

Bài 01:

TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU



Nội dung

1. Dữ liệu là gì
2. Định nghĩa một CSDL
3. Các đối tượng sử dụng CSDL
4. Hệ quản trị CSDL
- 5 Kiến trúc của một hệ CSDL

Nội dung

1. Dữ liệu là gì
2. Định nghĩa một CSDL
3. Các đối tượng sử dụng CSDL
4. Hệ quản trị CSDL
- 5 Kiến trúc của một hệ CSDL

What is Database



DBMS

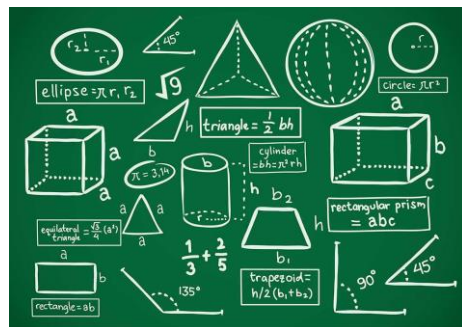
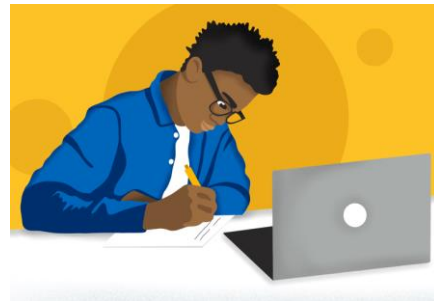
Relational Database

Non-Relational Database



1. Giới thiệu

Dữ liệu (Data) là gì?

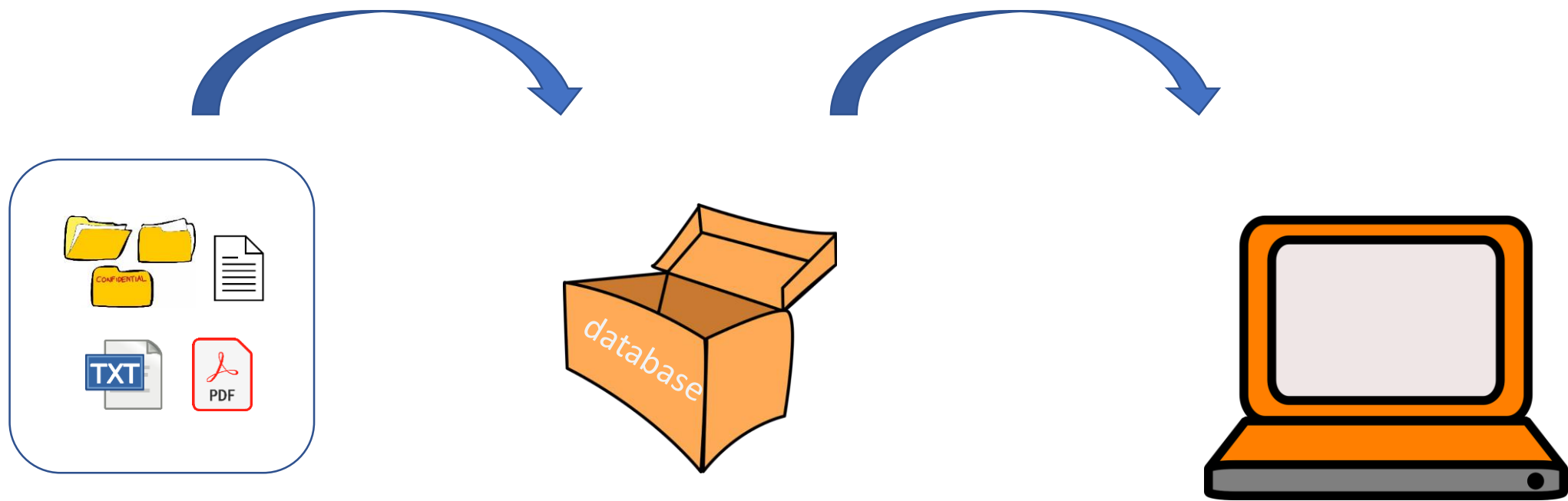


- ✓ Dữ liệu là những gì chúng ta biết có thể ghi lại được và có ý nghĩa (By data, we mean known facts that can be recorded and that have implicit meaning - Elmasri, Navathe).
- ✓ Ví dụ:
 - Dữ liệu về nhân viên, quá trình công tác và chuyên môn trong một công ty
 - Dữ liệu về thời tiết, khí hậu
 - Dữ liệu về công dân của một quốc gia

