



**TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI**  
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

# GIỚI THIỆU CHUNG VỀ CÁC HỆ CƠ SỞ DỮ LIỆU

GV: Đỗ Bá Lâm

Email: lamdb@soict.hust.edu.vn

Viện Công nghệ thông tin và Truyền thông,  
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

- Tập trung vào các nhu cầu xử lý dữ liệu riêng lẻ của từng đơn vị trong tổ chức
- Mỗi ứng dụng :
  - Quản lý một tập các tệp dữ liệu của riêng nó
  - Quy định cách thức biểu diễn dữ liệu khác nhau
- Ví dụ:
  - Phòng Đào tạo: các tệp quản lý thông tin học tập
  - Kí túc xá: các tệp quản lý thông tin thuê phòng
  - Khoa, Viện đào tạo: các tệp quản lý thông tin sinh viên
  - Thư viện: các tệp quản lý thông tin bạn đọc

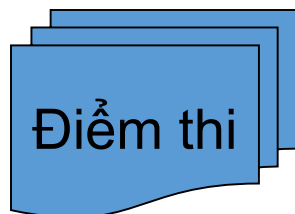
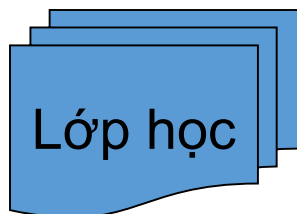
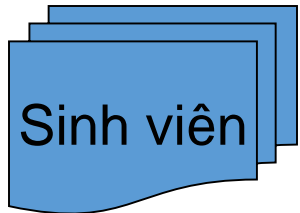
# Ví dụ Hệ thống xử lý tệp truyền thống



Sinh viên

Học tập

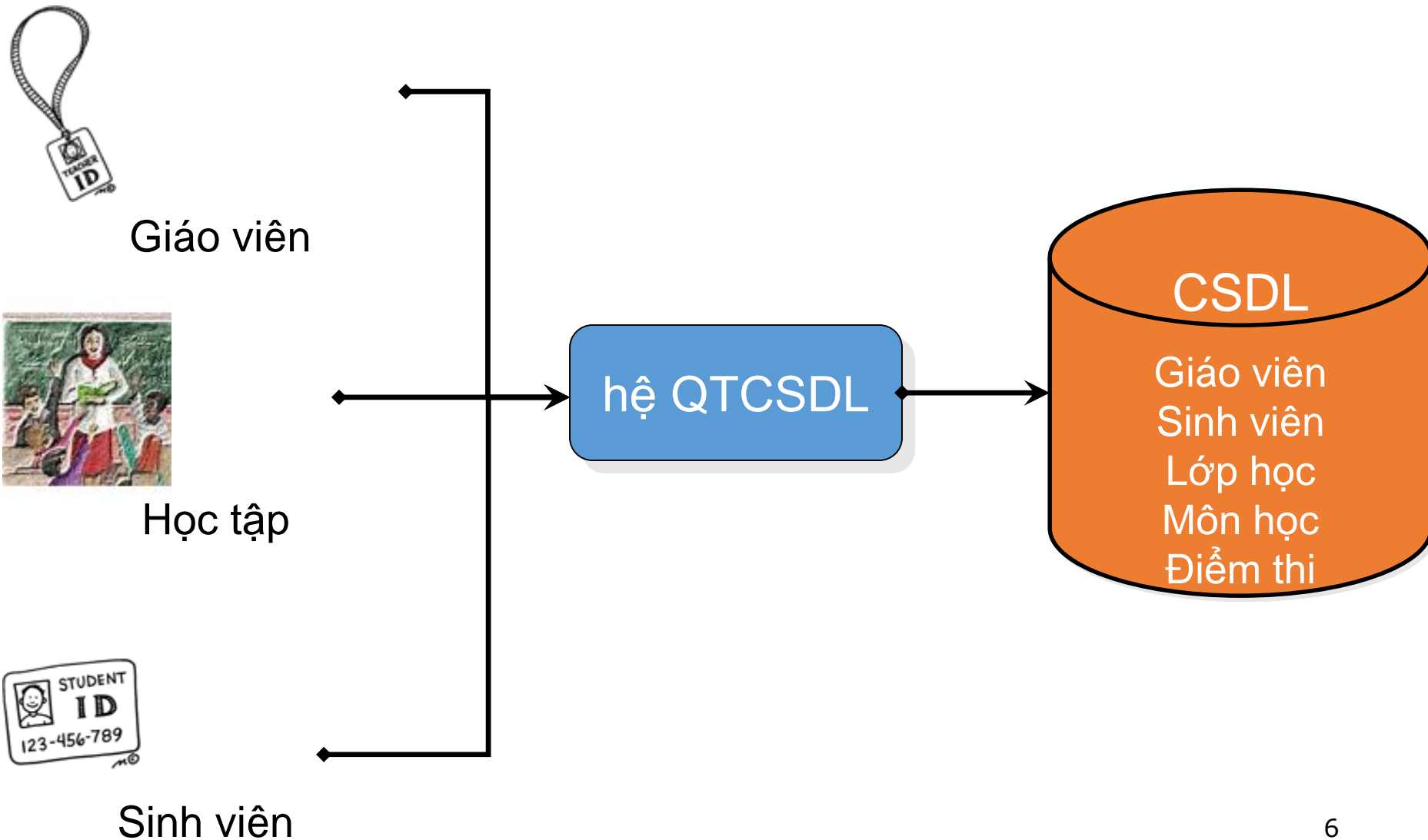
Giáo viên



# Hạn chế

- Dư thừa và không nhất quán dữ liệu
- Khó khăn trong truy cập dữ liệu với các yêu cầu không được dự tính trước
- Cô lập và hạn chế chia sẻ dữ liệu
- Các vấn đề về toàn vẹn
- Các vấn đề về độ tin cậy: sự cố khi đang thực hiện chuyển tiền
- Các dị thường khi truy nhập đồng thời: 2 yêu cầu được thực hiện đồng thời: rút tiền 50\$, 100\$
- Các vấn đề về an toàn: khó quản lý việc truy nhập dữ liệu của người sử dụng
- Sự phụ thuộc dữ liệu của các chương trình 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

- Cơ sở dữ liệu (CSDL) là gì?
- Tại sao phải sử dụng CSDL?
- Tại sao phải tìm hiểu về các hệ CSDL?

