



TRƯỜNG ĐẠI HỌC CẦN THƠ
TRƯỜNG CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG
KHOA CÔNG NGHỆ THÔNG TIN

QUẢN TRỊ DỮ LIỆU - CT467

**Chương 1: GIỚI THIỆU
HỆ QUẢN TRỊ CSDL**

Biên soạn:



Ths. Nguyễn Thị Kim Yến



Ntkyen@ctu.edu.vn



MỤC TIÊU CỦA CHƯƠNG 1

- Hiểu được khái niệm về hệ QT CSDL và các khái niệm cơ bản trong môi trường CSDL
- Hiểu được các ưu, nhược điểm của việc quản lý dữ liệu bằng hệ QT CSDL
- Biết được các kiến trúc hệ CSDL, hệ QT CSDL



NỘI DUNG

1

Giới thiệu hệ QT CSDL

2

Ưu, nhược điểm của hệ QT CSDL

3

Góc nhìn dữ liệu (Kiến trúc của CSDL)

4

Mô hình hệ cơ sở dữ liệu

5

Kiến trúc của một hệ QT CSDL

6

Các loại kiến trúc của hệ CSDL

7

Quản trị giao dịch



1. Giới thiệu hệ QT CSDL

What is Data?

Dữ liệu: có thể ký tự, văn bản, hình ảnh, âm thanh,... chưa được tổ chức và xử lý.



Thông tin: dữ liệu đã được xử lý

1. Giới thiệu hệ QT CSDL (tt)

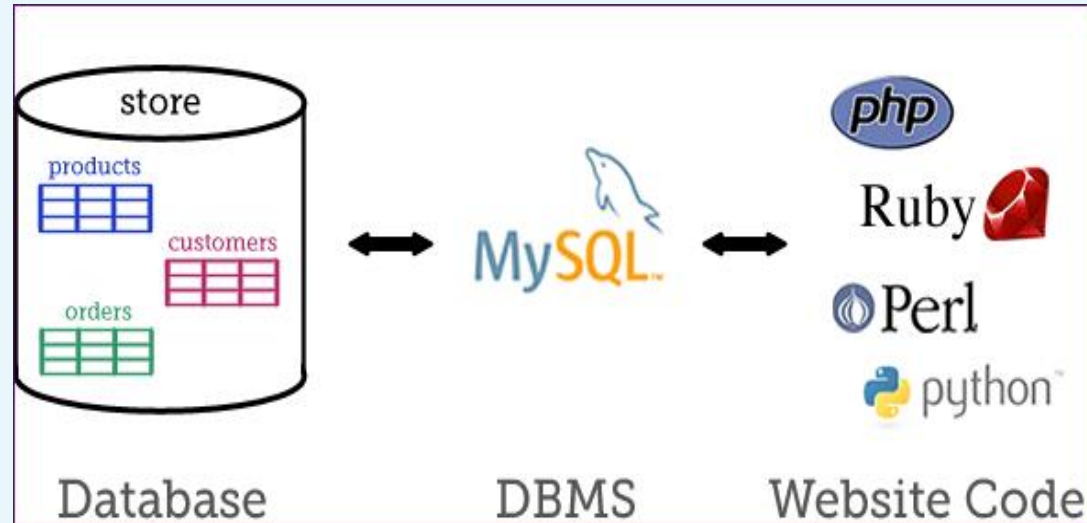
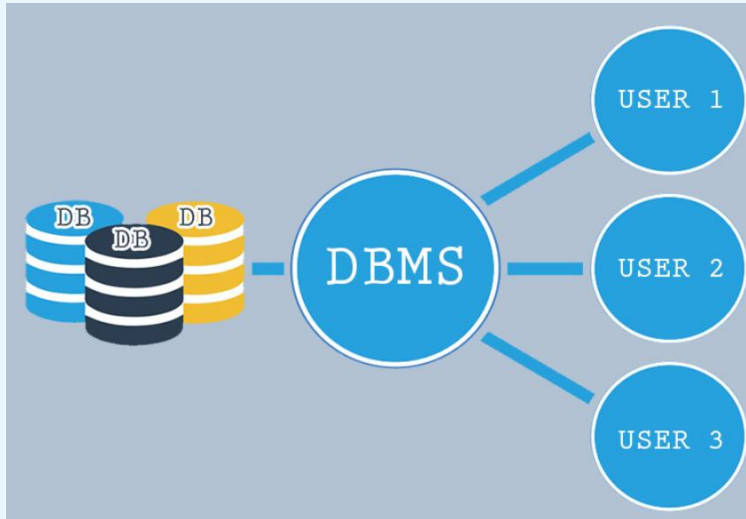


- Tập hợp các dữ liệu có cấu trúc, có tổ chức liên quan với nhau và được lưu trữ trong máy tính.
- Cho phép người dùng dễ dàng được truy cập và thay đổi bằng cách sử dụng các **hệ QT CSDL**.



1. Giới thiệu hệ QT CSDL (tt)

- Hệ QT CSDL (Database Management System - DBMS)



- **DBMS:** 1 bộ phần mềm cho phép người sử dụng định nghĩa, tạo lập, bảo trì và điều khiển truy xuất CSDL.



Quá trình phát triển các HTQCSDL

- **Đầu 1960:** mô hình phân cấp & mạng (**IMS** của IBM)
- **1976:** mô hình thực thể - liên kết (**System-R** của IBM)
- **1980:** các hệ quản trị CSDL quan hệ khác như **Access**, **Oracle**, **SQL-Server**,...
- **1990s – đến nay:** HQTCSDL hướng đối tượng (**Orion**, **Illustra**, **Itasca**...)
- **Đầu thế kỷ 21:** các HQTCSDL theo hướng NoSQL (**ODMG**)



CANTHO UNIVERSITY

Quá trình phát triển các HTQCSDL

- Hiện nay có khá nhiều HTQCSDL

Rank			DBMS	Database Model	Score		
Nov 2021	Oct 2021	Nov 2020			Nov 2021	Oct 2021	Nov 2020
1.	1.	1.	Oracle +	Relational, Multi-model ⓘ	1272.73	+2.38	-72.27
2.	2.	2.	MySQL +	Relational, Multi-model ⓘ	1211.52	-8.25	-30.12
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model ⓘ	954.29	-16.32	-83.35
4.	4.	4.	PostgreSQL + 💬	Relational, Multi-model ⓘ	597.27	+10.30	+42.22
5.	5.	5.	MongoDB +	Document, Multi-model ⓘ	487.35	-6.21	+33.52
6.	6.	↑ 7.	Redis +	Key-value, Multi-model ⓘ	171.50	+0.15	+16.08
7.	7.	↓ 6.	IBM Db2	Relational, Multi-model ⓘ	167.52	+1.56	+5.90
8.	8.	8.	Elasticsearch	Search engine, Multi-model ⓘ	159.09	+0.84	+7.54
9.	9.	9.	SQLite +	Relational	129.80	+0.43	+6.48
10.	10.	10.	Cassandra +	Wide column	120.88	+1.61	+2.13
11.	11.	11.	Microsoft Access	Relational	119.24	+2.86	+2.01
12.	12.	12.	MariaDB +	Relational, Multi-model ⓘ	102.19	-0.41	+9.90
13.	13.	13.	Splunk	Search engine	92.31	+1.69	+2.60
14.	14.	↑ 15.	Hive +	Relational	83.31	-1.43	+13.05
15.	15.	↑ 17.	Microsoft Azure SQL Database	Relational, Multi-model ⓘ	81.32	+1.60	+14.33

