

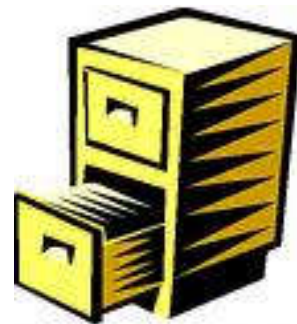
The background features abstract, colorful swirls in shades of green, purple, and blue, interspersed with small yellow triangles. The text is centered and underlined with a red line.

Chương 1

Đại cương về các hệ cơ sở dữ liệu

Nội dung chương

- 1.1 Các hệ thống xử lý tệp truyền thống.
- 1.2 Các hệ CSDL: các khái niệm cơ bản, các tính năng của một hệ quản trị CSDL và những người dùng của một hệ 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 được viết bởi một nhà tin học (kỹ sư lập trình) sử dụng một ngôn ngữ lập trình nào đó.
- 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 đảm bảo 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

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





1.2 Các hệ cơ sở dữ liệu

- Cơ sở dữ liệu (CSDL)
(tiếng Anh: DataBase-DB) là gì ?
- Hệ quản trị CSDL (DataBase Management System-DBMS) là gì ?
- 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 của Khoa Công nghệ thông tin ?
 - Điểm thi môn “Hệ cơ sở dữ liệu” ?
- Giáo viên
 - Danh sách sinh viên lớp Tin2-K49 ?
 - Thời khóa biểu của lớp Tin2-K49 ?
- Giáo vụ
 - Danh sách sinh viên K47 tốt nghiệp loại giỏi ?...

