

# CƠ SỞ DỮ LIỆU (DATABASE)



BỘ MÔN HỆ THỐNG THÔNG TIN

# TỔNG QUAN VỀ MÔN HỌC

- Môn học cung cấp những kiến thức về:
  - Kiến trúc hệ thống cơ sở dữ liệu.
  - Mô hình quan hệ-thực thể, và các nguyên tắc thiết kế cơ sở dữ liệu.
  - Mô hình quan hệ, đại số quan hệ, SQL và khung nhìn.
  - Phụ thuộc hàm và khóa, các ràng buộc toàn vẹn, các dạng chuẩn, và chuẩn hóa.
  - Tối ưu hóa câu hỏi truy vấn

# TỔNG QUAN VỀ MÔN HỌC...

- **Phân bổ thời gian:** Tổng số: 3TC 45 tiết
  - Lý thuyết: 30 tiết;
  - Bài tập, thảo luận: 15 tiết.
- **Yêu cầu:** Đối với sinh viên:
  - Dự lớp đầy đủ, Nộp bài tập lớn
  - Tham gia thảo luận, Dự kiểm tra và thi

# TÀI LIỆU THAM KHẢO

1. Đặng Thị Thu Hiền, *Cơ sở dữ liệu*, NXB Giao Thông Vận Tải, 2013.
2. Phạm Thị Hoàng Nhung, *Cơ sở dữ liệu I*, Trường Đại học Thủy Lợi.
3. Phạm Thị Hoàng Nhung, *Cơ sở dữ liệu II*, Trường Đại học Thủy Lợi.
4. Gillenson, Mark L. : Fundamentals of database management systems //Mark L. Gillenson. - Hoboken, NJ ::Wiley,,2005

# ĐỀ CƯƠNG TỔNG QUAN MÔN HỌC

TT	Tên chương	Số tiết			
		Tổng số	Lý thuyết	Thảo luận, BT, TH	Tiểu luận, KTra
1	TỔNG QUAN HỆ THỐNG CƠ SỞ DỮ LIỆU	2	2		
2	GIỚI THIỆU THIẾT KẾ CƠ SỞ DỮ LIỆU	5	3	2	
3	MÔ HÌNH QUAN HỆ	8	6	2	
4	SQL	14	8	6	
5	PHỤC THUỘC HÀM VÀ KHÓA	7	5	2	
6	DẠNG CHUẨN VÀ CHUẨN HÓA	9	6	3	
	Cộng:	45	30	15	

# BÀI TẬP LỚN

- Mỗi sinh viên tự chọn một đề tài thực tế để làm
- Yêu cầu trình bày báo cáo:
  - Nội dung từ 15-25 trang
  - Font: Time New Roman, size 13, dẫn dòng 1.3 cm
  - Định dạng trang: Đầu cuối trang 2cm, lề trái 3.5, lề phải 2.

# BÀI TẬP LỚN ...

## ■ Nội dung báo cáo:

*Lời nói đầu*

*Mục lục:* yêu cầu đánh mục lục tự động

I. Xác định các quy tắc/ràng buộc, xây dựng mô hình thực thể liên kết ER

II. Chuyển từ mô hình thực thể liên kết sang mô hình quan hệ

III. Xác định khóa

IV. Chuẩn hóa lược đồ quan hệ thành dạng chuẩn 3NF hoặc BCNF

V. Câu lệnh truy vấn dữ liệu SQL

*Kết luận*

*Tài liệu tham khảo*

# ĐÁNH GIÁ MÔN HỌC

TT	Các hình thức đánh giá	Trọng số
1	QT= Bài tập, Chuyên cần, xây dựng bài, Kiểm tra, BTL  (BT ở nhà 20%, KT trên lớp 30%, Chuyên cần+ Thường xuyên 20%, BTL 30%)	0.5
2	Thi hết môn (THM): Thi viết	0.5
	Điểm môn học = QT x 0.4 + THM x 0.6	



## Chương 1

# TỔNG QUAN HỆ THỐNG CƠ SỞ DỮ LIỆU

# TỔNG QUAN VỀ HỆ CSDL

- Sự cần thiết của một CSDL
- Khái niệm CSDL, Hệ CSDL
- Hệ quản trị CSDL (DBMS – DataBase Management System)
- Mô hình dữ liệu (Data model)

# 1. SỰ CẦN THIẾT CỦA CSDL

## ■ CSDL đến từ đâu? Nó có vai trò ntn?



"Chỉ Thượng đế là đáng tin, mọi thứ khác đều phải dựa vào dữ liệu"

- **Từ các phương tiện xã hội**  
Nhìn thấu (insights) được hành vi và ý kiến của khách hàng của công ty.



**Each day:**  
230M tweets,  
2.7B comments to FB,  
86400 hours of video  
to YouTube

- **Từ máy móc**  
Thiết bị công nghiệp, các sensors và dụng cụ giám sát, web logs...



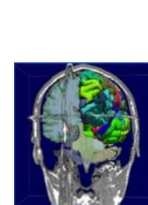
Large Hadron  
Collider generates  
40 terabytes/sec

- **Từ giao dịch kinh doanh**  
ID và giá cả sản phẩm, thanh toán, dữ liệu chế tạo và phân bố, ... ,



Amazon.com: \$10B in  
sales in Q3 2011, US  
pizza chain Domino's:  
1 million customers  
per day

- Nhiều loại khác



1 human  
brain at the  
micron level  
= 1 PetaByte



Large Hadron  
Collider,  
(PetaBytes/day)



Human Genomics  
= 7000 PetaBytes  
1GB / person



Printed materials in the Library of  
Congress = 10 TeraBytes



200 of  
London's  
Traffic  
Cams  
(8TB/day)