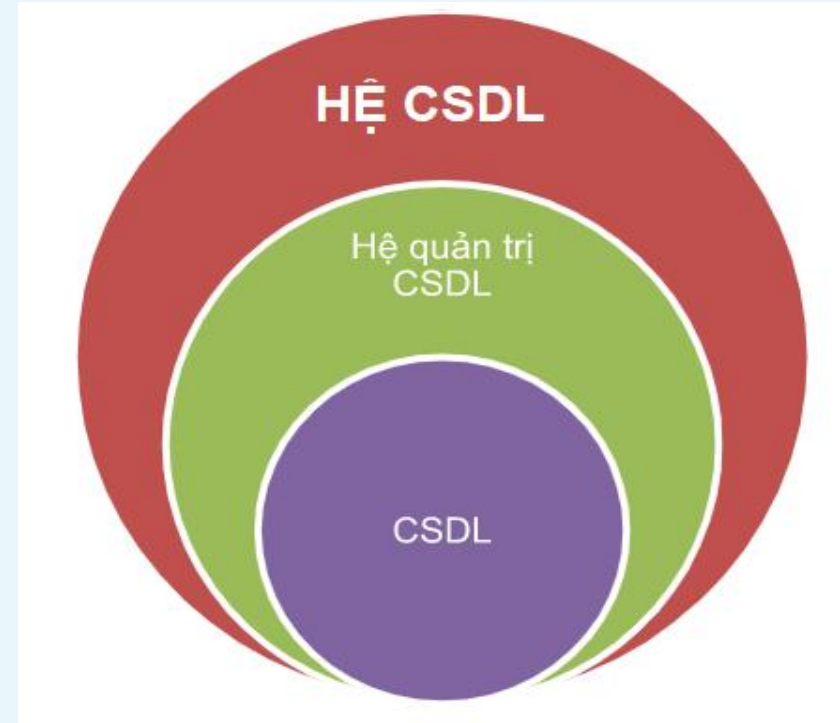
Top 15 hệ quản trị cơ sở dữ liệu phổ biến nhất trên thế giới. (Cập nhật tháng 11/2021)



CANTHO UNIVERSITY

Khái niệm hệ CSDL

- **Hệ CSDL:** Phần mềm cung cấp môi trường thuận lợi và hiệu quả để tạo lập, lưu trữ và khai thác thông tin của CSDL



$$\text{HỆ CSDL} = \text{CSDL} + \text{Hệ quản trị CSDL}$$



Các yêu cầu cơ bản của hệ CSDL

- **Tính cấu trúc:** thông tin trong CSDL được lưu trữ theo một cấu trúc xác định.
- **Tính toàn vẹn:** các giá trị DL được lưu trữ trong CSDL phải thỏa mãn 1 số ràng buộc thực tế.
- **Tính nhất quán:** hệ CSDL phải đảm bảo tính chính xác của DL ngay cả khi có sự cố xảy ra.



Các yêu cầu cơ bản của hệ CSDL

- **Tính an toàn và bảo mật thông tin:** CTƯ'D phải được phân quyền người dùng, ngăn chặn truy cập trái phép bảo đảm an toàn DL.
- **Tính độc lập:** DL độc lập với các ứng dụng, phương tiện lưu trữ và xử lí.
- **Tính không dư thừa:** trong CSDL không lưu trữ những DL có thể tính từ những DL khác (tuổi).

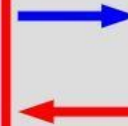
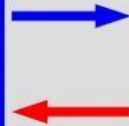
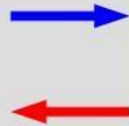
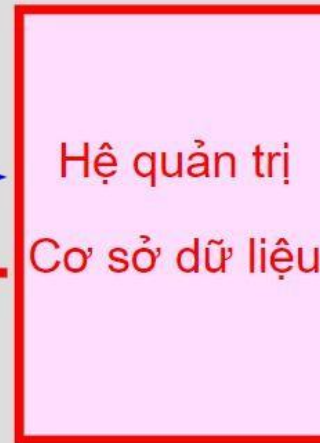


CANTHO UNIVERSITY

Cấu trúc hệ CSDL

Người dùng

Giao diện



Người sử dụng/Lập trình viên



Chương trình ứng dụng/Truy vấn



HQT CSDL

Xử lý truy vấn



Truy xuất dữ liệu



Catalog

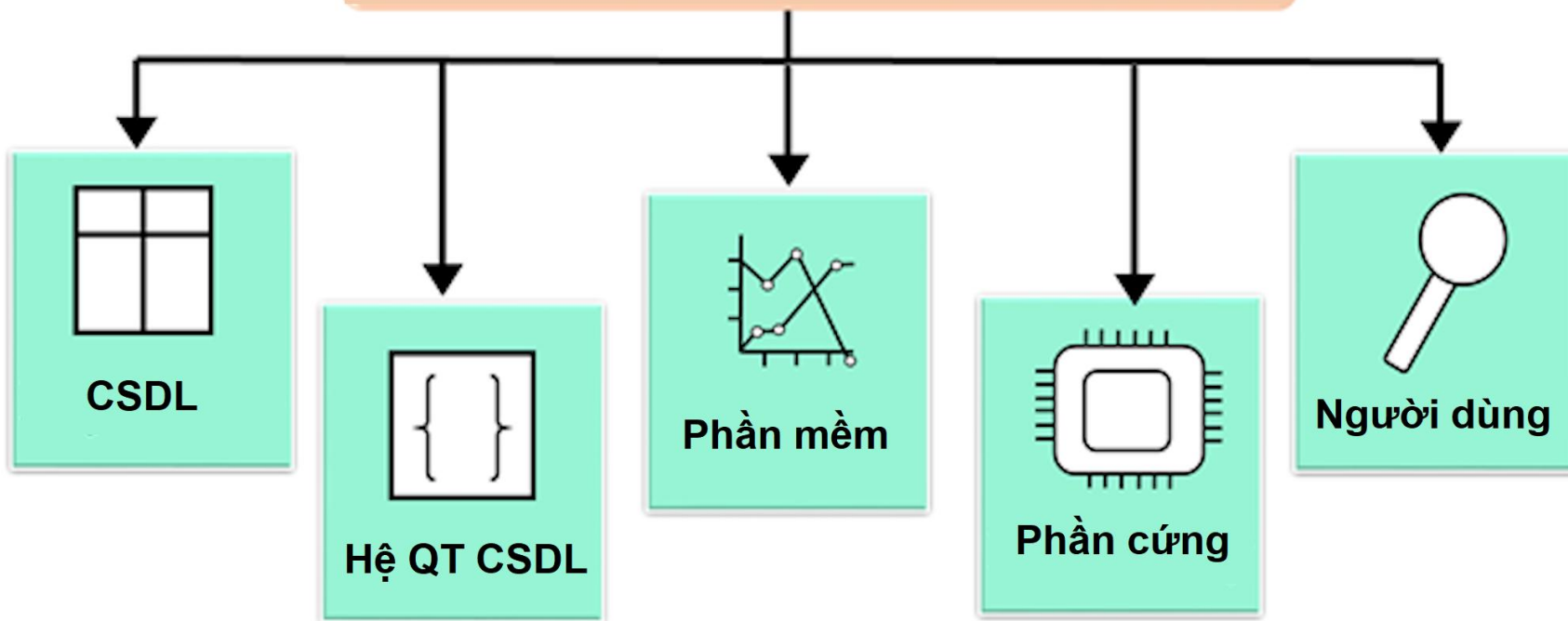
**Định nghĩa
CSDL**

CSDL

Hệ CSDL

Các thành phần hệ CSDL

Các thành phần của một hệ CSDL





CANTHO UNIVERSITY

Các thành phần hệ CSDL (tt)

❖ Người dùng trực tiếp:

