

CSC10006 – Introduction to Database

Chapter 3

Relational Data Model

(Mô hình dữ liệu quan hệ)

Instructor: Vũ Thị Mỹ Hằng, Ph.D. (vtmhang@fit.hcmus.edu.vn)

Lab Assistants: Tiết Gia Hồng, M.Sc. (tghong@fit.hcmus.edu.vn)

Lương Hán Cơ, M.Sc. (lhco@fit.hcmus.edu.vn)

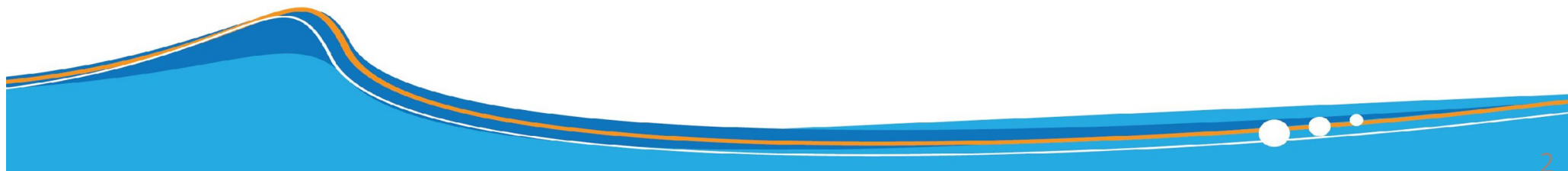


KHOA CÔNG NGHỆ THÔNG TIN
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

fit@hcmus

Content

- ☐ Recall: Database Design Process
- ☐ Relational Data Model
- ☐ ER-to-Relational Mapping
- ☐ Practice



Content

- ☐ Recall: Database Design Process
- ☐ Relational Data Model
- ☐ ER-to-Relational Mapping
- ☐ Practice



Database Design Process



Analyze & Design

Data



Functionalities, Interfaces



HOW to represent data structures and their characteristics, relations, constraints, etc.?

Data models
Entity Relationship Model
Relational Model
Object-oriented Model
Network Model
...

Implementation



Database



Application

Database Design Process

HOW to represent data structures and their characteristics, relations, constraints, etc.?

Real
world



Data Modeling

DATA MODEL

utilizes



DBMS

Conceptual

Easy to understand
for users

Logical

Understood by users
but not too far from
physical organization

Physical

Understood and
executed by computers

ER Model

(Mô hình thực thể-kết hợp)

Relational Data Model

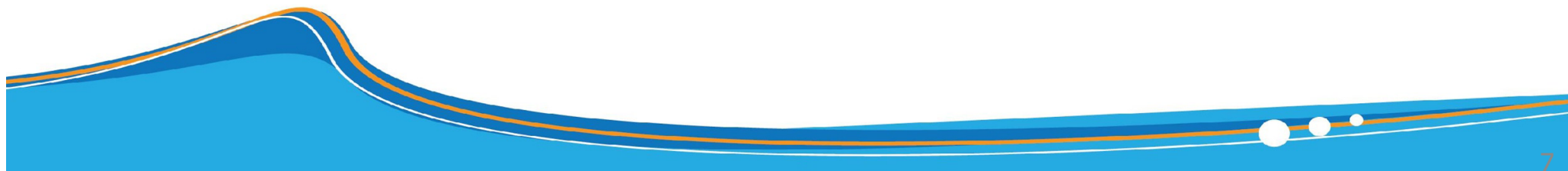
(Mô hình dữ liệu quan hệ)

Content

- Recall: Database Design Process
- **Relational Data Model**
 - Main Concepts
 - Characteristics of Relations
 - Integrity Constraints
- ER-to-Relational Mapping

Content

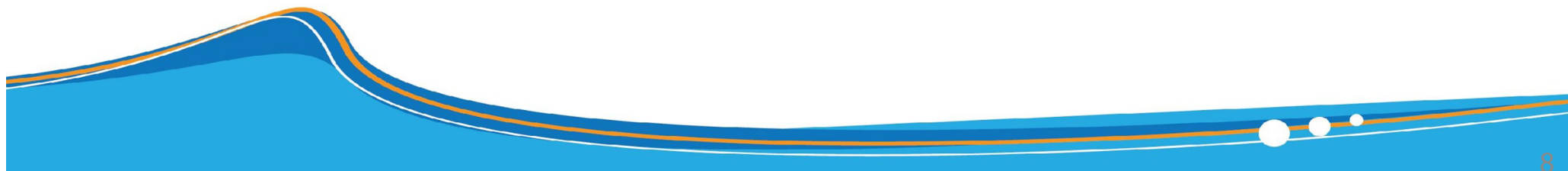
- ❑ Recall: Database Design Process
- ❑ **Relational Data Model**
 - **Main Concepts**
 - Characteristics of Relations
 - Integrity Constraints
- ❑ ER-to-Relational Mapping



Relational (Data) Model

Mô hình (dữ liệu) quan hệ

- ❑ Was introduced by E. F. Codd, 1970.
- ❑ Represents data in the form of **relations** (tables) based on mathematical foundation (set theory).
- ❑ Used in many DBMS.
 - Commercial: Oracle, SQL Server, DB2 ...
 - Open source: MySQL, PostgreSQL, ...



Relation – Attribute – Tuple

Quan hệ – Thuộc tính – Bộ (dữ liệu)

- ❑ The relational data model represents data in the form of relations based on mathematical foundation.
- ❑ A relation is a **named**, two-dimensional table of data, with rows (tuples) and **named** columns (attributes).
 - Relation \Leftrightarrow Table.
 - Tuples in a relation \Leftrightarrow Rows of a table.
 - Attributes of a relation \Leftrightarrow Columns of a table.

The relation **Employee** with four attributes (columns) and six tuples (rows).

Employee

staffNo	sName	position	salary
SL21	John White	Manager	30000
SG37	Ann Beech	Assistant	12000
SG14	David Ford	Supervisor	18000
SA9	Mary Howe	Assistant	9000
SG5	Susan Brand	Manager	24000
SL41	Julie Lee	Assistant	9000

Relation – Attribute – Tuple

Quan hệ – Thuộc tính – Bộ (dữ liệu)

Informal Definition

- A relation may be thought of as a set of rows.
- A relation may alternately be thought of as a set of columns.
- Each row represents a fact that corresponds to a real-world entity or relationship.

The relation **Employee** with
four attributes (columns)
and six tuples (rows).

Employee

staffNo	sName	position	salary
SL21	John White	Manager	30000
SG37	Ann Beech	Assistant	12000
SG14	David Ford	Supervisor	18000
SA9	Mary Howe	Assistant	9000
SG5	Susan Brand	Manager	24000
SL41	Julie Lee	Assistant	9000

Domain of Attributes

Miền giá trị

- ❑ Domain of an attribute: Set of values that may be assigned to an attribute.
- ❑ All values appear in a column of a relation must be from the same domain.

Employee

staffNo	sName	position	salary
SL21	John White	Manager	30000
SG37	Ann Beech	Assistant	12000
SG14	David Ford	Supervisor	18000
SA9	Mary Howe	Assistant	9000
SG5	Susan Brand	Manager	24000
SL41	Julie Lee	Assistant	9000

Possible values of the attribute "Salary" is a set of **positive integer**.

Relational Database

Cơ sở dữ liệu quan hệ

- A relational database: A set of relations storing meaningful data for a particular business.
- E.g.,: The database Student Management consists of five relations: MonHoc, HocPhan, KetQua, DieuKien, SinhVien.