➔ **Chương trình ứng dụng**



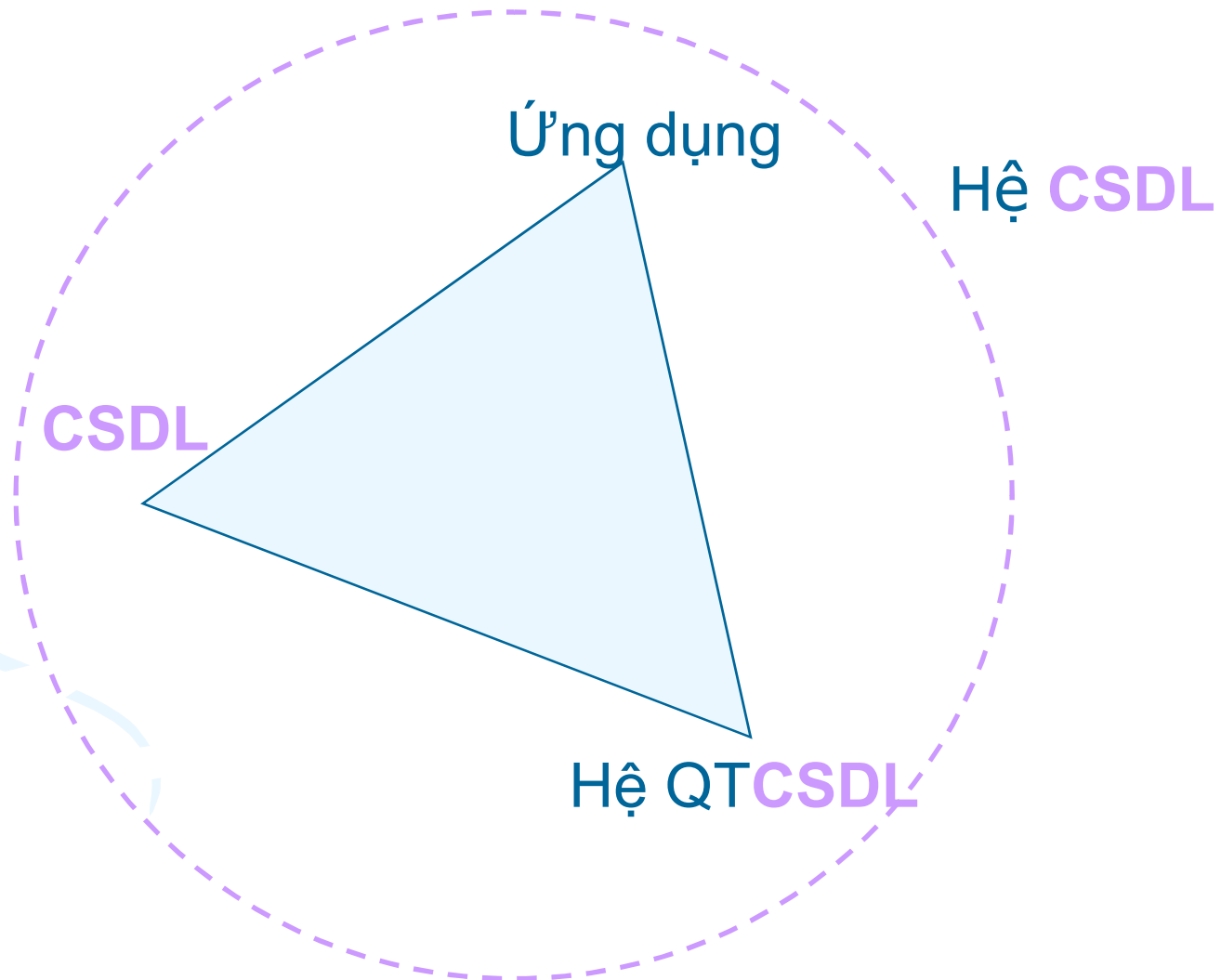
“Hình dung” về xây dựng một CSDL

- 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**



Các khái niệm cơ bản



Cơ sở dữ liệu (DataBase)

- Là một bộ sưu tập các dữ liệu tác nghiệp
 - Là các dữ liệu của một xí nghiệp
 - Có quan hệ logic với nhau
 - Được lưu trữ để phục vụ cho một ứng dụng cần tin học hóa.
- CSDL là một bộ sưu tập các dữ liệu tác nghiệp được lưu trữ lại để phục vụ cho các ứng dụng của một xí nghiệp cụ thể nào đó.

Hệ quản trị cơ sở dữ liệu

(DataBase Management System-DBMS)

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



Hệ cơ sở dữ liệu

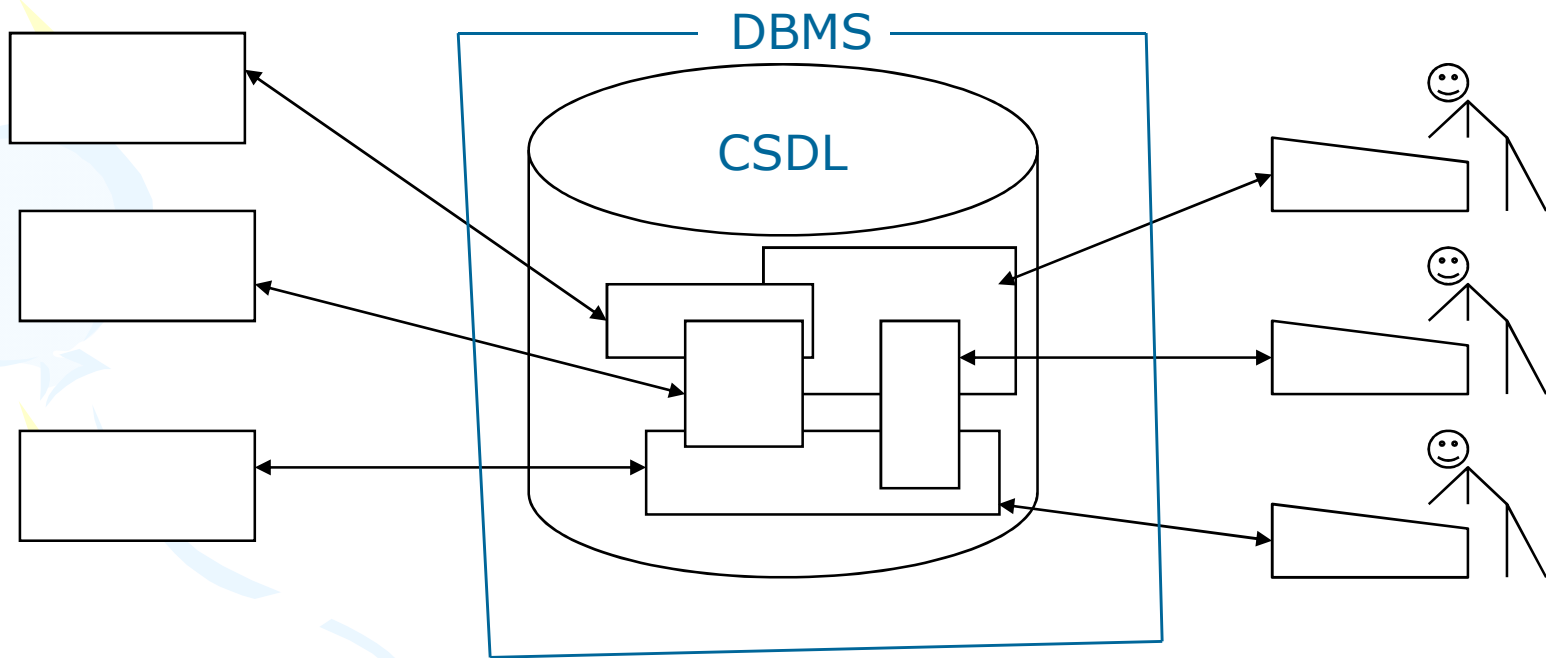
- Là một hệ thống gồm 4 thành phần
 - CSDL và các ứng dụng
 - Những người sử dụng
 - Phần mềm DBMS
 - Phần cứng
- Ví dụ: Hệ quản lý đào tạo, hệ quản lý nhân sự, hệ quản lý kinh doanh,...