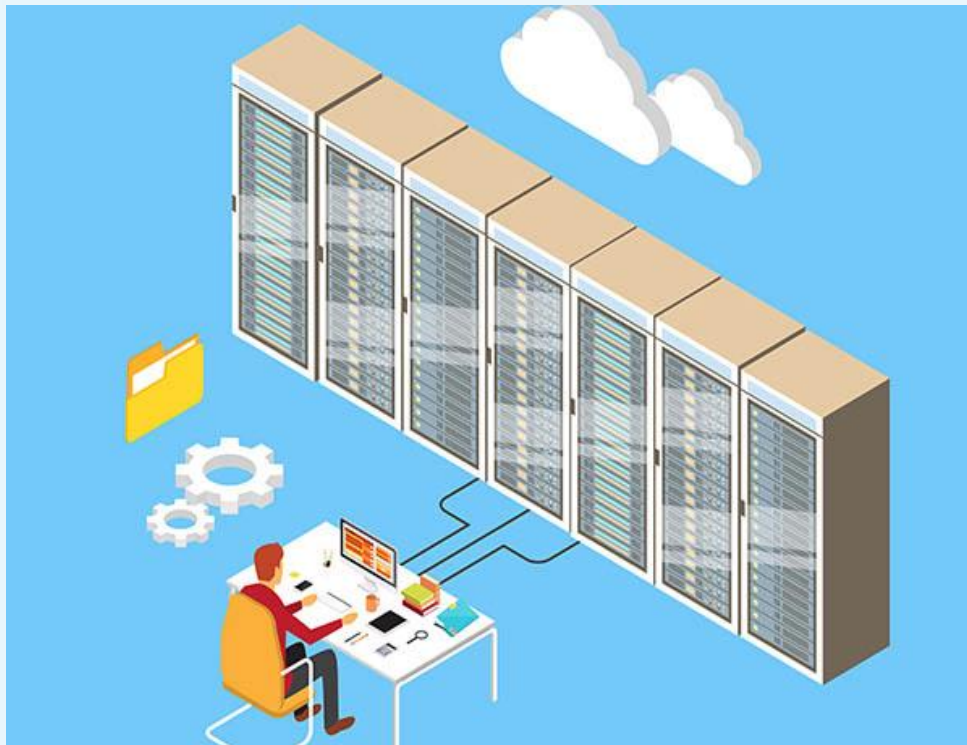
- Người quản trị hệ cơ sở dữ liệu (Database Administrator – DBA)
- Người thiết kế cơ sở dữ liệu (Database Designer)
- Những người sử dụng cuối (End User)
- Người phân tích hệ thống và lập trình ứng dụng (Software Engineers)



Các thành phần hệ CSDL (tt)

❖ Người dùng gián tiếp:

- Người thiết kế và cài đặt hệ quản trị dữ liệu
- Những người phát triển công cụ
- Các thao tác viên và những người bảo trì





Các ngôn ngữ giao tiếp CSDL

- **Ngôn ngữ mô tả dữ liệu (DDL-Data Definition Language):** khai báo cấu trúc CSDL, các mối liên hệ DL, cấu trúc ràng buộc DL.
- **Ngôn ngữ thao tác dữ liệu (DML-Data Manipulation Language):** thực hiện thao tác thêm, sửa, xóa DL.



Các ngôn ngữ giao tiếp CSDL (tt)

- **Ngôn ngữ truy vấn có cấu trúc (SQL - Structured Query Language):** cho phép người khai thác sử dụng để truy vấn thông tin cần thiết.
- **Ngôn ngữ quản lý dữ liệu (DCL - Data Control Language):** thay đổi cấu trúc bảng, khai báo bảo mật, cấp quyền cho người sử dụng.



2. Ưu, nhược điểm của hệ QT CSDL

■ Hệ thống xử lý tập tin truyền thống

Ưu điểm:

- Dễ dàng truy cập
- Dễ hiểu

Nhược điểm:

- Sự cô lập dữ liệu
- Trùng lặp => Dư thừa
- Thời gian phát triển lâu
- Chi phí bảo trì cao
- Truy xuất khó khăn

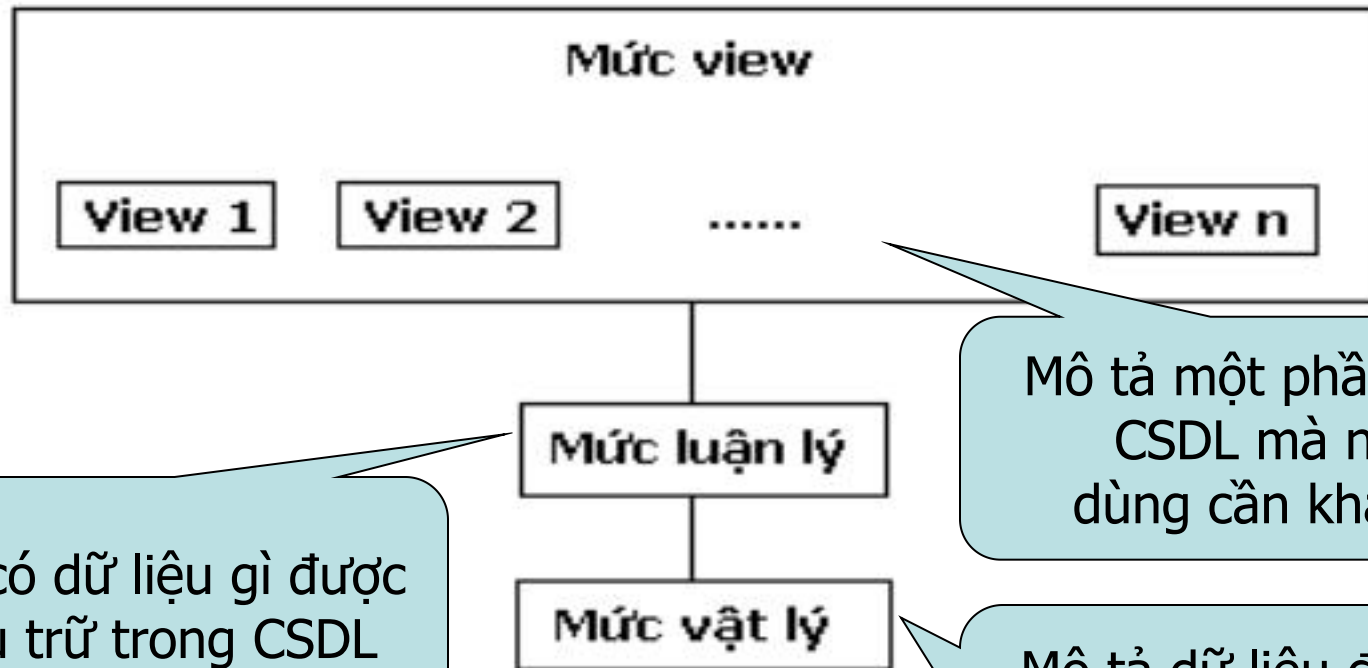
2. Ưu, nhược điểm của hệ QT CSDL (tt)

Ưu điểm	Nhược điểm
- Quản lý được DL dư thừa	- Khá phức tạp
- Đảm bảo tính nhất quán DL	- Kích thước lớn
- Cải tiến tính toàn vẹn cho DL	- Giá cả cao
- Chia sẻ DL được nhiều hơn	- Hiệu năng hệ thống giảm



3. Góc nhìn dữ liệu

- Để **che dấu sự phức tạp** của CSDL bằng cách thiết kế nó thành các lớp trừu tượng