MÔN HỌC	TênMH	MãMH	SốTC	Khoa
	Khoa học máy tính	CS1310	4	CNTT
	Cấu trúc dữ liệu	CS3320	4	CNTT
	Toán rời rạc	MATH2410	3	TOÁN
	Cơ sở dữ liệu	CS3380	3	CNTT

KẾT QUẢ	MSSV	MãHP	Điểm
	17	112	10
	17	119	7
	8	85	6
	8	92	9

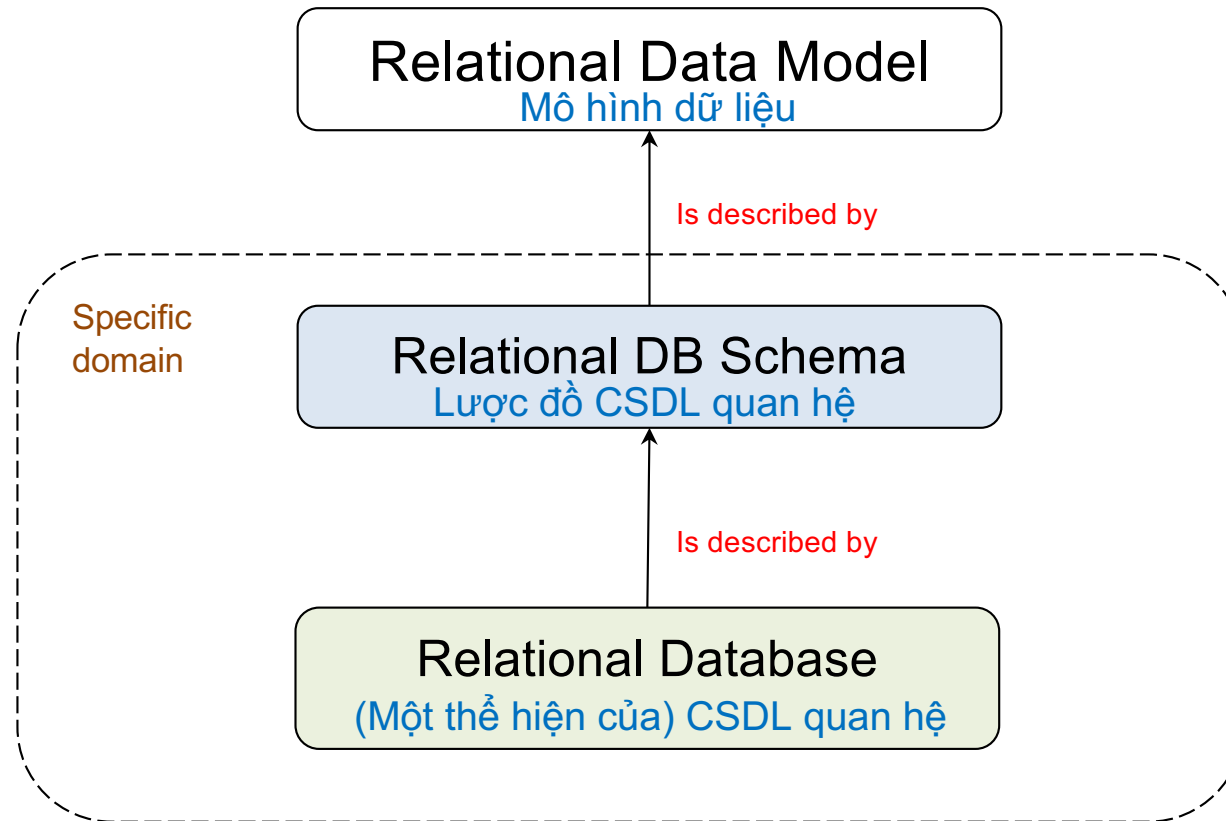
ĐIỀU KIỆN	MãMH	MãMH_Trước
	CS3380	CS3320
	CS3380	MATH2410
	CS3320	CS1310

HỌC PHẦN	MãHP	MãMH	Học Kỳ	Năm	Giáo Viên
	85	MATH2410	1	2008	Anh
	92	CS1310	1	2007	Tiền
	112	MATH2410	2	2008	Anh
	119	CS1310	2	2007	Tiền

SINH VIÊN	Tên	MSSV	Lớp	Khoa
	Trang	17	1	CNTT
	Ngọc	8	2	CNTT

Relational (Database) Schema

Lược đồ CSDL quan hệ



Relational (Database) Schema

fit@hcmus

Lược đồ CSDL quan hệ

The (relational)
database schema of
the database Student
Management

SINH VIÊN	TÊN SV	<u>MÃ SV</u>	LỚP	NGÀNH
MÔN HỌC	TÊN MH	<u>MÃ MH</u>	KHOA	TỈNH
ĐIỀU KIỆN	<u>MÃ MH TRƯỚC</u>	<u>MÃ MH</u>		
HỌC PHẦN	<u>MÃ HP</u>	GIÁO VIÊN	HỌC KỲ	NĂM
KQ_HỌC	<u>MÃ SV</u>	<u>MÃ HP</u>	ĐIỂM	

Is described by

Database Student Management

MÔN HỌC	Tên MH	Mã MH	Số TC	Khoa
	Khoa học máy tính	CS1310	4	CNTT
	Cấu trúc dữ liệu	CS3320	4	CNTT
	Toán rời rạc	MATH2410	3	TOÁN
	Cơ sở dữ liệu	CS3380	3	CNTT

KẾT QUẢ	MSSV	Mã HP	Điểm
	17	112	10
	17	119	7
	8	85	6
	8	92	9

ĐIỀU KIỆN	Mã MH	Mã MH Trước
	CS3380	CS3320
	CS3380	MATH2410
	CS3320	CS1310

HỌC PHẦN	Mã HP	Mã MH	Học Kỳ	Năm	Giáo Viên
	85	MATH2410	1	2008	Anh
	92	CS1310	1	2007	Tiên
	112	MATH2410	2	2008	Anh
	119	CS1310	2	2007	Tiên

SINH VIÊN	Tên	MSSV	Lớp	Khoa
	Trang	17	1	CNTT
	Ngọc	8	2	CNTT

Relation Schema

Lược đồ quan hệ

The (relational)
database schema
consists a set of
relation schemas.

SINH VIÊN	TÊN SV	<u>MÃ SV</u>	LỚP	NGÀNH
MÔN HỌC	TÊN MH	<u>MÃ MH</u>	KHOA	TỈNH
ĐIỀU KIỆN	<u>MÃ MH TRƯỚC</u>	<u>MÃ MH</u>		
HỌC PHẦN	<u>MÃ HP</u>	GIÁO VIÊN	HỌC KỲ	NĂM
KQ_HỌC	<u>MÃ SV</u>	<u>MÃ HP</u>	ĐIỂM	

Is described by

Database Student Management

MÔN HỌC	Tên MH	Mã MH	Số TC	Khoa
	Khoa học máy tính	CS1310	4	CNTT
	Cấu trúc dữ liệu	CS3320	4	CNTT
	Toán rời rạc	MATH2410	3	TOÁN
	Cơ sở dữ liệu	CS3380	3	CNTT

KẾT QUẢ	MSSV	Mã HP	Điểm
	17	112	10
	17	119	7
	8	85	6
	8	92	9

ĐIỀU KIỆN	Mã MH	Mã MH Trước
	CS3380	CS3320
	CS3380	MATH2410
	CS3320	CS1310

HỌC PHẦN	Mã HP	Mã MH	Học Kỳ	Năm	Giáo Viên
	85	MATH2410	1	2008	Anh
	92	CS1310	1	2007	Tiên
	112	MATH2410	2	2008	Anh
	119	CS1310	2	2007	Tiên

SINH VIÊN	Tên	MSSV	Lớp	Khoa
	Trang	17	1	CNTT
	Ngọc	8	2	CNTT

Formal Definition & Notation

- ❑ A **relation schema** R , denoted by $R(A_1, A_2, \dots, A_n)$, is made up of a **relation name** R and a **list of attributes**, A_1, A_2, \dots, A_n .
- ❑ Each attribute A_i has possible values defined in some domain D in the relation schema R . D is called the domain of A_i and is denoted by $DOM(A_i)$.
- ❑ The **degree (or arity) of a relation** is the **number of attributes** n of its relation schema.

