

INT1313

Cơ Sở Dữ Liệu

Các Hệ quản trị CSDL

Giảng viên: Lê Hà Thanh
lehathanh@ptithcm.edu.vn

Nội dung

- Các khái niệm về Cơ sở dữ liệu (CSDL) (Database - DB)
- Các mô hình Dữ liệu
- Hệ quản trị CSDL (Database Management Systems – DBMS)

Định nghĩa (1/3) – liên quan đến dữ liệu

- **Dữ liệu** (data): các nội dung, sự kiện, văn bản, hình ảnh, âm thanh, phim, ảnh có ý nghĩa được ghi lại và lưu trữ trong hệ thống tính toán
- **Cơ sở Dữ liệu** (database): một tập hợp có tổ chức các dữ liệu có liên quan luận lý với nhau
 - Phản ánh/trình bày một số khía cạnh của thế giới thực
 - Được thiết kế, xây dựng và lưu giữ dữ liệu cho một mục đích cụ thể. Phục vụ cho một nhóm người dùng.
- **Thông tin** (information): dữ liệu đã được xử lý, hữu ích trong việc ra quyết định

Định nghĩa (2/3) – liên quan đến chương trình

- **Hệ quản trị Cơ sở Dữ liệu** (database management system - DBMS): hệ thống tính toán cho phép người dùng tạo dựng và duy trì CSDL.
 - Siêu dữ liệu (meta-data): định nghĩa dữ liệu và thông tin mô tả dữ liệu
 - Xây dựng CSDL: quá trình lưu trữ dữ liệu trên một số phương tiện lưu trữ được DBMS kiểm soát
 - Thao tác CSDL: truy vấn dữ liệu, cập nhật CSDL, tạo báo cáo
 - Chia sẻ CSDL: cho phép nhiều người dùng và chương trình truy cập CSDL cùng lúc
- **Chương trình Ứng dụng** (application program): truy cập CSDL qua việc gửi các câu truy vấn (queries) hoặc yêu cầu (requests) dữ liệu tới DBMS
 - Câu truy vấn (query) để lấy dữ liệu
 - Giao tác (transaction) thực hiện đọc/ghi dữ liệu vào CSDL

Định nghĩa (3/3) - bảo vệ và duy trì CSDL

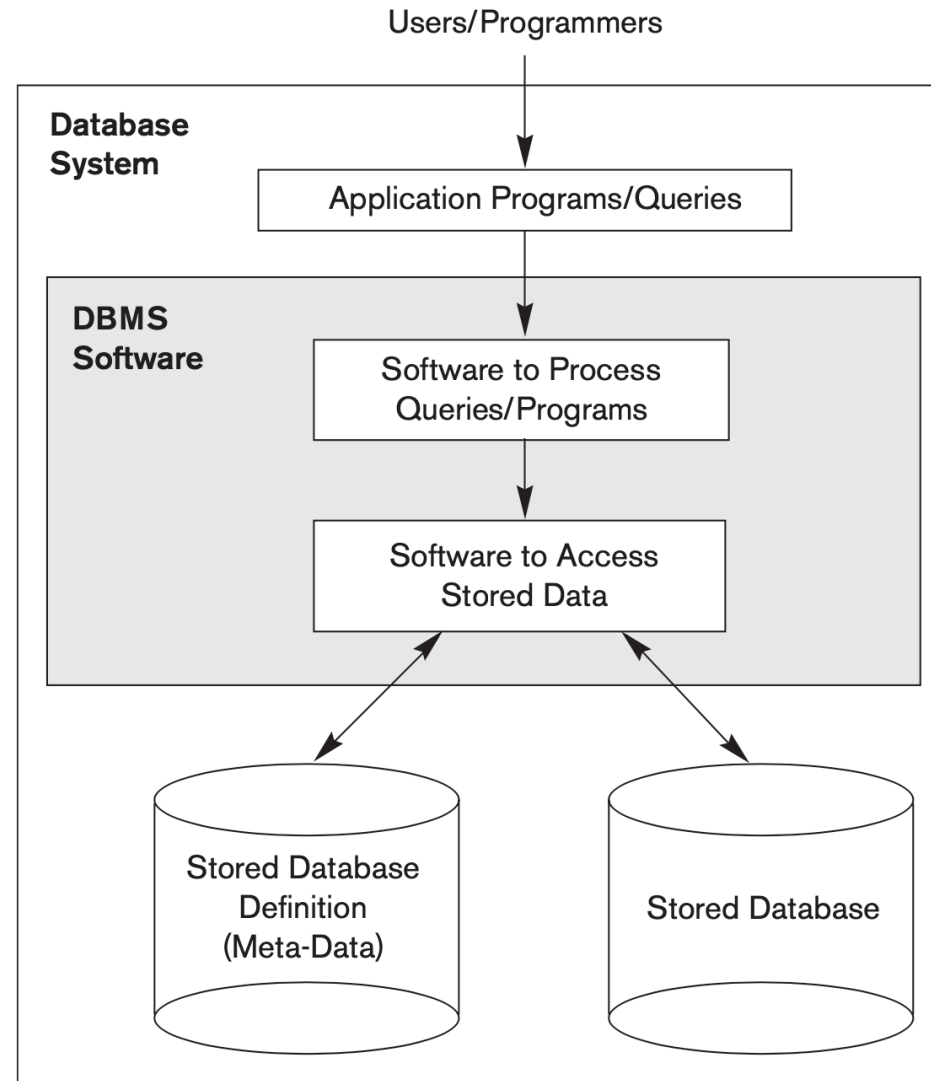
- **Bảo vệ CSDL:**

- Bảo vệ hệ thống khỏi các nguy cơ hư hỏng, sai sót trong phần cứng và phần mềm
- Bảo vệ an ninh chống lại các truy cập không được phép hoặc nhằm mục đích phá hoại