Hệ CSDL

Các chương trình ứng dụng

Người sử dụng trực tuyến



Các tính năng của DBMS

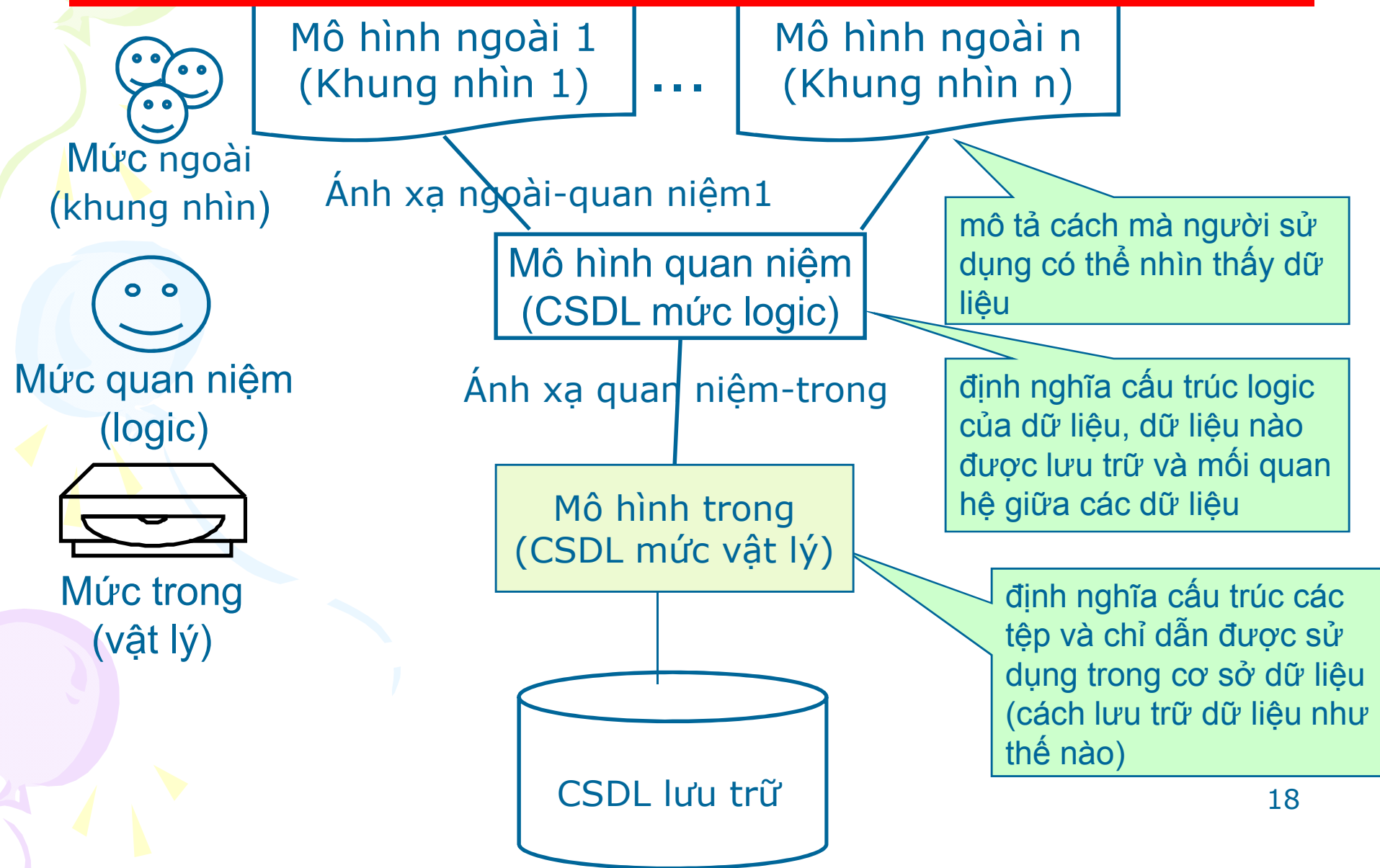
- Quản lý dữ liệu tồn tại lâu dài
 - Định nghĩa dữ liệu
 - Quản lý lưu trữ
- Truy xuất dữ liệu một cách hiệu quả
 - Cung cấp các thao tác dữ liệu
 - Xử lý và tối ưu hóa câu hỏi
 - Quản trị giao dịch