Formal Definition & Notation

- A **relation** of the relation schema $R(A_1, A_2, \dots, A_n)$, also denoted by $r(R)$, is a set of n-tuples: $r = \{t_1, t_2, \dots, t_m\}$.
- Each t is an ordered list of n-values: $t = \langle v_1, v_2, \dots, v_n \rangle$
 - Where each value v_i , $1 \leq i \leq n$, is an element of $\text{DOM}(A_i)$ or is a special NULL value.
 - $r(R) \subseteq (\text{DOM}(A_1) \times \text{DOM}(A_2) \times \dots \times \text{DOM}(A_n))$

The i^{th} -value of t is represented by $t.A_i$ or $t[i]$.

KHOA (MÃKHOA, TÊNKHOA, NĂM TL, PHÒNG, ..., NGÀY NHẬN CHỨC)

	MÃKHOA	TÊNKHOA	NĂM TL	PHÒNG	ĐIỆN THOẠI	TRƯỞNG KHOA	NGÀY NHẬN CHỨC
t_1	CNTT	Công nghệ thông tin	1995	B11	0838123456	002	20/02/2005
t_2	VL	Vật lý	1976	B21	0838223223	005	18/09/2003
t_3	SH	Sinh học	1980	B31	0838454545	004	11/10/2000
t_4	HH	Hóa học	1980	B41	NULL	007	15/10/2001

Summary of Notations

- The relation schema R of the degree n
 - $R(A_1, A_2, \dots, A_n)$
- Relations
 - r, q, s
- Tuples
 - t, u, v
- Domain of the attribute A
 - $\text{DOM}(A)$
- The value at the attribute A of the tuple t
 - $t.A$ hay $t[A]$

Example

GIÁOVIÊN (MÃGV, HỌTÊN, LƯƠNG, PHÁI, NGÀY SINH, SỐNHÀ, ĐƯỜNG, QUẬN, THÀNH PHỐ, GVQLCM, MÃBM)

GV_ĐT (MÃGV, ĐIÊNTHOẠI)

BỘMÔN (MÃBM, TÊNBM, PHÒNG, ĐIÊNTHOẠI, TRƯỞNGBM, MÃKHOA, NGÀYNHẬNCHỨC)

KHOA (MÃKHOA, TÊNKHOA, NĂM TL, PHÒNG, ĐIÊNTHOẠI, TRƯỞNGKHOA, NGÀYNHẬNCHỨC)

ĐỀTÀI (MÃĐT, TÊNĐT, KINH PHÍ, CẤP QL, NGÀYBD, NGÀYKT, MÃCĐ, GVCNĐT)

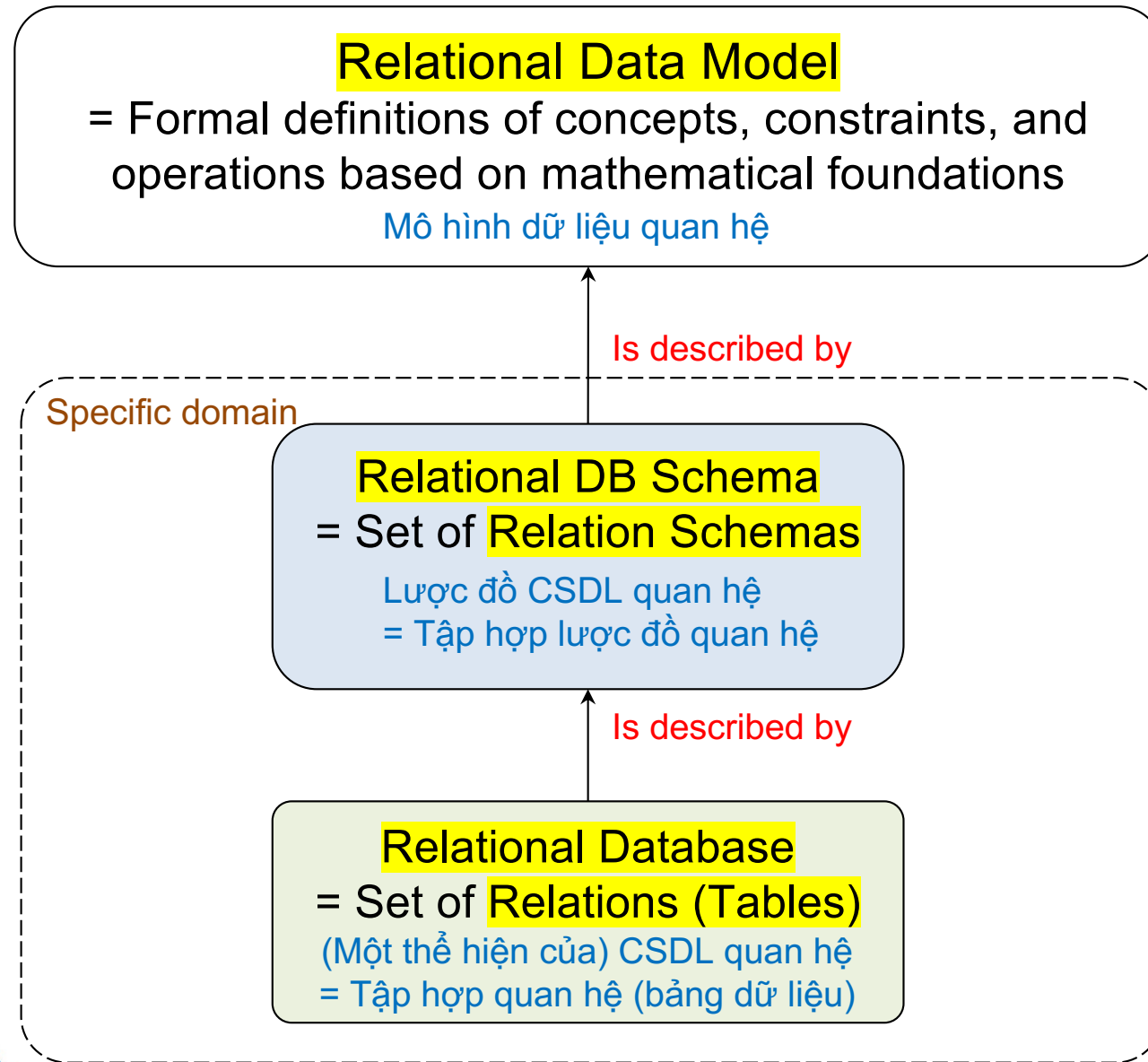
CHỦĐỀ (MÃCĐ, TÊN CĐ)

CÔNGVIỆC (MÃĐT, STT, TÊN CV, NGÀYBD, NGÀYKT)

THAMGIADT (MÃGV, MÃĐT, STT, PHỤC CẤP, KẾT QUẢ)

- Degree of the schema KHOA is 7.
- Domain of the attribute MÃKHOA is “String”.
- Domain of the attribute NĂM TL is “Integer”.

Summary of Main Concepts



Content

- Recall: Database Design Process
- **Relational Data Model**
 - Main Concepts
 - **Characteristics of Relations**
 - Integrity Constraints
- ER-to-Relational Mapping

Characteristics of Relations

Đặc trưng của quan hệ

- ❑ Allow distinguishing relations from nonrelation tables:
 - Each **relation** (or table) in a database has a **unique name**.
 - **No multivalued** attributes are allowed in a relation.
 - Each **attribute** (or column) within a table has a **unique name**.
 - The **order of columns** (left to right) is **insignificant**.
 - **No two rows** in a relation can be **identical**.
 - The **order of rows** (top to bottom) is **insignificant**.

