

CƠ SỞ DỮ LIỆU

CHƯƠNG 1

ĐẠI CƯƠNG VỀ CÁC HỆ CƠ SỞ DỮ LIỆU

Nội dung

1. Tổng quan về lý thuyết CSDL
2. Kiến trúc và người dùng của hệ quản trị CSDL
3. Phân loại các hệ CSDL
4. Ngôn ngữ CSDL

Dữ liệu

- Dữ liệu - Data
 - Dữ liệu là những **số liệu rời rạc** mô tả về sự kiện, sự vật, hiện tượng được chọn lọc để **lưu trữ** với một mục đích nào đó.
 - Ví dụ:
 - “Nguyễn Văn A” là tên của sinh viên
 - “11.12.008 ” là mã số sinh viên

Dữ liệu

- Dữ liệu cần quan tâm trong quản lý đào tạo
 - Khóa học, lớp học, sinh viên, giáo viên, học phần, kết quả...
 - Dữ liệu về sinh viên: thông tin cá nhân, thông tin học tập,...
 - Dữ liệu về môn học: khối lượng, giáo viên, lịch học,...
- Cần lưu trữ dữ liệu một cách **đa dạng** và **liên kết** nhau



Cơ sở dữ liệu - Database



Cơ sở dữ liệu

- 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 tập các dữ liệu tác nghiệp được lưu trữ lại và được các hệ ứng dụng của một doanh nghiệp cụ thể nào đó sử dụng.
- Tập ngẫu nhiên của các dữ liệu không thể xem là một CSDL

Cơ sở dữ liệu

MÔN HỌC	Tên MH	Mã MH	Số TC	Khoa
	Khoa học máy tính	CS1310	4	CNTT
	Cấu trúc dữ liệu	CS3320	4	CNTT
	Toán rời rạc	MATH2410	3	TOÁN
	Cơ sở dữ liệu	CS3380	3	CNTT

HỌC PHẦN	Mã HP	Mã MH	Học Kỳ	Năm	Giáo Viên
	85	MATH2410	1	2008	Anh
	92	CS1310	1	2007	Tiền
	112	MATH2410	2	2008	Anh
	119	CS1310	2	2007	Tiền

SINH VIÊN	Tên	MSSV	Lớp	Khoa
	Trang	17	1	CNTT
	Ngọc	8	2	CNTT

KẾT QUẢ	MSSV	Mã HP	Điểm
	17	112	10
	17	119	7
	8	85	6
	8	92	9

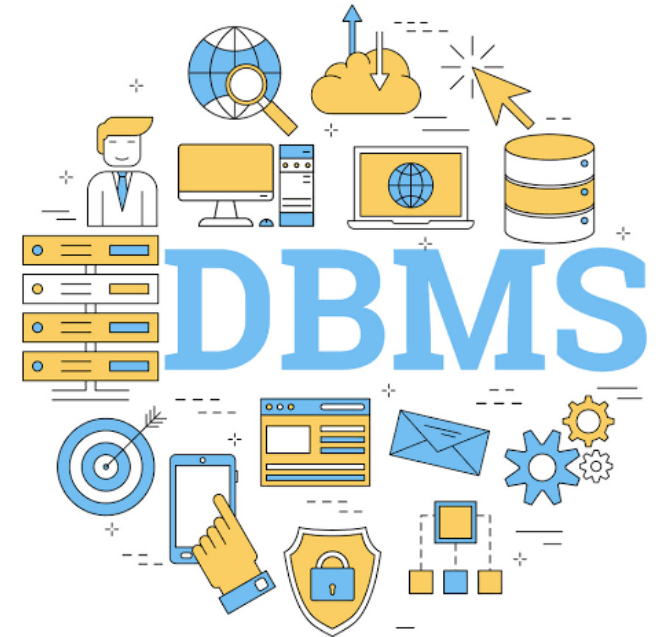
ĐIỀU KIỆN	Mã MH	Mã MH Trước
	CS3380	CS3320
	CS3380	MATH2410
	CS3320	CS1310

Xây dựng một CSDL

- Yêu cầu
 - Lưu trữ thông tin, dữ liệu 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



Hệ quản trị CSDL



Hệ quản trị CSDL

- Là một hệ thống phần mềm tạo và duy trì CSDL
 - Định nghĩa - khai báo bộ khung dữ liệu cùng với các mô tả chi tiết về dữ liệu
 - Xây dựng - lưu trữ dữ liệu lên các phương tiện lưu trữ
 - Xử lý - truy vấn, cập nhật và phát sinh báo cáo
 - Chia sẻ - cho phép nhiều người dùng và ứng dụng truy cập đồng thời CSDL
 - Bảo vệ - đảm bảo dữ liệu được lưu trữ an toàn từ các sự cố, ngăn cản truy cập không được phép...