2. Hệ thống tập tin

- ✓ Là tập hợp các tập tin riêng lẻ phục vụ cho một mục đích của đơn vị sử dụng.
- ✓ Ưu điểm:
 - Triển khai ứng dụng nhanh
 - Khả năng đáp ứng nhanh chóng, kịp thời (vì chỉ phục vụ cho một mục đích hạn hẹp).
- ✓ Nhược điểm:
 - Sự trùng lặp dữ liệu giữa các ứng dụng gây tốn không gian lưu trữ.
 - Dữ liệu tách biệt và chia sẻ dữ liệu kém.
 - Phụ thuộc dữ liệu.
 - Định dạng tập tin không tương thích.

2. Cơ sở dữ liệu



2. Cơ sở dữ liệu

✓ Định nghĩa:

- Cơ sở dữ liệu là một **hệ thống các thông tin có cấu trúc**, được lưu trữ trên các thiết bị lưu trữ thông tin thứ cấp nhằm thỏa mãn nhu cầu **khai thác thông tin đồng thời** của nhiều người sử dụng hay nhiều chương trình ứng dụng với những mục đích khác nhau.

✓ Ví dụ:

- CSDL nhân viên
- CSDL hàng hóa
- CSDL sinh viên, ...

2. Cơ sở dữ liệu

Số CMND	Họ tên	Ngày sinh	Giới tính
1	Đỗ Nam Trung	1990-06-07	Nam
2	Nguyễn Đông Anh	1991-02-10	Nam
3	Trần Thu Ngân	1995-10-17	Nữ
4	Hoàng Vũ	1980-01-01	Nam
5	Lê Quỳnh Chi	2000-10-10	Nữ

Biển số	Số máy	Số khung	Số chỗ	Nhãn hiệu	Màu sơn
BS1	1	3	5	Honda	Đen
BS2	5	9	5	Mazda	Đỏ
BS3	3	2	2	Hyundai	Xám
BS4	10	7	7	Toyota	Vàng

Số CMND	Biển số	Ngày sở hữu
1	BS1	2020-01-01
1	BS4	2019-06-09
3	BS3	2016-07-09
2	BS2	2018-01-20
5	BS3	2016-07-09

2. Cơ sở dữ liệu



✓ Lưu ý:

- Trước hết, CSDL phải là một **tập hợp các thông tin mang tính hệ thống** chứ không phải là các thông tin rời rạc, không có mối quan hệ với nhau.
- Các thông tin này **phải có cấu trúc** và tập hợp các thông tin này phải có khả năng đáp ứng các nhu cầu khai thác của nhiều người sử dụng một cách đồng thời.



- ✓ Một CSDL biểu thị một khía cạnh nào đó của thế giới thực. Thông tin được đưa vào trong CSDL tạo thành một không gian CSDL hoặc một thế giới nhỏ (miniworld)
- ✓ Một cơ sở dữ liệu được thiết kế và được phổ biến cho một mục đích riêng
- ✓ Một cơ sở dữ liệu là một tập hợp dữ liệu liên kết với nhau một cách logic và mang một ý nghĩa nào đó.
- ✓ Một CSDL không phải là một tập hợp dữ liệu tùy tiện.