Content

- Recall: Database Design Process
- **Relational Data Model**
 - Main Concepts
 - Characteristics of Relations
 - **Integrity Constraints**
- ER-to-Relational Mapping

Integrity Constraints

Ràng buộc toàn vẹn (trên dữ liệu)

- ❑ Specify (business) rules to maintain the accuracy and integrity of data stored in relations.
 - **Keys Constraint:** No two rows in a relation can be identical.
 - **Referential Constraint.** Some (attribute) values of a relation must match a (set of) value(-s) in another relation or must be null. This allows maintaining consistency among the rows of the two relations.
 - **Domain Constraint:** All values appear in a column of a relation must be from the same domain.

Keys (Entity Integrity)

Khoá (Đảm bảo tính toàn vẹn của dữ liệu)

(Siêu khoá)

- A **super key** of a relation schema $R = \{ A_1, A_2, \dots, A_n \}$ is a set of attributes from R ($S \subseteq R$) with the property that no two tuples t_1 and t_2 in any legal relation state r of R will have $t_1[S] = t_2[S]$.

- **Uniqueness:** Identify each row in the relation.

- Example:

SINHVIEN (MSSV, HOTEN, PHAI, CCCD, NGSINH)

- $\{MSSV\}$ is super key.
- $\{MSSV, HOTEN\}$ is super key.
- All attributes of a relation constitute a super key.

Keys (Entity Integrity)

Khoá (Đảm bảo tính toàn vẹn của dữ liệu)

(Khoá) (Khoá ứng viên)

- A **key** (or **candidate key**) K is a **super key** with the additional property that removal of any attribute from K will cause K not to be a super key anymore.

- **Uniqueness**: Identify each row in the relation.
- **Minimality**: $\neg \exists X' \subset X$ such as X' is a super key.
- Determined from the meaning of attributes.
- A relation can have many candidate keys.

- Example:

SINHVIEN (MSSV, HOTEN, PHAI, CCCD, NGSINH)

- {MSSV} is key.
- {CCCD} is a key.
- {MSSV, HOTEN} is not a key.

Keys (Entity Integrity)

Khoá (Đảm bảo tính toàn vẹn của dữ liệu)

(Khoá chính)

- A **primary key** is the (candidate key) selected for implementation to identify uniquely the tuples in relation.
 - Should choose the key with **few attributes**.
 - The attributes of primary key are **not null**.
 - A relation has **only one primary key**.
 - A primary key **can be a list of attributes** (a composite key).
 - The primary key is **underlined**.
- Example:
SINHVIEN (MSSV, HOTEN, PHAI, CCCD, NGSINH)
 - {MSSV} **is the unique primary key** of the relation SINHVIEN.
 - CCCD could be a secondary (unique) key at the implementation level.

Referential Constraint

Ràng buộc tham chiếu

- An attribute A of a tuple in relation R receives a value from an attribute B of relation S.
- **R** is called referencing relation (bảng tham chiếu).
- **S** is called referenced relation (bảng được tham chiếu).

S

TÊNKHOA	MÃKHOA
CNTT	5
VL	4
TOÁN	1

R

MÃBM	TÊNBM	PHÒNG	ĐIỆNTHOẠI	MÃKHOA
HTTT	Hệ thống thông tin	I84	838125125	5
CNPM	Công nghệ phần mềm	I82	838126126	4
VLDT	Vật lý điện tử	F203	838127127	4
VLƯD	Vật lý ứng dụng	F205	838128128	5

Foreign Key

Khoá ngoại

- ❑ A specific referential constraint.
- ❑ Supposing two relation schemas:
 - $R_1(A_1, A_2, \dots, A_n)$ and $R_2(B_1, B_2, \dots, B_m)$
 - $PK \subseteq \{A_1, \dots, A_n\}$ is the primary key of R_1 , $FK \subseteq \{B_1, \dots, B_m\}$
- ❑ FK is a foreign key of R_2 if:
 - Attributes in FK have the same domains as attributes in the primary key PK: $DOM(A_i) = DOM(B_i), 1 \leq i \leq n$
 - $\forall t_2 \text{ of } R_2, \exists t_1 \in R_1, t_2[FK] = t_1[PK].$

Foreign key

BỘMÔN (MÃBM, TÊNBM, PHÒNG, ĐIỆNTHOẠI, TRƯỞNGBM, MÃKHOA, NGÀYNHẬNCHỨC)

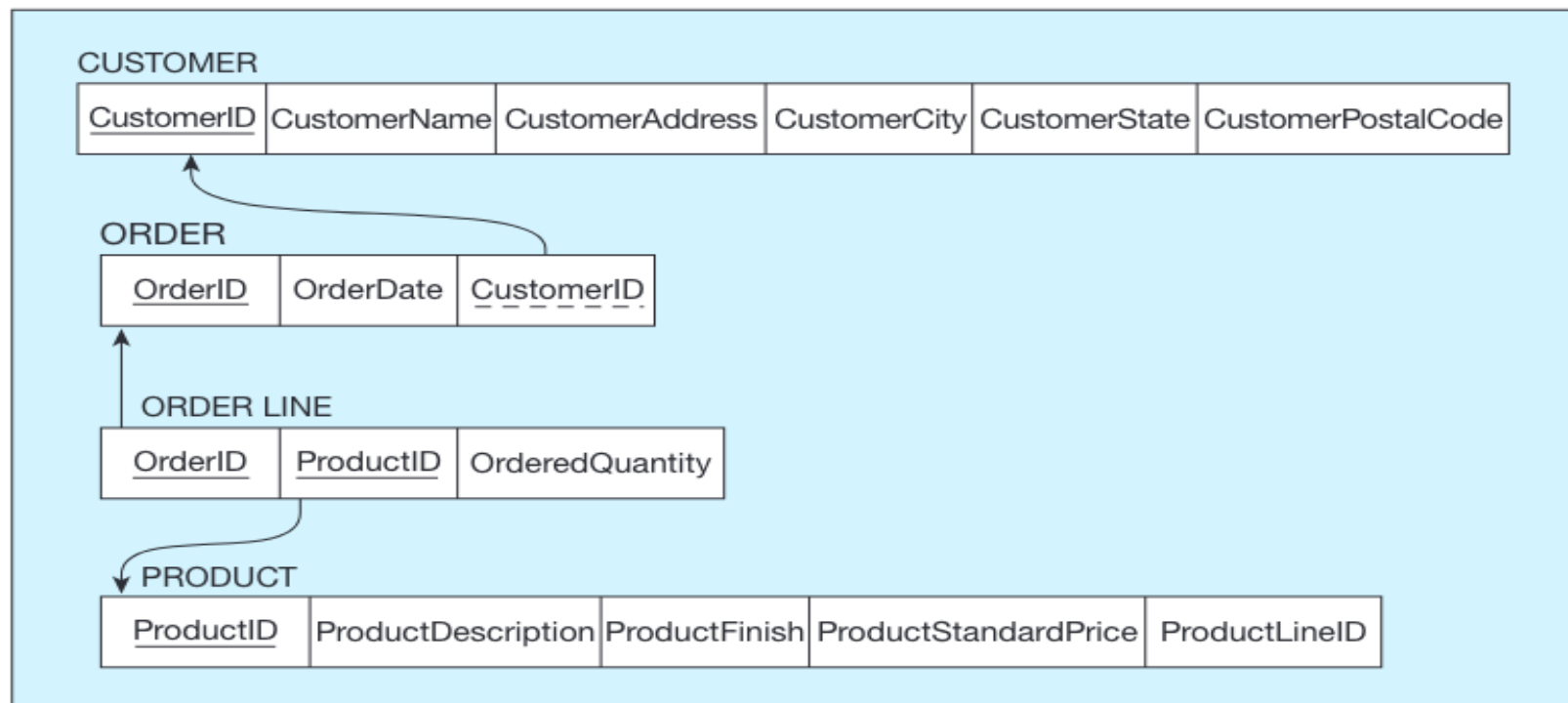
KHOA (MÃKHOA, TÊNKHOA, NĂMTL, PHÒNG, ĐIỆN THOẠI, TRƯỞNGKHOA, NGÀYNHẬNCHỨC)

Primary key

Foreign Key

Khoá ngoại

- ❑ An attribute can be part of a primary key and foreign key at the same time.
- ❑ A foreign key can refer to the primary key of the same relation.
- ❑ Many foreign keys can refer to the same primary key.
- ❑ Referential integrity constraint = foreign key constraint.



