

# MÔ HÌNH DỮ LIỆU QUAN HỆ

---

Nguyễn Đình Hóa

[dinhhoa@gmail.com](mailto:dinhhoa@gmail.com)

094-280-7711

# Các khái niệm chung

- **Dữ liệu (data)** là các thông tin được lưu trữ, có thể dưới dạng chữ số, văn bản, đồ họa, đa phương tiện,...
- **Tập dữ liệu (dataset)** là tập hợp các dữ liệu theo một cấu trúc cụ thể phục vụ một mục đích nào đó
- **Cơ sở dữ liệu (database)** là một tập hợp các dữ liệu có tổ chức, thường bao gồm nhiều tập dữ liệu.
- CSDL thường được lưu trữ và truy cập bằng các thiết bị/hệ thống điện tử (máy tính), mà trên đó các dữ liệu được dễ dàng truy cập, thao tác, cập nhật.
- Một CSDL được thiết kế, xây dựng và sử dụng cho một số mục đích cụ thể.

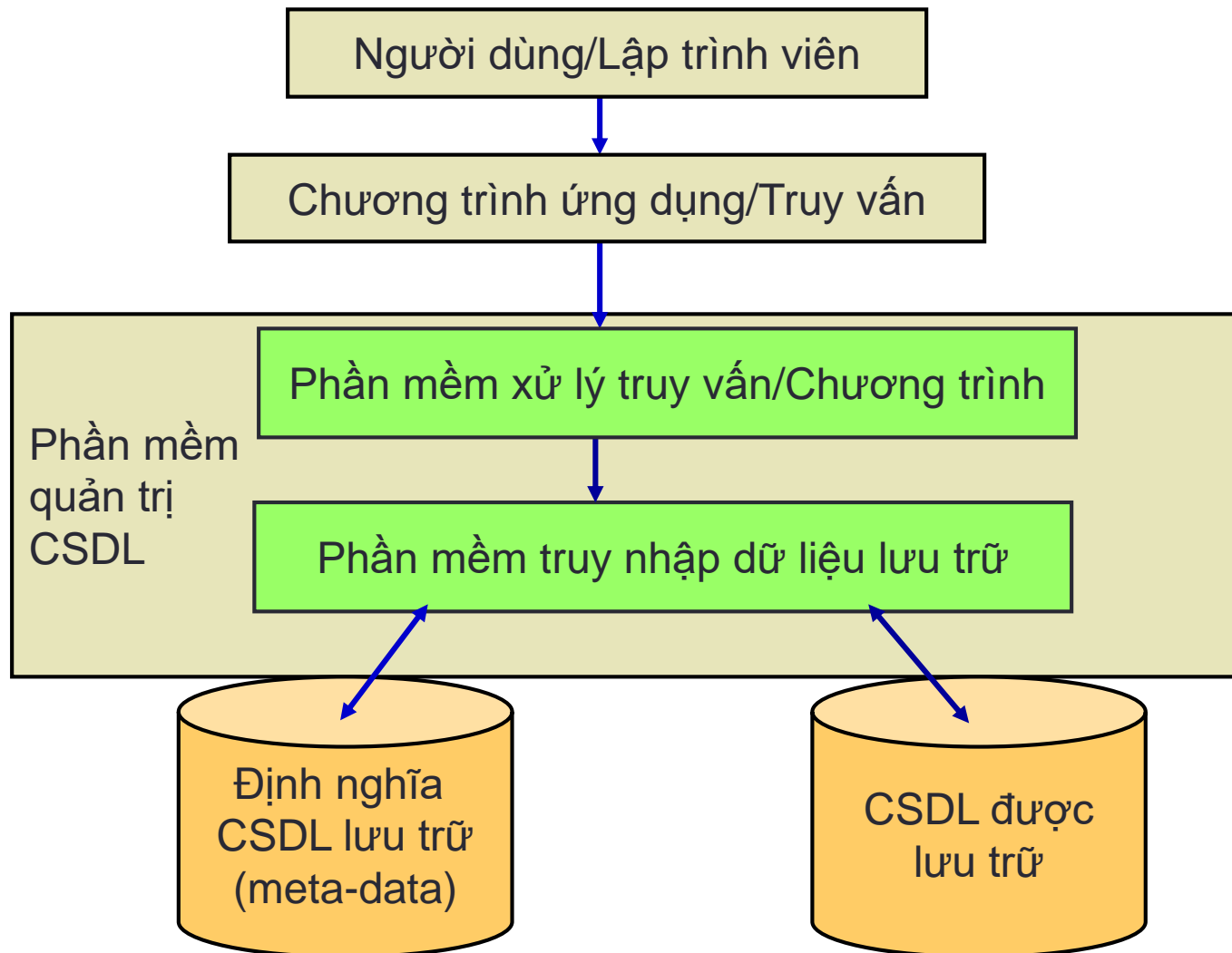
# Các khái niệm chung

- **Hệ quản trị CSDL** (DBMS – Database management system) là một hệ thống phần mềm cho phép tạo lập CSDL và điều khiển mọi truy nhập đến CSDL đó.
- Các đặc tính quan trọng của một hệ quản trị CSDL:
  1. Cho phép người dùng tạo mới CSDL, thông qua ngôn ngữ định nghĩa dữ liệu (DDLs – Data Definition Languages).
  2. Cho phép người dùng truy vấn cơ sở dữ liệu, thông qua ngôn ngữ thao tác dữ liệu (DMLs – Data Manipulation Languages).
  3. Hỗ trợ lưu trữ số lượng lớn dữ liệu, thường lên tới hàng Gigabytes hoặc nhiều hơn, trong một thời gian dài. Duy trì tính bảo mật và tính toàn vẹn trong quá trình xử lý.
  4. Kiểm soát truy nhập dữ liệu từ nhiều người dùng tại cùng một thời điểm.

