

Chương 1

Đại cương về các hệ cơ sở dữ liệu

Nguyễn Hồng Phương

phuongnh@soict.hust.edu.vn

<http://users.soict.hust.edu.vn/phuongnh>

<https://www.facebook.com/phuongnhbk>

Bộ môn Hệ thống thông tin
Viện Công nghệ thông tin và Truyền thông
Đại học Bách Khoa Hà Nội

Nội dung chương này

- 1.1 Các hệ thống xử lý tệp truyền thống và những hạn chế của nó.
- 1.2 Các hệ CSDL: khái niệm, khả năng, kiến trúc, người dùng của một hệ quản trị CSDL.
- 1.3 Sự phân loại các hệ CSDL.



1.1 Các hệ thống xử lý tệp truyền thống

- Bước khởi đầu của quá trình tin học hóa doanh nghiệp.
- Tập trung vào nhu cầu xử lý dữ liệu của các phòng riêng lẻ trong tổ chức mà không xem xét tổng thể tổ chức này.
- Viết một chương trình mới đối với mỗi ứng dụng đơn lẻ, không có kế hoạch, không có mô hình hướng đến sự tăng trưởng.



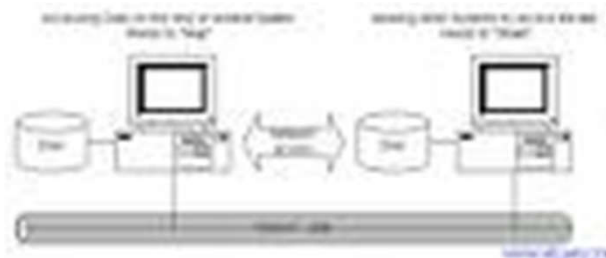
Các hệ thống xử lý tệp truyền thống

- Mỗi chương trình ứng dụng định nghĩa và quản lý các tệp dữ liệu của riêng nó.
- Trước khi xuất hiện các phần mềm hệ quản trị CSDL, trong quá khứ các hệ thống trên cơ sở tệp đã được tạo lập để xử lý một số lượng lớn dữ liệu.



Hạn chế của các hệ thống xử lý tệp truyền thống

- Dư thừa và không nhất quán dữ liệu
- Khó khăn trong truy nhập dữ liệu
- Cô lập và hạn chế chia sẻ dữ liệu
- Các vấn đề về an toàn và toàn vẹn
- Các vấn đề về độ tin cậy
- Sự phụ thuộc dữ liệu của các chương trình ứng dụng



Nội dung chương này

- 1.1 Các hệ thống xử lý tệp truyền thống và những hạn chế của nó.
- 1.2 Các hệ CSDL: khái niệm, khả năng, kiến trúc, người dùng của một hệ quản trị CSDL.
- 1.3 Sự phân loại các hệ CSDL.





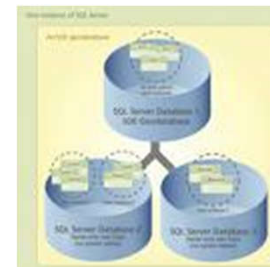
1.2 Các hệ cơ sở dữ liệu

- CSDL (database) là gì ?
- Tại sao phải sử dụng CSDL ?
- Tại sao phải tìm hiểu về các hệ CSDL (database systems) ?

Ví dụ về quản lý đào tạo

- Thông tin cần quan tâm
 - Khóa học, lớp học, sinh viên, giáo viên, môn học,...
 - Thông tin về sinh viên: thông tin cá nhân, thông tin học tập,...
 - Thông tin về môn học: khối lượng, giáo viên, lịch học,...
- Cần *lưu trữ* những thông tin đa dạng

→ **Cơ sở dữ liệu**



Ví dụ: khai thác thông tin

- Sinh viên
 - Các môn học trong CTĐT?
 - Điểm thi môn "Cơ sở dữ liệu"?
- Giáo viên
 - Danh sách sinh viên lớp 111565 ?
 - Thời khóa biểu của lớp KTMT1-K62 ?
- Giáo vụ
 - Danh sách sinh viên K59 tốt nghiệp loại giỏi ?...



Phần mềm ứng dụng



"Hình dung" về xây dựng một CSDL

- Yêu cầu
 - Lưu trữ thông tin **cần thiết** một cách **chính xác**
 - Truy xuất thông tin **hiệu quả**
- Thực hiện
 - Xác định yêu cầu nghiệp vụ
 - Xác định thông tin cần lưu trữ
 - Xác định cách thức lưu trữ
- Cần công cụ trợ giúp xây dựng một CSDL

➔ **Phần mềm quản trị CSDL**





Cơ sở dữ liệu (database)

- Là một tập hợp các dữ liệu
 - Biểu diễn một vài khía cạnh nào đó của thế giới thực
 - Có liên hệ logic thống nhất
 - Được thiết kế và bao gồm những dữ liệu phục vụ một mục đích nào đó.
- Là một bộ sưu tập các dữ liệu tác nghiệp được lưu trữ lại và được các hệ ứng dụng của một xí nghiệp cụ thể nào đó sử dụng.

