EMP_NAME	POSITION	SALARY	BRANCH_ID
AM021	พีรพล โพธิ์ทัย	MANAGER	30000	B001
AM035	ศิริรัตน์ สายคุณากร	ASSISTANT MANAGER	20000	B005
AM010	เพชรภรณ์ เปี่ยมศิริ	SECRETARY	20000	B004
AM005	ธวัชชัย ต่านลาเคน	MANAGER	30000	B005

<u>BRANCH_ID</u>	ADDRESS
B001	BANGKOK
B005	RATCHABURI
B004	CHIANGMAI

Normalization



รูปแบบบรรทัดฐาน

1NF “ขจัด Repeating Group”

2NF “ขจัด Partial Dependency”

3NF “ขจัด Transitive Dependency”

BCNF “ขจัดปัญหา Overlap Multiple Candidate Key”

4NF “ขจัด Nontrivial Multivalued Dependency”

5NF “ขจัด Join Dependency”

ตารางที่มีลักษณะข้อมูลเป็น Repeating Group

s_id	s_name	s_surname	c_id
542107030005	กฤษฎา	อิมรุ่งเรือง	CE354
			CE498
			CE305
554607030015	สาวิตรี	วิไลโรจน์	CE354
			CE356
			CE215
			CE327

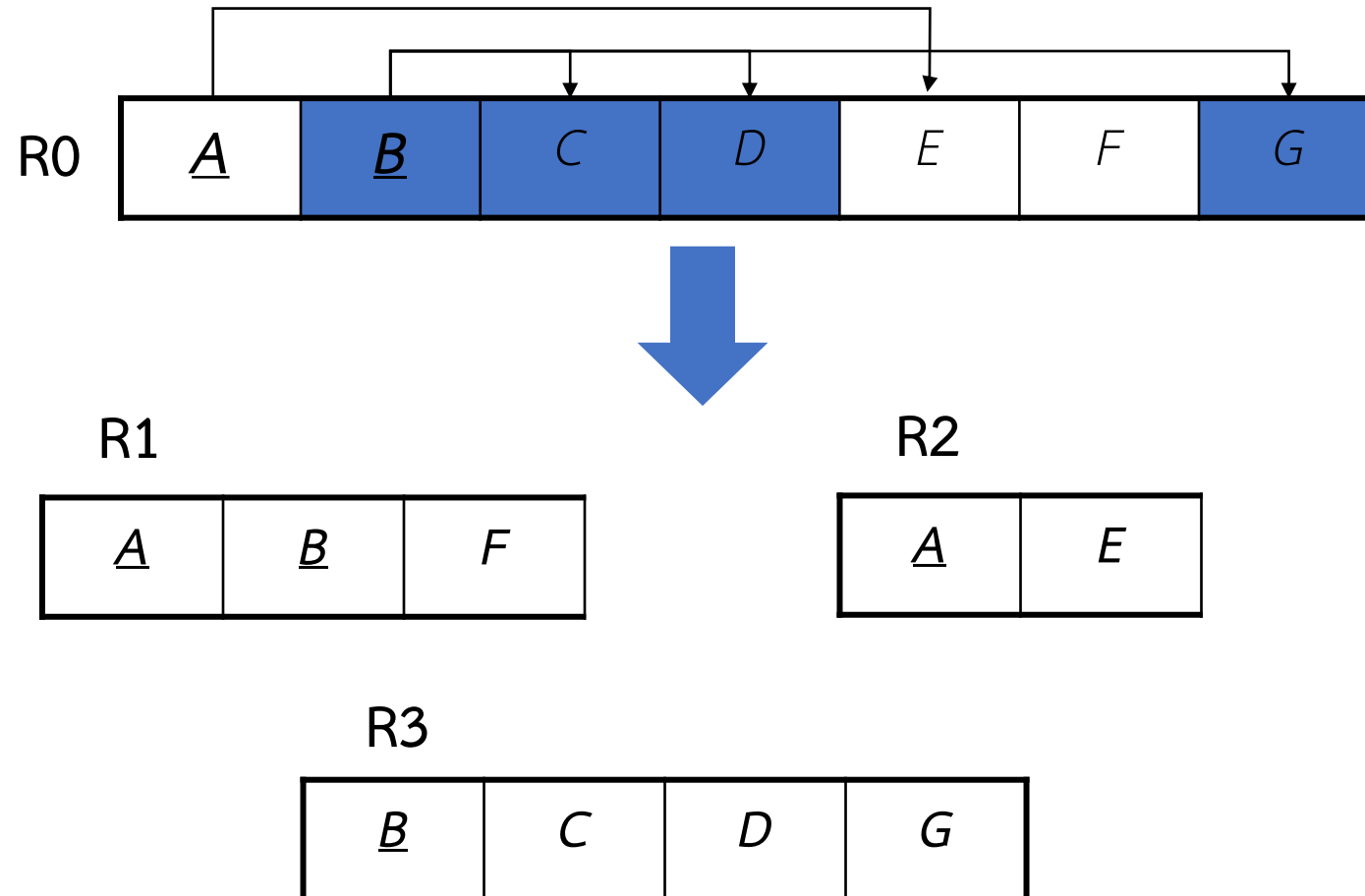
First Normal Form: 1NF

<u>s_id</u>	s_name	s_surname	<u>c_id</u>
542107030005	กฤษฎา	อิมรุ่งเรือง	CE354
542107030005	กฤษฎา	อิมรุ่งเรือง	CE498
