

Cơ sở dữ liệu

Databases



Ts. Nguyễn Đức Thuần
BM Hệ thống Thông Tin
ngducthuan@gmail.com



Giới thiệu môn học

- Thời lượng: 03TC
 - ❖ Lý thuyết : 02 TC
 - ❖ Bài tập: 01 TC
- Yêu cầu
 - ❖ Bài tập giữa kỳ (viết): 30%
 - ❖ Báo cáo: 20%
 - ❖ Bài thi kết thúc: 50%

Mục đích

- Trang bị kiến thức cơ bản về CSDLQH và hệ quản trị CSDL
- Giới thiệu sơ lược về PTTK một Hệ CSDL
- Kỹ năng: Thiết kế CSDL có chuẩn hóa 3NF

Tài liệu tham khảo chính

- [1] Bài giảng CSDL – ĐHNT - Nguyễn Đức Thuần
- [2] PP giải bài tập CSDLQH – NXB KHKT - Nguyễn Đức Thuần + Trương Ngọc Châu
- [3] Ho Thuan, “Contribution to The theory of Relational databases” , Tanulmányok 1984/1986, ISBN 963 311 213 3, ISSN 0237-0131, 1986
- [4] Hồ Thuần, Hồ Cẩm Hà, “Các hệ cơ sở dữ liệu Lý thuyết & thực hành”, NXB Giáo dục, 2005 (2 tập).
- [5] Maier, D, “The Theory of Relational Databases”, Computer Science Press, Rockville, Md, 1983.
- [6] Nguyễn Xuân Huy, Lê Hoài Bắc, “Bài tập cơ sở dữ liệu”, NXB Khoa học tự nhiên & Công nghệ, 2008.
- [7] Phạm Thế Quế, “ Giáo trình Cơ sở dữ liệu Lý thuyết và thực hành”, HV CNBCVT, NXB Bưu điện, 2004.
- [8] Raghu Ramakrishnan, Johannes Gehrke, Jeff Derstadt, Scott Selikoff, and Lin Zhu, “Database Management Systems Solutions Manual Third Edition”.
<http://www.cs.wisc.edu/~dbbook>
- [9] Ullman J.D, “Principle of Database and Knowledge-base Systems”, Vol I&II, Computer Science Press, 1998.

Nội dung

- Các khái niệm cơ bản
- Mô hình thực thể liên kết (ERM)
- Mô hình dữ liệu quan hệ (RDM)
- Ngôn ngữ SQL
- Phép tính quan hệ
- Thiết kế CSDLQH
- Chuẩn hóa CSDL

Chương I

Các khái niệm cơ bản

Chương 1: Các khái niệm cơ bản

1. Dữ liệu, Thông tin, CSDL
2. Quá trình phát triển
3. Một số tính chất của CSDL
4. Chu kỳ sống của CSDL
5. Người sử dụng CSDL
6. Kiến trúc một hệ CSDL
7. Hệ quản trị CSDL
8. Một số khái niệm cơ sở
9. Các mô hình CSDL

Chương 1: Các khái niệm cơ bản

- ▶ **Dữ liệu** (*data*): sự biểu diễn của các đối tượng và sự kiện được ghi nhận và được lưu trữ trên các thiết bị trữ tin.
 - Dữ liệu có cấu trúc: số, ngày, chuỗi ký tự, ...
 - Dữ liệu không có cấu trúc: hình ảnh, âm thanh, đoạn phim, ...
- ▶ **Thông tin** (*Information*): Thông tin là dữ liệu được xử lý và biểu diễn dưới dạng phù hợp với mục đích yêu cầu của người sử dụng
- ▶ **Cơ sở dữ liệu** (*Databases*) một tập hợp dữ liệu được tổ chức có cấu trúc, được lưu trữ trên những thiết bị trữ tin thỏa mãn một cách đồng thời và có chọn lọc cho nhiều người dùng khác nhau và cho những mục đích khác nhau.

Chương 1: Các khái niệm cơ bản

- ▶ **Siêu dữ liệu** (*metadata*) là dữ liệu dùng để mô tả các tính chất / đặc tính của dữ liệu khác (dữ liệu về dữ liệu)
- **Các đặc tính:** định nghĩa dữ liệu, cấu trúc dữ liệu, qui tắc / ràng buộc.

Ví dụ:

Siêu dữ liệu cho Sinh_viên

Data Item			Value		
Name	Type	Length	Min	Max	Description
MaSV	Character	8			Ma sinh viên
Hoten	Character	30			Họ tên sinh viên
Lop	Character	3			Lớp
Tuoi	Number	2	17	25	Tuổi