1 book = 1  
MegaByte



Family photo =  
586 KiloBytes

Kilo	$10^3$
Mega	$10^6$
Giga	$10^9$
Tera	$10^{12}$
Peta	$10^{15}$
Exa	$10^{18}$



All  
worldwide  
information  
in one year  
= 2  
ExaBytes

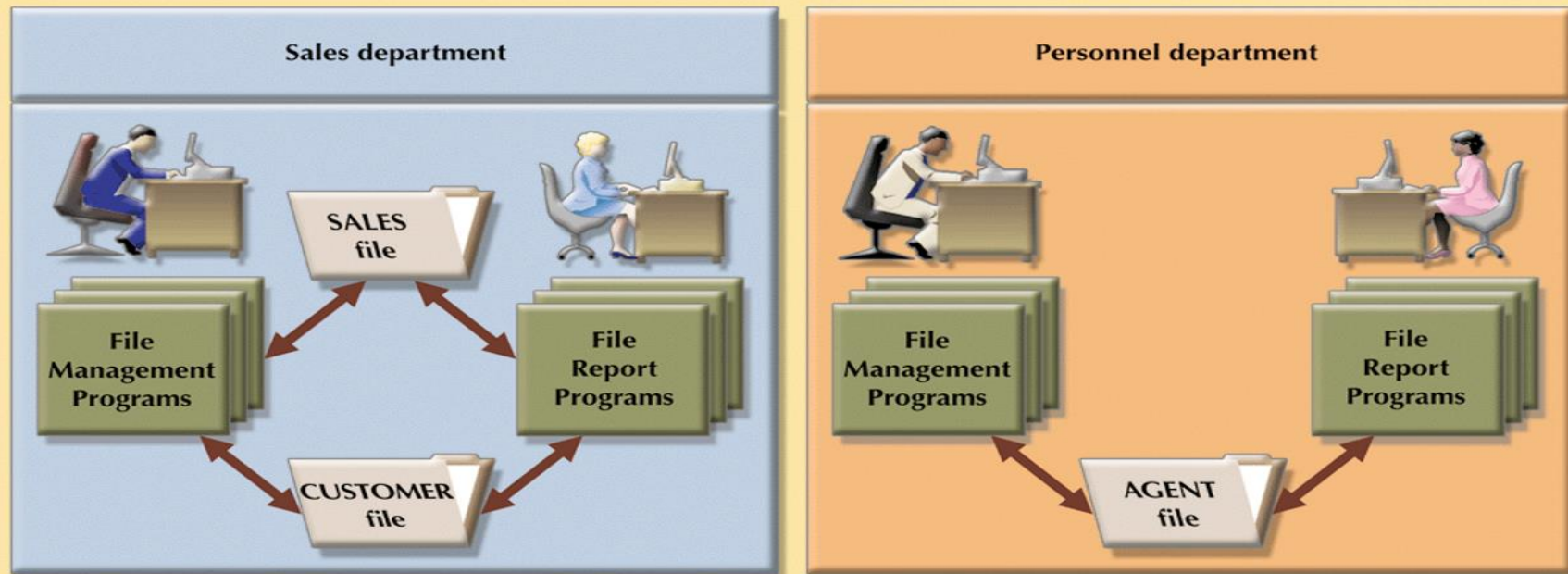


# 1. SỰ CẦN THIẾT CỦA CSDL...

- **Hệ thống các file cổ điển:** Được tổ chức riêng rẽ, phục vụ cho mục đích của đơn vị

FIGURE 1.7

A simple file system



SOURCE: Course Technology/Cengage Learning

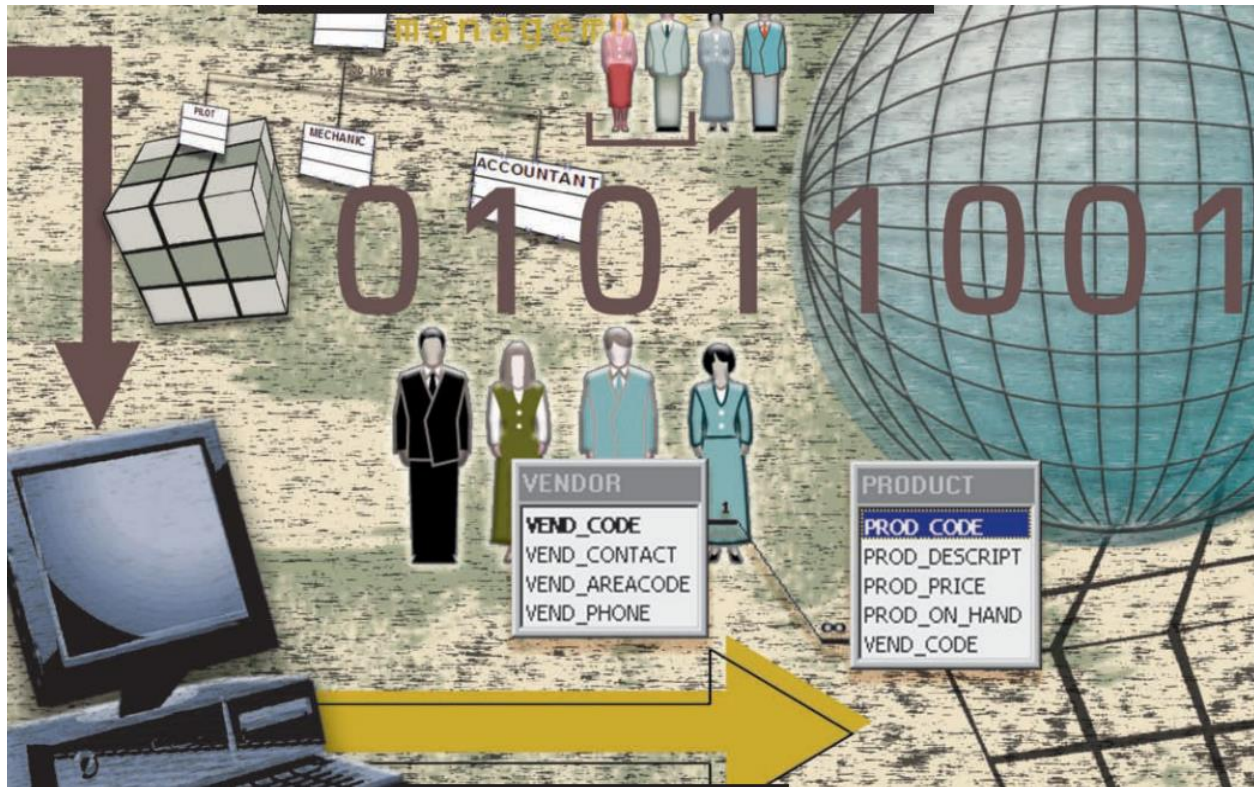
# 1. SỰ CẦN THIẾT CỦA CSDL...

- Hệ thống file là sự cải tiến của hệ thống thủ công (manual system)
- Ưu điểm: Đơn giản, nhanh chóng
- Nhược điểm:
  - Thông tin được tổ chức riêng rẽ nhiều nơi làm mất tính nhất quán, dư thừa dữ liệu.
  - Thiếu sự chia sẻ thông tin giữa các vị trí.
  - Không thực hiện được truy vấn phức tạp
  - Quản trị hệ thống khó và phức tạp
  - Khó mở rộng khi lập trình
  - Khó thay đổi cấu trúc.
  - Tính năng bảo mật thấp.



