



# Ch1. Đại cương về các hệ CSDL

---

Nguyễn Thị Oanh

[oanhnt@soict.hust.edu.vn](mailto:oanhnt@soict.hust.edu.vn)

Bộ môn Hệ thống thông tin (<http://is.hust.edu.vn/>)

Viện CNTT&TT

Trường Đại học Bách Khoa Hà Nội

# Nội dung

---

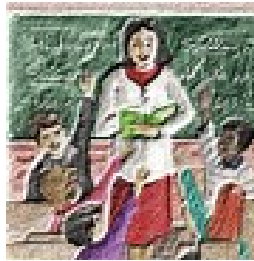
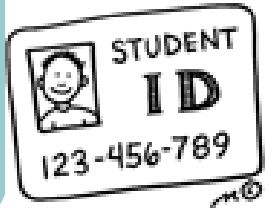
- ✓ Hệ thống xử lý tệp truyền thống
- Hệ CSDL: các khái niệm cơ bản
- Phân loại hệ CSDL

# Hệ thống xử lý tệp truyền thống

---

- Các ứng dụng phát triển do nhu cầu riêng rẽ của từng bộ phận phòng ban, từng nhóm cụ thể nào đó mà không có kế hoạch tổng thể. VD:
  - CT quản lý nhân sự (giảng viên, kỹ thuật viên, chuyên viên, ...)
  - CT quản lý thông tin SV
  - CT quản lý học tập của SV
  - CT đánh giá điểm rèn luyện của SV
- Hệ thống xử lý tệp được hỗ trợ bởi hệ điều hành, các bản ghi khá ổn định và được tồn tại lâu dài trong các tệp
- Mỗi ứng dụng :
  - định nghĩa và quản lý các tệp DL của riêng nó
  - ngôn ngữ khác nhau, quy cách biểu diễn DL trong các tệp khác nhau

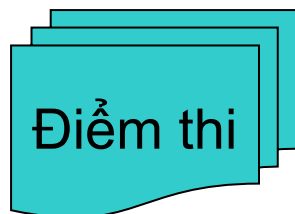
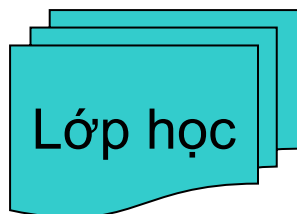
# Hệ thống xử lý tệp truyền thống



Sinh viên

Học tập

Giáo viên





# Hạn chế

---

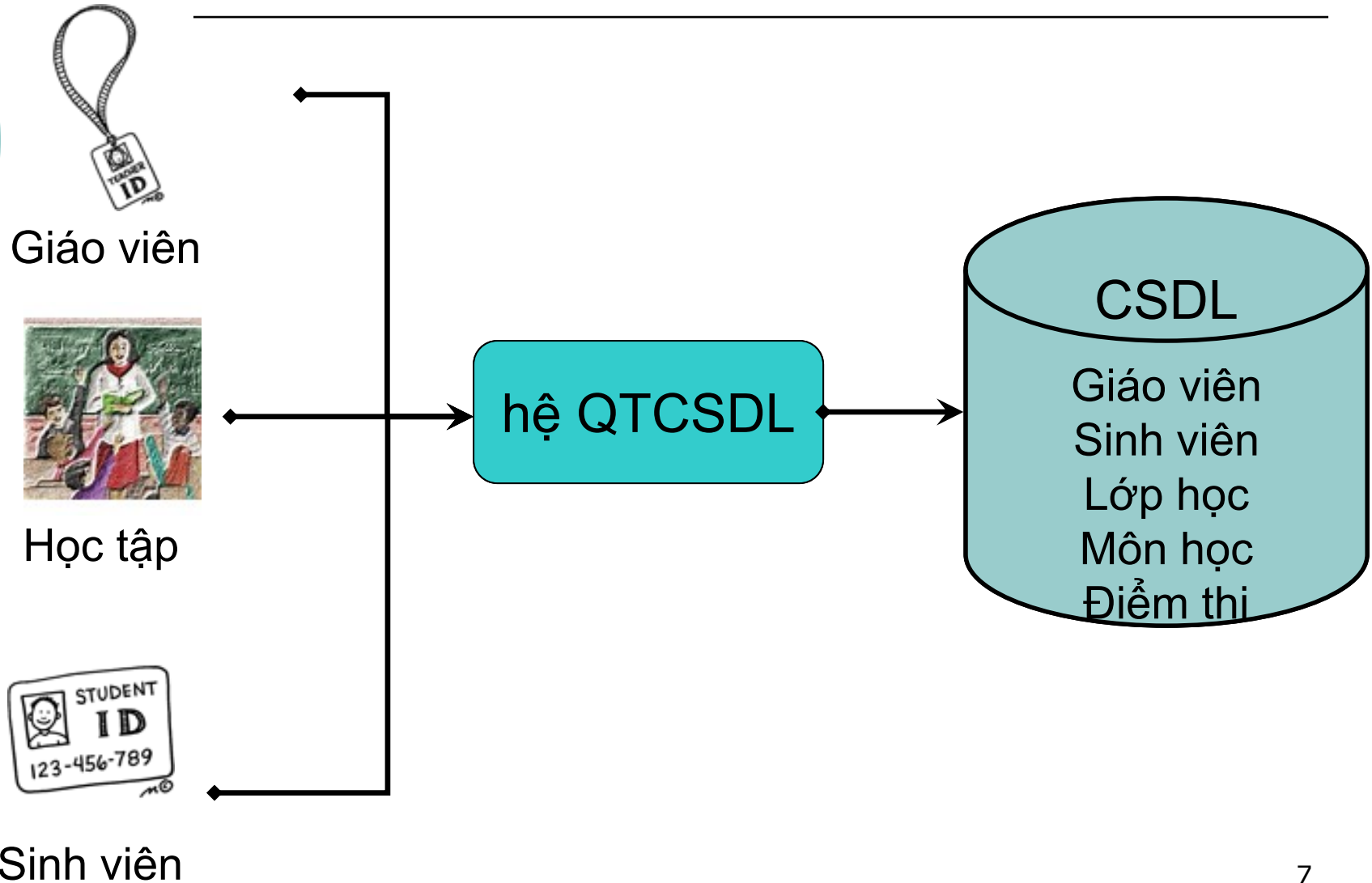
- Mức độ diễn tả ngữ nghĩa hạn chế
- Dữ liệu riêng lẻ, rời rạc
- Quản lý, khai thác ở mức thấp