Hệ quản trị cơ sở dữ liệu

(Database Management System-DBMS)

- Là một hệ thống phần mềm cho phép
 - Định nghĩa, tạo lập: xác định kiểu, cấu trúc, ràng buộc dữ liệu, lưu trữ dữ liệu trên các thiết bị nhớ.
 - Thao tác: truy vấn, cập nhật, kết xuất,... các CSDL cho các ứng dụng khác nhau
- Ví dụ: MS SQL Server, mySQL, PostGreSQL, DB2, MS Access, Oracle, FoxPro,...



Hệ cơ sở dữ liệu

- Là một hệ thống gồm 4 thành phần
 - Hệ quản trị CSDL
 - Phần cứng
 - CSDL và phần mềm ứng dụng
 - Những người sử dụng
- Ví dụ: Hệ quản lý đào tạo, hệ quản lý nhân sự, hệ quản lý kinh doanh,...



Hệ CSDL

Hệ
CSDL



Ứng dụng

Hệ Quản Trị CSDL

CSDL

CSDL

Các tính năng của hệ quản trị CSDL

- Quản lý dữ liệu tồn tại lâu dài
 - Định nghĩa dữ liệu
 - Quản lý lưu trữ
- Truy xuất dữ liệu một cách hiệu quả
 - Biểu diễn các thao tác dữ liệu
 - Xử lý câu hỏi
 - Quản trị giao dịch





Các tính năng của hệ quản trị CSDL

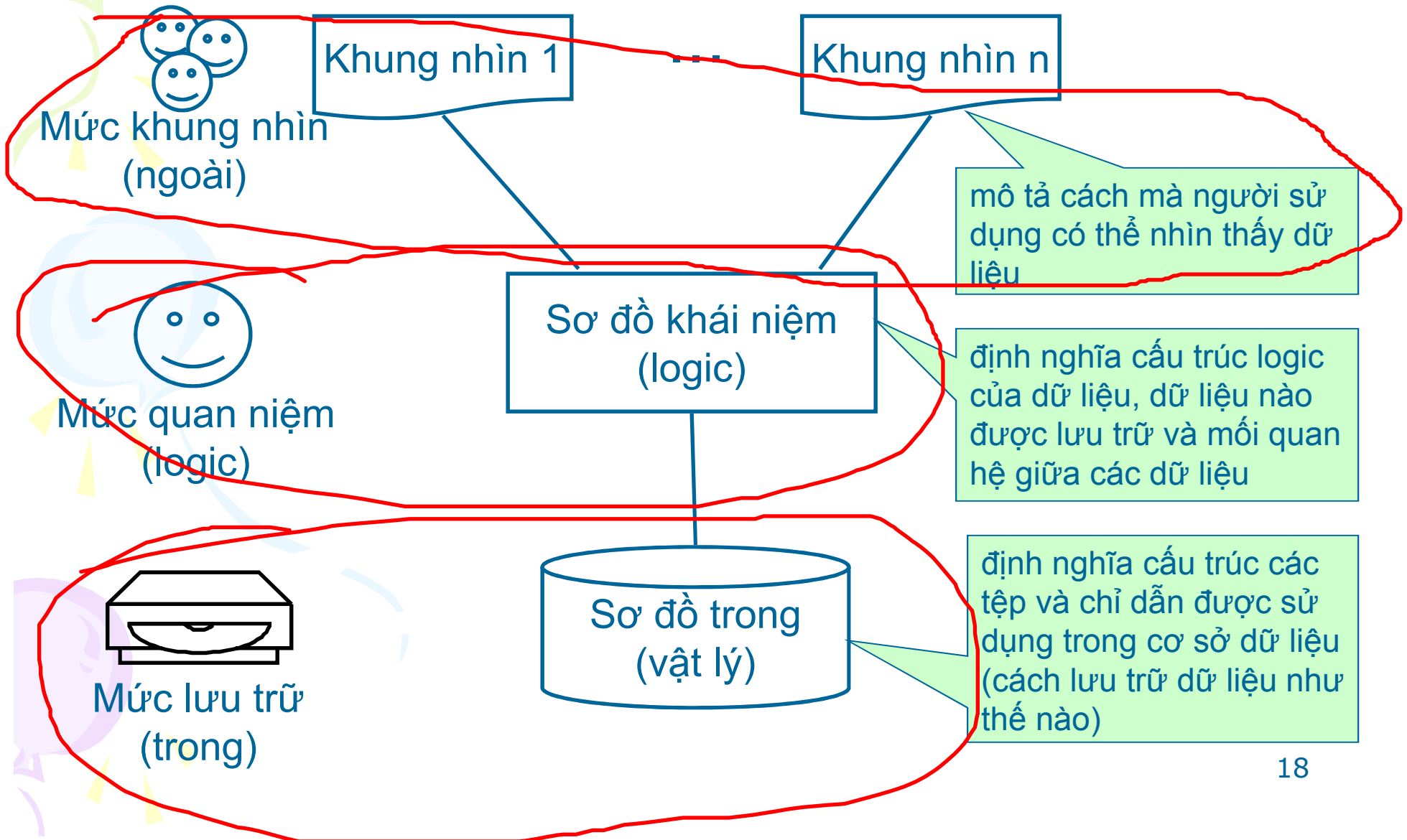
- Hỗ trợ ít nhất một mô hình dữ liệu
- Đảm bảo tính độc lập dữ liệu
- Hỗ trợ các ngôn ngữ cấp cao nhất định cho phép người sử dụng định nghĩa cấu trúc của dữ liệu, truy nhập và thao tác dữ liệu
- Điều khiển truy nhập
- Phục hồi dữ liệu