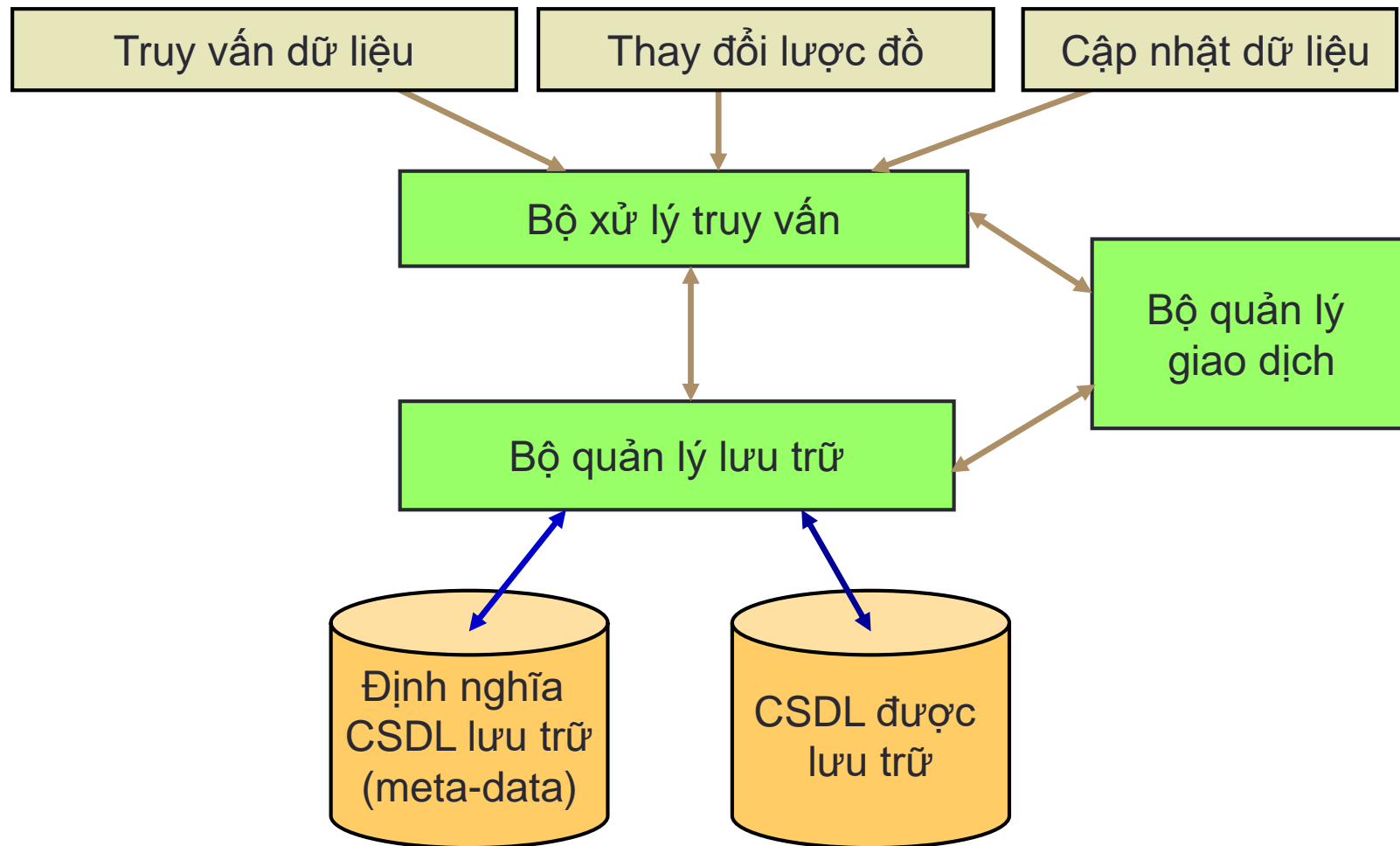
# Các khái niệm chung

- Một CSDL được quản lý bởi một hệ quản trị CSDL thường được gọi là một **hệ cơ sở dữ liệu**.
- Hệ CSDL gồm 4 thành phần:
  1. **CSDL hợp nhất**: có 2 tính chất là tối thiểu hóa dư thừa và được chia sẻ.
  2. **Người dùng**: là những người có nhu cầu truy nhập vào CSDL (người dùng cuối, người viết chương trình ứng dụng, người quản trị CSDL).
  3. **Phần mềm hệ quản trị CSDL**.
  4. **Phần cứng**: gồm các thiết bị nhớ thứ cấp được sử dụng để lưu trữ CSDL.

# Các thành phần của một hệ CSDL



# Kiến trúc hệ CSDL



# Các loại dữ liệu

- **Dữ liệu vật lý:** là những dữ liệu có thực, được nhập vào trong CSDL.
- **Dữ liệu dẫn xuất:** Là những dữ liệu được tính toán từ những dữ liệu nằm trong CSDL.
- Phụ thuộc vào mức độ phức tạp của chương trình ứng dụng và hệ quản trị CSDL, khối dữ liệu dẫn xuất có thể lớn hơn rất nhiều khối dữ liệu vật lý.
- Cần nhắc khi nào dữ liệu dẫn xuất trở thành dữ liệu vật lý?

**Ví dụ:** CSDL lưu thông tin về Sinh viên, trong đó lưu điểm thi của SV. Giá trị trung bình điểm thi của SV có cần lưu trong CSDL hay sẽ được tính toán khi cần?