Domain Constraint

Ràng buộc miền dữ liệu

TABLE 4-1 Domain Definitions for INVOICE Attributes

Attribute	Domain Name	Description	Domain
CustomerID	Customer IDs	Set of all possible customer IDs	character: size 5
CustomerName	Customer Names	Set of all possible customer names	character: size 25
CustomerAddress	Customer Addresses	Set of all possible customer addresses	character: size 30
CustomerCity	Cities	Set of all possible cities	character: size 20
CustomerState	States	Set of all possible states	character: size 2
CustomerPostalCode	Postal Codes	Set of all possible postal zip codes	character: size 10
OrderID	Order IDs	Set of all possible order IDs	character: size 5
OrderDate	Order Dates	Set of all possible order dates	date: format mm/dd/yy
ProductID	Product IDs	Set of all possible product IDs	character: size 5
ProductDescription	Product Descriptions	Set of all possible product descriptions	character: size 25
ProductFinish	Product Finishes	Set of all possible product finishes	character: size 15
ProductStandardPrice	Unit Prices	Set of all possible unit prices	monetary: 6 digits
ProductLineID	Product Line IDs	Set of all possible product line IDs	integer: 3 digits
OrderedQuantity	Quantities	Set of all possible ordered quantities	integer: 3 digits

Concept Mapping to Physical Design

Khái niệm ở mức cài đặt trong HQTCSDL (DBMS)

- A relation is a **named** (two-dimensional) **table of data**, with **rows** (tuples) and **named columns** (attributes).
- Relation \Leftrightarrow **Table**.
- Tuples in a relation \Leftrightarrow **Rows** of a table.
- Attributes of a relation \Leftrightarrow **Columns** of a table.

The relation **Employee** with
four attributes (columns)
and six tuples (rows).

staffNo	sName	position	salary
SL21	John White	Manager	30000
SG37	Ann Beech	Assistant	12000
SG14	David Ford	Supervisor	18000
SA9	Mary Howe	Assistant	9000
SG5	Susan Brand	Manager	24000
SL41	Julie Lee	Assistant	9000

Content

- ☐ Recall: Database Design Process
- ☐ Relational Data Model
- ☐ ER-to-Relational Mapping
- ☐ Practice

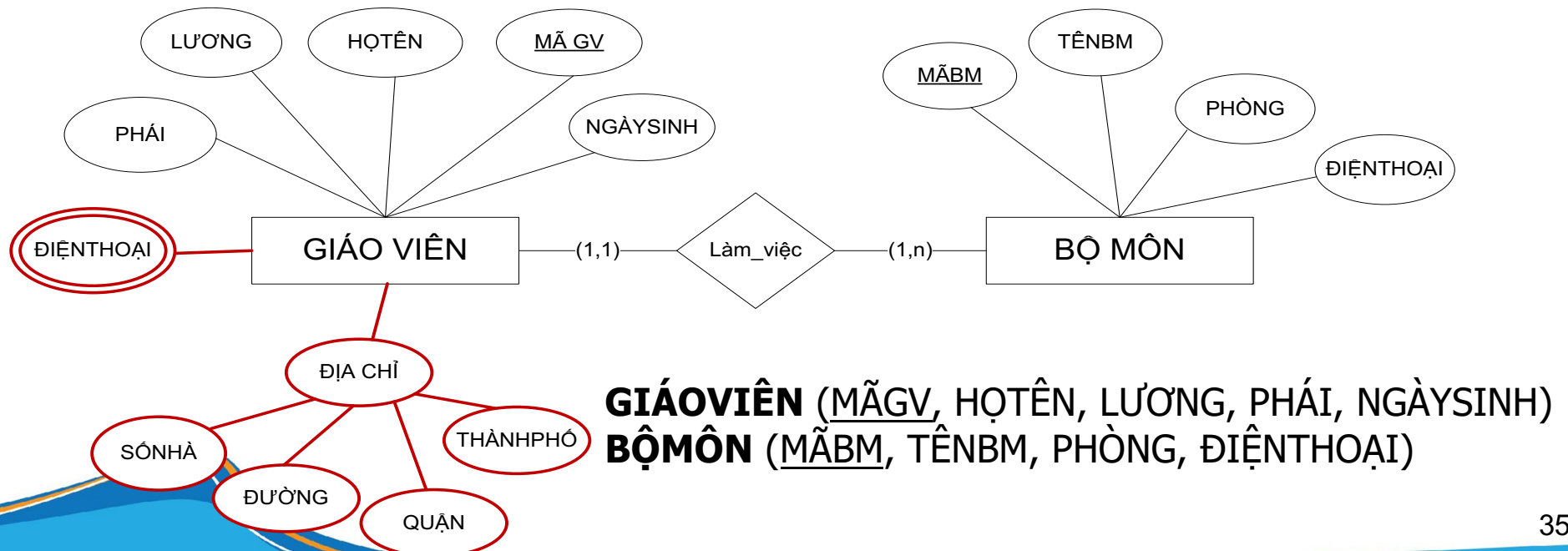
Content

- ☐ Recall: Database Design Process
- ☐ Relational Data Model
- ☐ ER-to-Relational Mapping
- ☐ Practice

R#1: Regular Entity Set

For each Entity set (except weak entity set), create a corresponding relation:

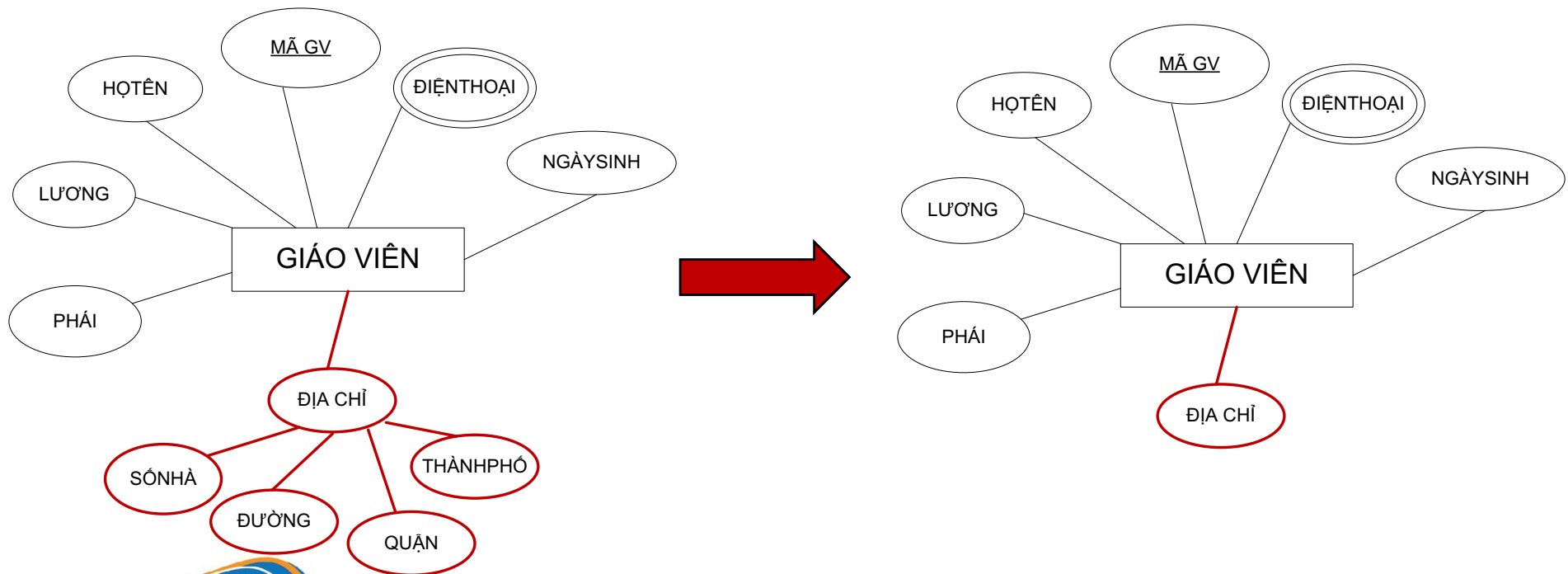
- ❓ Same name
- ❓ Same set of attribute (except composite and multi-valued attribute)