Các ngôn ngữ

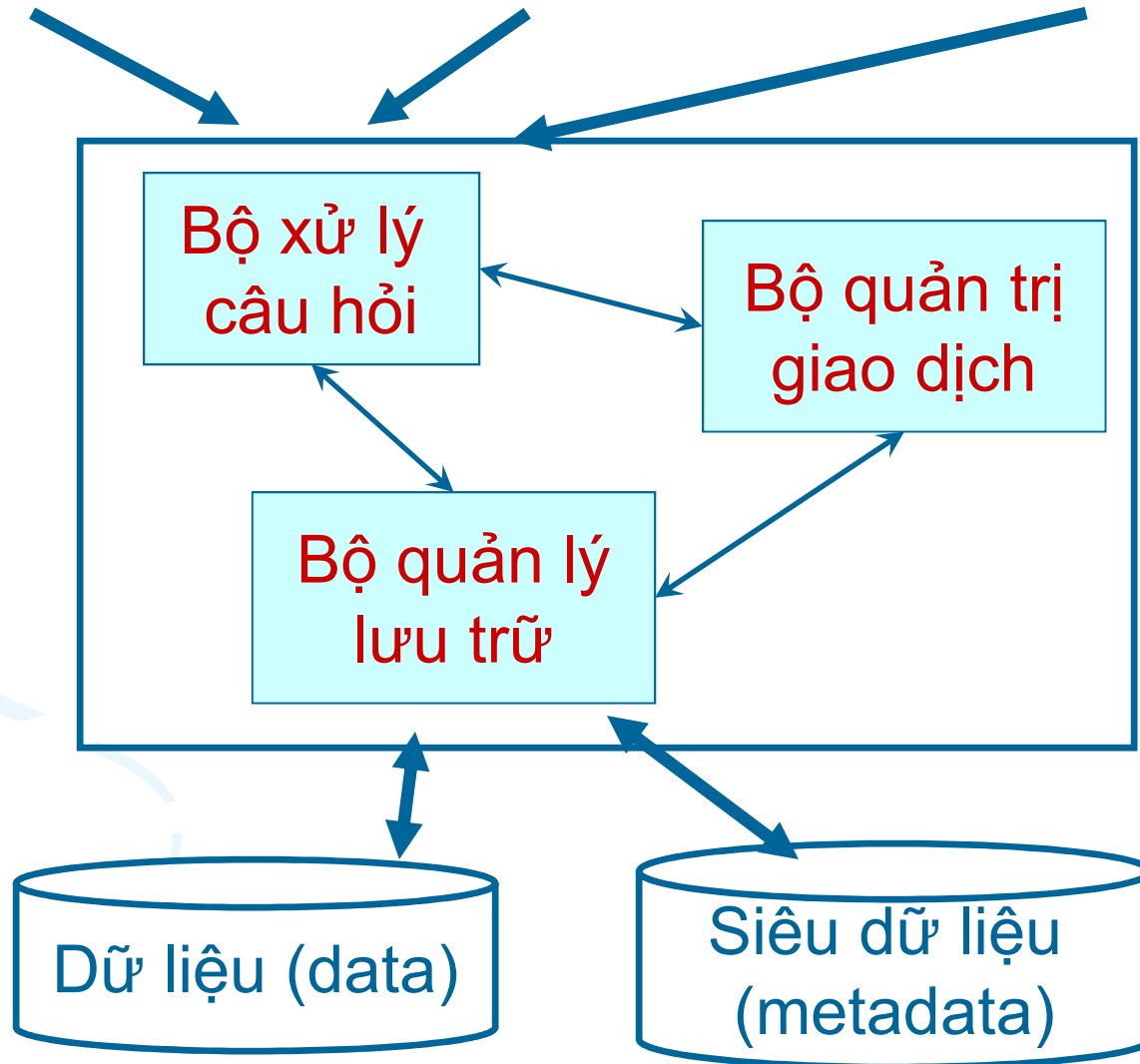
- Ngôn ngữ định nghĩa dữ liệu (**D**ata **D**efinition **L**anguage - DDL)
 - Cấu trúc dữ liệu
 - Mỗi liên hệ giữa các dữ liệu và quy tắc, ràng buộc áp đặt lên dữ liệu
- Ngôn ngữ thao tác dữ liệu (**D**ata **M**anipulation **L**anguage - DML)
 - Tìm kiếm, thêm, xóa, sửa dữ liệu trong CSDL
- Ngôn ngữ điều khiển dữ liệu (**D**ata **C**ontrol **L**anguage - DCL)
 - Thay đổi cấu trúc của các bảng dữ liệu
 - Khai báo bảo mật thông tin
 - Quyền hạn của người dùng trong khai thác CSDL