Các tính năng của DBMS

- 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 đồng thời có cạnh tranh
- Sao lưu và phục hồi dữ liệu
- Đảm bảo an toàn và toàn vẹn dữ liệu

Sự trừu tượng hóa dữ liệu (Kiến trúc của CSDL)



Tương ứng 3 mức với ngôn ngữ Pascal

```
Type khách_hang = record
```

```
    ten:string;
```

```
    ngay_sinh:string;
```

```
    dia_chi:string;
```

```
end;
```

- Mức vật lý: một bản ghi khách_hang được mô tả như một khối nhớ, chương trình dịch che dấu các chi tiết mức này đối với người lập trình.
- Mức logic: mỗi bản ghi được mô tả bởi một định nghĩa kiểu, người lập trình sử dụng ngôn ngữ lập trình làm việc tại mức trừu tượng này.
- Mức khung nhìn: người sử dụng máy tính thấy một tập các chương trình ứng dụng, che dấu những chi tiết về các kiểu dữ liệu



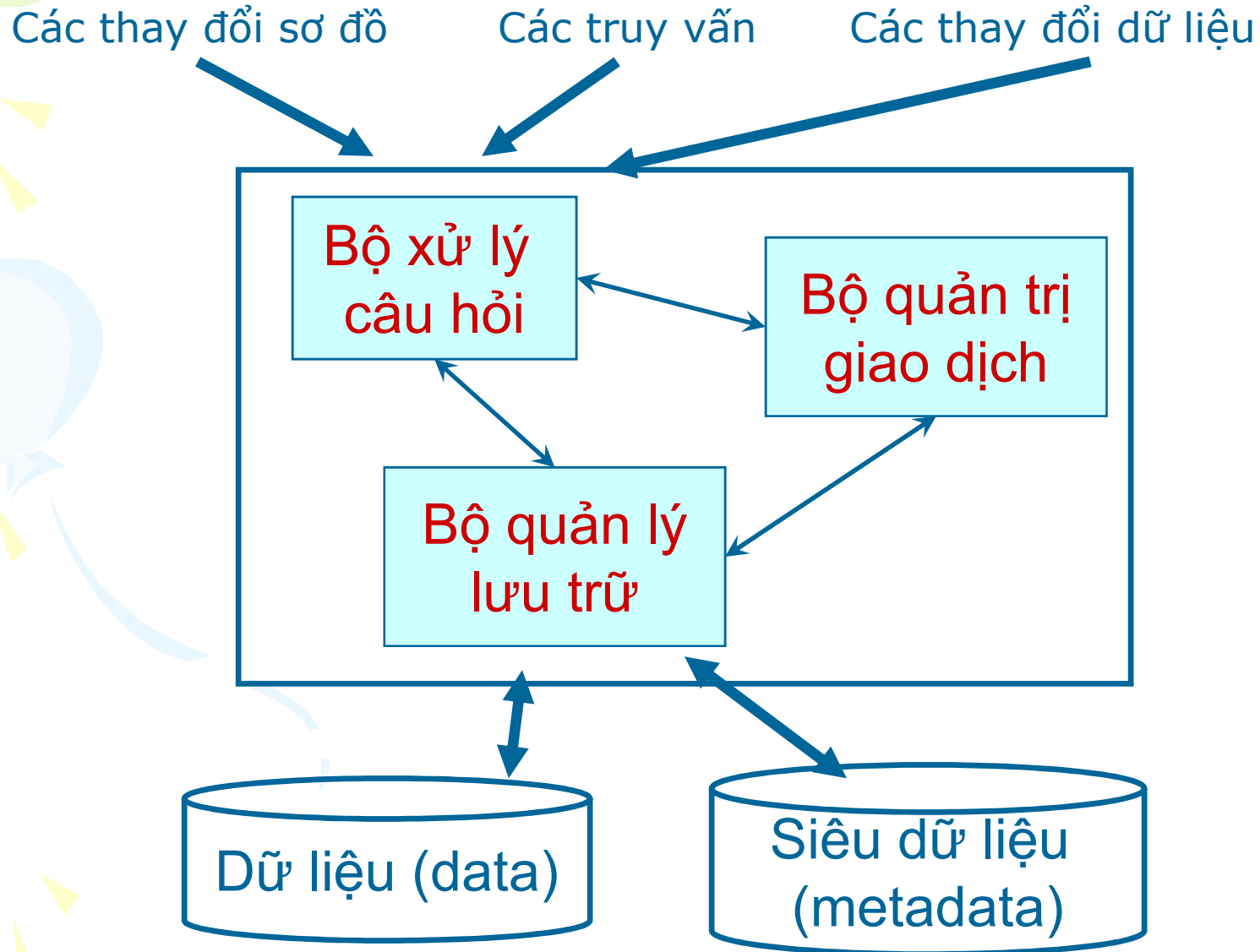
Tính độc lập dữ liệu

- Độc lập dữ liệu mức vật lý: các chương trình ứng dụng là bất biến đối với các thay đổi trong sơ đồ vật lý.
- Độc lập dữ liệu mức logic: các chương trình ứng dụng là bất biến đối với các thay đổi trong sơ đồ logic.

Các ngôn ngữ được hỗ trợ bởi DBMS

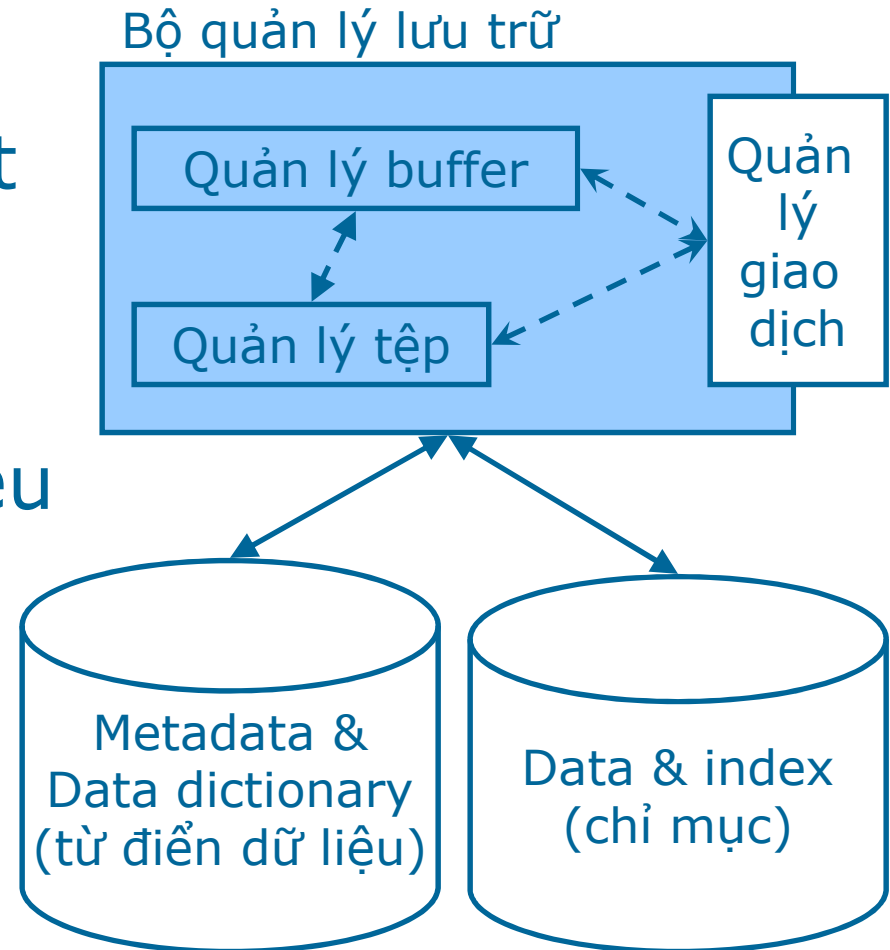
- Ngôn ngữ định nghĩa dữ liệu (**D**ata **D**efinition **L**anguage - DDL)
 - Mô tả cấu trúc các đối tượng trong CSDL
 - Khai báo mối liên hệ giữa các dữ liệu và các quy tắc, ràng buộc áp đặt lên dữ liệu
 - Thay đổi các mô tả cấu trúc
- Ngôn ngữ thao tác dữ liệu (**D**ata **M**anipulation **L**anguage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