Hệ quản trị CSDL

- Ví dụ:
 - MS SQL Server
 - Oracle
 - mySQL
 - PostgreSQL
 - SQLite
 - H2 DB



ORACLE®



SYBASE®
An SAP Company



INGRES™

Apache Derby 



HyperSQL



Informix®



 PROGRESS



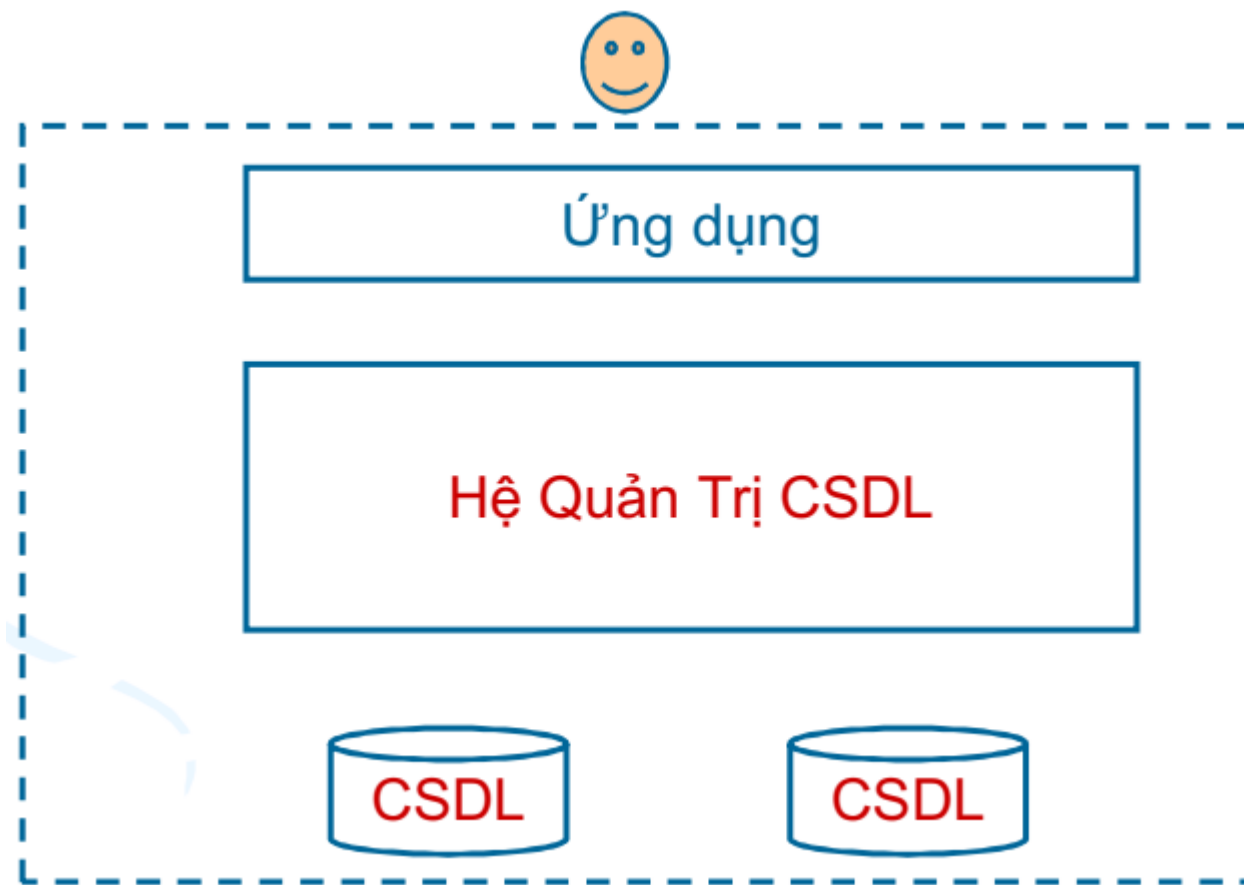

FRONTBASE™



Hệ cơ sở dữ liệu

- Là một hệ thống gồm các thành phần
 - Phần cứng
 - CSDL
 - Hệ quản trị CSDL
 - Phần mềm ứng dụng
 - Người dùng
- Ví dụ: Hệ quản lý đào tạo, hệ quản lý nhân sự,...