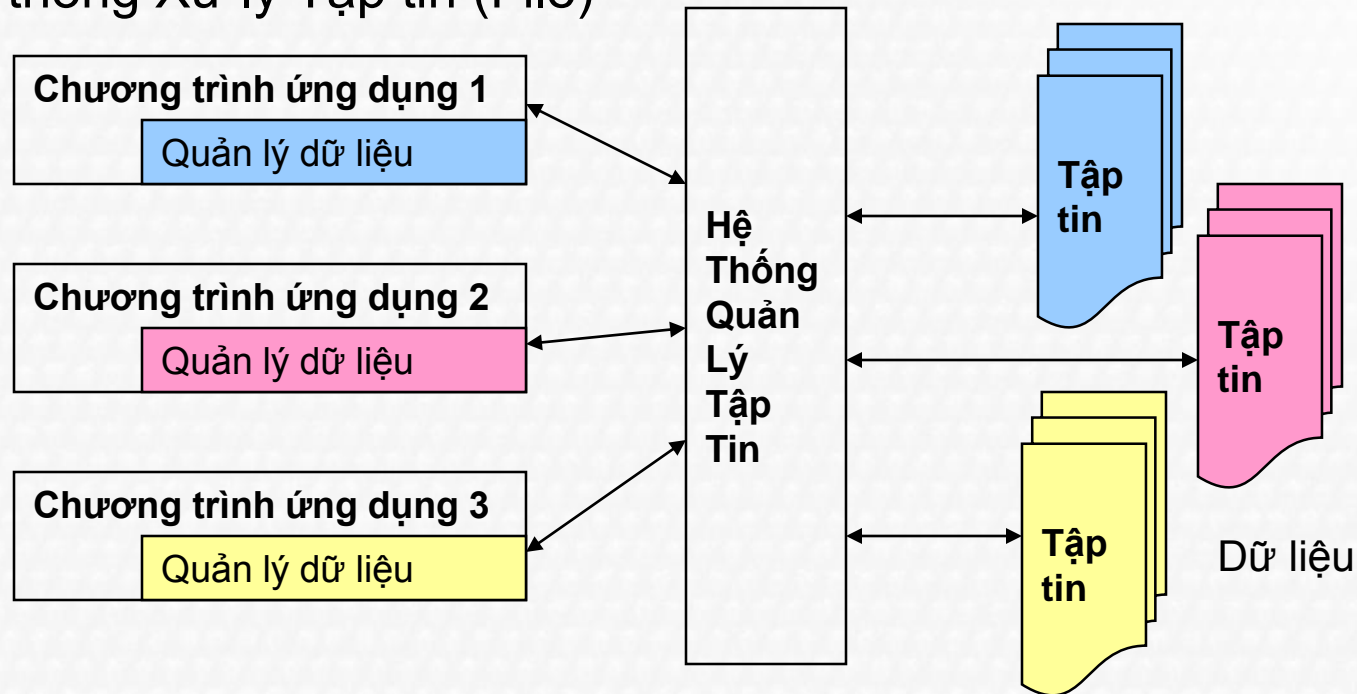
Chương 1: Các khái niệm cơ bản

1. Dữ liệu, Thông tin, CSDL
2. Quá trình phát triển
3. Một số tính chất của CSDL
4. Chu kỳ sống của CSDL
5. Người sử dụng CSDL
6. Kiến trúc một hệ CSDL
7. Hệ quản trị CSDL
8. Một số khái niệm cơ sở
9. Các mô hình CSDL

Chương 1: Các khái niệm cơ bản

❖ Quá trình phát triển

▪ Hệ thống Xử lý Tập tin (File)



Chương 1: Các khái niệm cơ bản

❖ Hệ thống xử lý tập tin (*file processing system*)

- ▶ *Hệ thống xử lý tập tin* là tập hợp các chương trình dùng để lưu trữ, thao tác và truy xuất các tập tin dữ liệu có kích thước lớn.

❖ Các thành phần của hệ thống xử lý tập tin

- ▶ Phần cứng: các máy tính.
- ▶ Phần mềm:
 - Hệ điều hành
 - Các tiện ích
 - Các tập tin
 - Các chương trình quản lý tập tin
 - Các chương trình ứng dụng tạo các báo cáo từ các dữ liệu được lưu trữ trong các tập tin.

Chương 1: Các khái niệm cơ bản

❖ Các thành phần của hệ thống xử lý tập tin (tt)

- ▶ **Con người:** người quản lý, chuyên gia, người lập trình, người sử dụng cuối cùng.
- ▶ **Các thủ tục:** các lệnh và các qui tắc chi phối việc thiết kế và sử dụng các thành phần của phần mềm.
- ▶ **Dữ liệu:** tập hợp các sự kiện.

❖ Quản lý dữ liệu của hệ thống tập tin

- ▶ **Ngôn ngữ lập trình:** 3GL (*third-Generation Language*).
 - Các ngôn ngữ: COBOL (*COmmon Business-Oriented Language*), BASIC (*Beginner's All-purpose Symbolic Instruction Code*), FORTRAN (*FORmula TRANslation*)

Chương 1: Các khái niệm cơ bản

❖ Quản lý dữ liệu của hệ thống tập tin (tt)

▶ Các chương trình xử lý tập tin

- Tạo cấu trúc tập tin.
- Thêm dữ liệu vào tập tin.
- Xóa dữ liệu của tập tin
- Sửa dữ liệu của tập tin.
- Liệt kê dữ liệu của tập tin.

Chương 1: Các khái niệm cơ bản

❖ Nhược điểm của hệ thống xử lý tập tin

▶ Phụ thuộc dữ liệu – chương trình (*Program-Data Dependence*)

- Tất cả các chương trình ứng dụng phải duy trì siêu dữ liệu (phần mô tả) của các tập tin mà chúng sử dụng.
- ❖ Mỗi người lập trình phải duy trì dữ liệu riêng biệt.
- ❖ Mỗi chương trình ứng dụng phải có mã lệnh cho siêu dữ liệu của mỗi tập tin.
- ❖ Mỗi chương trình ứng dụng phải có các chương trình con xử lý để đọc, thêm, sửa và xóa dữ liệu.
- ❖ Không có các điều khiển chung và phối hợp.
- ❖ Các dạng thức tập tin không có cùng chuẩn.

