

Chương 1

TỔNG QUAN VỀ CÁC HỆ CƠ SỞ DỮ LIỆU

NỘI DUNG

- Khái niệm CSDL
- Hệ quản trị CSDL
- Tính nhất quán và toàn vẹn dữ liệu
- Lược đồ và Thể hiện của CSDL
- Tính độc lập dữ liệu
- Kiến trúc của một hệ CSDL
- Mô hình dữ liệu
- Kiến trúc của hệ CSDL đa người dùng
- Vai trò của con người đối với các hệ CSDL

Khái niệm cơ sở dữ liệu, hệ cơ sở dữ liệu

- Một CSDL (Database) là một tập hợp các dữ liệu cần được quản lý, được lưu trữ đồng thời trên thiết bị nhớ thứ cấp, được quản lý theo một cơ chế thống nhất, nhằm thực hiện ba chức năng sau một cách tối ưu: mô tả dữ liệu, tìm kiếm dữ liệu, cập nhật dữ liệu.

Tính nhất quán và toàn vẹn dữ liệu

■ Tính nhất quán dữ liệu:

- ◆ Dữ liệu được lưu trữ trong CSDL phải phù hợp với yêu cầu thực tiễn và không mâu thuẫn nhau
- ◆ Ràng buộc dữ liệu: Những quy định mà dữ liệu trong CSDL phải thỏa mãn.

■ Ví dụ

- ◆ $Tuoi > 0$; Ngày_đến < ngày_đi
- ◆ $Tổng_điểm = Toán + lý + hoá$; $Thành_tiền = giá * slượng$
- ◆ Mỗi sinh viên có một mã số duy nhất và không có 2 sinh viên nào cùng mã số
- ◆

■ Database Management System

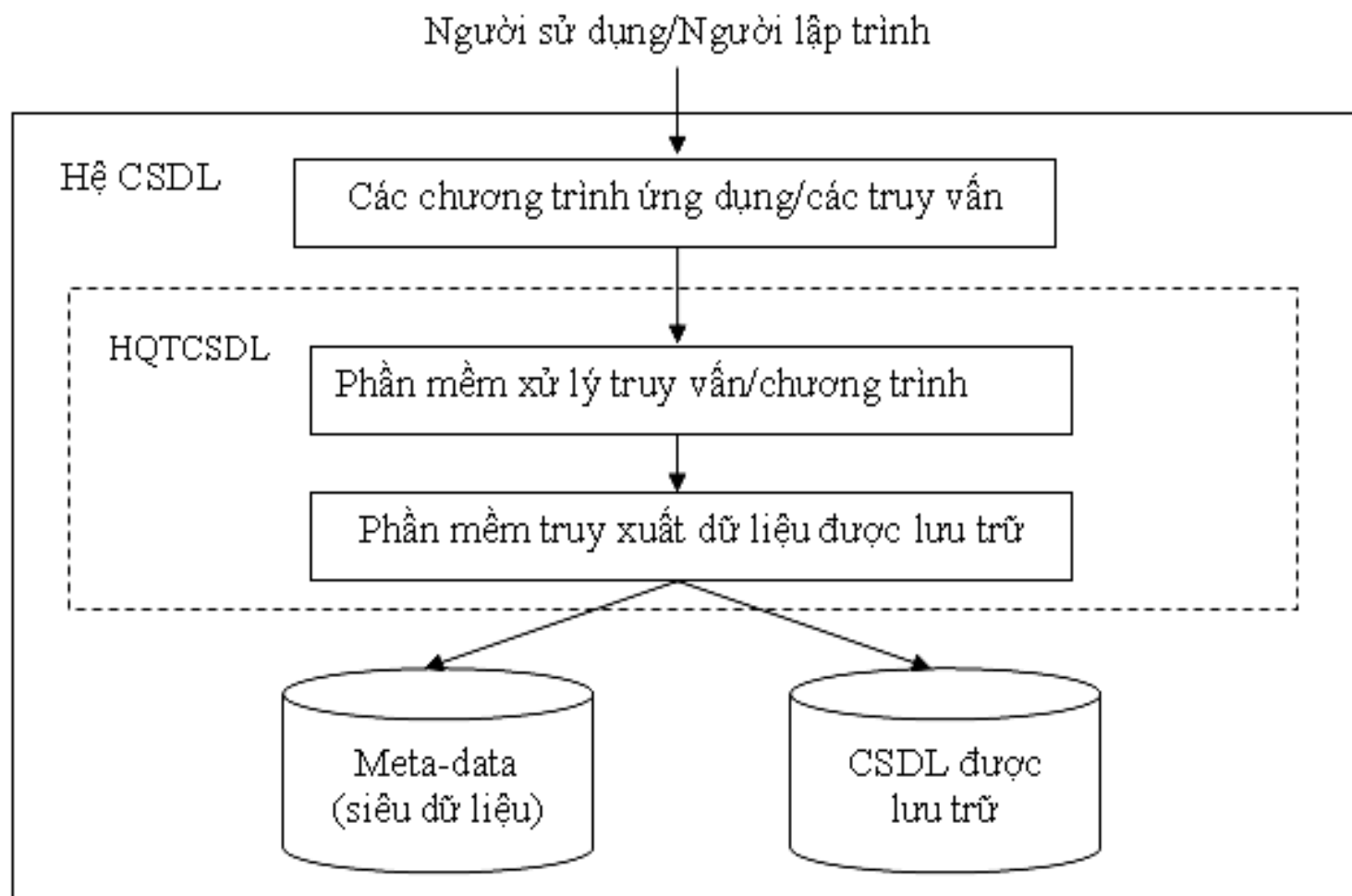
- ◆ *phần mềm cho phép người sử dụng tạo lập và duy trì CSDL một cách thuận lợi và hiệu quả*