Sự trừu tượng hóa dữ liệu



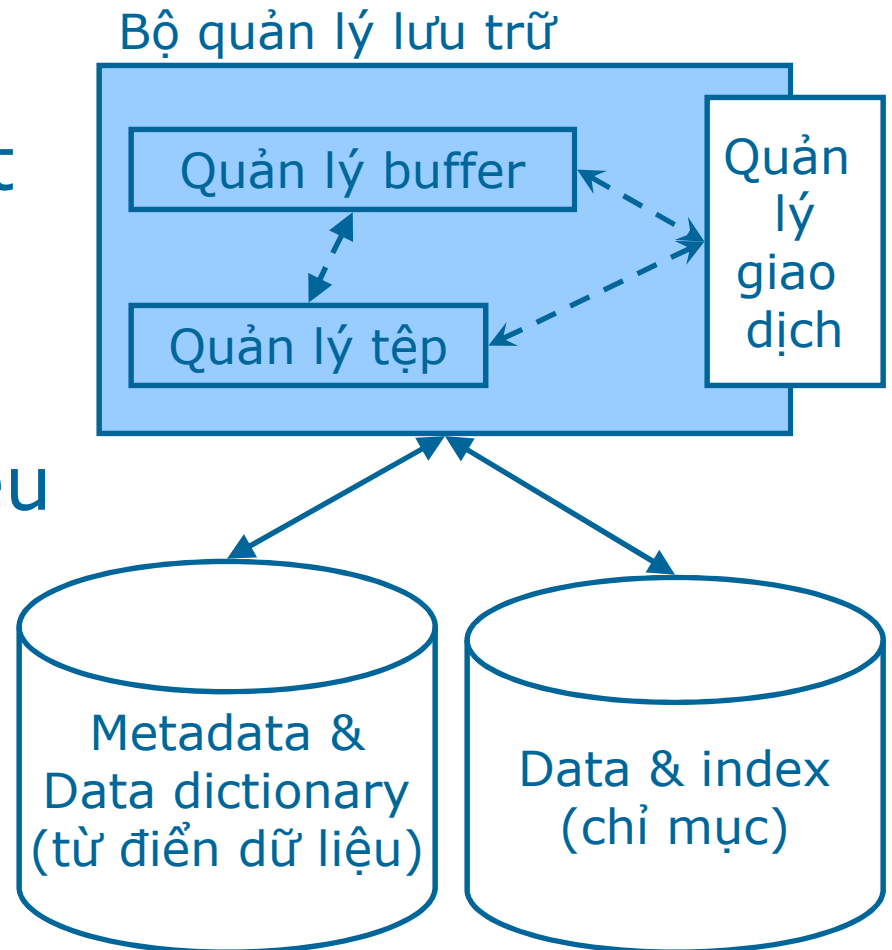
Kiến trúc của một hệ quản trị CSDL

Các thay đổi sơ đồ Các truy vấn Các thay đổi dữ liệu



Quản lý lưu trữ

- Yêu cầu
 - lưu trữ và truy xuất dữ liệu trên các thiết bị nhớ
- Thực hiện
 - Tổ chức tối ưu dữ liệu trên thiết bị nhớ
 - Tương tác hiệu quả với bộ quản lý tệp