## 2. CƠ SỞ DỮ LIỆU (DATABASE), HỆ CSDL

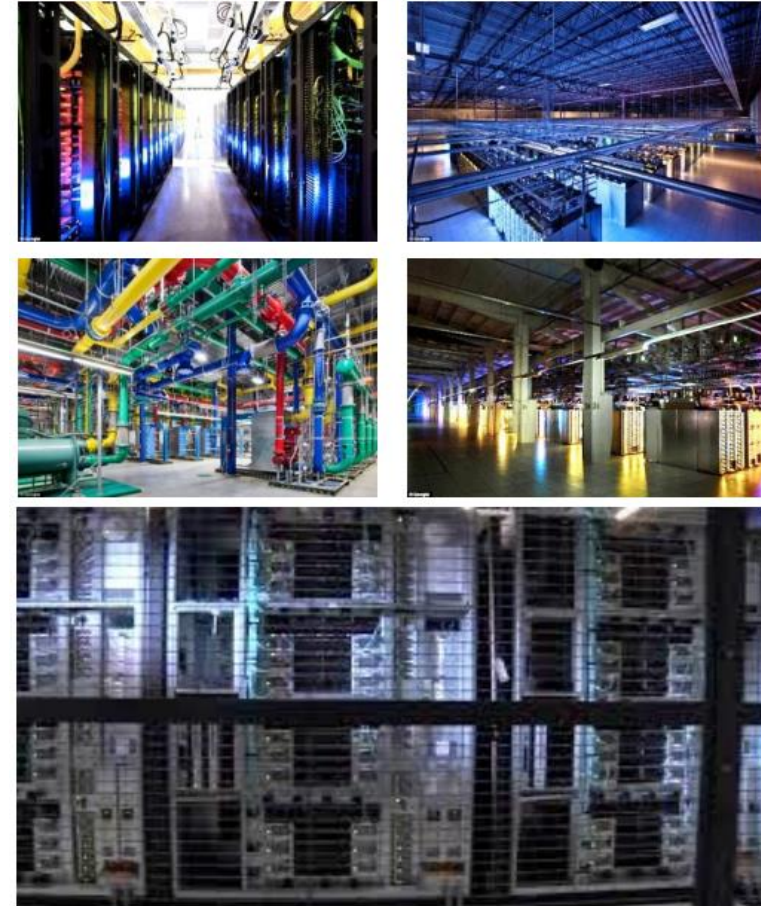
- **CSDL:** là một hệ thống các thông tin có cấu trúc được lưu trữ trên các thiết bị lưu trữ thông tin thứ cấp (như băng từ, đĩa từ,...), để thỏa mãn yêu cầu khai thác thông tin đồng thời của nhiều người sử dụng/nhiều chương trình ứng dụng với những mục đích khác nhau.



## 2. CƠ SỞ DỮ LIỆU, HỆ CSDL...

- **Hệ cơ sở dữ liệu (Database System)** gồm:
  - Phần cứng (Hardware)
  - Phần mềm (Software)
  - Người sử dụng (People)
  - Những thủ tục (Procedures)
  - Cơ sở dữ liệu (Database)