Hệ cơ sở dữ liệu



Hệ cơ sở dữ liệu

- Ví dụ về Hệ quản lý đào tạo
 - Định nghĩa CSDL
 - Định nghĩa cấu trúc cho: MÔN HỌC, SINH VIÊN, HỌC PHẦN, KẾT QUẢ
 - Xây dựng CSDL
 - Đưa dữ liệu vào các bảng
 - Xử lý CSDL
 - Thực hiện các truy vấn: Tìm kiếm sinh viên được điểm 10 môn Toán rồi rạc
 - Thực hiện cập nhật: Chuyển sinh viên Trang sang lớp số 2

Một số đặc tính của CSDL

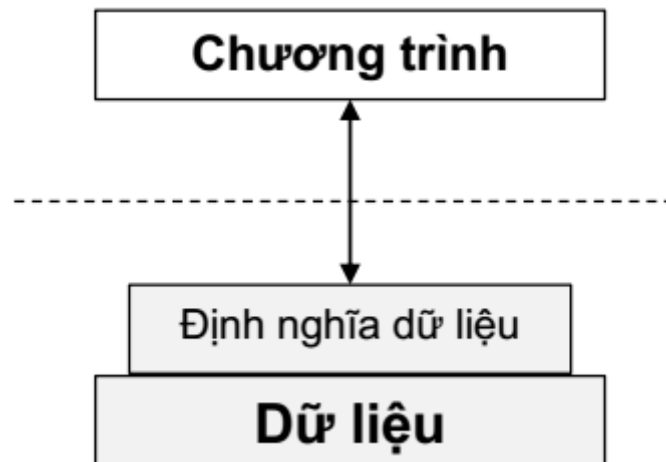
- Tính tự mô tả
- Tính độc lập giữa ứng dụng và dữ liệu
- Tính trừu tượng hóa dữ liệu
- Hỗ trợ nhiều khung nhìn dữ liệu

Tính tự mô tả

- Hệ CSDL không chỉ chứa bản thân CSDL mà còn chứa thông tin định nghĩa của CSDL
- Các định nghĩa được lưu trữ gọi là metadata
 - Chứa các thông tin về cấu trúc tập tin, kiểu và dạng thức lưu trữ của mỗi thành phần dữ liệu và những ràng buộc dữ liệu
- Các ứng dụng có thể truy xuất đến nhiều CSDL nhờ thông tin metadata của CSDL đó.

Tính độc lập giữa ứng dụng và dữ liệu

- CSDL và metadata được quản lý chung
 - Quản lý CSDL độc lập tương đối với chương trình ứng dụng