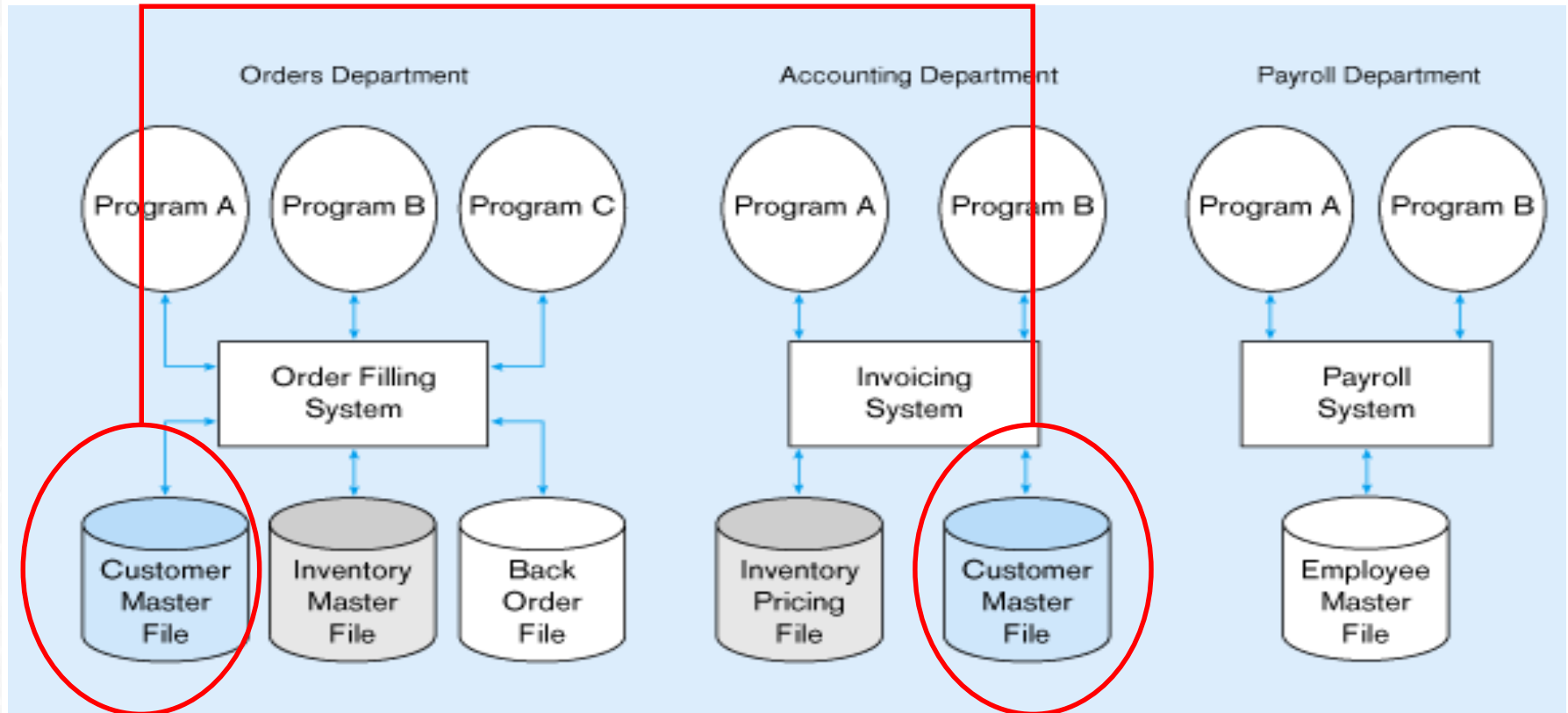
Chương 1: Các khái niệm cơ bản

❖ Nhược điểm của hệ thống xử lý tập tin

- ▶ Dư thừa dữ liệu / Trùng lặp dữ liệu (*Data Redundancy / Duplication of Data*)
 - Các hệ thống / chương trình khác nhau có các bản dữ liệu riêng biệt của cùng dữ liệu.
- ❖ Tốn vùng nhớ để lưu trữ dữ liệu dư thừa.
- ❖ Gây ra các vấn đề khó về bảo trì dữ liệu.
- ❖ Vấn đề chính:
 - ▶ Việc cập nhật dữ liệu của một tập tin có thể dẫn đến các mâu thuẫn dữ liệu.
 - ▶ Vi phạm tính toàn vẹn dữ liệu.

Chương 1: Các khái niệm cơ bản

Duplicate Data



Chương 1: Các khái niệm cơ bản

❖ Nhược điểm của hệ thống xử lý tập tin (tt)

▶ Hạn chế việc dùng chung dữ liệu

- Mỗi ứng dụng có các tập tin riêng biệt, ít sử dụng chung dữ liệu với các ứng dụng khác.

▶ Thời gian phát triển lâu

- Người lập trình phải thiết kế các dạng tập tin dữ liệu riêng và viết cách truy xuất tập tin cho mỗi ứng dụng mới.

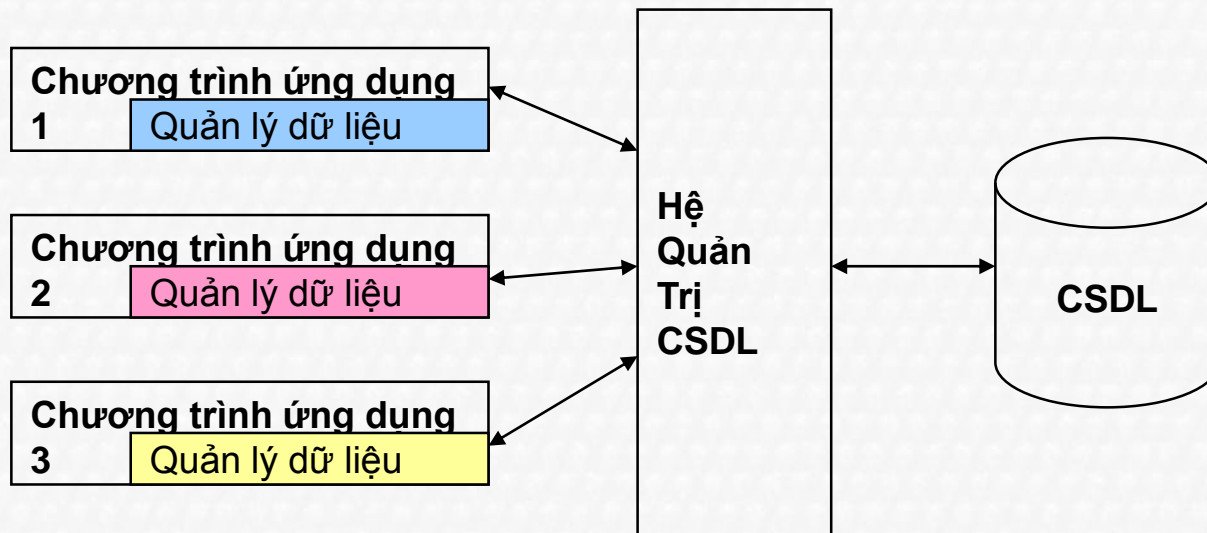
▶ Chi phí bảo trì chương trình cao

- Các nhược điểm nêu trên làm cho việc bảo trì chương trình gặp nhiều khó khăn, thường chiếm khoảng 80% ngân sách phát triển HTTT.

Chương 1: Các khái niệm cơ bản

❖ Tiếp cận Cơ sở dữ liệu

- Kho dữ liệu trung tâm chứa các dữ liệu dùng chung.
- Dữ liệu được quản lý bởi một đơn vị điều khiển (*controlling agent*).
- Dữ liệu được lưu trữ theo một dạng thức chuẩn và thích hợp.
- Cần phải có một hệ quản trị CSDL.



Chương 1: Các khái niệm cơ bản

❖ Ưu điểm của cách tiếp cận CSDL

- ▶ Độc lập dữ liệu – chương trình (*data - program independence*).
 - DBMS chứa siêu dữ liệu (*metadata*), do đó các ứng dụng không cần quan tâm đến các dạng thức của dữ liệu.
 - DBMS quản lý các truy vấn và cập nhật dữ liệu, do đó ứng dụng không cần xử lý việc truy xuất dữ liệu.
- ▶ Giảm tối thiểu sự dư thừa dữ liệu (*data redundancy*).
- ▶ Nâng cao tính nhất quán (*data consistency*) / toàn vẹn dữ liệu (*data integrity*).

Chương 1: Các khái niệm cơ bản

- ▶ Nâng cao việc dùng chung dữ liệu (*data sharing*).
 - Những người sử dụng khác nhau có những cái nhìn khác nhau về dữ liệu.
- ▶ Tăng hiệu suất phát triển ứng dụng.
- ▶ Tuân thủ các tiêu chuẩn.
 - Tất cả các truy xuất dữ liệu đều được thực hiện theo cùng một cách.
- ▶ Nâng cao chất lượng của dữ liệu.
 - Các ràng buộc (*constraint*), các qui tắc hợp lệ của dữ liệu (*data validation rule*).

Chương 1: Các khái niệm cơ bản

- ▶ Nâng cao tính truy xuất và tính đáp ứng của dữ liệu.
 - Sử dụng ngôn ngữ truy vấn dữ liệu chuẩn (SQL - *Structured Query Language*).
- ▶ Giảm chi phí bảo trì chương trình.
- ▶ Bảo mật (*security*).
- ▶ Chép lưu (*backup*) và phục hồi (*recovery*).
- ▶ Điều khiển tương tranh (*concurrency control*).