- **Duy trì CSDL:** CSDL thường có vòng đời rất dài

- DBMS phải có khả năng duy trì và bảo đảm hoạt động của hệ thống CSDL
- Thích ứng với thay đổi trong các yêu cầu đặt ra cho CSDL

Mô hình Hệ thống CSDL



- Đây là mô hình đơn giản để minh họa

Ví dụ Dữ liệu UNIVERSITY

- Dữ liệu được đặt trong ngữ cảnh cụ thể: môi trường một trường đại học
- Cơ sở dữ liệu UNIVERSITY
- Lưu trữ thông tin liên quan đến sinh viên, khóa học, bảng điểm

Ví dụ Dữ liệu UNIVERSITY

STUDENT

Name	Student_number	Class	Major
Smith	17	1	CS
Brown	8	2	CS

COURSE

Course_name	Course_number	Credit_hours	Department
Intro to Computer Science	CS1310	4	CS
Data Structures	CS3320	4	CS
Discrete Mathematics	MATH2410	3	MATH
Database	CS3380	3	CS

GRADE_REPORT

Student_number	Section_identifier	Grade
17	112	B
17	119	C
8	85	A
8	92	A
8	102	B
8	135	A

SECTION

Section_identifier	Course_number	Semester	Year	Instructor
85	MATH2410	Fall	07	King
92	CS1310	Fall	07	Anderson
102	CS3320	Spring	08	Knuth
112	MATH2410	Fall	08	Chang
119	CS1310	Fall	08	Anderson
135	CS3380	Fall	08	Stone

PREREQUISITE

Course_number	Prerequisite_number
CS3380	CS3320
CS3380	MATH2410
CS3320	CS1310

Ví dụ Dữ liệu UNIVERSITY – các chi tiết

- Bản ghi dữ liệu (data records)
- Thành phần dữ liệu (data elements)
- Kiểu dữ liệu (data type)
 - **Name** - Tên SV: string, gồm các ký tự alphabet
 - **Student_number** – mã SV: số nguyên
 - **Grade** – cấp đánh giá: ký tự, trong tập {'A', 'B', 'C', 'D', 'F', 'I'}
- Dùng lược đồ mã hóa (coding scheme) để biểu diễn các giá trị của một mục dữ liệu (data item)
 - Ví dụ: 1– freshman, 2 – sophomore, 3 – junior, 4 – senior, và 5 – graduate student

Thiết lập CSDL UNIVERSITY – Bản ghi và Quan hệ

- Lưu dữ liệu của một chủ thể thành một bản ghi dữ liệu trong từng tập tin riêng.
- Tồn tại các mối quan hệ (relationship) giữa các bản ghi ở các tập tin.
 - SV – điểm xếp hạng
 - Học phần – Điều kiện tiên quyết – Học phần (khác, là điều kiện tiên quyết)
- CSDL chứa nhiều kiểu bản ghi và nhiều quan hệ

Thiết lập CSDL UNIVERSITY – Thao tác với CSDL

- Trích xuất dữ liệu, thực hiện theo nhu cầu theo dõi thông tin cụ thể
 - Lấy bảng điểm {(học phần, điểm)} của 'Smith'
 - Xem danh sách SV tham gia nhóm học phần 'Database' được tổ chức vào mùa thu 2008; lấy kèm điểm xếp hạng của mỗi SV
 - Liệt kê các môn tiên quyết của học phần 'Database'
- Cập nhật dữ liệu, thay đổi nội dung dữ liệu
 - Đổi lớp của 'Smith' thành 'sophomore' – 'năm hai'
 - Tạo nhóm học mới cho 'Database' trong học kỳ này
 - Nhập xếp hạng 'A' cho 'Smith' trong nhóm học của 'Database' trong học kỳ vừa qua
- Phải được mô tả chính xác bằng ngôn ngữ truy vấn dữ liệu

Xây dựng CSDL trong Hệ thống Thông tin

- HTTT gồm các hệ thống máy tính, lưu trữ, phần mềm ứng dụng, CSDL
- Quy trình
 1. Xác định và phân tích yêu cầu
 - Tài liệu chi tiết các yêu cầu CSDL
 2. Thiết kế ở mức khái niệm
 - Mô tả CSDL bằng các mô hình khái niệm; sử dụng công cụ; định hướng cài đặt CSDL
 3. Thiết kế ở mức logic
 - Mô hình dữ liệu cài đặt được trong một DBMS cụ thể
 4. Thiết kế ở mức vật lý
 - Bổ sung các tiêu chí cài đặt cho việc lưu trữ và truy cập dữ liệu
 5. Cài đặt CSDL, nhập dữ liệu thực tế; duy trì liên tục

Đặc Tính của CSDL

- Bản chất tự mô tả của hệ thống cơ sở dữ liệu
- Cách ly giữa các chương trình và dữ liệu, và tính trừu tượng hóa dữ liệu
- Hỗ trợ nhiều chế độ xem dữ liệu
- Chia sẻ dữ liệu và xử lý giao dịch nhiều người dùng

Tính tự mô tả

- Dữ liệu
- Thông tin hoàn chỉnh các định nghĩa/mô tả (meta-data) được lưu trữ trong danh mục DBMS – DBMS catalog, gồm:
 - thông tin cấu trúc của từng tệp, loại và định dạng lưu trữ của từng mục dữ liệu,
 - nhiều ràng buộc khác nhau đối với dữ liệu.
- DBMS làm việc với nhiều CSDL
- Chú ý: NOSQL không yêu cầu thiết lập meta-data

Tính tự mô tả - Catalog

RELATIONS

Relation_name	No_of_columns
STUDENT	4
COURSE	4
SECTION	5
GRADE_REPORT	3
PREREQUISITE	2

- Major_type: kiểu liệt kê các thành phần thuộc các kiểu dữ liệu đã biết
- XXXXNNNN: định nghĩa chuỗi gồm 4 ký tự, và 4 ký số