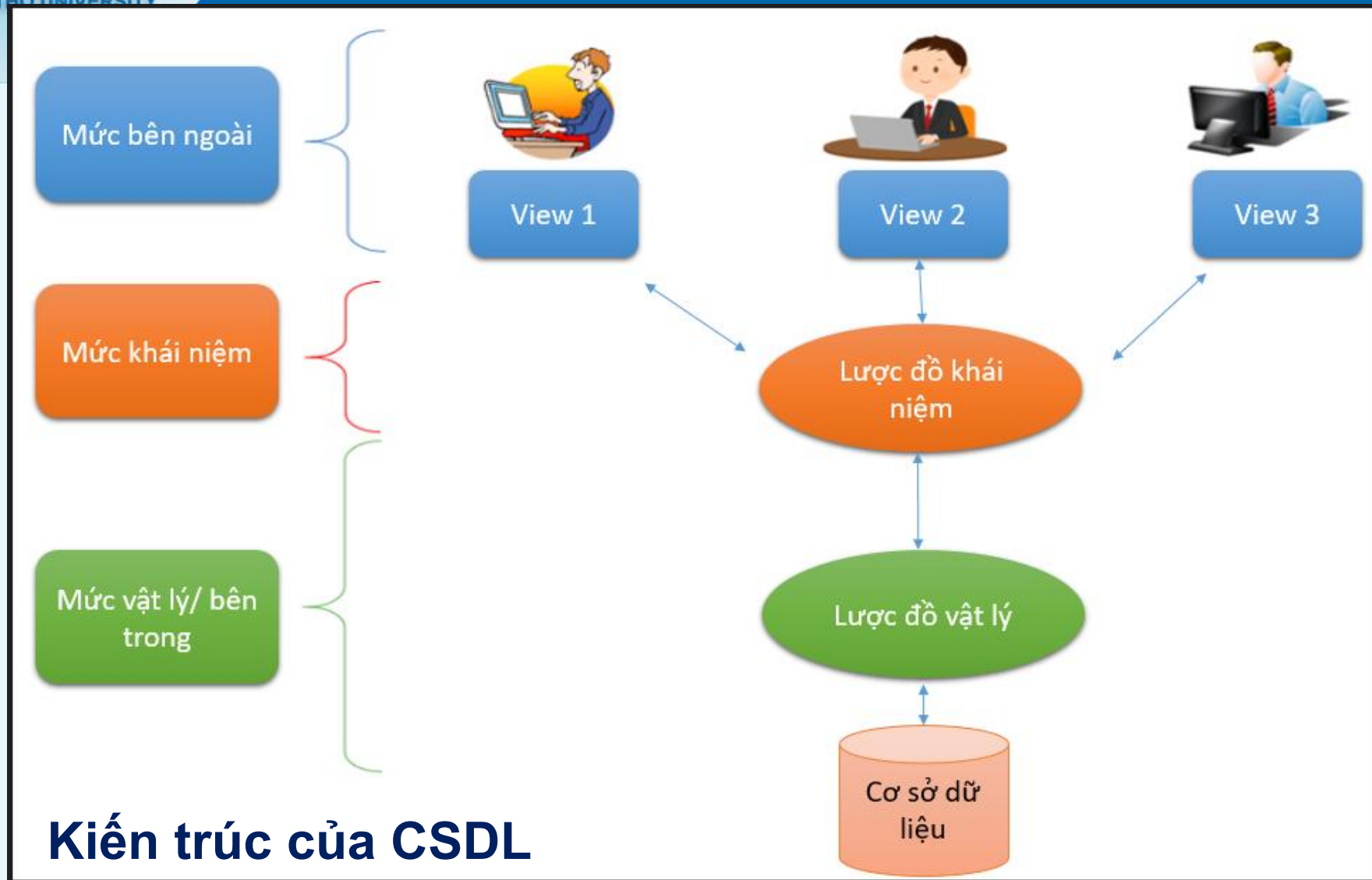


Mô tả có dữ liệu gì được lưu trữ trong CSDL

Mô tả một phần của CSDL mà người dùng cần khai thác

Mô tả dữ liệu được lưu trữ như thế nào trong thiết bị lưu trữ

3. Góc nhìn dữ liệu (tt)





Ví dụ: Kiến trúc CSDL Nhân viên

Khung nhìn mức ngoài 1

Maso Hodem Ten Tuoi Luong

Khung nhìn mức ngoài 2

MaNV Ten Ma_chi_nhanh

Mức khái niệm

MaNV Hodem Ten Ngaysinh Luong Ma_chi_nhanh

Mức trong

```
struct NHANVIEN {  
    int MaNV;  
    int Ma_chi_nhanh  
    char Hodem[15];  
    char Ten[15];  
    struct date Ngaysinh;
```


3. Góc nhìn dữ liệu (tt)

Mức View

View 1

View 2

View 3

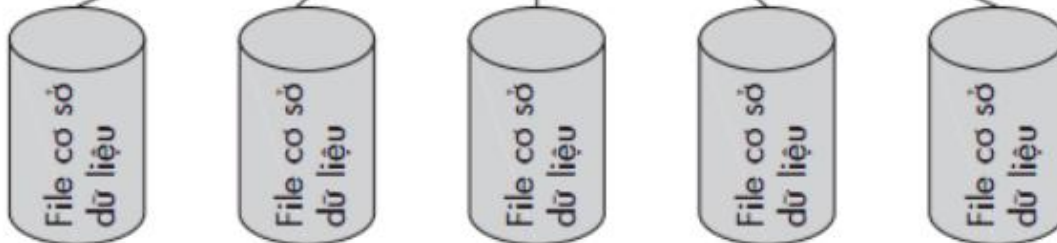
Schema (sơ đồ)

Mức luận lý

Độc lập dữ
liệu luận lý

Mức vật lý

Độc lập dữ
liệu vật lý





3. Góc nhìn dữ liệu (tt)

- Khả năng sửa đổi một định nghĩa ở một mức không ảnh hưởng đến mức cao hơn được gọi là **sự độc lập dữ liệu**.
 - **Độc lập dữ liệu vật lý**: khả năng sửa đổi sơ đồ vật lý mà không viết lại chương trình.
 - **Độc lập dữ liệu luận lý**: khả năng thay đổi sơ đồ luận lý mà không phải viết lại chương trình.



4. Mô hình hệ CSDL

- **Mô hình dữ liệu** là một bộ các công cụ quan niệm để mô tả dữ liệu, quan hệ dữ liệu và các ràng buộc