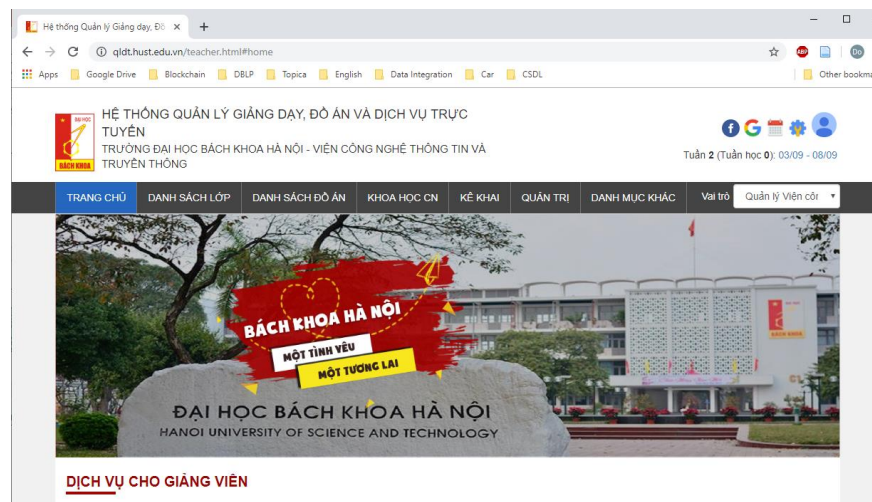
# Hệ cơ sở dữ liệu (Database System)

- Là một hệ thống gồm 4 thành phần :
  - CSDL: tối thiểu hóa dư thừa và được chia sẻ
  - Hệ quản trị CSDL
  - Người sử dụng
  - Phần cứng: các thiết bị nhớ dùng lưu trữ CSDL

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

# Ví dụ: Dữ liệu 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**



# 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 tập hợp các dữ liệu có liên quan đến nhau, được thiết kế và lưu trữ nhằm đáp ứng các nhu cầu về thông tin của một tổ chức, cá nhân nào đó.*

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 và điều khiển mọi truy cập tới các CSDL đó.*
- **Quản lý:** định nghĩa dữ liệu và lưu trữ dữ liệu lâu dài trên các thiết bị nhớ
- **Thao tác:** truy cập dữ liệu một cách hiệu quả

Ví dụ: MS. Access, MS. SQL Server, ORACLE, IBM DB2, ...



# Top 10 Hệ QTCSDL phổ biến năm 2019

Rank			DBMS
Sep 2019	Aug 2019	Sep 2018	
1.	1.	1.	Oracle +
2.	2.	2.	MySQL +
3.	3.	3.	Microsoft SQL Server +
4.	4.	4.	PostgreSQL +
5.	5.	5.	MongoDB +
6.	6.	6.	IBM Db2 +
7.	7.	7.	Elasticsearch +
8.	8.	8.	Redis +
9.	9.	9.	Microsoft Access
10.	10.	10.	Cassandra +

# Các ngôn ngữ trong Hệ QTCSDL

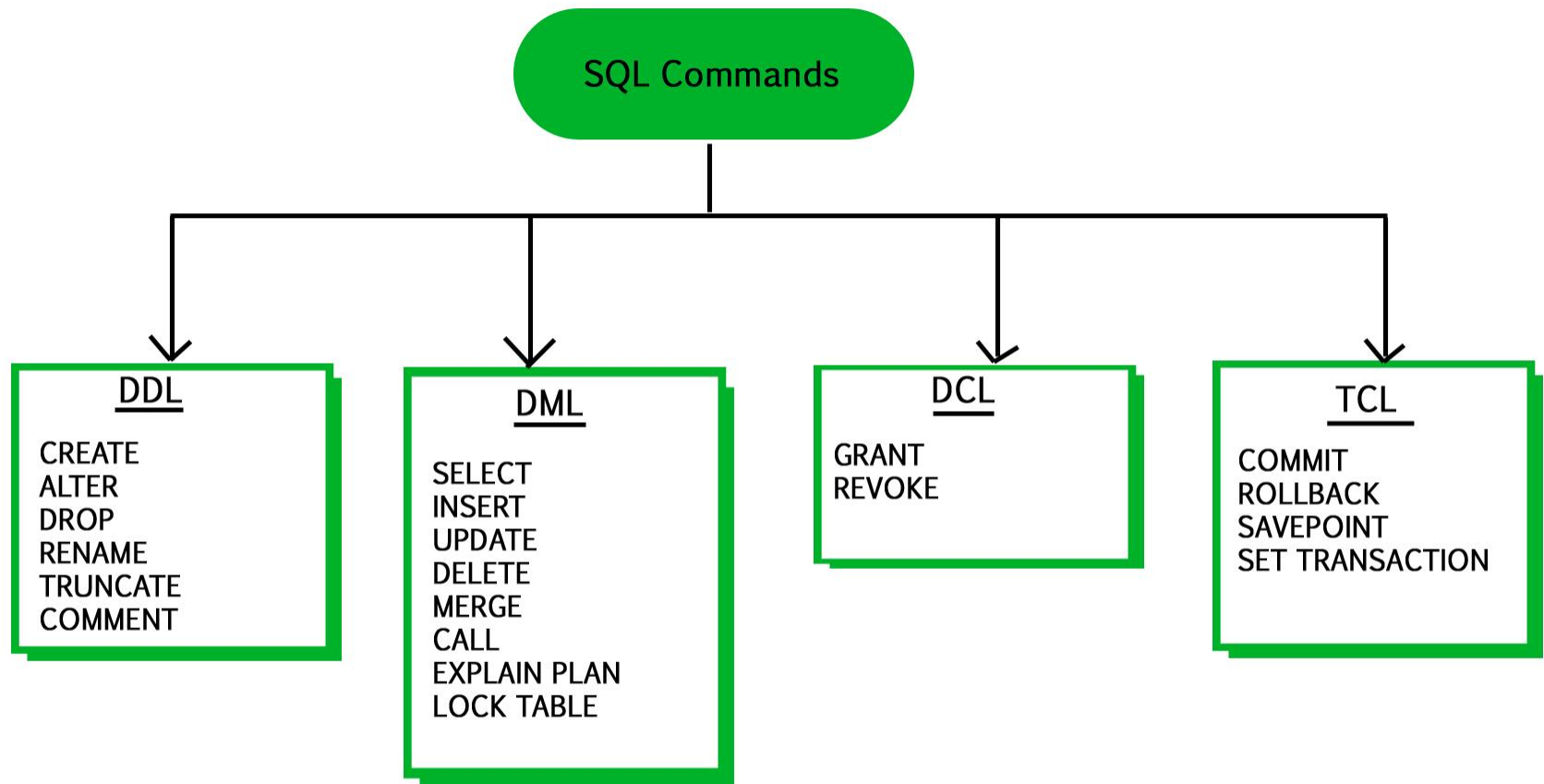
- Các Hệ QTCSDL có 2 ngôn ngữ chính
- Ngôn ngữ định nghĩa dữ liệu (Data Definition Language - DDL)
  - Cấu trúc dữ liệu
  - Mối liên hệ giữa các dữ liệu và các quy tắc, ràng buộc áp đặt lên dữ liệu
- Ngôn ngữ thao tác dữ liệu (Data Manipulation Language - DML)
  - Tìm kiếm, thêm, xoá, sửa dữ liệu trong CSDL