2. Cơ sở dữ liệu

✓ Ưu điểm:

- Giảm sự trùng lặp thông tin xuống mức thấp nhất và do đó bảo đảm được tính nhất quán và toàn vẹn dữ liệu
- Đảm bảo dữ liệu có thể được truy xuất theo nhiều cách khác nhau
- Khả năng chia sẻ thông tin cho nhiều người sử dụng và nhiều ứng dụng khác nhau

2. Cơ sở dữ liệu

Vấn đề cần giải quyết:

1. Tính chủ quyền của dữ liệu

- Dữ liệu chia sẻ → người có chức năng thay đổi thường xuyên
- Để có dữ liệu mới nhất phải cập nhật thường xuyên

2. Tính bảo mật và quyền khai thác thông tin của người dùng

- Dữ liệu được bảo vệ
- Người dùng chỉ truy xuất thông tin mình có quyền

3. Tranh chấp dữ liệu

- Dữ liệu được thêm vào đồng thời
- Dữ liệu được sửa đồng thời

4. Đảm bảo dữ liệu khi có sự cố

- Đang thực hiện thao tác thì mất điện
- ...

3. Các đối tượng sử dụng CSDL



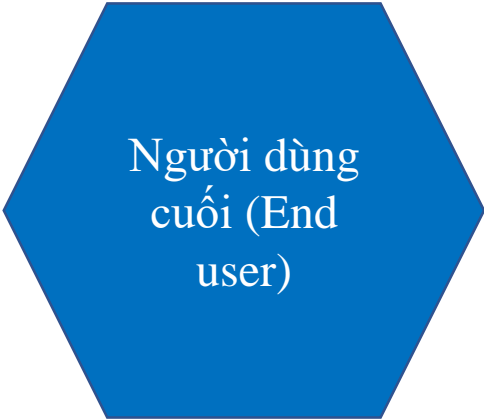
Người quản trị
CSDL
(Database
Administrator)

- ✓ Tạo lập CSDL
- ✓ Quản lý CSDL:
 - Cấp quyền truy cập
 - Điều phối việc sử dụng
 - Giám sát việc sử dụng
- ✓ Tính an toàn và khả năng đáp ứng của CSDL



Người thiết kế
CSDL
(Database
Designer)

- ✓ Quyết định những dữ liệu nào cần được lưu trữ
- ✓ Lựa chọn cấu trúc phù hợp để lưu trữ dữ liệu
- ✓ Xây dựng ứng dụng kết nối CSDL
- ✓ Trao đổi với người dùng để nắm bắt được những yêu cầu và đưa ra một thiết kế CSDL thỏa yêu cầu này.



Người dùng
cuối (End
user)

- ✓ Sử dụng những công cụ hỗ trợ để khai thác
- ✓ Ứng dụng này do chuyên viên tin học tạo ra

4. Hệ quản trị CSDL

Định nghĩa:

- ✓ Database Management System (DBMS)
- ✓ Là tập hợp các chương trình dùng để quản lý cấu trúc và dữ liệu của CSDL đồng thời điều khiển việc truy xuất dữ liệu trong CSDL
- ✓ Cung cấp cho người dùng và ứng dụng một môi trường thuận tiện, sử dụng hiệu quả tài nguyên dữ liệu

4. Hệ quản trị CSDL



ORACLE



dBASE™

4. Hệ quản trị CSDL

Các thành phần của hệ CSDL:

1. Ngôn ngữ giao tiếp giữa người sử dụng (NSD) và CSDL

- Ngôn ngữ mô tả dữ liệu (Data Definition Language - DDL): cho phép khai báo cấu trúc CSDL, các mối liên hệ của dữ liệu, các quy định, ràng buộc dữ liệu.
- Ngôn ngữ thao tác dữ liệu (Data Manipulation Language - DML): cho phép thực hiện thao tác thêm, xóa, sửa dữ liệu.
- Ngôn ngữ truy vấn dữ liệu, hay ngôn ngữ hỏi đáp có cấu trúc (Structured Query Language - SQL): cho phép người khai thác sử dụng để truy vấn thông tin cần thiết
- Ngôn ngữ điều khiển dữ liệu (Data Control Language - DCL): cho phép thay đổi cấu trúc bảng, khai báo bảo mật, cấp quyền cho người dùng.