# Hạn chế

---

- Dư thừa và không nhất quán DL
- Khó khăn trong truy cập DL
  - y/c DL không được dự tính trước
- Cô lập và hạn chế chia sẻ DL
- Các vấn đề về toàn vẹn
- Các vấn đề về độ tin cậy:
  - Sự cố khi đang thực hiện CV: chuyển tiền
- Các dị thường khi truy nhập đồng thời:
  - 2 y/c được thực hiện đồng thời: rút tiền 50\$, 100\$
- Các vấn đề về an toàn:
  - khó QL việc truy nhập DL của người sử dụng
- Sự phụ thuộc DL của các CT ứng dụng

# Hệ cơ sở dữ liệu





# Ưu điểm của cách tiếp cận CSDL

---

- Tính trừu tượng hoá của dữ liệu
- Hỗ trợ nhiều khung nhìn dữ liệu
- Chia sẻ dữ liệu giữa nhiều người dùng và hỗ trợ quản lý giao dịch



# Đặc điểm của cách tiếp cận CSDL

---

- Biểu diễn ngữ nghĩa phong phú và những quan hệ phức tạp của các dữ liệu
- Kiểm soát tính dư thừa và đảm bảo các ràng buộc toàn vẹn dữ liệu
- Hỗ trợ lưu trữ và truy vấn dữ liệu một cách hiệu quả
- Chia sẻ dữ liệu giữa nhiều người dùng
- Phân quyền người dùng và kiểm soát tính hợp lệ của các truy xuất dữ liệu
- Hỗ trợ dịch vụ sao lưu và phục hồi dữ liệu



# Nội dung

---

- Hệ thống xử lý tệp truyền thống
- ✓ Hệ CSDL: các khái niệm cơ bản
- Phân loại hệ CSDL

# Đặt vấn đề

---

- CSDL là gì?
- Tại sao phải sử dụng CSDL?
- Tại sao phải tìm hiểu về các hệ CSDL?

# Ví dụ: quản lý đào tạo

---

- Thông tin cần quan tâm
  - Khoá học, lớp học, sinh viên, môn học, giáo viên, ...
  - Thông tin về sinh viên: thông tin cá nhân, thông tin học tập
  - Thông tin về môn học: khối lượng học tập, 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 CNTT?
  - Điểm thi môn « Nhập môn CSDL »?
- Giáo viên
  - Danh sách sinh viên lớp Tin 1?
  - Thời khoá biểu của lớp Tin 1?
- Giáo vụ
  - Danh sách sinh viên khoá K47?
  - Tỷ lệ sinh viên thi đạt của từng môn học?
- Cần xây dựng một phần mềm cho phép khai thác một CSDL
  - Phần mềm ứng dụng

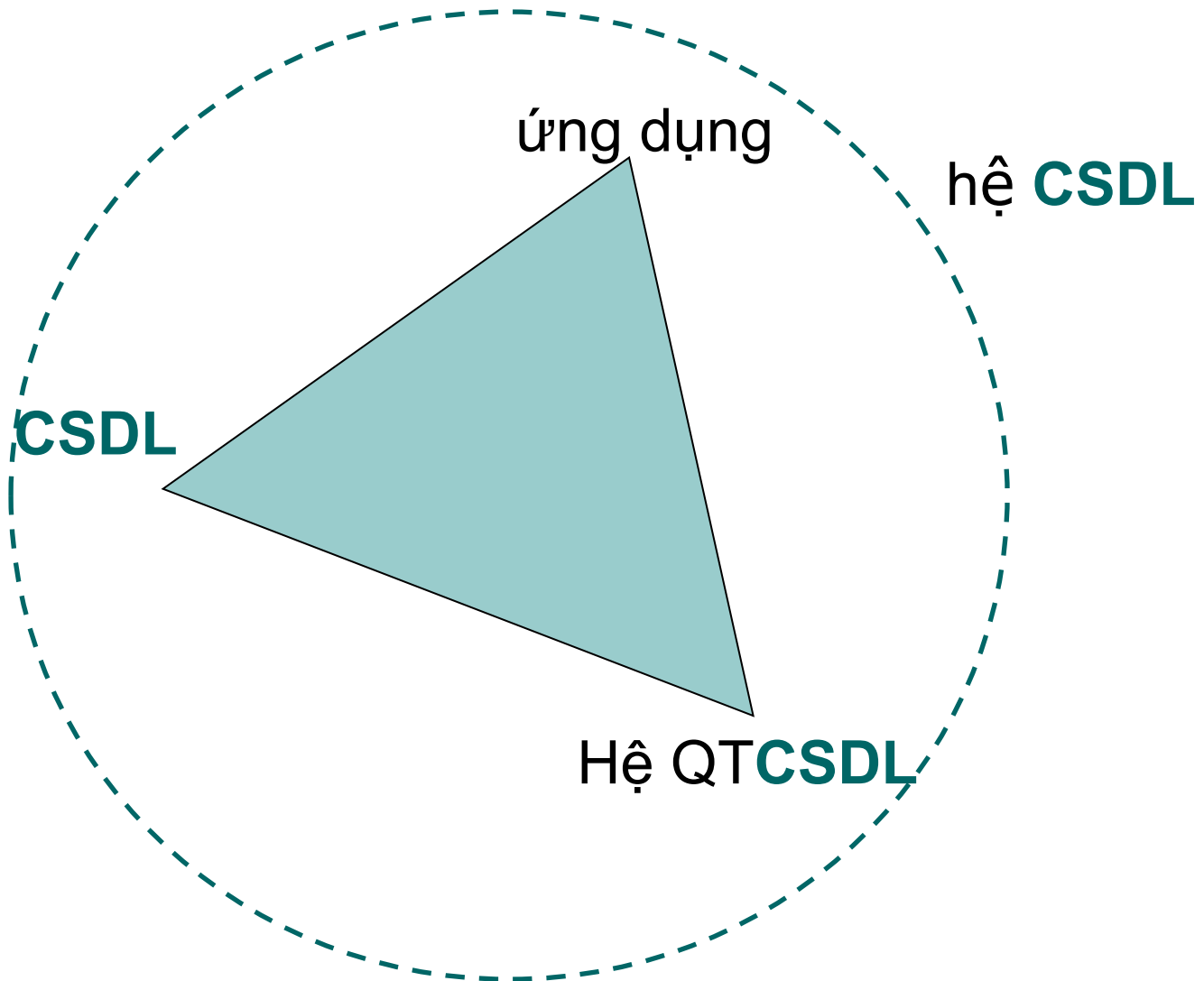
# «Hình dung» về xây dựng một CSDL

---

- 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 những 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 (QTCSDL)

# Các khái niệm cơ bản

---



# Cơ sở dữ liệu (Database)

---

- Một tập hợp các dữ liệu
  - Biểu diễn một vài khía cạnh 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
- *CSDL là một bộ các dữ liệu tác nghiệp được lưu trữ lại và được các hệ ứng dụng của một đơn vị cụ thể nào đó sử dụng*

Ví dụ: CSDL về quản lý đào tạo gồm thông tin về

- giáo viên
- sinh viên
- môn học
- lớp học
- điểm thi, ...