# Các ngôn ngữ trong Hệ QTCSDL

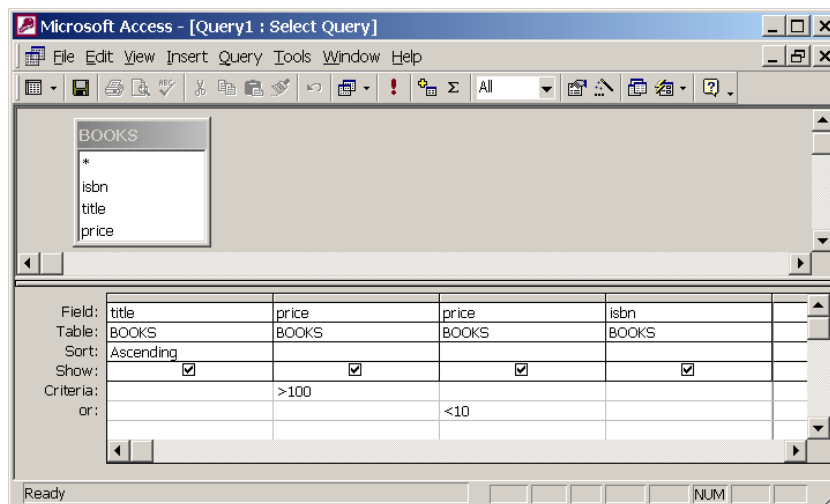
- Các ngôn ngữ khác (tùy thuộc vào từng Hệ QTCSDL)
- Ngôn ngữ điều khiển dữ liệu (**D****C****L** - *DCL*)
  - Cấp quyền người dùng để truy cập tới CSDL
  - Thu hồi quyền người dùng
- Ngôn ngữ điều khiển giao dịch (**T****C****L** - *TCL*)
  - Lưu các giao dịch trong CSDL
  - Phục hồi/quay lui lại CSDL đến thời điểm trước giao dịch