4. Hệ quản trị CSDL



2. Từ điển dữ liệu (Data Dictionary)

- Dùng để mô tả các ánh xạ liên kết

3. Ghi nhận các thành phần cấu trúc của CSDL, các chương trình ứng dụng, mật mã, quyền hạn sử dụng v.v...

4. Cơ chế giải quyết vấn đề tranh chấp dữ liệu

- Cấp quyền ưu tiên cho từng người sử dụng (người quản trị CSDL thực hiện)
- Đánh dấu yêu cầu truy xuất dữ liệu, phân chia thời gian, người nào có yêu cầu trước thì có quyền truy xuất dữ liệu trước



4. Hệ quản trị CSDL

5. Cơ chế sao lưu (Backup) và phục hồi (Restore) dữ liệu khi có sự cố xảy ra.

- Định kỳ kiểm tra CSDL, sau một thời gian nhất định hệ quản trị CSDL sẽ tự động tạo ra một bản sao CSDL. Cách này hơi tốn kém, nhất là đối với các CSDL lớn
- Tạo nhật ký (LOG) thao tác CSDL. Mỗi thao tác trên CSDL đều được hệ thống ghi lại, khi có sự cố xảy ra thì tự động lần ngược lại (RollBack) để phục hồi CSDL

4. Hệ quản trị CSDL

6. Hệ quản trị CSDL phải cung cấp một giao diện (Interface) tốt, dễ sử dụng, dễ hiểu cho những người sử dụng không chuyên
7. **Đảm bảo tính độc lập giữa dữ liệu và chương trình**
 - Khi có sự thay đổi dữ liệu (như sửa đổi cấu lưu trữ các bảng dữ liệu, thêm các chỉ mục (Index) ...) thì các chương trình ứng dụng (Application) đang chạy trên CSDL đó vẫn không cần phải được viết lại, hay cũng không làm ảnh hưởng đến những NSD khác.

4. Hệ quản trị CSDL

Đặc trưng theo hướng CSDL

a. Tính tự mô tả

b. Tính độc lập

c. Tính trừu tượng

d. Tính nhất quán

e. Đa khung hình

4. Hệ quản trị CSDL

a. Tính tự mô tả

- ✓ Hệ Cơ sở dữ liệu:
 - Chứa dữ liệu của CSDL
 - Chứa Mô tả (cấu trúc + ràng buộc) của CSDL
- ✓ Thông tin mô tả được lưu trong Catalog
- ✓ Thông tin mô tả = meta data
- ✓ Thông tin lưu trong catalog được sử dụng bởi :
 - Hệ quản trị CSDL
 - Người lập trình CSDL

5. Hệ quản trị CSDL

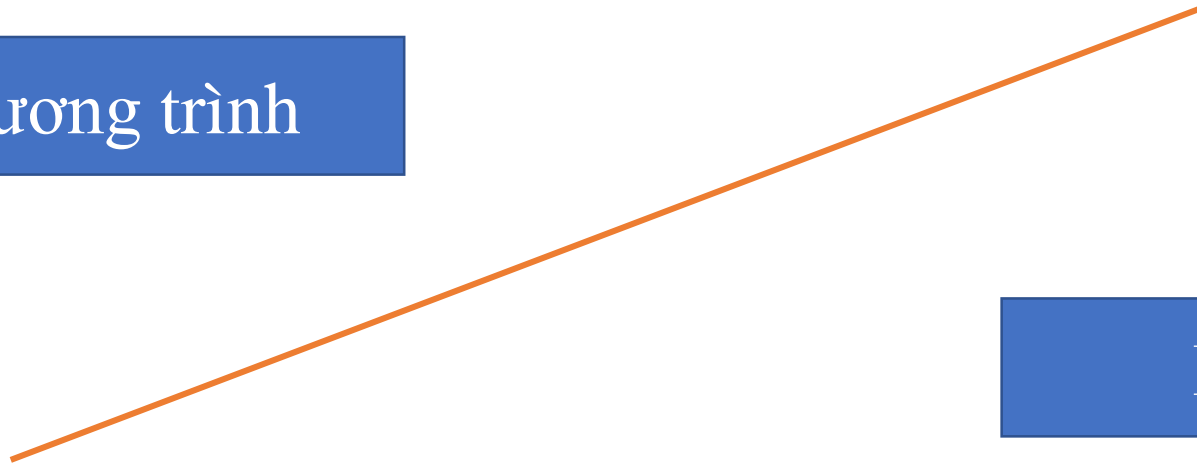


b. Tính độc lập

- ✓ Vì định nghĩa về cấu trúc CSDL được lưu trữ trong catalog nên khi có thay đổi nhỏ về cấu trúc , ta ít phải sửa lại chương trình.