COLUMNS

Column_name	Data_type	Belongs_to_relation
Name	Character (30)	STUDENT
Student_number	Character (4)	STUDENT
Class	Integer (1)	STUDENT
Major	Major_type	STUDENT
Course_name	Character (10)	COURSE
Course_number	XXXXNNNN	COURSE
....
....
....
Prerequisite_number	XXXXNNNN	PREREQUISITE

Tính cách ly, Trừu tượng của Dữ liệu

- Tính độc lập giữa chương trình – dữ liệu: các thay đổi trong cấu trúc (lưu trữ) dữ liệu không ảnh hưởng đến cách (lập trình) thiết lập kết nối CSDL và thao tác truy xuất dữ liệu của chương trình
- Tính độc lập giữa chương trình và cách thức hoạt động của CSDL

→ Trừu tượng hóa dữ liệu (**data abstraction**)

- Biểu diễn khái niệm dữ liệu trong DBMS bằng mô hình dữ liệu (**data model**)
 - Dùng các khái niệm logic mô tả đối tượng dữ liệu
 - Che các chi tiết lưu trữ, triển khai

Ví dụ: lưu trữ bản ghi SV - STUDENT

STUDENT

Name	Student_number	Class	Major
Smith	17	1	CS
Brown	8	2	CS

Data Item Name	Starting Position in Record	Length in Characters (bytes)
Name ←	1	30
Student_number ←	31	4
Class ←	35	1
Major ←	36	4

Quan sát Dữ liệu đa chiều

- Phụ thuộc nhu cầu của người dùng
- Là tập dữ liệu con
- Dữ liệu ảo (virtual data) được lấy ra từ CSDL nhưng không được lưu trữ tường minh
- Có nhiều ứng dụng truy xuất và sử dụng dữ liệu

Quan sát Dữ liệu đa chiều

TRANSCRIPT

(a)

Student_name	Student_transcript				
	Course_number	Grade	Semester	Year	Section_id
Smith	CS1310	C	Fall	08	119
	MATH2410	B	Fall	08	112
Brown	MATH2410	A	Fall	07	85
	CS1310	A	Fall	07	92
	CS3320	B	Spring	08	102
	CS3380	A	Fall	08	135

COURSE_PREREQUISITES

(b)

Course_name	Course_number	Prerequisites
Database	CS3380	CS3320
		MATH2410
Data Structures	CS3320	CS1310

Chia sẻ Dữ liệu và Xử lý Giao dịch đa người dùng

- Cho phép nhiều người dùng truy xuất CSDL (bằng các ứng dụng khác nhau) tại cùng một thời điểm
- Tích hợp và duy dữ liệu trong một CSDL duy nhất
- Cơ chế Điều khiển và kiểm soát đồng hành (concurrency) để truy xuất và cập nhật dữ liệu đúng
 - Tại một thời điểm chỉ có một truy xuất đến đối tượng dữ liệu
- OLTP – Online Transaction Processing
- Giao tác (transaction): ACID

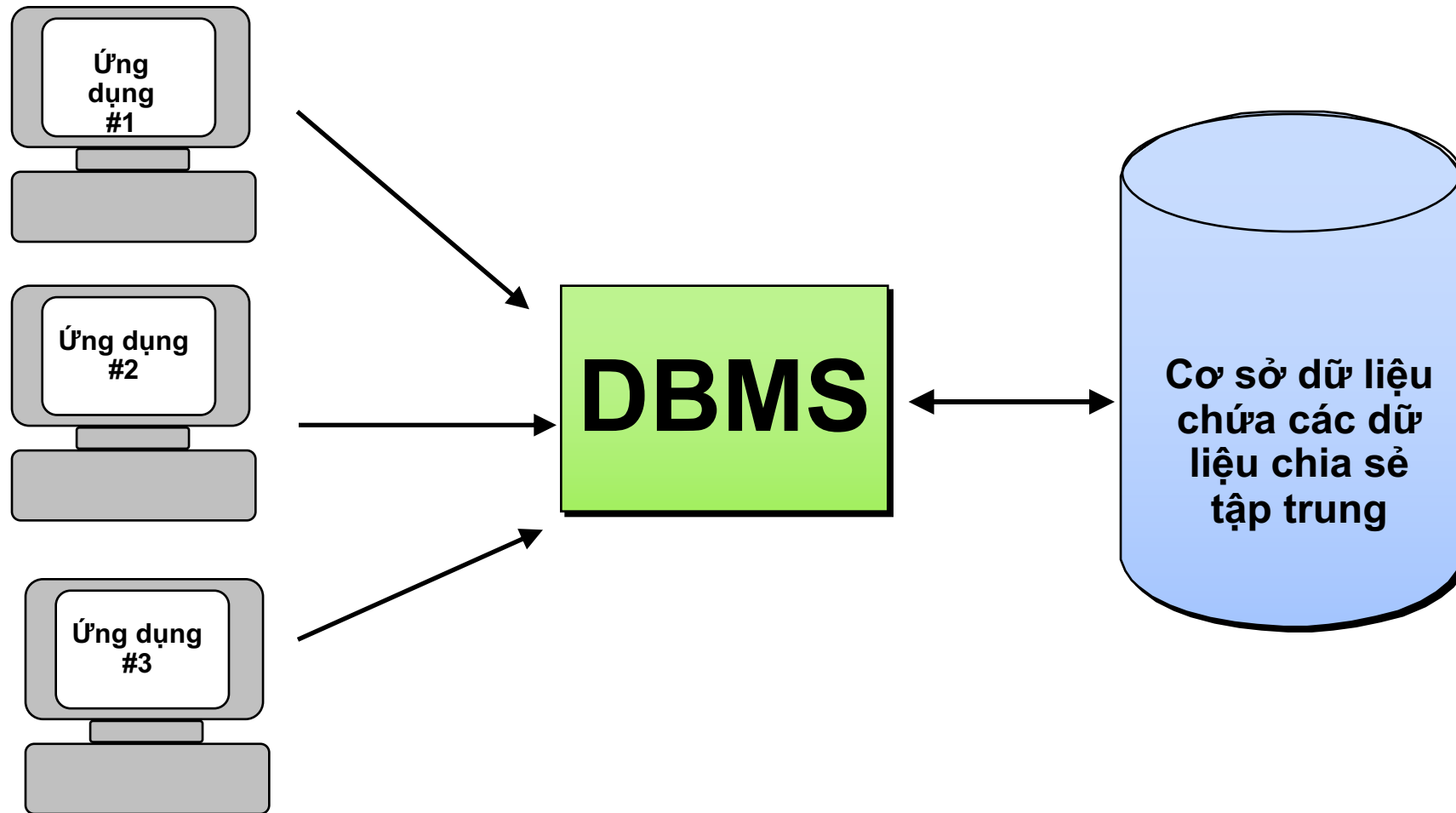
Tại sao phải có DBMS?