# Vai trò con người trong môi trường CSDL

- **Người quản trị dữ liệu (DA – Data Administrator):** có trách nhiệm quản lý tài nguyên dữ liệu, bao gồm lập kế hoạch cho CSDL, phát triển và duy trì các chuẩn, chính sách và thủ tục, và thiết kế CSDL mức khái niệm/logic.
- **Người quản trị CSDL (DBA – Database Administrator):** có trách nhiệm với việc lưu trữ vật lý CSDL, bao gồm thiết kế và cài đặt CSDL vật lý, kiểm soát bảo mật và toàn vẹn dữ liệu, duy trì hệ điều hành, và đảm bảo thỏa mãn hiệu năng cho các ứng dụng người dùng.

*=> Vai trò của DBA liên quan đến nhiều đặc tính kỹ thuật hơn DA.*



# Vai trò con người trong môi trường CSDL

- **Người thiết kế CSDL:** trong các dự án thiết kế CSDL lớn, cần phân biệt 2 loại thiết kế:
  - **Người thiết kế CSDL logic** liên quan tới việc xác định CSDL (các thực thể và thuộc tính), các mối quan hệ giữa dữ liệu, và các ràng buộc đối với dữ liệu sẽ được lưu trữ trong CSDL.
  - **Người thiết kế CSDL vật lý** phụ thuộc nhiều vào hệ quản trị CSDL đích, và có thể có nhiều cách để cài đặt CSDL. Người thiết kế CSDL vật lý phải có hiểu biết đầy đủ về tính năng của hệ quản trị CSDL đích.
- **Người phát triển ứng dụng:** Có trách nhiệm xây dựng chương trình ứng dụng cung cấp các chức năng cần thiết cho người dùng cuối, sau khi CSDL đã được cài đặt.