Chương 1: Các khái niệm cơ bản

❖ Chi phí và rủi ro của cách tiếp cận CSDL

- ▶ Chi phí ban đầu
 - Chi phí cài đặt và quản lý
 - Chi phí chuyển đổi (*conversion cost*)
- ▶ Chi phí vận hành
 - Cần nhân viên mới có chuyên môn.
 - Cần phải chép lưu và phục hồi.
- ▶ Mâu thuẫn về mặt tổ chức
 - Rất khó thay đổi các thói quen cũ.

Chương 1: Các khái niệm cơ bản

1. Dữ liệu, Thông tin, CSDL
2. Quá trình phát triển
3. Một số tính chất của CSDL
4. Chu kỳ sống của CSDL
5. Người sử dụng CSDL
6. Kiến trúc một hệ CSDL
7. Hệ quản trị CSDL
8. Một số khái niệm cơ sở
9. Các mô hình CSDL

Chương 1: Các khái niệm cơ bản

❖ Một số tính chất của CSDL

□ Tính tự mô tả

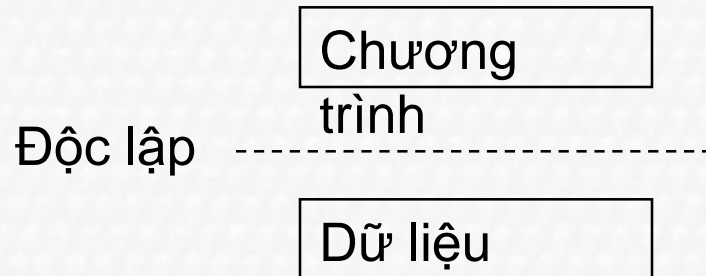
- Hệ CSDL không chỉ chứa bản thân CSDL mà còn chứa định nghĩa đầy đủ (mô tả) của CSDL
- Các định nghĩa được lưu trữ trong catalog
 - Chứa các thông tin về cấu trúc tập tin, kiểu và dạng thức lưu trữ của mỗi thành phần dữ liệu và những ràng buộc dữ liệu
- Dữ liệu trong catalog gọi là meta-data (data of data)
- Các CTƯD có thể truy xuất đến nhiều CSDL nhờ thông tin cấu trúc được lưu trữ trong catalog

Chương 1: Các khái niệm cơ bản

❖ Một số tính chất của CSDL

☐ Tính độc lập

Vì định nghĩa về cấu trúc CSDL được lưu trữ trong catalog nên khi có thay đổi nhỏ về cấu trúc ta ít phải sửa lại chương trình



Chương 1: Các khái niệm cơ bản

□ Tính nhất quán

- Lưu trữ dữ liệu thống nhất
 - Tránh được tình trạng trùng lặp thông tin
- Có cơ chế điều khiển truy xuất dữ liệu hợp lý
 - Tránh được việc tranh chấp dữ liệu
 - Bảo đảm dữ liệu luôn đúng tại mọi thời điểm

Chương 1: Các khái niệm cơ bản

□ Các cách nhìn CSDL

- Hệ CSDL cho phép nhiều người dùng thao tác lên cùng một CSDL
- Mỗi người đòi hỏi một cách nhìn (view) khác nhau về CSDL
- Một view là
 - Một phần của CSDL hoặc
 - Dữ liệu tổng hợp từ CSDL

Chương 1: Các khái niệm cơ bản

1. Dữ liệu, Thông tin, CSDL
2. Quá trình phát triển
3. Một số tính chất của CSDL
4. Chu kỳ sống của CSDL
5. Người sử dụng CSDL
6. Kiến trúc một hệ CSDL
7. Hệ quản trị CSDL
8. Một số khái niệm cơ sở
9. Các mô hình CSDL

Chương 1: Các khái niệm cơ bản

❖ Chu kỳ sống của CSDL

- Lập kế hoạch CSDL (database planning)
- Khảo sát, phân tích (study & analysis)
- Thiết kế CSDL dữ liệu (database design)
- Cài đặt CSDL (database implementation)
- Bảo trì CSDL (Post - Implementation)

Chương 1: Các khái niệm cơ bản

1. Dữ liệu, Thông tin, CSDL
2. Quá trình phát triển
3. Một số tính chất của CSDL
4. Chu kỳ sống của CSDL
5. Người sử dụng CSDL
6. Kiến trúc một hệ CSDL
7. Hệ quản trị CSDL
8. Một số khái niệm cơ sở
9. Các mô hình CSDL

Chương 1: Các khái niệm cơ bản

– Quản trị viên (Database Administrator - DBA)

- Có trách nhiệm quản lý hệ CSDL
- Cấp quyền truy cập CSDL
- Điều phối và giám sát việc sử dụng CSDL

– Thiết kế viên (Database Designer)