- CSDL đặt ra những vấn đề cần phải giải quyết:
 - Tính chủ quyền
 - Cơ chế bảo mật hay phân quyền khi khai thác CSDL
 - Giải quyết tranh chấp trong quá trình truy nhập dữ liệu
 - Phục hồi dữ liệu khi xảy ra sự cố
- Phần mềm có khả năng giải quyết các vấn đề trên được gọi là hệ quản trị CSDL (DBMS)

DBMS

- Cung cấp các công cụ hỗ trợ tích cực cho các nhà phân tích, thiết kế và người khai thác CSDL
- Trợ giúp cho quá trình định nghĩa, xây dựng và xử lý CSDL cho các ứng dụng khác nhau
- MS SQL Server, Oracle, MySQL, PostgreSQL, Sybase, DB-2, MongoDB, ...

DBMS



Lợi ích của cách tiếp cận CSDL

- Sự độc lập chương trình với dữ liệu
 - Siêu dữ liệu được lưu trữ trong DBMS vì thế các ứng dụng không gặp khó khăn với dạng thức dữ liệu
 - Việc truy vấn/cập nhật dữ liệu được quản lý bởi DBMS vì thế chương trình không cần phải có các quy trình xử lý dữ liệu
 - Kết quả: tăng hiệu quả phát triển và bảo trì ứng dụng
- Giảm thiểu sự dư thừa dữ liệu
 - Tăng cường sự bền vững/tích hợp dữ liệu

Lợi ích của cách tiếp cận Cơ sở dữ liệu (tt)

- Tăng cường khả năng chia sẻ dữ liệu
- Bắt buộc phải có tính chuẩn hoá
- Nâng cao chất lượng dữ liệu
- Khả năng truy xuất và đáp ứng dữ liệu tốt hơn
- Sự bảo mật, sao lưu/phục hồi.

Phí tổn và rủi ro của cách tiếp cận CSDL

- Phí tổn ban đầu
 - Phí tổn cài đặt ban đầu cao, phức tạp
 - Chuyển đổi dữ liệu từ hệ thống cũ sang hệ thống CSDL mới
- Phí tổn vận hành
 - Yêu cầu nhân lực chuyên môn mới
 - Sao lưu và phục hồi
- Xung đột trong tổ chức
 - Sự khó khăn khi phải thay đổi thói quen, lề lối làm việc cũ

Mô hình Dữ liệu,
Lược đồ
Thực thể hóa Dữ liệu

Mô hình Dữ liệu

- Mô hình dữ liệu khái niệm
 - Mô hình thực thể - quan hệ (entity-relationship ER)
- Mô hình dữ liệu vật lý
- Mô hình dữ liệu biểu diễn (mô hình triển khai)
 - Mô hình dữ liệu quan hệ

Lược đồ, Thực thể và Trạng thái CSDL

- Lược đồ CSDL: mô tả CSDL
 - Dùng trong giai đoạn thiết kế CSDL, ít thay đổi
- Được biểu diễn bằng các sơ đồ lược đồ
- Dữ liệu thay đổi theo thời điểm
 - Trạng thái dữ liệu
- Trạng thái dữ liệu:
 - Khởi tạo với dữ liệu ban đầu,
 - Trạng thái mới mỗi khi có cập nhật dữ liệu

Sơ đồ Lược đồ

- Đối tượng trong lược đồ:
Cấu trúc lược đồ
- Chỉ mô tả một số khía cạnh nhất định của lược đồ, một số kiểu ràng buộc.
- Ví dụ:
 - Không có mô tả kiểu dữ liệu
 - Không mô tả quan hệ giữa các tập tin lưu trữ dữ liệu

STUDENT

Name	Student_number	Class	Major
------	----------------	-------	-------

COURSE

Course_name	Course_number	Credit_hours	Department
-------------	---------------	--------------	------------