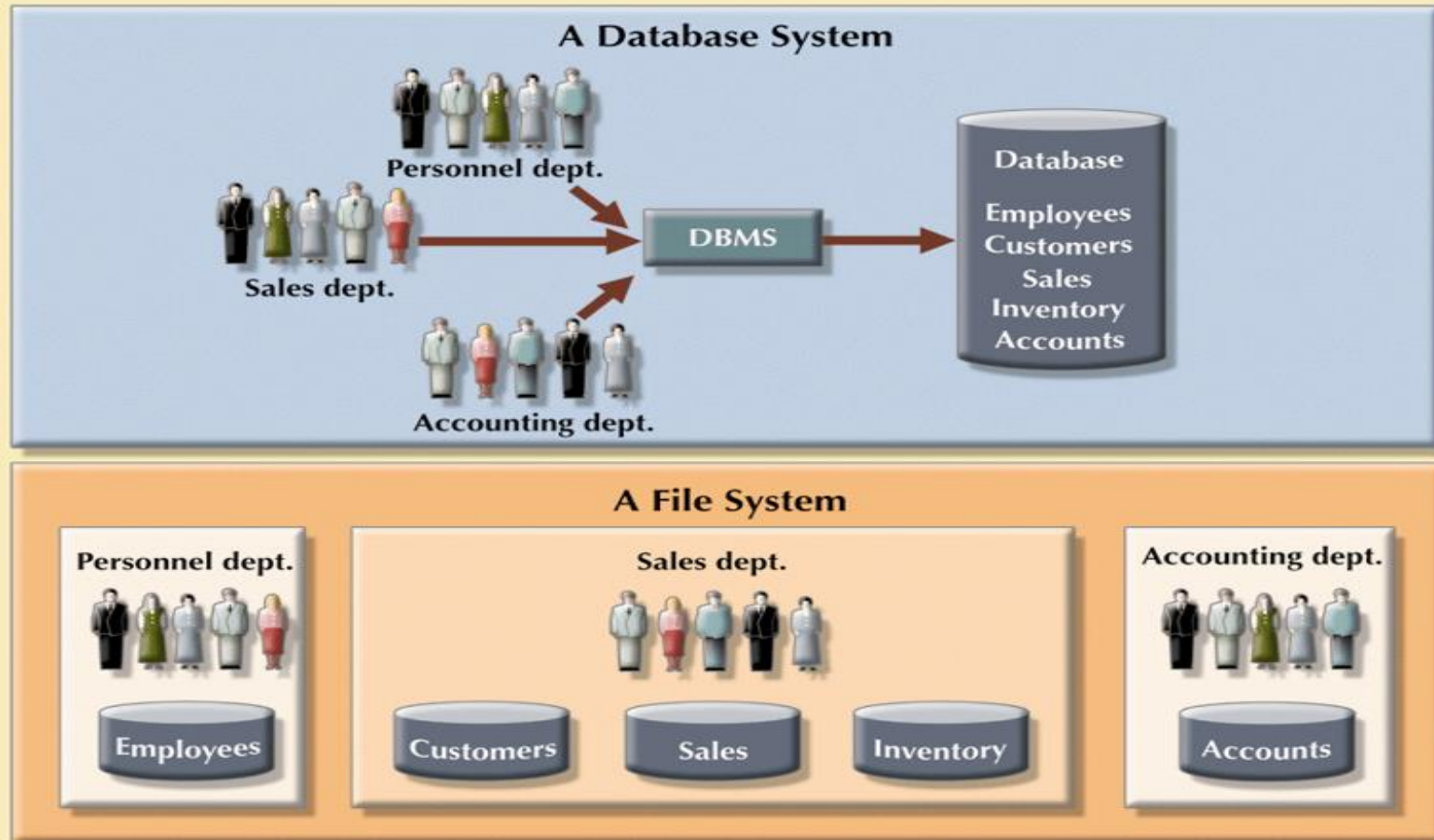
Google Data Center



## 2. CƠ SỞ DỮ LIỆU, HỆ CSDL...

**FIGURE 1.8**

**Contrasting database and file systems**



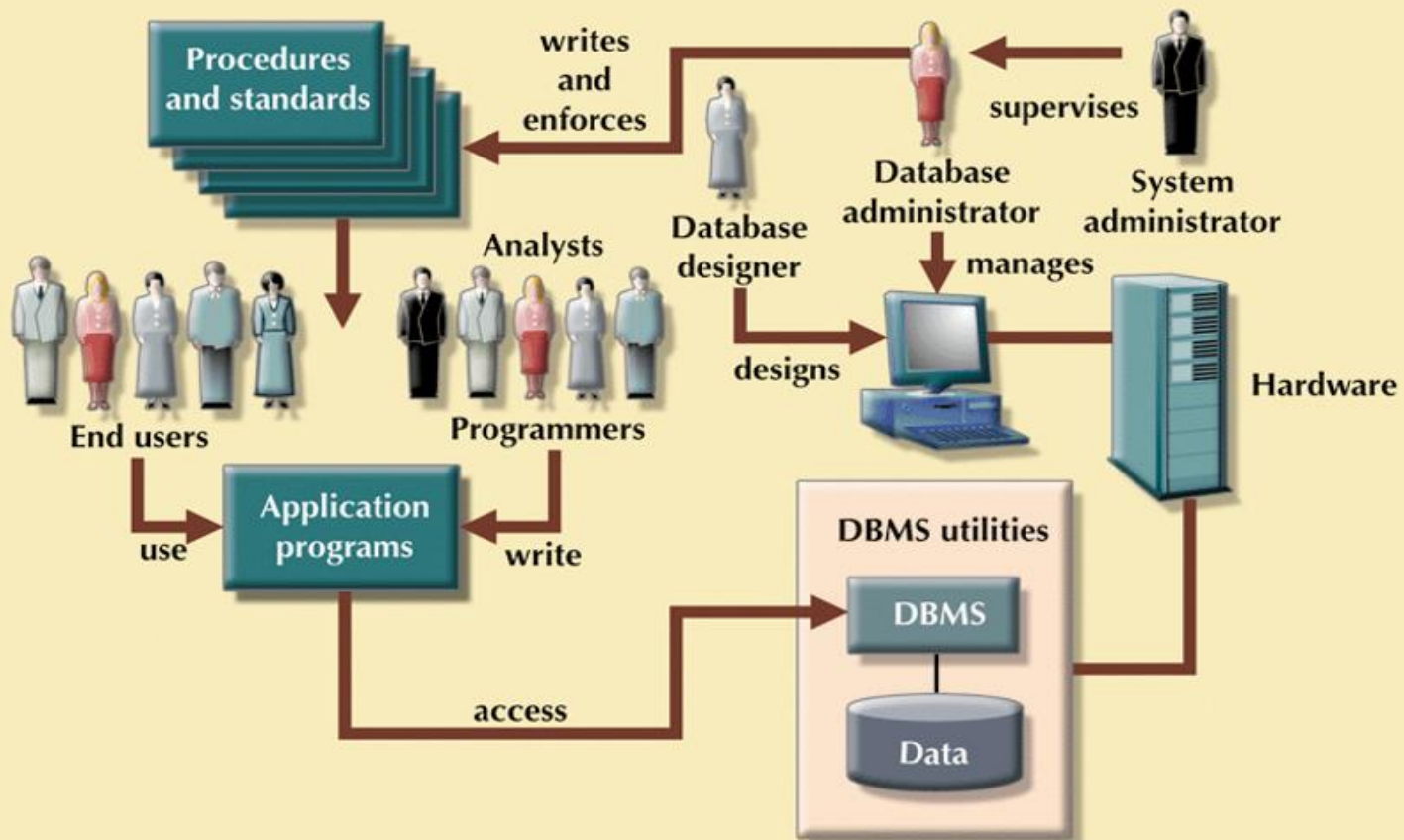
SOURCE: Course Technology/Cengage Learning



## 2. CƠ SỞ DỮ LIỆU, HỆ CSDL ...

FIGURE  
1.9

The database system environment



SOURCE: Course Technology/Cengage Learning

## 2. CƠ SỞ DỮ LIỆU, HỆ CSDL ...

- Phần cứng: Gồm các thiết bị vật lý của hệ thống
- Phần mềm:
  - Hệ điều hành; Hệ quản trị CSDL; Chương trình ứng dụng và phần mềm tiện ích
- Người dùng: Tất cả những người dùng hệ thống CSDL
  - Quản trị hệ thống; Thiết kế dữ liệu; Phân tích hệ thống, lập trình viên và Người sử dụng
- Thủ tục: Những quy tắc chung trong việc thiết kế và sử dụng hệ thống
- Dữ liệu: Những dữ liệu thực tế thu thập và lưu trữ trong CSDL

## 2. CƠ SỞ DỮ LIỆU, HỆ CSDL ...

- **Chú ý:**
  - CSDL phải là một tập hợp các thông tin mang tính hệ thống chứ không phải là các thông tin rời rạc, không có mối quan hệ với nhau.
  - Các thông tin này phải có cấu trúc và tập hợp các thông tin này phải có khả năng đáp ứng các nhu cầu khai thác của nhiều người sử dụng một cách đồng thời.

## 2. CƠ SỞ DỮ LIỆU, HỆ CSDL...

- Ưu điểm nổi bật của CSDL.
  - Giảm sự trùng lặp thông tin xuống mức thấp nhất và do đó bảo đảm được tính nhất quán và toàn vẹn dữ liệu.
  - Đảm bảo dữ liệu có thể được truy xuất theo nhiều cách khác nhau.
  - Khả năng chia sẻ thông tin cho nhiều người sử dụng và nhiều ứng dụng khác nhau....

## 2. CƠ SỞ DỮ LIỆU, HỆ CSDL...

- CSDL đặt ra những vấn đề cần giải quyết

### 1- Tính chủ quyền của dữ liệu.

- Do tính chia sẻ của CSDL nên chủ quyền của CSDL dễ bị xâm phạm.

### 2- Tính bảo mật và quyền khai thác thông tin.

- Cần phải có một cơ chế bảo mật và phân quyền hạn khai thác CSDL.

### 3- Tranh chấp dữ liệu.

- Cần phải có một cơ chế ưu tiên truy nhập dl, cơ chế giải quyết tình trạng khóa chết (DeadLock) trong quá trình khai thác cạnh tranh.