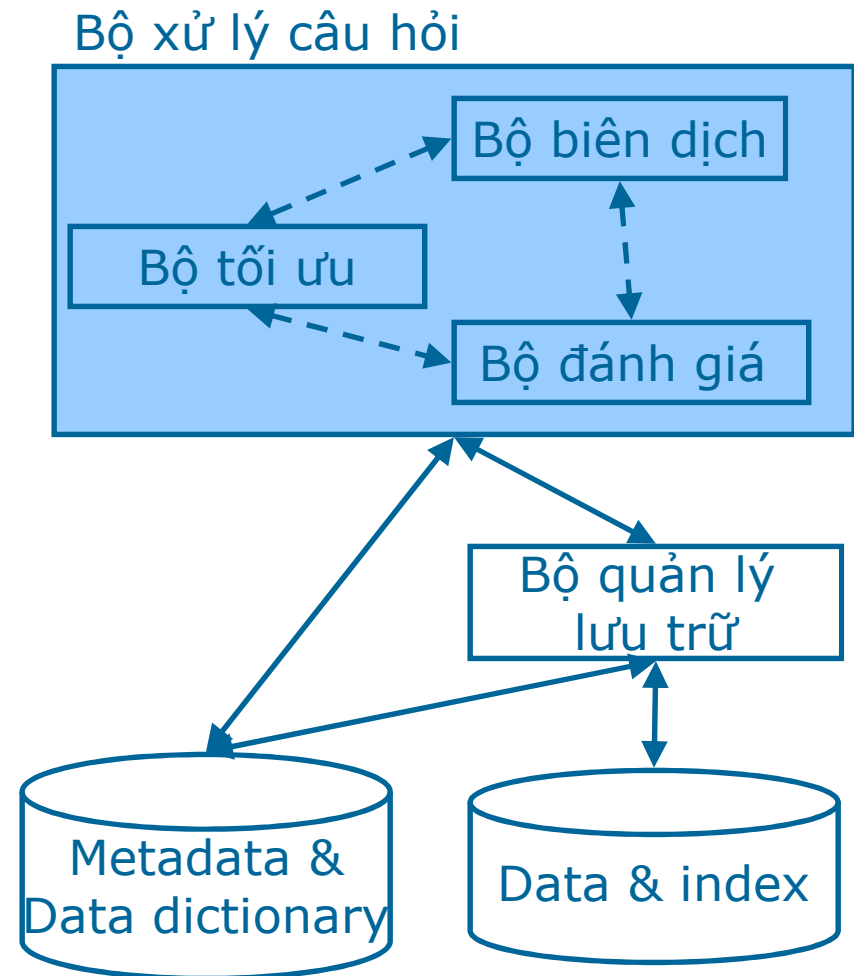
Xử lý câu hỏi

- Yêu cầu

- Tìm kiếm dữ liệu trả lời cho một yêu cầu truy vấn.

- Thực hiện

- Biến đổi truy vấn ở mức cao thành các yêu cầu có thể hiểu được bởi hệ CSDL.
- Lựa chọn một kế hoạch tốt nhất để trả lời truy vấn này.





Quản trị giao dịch

- Yêu cầu

- Định nghĩa giao dịch: một tập các thao tác được xử lý như một đơn vị không chia cắt được.
- Đảm bảo tính đúng đắn và tính nhất quán của dữ liệu.

- Thực hiện

- Quản lý điều khiển tương tranh.
- Phát hiện lỗi và phục hồi CSDL



Người dùng

- **Người thiết kế và cài đặt hệ QTCSDL:** chịu trách nhiệm thiết kế và cài đặt các module của hệ QTCSDL và các giao diện dưới hình thức các gói phần mềm
- **Người phát triển công cụ:** chịu trách nhiệm thiết kế và cài đặt các gói phần mềm hỗ trợ cho việc thiết kế, sử dụng cũng như tăng cường hiệu năng của các hệ CSDL.