Mô hình
mức cao

Mô hình cài
đặt

Mô hình
mức thấp

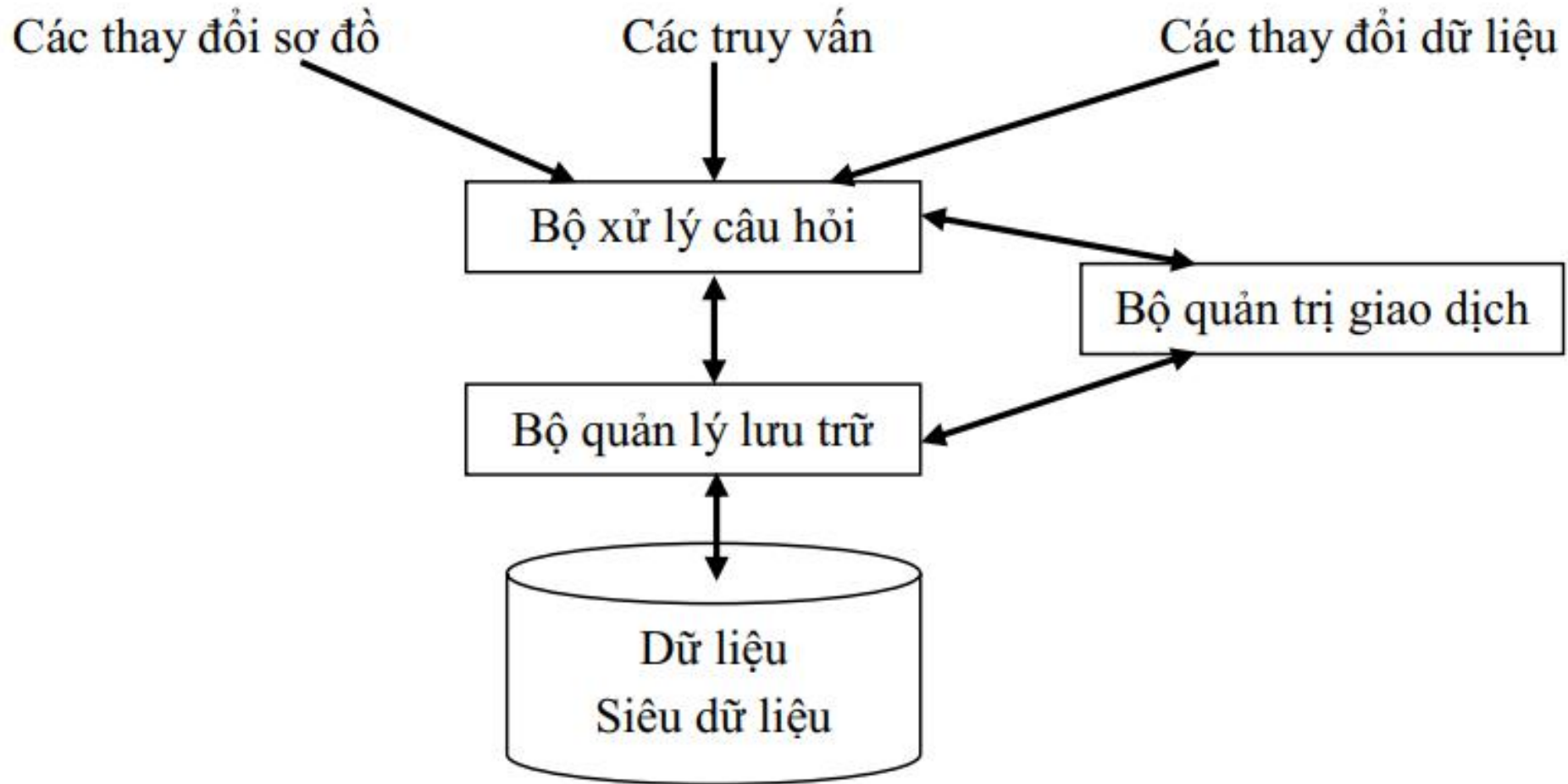


4. Mô hình hệ CSDL (tt)

- **Mô hình mức cao**
 - Cung cấp các khái niệm gần gũi với người dùng
 - Mô hình phải tự nhiên và giàu ngữ nghĩa
 - VD: mô hình thực thể kết hợp (ER), mô hình đối tượng
- **Mô hình cài đặt**
 - Đưa ra các khái niệm người dùng có thể hiểu được nhưng không quá xa với cách DL được tổ chức thật trên máy tính.
 - VD: mô hình quan hệ, mô hình mạng, mô hình phân cấp
- **Mô hình mức thấp**
 - Đưa ra các khái niệm mô tả chi tiết về cách thức dữ liệu được lưu trữ trong máy tính.



5. Kiến trúc của một hệ QT CSDL





5. Kiến trúc hệ QT CSDL (tt)

❖ Các thao tác với hệ QT CSDL:

- Các truy vấn: những hỏi đáp, yêu cầu về dữ liệu
- Các thay đổi sơ đồ dữ liệu: những thay đổi của người quản trị CSDL
- Các cập nhật dữ liệu: các thao tác xóa, sửa dữ liệu trong CSDL



5. Kiến trúc hệ QT CSDL (tt)

❖ Bộ xử lý câu hỏi:

- Tìm kiếm dữ liệu trả lời cho một yêu cầu truy vấn
- Lựa chọn cách tốt nhất để trả lời truy vấn.
- Phát ra lệnh đối với **bộ quản lý lưu trữ** và thực thi truy vấn đó.

❖ Bộ quản lý lưu trữ

- Lấy các thông tin dữ liệu được yêu cầu.
- Thực thi các thao tác thay đổi với CSDL.



5. Kiến trúc hệ QT CSDL (tt)

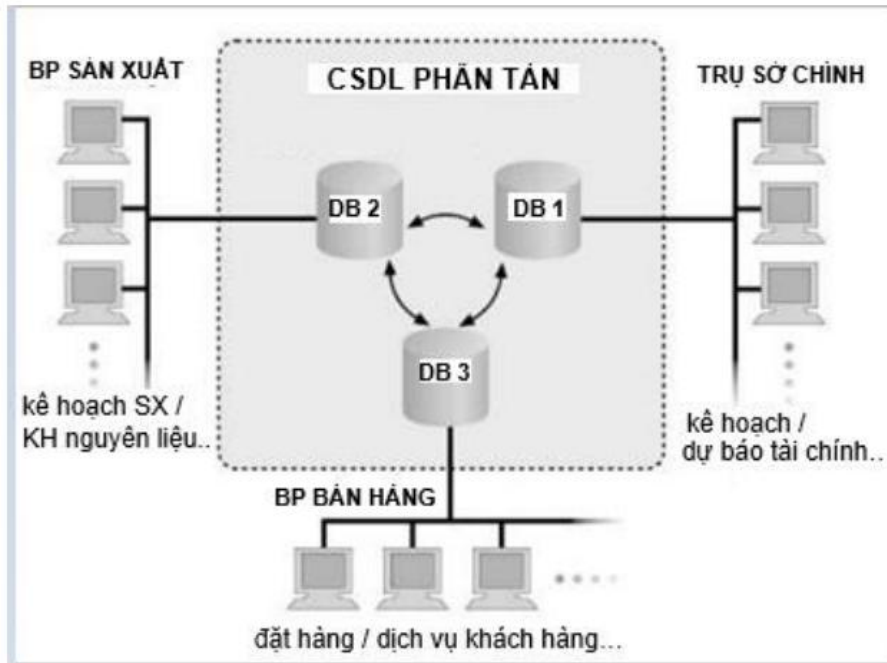
❖ Bộ quản lý giao dịch:

- 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.

❖ Dữ liệu, siêu dữ liệu:

- Thiết bị nhớ ngoài chứa dữ liệu được lưu trữ trong CSDL và thông tin cấu trúc của CSDL.
- Đây là thành phần chính của một hệ QT CSDL.

6. Các loại kiến trúc của hệ CSDL



①



②



6.1 Các hệ CSDL tập trung

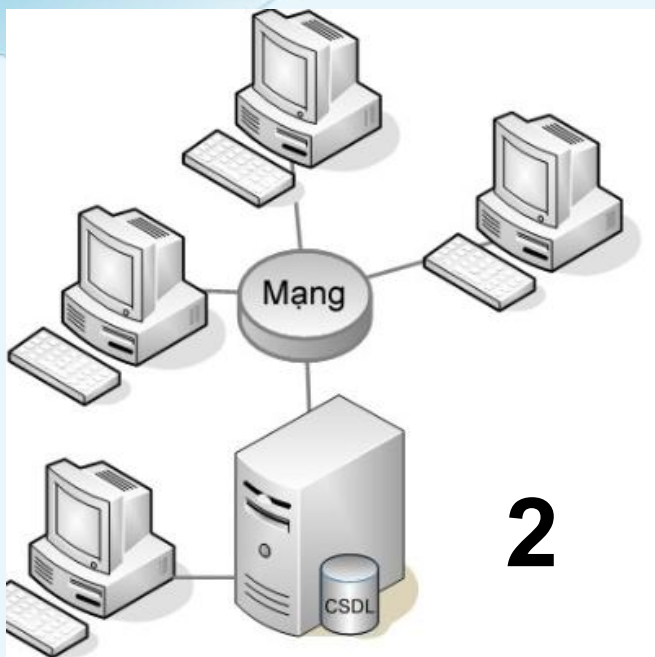
- Toàn bộ dữ liệu được lưu tại một máy hoặc một dàn máy. Những người dùng từ xa có thể truy cập vào CSDL thông qua các phương tiện truyền thông dữ liệu.
- Chia thành 3 loại:
 - Hệ CSDL cá nhân
 - Hệ CSDL trung tâm
 - Hệ CSDL khách - chủ



