Đại cương về các hệ CSDL

Vũ Tuyết Trinh
trinhvt-fit@mail.hut.edu.vn

Bộ môn Hệ thống thông tin, Viện CNTT&TT
Đại học Bách Khoa Hà Nội

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?

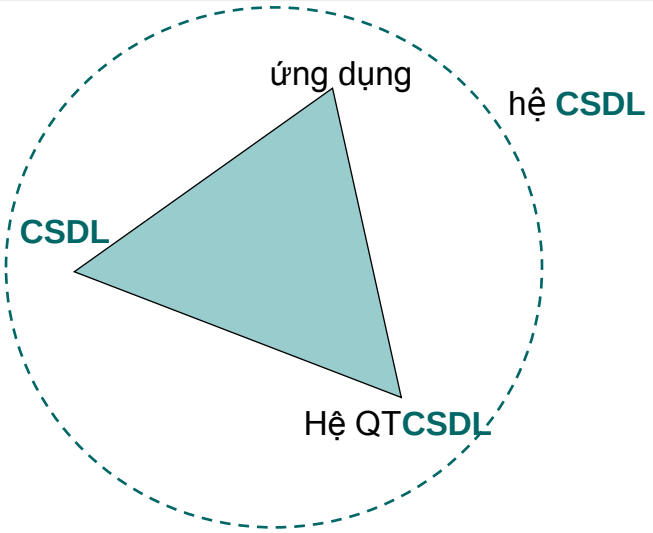
- Cần xây dựng một phần mềm cho phép khai thác một CSDL
  - Phần mềm ứng dụng

«*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

- 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

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

- Một phần mềm 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

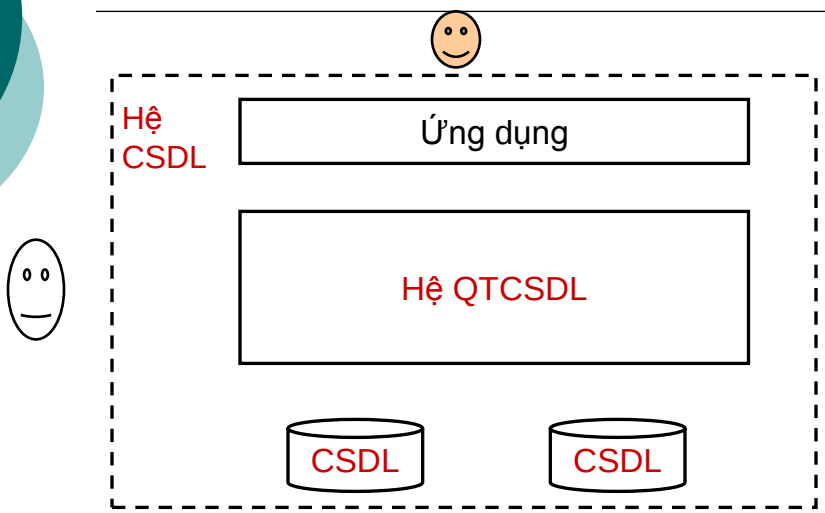
Ví dụ: MS. Access, MS. SQL Server, ORACLE, IBM DB2, ...

Hệ cơ sở dữ liệu

- Một hệ thống gồm 4 thành phần :
  - Hệ QTCSDL
  - Phần cứng
  - CSDL và phần mềm ứng dụng
  - Những người sử dụng