Người dùng

- **Người phân tích hệ thống và phát triển ứng dụng:** chịu trách nhiệm xác định yêu cầu của người dùng cuối, xác định các giao dịch cần thiết để đáp ứng các yêu cầu người dùng. Người lập trình ứng dụng cài đặt những yêu cầu này trong chương trình, kiểm thử, gỡ rối, lập tài liệu cho chương trình
- **Người thiết kế CSDL:** chịu trách nhiệm xác định dữ liệu lưu trữ trong CSDL và cấu trúc biểu diễn và lưu trữ những dữ liệu này



Người dùng

- **Người sử dụng cuối:** là người khai thác các hệ CSDL
- **Người quản trị CSDL:** chịu trách nhiệm cho phép truy nhập CSDL, điều phối và kiểm tra sử dụng CSDL, quản lý tài nguyên phần cứng và phần mềm khi cần thiết
- **Người bảo trì hệ thống:** là những người quản trị hệ thống chịu trách nhiệm việc hoạt động và bảo trì môi trường (phần cứng và phần mềm) cho hệ CSDL

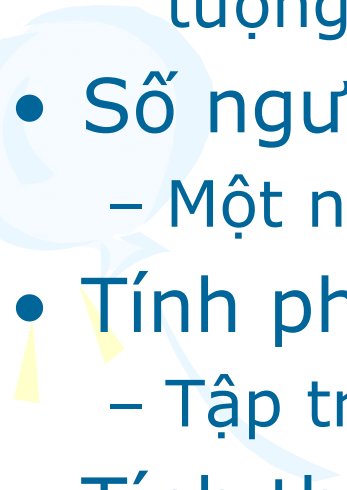

Nội dung chương này

- 1.1 Các hệ thống xử lý tệp truyền thống và những hạn chế của nó.
- 1.2 Các hệ CSDL: khái niệm, khả năng, kiến trúc, người dùng của một hệ quản trị CSDL.
- 1.3 Sự phân loại các hệ CSDL.





1.3 Phân loại các hệ CSDL

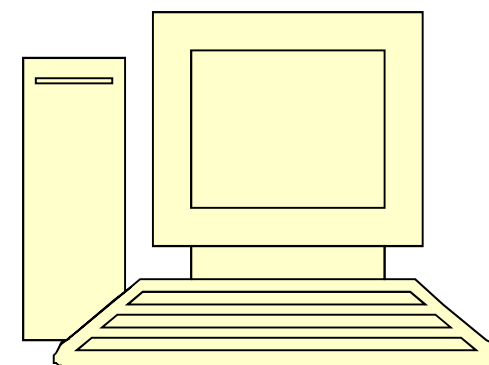
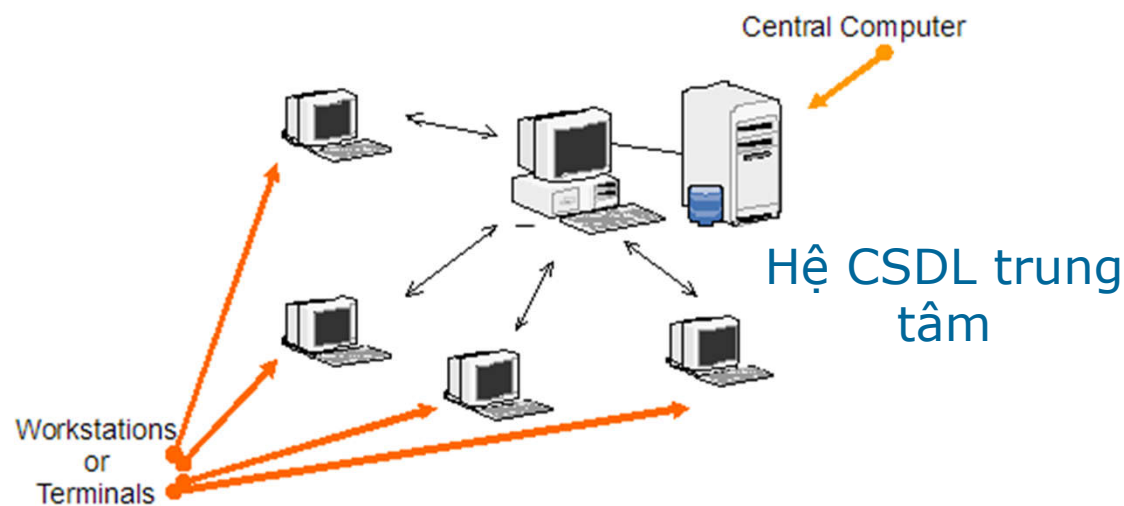
- Mô hình dữ liệu
 - Mạng vs. phân cấp vs. quan hệ vs. hướng đối tượng vs. ...
 - Số người sử dụng
 - Một người dùng vs. nhiều người dùng
 - Tính phân tán của CSDL
 - Tập trung vs. Phân tán
 - Tính thống nhất của dữ liệu
 - Đồng nhất vs. Không đồng nhất
 - ...
- 
- 