### 4- Đảm bảo dữ liệu khi có sự cố.

- Khi có sự cố như mất điện đột xuất, đĩa hỏng...

## 2. CÁC ĐỐI TƯỢNG SỬ DỤNG CSDL

- Những người sử dụng CSDL không chuyên về lĩnh vực tin học và CSDL.
- Các chuyên viên tin học biết khai thác CSDL.
- Những người quản trị CSDL. Họ là người tổ chức CSDL (khai báo cấu trúc CSDL, ghi nhận các yêu cầu bảo mật cho các dl cần bảo vệ ...).

### 3. HỆ QUẢN TRỊ CSDL - DBMS (DATABASE MANAGEMENT SYSTEM)

- **Hệ phần mềm quản trị CSDL**

- Hệ quản trị CSDL là hệ thống các phần mềm hỗ trợ CSDL giải quyết các vấn đề như tính chủ quyền, cơ chế bảo mật/phân quyền khai thác CSDL, giải quyết tranh chấp, và phục hồi dl khi có sự cố ...
- Một số DBMS: Visual FoxPro, MicroSoft Access, SQL-Server, MySQL, DB2, Sybase, Paradox, Informix, Oracle...

### 3. HỆ QUẢN TRỊ CSDL - DBMS ...

- **Quá trình phát triển các hệ quản trị CSDL:**

- DBMS ra đời năm 60s dựa trên mô hình DL phân cấp và mạng, có IMS của IBM dựa trên mô phân cấp.
- 1976, System-R ra đời mô hình đầu tiên dựa trên mô hình quan hệ của IBM.
- Từ năm 1980, IBM cho ra đời DBMS trên các máy Main Frame mang tên DB2, tiếp theo là Dbase, Sybase, Oracle, Informix, SQL-Server ...



### 3. HỆ QUẢN TRỊ CSDL - DBMS ...

- 1990 bắt đầu xây dựng các DBMS hướng đối tượng (Oriented Object DataBase Management System) như Orion, Illustra, Itasca, Objectstore, Versant, Oracle, Informix, ...
- Hầu hết các hệ này đều vẫn là quan hệ - hướng đối tượng, DBMS hướng đối tượng thuần nhất là ODMG ra đời vào năm 1996.

# 3. HỆ QUẢN TRỊ CSDL - DBMS ...

- Một số DBMS hỗ trợ DL lớn



# 3. HỆ QUẢN TRỊ CSDL - DBMS ...

- Một DBMS phải có:

1) Ngôn ngữ giao tiếp giữa người sử dụng (NSD) và CSDL, gồm:

- *Ngôn ngữ mô tả dữ liệu (Data Definition Language - DDL).*
- *Ngôn ngữ thao tác dữ liệu (Data Manipulation Language - DML)*
- *Ngôn ngữ truy vấn dữ liệu, hay ngôn ngữ hỏi đáp có cấu trúc (Structured Query Language - SQL)*
- *Ngôn ngữ quản lý dữ liệu (Data Control Language - DCL)*

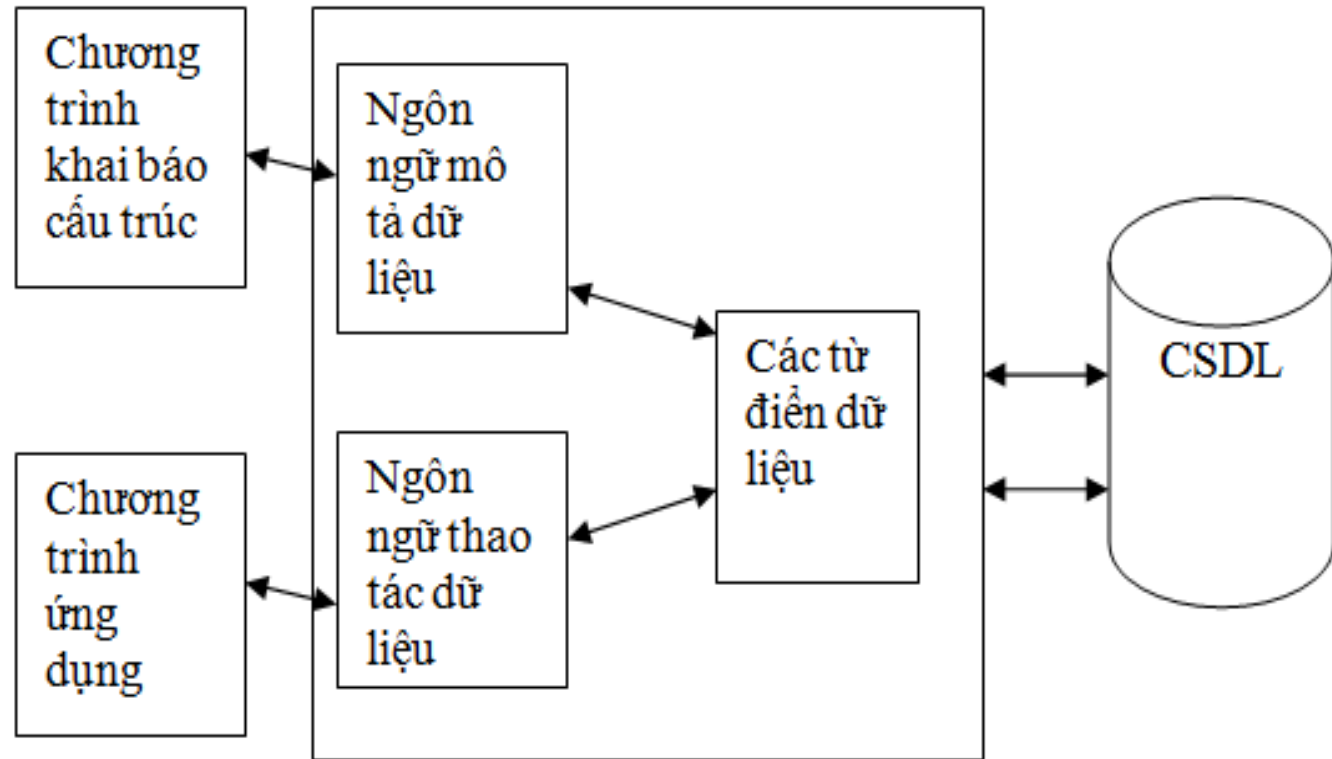
2) Từ điển dữ liệu (Data Dictionary) dùng để mô tả các ánh xạ liên kết, ghi nhận các thành phần cấu trúc của CSDL, các chương trình ứng dụng, mật mã, quyền hạn sử dụng v.v....