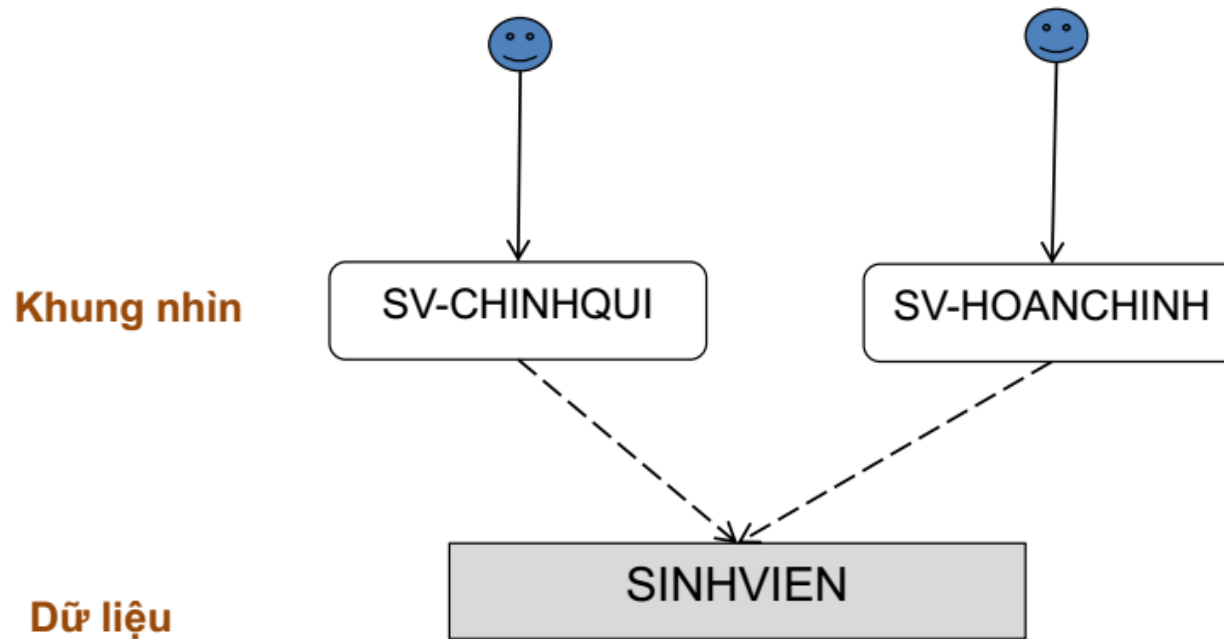


Trừu tượng hóa dữ liệu

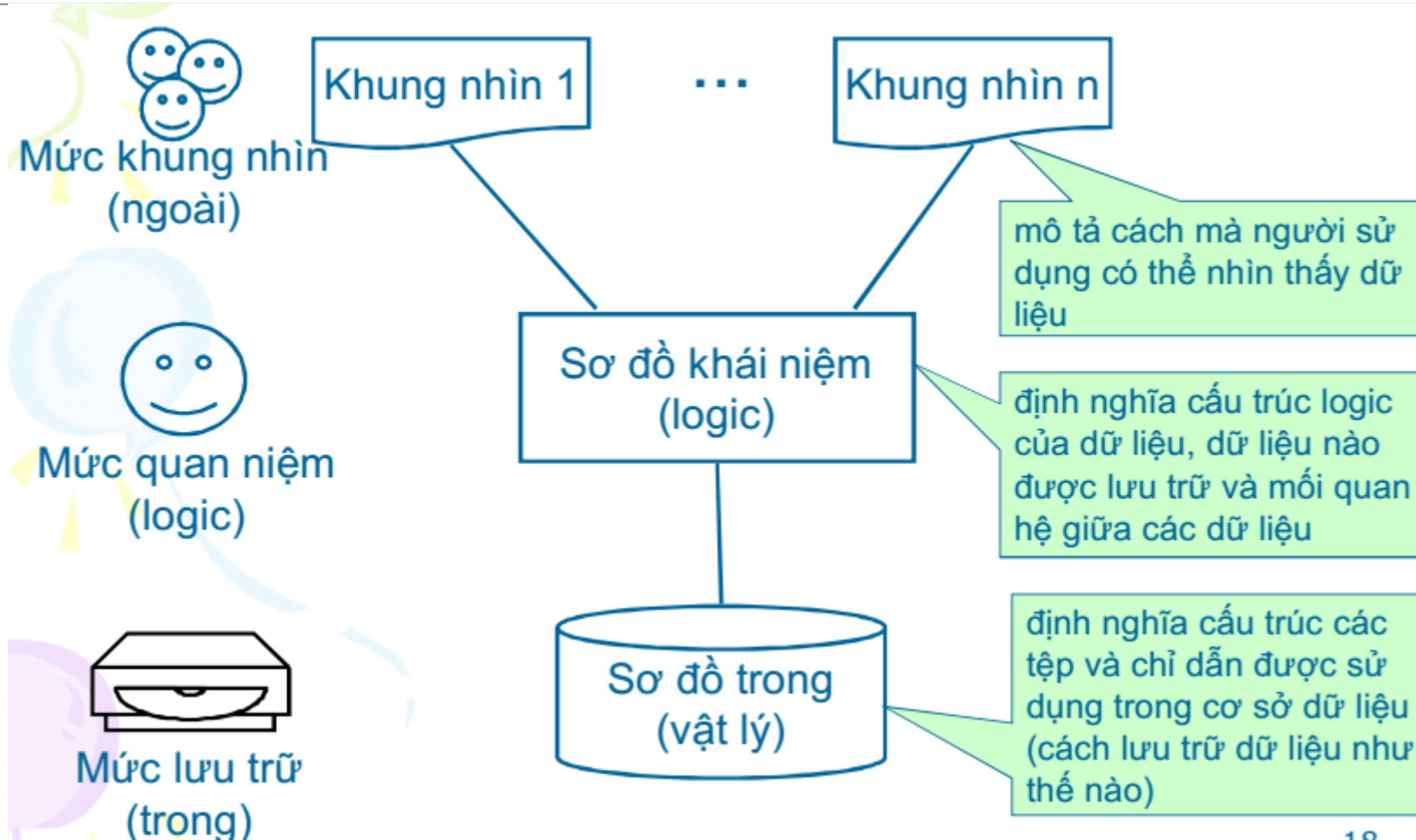
- Hệ CSDL cho phép trình bày dữ liệu ở một mức trừu tượng
 - Che bớt những chi tiết lưu trữ và cài đặt của dữ liệu
 - Người dùng và ứng dụng truy xuất đến các thành phần trừu tượng thay vì các chi tiết vật lý.
- Trừu tượng hóa dữ liệu thông qua mô hình dữ liệu
 - Đối tượng, Thuộc tính, Liên kết