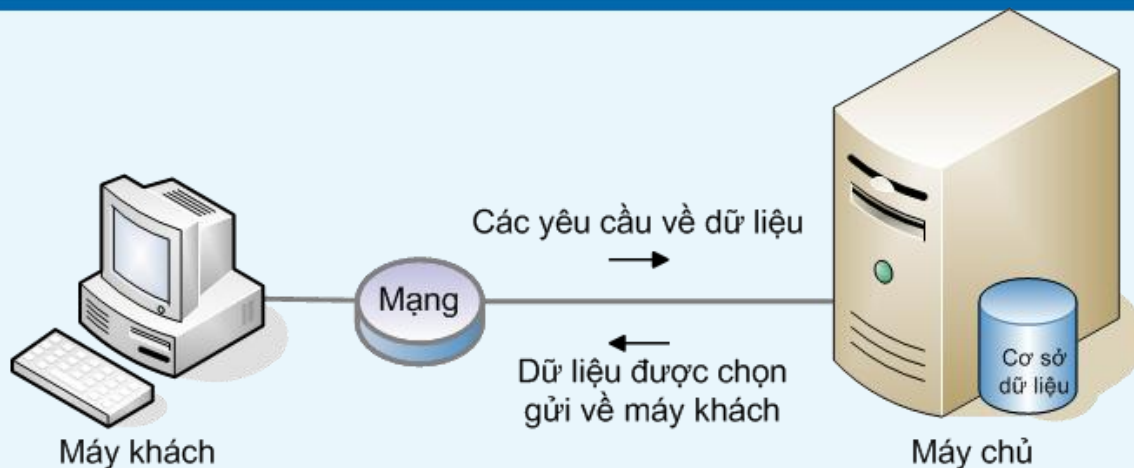


CANTHO UNIVERSITY

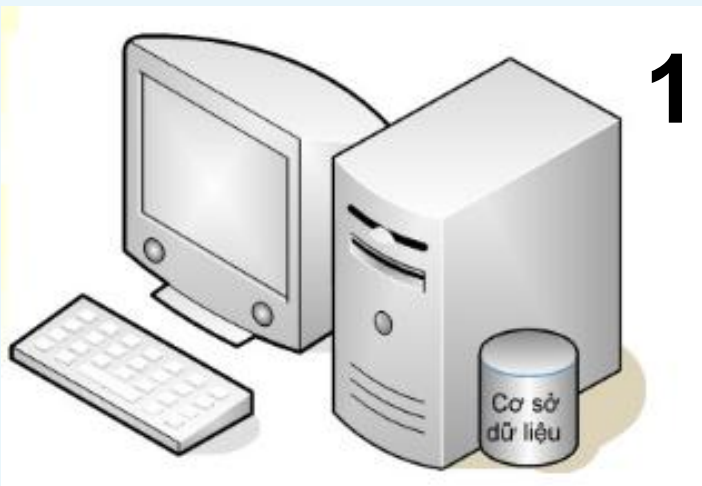
6.1 Các hệ CSDL tập trung (tt)



Hệ CSDL trung tâm



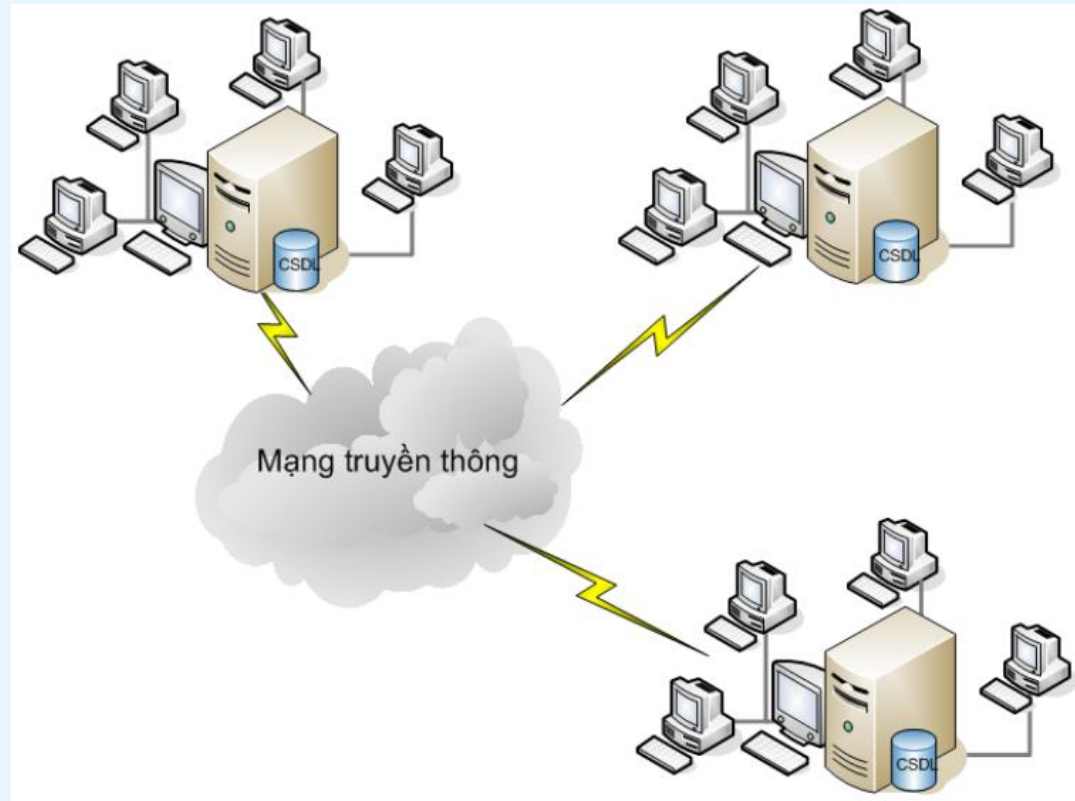
3 Hệ CSDL khách/chủ



Hệ CSDL cá nhân

6.2 Các hệ CSDL phân tán

- Là một tập hợp dữ liệu có liên quan (về logic) được dùng chung và phân tán về mặt vật lí trên một mạng máy tính.
- Một **HQT CSDL** phân tán là một hệ thống phần mềm cho phép quản trị CSDL phân tán và làm cho người dùng không nhận thấy sự phân tán.