### 3. HỆ QUẢN TRỊ CSDL - DBMS ...

- 3) Có biện pháp bảo mật.
- 4) Cơ chế giải quyết vấn đề tranh chấp dữ liệu.
- 5) Phải có cơ chế sao lưu (Backup) và phục hồi (Restore) dữ liệu khi có sự cố xảy ra.
- 6) DBMS phải cung cấp một giao diện (Interface) tốt, dễ sử dụng, dễ hiểu cho những người sử dụng không chuyên.
- 7) Bảo đảm tính độc lập giữa dữ liệu và chương trình.

### 3. SƠ ĐỒ TỔNG QUÁT CỦA DBMS

- 3 mức:
  - Mức chương trình khai báo cấu trúc và ứng dụng;
  - Mức mô tả CSDL, thao tác CSDL và các từ điển di;
  - Mức CSDL.



*Hình 1.3 Sơ đồ tổng quát của một hệ quản trị CSDL*

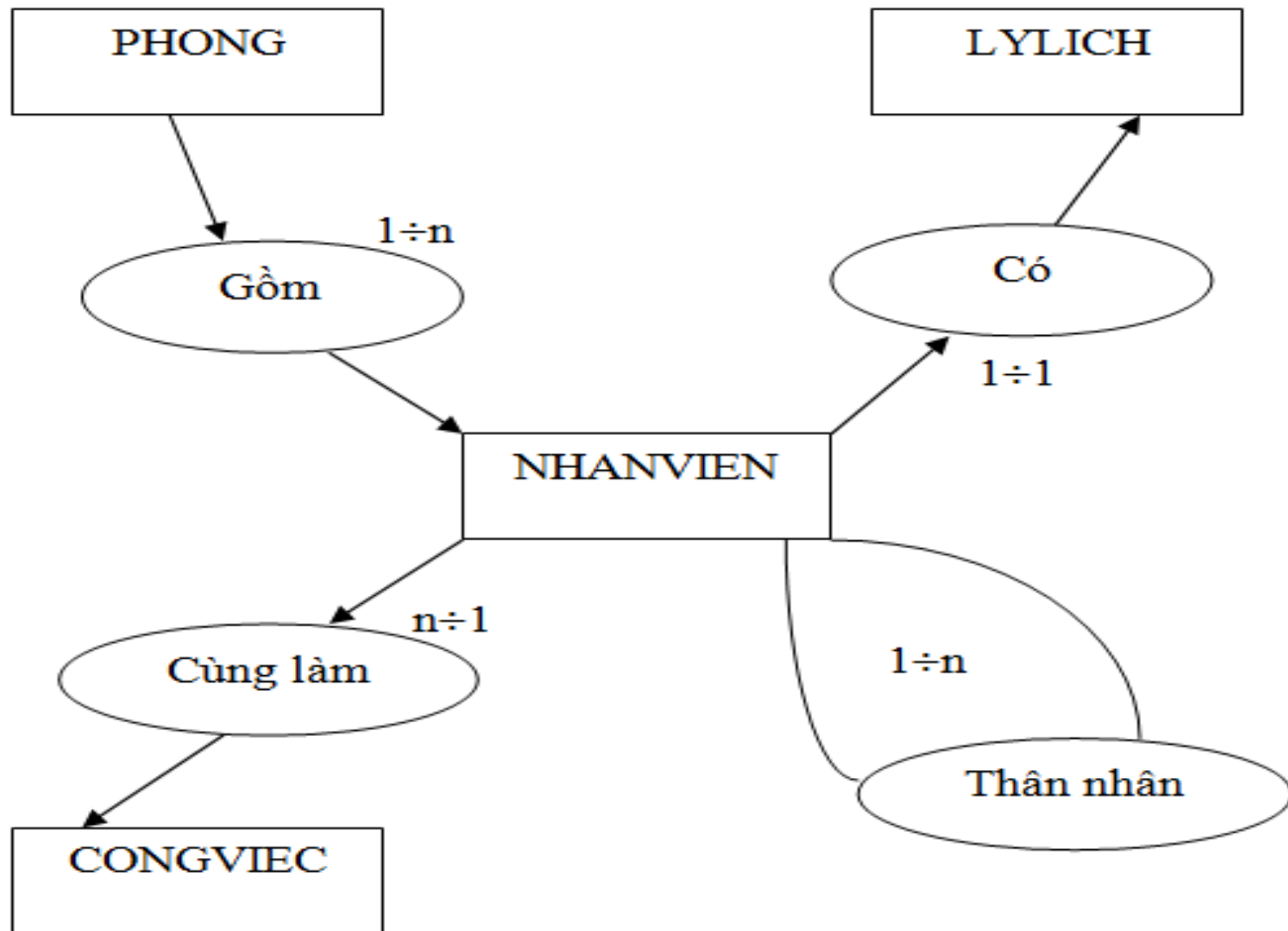
# 4. MÔ HÌNH DỮ LIỆU

- Mô hình dữ liệu (Data Model): những khái niệm dùng để biểu diễn cấu trúc của CSDL (data types, relationships, constraints).
  - Mô hình mạng (Network Data Model)
  - Mô hình phân cấp (Hierarchical Data Model)
  - Mô hình quan hệ (Relational Data Model)
  - Mô hình thực thể - liên kết (Entity - Relationship Model)
  - Mô hình hướng đối tượng (Object Oriented Data Model)
- => *Mỗi DBMS đều phải xây dựng dựa trên một mô hình DL nhất định.*

# 4. MÔ HÌNH MẠNG (NETWORK DATA MODEL)

- Do nhóm DBGIT của CODASYL đề xuất 1971
- Là mô hình được biểu diễn bởi một đồ thị có hướng
- Các khái niệm: mẫu tin (Record), loại mẫu tin (Record Type) và loại liên kết (Set Type):
  - *Loại mẫu tin (Record Type)* là mẫu đặc trưng cho 1 loại đối tượng riêng biệt
  - *Loại liên kết (Set Type)* là sự liên kết giữa một loại mẫu tin chủ với mẫu tin thành viên
    - 1 - 1 (One-to-One)
    - 1 - n (One-to-Many).
    - n -1 (Many-to-One)
    - Đệ quy (Recursive)