■ Cho phép người dùng định nghĩa CSDL (DDL)

■ Cho phép người dùng cập nhật, tìm kiếm dữ liệu (DML)

■ Cung cấp các kiểm soát, truy nhập điều khiển,

- ◆ Bảo mật dữ liệu
- ◆ Toàn vẹn dữ liệu (Integrity)
- ◆ Điều khiển tương tranh
- ◆ Phục hồi dữ liệu (Recovery)
- ◆ Từ điển dữ liệu cho phép người dùng truy nhập



Hình 1.1 Một hệ CSDL đơn giản

Lược đồ và Thể hiện

■ Lược đồ CSDL (database schema):

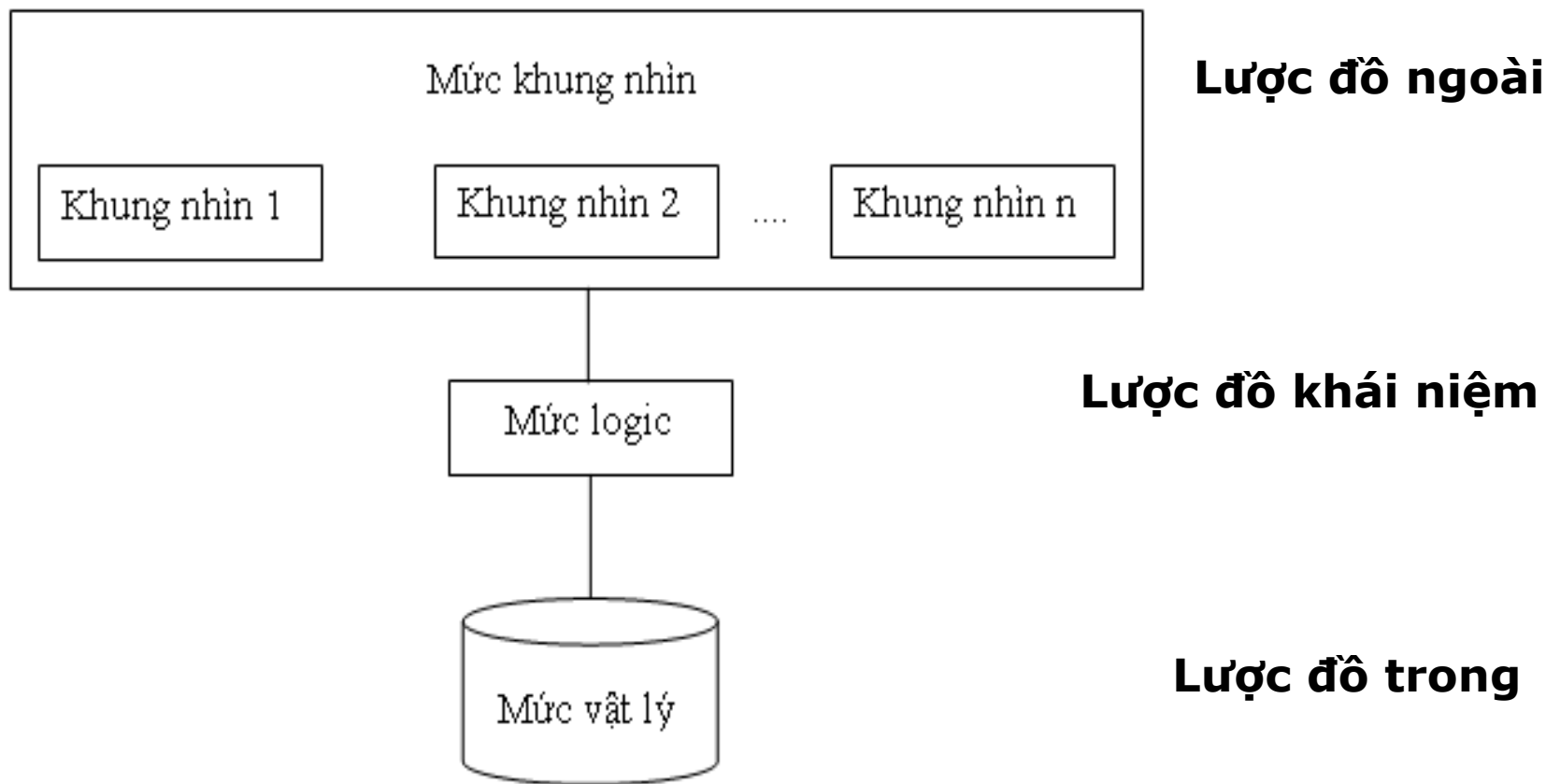
- ◆ Toàn bộ mô tả về CSDL.
- ◆ Lược đồ được xác định trong quá trình thiết kế CSDL
- ◆ Ít khi thay đổi

■ Thể hiện của CSDL (database instance):

- ◆ Toàn bộ dữ liệu trong CSDL tại một thời điểm (trạng thái của CSDL).
- ◆ Thường xuyên thay đổi theo thời gian bởi các thao tác bổ sung, xóa hay sửa đổi.