542107030005	กฤษฎา	อิมรุ่งเรือง	CE305
554607030015	สาวิตรี	วิไลโรจน์	CE354
554607030015	สาวิตรี	วิไลโรจน์	CE356
554607030015	สาวิตรี	วิไลโรจน์	CE215
554607030015	สาวิตรี	วิไลโรจน์	CE327

รูปแบบบรรทัดฐานที่สอง (Second Normal Form: 2NF)

- ความสัมพันธ์ใด ๆ จะเป็น 2NF ก็ต่อเมื่อ
 - ความสัมพันธ์นั้นเป็น 1NF และ
 - Attribute ที่ไม่ใช่กุญแจ (non-key attribute) มีการพึ่งพิงเชิงฟังก์ชันแบบเต็มกับกุญแจหลัก
 - ต้องไม่มี Partial Dependency (การขึ้นต่อกันบางส่วน)
- สรุปก็คือ Second Normal form : 2NF เป็นการขจัด Attribute ที่ ไม่ขึ้นกับทั้งส่วนของคีย์หลักออกไป เพื่อให้ Attribute อื่นทั้งหมดขึ้นตรงกับส่วนที่เป็นคีย์หลักทั้งหมดเท่านั้น

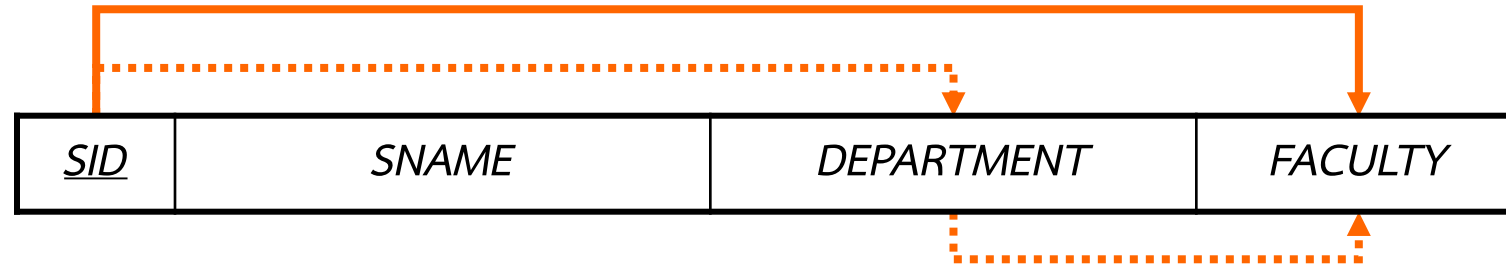
ตัวอย่าง



รูปแบบบรรทัดฐานที่สาม (Third Normal Form: 3NF)

- ความสัมพันธ์ใด ๆ จะเป็น 3NF ก็ต่อเมื่อ
 - ความสัมพันธ์นั้นเป็น 2NF และ
 - ไม่มี Transitive FD ระหว่าง primary key กับ non-key attribute
 - Attribute ที่ไม่ใช่ Primary key ต้องไม่ขึ้นต่อกันเอง