R#2: Composite Attribute

We have 2 case:

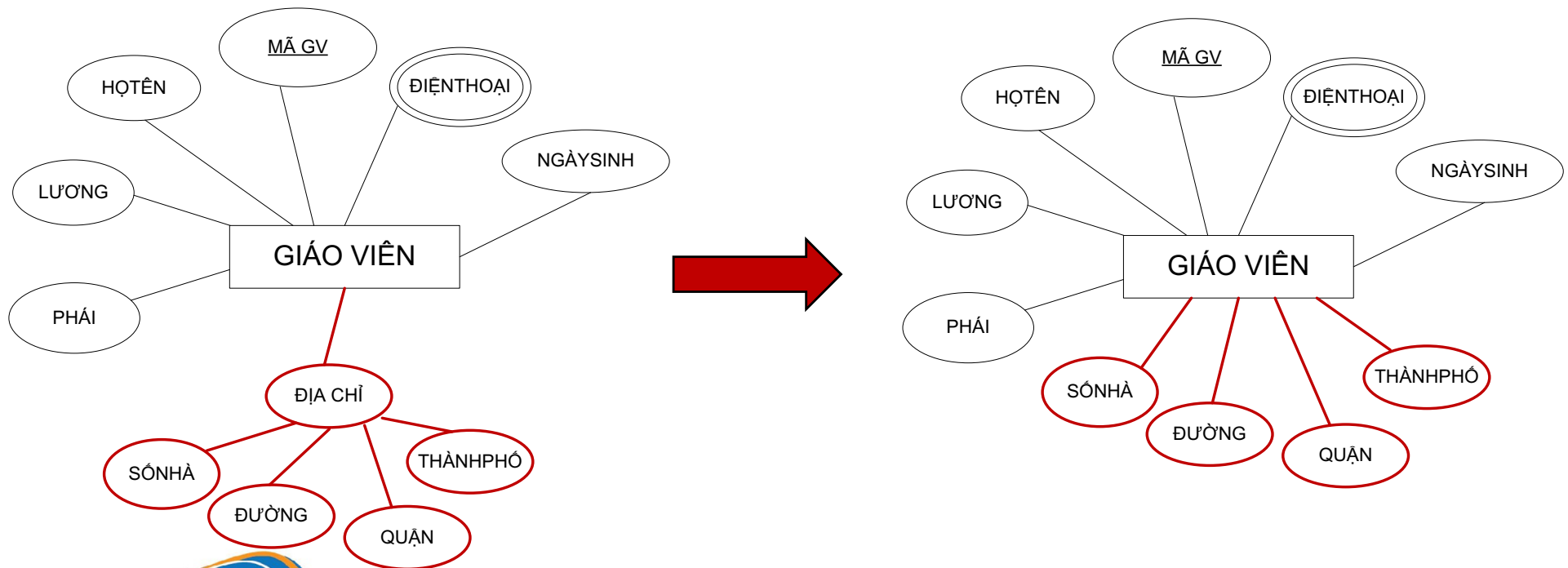
Case 1: a composite attribute is transformed into a single-valued attribute



GIÁOVIÊN (MÃGV, HỌTÊN, LƯƠNG, PHÁI, NGÀYSINH, *ĐỊA CHỈ*)

R#2: Composite Attribute (cont.)

Case 2: a composite attribute is transformed into a set of single valued attributes



GIÁOVIÊN (MÃGV, HỌTÊN, LƯƠNG, PHÁI, NGÀYSINH, *SỐNHÀ, ĐƯỜNG, QUẬN, THÀNHHPỐ*)₃₇

R#3: Multi-valued Attribute

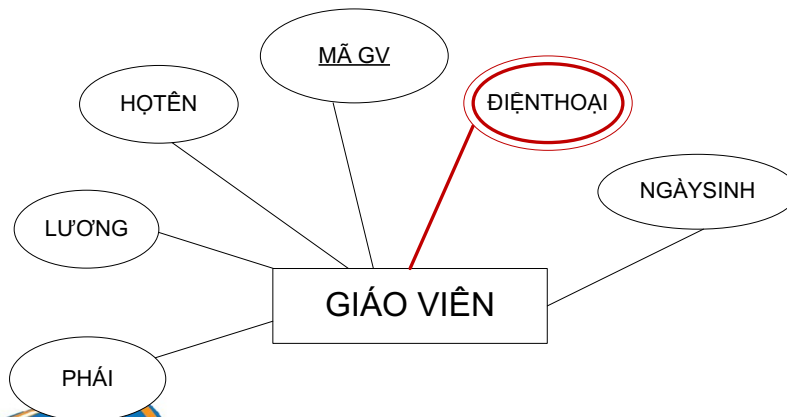
For each multi-valued attribute A of relation R , create a relation S with the following attribute:

- primary key of R
- attribute A (as a single-valued attribute)

Primary key of new relation S :

either A

or $A + \text{primary key of } R$



ĐIỆN THOẠI (SỐ ĐT, MÃ GV)

hoặc

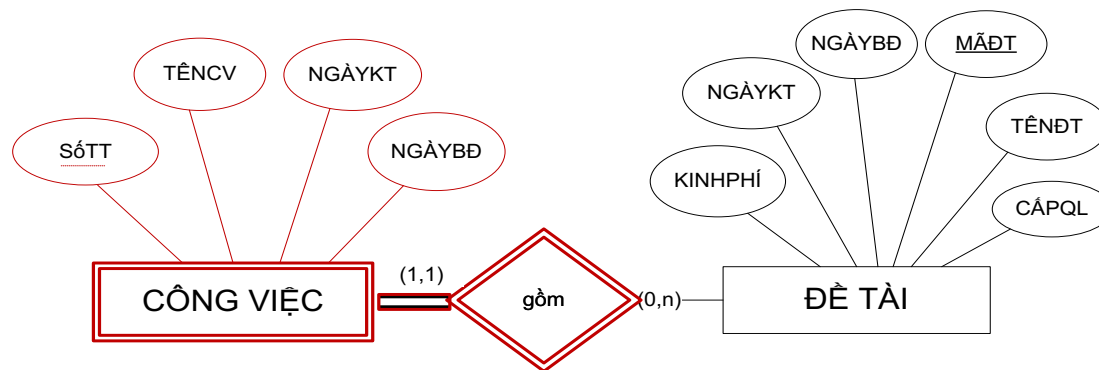
ĐIỆN THOẠI (SỐ ĐT, MÃ GV)