# Hệ quản trị cơ sở dữ liệu

## (Database Management System - DBMS)

---

- Hệ QT CSDL là một phần mềm cho phép *tạo lập* các CSDL cho các ứng dụng khác nhau và *điều khiển* mọi truy cập tới các CSDL đó. Nghĩa là, hệ QTCSDL cho phép:
  - **Định nghĩa**  
xác định *kiểu, cấu trúc, ràng buộc* dữ liệu
  - **Tạo lập**  
*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, ORACLE, MySQL, PostgreSQL  
IBM DB2, MS. Access,...

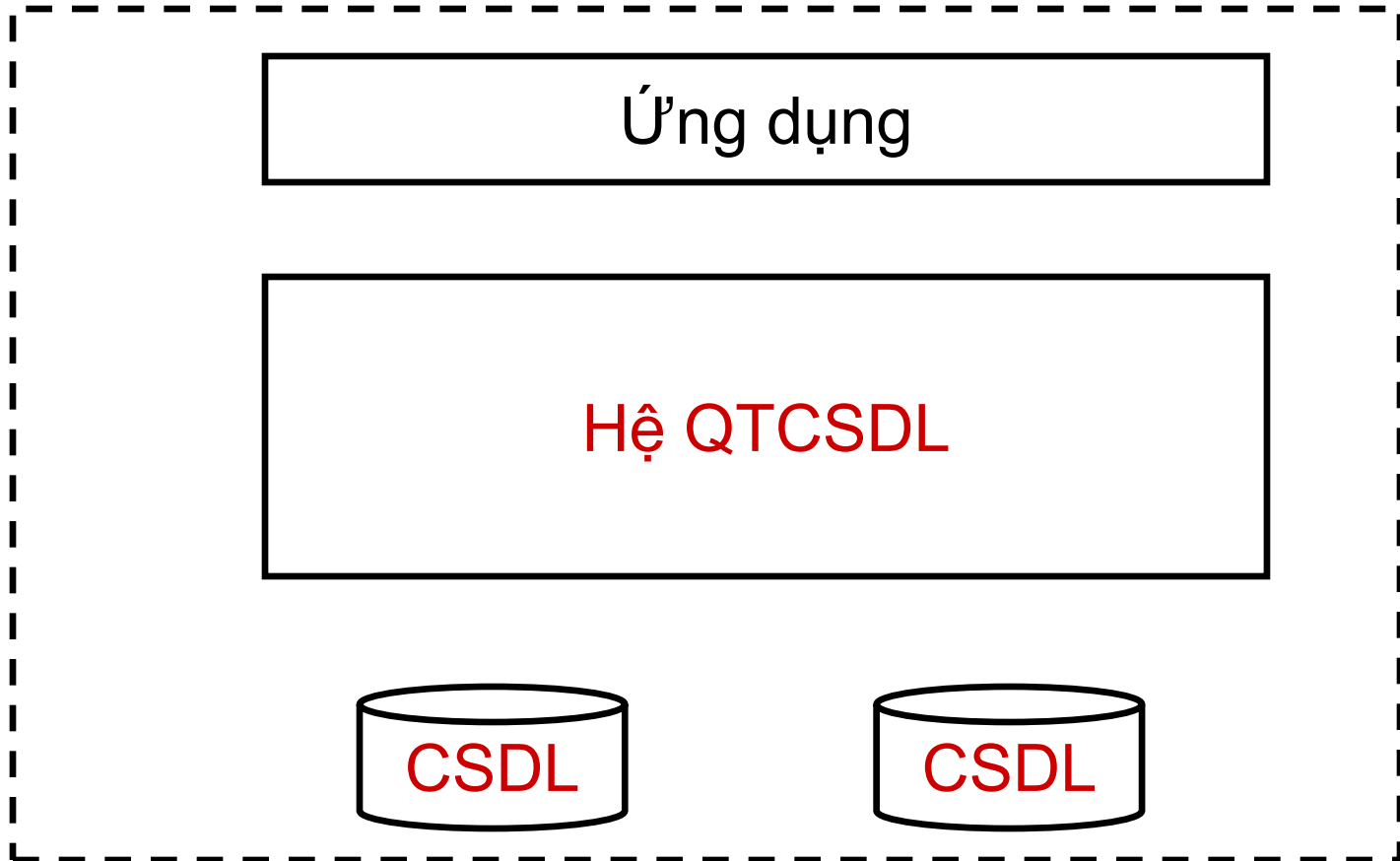
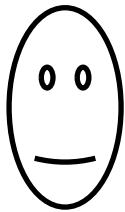
# Hệ cơ sở dữ liệu (Database System)

---

- Là một hệ thống gồm 4 thành phần :
  - Hệ QTCSDL
  - Phần cứng: các thiết bị nhớ thứ cấp dùng lưu trữ CSDL
  - CSDL và phần mềm ứng dụng
  - Những người sử dụng : người sử dụng cuối, người viết chương trình ứng dụng, người điều khiển toàn bộ hệ thống, ...

Ví dụ: hệ quản lý đào tạo, quản lý nhân sự, ...