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

Môi trường hệ CSDL



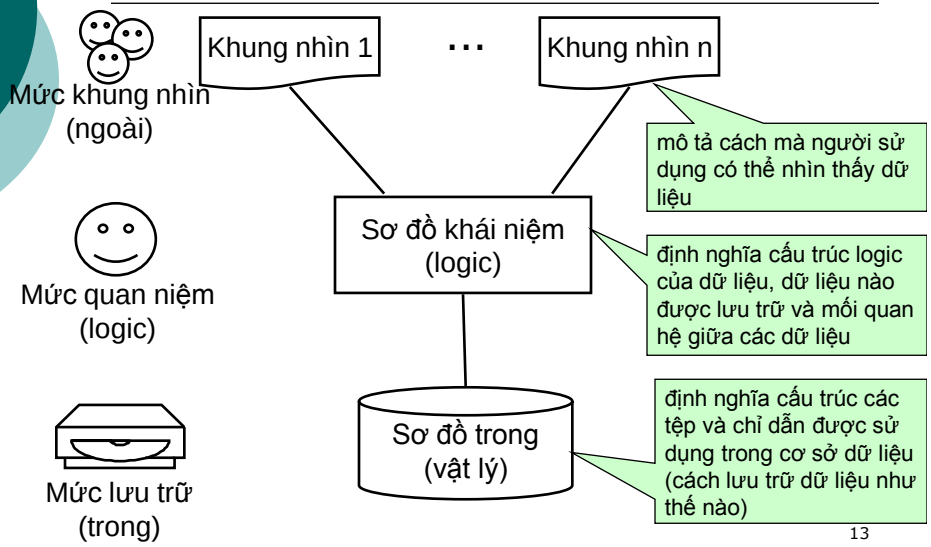
Chức năng của hệ QTCSDL

- 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 (***D**ata **D**efinition **L**anguage - 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 (***D**ata **M**anipulation **L**anguage - DML*)
  - Tìm kiếm, thêm, xoá, 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 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

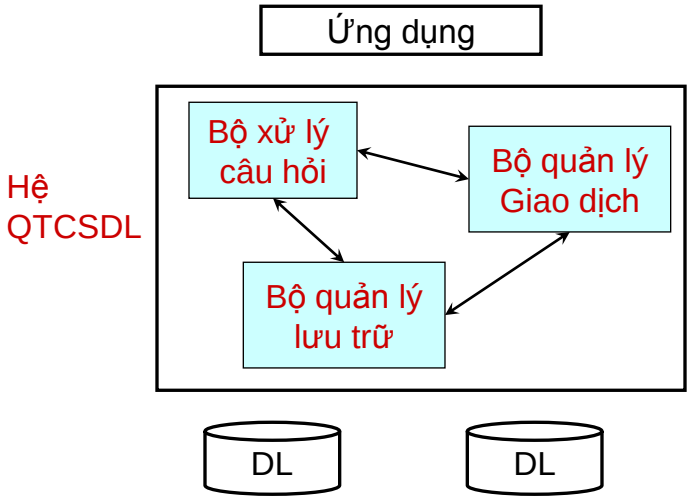
```
type ds_sinhvien = record
    ten: string;
    nam: boolean;
    ngay_sinh: date;
end;
```

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

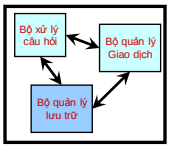

Sơ đồ và thể hiện (*schema vs. instance*)