Kiến trúc của một hệ quản trị CSDL



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



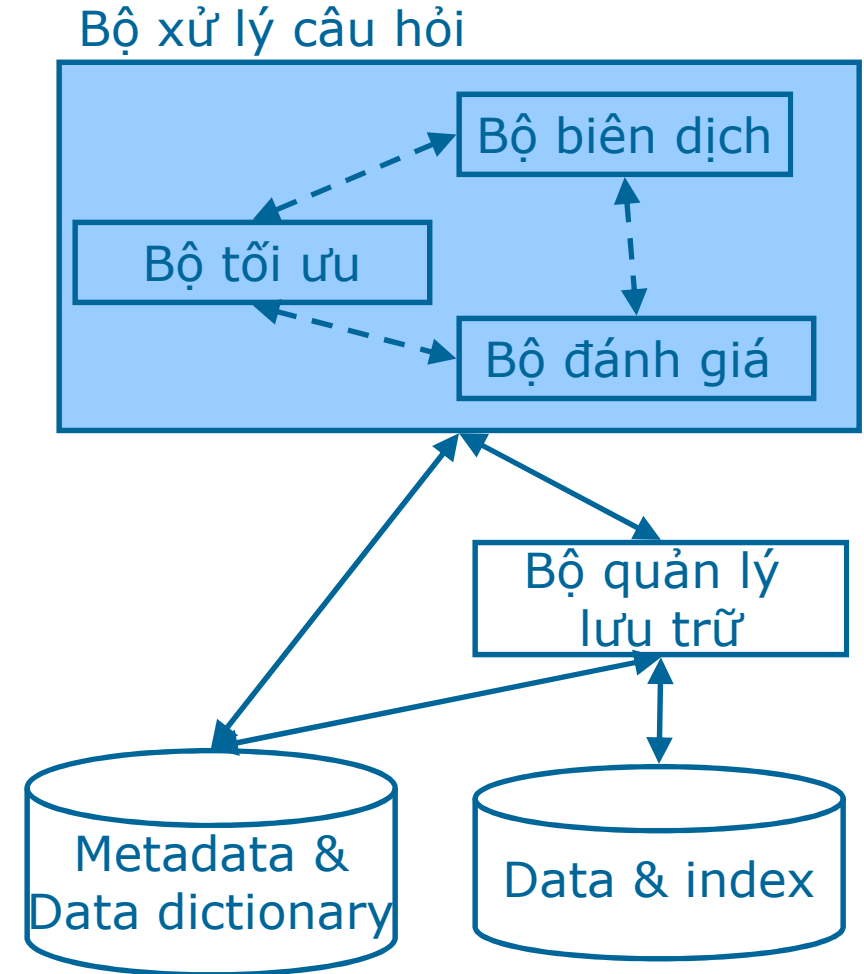
Xử lý câu hỏi

- 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 thiết kế CSDL:** chịu trách nhiệm xác định các dữ liệu được lưu trữ trong CSDL và các cấu trúc biểu diễn cũng như lưu trữ những dữ liệu này
- **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 cho việc hoạt động và bảo trì môi trường (phần cứng và phần mềm) cho hệ CSDL

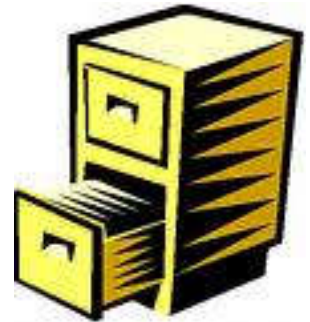


Người dùng

- **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, lập trình ứng dụng cài đặt những yêu cầu này trong chương trình, kiểm thử, gỡ rối, lập tài liệu cho chương trình
- **Người sử dụng cuối:** là người khai thác các hệ CSDL từ terminal.

Nội dung chương

- 1.1 Các hệ thống xử lý tệp truyền thống và những hạn chế của nó.
- 1.2 Các hệ CSDL: các khái niệm cơ bản, các tính năng của một hệ quản trị CSDL và những người dùng của một hệ CSDL.
- 1.3 Phân loại các hệ CSDL.