Kiến trúc của một hệ CSDL

■ Kiến trúc 3 mức ANSI / SPARC



Hình 1.2. Các mức trừu tượng của một hệ CSDL

Tính độc lập dữ liệu

- Tính độc lập dữ liệu (data independence):
 - ◆ sự thay đổi các lược đồ ở một mức không làm ảnh hưởng đến các lược đồ ở mức cao hơn.
- Hai loại độc lập dữ liệu:
 - ◆ Độc lập dữ liệu mức vật lý
 - ◆ Độc lập dữ liệu mức logic
- Theo C.J. Date: *“Độc lập dữ liệu là tính bất biến của các hệ ứng dụng đối với các thay đổi trong chiến lược lưu trữ và chiến lược truy nhập dữ liệu”*

Tính độc lập dữ liệu

- **Độc lập dữ liệu mức vật lý (*Physical Data Independence*):**
khả năng thay đổi lược đồ mức vật lý nhưng không làm thay đổi lược đồ mức logic
 - ◆ Chương trình ứng dụng phụ thuộc vào lược đồ mức logic
- **Độc lập dữ liệu mức logic (*Logical Data Independence*):**
khả năng thay đổi lược đồ mức logic nhưng không làm thay đổi lược đồ mức ngoài
- Nói chung, giao diện giữa các mức khác nhau và các thành phần trong chương trình ứng dụng cần được định nghĩa/ cài đặt sao cho thay đổi ở một phần ít ảnh hưởng nghiêm trọng đến các phần khác!

Mô hình dữ liệu

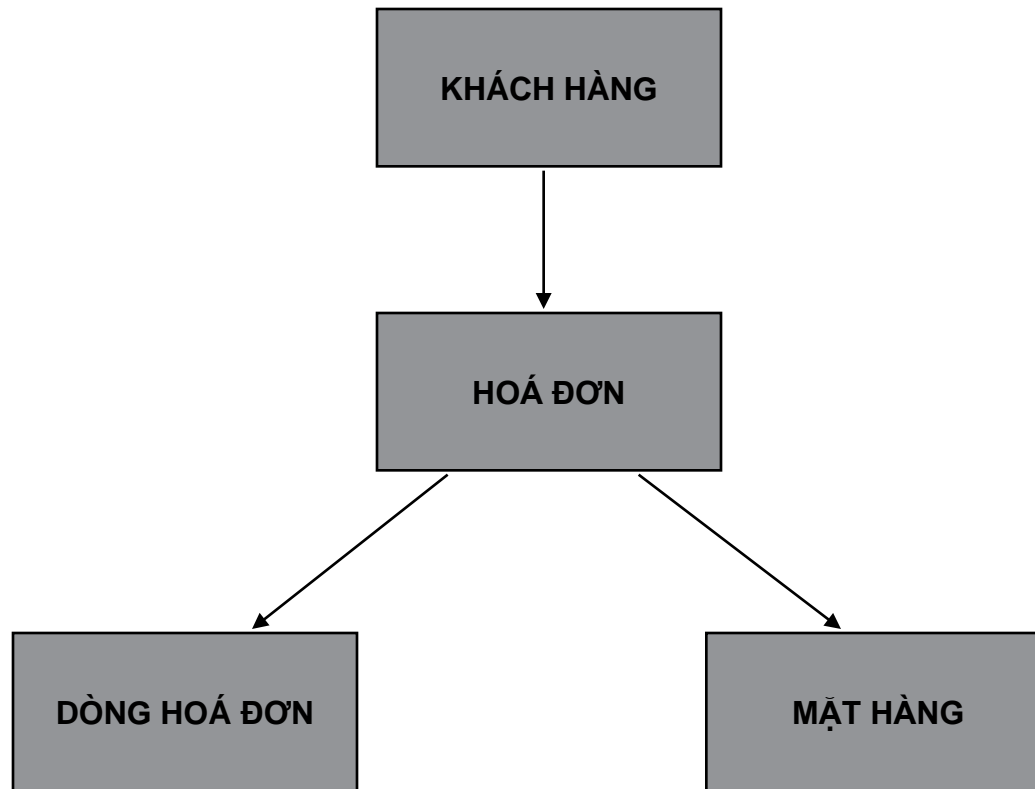
■ Mô hình dữ liệu (Data Model)

- ◆ tập các khái niệm và ký pháp dùng để mô tả dữ liệu, các mối quan hệ của dữ liệu, các ràng buộc trên dữ liệu của một CSDL
- ◆ tập các phép toán để thao tác trên dữ liệu