Hỗ trợ nhiều khung nhìn dữ liệu

- Hỗ trợ định nghĩa nhiều khung nhìn (view) khác nhau về cùng một dữ liệu



Kiến trúc ba mức của hệ CSDL



Kiến trúc ba mức của hệ CSDL

- Mức lưu trữ (mức trong)
 - Mô tả cấu trúc lưu trữ vật lý CSDL
- Mức quan niệm (mức logic)
 - Mô tả cấu trúc của toàn thể CSDL cho 1 cộng đồng người sử dụng, gồm thực thể, kiểu dữ liệu, mối liên hệ và ràng buộc
 - Che bớt các chi tiết của cấu trúc lưu trữ vật lý
- Mức khung nhìn (mức ngoài)
 - Mô tả một phần của CSDL mà 1 nhóm người dùng quan tâm đến và che dấu phần còn lại của CSDL đối với nhóm người dùng đó

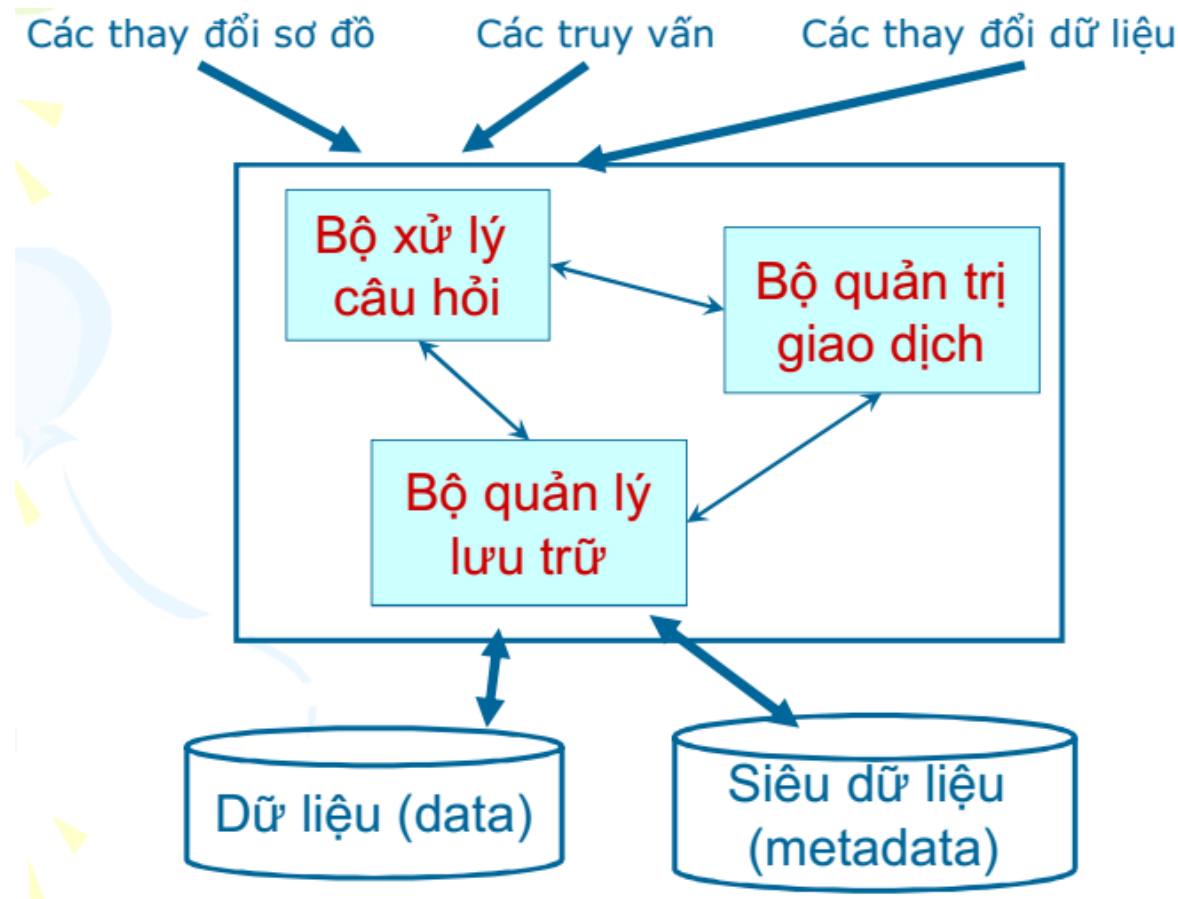
Các tính năng của hệ quản trị CSDL

- 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

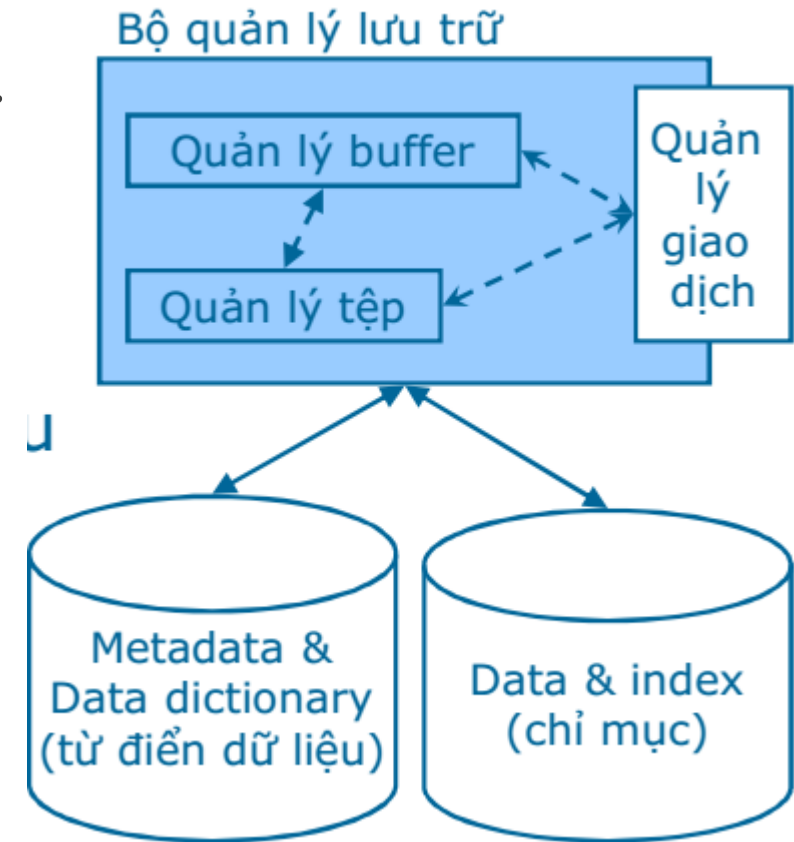
- Hỗ trợ ít nhất một mô hình dữ liệu
- Đảm bảo tính độc lập dữ liệu
- Hỗ trợ các ngôn ngữ cấp cao nhất định cho phép người sử dụng định nghĩa cấu trúc của dữ liệu, truy nhập và thao tác dữ liệu
- Điều khiển truy nhập
- Phục hồi dữ liệu