# Môi trường hệ CSDL



# Chức năng của hệ QTCSDL

---

- 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 ngôn ngữ

---

- Ngôn ngữ định nghĩa dữ liệu (Data Definition Language - *DDL*)
  - Cấu trúc dữ liệu
  - Mối liên hệ giữa các dữ liệu và các quy tắc, ràng buộc áp đặt lên dữ liệu
- Ngôn ngữ thao tác dữ liệu (Data Manipulation Language - *DML*)
  - Tìm kiếm, thêm, xóa, sửa dữ liệu trong CSDL
- Ngôn ngữ điều khiển dữ liệu (Data Control Language - *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

# Các ngôn ngữ

- Ngôn ngữ thao tác dữ liệu (Data Manipulation Language - *DML*) có 2 loại:

- DML thủ tục: **WHAT + HOW**

NSD đặc tả DL được tìm kiếm và tìm kiếm thế nào  
(*SQL, đại số quan hệ*)

```
SELECT Name  
FROM Student, Enrol  
WHERE ID = SID
```

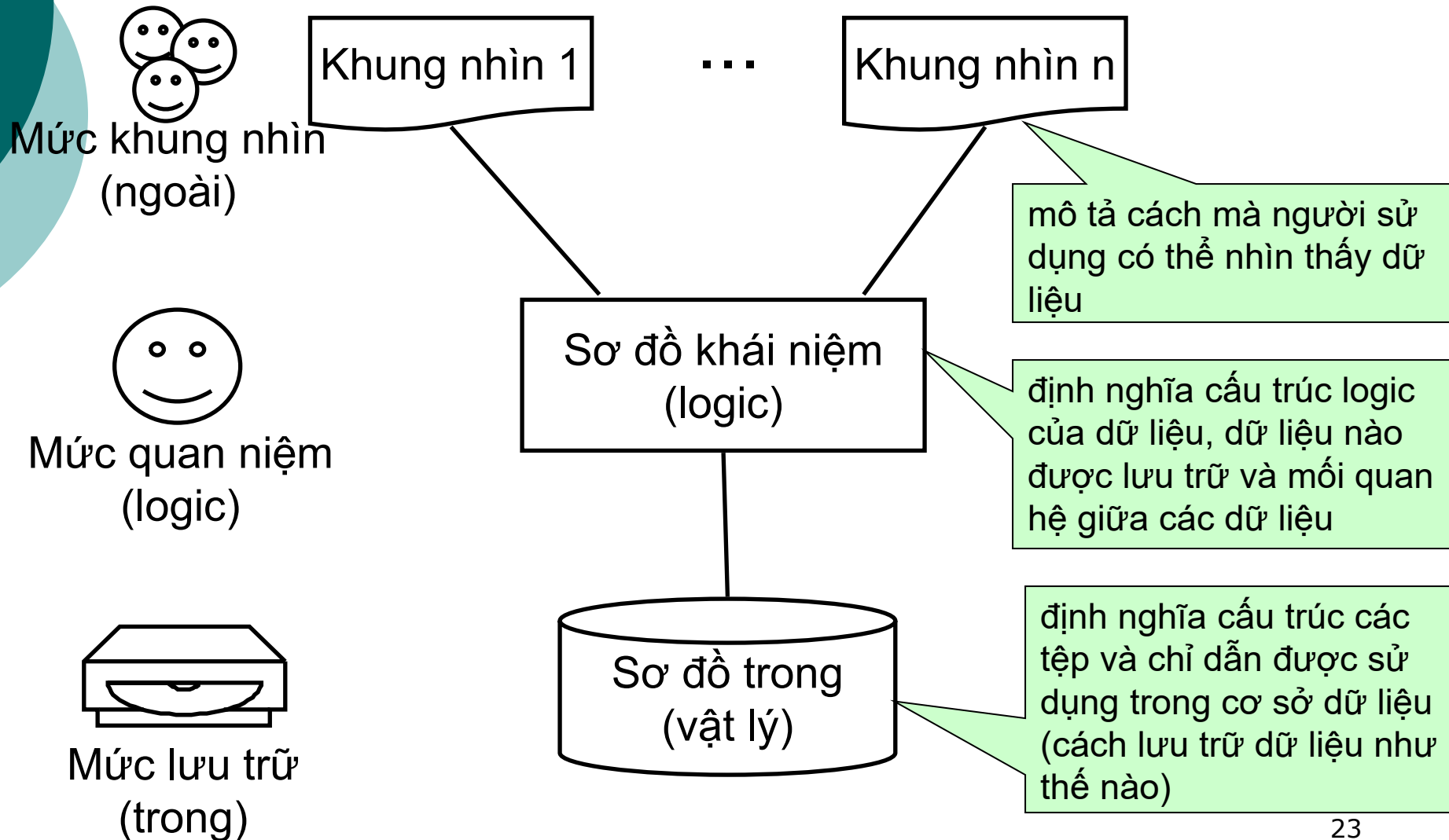
- DML phi thủ tục: **WHAT**

NSD chỉ cần đặc tả DL cần tìm kiếm mà không cần mô tả tìm kiếm thế nào (*QBE, vị từ*)

Student	ID	Name	Suburb
	_id	P._name	

Enrol	SID	Course
	_id	

# Sự trừu tượng hoá dữ liệu



# Ví dụ

---

- Mức quan niệm

```
type lop = record  
    ma_lop : string;  
    ten: string;  
    heDT: string;  
    dia_diem: string;  
end;
```

```
type sinh_vien = record  
    maSV : string;  
    ten: string;  
    nam: boolean;  
    ngay_sinh: date;  
    dia_chi: string ;  
    ma_lop: string;  
end;
```

```
Tin1 = {  
    ma_lop=1  
    ten=« CNTT1_K47 »  
    heDT=« chinh quy »  
    dia_diem=« DHBKHN »  
}
```

```
NVA = {  
    maSV=« SV001 »  
    ten=« Nguyễn Văn A »  
    nam=1;  
    ngay_sinh=« 1/4/1983 »  
    dia_chi=« 1 Tạ Quang Bửu »  
    ma_lop=1  
}
```



## Ví dụ (2)

---

- Mức khung nhìn

```
type ds_sinhvien = record  
    ten: string;  
    nam: boolean;  
    ngay_sinh: date;  
end;
```