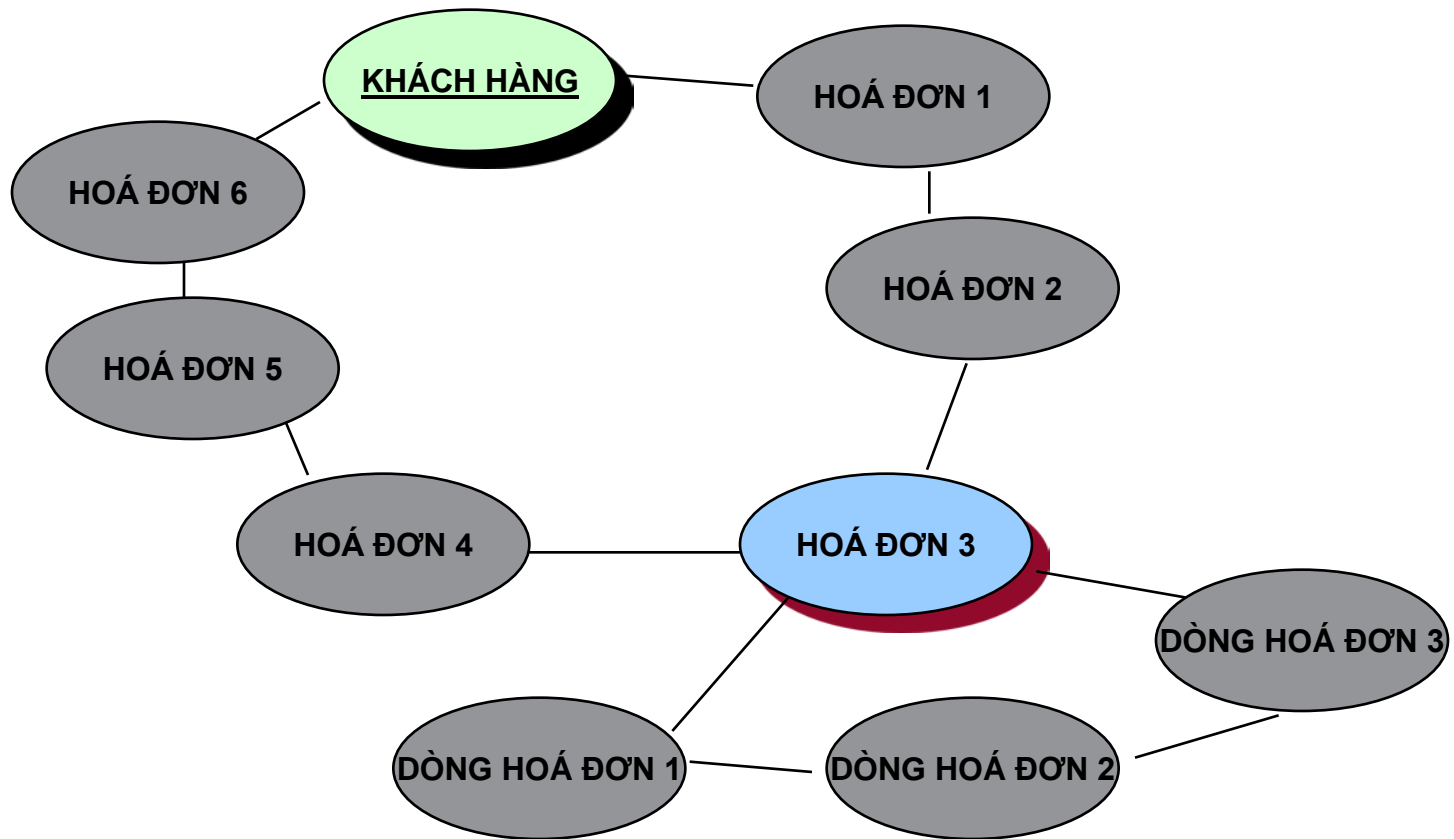
■ Các cách tiếp cận được biết đến nhiều nhất:

- ◆ mô hình thực thể-quan hệ (entity-relationship model)
- ◆ mô hình hướng đối tượng (object-oriented model)
- ◆ mô hình mạng (network data model)
- ◆ mô hình phân cấp (hierarchical data model)
- ◆ mô hình quan hệ (relational data model)

Mô hình phân cấp



MÔ HÌNH MẠNG



MÔ HÌNH DỮ LIỆU QUAN HỆ

- Hệ CSDL quan hệ đầu tiên do Codd đưa ra năm 1970
- Cố gắng kế thừa một mô hình chuẩn có nền tảng toán học vững chắc
 - ◆ Tại sao điều này tạo nên sự khác biệt so với các hệ thống trước?
- Mô hình dữ liệu thành công nhất (phổ biến nhất)
- Chủ yếu sử dụng ngôn ngữ vấn tin (query language) SQL

MÔ HÌNH DỮ LIỆU QUAN HỆ

■ Quan hệ Khách hàng

Mskh	Tenkh	Diachi
<i>KH1</i>	<i>Nguyễn Trung Thành</i>	<i>Hà Nội</i>
<i>KH2</i>	<i>Mai Chung Thuỷ</i>	<i>Huế</i>

■ Quan hệ Hoá đơn

Sohd	Ngay	Mskh	Tongtien
<i>HD1</i>	<i>2/2/2009</i>	<i>KH1</i>	<i>1.000</i>
<i>HD2</i>	<i>3/3/2009</i>	<i>KH2</i>	<i>2.300</i>
<i>HD3</i>	<i>3/3/2009</i>	<i>KH1</i>	<i>950</i>