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

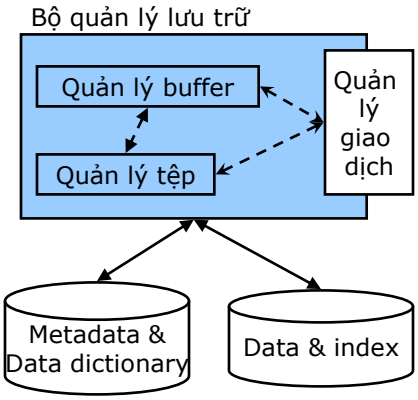


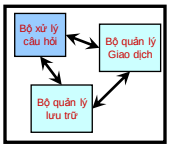
Quản lý lưu trữ



- Yêu cầu

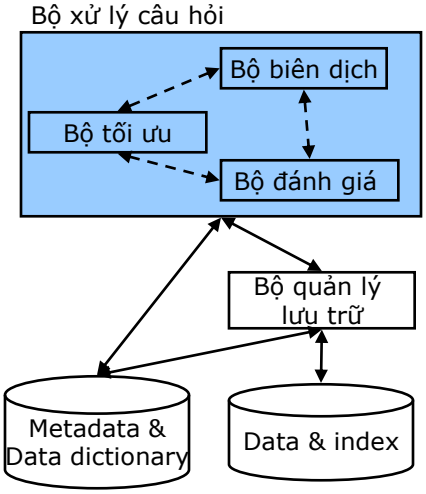
  lưu trữ và truy xuất dữ liệu trên các *thiết bị nhớ (ngoài)*
- Thực hiện
  - Tổ chức tối ưu dữ liệu trên thiết bị nhớ (*ngoài*)
  - Tương tác hiệu quả với bộ quản lý tệp

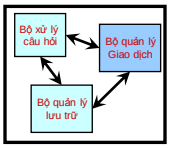




Xử lý câu hỏi

- 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

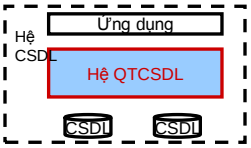




Quản lý giao dịch

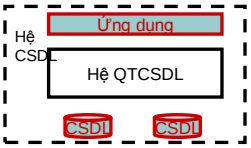
- Yêu cầu
  - Đ/n giao dịch: một tập các thao tác được xử lý như một đơn vị không chia cắt được
  - Đảm bảo tính nhất quán và tính đúng đắn của dữ liệu
- Thực hiện
  - Quản lý điều khiển tương tranh
  - Phát hiện lỗi và phục hồi cơ sở dữ liệu

Người dùng



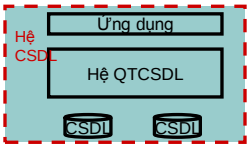
- **Người thiết kế và cài đặt hệ QTCSDL**: chịu trách nhiệm thiết kế và cài đặt các module của hệ QTCSDL và các giao diện dưới hình thức các gói phần mềm
- **Người phát triển công cụ**: chịu trách nhiệm thiết kế và cài đặt các gói phần mềm hỗ trợ cho việc thiết kê, sử dụng cũng như tăng cường hiệu năng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

Người dùng



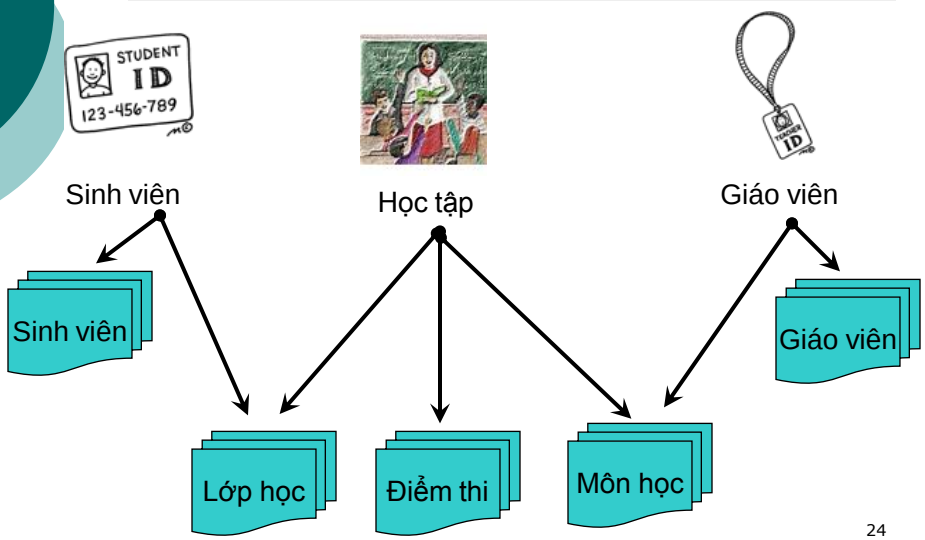
- **Người phân tích hệ thống và phát triển ứng dụng**: chịu trách nhiệm xác định yêu cầu của người dung cuối, xác định các giao dịch cần thiết để đáp ứng các yêu cầu người dùng. Người lập trình ứng dụng cài đặt những yêu cầi này trong chương trình, kiêm thử, gỡ rối, lập tài liệu cho chương trình
- **Người 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 phân 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