Chương trình

Dữ liệu



5. Hệ quản trị CSDL



c. Tính trừu tượng

- ✓ Hệ CSDL cho phép trình bày dữ liệu ở một mức trừu tượng cho phép, nhằm che bớt những chi tiết lưu trữ thật của dữ liệu
- ✓ Trừu tượng hóa dữ liệu
 - Mô hình dữ liệu
 - Đối tượng
 - Thuộc tính của đối tượng
 - Mối liên hệ



5. Hệ quản trị CSDL

d. Tính nhất quán

- ✓ Lưu trữ dữ liệu thống nhất
 - Tránh được tình trạng trùng lặp thông tin
- ✓ Có cơ chế điều khiển truy xuất dữ liệu hợp lý
 - Tránh được việc tranh chấp dữ liệu
 - Bảo đảm dữ liệu luôn đúng khi muốn truy xuất

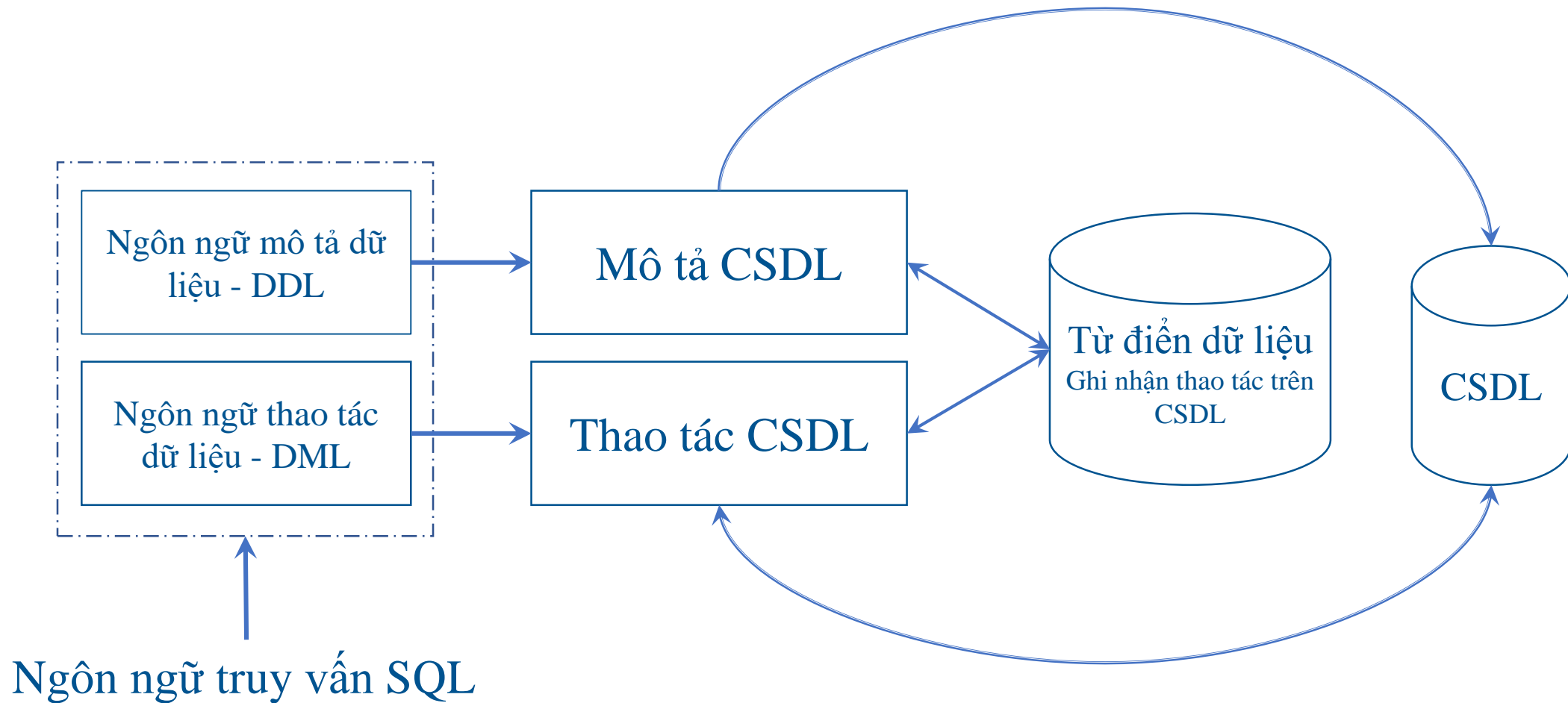
5. Hệ quản trị CSDL



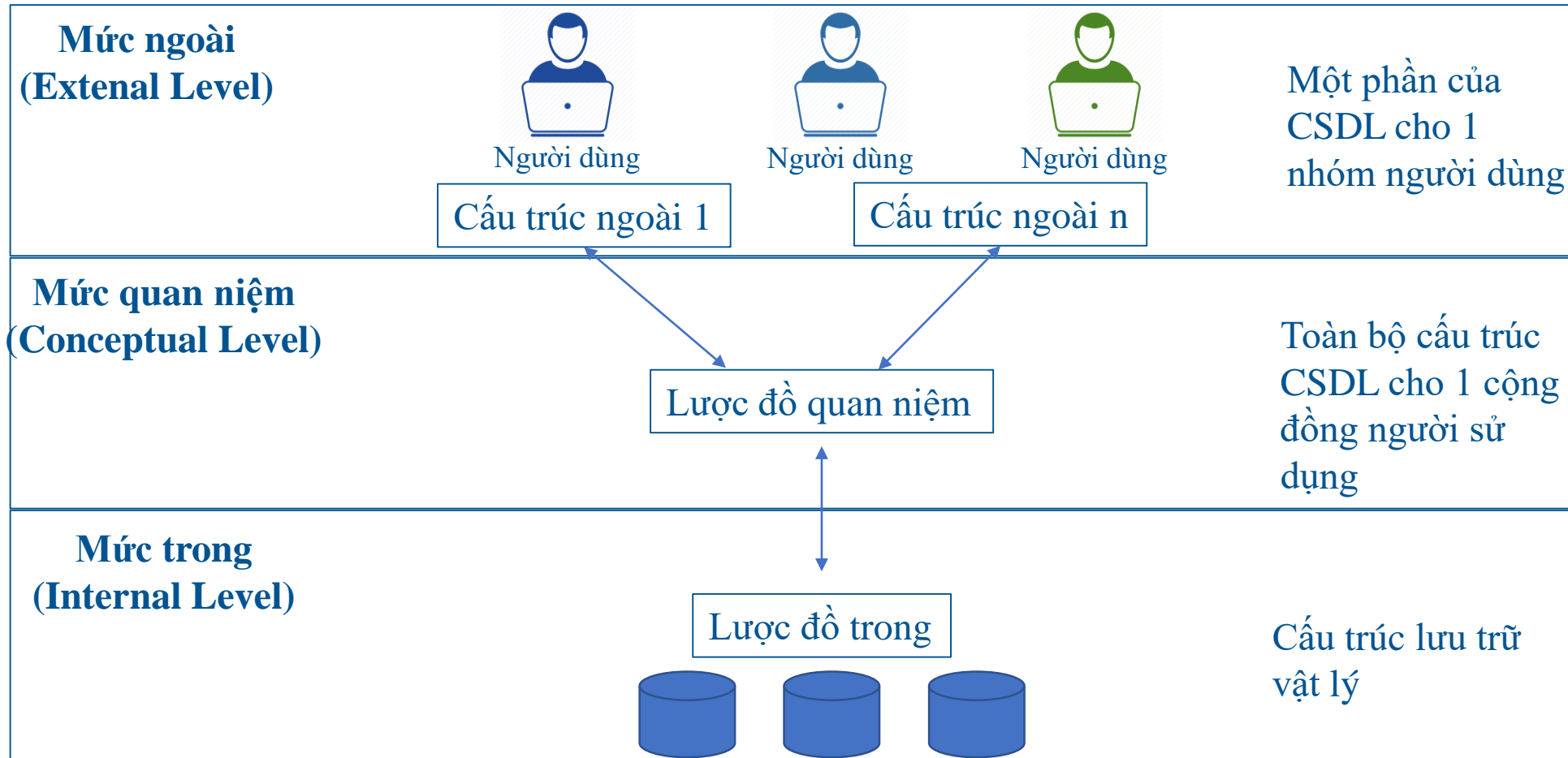
e. Đa khung nhìn

- ✓ Hệ CSDL cho phép nhiều người dùng thao tác lên cùng một CSDL
- ✓ Mỗi người dùng có một cách nhìn (view) khác nhau về CSDL
- ✓ Một khung nhìn (view) là
 - Một phần của CSDL
 - hoặc
 - Dữ liệu tổng hợp từ CSDL

6. Kiến trúc của hệ quản trị CSDL



6. Kiến trúc của hệ quản trị CSDL



6. Kiến trúc của hệ quản trị CSDL