GIÁO VIÊN (MÃ GV, HỌ TÊN, LƯƠNG, PHÁI, NGÀY SINH, SỐ NHÀ, ĐƯỜNG, QUẬN, THÀNH PHỐ)³⁸

R#4: Weak Entity Set

For each entity set, create a corresponding relation:

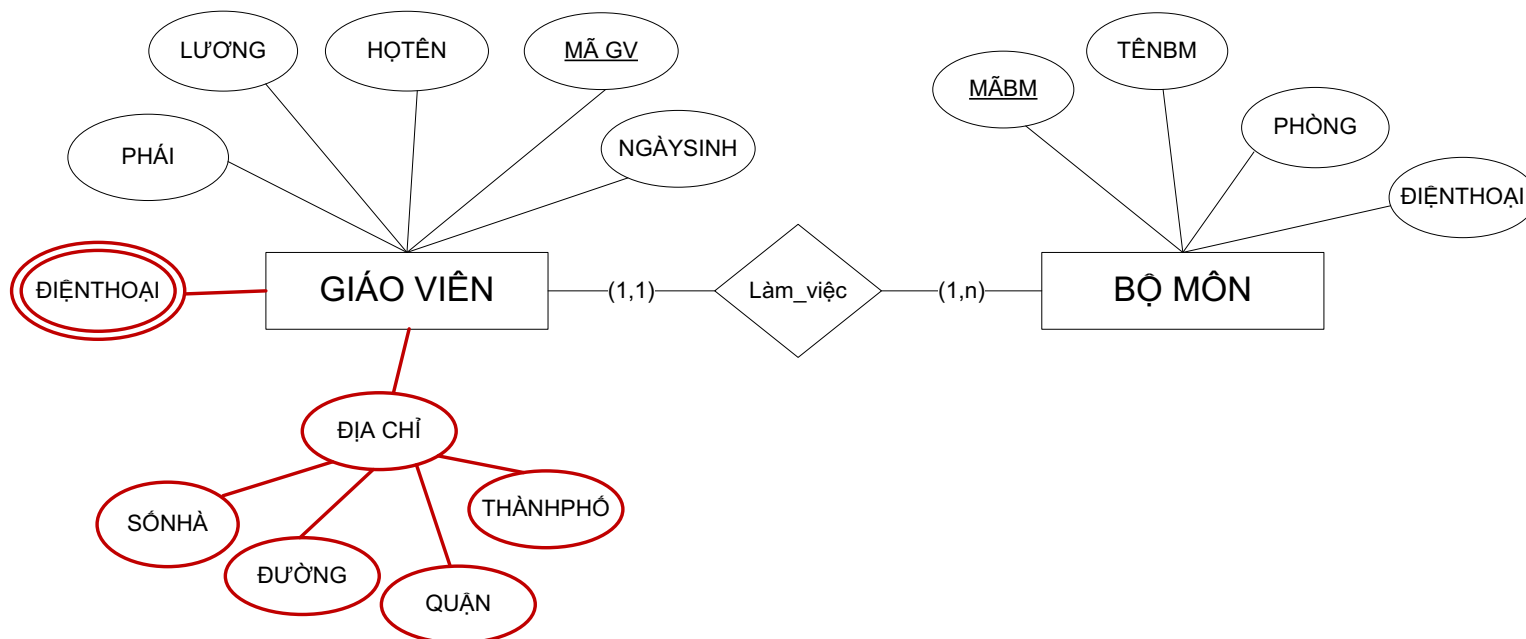
- ❓ Same name
- ❓ Same set of attribute (except composite and multi valued attribute)
- ❓ Adding the key attributes of the entity sets which the weak entity set depends on
- ❓ The key of relation consists of
weak key attributes
key attributes of the entity sets which the weak entity set depends on



ĐỀ TÀI (MÃ DT, TÊN DT, KINH PHÍ, CẤP QL, NGÀY BD, NGÀY KT)
 CÔNG VIỆC (MÃ DT, STT, TÊN CV, NGÀY BD, NGÀY KT)

R#5: Relationship Set 1-n

Adding the key of the many-relation to the one-relation

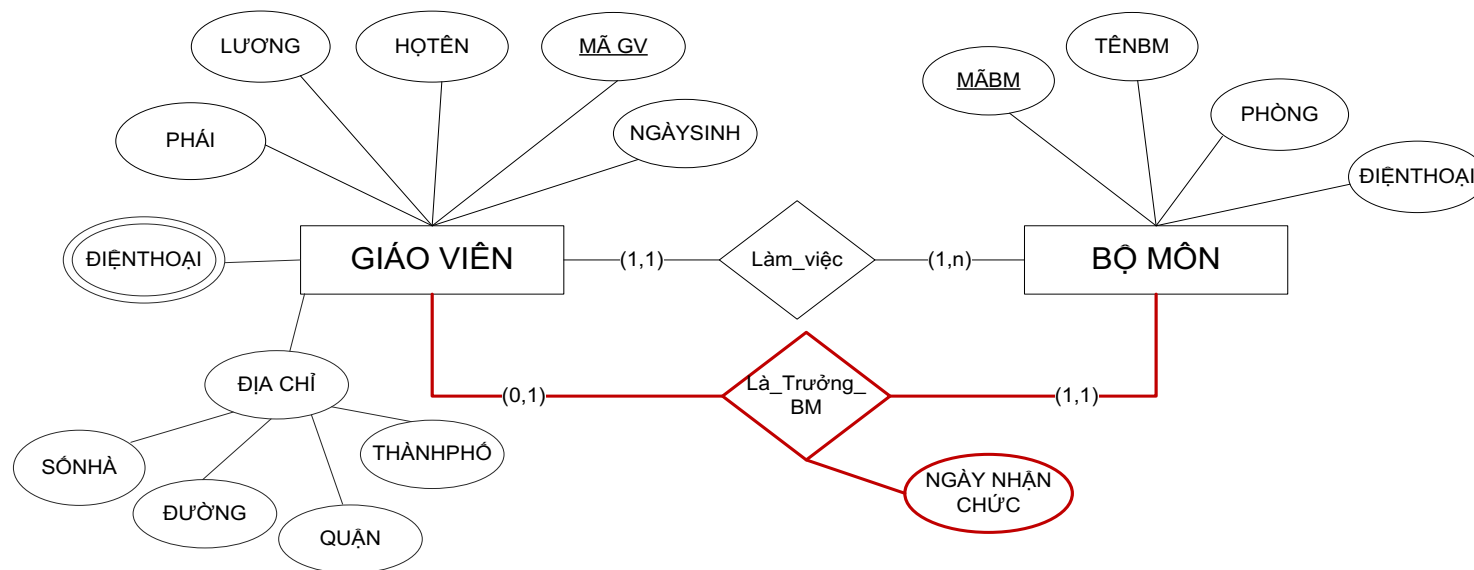


GIÁOVIÊN (MÃGV, HỌTÊN, LƯƠNG, PHÁI, NGÀYSINH, SÓNHÀ, ĐƯỜNG, QUẬN, THÀNHPHỐ, **MÃBM**)
BỘMÔN (MÃBM, TÊNBM, PHÒNG, ĐIỆNTHOẠI)

R#6: Relationship set: 1-1

Either adding the key of a relation to another relation + the attributes on the relationship