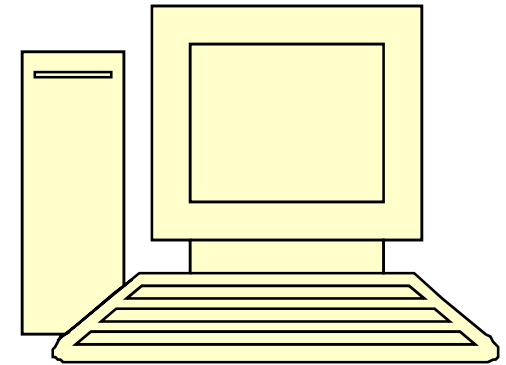
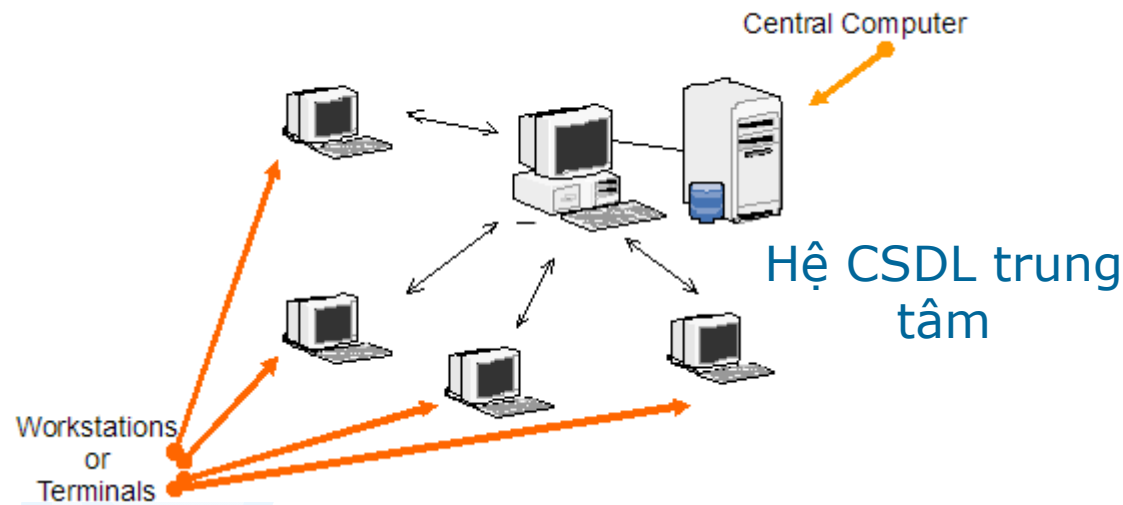
1.3 Phân loại các hệ CSDL

- 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
- Nhu cầu khai thác CSDL
 - OLTP vs. OLAP, KDD

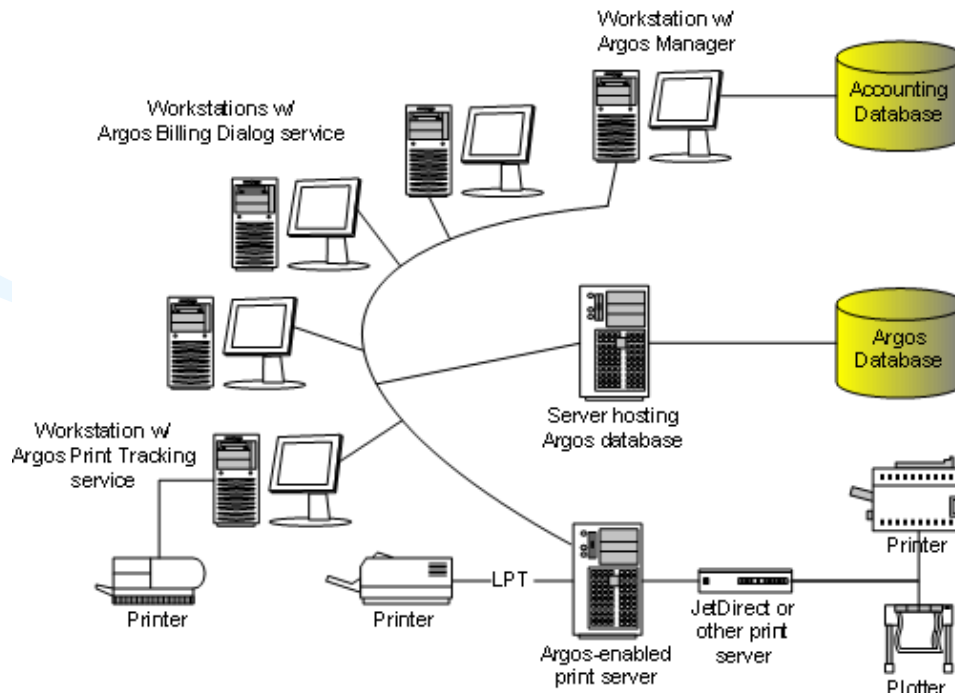
Các hệ CSDL tập trung

- Hệ CSDL cá nhân: một người sử dụng vừa thiết kế, tạo lập CSDL, viết các chương trình tin học hóa một vài ứng dụng đơn lẻ và cập nhật, khai thác dữ liệu.
 - 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, người dùng truy cập CSDL từ các terminal.
- 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