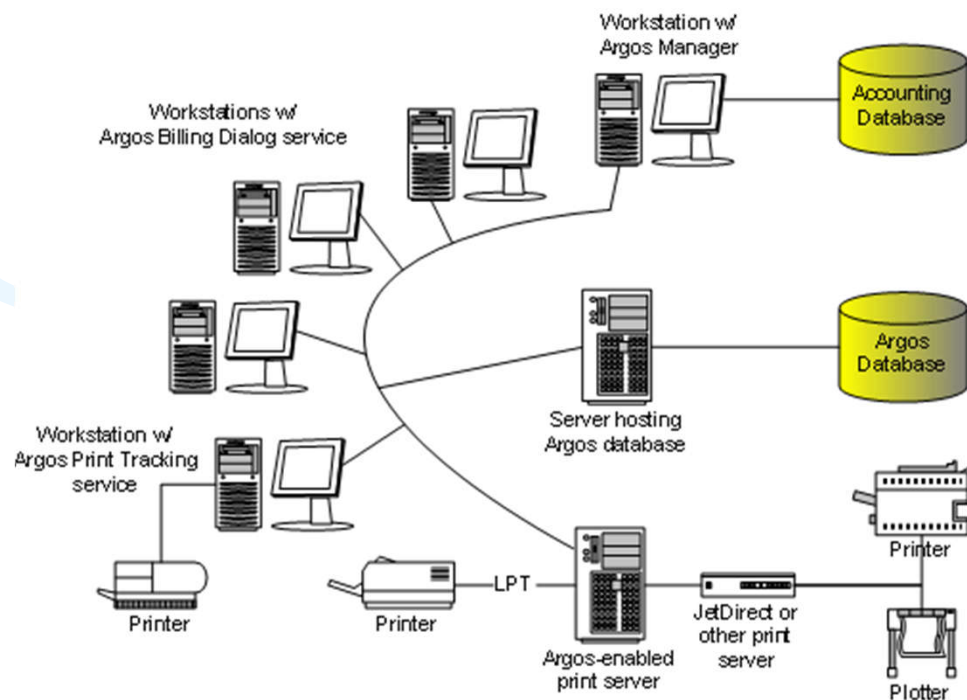
Các hệ CSDL tập trung

- Hệ CSDL cá nhân: một người sử dụng đơn lẻ vừa thiết kế, tạo lập CSDL, cập nhật, bảo trì dữ liệu, lập và hiển thị báo cáo.
 - đảm nhiệm vai trò: người quản trị CSDL, người viết chương trình ứng dụng, end-user.
- Hệ CSDL trung tâm: dữ liệu được lưu trữ trên một máy tính trung tâm.
- Hệ CSDL khách-chủ:
 - Các máy tính trung tâm lớn → đặt so với các máy nhỏ và máy trạm.
 - Các ứng dụng máy khách truy nhập dữ liệu được quản lý bởi máy chủ.

Các hệ CSDL tập trung (tiếp)