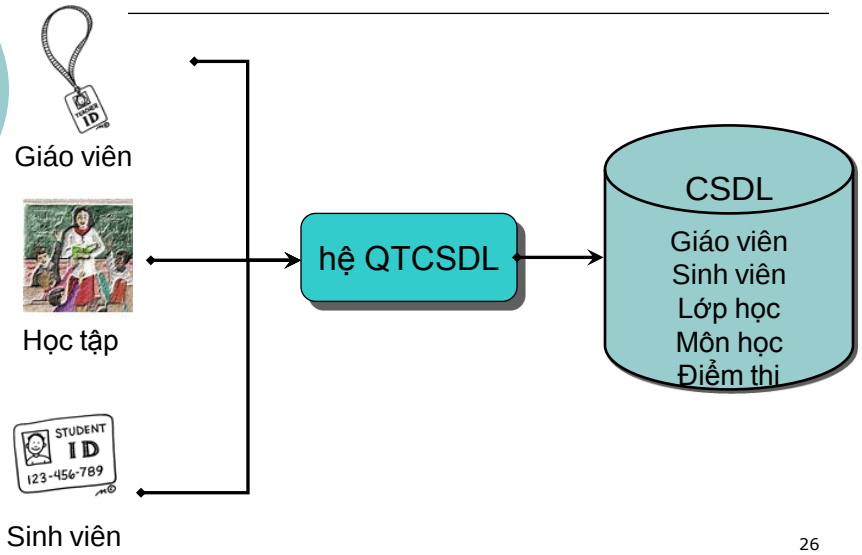
Hệ thống xử lý tệp



Hạn chế

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

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

27

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

28

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
- ...

29

Kết luận

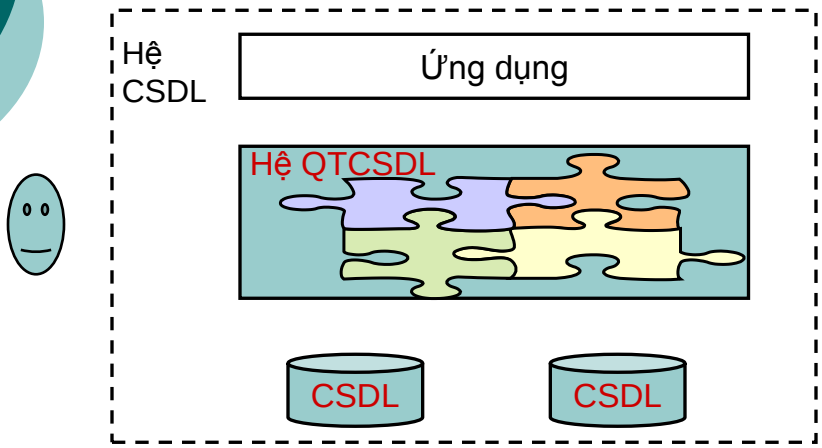
- CSDL cho phép lưu trữ và khai thác dữ liệu một cách thống nhất và hiệu quả (đặc biệt trong trường hợp khối lượng dữ liệu lớn)
- Sự trừu tượng về dữ liệu và tính độc lập dữ liệu cho phép phát triển ứng dụng « dễ dàng »
- Hệ QTCSDL cung cấp các công cụ hữu hiệu trợ giúp việc tạo lập CSDL và phát triển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

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

* Korth & Silberschatz. Database System Concepts.

30

Trọng tâm của môn học này



31

Sử dụng kiến thức của môn học này trong tương lai

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

32

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

33

34