## 4. MÔ HÌNH MẠNG(NETWORK DATA MODEL)...



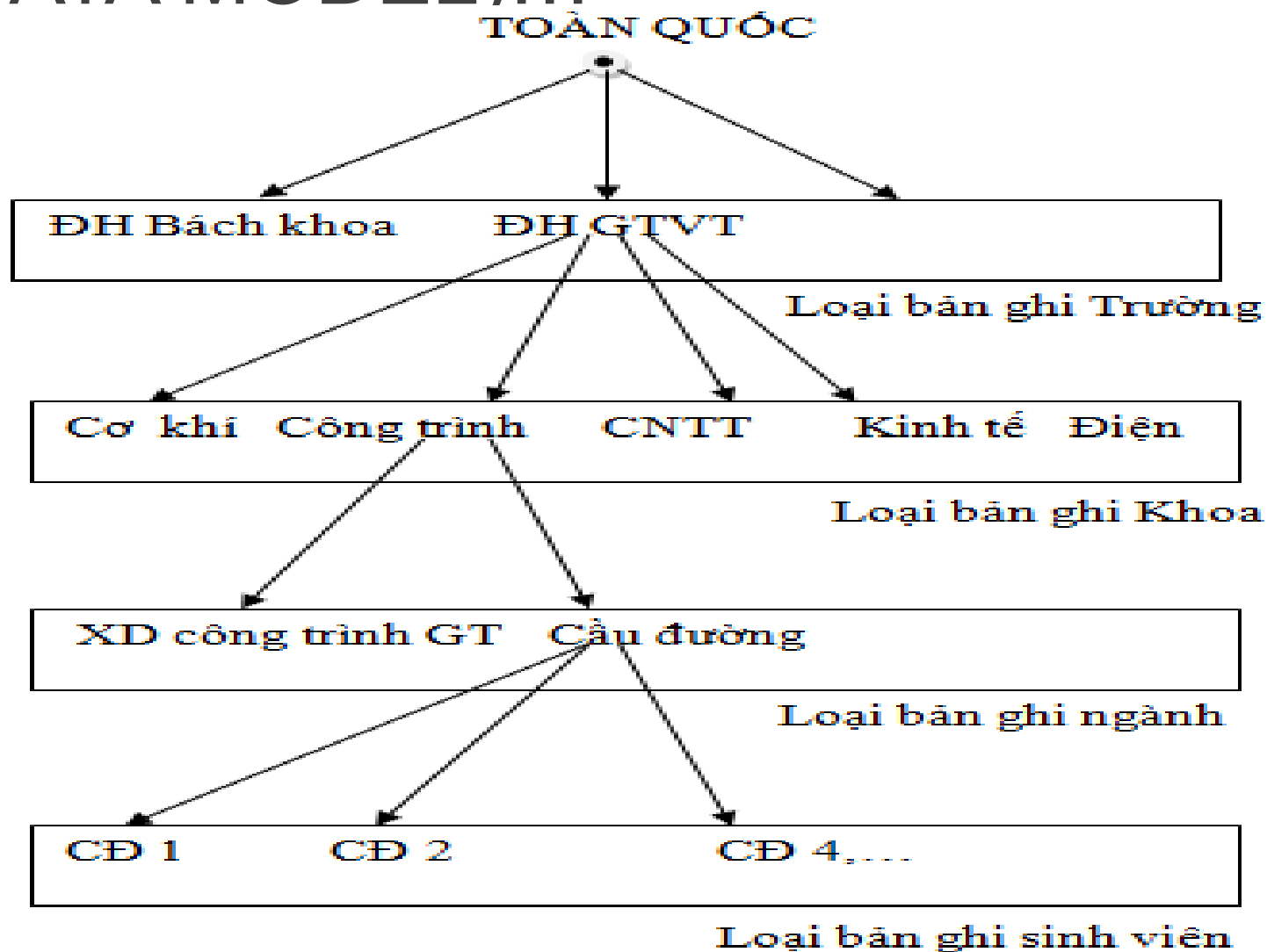
Hình 1.4 Mô hình dữ liệu mạng (Network Model)



# 4. MÔ HÌNH PHÂN CẤP (HIERACHICAL DATA MODEL)

- Có cấu trúc cây (Tree), trong đó mỗi nút của cây biểu diễn một thực thể, giữa nút con và nút cha được liên kết với nhau theo một mối quan hệ xác định.
- Mô hình dữ liệu phân cấp sử dụng các khái niệm sau:
  - *Loại mẫu tin*: giống khái niệm mẫu tin trong mô hình dữ liệu mạng.
  - *Loại mối liên kết*: Kiểu liên kết là phân cấp, theo cách:
  - Mẫu tin thành viên chỉ đóng vai trò thành viên của một mối liên kết duy nhất, tức là nó thuộc một chủ duy nhất. Như vậy, mối liên kết từ mẫu tin chủ tới các mẫu tin thành viên là 1 - n, và từ mẫu tin (hay bản ghi - record) thành viên với mẫu tin chủ là 1 –1.
- Giữa 2 loại mẫu tin chỉ tồn tại 1 mối liên kết duy nhất.

# 4. MÔ HÌNH PHÂN CẤP(HIERACHICAL DATA MODEL)...



Hình 1.5 Mô hình dữ liệu phân cấp (Hierarchical Model)

# 4. MÔ HÌNH DL QUAN HỆ (RELATIONAL DATA MODEL)

- Do E.F.Codd [2] đề xuất năm 1970. Nền tảng cơ bản là lý thuyết tập hợp trên các quan hệ, tức là tập của các bộ giá trị (Value Tuples).

- Các khái niệm:

- Thuộc tính (Attribute)
- Quan hệ (Relation)
- Lược đồ quan hệ (Relation Schema)
- Bộ/bản ghi (Tuple)
- Khóa (Key).

**SINHVIEN**

Masv	Tensv	Que
Sv1	Nguyễn Văn Anh	Hà Nội
Sv2	Phạm Ngọc Bình	Hải phòng
Sv3	Nguyễn Hoa Cúc	Quảng Ninh
Sv4	Đinh Gia Linh	Hà Nội

# 4. MÔ HÌNH DL THỰC THỂ LIÊN KẾT ER (ENTITY - RELATIONSHIP MODEL)

- Do P.P.Chen đề xuất 1976. Các khái niệm chủ yếu được sử dụng trong mô hình này là:
- **Thực thể (Entity):** Là khái niệm mô tả một lớp các đối tượng có đặc trưng chung mà chúng ta cần quan tâm.