MÔ HÌNH HƯỚNG ĐỐI TƯỢNG

- **Các hệ thống CSDL cho mục đích chung phải quản lý những dữ liệu động, phức tạp với các ứng dụng phức tạp:**
 - ◆ thiết kế có sự trợ giúp của máy tính (computer aided design-CAD)
 - ◆ công nghệ phần mềm có sự trợ giúp của máy tính (computer aided software engineering - CASE)
 - ◆ hệ thống thông tin địa lý
 - ◆ Các web sites tương tác
 - ◆
- **Mô hình dữ liệu quan hệ không đáp ứng được các hệ thống này.**

MÔ HÌNH HƯỚNG ĐỐI TƯỢNG

■ Các loại CSDL hướng đối tượng

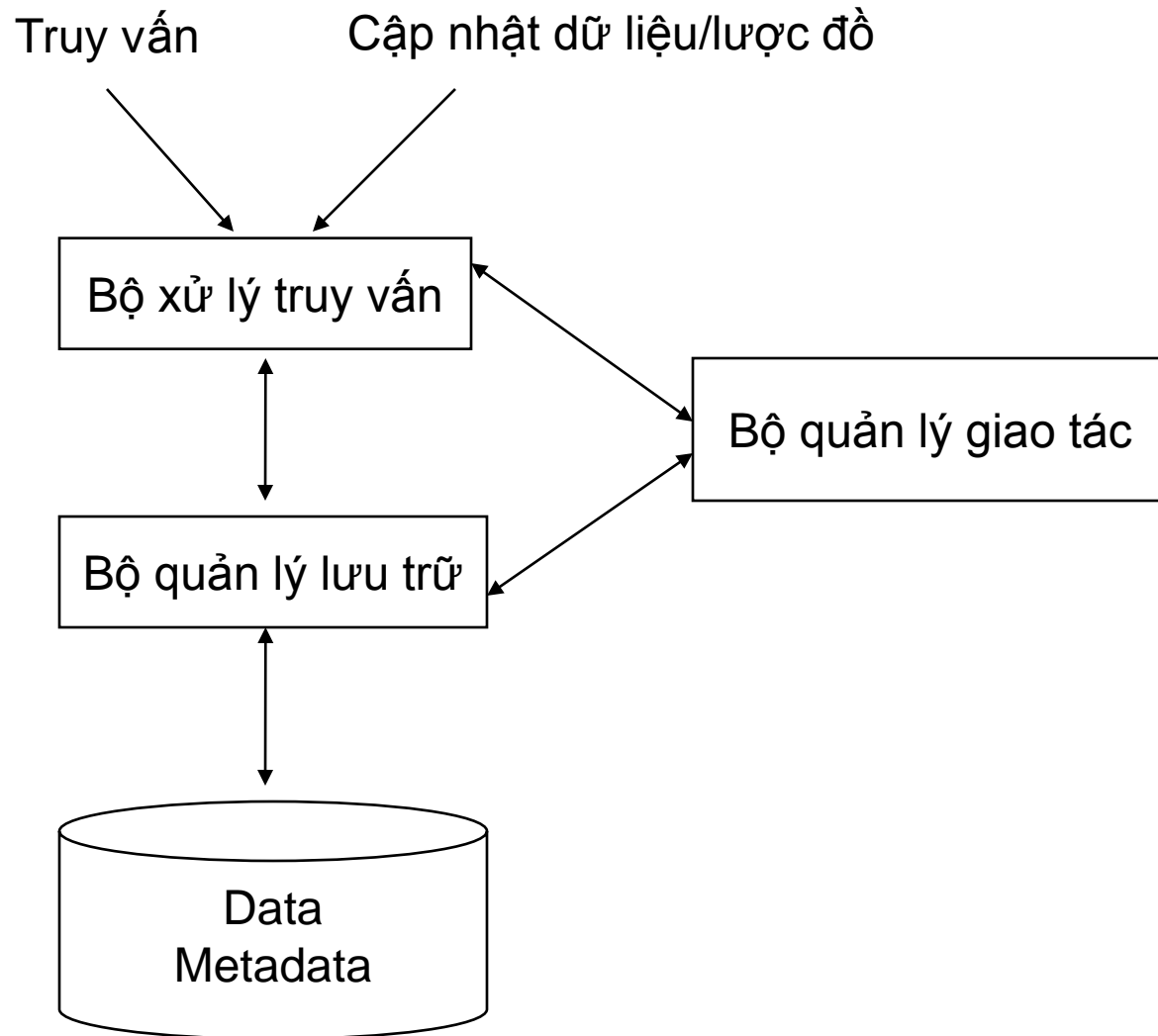
◆ Object-oriented

- Mở rộng từ ngôn ngữ lập trình như Java và ngôn ngữ văn tin

◆ Object-relational

- Mở rộng từ các HQTCSDL quan hệ hiện thời, thêm tính năng hướng đối tượng

Hệ quản trị CSDL – Kiến trúc của một DBMS



Hệ quản trị CSDL

■ Ưu điểm

- ◆ Điều khiển được sự dư thừa dữ liệu
- ◆ Cung cấp cấu trúc lưu trữ cho xử lý truy vấn một cách hiệu quả
- ◆ Ngăn cấm truy nhập trái phép của người dùng
- ◆ Điều khiển tương tranh
- ◆ Sao lưu và phục hồi dữ liệu.
- ◆ Buộc tuân thủ các ràng buộc toàn vẹn
- ◆ Giảm thời gian phát triển ứng dụng