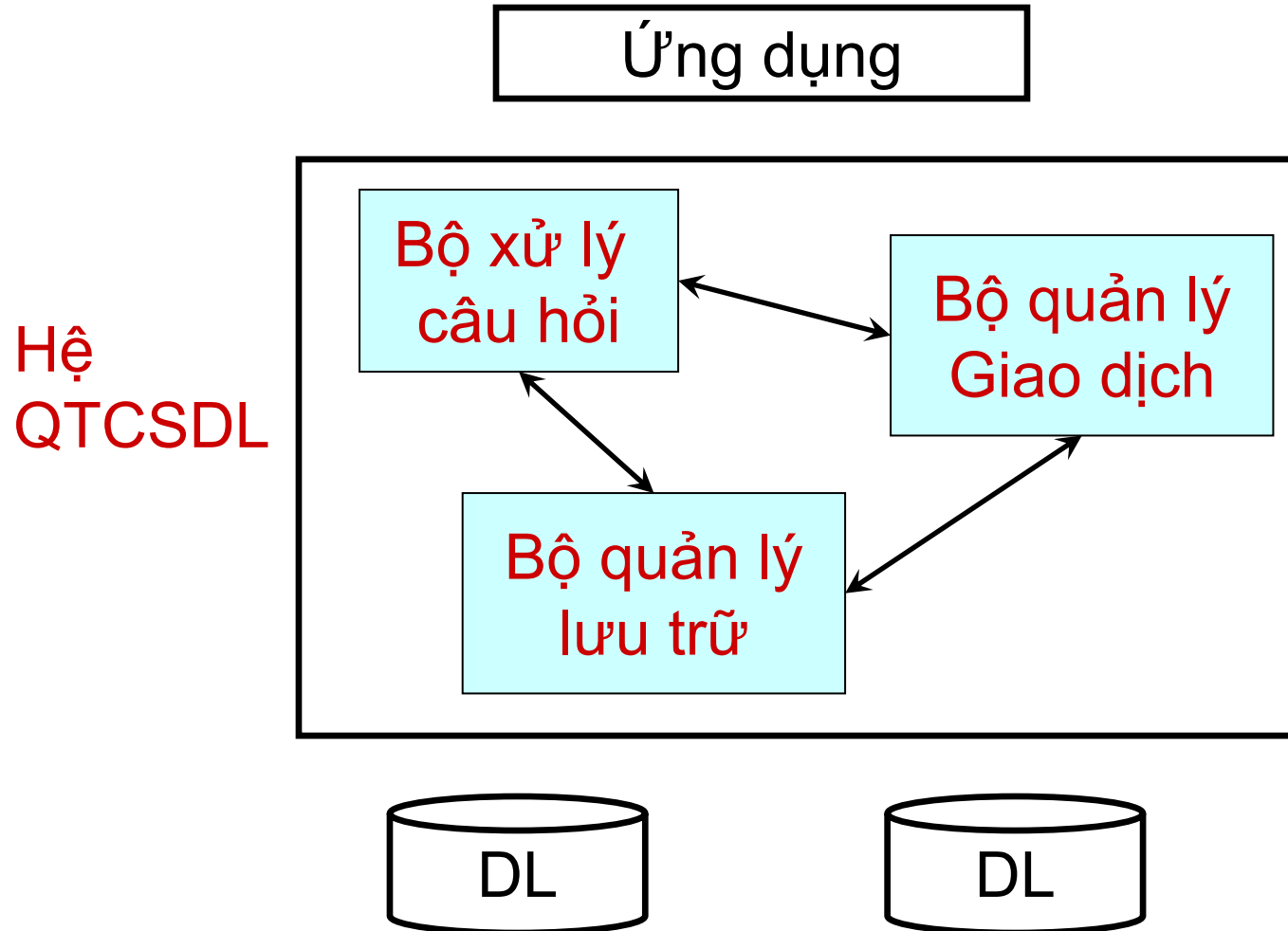
```
CN_NVA = {  
    ten=« Nguyễn Văn A »  
    nam=1;  
    ngay_sinh=« 1/4/1983 »  
}
```

# Sơ đồ và thể hiện (*schema vs. instance*)

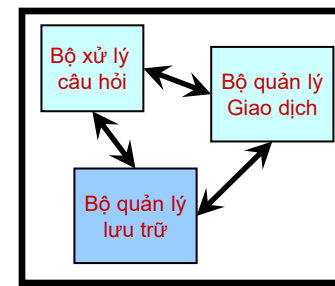
---

Sơ đồ	Thể hiện
-cấu trúc/kiểu dữ liệu	- giá trị/đối tượng dữ liệu
-ít thay đổi -thay đổi phức tạp	-thường xuyên thay đổi -dễ dàng thay đổi

# Các module chính của hệ QTCSDL



# Quản lý lưu trữ

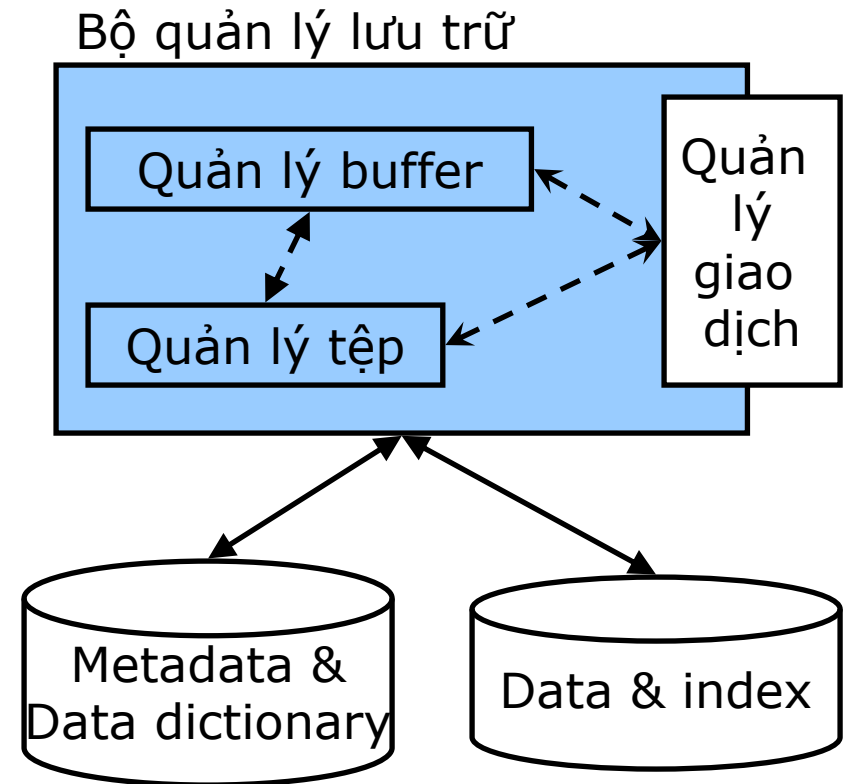


## ○ Yêu cầu

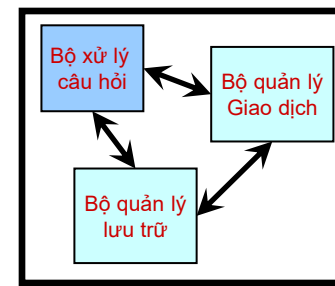
lưu trữ và truy xuất dữ liệu trên các *thiết bị nhớ (ngoài)*