Sinhvien

Khachhang

- **Liên kết:** *X là thực thể yếu nếu sự tồn tại của X phụ thuộc vào sự tồn tại của thực thể Y. Được ký hiệu bằng hình chữ nhật kép*

# 4. MÔ HÌNH DỮ LIỆU HƯỚNG ĐỐI TƯỢNG (OBJECT ORIENTED DATA MODEL)

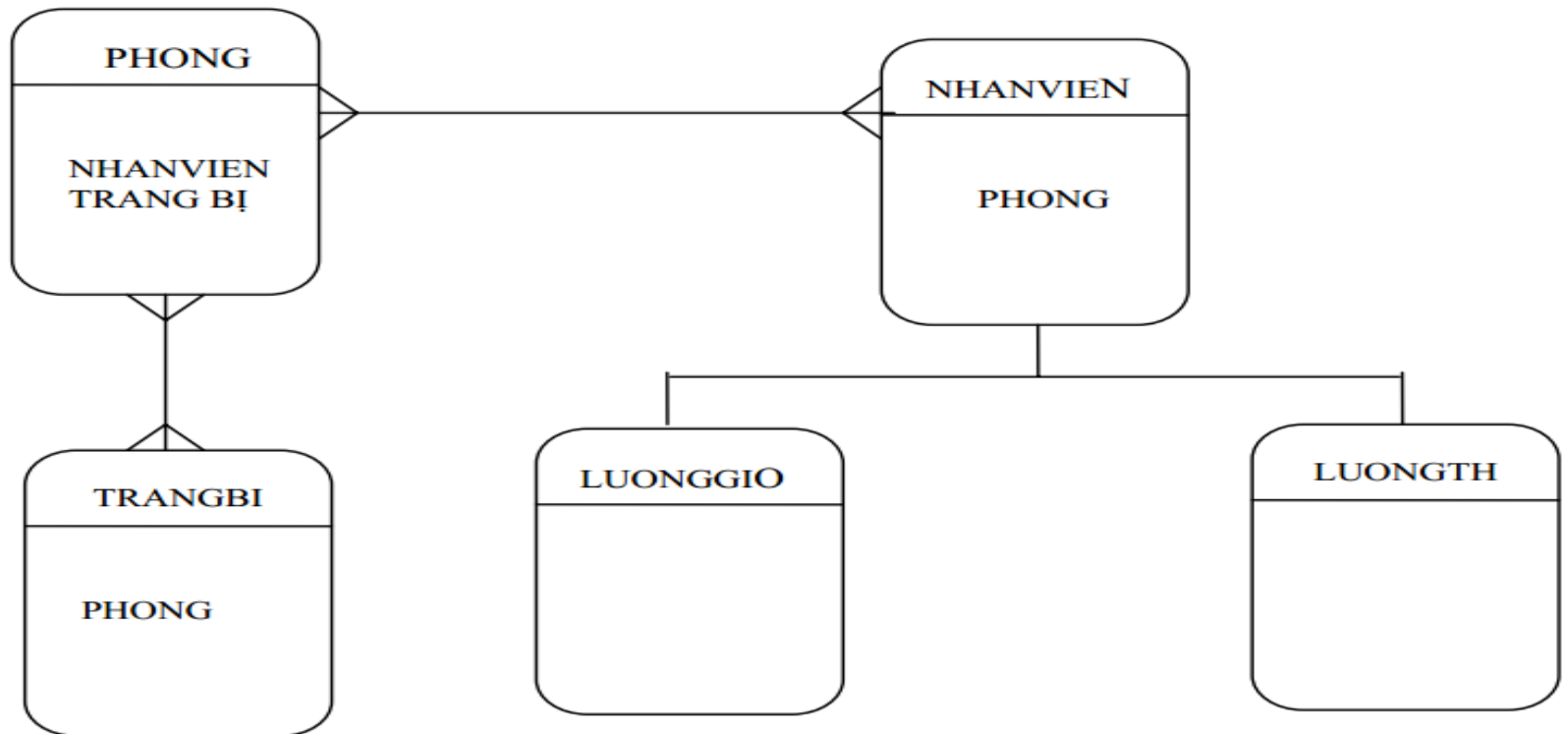
- Ra đời khoảng cuối những năm 80.
- Là loại mô hình tiên tiến nhất hiện nay dựa trên cách tiếp cận hướng đối tượng.
- Sử dụng các khái niệm như lớp (*class*), sự kế thừa (*inheritance*), kế thừa bội (tức là kế thừa từ nhiều lớp cơ sở *multi-inheritance*).
- Đặc trưng cơ bản: tính đóng gói (*encapsulation*), tính đa hình (*polymorphism*) và tính tái sử dụng (*Reusability*).
- Lớp: là một kiểu dl có cấu trúc bao gồm các thành phần dl và các phương thức xử lý thao tác trên cấu trúc dl đó.

# 4. MÔ HÌNH DỮ LIỆU HƯỚNG ĐỐI TƯỢNG (OBJECT ORIENTED DATA MODEL)...

- Các thuộc tính dữ liệu và các thao tác trên các dữ liệu này được bao gói trong một cấu trúc gọi là đối tượng.
- Đối tượng có thể chứa các dữ liệu phức hợp như văn bản, hình ảnh, tiếng nói và hình ảnh động.
- Một đối tượng có thể yêu cầu hoặc xử lý dữ liệu từ một đối tượng khác bằng việc gửi đi một thông báo đến đối tượng đó.
- Mô hình hướng đối tượng biểu diễn một sơ đồ mới để lưu trữ và thao tác dữ liệu. Từ một đối tượng có thể sinh ra một đối tượng khác.

# 4. MÔ HÌNH DỮ LIỆU HƯỚNG ĐỐI TƯỢNG (OBJECT ORIENTED DATA MODEL)...

- Theo [71]



Hình 2.4. Minh họa mô hình cơ sở dữ liệu hướng đối tượng

# BÀI TẬP CHƯƠNG 1

**1.1.** Định nghĩa các thuật ngữ sau:

- Cơ sở dữ liệu
- Hệ cơ sở dữ liệu
- Hệ quản trị cơ sở dữ liệu

**1.2.** Nêu sơ đồ kiến trúc của một hệ cơ sở dữ liệu

**1.3.** Hiểu và lấy ví dụ của các mô hình cơ sở dữ liệu.



# BÀI TẬP CHƯƠNG 1...

**1.4.** Dựa vào những khái niệm đã học hãy biểu diễn CSDL có các loại bản ghi PHONG, NHANVIEN, CONGVIEC, LYlich đã trình bày trong mô hình mạng theo cách tiếp cận phân cấp.

Loại liên kết là phân cấp:

Phòng có nhiều nhân viên; mỗi nhân viên chỉ thuộc 1 phòng duy nhất.

Công việc có nhiều nhân viên cùng làm, mỗi nhân viên chỉ làm một công việc duy nhất.

Mỗi nhân viên có một lý lịch; mỗi lý lịch chỉ thuộc 1 nhân viên duy nhất.

**1.5.** Dựa vào những khái niệm đã học, hãy biểu diễn CSDL về quản lý Sinh viên đã trình bày trong mô hình phân cấp theo cách tiếp cận mạng.

Loại liên kết phân mạng là loại "thuộc về"