PREREQUISITE

Course_number	Prerequisite_number
---------------	---------------------

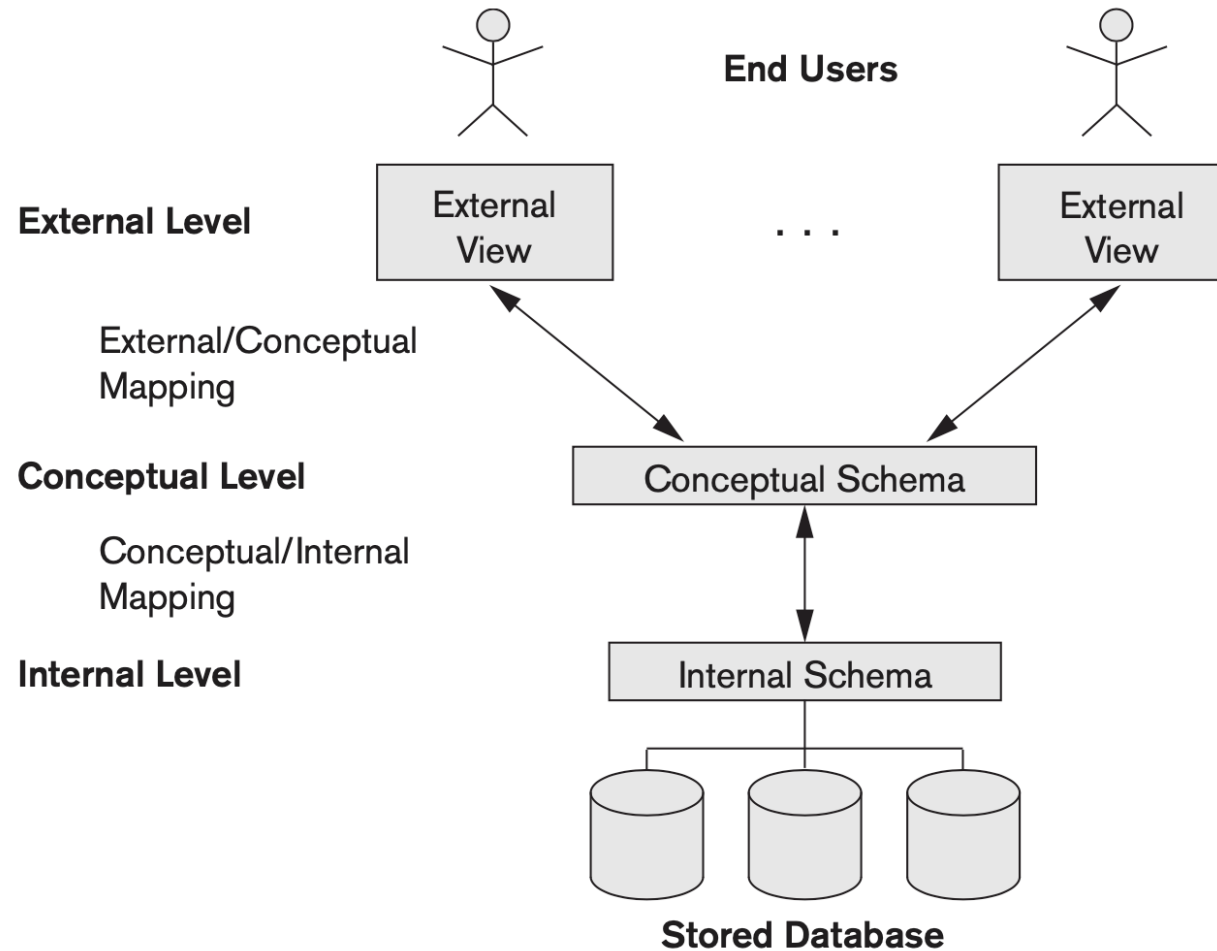
SECTION

Section_identifier	Course_number	Semester	Year	Instructor
--------------------	---------------	----------	------	------------

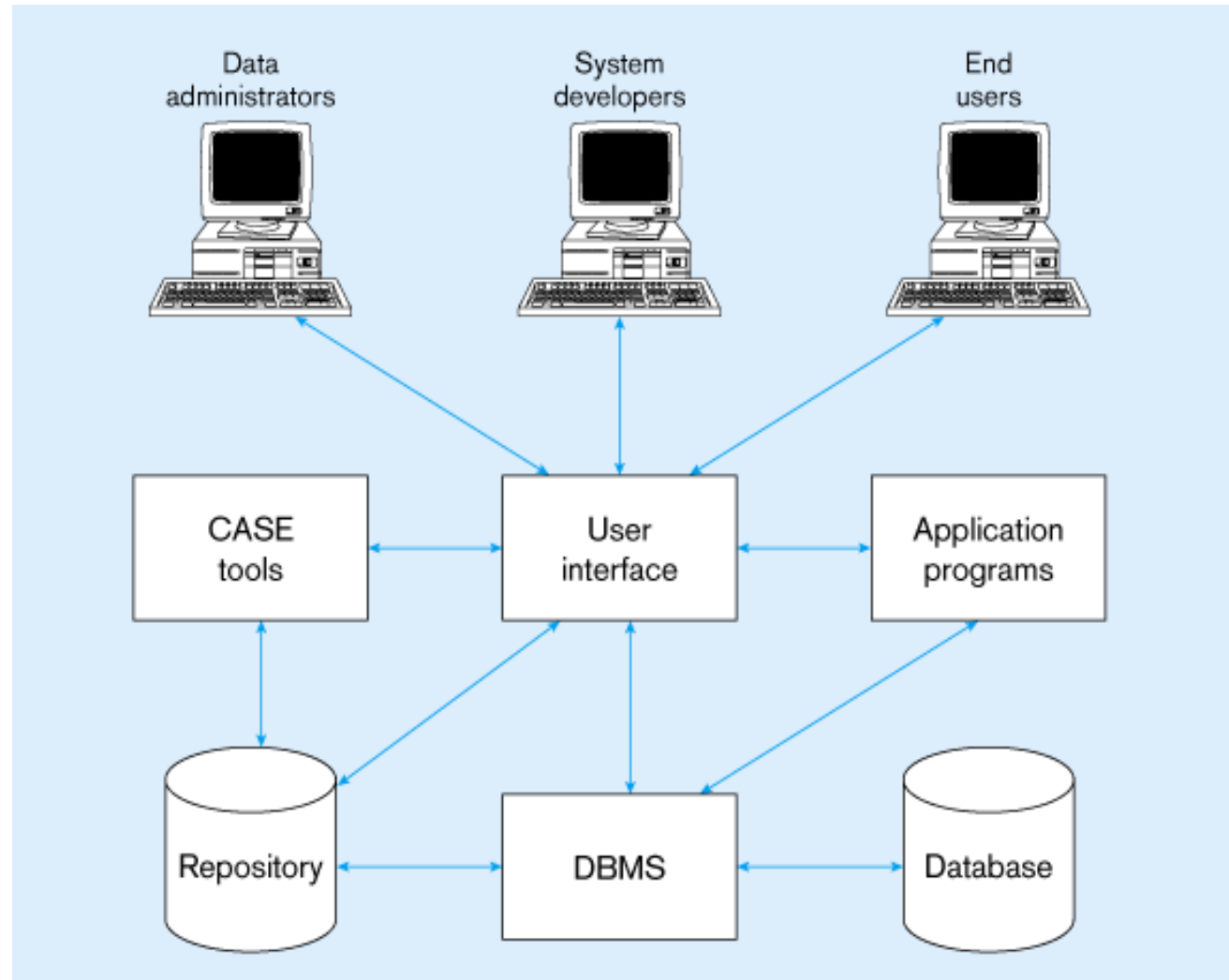
GRADE_REPORT

Student_number	Section_identifier	Grade
----------------	--------------------	-------