6.2 Các hệ CSDL phân tán (tt)

- Người dùng truy cập vào CSDL phân tán thông qua **chương trình ứng dụng**:
 - **Không** yêu cầu dữ liệu từ nơi khác
 - **Có** yêu cầu dữ liệu từ nơi khác.
- Hệ CSDL phân tán được chia làm 2 loại:
 - Hệ QT CSDL phân tán **thuần nhất**
 - Hệ QT CSDL phân tán **không thuần nhất** (hỗn hợp)



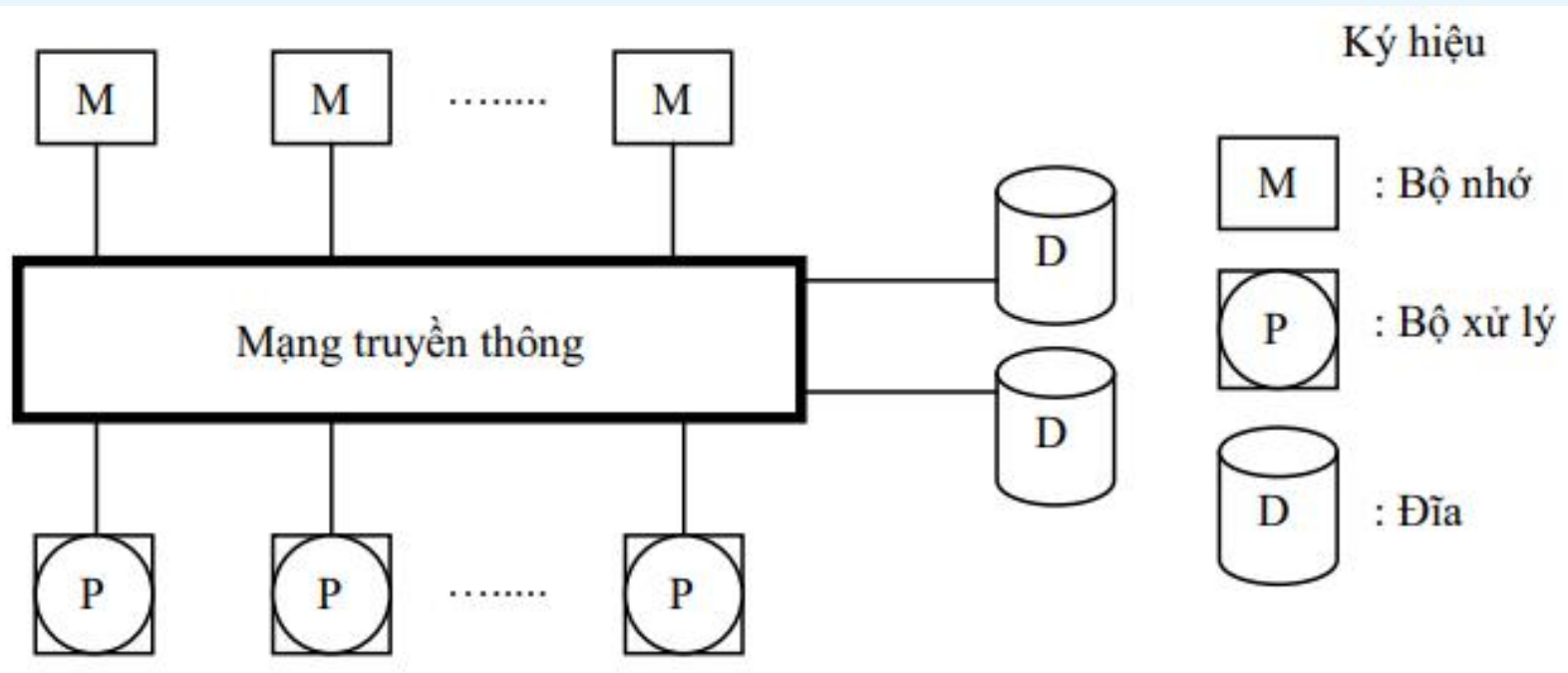
Kiến trúc hệ CSDL song song

- Cải tiến tốc độ xử lý và tốc độ I/O bằng cách sử dụng nhiều CPU và đĩa song song.
- Có hai biện pháp chính để đánh giá hiệu năng của một DBS:
 - Năng lực truyền qua (throughput): Số công việc được hoàn tất trong một thời gian đã cho.
 - Thời gian đáp ứng (response time): Lượng thời gian cần thiết để hoàn thành một công việc từ khi nó được đệ trình.



Kiến trúc hệ CSDL song song (tt)

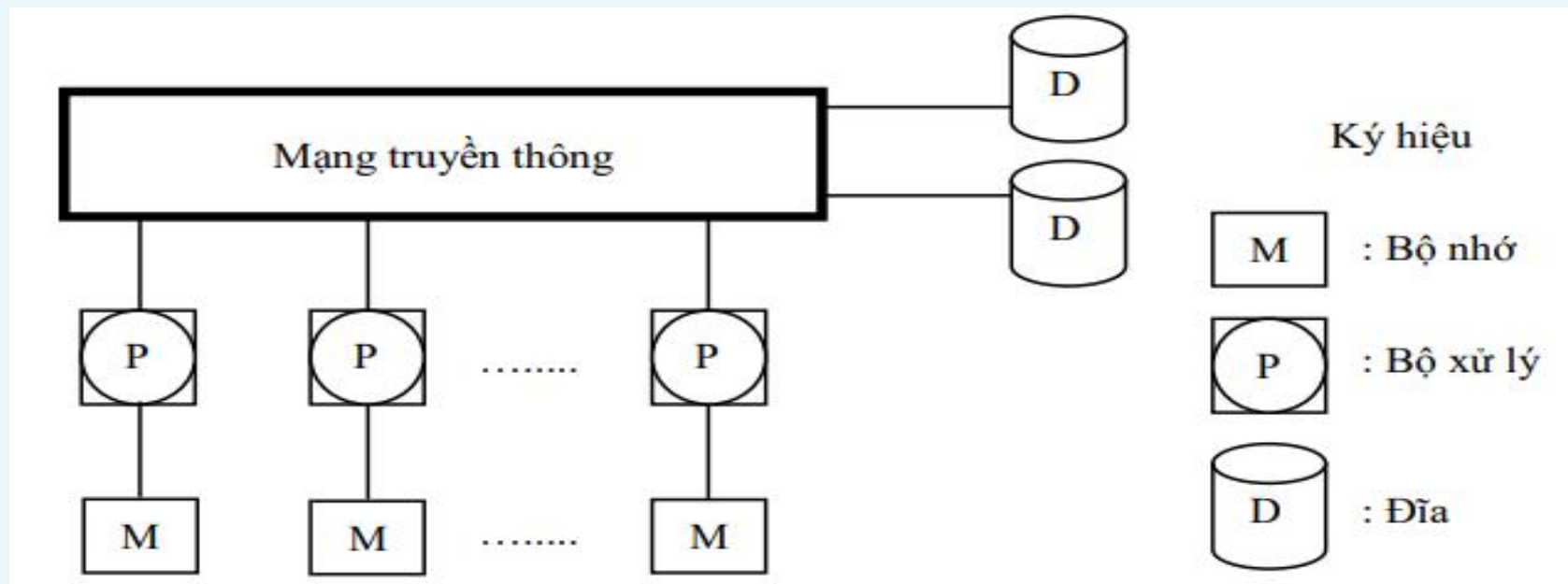
- **Kiến trúc mọi thứ dùng chung (shared everything)**



- **Ưu điểm:** Tốc độ liên lạc giữa các P là hiệu quả
- **Hạn chế:** Mỗi P thường có cache lớn => dễ bị nghẽn

Kiến trúc hệ CSDL song song (tt)

- Kiến trúc dùng chung đĩa (shared disk)



Ưu điểm: M không bị nghẽn. P không ảnh hưởng (lượng thứ lỗi)