Hệ quản trị CSDL

■ Nhược điểm

- ◆ Sự tập trung: Việc sử dụng cùng 1 chương trình, tại cùng một thời điểm bởi nhiều người dùng, đôi khi dẫn đến “mất mát” một số dữ liệu.
- ◆ Chi phí:
 - phần cứng và phần mềm
 - chuyển đổi dữ liệu (nếu phải chuyển từ hệ thống xử lý tệp sang hệ CSDL)
 - đào tạo nhân lực (Đòi hỏi kinh nghiệm!)
 - lập trình, nhập liệu, quản trị CSDL
- ◆ Sự phức tạp của sao lưu và phục hồi dữ liệu
- ◆

Lịch sử phát triển DBMS

1957 First commercial computer installed.

1961 - First generalised DBMS (Bachmann - GE's IDS)

1970 - Relational model - Ted Codd, IBM.

1970s ...

- ◆ General DB growth, theory development
- ◆ Custom commercial DMS (Digital, Honeywell, Burroghs)

1976 - ER model introduced by Chen.

1979 – First commercial version of SQL (Structured Query Language)

1980s ...

- ◆ Emergence of commercial RDBMS: Oracle, DB2, Ingres, Sybase, Informix....
- ◆ PC DBMS: DBASE, PARADOX

1985 - Preliminary SQL standard.

1990s ...

- ◆ Extension of RDBMS for new data types: Spatial, multimedia, ...

- ◆ OODBMS! ORDBMS...
- ◆ UML
- ◆ Improved performance; parallel machines.

1992

- ◆ SQLstandard: SQL2 (SQL92)
 - Most DBMS support this standard

1999

- ◆ SQLstandard: SQL3 (SQL99)
 - Most DBMS claim they will soon support this standard

Phân loại Hệ quản trị CSDL

■ Thế hệ thứ nhất:

- ◆ IMS (Information Management System) của hãng IBM [1960's]: HQTCSDL đầu tiên dựa trên mô hình phân cấp
- ◆ IDS (Intergrated Data Store) [Giữa những năm 60]: dựa trên mô hình mạng

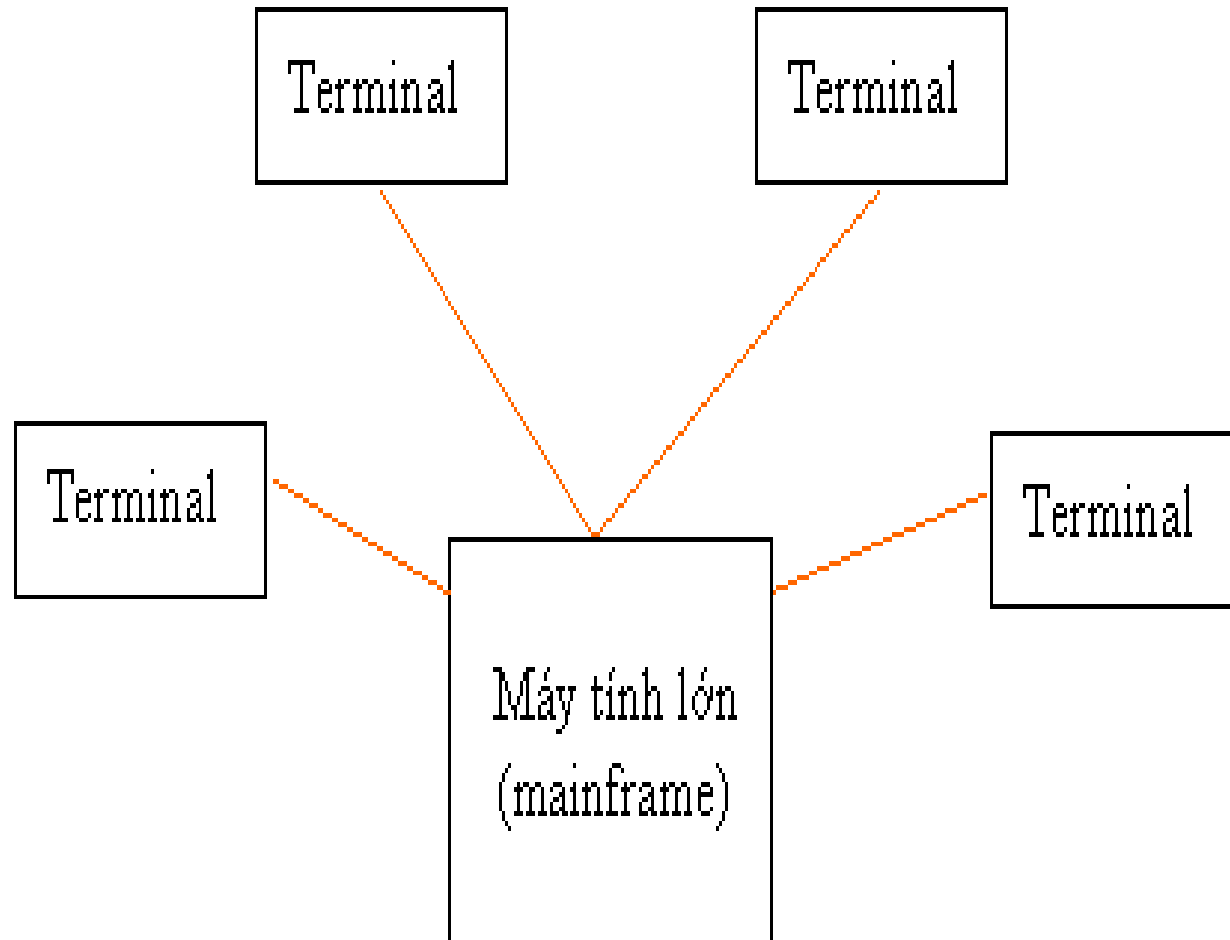
■ Thế hệ thứ hai:

- ◆ System-R của IBM [1976]: dựa trên mô hình quan hệ
- ◆ [Từ 1980 đến nay]: DB2 cho máy tính lớn của IBM, ORACLE của hãng ORACLE, DBase, Sybase, Informix, Foxpro, SQL Server, Access, Paradox

■ Thế hệ thứ ba:

- ◆ [Từ 1990]: Các HQTCSDL hướng đối tượng và hướng đối tượng quan hệ (vd: ORACLE)

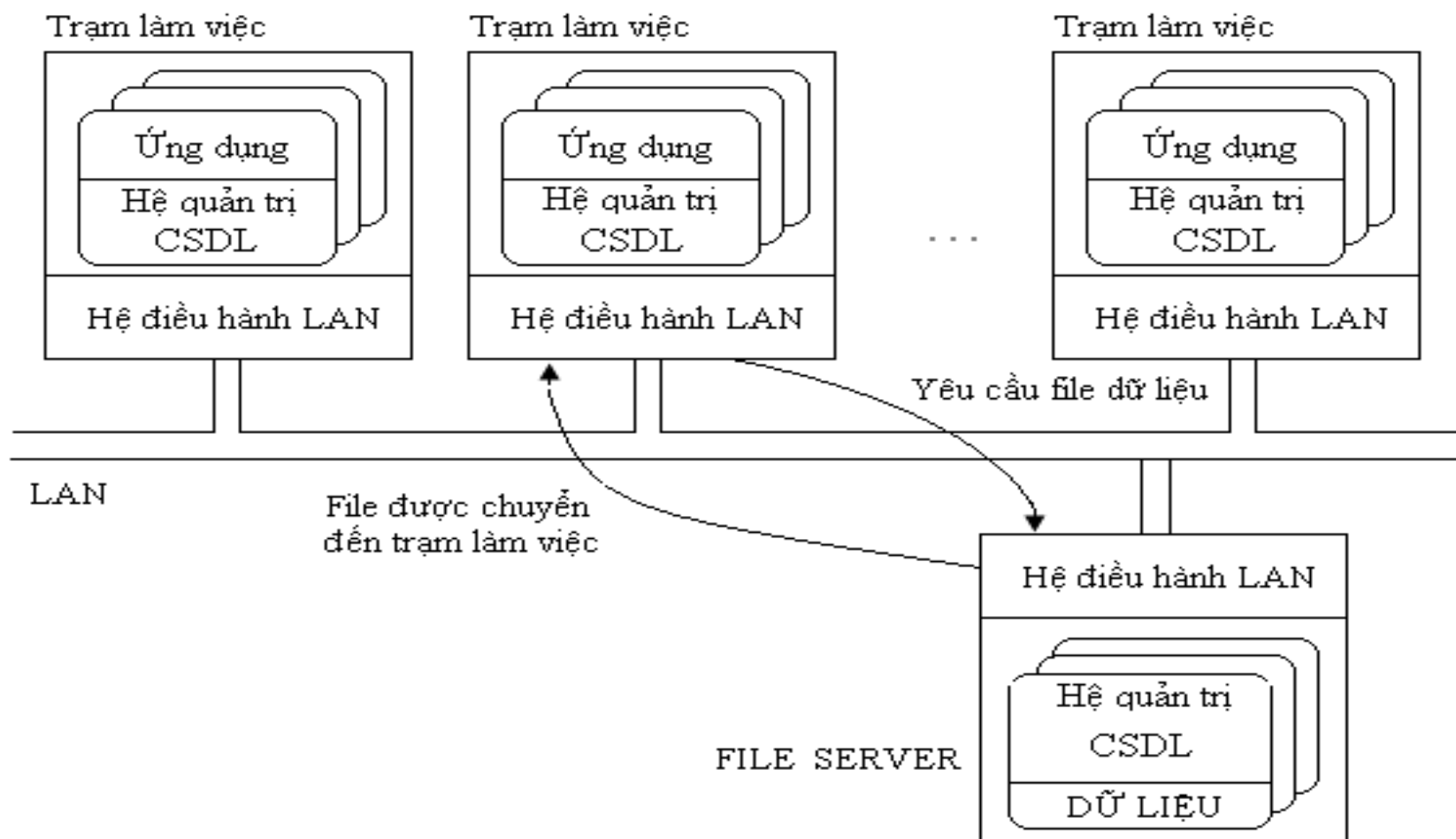
KIẾN TRÚC HQTCSDL ĐA NGƯỜI DÙNG



Kiến trúc hệ xử lý từ xa

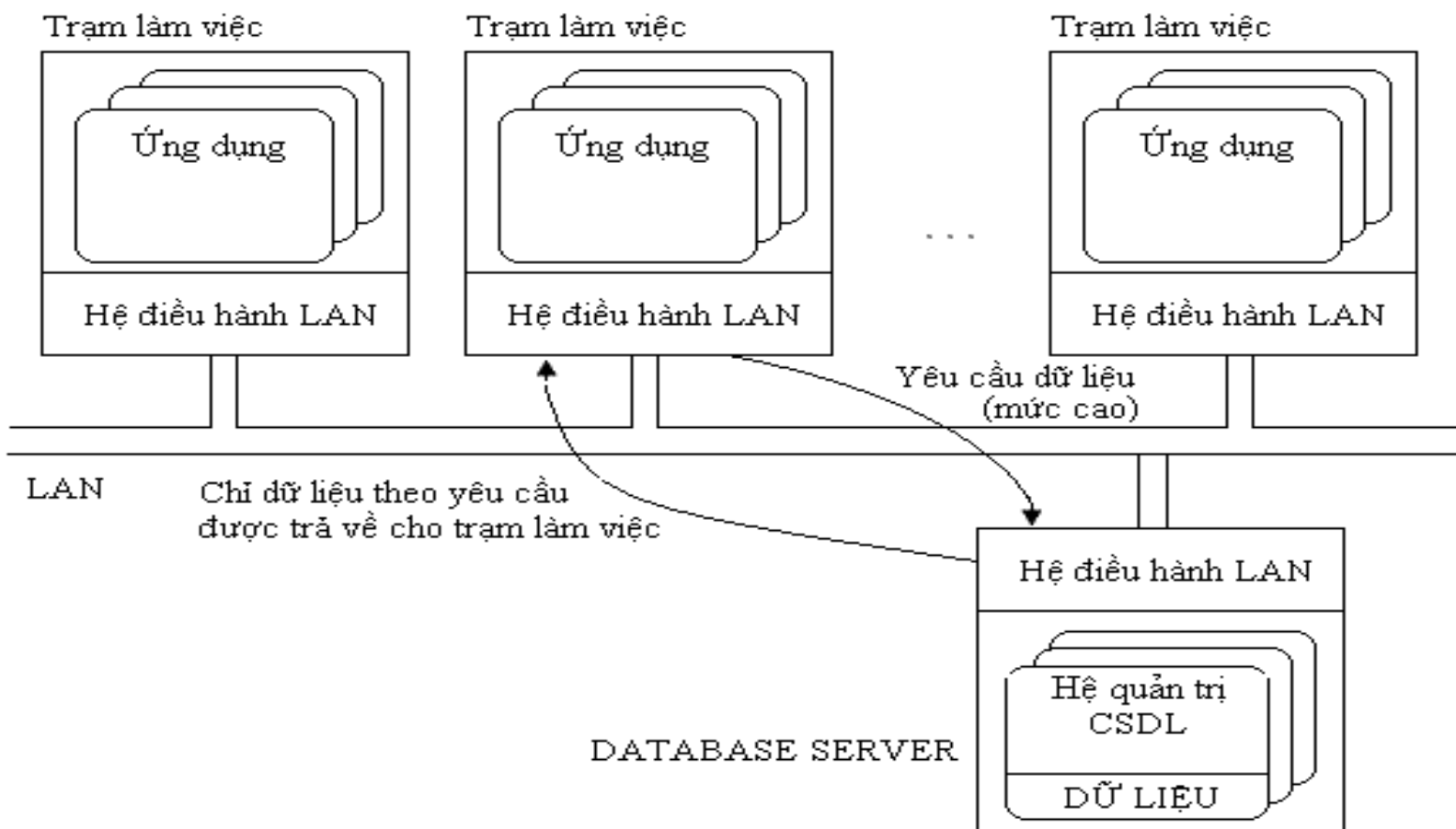
KIẾN TRÚC HQTCSDL ĐA NGƯỜI DÙNG

■ Kiến trúc Client-Server: File Server



KIẾN TRÚC HQTCSDL ĐA NGƯỜI DÙNG

■ Kiến trúc Client-Server: Database Server



VAI TRÒ CỦA CON NGƯỜI ĐỐI VỚI HỆ CSDL

■ Người quản trị CSDL (DBA – Database Administrator)

- ◆ Quản lý quyền truy nhập vào CSDL
- ◆ Điều phối và giám sát việc sử dụng CSDL
- ◆ Yêu cầu tài nguyên phần cứng và phần mềm

■ Người thiết kế CSDL (DB Designer)

- ◆ Xác định dữ liệu được lưu trong CSDL, các ràng buộc, các mối liên hệ giữa dữ liệu
- ◆ Lựa chọn cấu trúc phù hợp để biểu diễn dữ liệu và lưu dữ liệu này

■ Người sử dụng đầu cuối (End–User)

- ◆ Truy xuất CSDL để vấn tin, cập nhật, tạo báo cáo;

■ Kỹ sư phần mềm (Software Engineers)

- ◆ Gồm: Nhà phân tích hệ thống và lập trình viên ứng dụng
- ◆ Nhà phân tích hệ thống xác định yêu cầu của người dùng cuối; đặc tả các giao tác đáp ứng các yêu cầu trên.
- ◆ Lập trình viên: cài đặt các đặc tả trên thành chương trình.