Kiến trúc của hệ quản trị CSDL



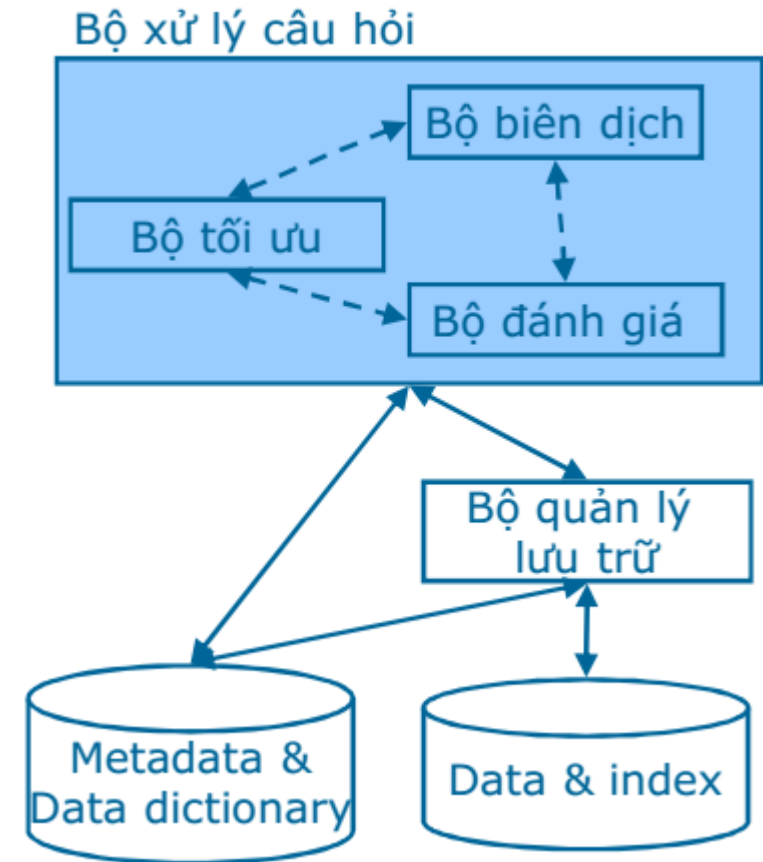
Quản lý lưu trữ

- 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 hệ quản trị CSDL

- **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 hệ quản trị CSDL

- **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 hệ quản trị CSDL

- **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

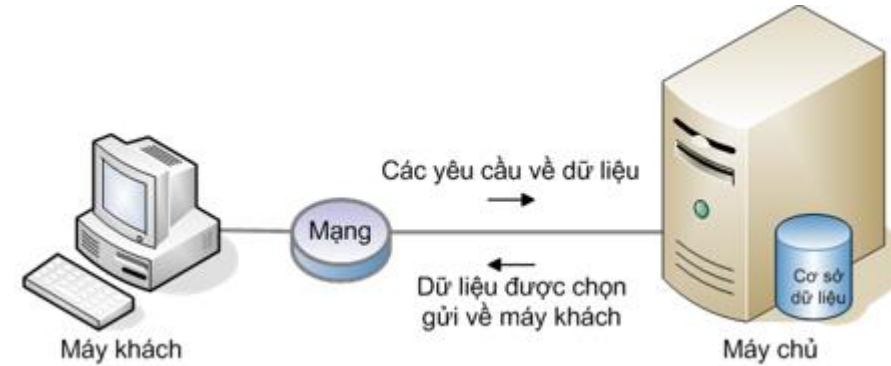
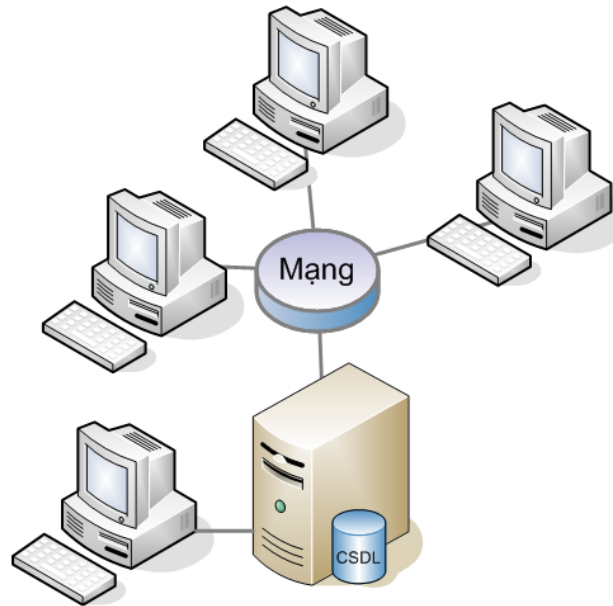
Phân loại các hệ CSDL

- 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 máy tính trung tâm.
- Hệ CSDL khách-chủ:
 - Các máy tính trung tâm lớn
 - Các ứng dụng máy khách truy nhập dữ liệu được quản lý bởi máy chủ.