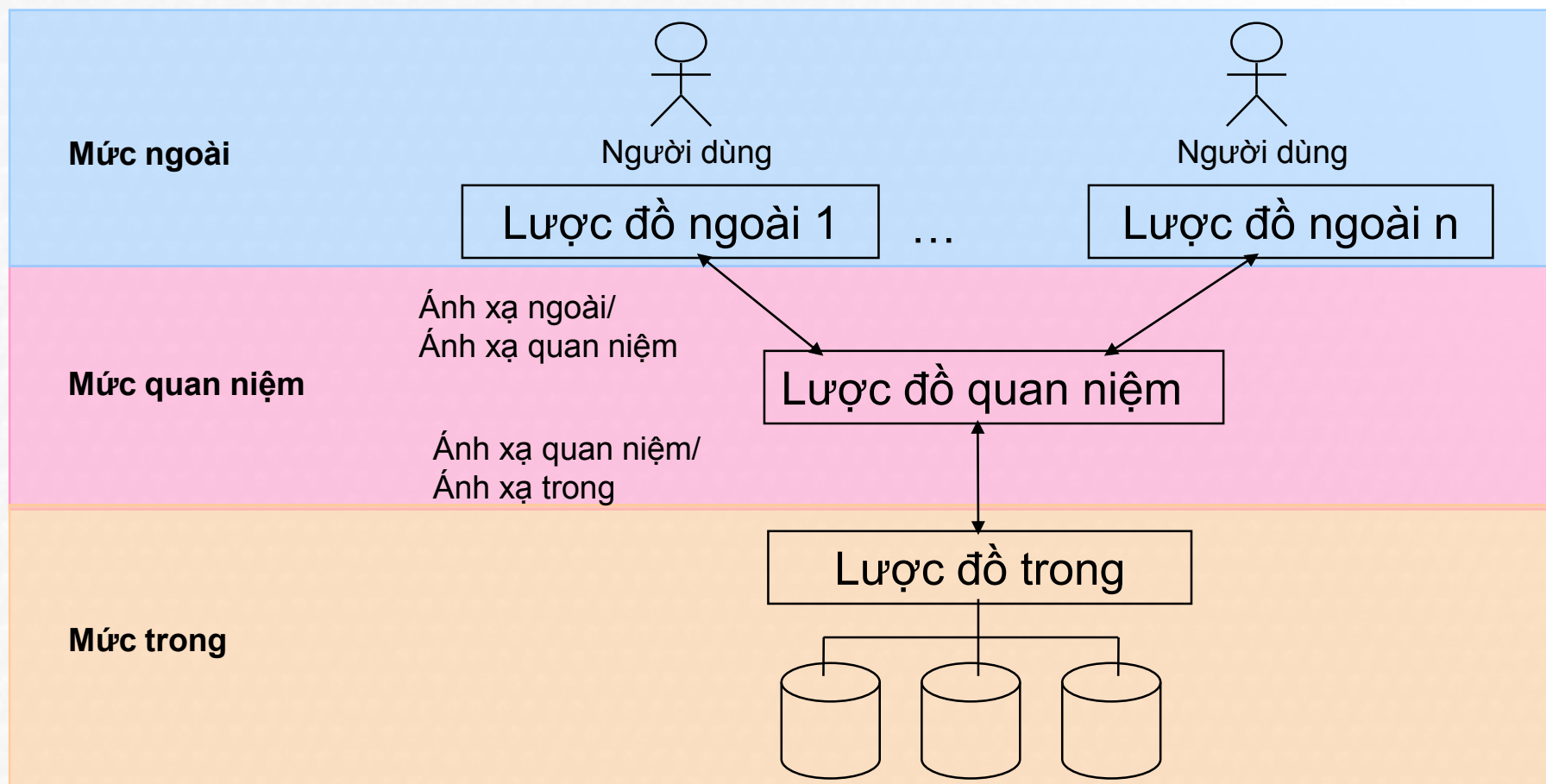
- Chịu trách nhiệm về
 - Lựa chọn cấu trúc phù hợp để lưu trữ dữ liệu
 - Quyết định những dữ liệu nào cần được lưu trữ
- Liên hệ với người dùng để nắm bắt được những yêu cầu và đưa ra một thiết kế CSDL thỏa yêu cầu này
- Có thể là 1 nhóm các DBA quản lý các CSDL sau khi việc thiết kế hoàn tất

– Người dùng cuối (End User)

Chương 1: Các khái niệm cơ bản

1. Dữ liệu, Thông tin, CSDL
2. Quá trình phát triển
3. Một số tính chất của CSDL
4. Chu kỳ sống của CSDL
5. Người sử dụng CSDL
6. Kiến trúc một hệ CSDL
7. Hệ quản trị CSDL
8. Một số khái niệm cơ sở
9. Các mô hình CSDL

Chương 1: Các khái niệm cơ bản



Chương 1: Các khái niệm cơ bản

- **Mức trong** (lược đồ trong)
 - Mô tả cấu trúc lưu trữ vật lý CSDL
- **Mức quan niệm** (lược đồ quan niệm)
 - Mô tả cấu trúc của toàn thể CSDL cho 1 cộng đồng người sử dụng, gồm thực thể, kiểu dữ liệu, mối liên hệ và ràng buộc
 - Che bớt các chi tiết của cấu trúc lưu trữ vật lý
- **Mức ngoài** (lược đồ ngoài)
 - Còn gọi là mức khung nhìn (view)
 - Mô tả một phần của CSDL mà 1 nhóm người dùng quan tâm đến và che dấu phần còn lại của CSDL đối với nhóm người dùng đó

Chương 1: Các khái niệm cơ bản

1. Dữ liệu, Thông tin, CSDL
2. Quá trình phát triển
3. Một số tính chất của CSDL
4. Chu kỳ sống của CSDL
5. Người sử dụng CSDL
6. Kiến trúc một hệ CSDL
7. **Hệ quản trị CSDL**
8. Một số khái niệm cơ sở
9. Các mô hình CSDL

Chương 1: Các khái niệm cơ bản

- Hệ quản trị CSDL (Database Management Systems)

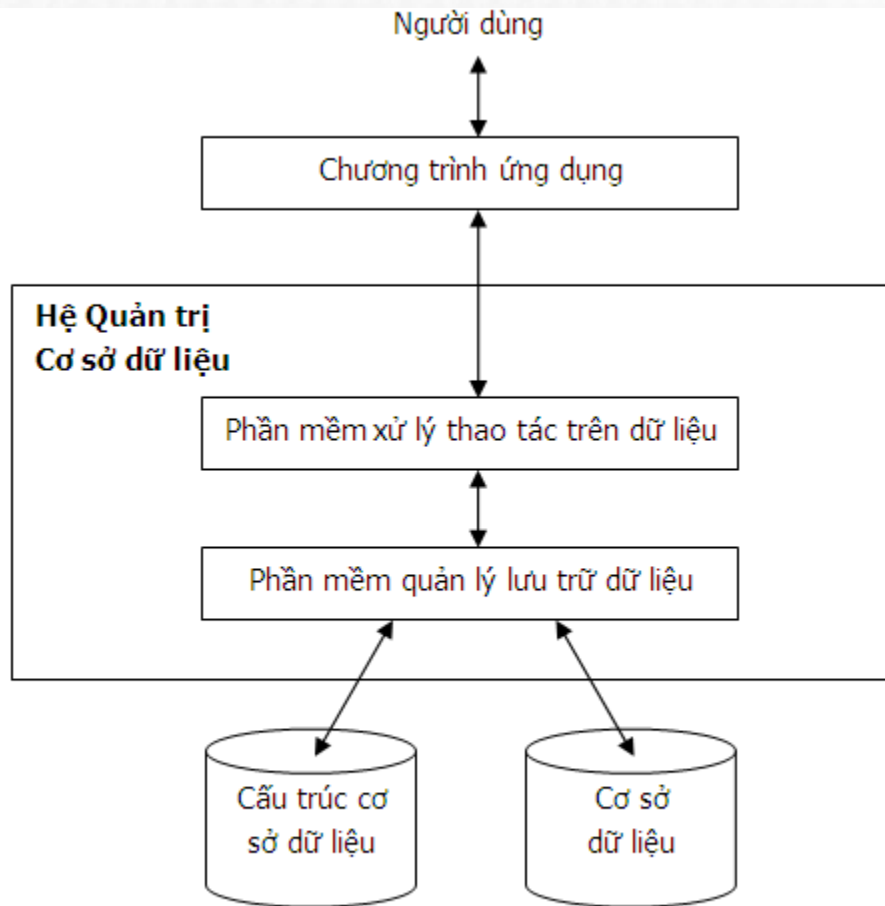
- Phần mềm: Tổ chức, thiết kế, truy xuất,.. dữ liệu

Tính chất:

- **Định nghĩa dữ liệu:** tạo cấu trúc để lưu trữ dữ liệu tùy theo mục đích của cơ sở dữ liệu.
- **Thao tác dữ liệu:** cho phép người dùng có thể truy vấn các dữ liệu trong cơ sở dữ liệu, hay cập nhật các thay đổi trong thế giới thực vào cơ sở dữ liệu.
- **Chia sẻ:** cho phép nhiều người dùng có thể khai thác dữ liệu một cách đồng thời.
- **Bảo vệ hệ thống:** nhằm bảo đảm cho cơ sở dữ liệu luôn an toàn trong trường hợp xảy ra hỏng hóc phần cứng cũng như phần mềm máy tính. Bảo vệ khỏi các hành vi phá hoại từ bên ngoài.