# Vai trò con người trong môi trường CSDL

- **Người dùng cuối:** là “khách hàng” của CSDL, và có thể được phân thành 2 nhóm dựa theo cách mà họ sử dụng hệ thống:
  - Nhóm người dùng không biết đến khái niệm CSDL hoặc hệ quản trị CSDL: Truy nhập CSDL thông qua chương trình ứng dụng được viết riêng biệt, giúp cho các thao tác của người dùng đơn giản nhất có thể.
  - Nhóm người dùng nhận biết được cấu trúc CSDL và các phương tiện được cung cấp bởi hệ quản trị CSDL: Thường dùng các ngôn ngữ truy vấn bậc cao như SQL để thực hiện những thao tác được yêu cầu và thậm chí có thể viết những chương trình ứng dụng để phục vụ cho mục đích riêng.

# Các ngôn ngữ cơ sở dữ liệu

- Một ngôn ngữ cơ sở dữ liệu bao gồm 2 phần: Một **ngôn ngữ định nghĩa dữ liệu (DDL - Data Definition Language)** và một **ngôn ngữ thao tác dữ liệu (DML - Data Manipulation Language)**.
- DDL được dùng để xác định lược đồ CSDL, và DML dùng để đọc và cập nhật CSDL.
- Các ngôn ngữ này được gọi là *ngôn ngữ cơ sở dữ liệu* vì chúng không bao gồm các cấu trúc lập trình cần thiết cho việc tính toán như là các cấu trúc điều khiển hoặc câu lệnh lặp (được cung cấp bởi các ngôn ngữ lập trình bậc cao).

# Các ngôn ngữ cơ sở dữ liệu

- Hầu hết các hệ quản trị CSDL đều có một môi trường cho phép nhúng các ngôn ngữ con dữ liệu vào trong một ngôn ngữ lập trình bậc cao như COBOL, Pascal, C, C++, Java hay Visual Basic (được gọi là các *ngôn ngữ chủ*).
- Nhiều ngôn ngữ con dữ liệu cũng cung cấp một phiên bản tương tác hoặc không nhúng vào ngôn ngữ mà có thể được đưa vào trực tiếp từ một thiết bị đầu cuối.

# Ngôn ngữ định nghĩa dữ liệu (DDL)

- DDL là ngôn ngữ cho phép người quản trị CSDL hoặc người dùng mô tả và đặt tên các thực thể, thuộc tính và các quan hệ cần thiết cho ứng dụng, cùng với những ràng buộc về bảo mật và toàn vẹn liên quan.
- Kết quả của việc thực thi/biên dịch câu lệnh DDL là một tập các bảng được lưu trong các tệp đặc biệt, được gọi là **danh mục hệ thống** (system catalog) (hay **từ điển dữ liệu/ thư mục dữ liệu**).

# Ngôn ngữ thao tác dữ liệu (DML)

- DML là ngôn ngữ cung cấp một tập các thao tác hỗ trợ cho các phép toán thao tác dữ liệu cơ bản trên dữ liệu được lưu trong CSDL.
- Các thao tác của DML bao gồm:
  - Chèn dữ liệu mới vào CSDL
  - Sửa đổi dữ liệu đã được lưu trữ trong CSDL
  - Lấy dữ liệu từ CSDL
  - Xóa dữ liệu trong CSDL
- Phần thao tác lấy dữ liệu ra được gọi là **ngôn ngữ truy vấn**.

# Ngôn ngữ thao tác dữ liệu (DML)

- DML là ngôn ngữ cung cấp một tập các thao tác hỗ trợ cho các phép toán thao tác dữ liệu cơ bản trên dữ liệu được lưu trong CSDL.
- Các thao tác của DML bao gồm:
  - Chèn dữ liệu mới vào CSDL
  - Sửa đổi dữ liệu đã được lưu trữ trong CSDL
  - Lấy dữ liệu từ CSDL
  - Xóa dữ liệu trong CSDL
- Phần thao tác lấy dữ liệu ra được gọi là **ngôn ngữ truy vấn**.