## ○ Thực hiện

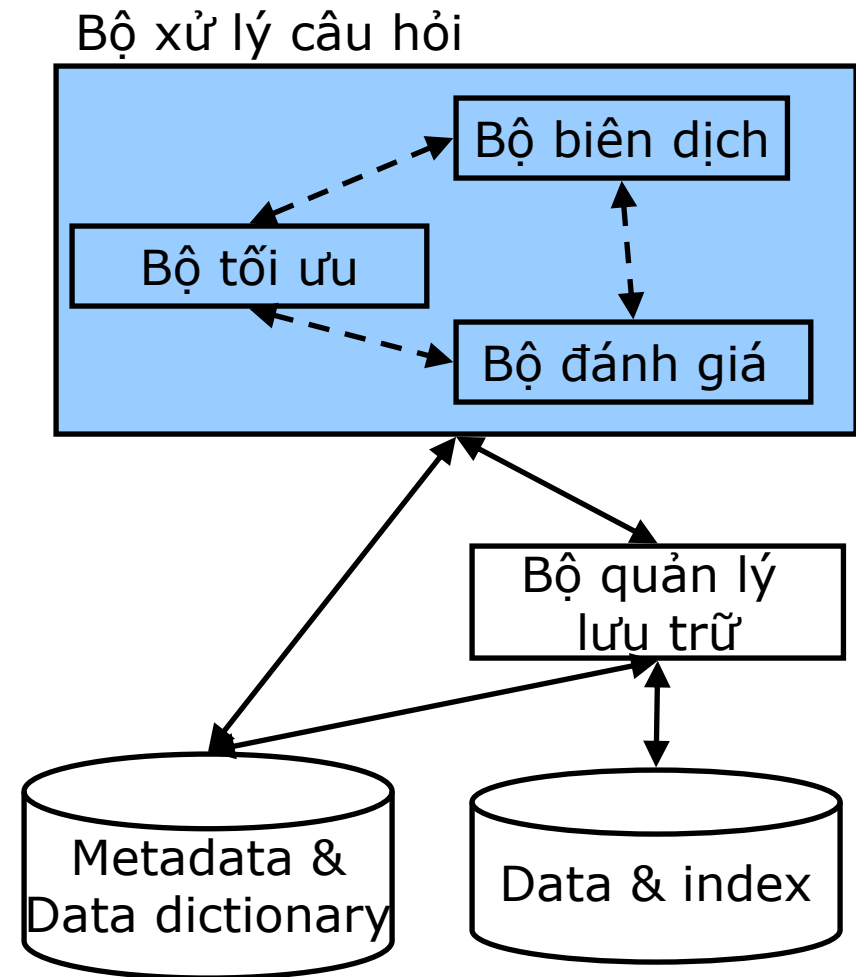
- Tổ chức tối ưu dữ liệu trên thiết bị nhớ (*ngoài*)
- Tương tác hiệu quả với bộ quản lý tệp



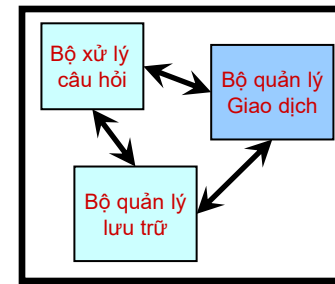
# Xử lý câu hỏi



- Yêu cầu:
  - Tìm kiếm dữ liệu trả lời cho một yêu cầu truy vấn
- Thực hiện
  - Biến đổi truy vấn biểu diễn ở một 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 lý giao dịch



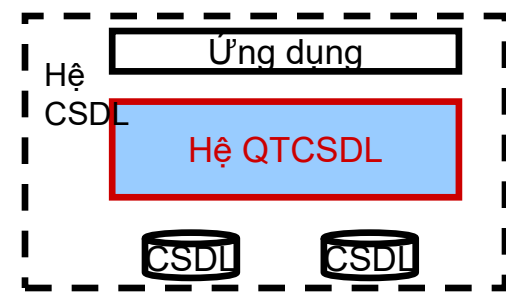
## ○ Yêu cầu

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

## ○ Thực hiện

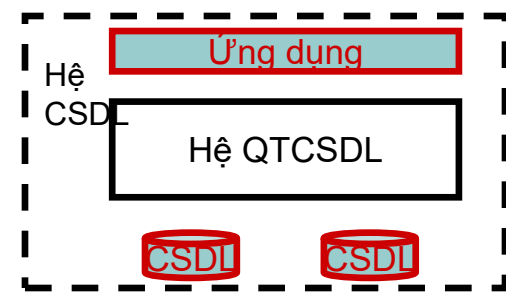
- Quản lý điều khiển tương tranh
- Phát hiện lỗi và phục hồi cơ sở dữ liệu

# 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.

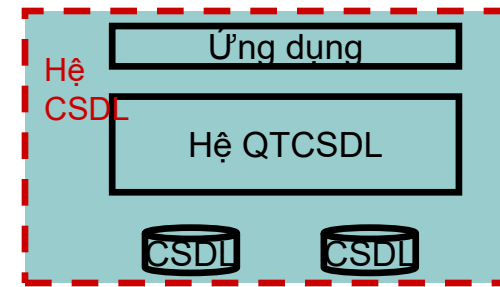
# Người dùng



- **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

---

- Hệ thống xử lý tệp truyền thống
- Hệ CSDL: các khái niệm cơ bản
- ✓ Phân loại hệ CSDL