- Ví dụ:

Chương 1: Các khái niệm cơ bản



Hình 1-1. Mô tả đơn giản một hệ thống cơ sở dữ liệu

Chương 1: Các khái niệm cơ bản

❖ Các chức năng của hệ quản trị CSDL

- ▶ Lưu trữ, truy xuất và cập nhật dữ liệu
 - Ngôn ngữ định nghĩa dữ liệu (DDL - *Data Definition Language*)
 - Ngôn ngữ thao tác dữ liệu (DML - *Data Manipulation Language*).
- ▶ Quản lý giao tác (*transaction management*).
- ▶ Điều khiển tương tranh (*concurrency control*)
- ▶ Chép lưu và phục hồi dữ liệu.
- ▶ Bảo mật dữ liệu
 - Ngôn ngữ điều khiển dữ liệu (DCL - *Data Control Language*).
- ▶ Hỗ trợ truyền thông dữ liệu.
- ▶ Duy trì tính toàn vẹn / nhất quán dữ liệu.
- ▶ Cung cấp các tiện ích.

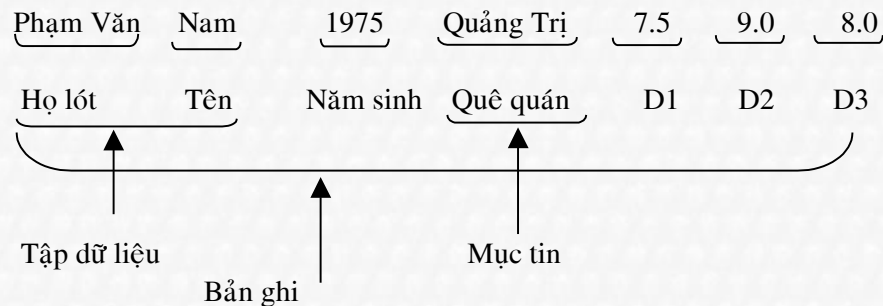
Chương 1: Các khái niệm cơ bản

1. Dữ liệu, Thông tin, CSDL
2. Quá trình phát triển
3. Một số tính chất của CSDL
4. Chu kỳ sống của CSDL
5. Người sử dụng CSDL
6. Kiến trúc một hệ CSDL
7. Hệ quản trị CSDL
8. Một số khái niệm cơ sở
9. Các mô hình CSDL

Chương 1: Các khái niệm cơ bản

❖ Các khái niệm cơ sở:

- **Mục tin** (*Data Item*): là đơn vị bé nhất của dữ liệu có một tên.
- **Tập dữ liệu** (*Data Aggregate*): tập hợp các mục tin xếp tuần tự trong dữ liệu và giữ một tên.
- **Bản ghi** (*Record*): tập các dữ liệu gộp lớn và các mục tin xếp kế tiếp nhau của thông tin về một đối tượng, tạo thành đơn vị trao đổi giữa cơ sở dữ liệu và các ứng dụng.
- **File dữ liệu**: là tập hợp các bản ghi có cùng một cấu trúc xác định, thường được lưu trữ thành từng file.



Chương 1: Các khái niệm cơ bản

1. Dữ liệu, Thông tin, CSDL
2. Quá trình phát triển
3. Một số tính chất của CSDL
4. Chu kỳ sống của CSDL
5. Người sử dụng CSDL
6. Kiến trúc một hệ CSDL
7. Hệ quản trị CSDL
8. Một số khái niệm cơ sở
9. Các mô hình CSDL

Chương 1: Các khái niệm cơ bản

■ Mô hình dữ liệu

- Mô hình: Sự trừu tượng hóa / tổng quát hóa một đối tượng nào đó.
- Mô hình dữ liệu: Biểu diễn / Thể hiện dữ liệu ở mức trừu tượng hóa đối tượng cần xử lý.