# Các ngôn ngữ trong Hệ QTCSDL

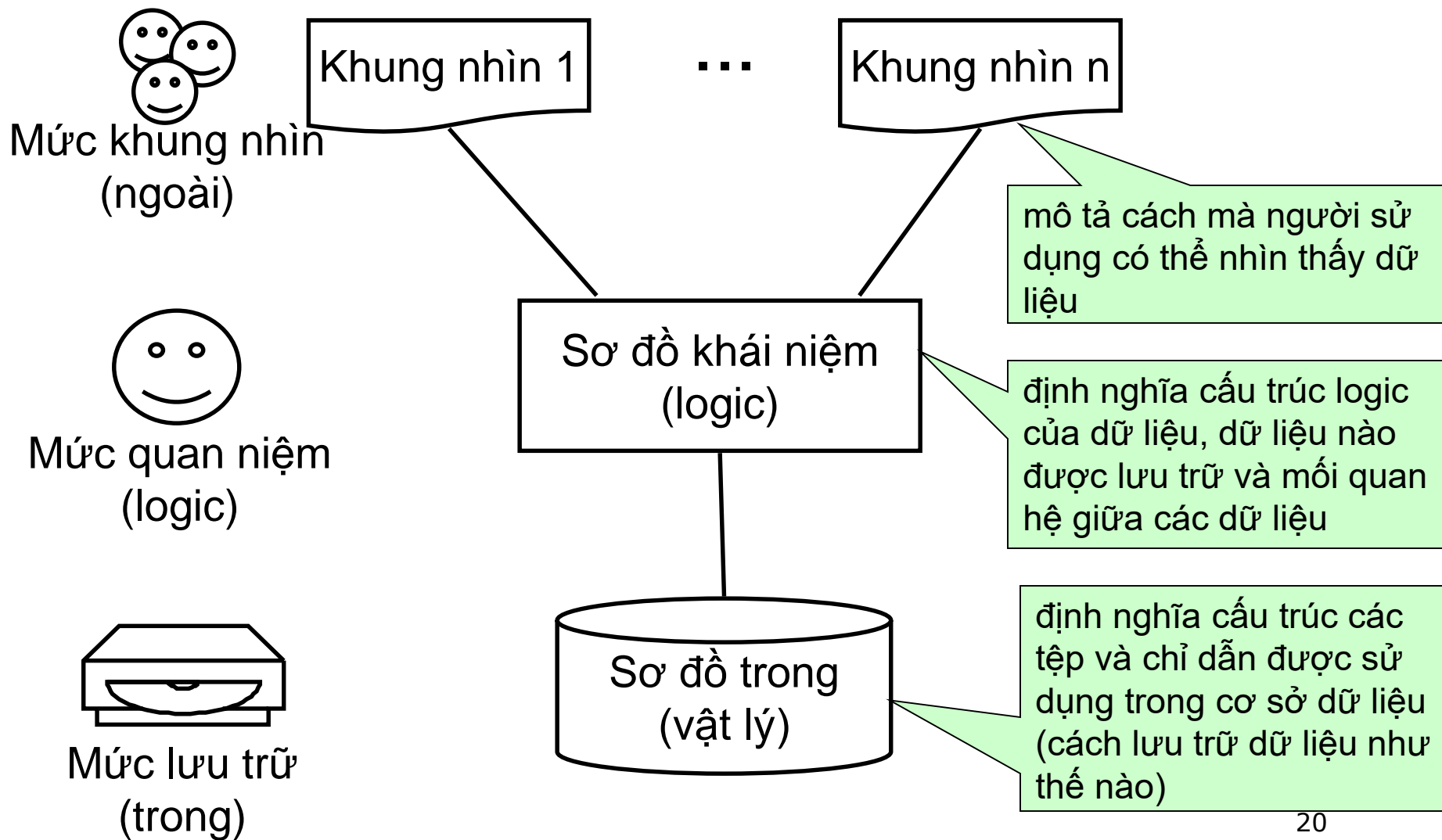


# Các ngôn ngữ

- Ngôn ngữ thao tác dữ liệu (**D**ata **M**anipulation **L**anguage - *DML*) có 2 loại:
  - Thủ tục: **WHAT + HOW**: đặc tả DL được tìm kiếm và tìm kiếm thế nào (*SQL, đại số quan hệ*)  
**SELECT Name FROM Student, Enrol**  
**WHERE ID = SID**
  - Phi thủ tục: **WHAT**: đặc tả DL cần tìm kiếm mà không cần mô tả tìm kiếm thế nào (*QBE, vị từ*)



# Sự trừu tượng hoá dữ liệu



# Ví dụ

- Mức logic

```
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  
}
```

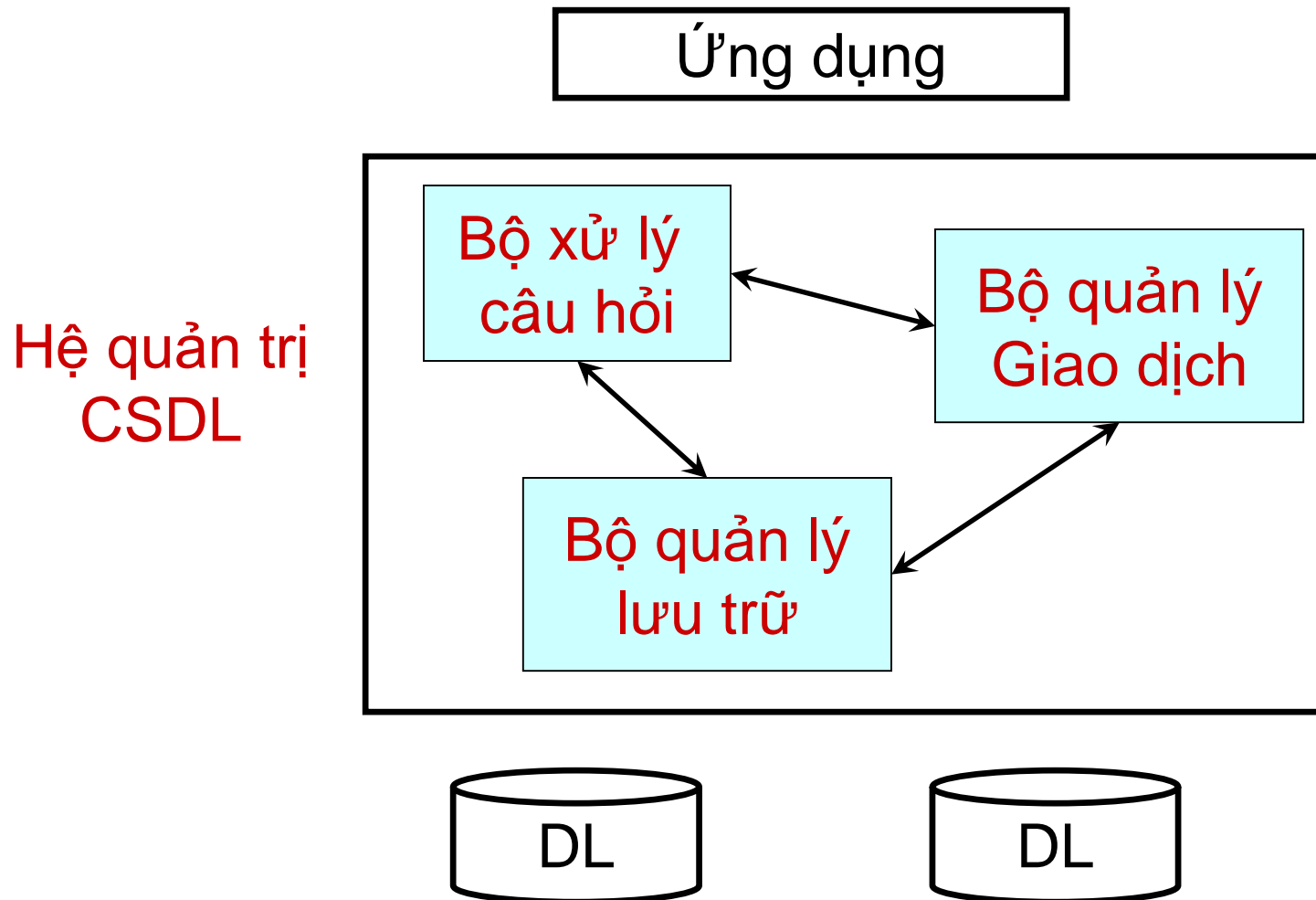
# Ví dụ (2)