- ✓ Theo kiến trúc ANSI-PARC, một CSDL có 3 mức biểu diễn
 - Mức trong (còn gọi là mức vật lý - Physical)
 - Mức quan niệm (Conception hay Logical)
 - Mức ngoài (External Level)

6. Kiến trúc của hệ quản trị CSDL



Mức trong

- ✓ Đây là mức lưu trữ CSDL
- ✓ Vấn đề cần giải quyết ở mức này là
 - Dữ liệu gì và được lưu trữ như thế nào?
 - Ở đâu (đĩa từ, băng từ, track, sector ... nào)?
 - Cần các chỉ mục gì? Việc truy xuất là tuần tự (Sequential Access) hay ngẫu nhiên (Random Access) đối với từng loại dữ liệu
- ✓ Những người hiểu và làm việc với CSDL tại mức này là người quản trị CSDL (Administrator), gọi là những người sử dụng chuyên môn.



6. Kiến trúc của hệ quản trị CSDL

Mức quan niệm - logical

- ✓ Tại mức này sẽ giải quyết cho câu hỏi CSDL:
 - Cần phải lưu giữ bao nhiêu loại dữ liệu?
 - Đó là những dữ liệu gì?
 - Mối quan hệ giữa các loại dữ liệu này như thế nào?
- ✓ Chuyên viên tin học khảo sát và phân tích, cùng với những người sẽ đảm nhận vai trò quản trị CSDL xác định những loại thông tin gì được cho là cần thiết phải đưa vào CSDL, đồng thời mô tả rõ mối liên hệ giữa các thông tin này.
- ✓ Có thể nói cách khác, CSDL mức quan niệm là một sự biểu diễn trừu tượng CSDL mức vật lý; hoặc ngược lại, CSDL vật lý là sự cài đặt cụ thể của CSDL mức quan niệm

6. Kiến trúc của hệ quản trị CSDL



Mức ngoài

- ✓ Đó là mức của người sử dụng và các chương trình ứng dụng.
- ✓ Mỗi người sử dụng hay mỗi chương trình ứng dụng có thể được "nhìn" (View) CSDL theo một góc độ khác nhau.
- ✓ Có thể "nhìn" thấy toàn bộ hay chỉ một phần hoặc chỉ là các thông tin tổng hợp từ CSDL hiện có.
- ✓ Người sử dụng hay chương trình ứng dụng có thể hoàn toàn không được biết về cấu trúc tổ chức lưu trữ thông tin trong CSDL, thậm chí ngay cả tên gọi của các loại dữ liệu hay tên gọi của các thuộc tính.
- ✓ Họ chỉ có thể làm việc trên một phần CSDL theo cách "nhìn" do người quản trị hay chương trình ứng dụng quy định, gọi là khung nhìn (View)





Hết