Or adding the key to both relations + the attributes on the relationship

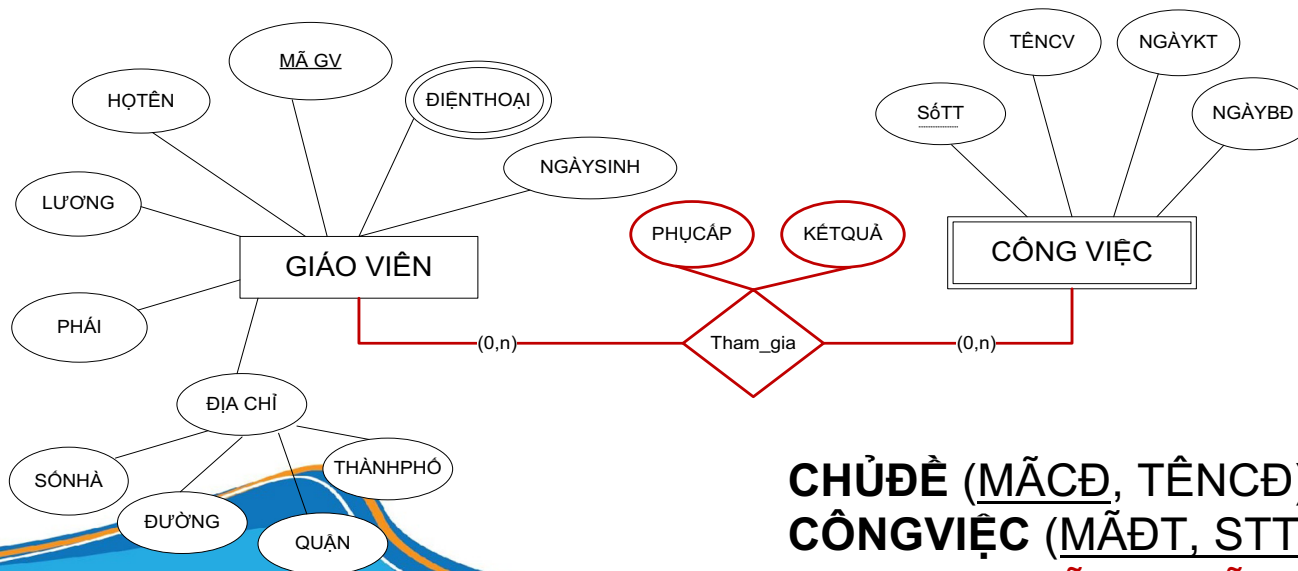


T GIÁO VIÊN (MÃ_GV, HỌTÊN, LƯƠNG, PHÁI, NGÀY_SINH, SÓNHÀ, ĐƯỜNG, QUẬN, THÀNH PHỐ, MÃ_BM)
S BỘ MÔN (MÃ_BM, TÊN_BM, PHÒNG, ĐIỆN THOẠI, NGÀY_NHẬN_CHỨC, **TRƯỞNG_BM**)

R#&: Relationship Set: n-n

Create a new relation

- ❓ Relation name is the name of the relationship
- ❓ Set of attribute consists of
 - the key attributes of connected entity sets
 - The attributes on the relationship
- ❓ Primary key consists of
 - the key attributes of connected entity sets



CHỦ ĐỀ (MÃ Đ, TÊN Đ)

CÔNG VIỆC (MÃ ĐT, STT, TÊN CV, NGÀY BD, NGÀY KT)

THAM GIA (MÃ GV, MÃ ĐT, STT, PHỤ CẤP, KẾT QUẢ)

Content

- ☐ Recall: Database Design Process
- ☐ Relational Data Model
- ☐ ER-to-Relational Mapping
- ☐ Practice

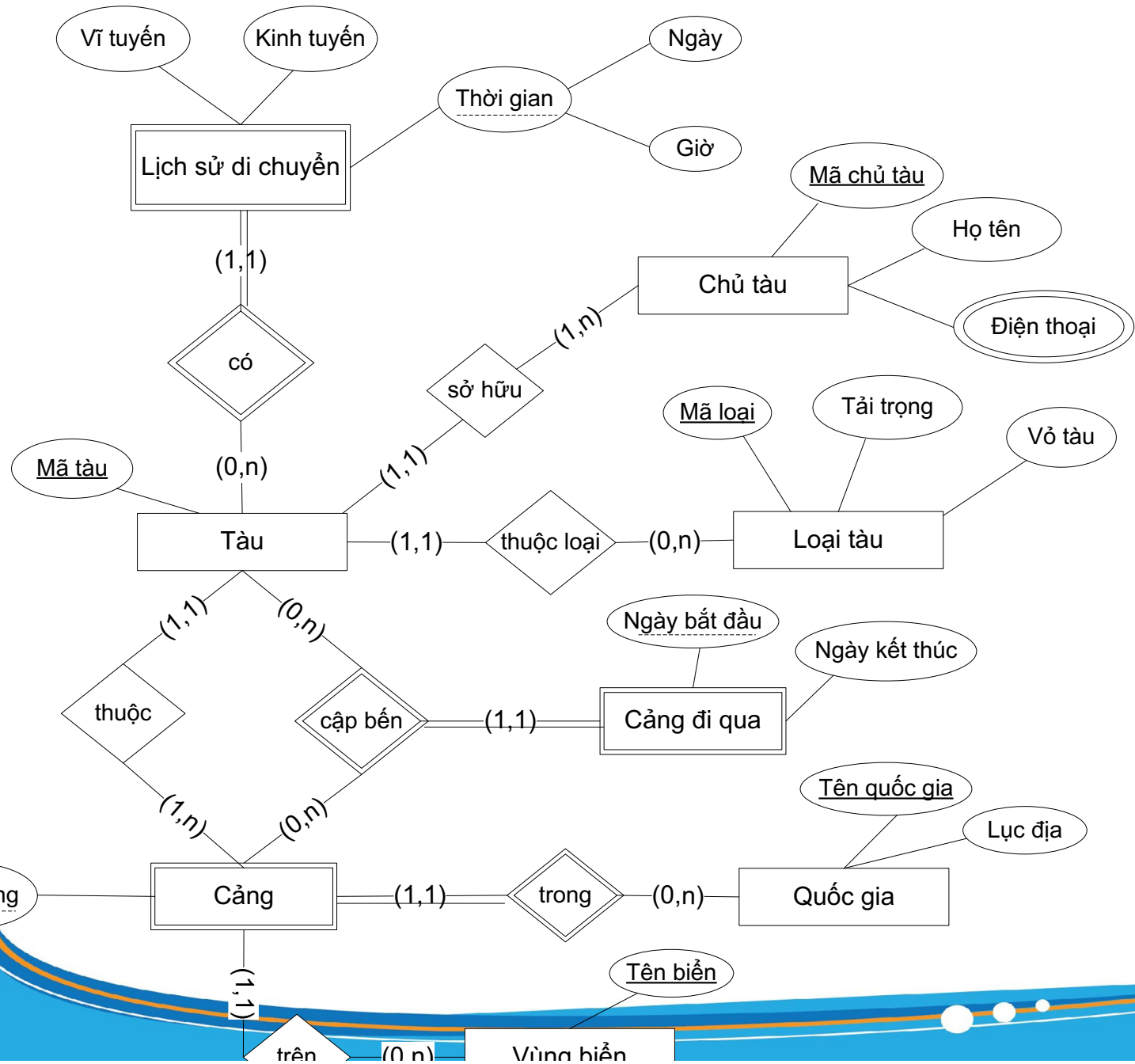
Practice #1

Consider the following relations for a database that keeps track of student enrollment in courses and the books adopted for each course:

- ❑ STUDENT(SSN, Name, Major, Bdate)
- ❑ COURSE(Course#, Cname, Dept)
- ❑ ENROLL(SSN, Course#, Quarter, Grade)
- ❑ BOOK_ADOPTION(Course#, Quarter, Book_ISBN)
- ❑ TEXT(Book_ISBN, Book_Title, Publisher, Author)

Draw a relational schema diagram specifying the foreign keys for this schema.

Practice #2



Practice #3