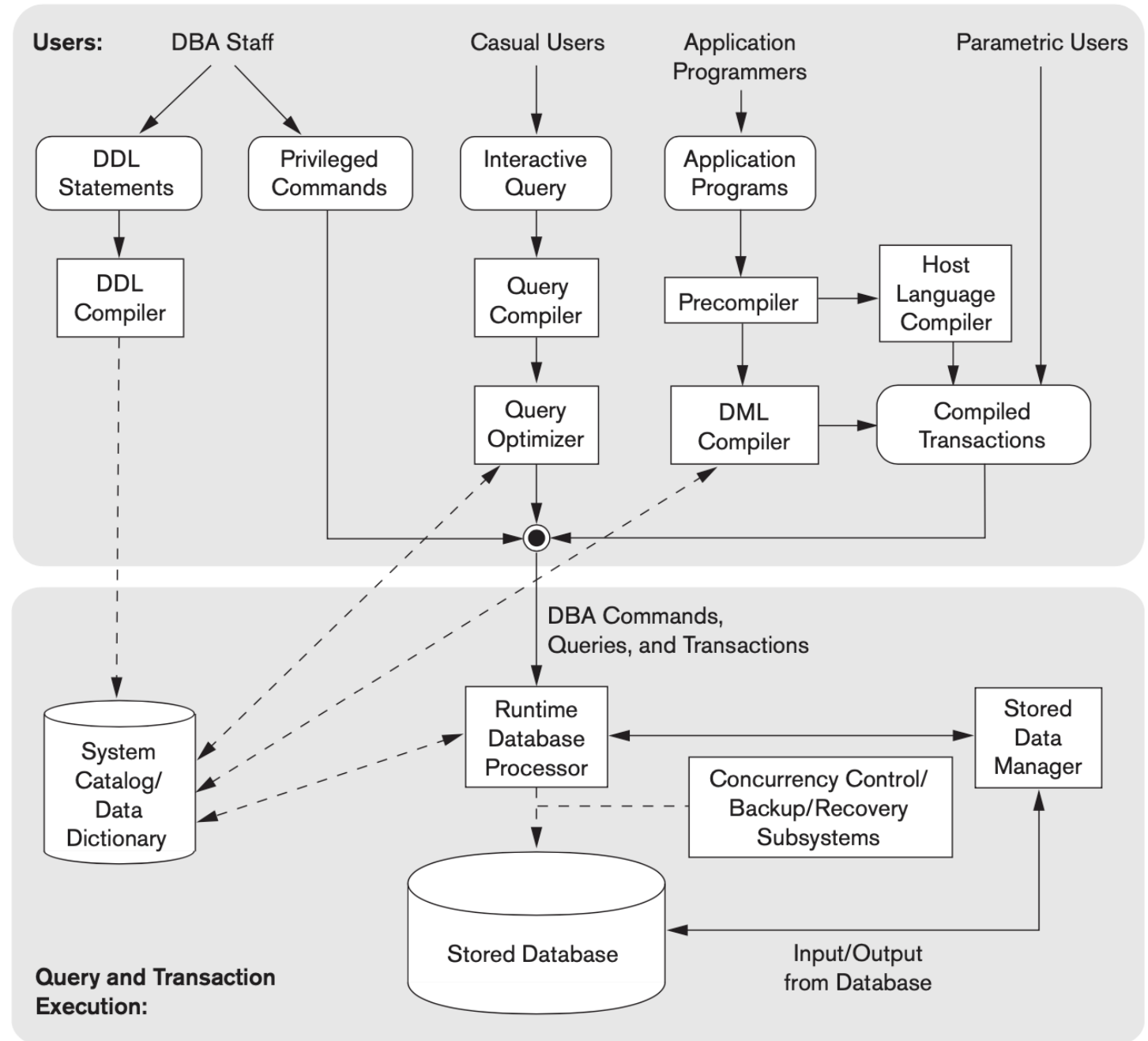
Kiến trúc 3-lược đồ dữ liệu



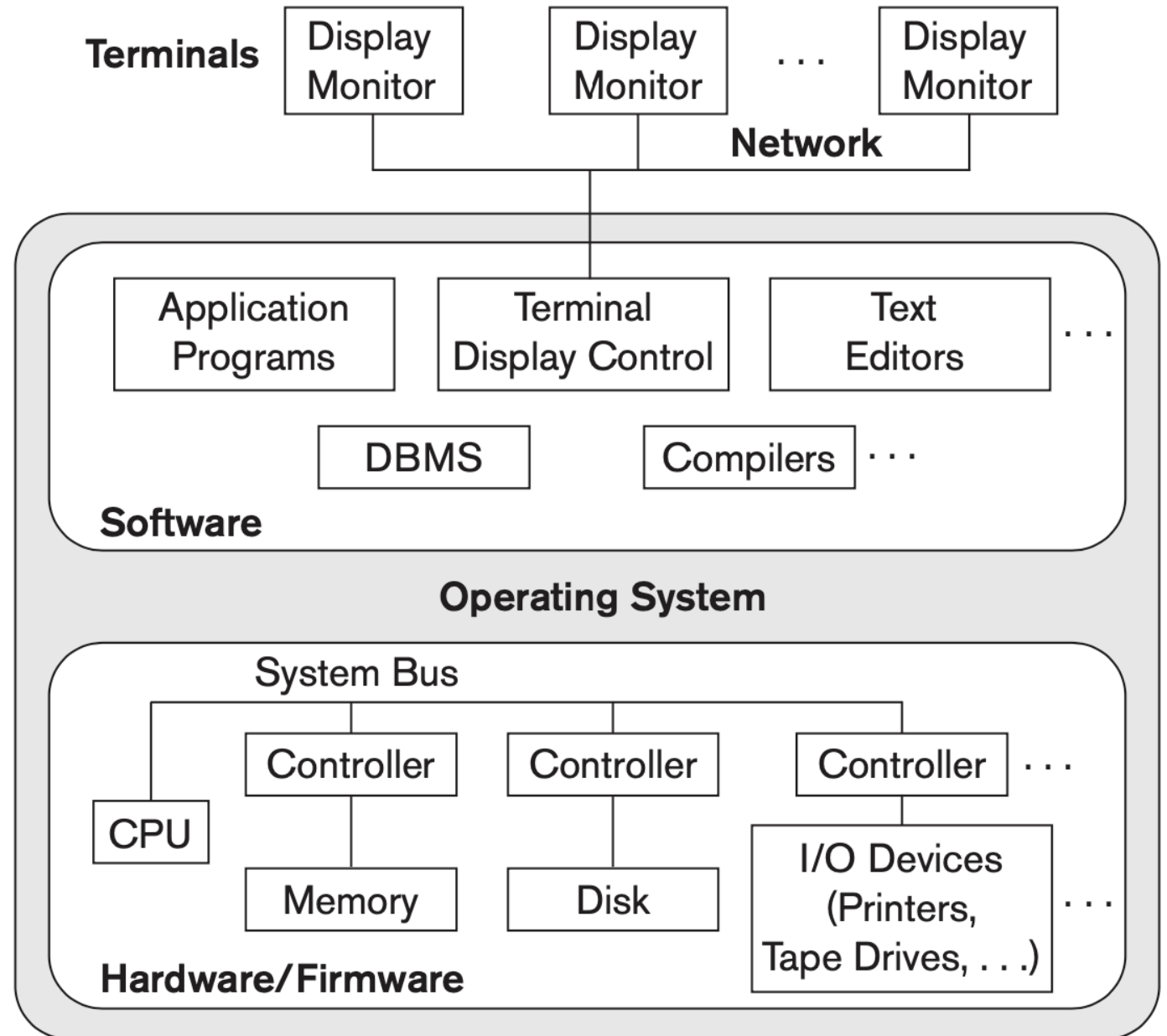
Các thành phần của môi trường CSDL