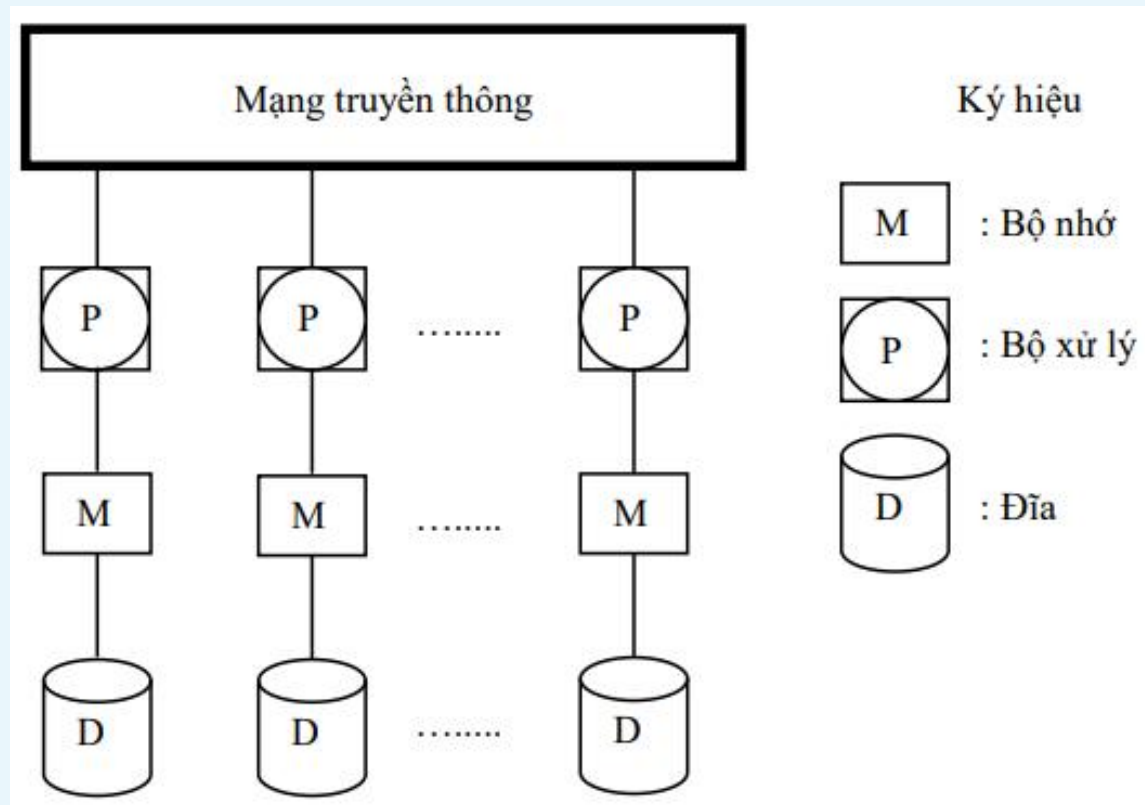
Hạn chế: Các đĩa dễ bị nghẽn. Tốc độ liên lạc giữa các P chậm.

Kiến trúc hệ CSDL song song (tt)

- Kiến trúc không chia sẻ (shared nothing)

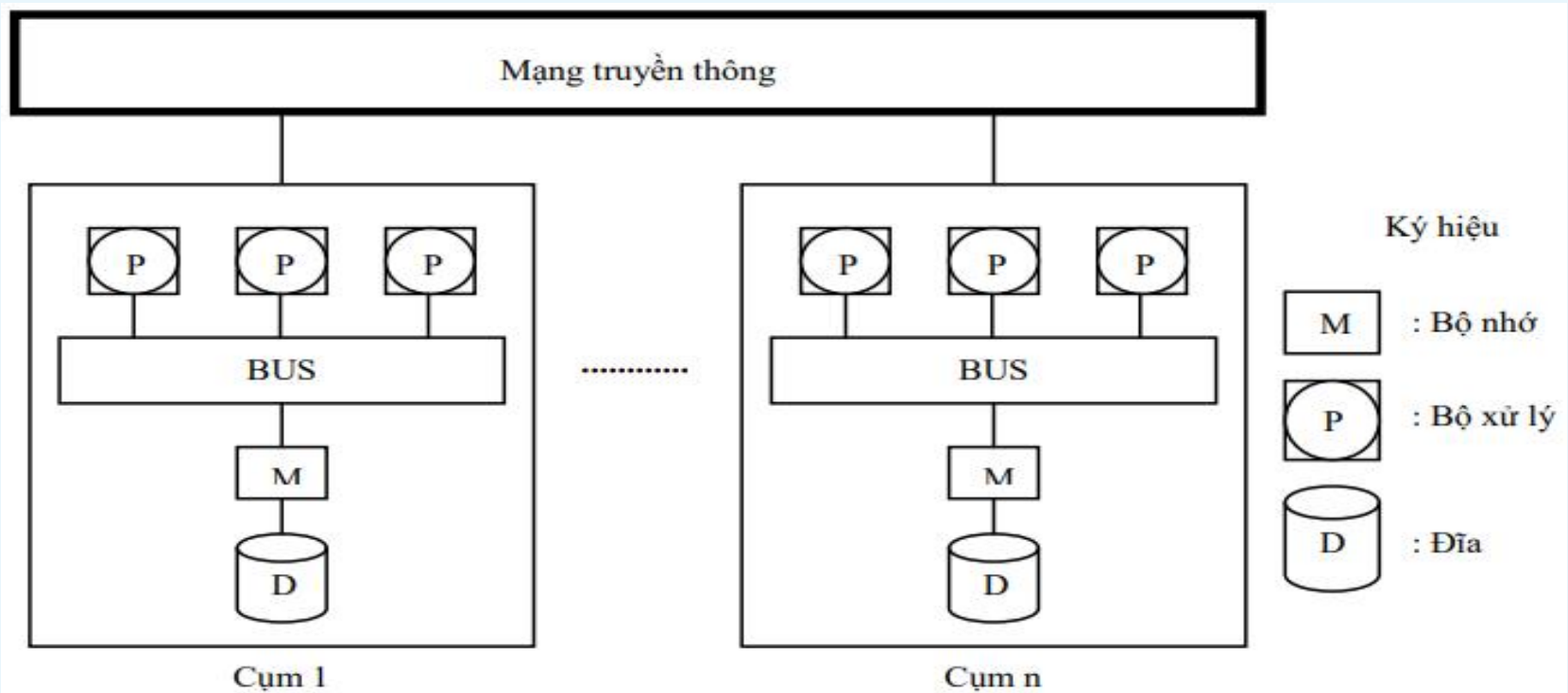
Hạn chế:

- Vấn đề phát sinh khi các CPU truy xuất các đĩa không cục bộ
- Việc truyền các quan hệ kết quả qua mạng



Kiến trúc hệ CSDL song song (tt)

- Kiến trúc phân cấp**



Tổ hợp các đặc trưng của các kiến trúc chia sẻ bộ nhớ, chia sẻ đĩa và không chia sẻ gì



7. Quản trị giao dịch

- **Giao dịch:** tập các thao tác trên CSDL tạo thành một đơn vị công việc logic.
- Khi một GD xảy ra, để đảm bảo không ảnh hưởng đến tính chính xác của dữ liệu trong CSDL. GD đòi hỏi phải thỏa 4 tính chất: nguyên tử (**A**tomicity), nhất quán (**C**onsistency), cô lập (**I**solate), bền vững (**D**urability).



7. Quản trị giao dịch (tt)

❖ Ý nghĩa của các tính chất:

- Tính nguyên tử: đòi hỏi toàn thể - hoặc - không.
- Tính bền vững: đòi hỏi về sự tồn tại bền vững của DL.
- Tính cô lập: đòi hỏi các GD trong hệ thống không cạnh tranh dữ liệu lẫn nhau (thực hiện tuần tự).
- Tính nhất quán: đòi hỏi tính chính xác.



7. Quản trị giao dịch (tt)

❖ Việc đảm bảo các tính chất:

- Tính nguyên tử + bền vững: bộ điều khiển phục hồi.
- Tính cô lập: bộ điều khiển cạnh tranh.
- Tính nhất quán: người sử dụng.



CANTHO UNIVERSITY

HẾT CHƯƠNG 1