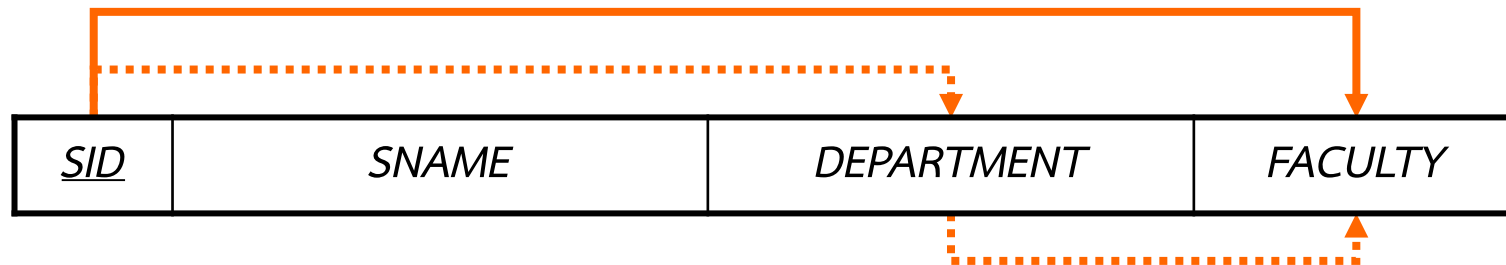
Transitive FDs



- ***SID* → *SNAME***
- ***SID* → *DEPARTMENT***
- ***SID* → *FACULTY***
- ***DEPARTMENT* → *FACULTY***

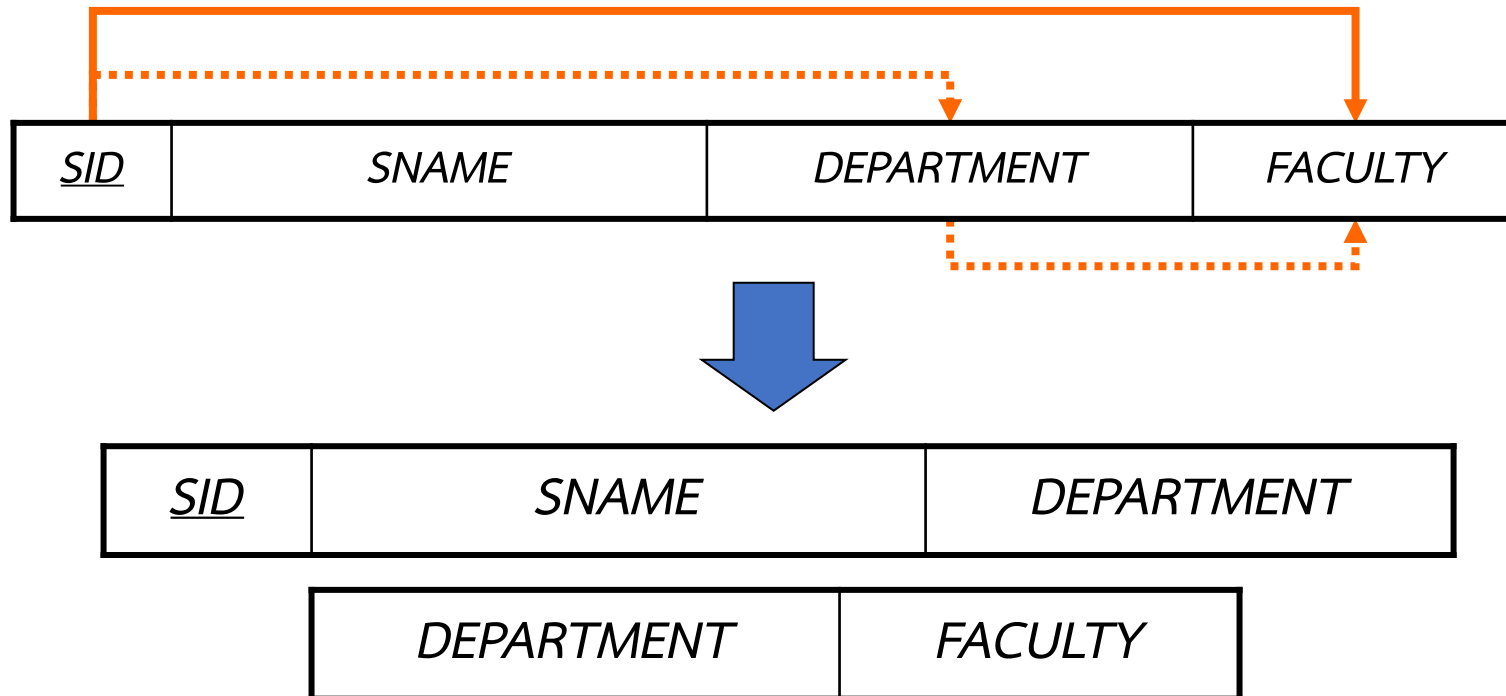
Example

- ความสัมพันธ์นี้ไม่เป็น 3NF เนื่องจากมี Transitive FD ระหว่าง primary key กับ non-key attribute



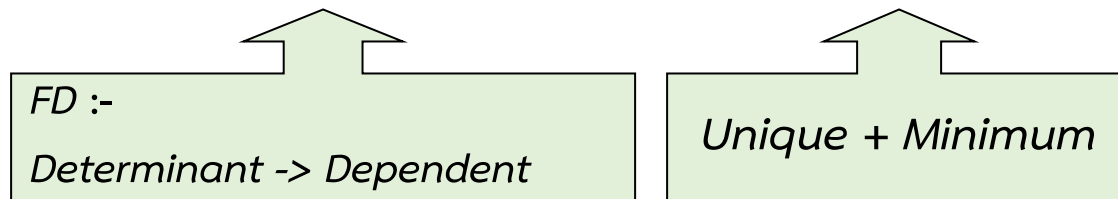
ทำให้เป็น 3FD...

- ต้องกำจัด Transitive FD ระหว่าง primary key กับ non-key attribute



Boyce/Codd Normal Form: BCNF

- ความสัมพันธ์ใด ๆ จะเป็น BCNF ก็ต่อเมื่อ
 - ทุก ๆ *determinant* ต้องเป็น *candidate key*



Conclusion

Database

DBMS

ER Diagram

Convert ER Diagram to Relational Database

Normalization