- 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 »  
}
```

# Các module chính của hệ QTCSDL



# Quản lý lưu trữ

- Yêu cầu

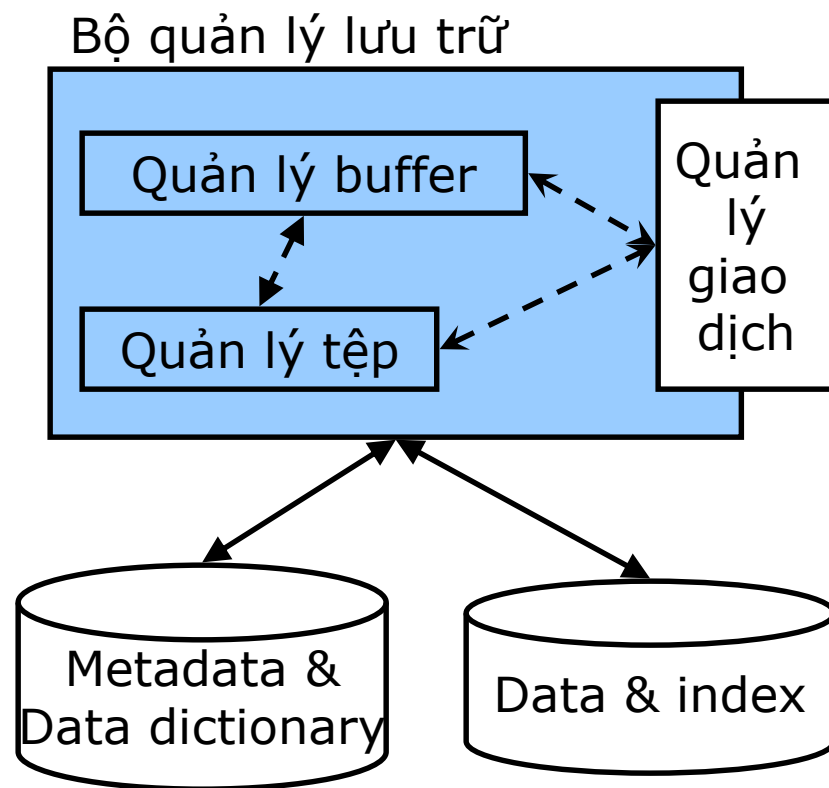
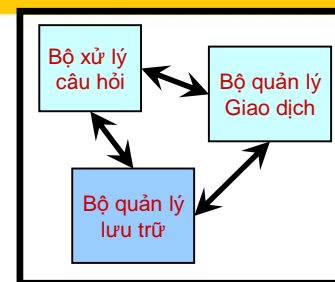
Lưu trữ và truy xuất dữ liệu trên các *thiết bị nhớ (ngoài)*

- Thực hiện

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

- Data dictionary (metadata): tập hợp các mô tả về dữ liệu

- Cấu trúc của CSDL
- Các ràng buộc về dữ liệu
- Các khung nhìn

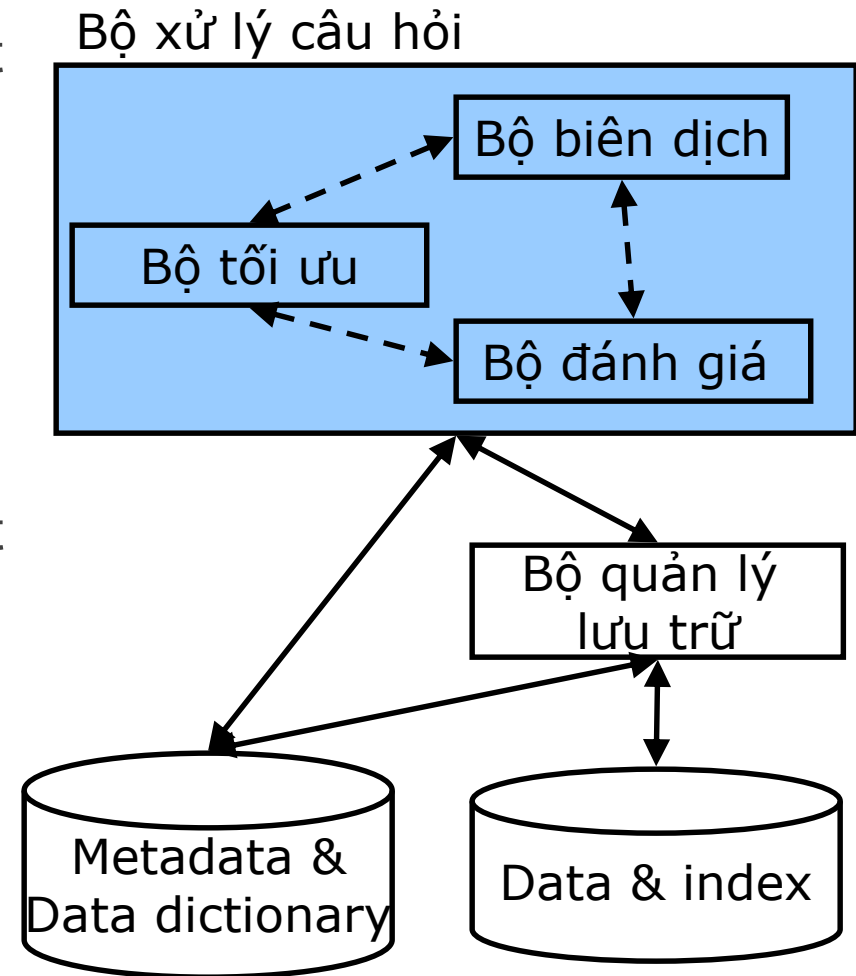




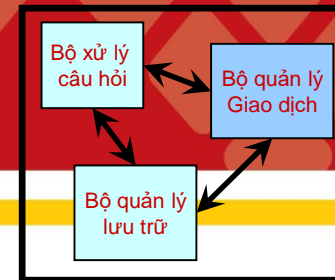
# Xử lý câu hỏi



- 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



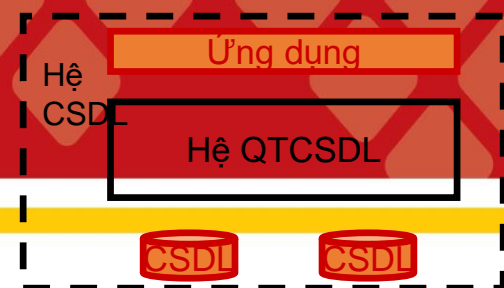
- Yêu cầu

- Định nghĩa 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: cho phép nhiều thao tác thực hiện đồng thời mà không xảy ra sự tranh chấp