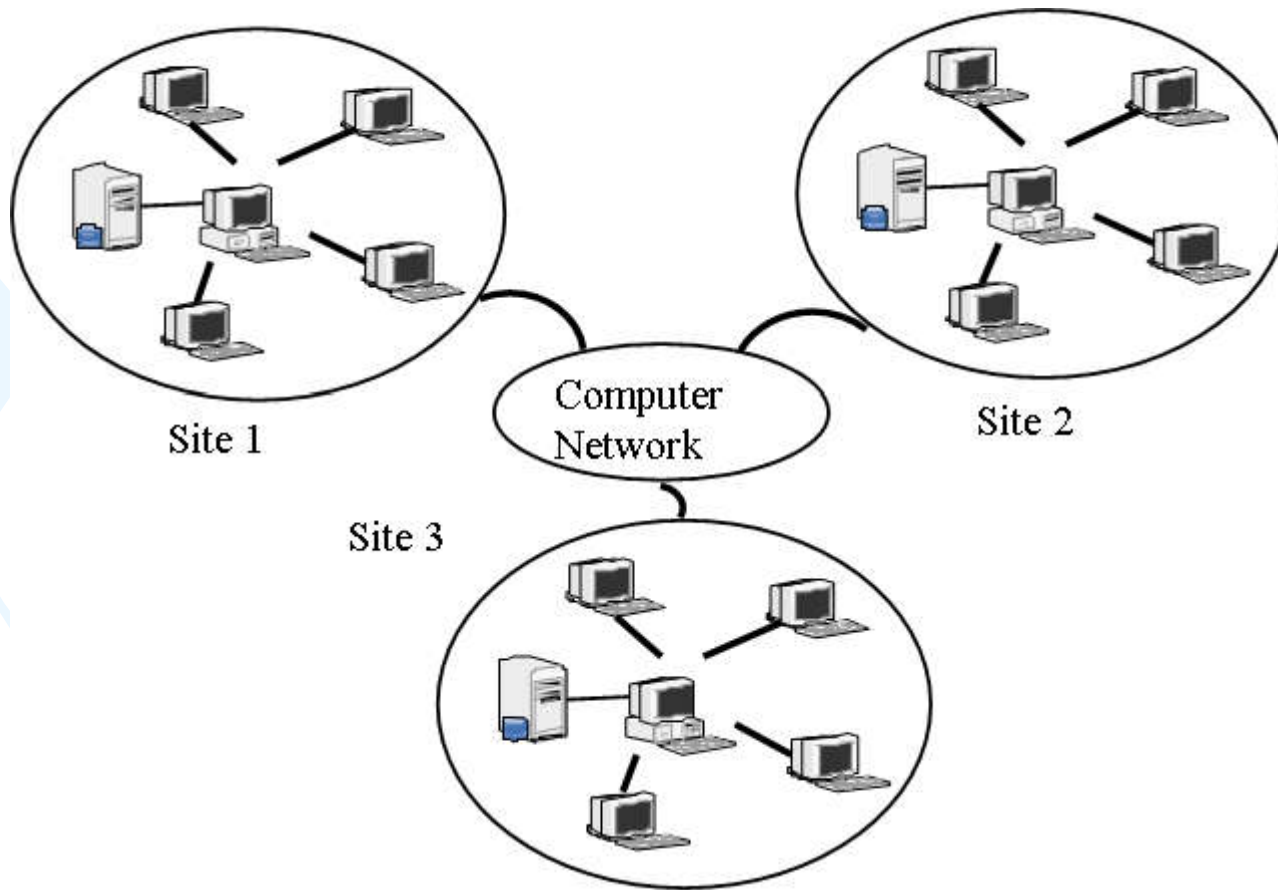


Hệ CSDL khách-chủ

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

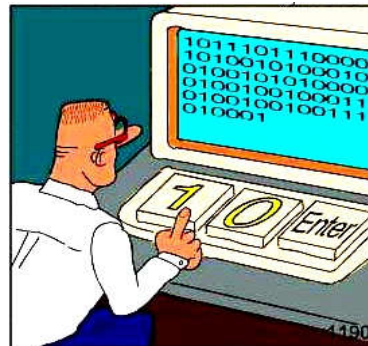
Sử dụng kiến thức môn học này trong tương lai

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

* Korth & Silberschatz. Database System Concepts.

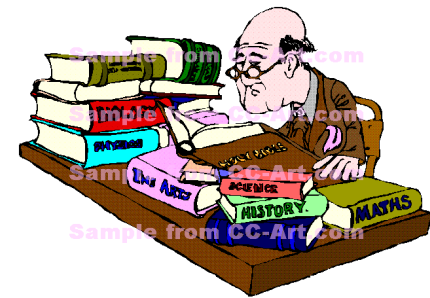


Phát triển ứng dụng



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