# Phân loại hệ CSDL

---

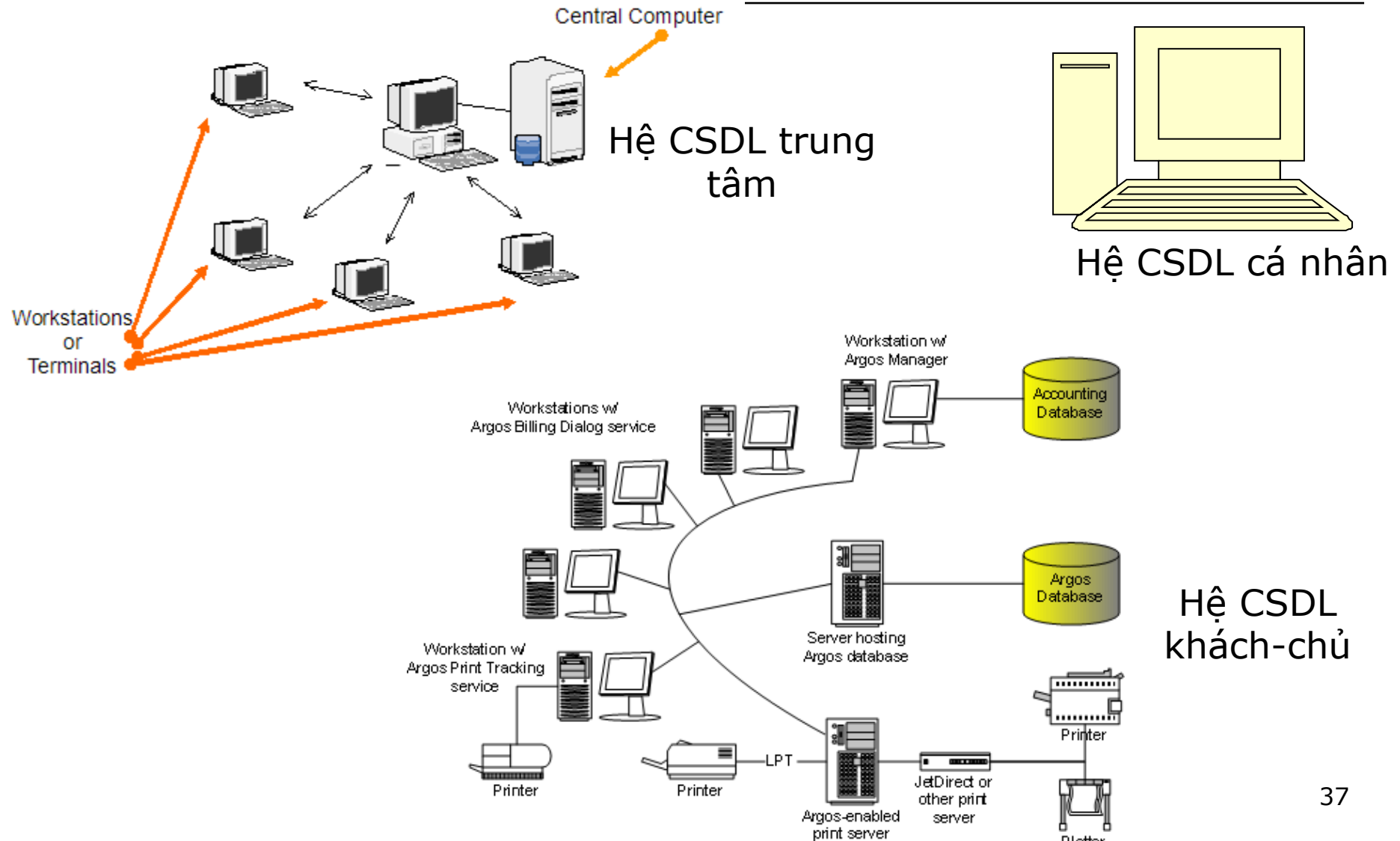
- Mô hình dữ liệu
  - Mạng vs. phân cấp vs. quan hệ vs. hướng đối tượng vs. ...
- Số người sử dụng
  - Một người dùng vs. nhiều người dùng
- Tính phân tán của CSDL
  - Tập trung vs. Phân tán
- Tính thống nhất của dữ liệu
  - Đồng nhất vs. Không đồng nhất
- ...

# Các hệ CSDL tập trung

---

- Hệ CSDL cá nhân:
  - một người sử dụng đơn lẻ
  - 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



# 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 ở nhiều vị trí địa lý khác nhau của một mạng máy tính

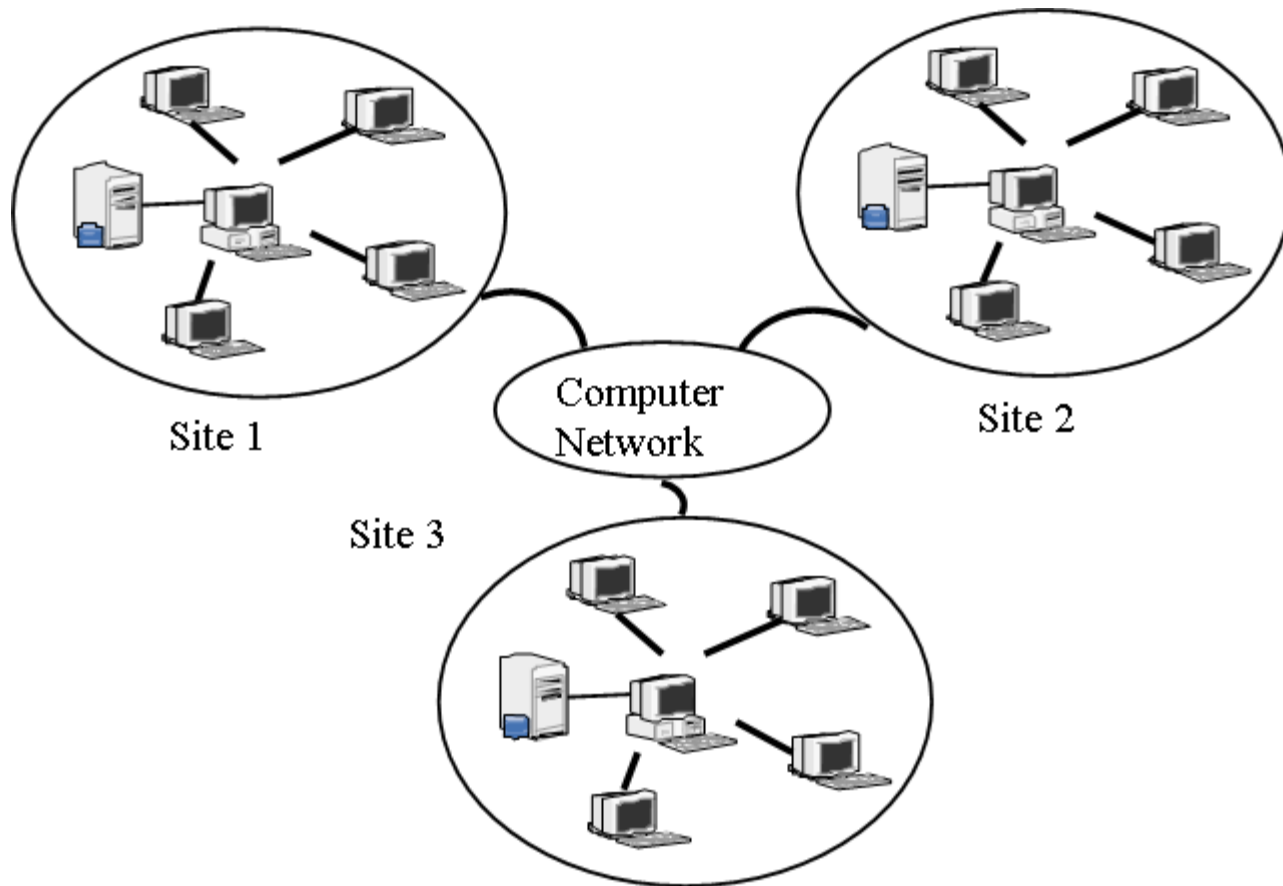
- Hệ QT CSDL phân tán:

Là một hệ thống phần mềm cho phép tạo lập CSDLPT và điều khiển các truy nhập đối với CSDLPT này.