- Phát hiện lỗi và phục hồi cơ sở dữ liệu

# Người dùng



- **Người lập trình ứng dụng:** chịu trách nhiệm xác định yêu cầu của người dùng cuối, xác định các chức năng 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 phân tích 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 sử dụng cuối:** là người khai thác các hệ CSDL

# Nội dung

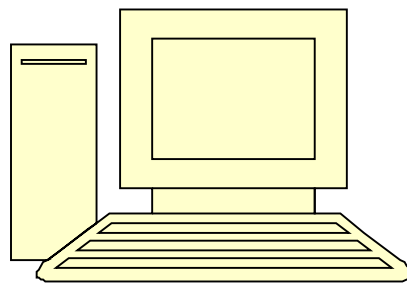
- Hệ thống xử lý tệp truyền thống
- Hệ CSDL: các khái niệm cơ bản
- ✓ Phân loại hệ CSDL

# Phân loại hệ CSDL

- Tính phân tán của CSDL
  - Tập trung vs. Phân tán
- 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 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 viết chương trình ứng dụng, thiết kế CSDL, end-user.

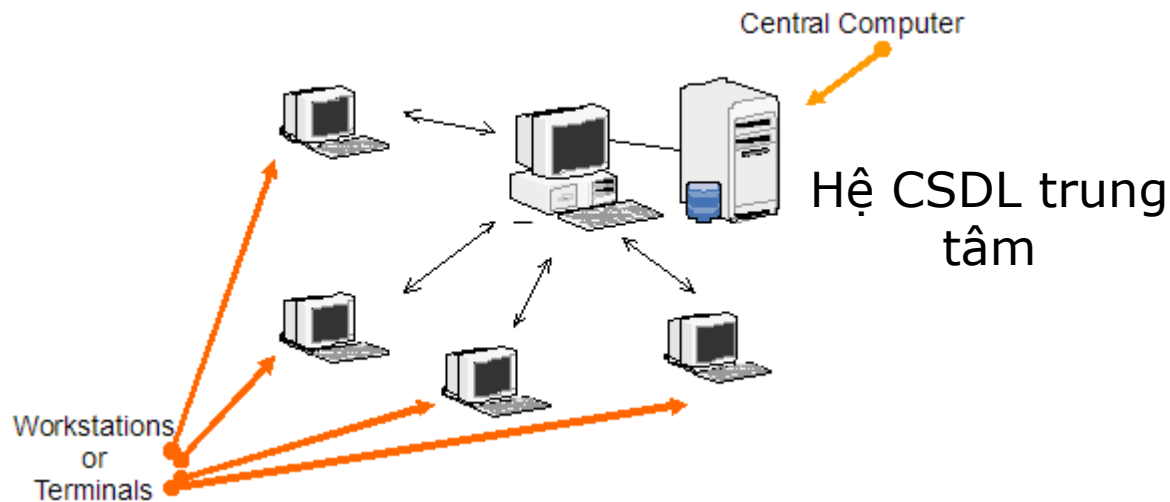


Hệ CSDL cá nhân

# Các hệ CSDL tập trung

- **Hệ CSDL trung tâm:**

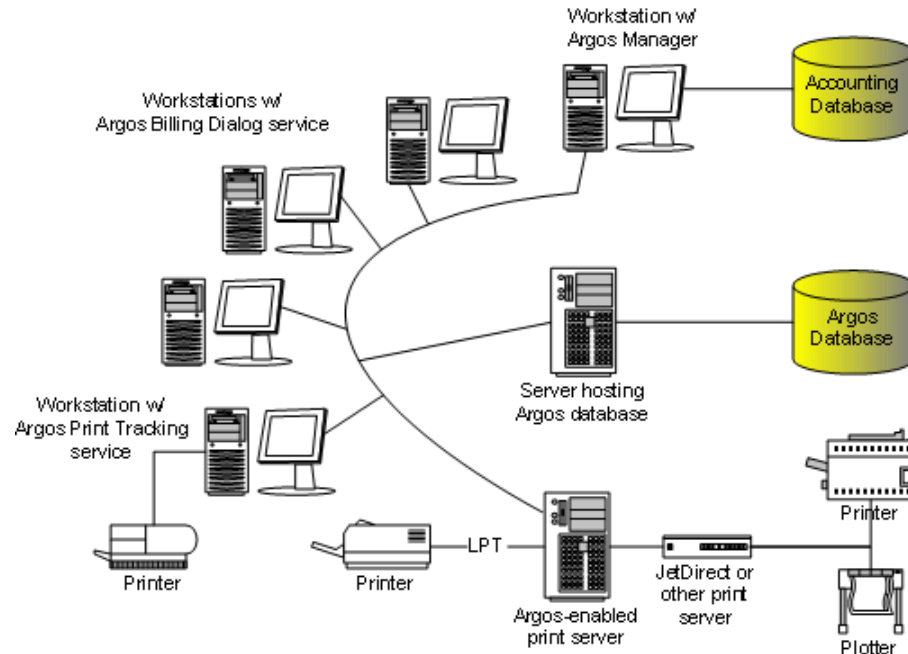
- dữ liệu được lưu trữ trên một máy tính trung tâm



# Các hệ CSDL tập trung

- **Hệ CSDL khách-chủ:**

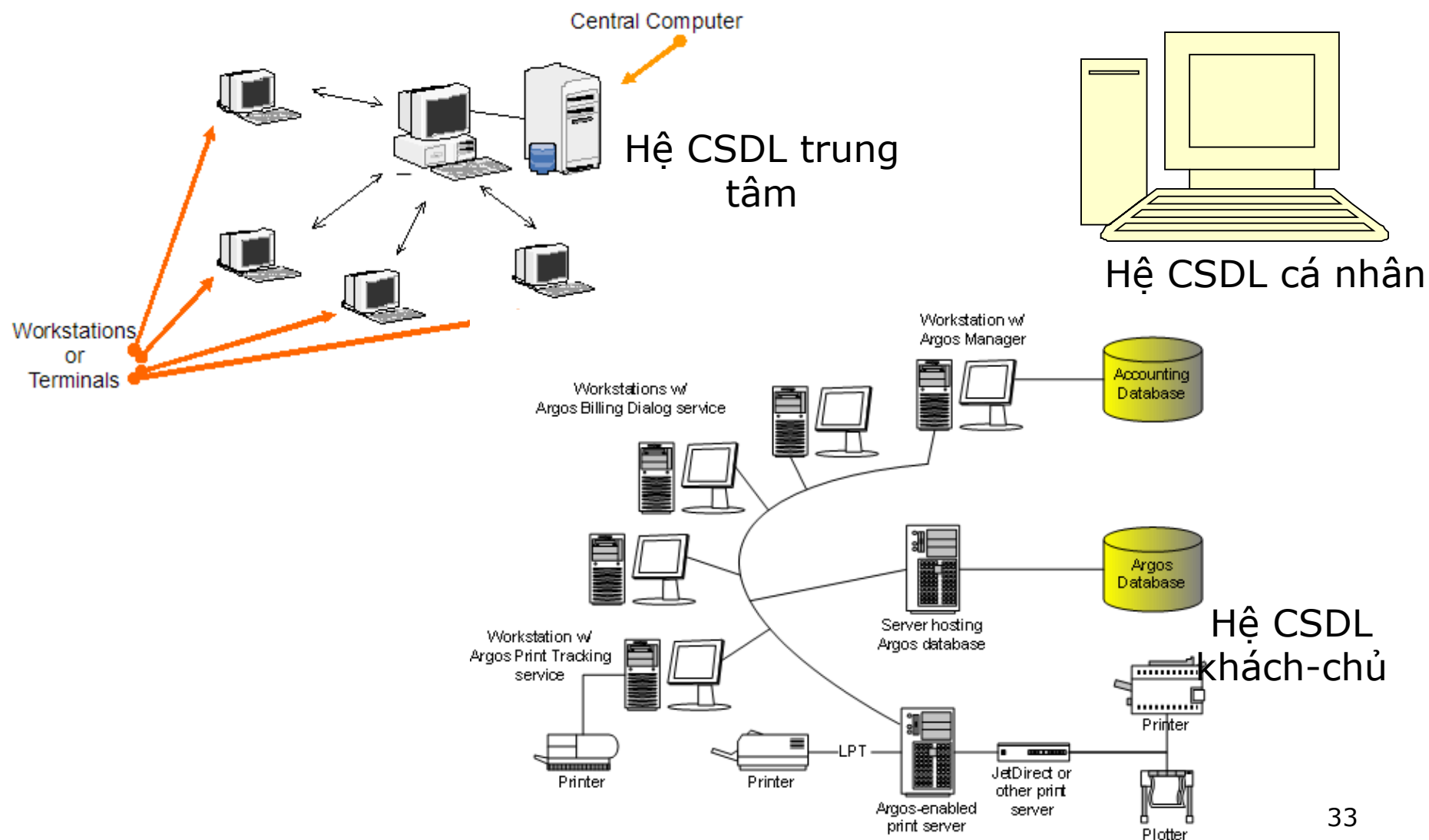
- Các máy tính trung tâm lớn → đặt so với các máy nhỏ và máy trạm.
- Chuyển yêu cầu cung cấp tài nguyên/dữ liệu từ máy tính trung tâm trong hệ CSDL trung tâm cho máy khách



Hệ CSDL  
khách-chủ



# Các hệ CSDL tập trung



# Các hệ CSDL phân tán

- **CSDL phân tán:**

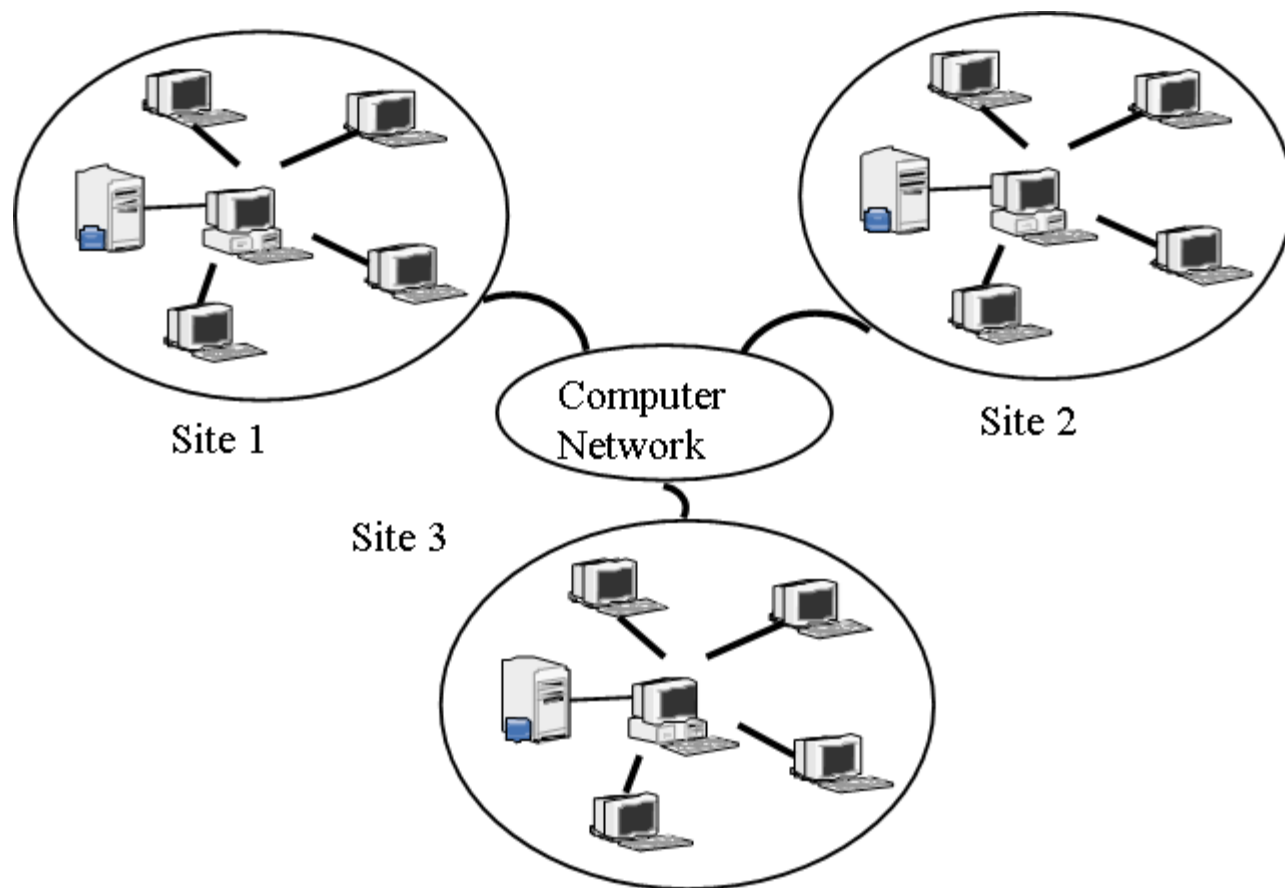
- Là một tập các CSDL có quan hệ logic với nhau nhưng được trải ra trên nhiều trạm làm việc ở nhiều vị trí địa lý khác nhau của một mạng máy tính