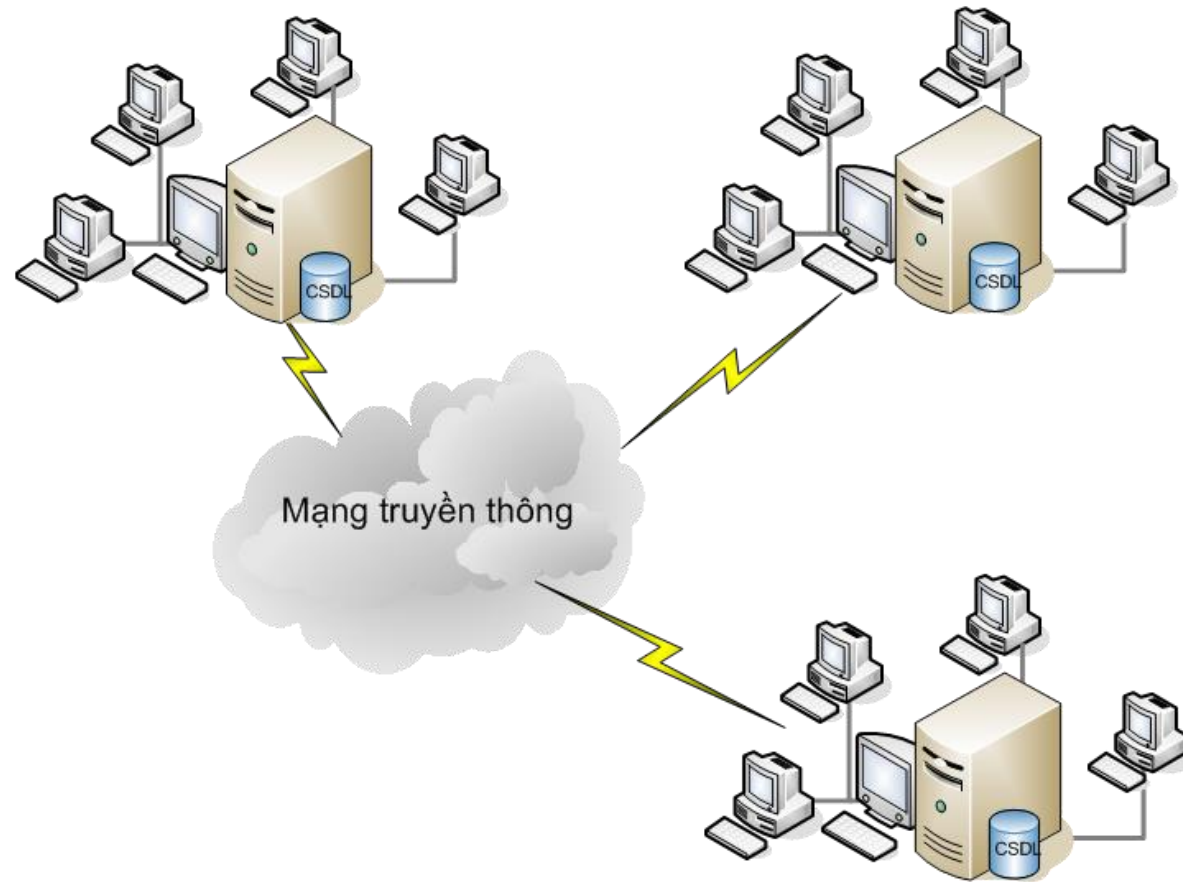
Các hệ CSDL tập trung



Các hệ CSDL phân tán

- CSDL phân tán cho phép người dùng truy cập dữ liệu tại nhiều điểm khác nhau.
- Trong hệ CSDL phân tán, dữ liệu được sử dụng nhiều nhất sẽ lưu tại điểm sử dụng
- CSDL phân tán là một tập hợp dữ liệu có liên quan về mặt logic được dùng chung và phân tán về mặt vật lý trên một mạng máy tính.

Các hệ CSDL phân tán



Ngôn ngữ CSDL

- Ngôn ngữ định nghĩa dữ liệu (Data Definition Language - DDL)
 - Cấu trúc dữ liệu
 - Mối liên hệ giữa các dữ liệu và 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

Tổng kết

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