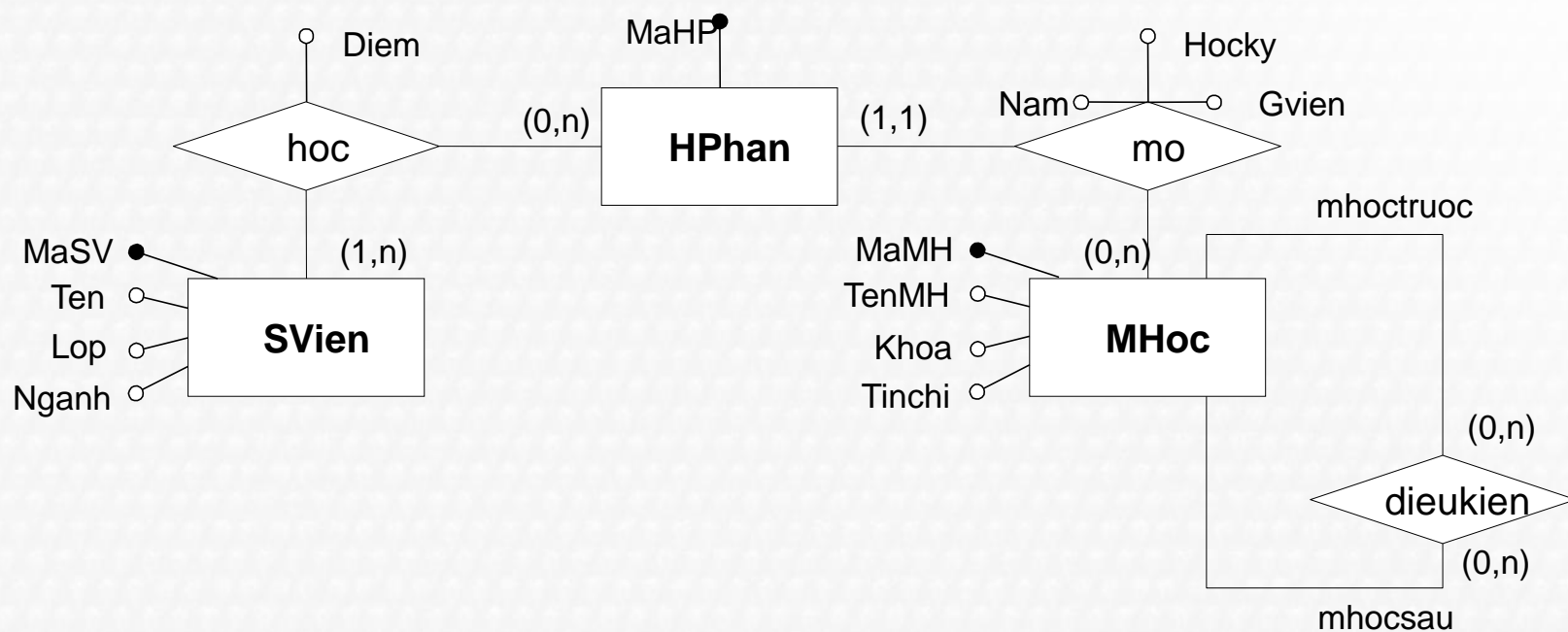
■ Mô hình dữ liệu

Các mô hình dữ liệu thông dụng

- Mô hình thực thể kết hợp (Entity-Relationship Data Model)
- Mô hình dữ liệu mạng (Network Data Model)
- Mô hình dữ liệu phân cấp (Hierarchical Data Model)
- **Mô hình dữ liệu quan hệ** (Relational Data Model)
- Mô hình dữ liệu hướng đối tượng (Object-Oriented Model)

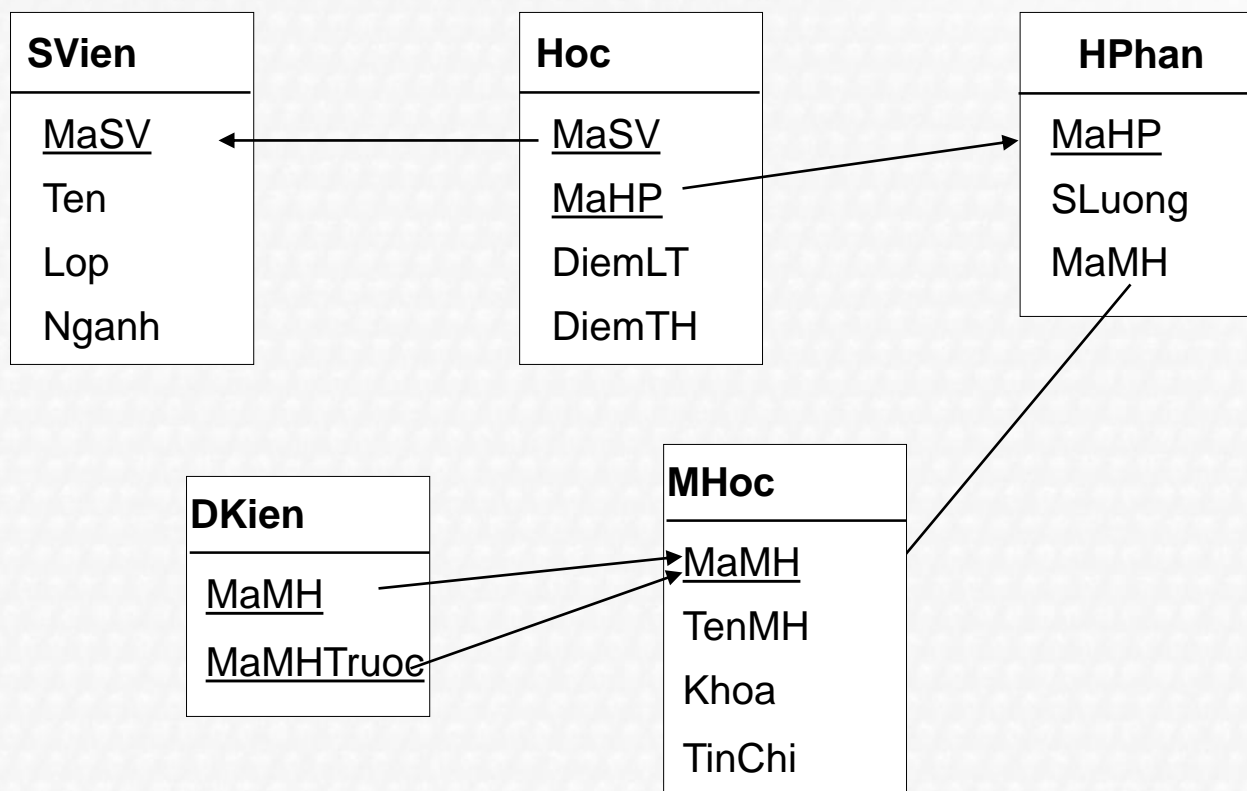
Chương 1: Các khái niệm cơ bản

Ví dụ mô hình Thực thể kết hợp (ER)



Chương 1: Các khái niệm cơ bản

Ví dụ mô hình dữ liệu quan hệ



Chương 1: Các khái niệm cơ bản

Ví dụ Mô hình phân cấp

Mức 1:

KQua	
DiemTH	DiemLT

Mức 2:

HPhan	
TenHP	SLuong

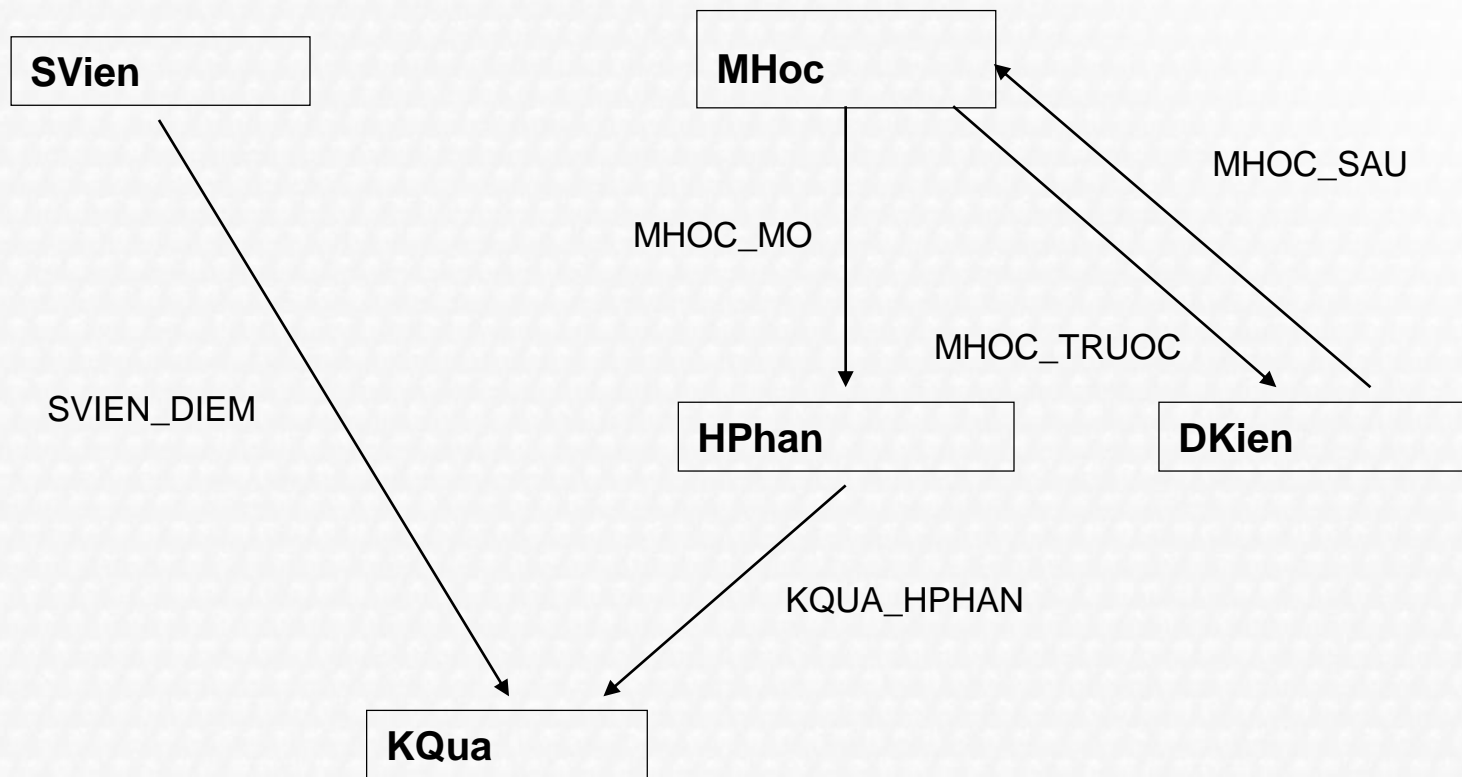
SVien		
TenSV	Lop	Nganh

Mức 3:

MHoc		
TenMH	Khoa	TinChi

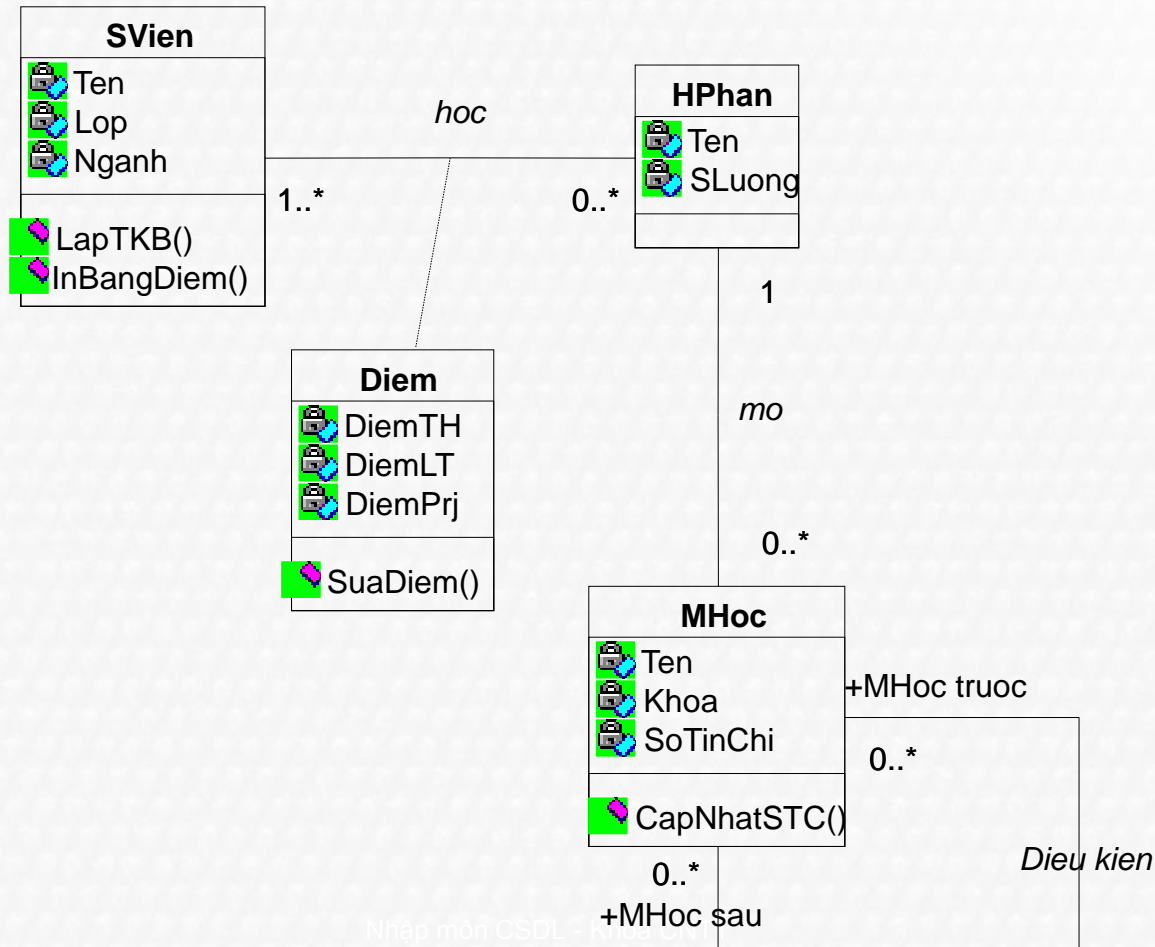
Chương 1: Các khái niệm cơ bản

Ví dụ Mô hình mạng



Chương 1: Các khái niệm cơ bản

Ví dụ mô hình dữ liệu hướng đối tượng



Chương 1: Các khái niệm cơ bản

Lịch sử phát triển các hệ CSDL

- ❖ Hệ thống tập tin (*flat file*): 1960 - 1980
- ❖ Hệ CSDL phân cấp (*hierarchical*): 1970 - 1990
- ❖ Hệ CSDL mạng (*network*): 1970 - 1990
- ❖ Hệ CSDL quan hệ (*relational*): 1980 - nay
- ❖ Hệ CSDL hướng đối tượng (*object-oriented*): 1990 - nay
- ❖ Hệ CSDL đối tượng - quan hệ (*object-relational*): 1990 - nay
- ❖ Kho dữ liệu (*data warehouse*): 1980 - nay
- ❖ *Web-enabled*: 1990 - nay

Cám ơn đã theo dõi