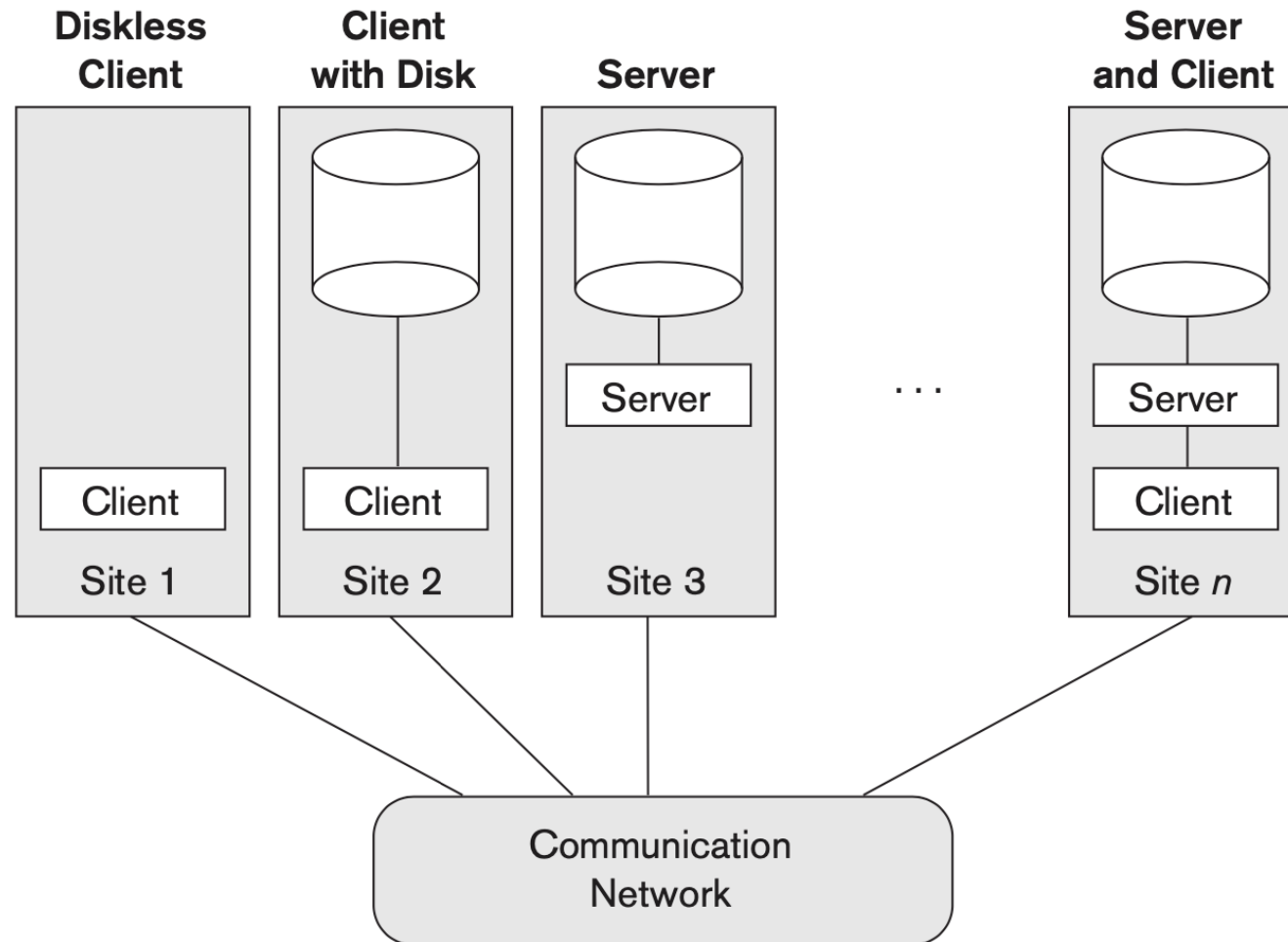
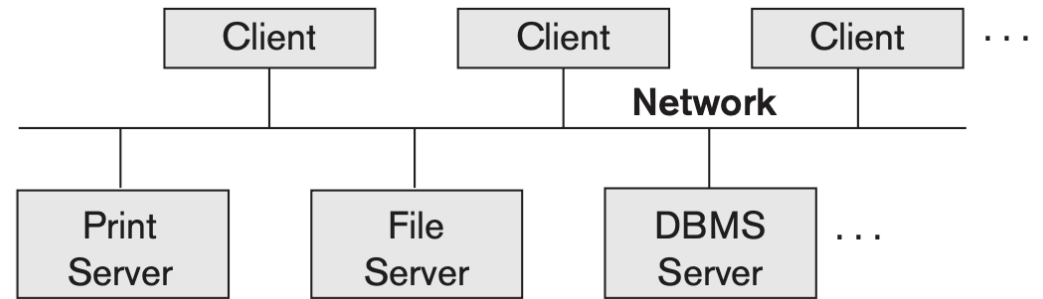
Các Modules thành phần của DBMS