- **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 phân tán 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 cao
  - 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 phân tán 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

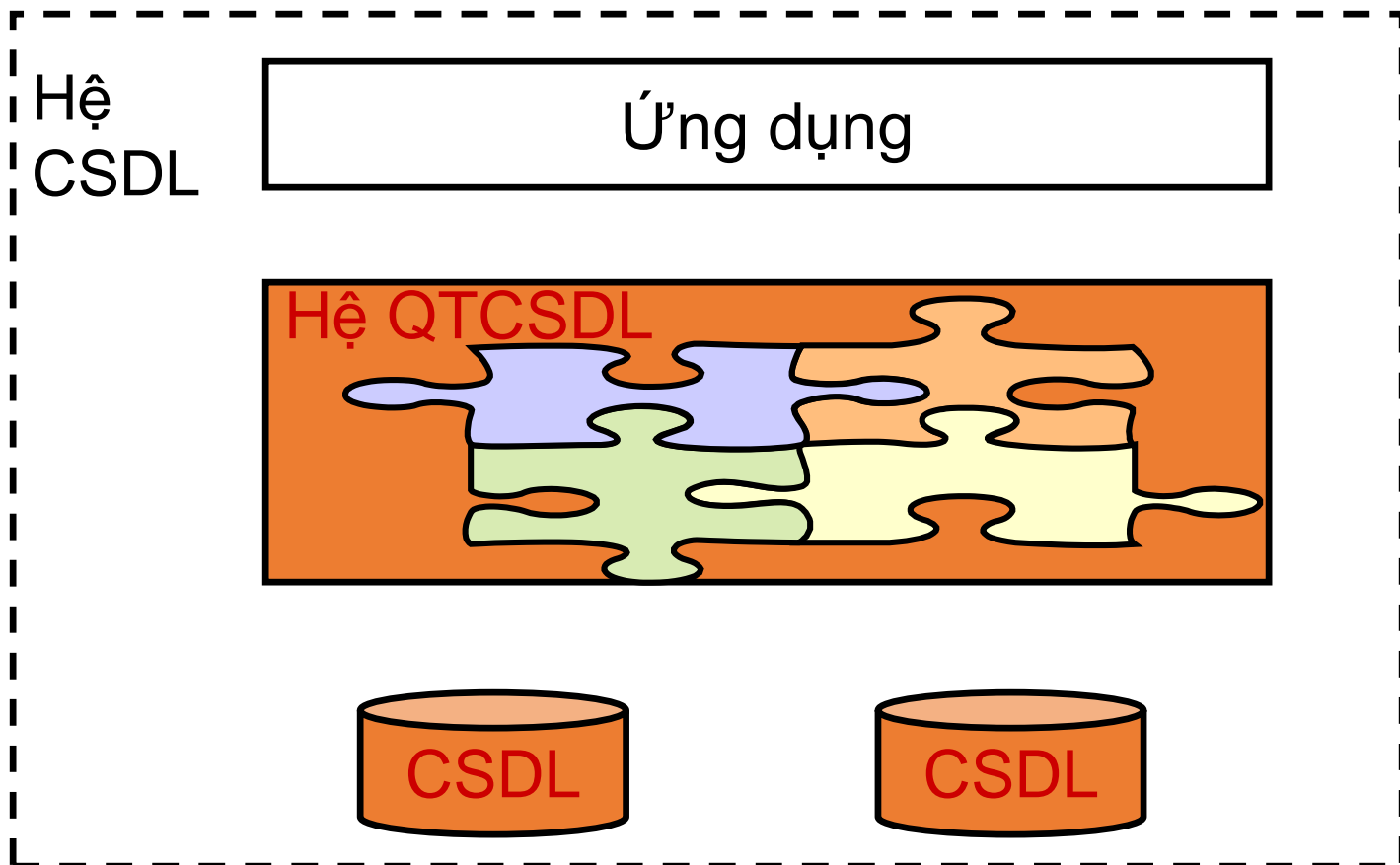
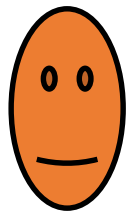
# Kết luận

- 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)
- 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"\*

\* Korth & Silberschatz. Database System Concepts.

# Trọng tâm của môn học này



# Các điểm cần lưu ý

- Cách tiếp cận tệp *vs.* cách tiếp cận CSDL
- CSDL *vs.* hệ QTCSDL *vs.* hệ CSDL
- Các chức năng chính của một hệ QTCSDL
- Người sử dụng trong một hệ CSDL
- Đặc điểm của hệ CSDL