- CSDL phân tán thuần nhất và không thuần nhất

# Các hệ CSDL phân tán

---



# Hệ CSDL PT thuần nhất

---

- Thuần nhất:
  - Công nghệ giống nhau hoặc tương thích cao ở mỗi trạm
  - DL trên các trạm cũng có thể tương thích
- Điều kiện:
  - Các hệ điều hành là như nhau hoặc tương thích các
  - Các mô hình DL ở mỗi trạm là như nhau
  - Các hệ QTCSDL ở mỗi trạm là như nhau hoặc tương thích cao
  - DL ở các trạm có các định nghĩa và khuôn dạng chung



# Hệ CSDL PT không thuần nhất

---

- Không thuần nhất:
  - Các máy tính có hệ điều hành khác nhau, mô hình DL, hệ QT CSDL khác nhau
  - DL trên các trạm cũng không tương thích:
    - Về cú pháp: biểu diễn các khoản mục DL khác nhau
    - Về ngữ nghĩa: ngữ nghĩa khác nhau với cùng 1 khoản mục trên
- ➔ Xây dựng CSDL mới để hợp nhất :
  - không dễ về mặt kỹ thuật và kinh tế
  - hạn chế các kiểu xử lý mà người dùng có thể thực hiện

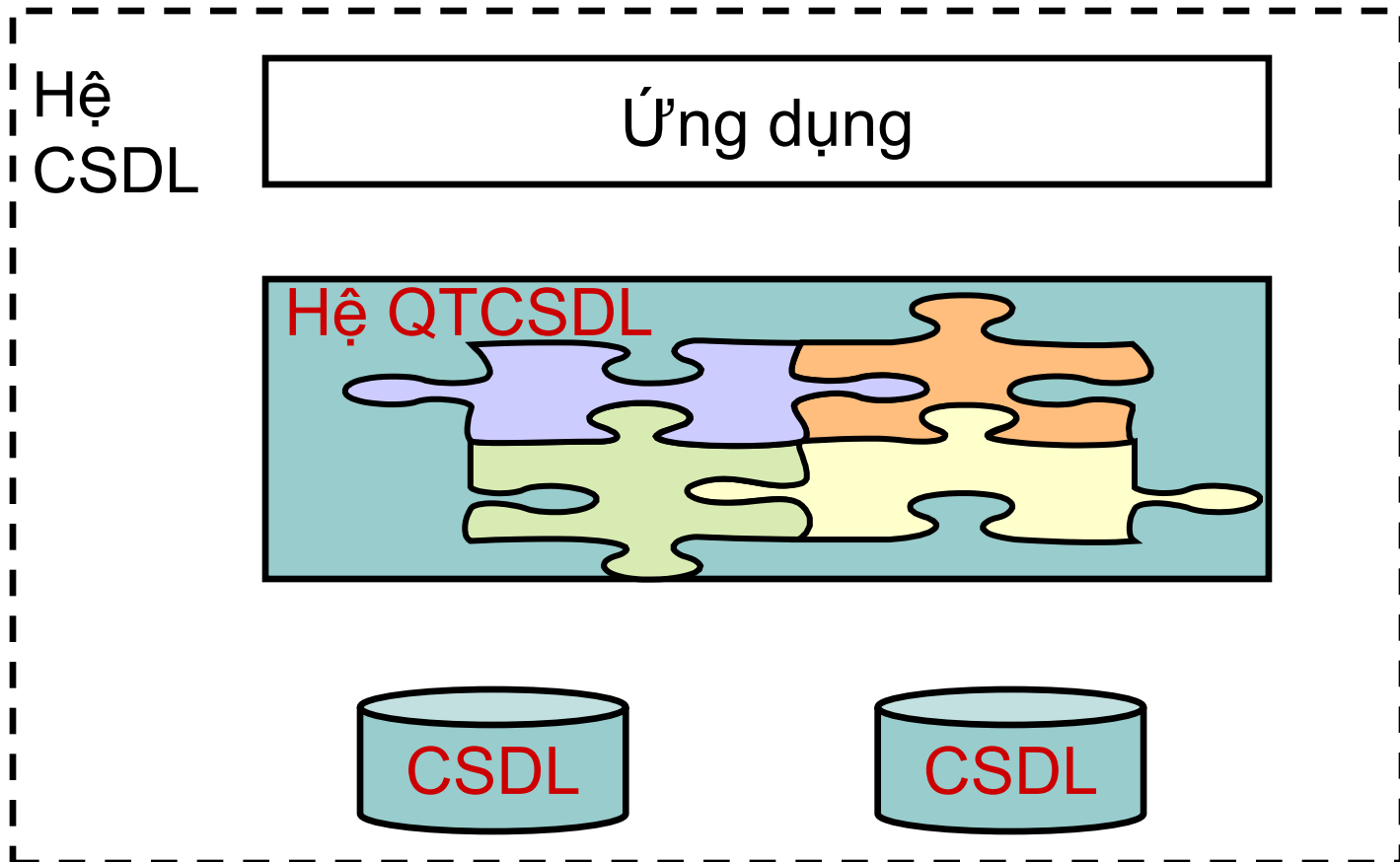
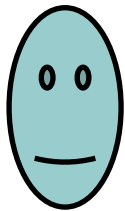
# Kết luận

---

- 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 »
- Hệ QTCSDL cung cấp các công cụ hữu hiệu trợ giúp việc tạo lập CSDL và phát triển ứng dụng

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

# Trọng tâm của môn học này





# Sử dụng kiến thức của môn học này trong tương lai

---

- Phát triển ứng dụng
- Nghiên cứu và phát triển (R&D)
- Nghiên cứu

# Các điểm cần lưu ý

---

- Cách tiếp cận tệp *vs.* cách tiếp cận CSDL
- CSDL *vs.* hệ QTCSDL *vs.* hệ CSDL
- Kiến trúc ANSI/SPARC (hay 3 mức) của hệ CSDL
- Sơ đồ *vs.* thể hiện
- Các chức năng chính của một hệ QTCSDL
- Người sử dụng trong một hệ CSDL
- Đặc điểm của hệ CSDL