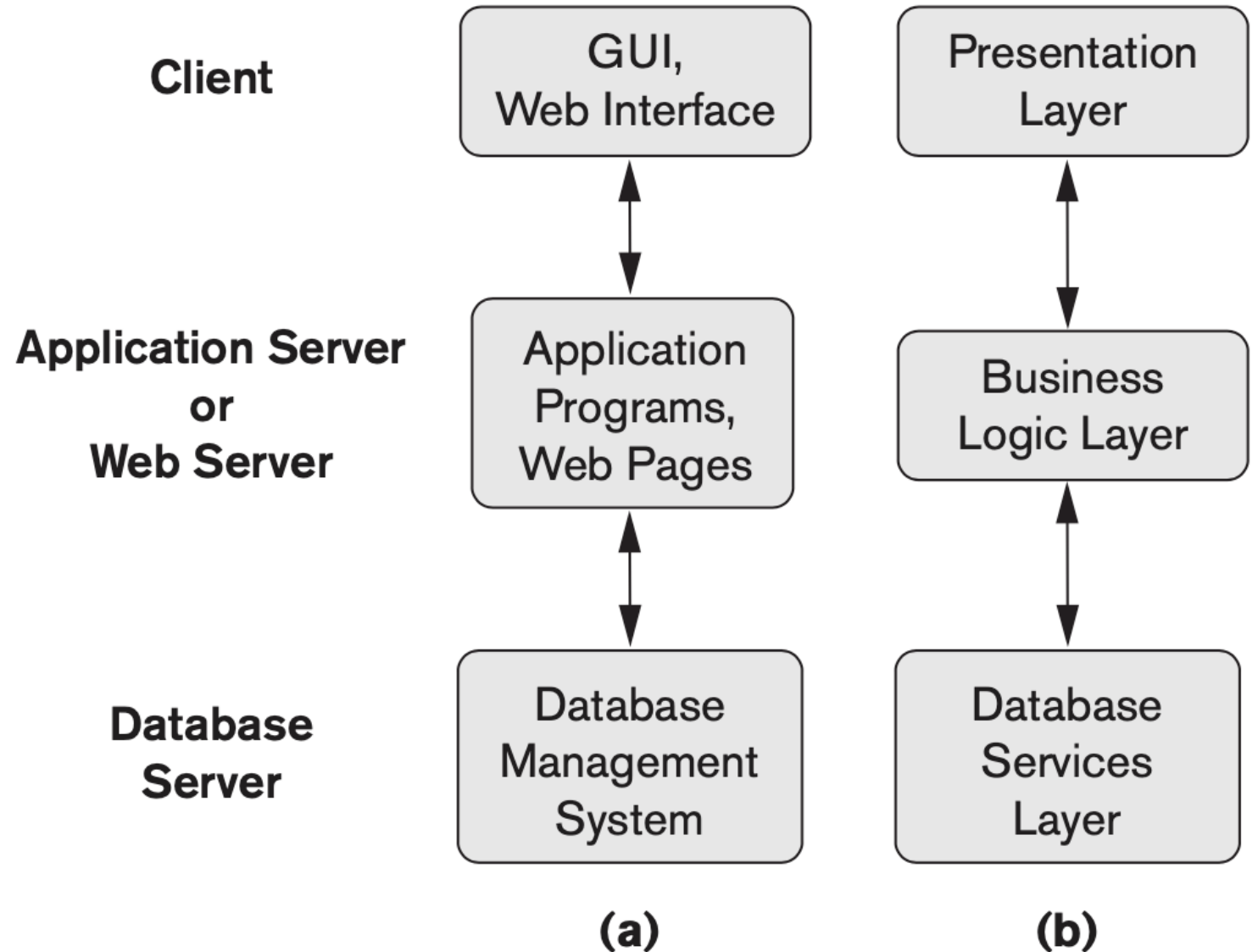
Kiến trúc Vật lý của DBMS



Kiến trúc Client/Server 2-lớp



Kiến trúc Client/Server 3-lớp



Phân loại DBMS

- Dựa trên mô hình dữ liệu được dùng:
 - Truyền thống: quan hệ, mạng, phân cấp
 - Đang thịnh hành: hướng đối tượng, quan hệ đối tượng
- Các sự phân loại khác:
 - Single-user và Multi-user
 - Tập trung và Phân tán