Hệ CSDL cá nhân



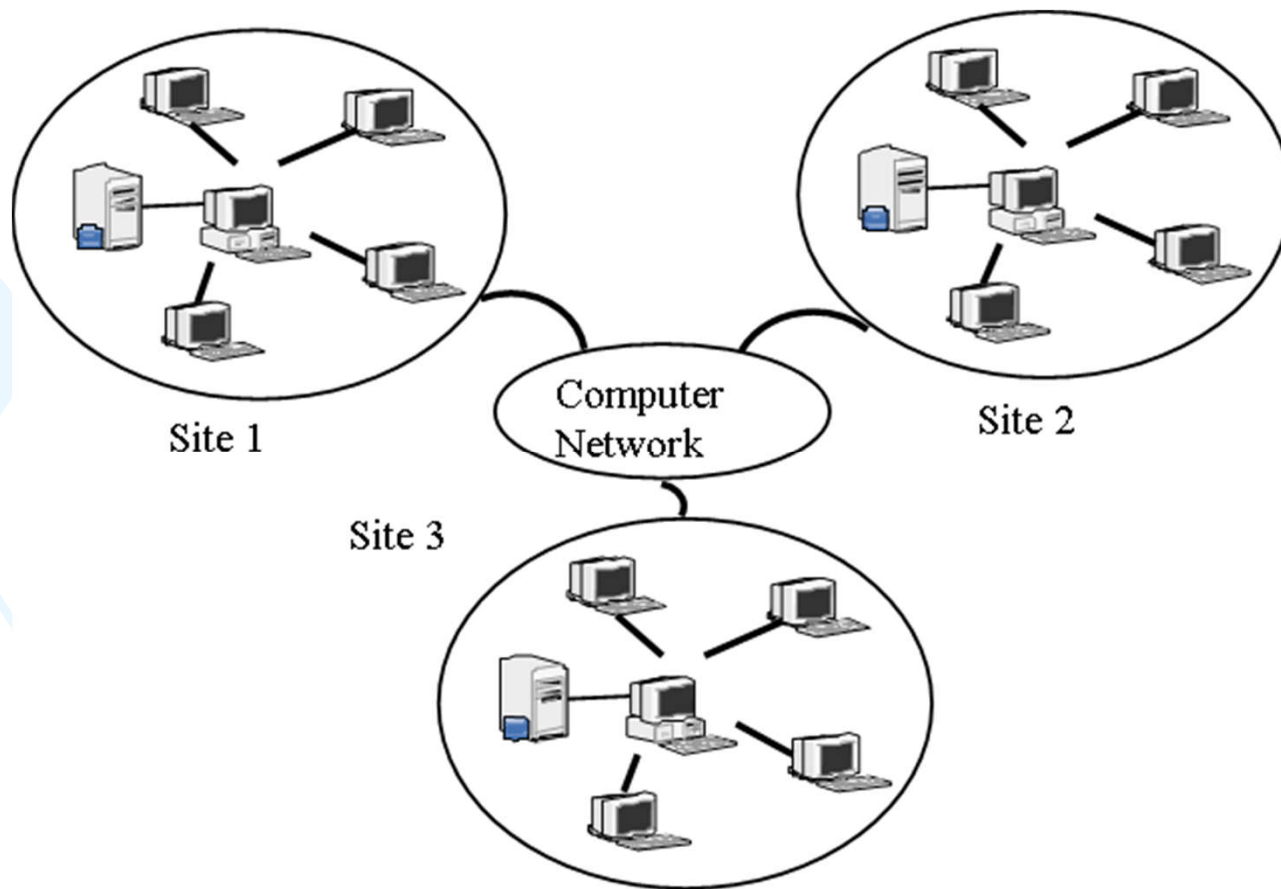
Hệ CSDL khách-chủ



Các hệ CSDL phân tán

- CSDL phân tán? Là một tập các CSDL có quan hệ logic với nhau nhưng được trải ra trên nhiều trạm làm việc của một mạng máy tính.
- Có 2 tính chất: quan hệ logic và phân tán
- Hệ QTCSDL phân tán: Là một hệ thống phần mềm cho phép tạo lập CSDLPT và điều khiển các truy nhập đối với CSDLPT này.
- Chia ra 2 loại: CSDLPT thuần nhất và không thuần nhất

Các hệ CSDLPT (tiếp)





Kết luận

- CSDL cho phép lưu trữ và khai thác dữ liệu một cách thống nhất và hiệu quả (đặc biệt trong trường hợp khối lượng dữ liệu lớn).
- Sự trừu tượng về dữ liệu và tính độc lập dữ liệu cho phép phát triển ứng dụng “dễ dàng hơn”.
- Hệ quản trị CSDL cung cấp các công cụ hữu hiệu trợ giúp việc tạo lập CSDL và phát triển ứng dụng

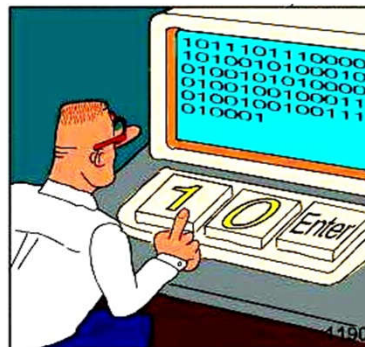
Sử dụng kiến thức môn học này trong tương lai

"More than 80 % of real world computer applications are associated with databases"*

* Korth & Silberschatz. Database System Concepts.



Phát triển ứng dụng



REAL Programmers code in BINARY.

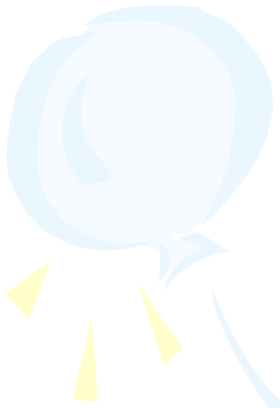
nghiên cứu và phát triển



Nghiên cứu

Các điểm cần lưu ý trong chương này

- Cách tiếp cận tệp vs. cách tiếp cận CSDL
- CSDL vs. hệ QTCSDL vs. hệ CSDL
- Kiến trúc 3 mức của hệ CSDL
- Các chức năng chính của một hệ QTCSDL
- Người sử dụng trong một hệ CSDL
- Phân loại các hệ CSDL





Lời hay ý đẹp

**Điều chúng ta biết chỉ là một giọt nước,
điều không biết mênh mông như đại dương**

Einstein