# Các loại hệ CSDL

***Có 2 loại hệ CSDL dựa trên kiến trúc của chúng:***

- Hệ cơ sở dữ liệu tập trung:
  - Hệ CSDL cá nhân
  - Hệ CSDL trung tâm
  - Hệ CSDL khách/chủ (client/server)
- Hệ cơ sở dữ liệu phân tán
  - Hệ CSDL phân tán thuần nhất
  - Hệ CSDL phân tán không thuần nhất



# Mô hình dữ liệu

- **Mô hình dữ liệu** là một tập hợp các khái niệm dùng cho việc mô tả và thao tác dữ liệu, các mối quan hệ và các ràng buộc trên dữ liệu của tổ chức.
- Mô hình dữ liệu phải cung cấp các khái niệm và ký hiệu cơ bản, cho phép người thiết kế CSDL và người dùng trao đổi với nhau những hiểu biết về dữ liệu của tổ chức một cách chính xác và không đa nghĩa.

# Các thành phần của mô hình dữ liệu

- Một mô hình dữ liệu có thể bao gồm 3 thành phần:
  1. **Phần cấu trúc** gồm tập các luật mà theo đó CSDL được xây dựng.
  2. **Phần thao tác, định nghĩa các thao tác** được phép trên dữ liệu (gồm cả các thao tác cập nhật, lấy dữ liệu từ CSDL, và thay đổi cấu trúc của CSDL).
  3. Có thể có một **tập các luật về tính toàn vẹn**, nhằm đảm bảo CSDL luôn chính xác.

# Mô hình dữ liệu quan hệ

- **Mô hình dữ liệu quan hệ (Relational Data Model)** là mô hình được xây dựng dựa trên mối quan hệ giữa các dữ liệu theo đó dữ liệu được biểu diễn dưới dạng tập hợp.
- Hệ thống phần mềm được sử dụng để duy trì sự tồn tại của một cơ sở dữ liệu theo mô hình quan hệ được gọi là hệ quản trị cơ sở dữ liệu quan hệ (Relational database management system – RDBMS).
- Structural Query Language (SQL) là ngôn ngữ CSDL thường được sử dụng trong các RDBMS.

# Các bước thiết kế một hệ CSDL

## **1. Phân tích yêu cầu:** Phải xác định được:

- Dữ liệu nào được lưu trữ trong CSDL,
- Ứng dụng nào sẽ được xây dựng trên CSDL này,
- Các thao tác nào được sử dụng thường xuyên và các yêu cầu về hiệu năng của hệ thống.

*=> Quá trình này liên quan đến những trao đổi của các nhóm người dùng và nhóm nghiên cứu môi trường hiện tại. Tìm hiểu các ứng dụng đang có xem có cần thay thế hoặc bổ trợ cho hệ CSDL không.*

# Các bước thiết kế một hệ CSDL

- 2. Thiết kế CSDL mức khái niệm:** Những thông tin có được từ bước phân tích yêu cầu sẽ được dùng để phát triển một mô tả mức tổng quát dữ liệu được lưu trong CSDL, cùng với các ràng buộc cần thiết trên dữ liệu này.
- 3. Thiết kế CSDL mức logic:** Một hệ quản trị CSDL sẽ được chọn để cài đặt CSDL và chuyển thiết kế CSDL mức khái niệm thành một lược đồ CSDL với mô hình dữ liệu của hệ quản trị CSDL đã chọn.

# Các bước thiết kế một hệ CSDL

- 4. Cải tiến lược đồ:** Các lược đồ được phát triển ở bước 3 sẽ được phân tích các vấn đề tiềm ẩn. Tại đây, CSDL sẽ được **chuẩn hóa**, dựa trên lý thuyết toán học.
- 5. Thiết kế CSDL mức vật lý:** Khối lượng công việc tiềm ẩn và các phương pháp truy nhập được mô phỏng để xác định các điểm yếu tiềm ẩn trong CSDL mức khái niệm. Quá trình này thường là nguyên nhân tạo ra các tệp chỉ mục hoặc/và các quan hệ nhóm. Trong trường hợp đặc biệt, toàn bộ mô hình khái niệm sẽ được xây dựng lại.

# Các bước thiết kế một hệ CSDL

**6. Thiết kế an toàn bảo mật:** Xác định các nhóm người dùng và phân tích vai trò của họ để định nghĩa các phương pháp truy nhập dữ liệu.

Trong quá trình phát triển, thường có bước cuối cùng (bước thứ 7), gọi là **pha điều chỉnh (tuning phase)**, trong đó CSDL sẽ được thực hiện (mặc dù nó có thể chỉ được chạy mô phỏng) và sẽ được cải tiến, chỉnh sửa để đáp ứng nhu cầu thực thi trong môi trường mong đợi.