

# Cơ sở dữ liệu

Trịnh Thành

Khoa Công Nghệ Thông Tin – Đại học Phenikaa

[thanh.trinh@phenikaa-uni.edu.vn](mailto:thanh.trinh@phenikaa-uni.edu.vn)

# Chương 1. Tổng quan về cơ sở dữ liệu

- Chương 1: Tổng quan về cơ sở dữ liệu
  - 1.1. Vấn đề cần thiết lưu trữ dữ liệu
  - 1.2. Hệ quản trị CSDL
  - 1.3 Các đặc tính CSDL
  - 1.4 Quản lý dữ liệu

# 1. Vấn đề cần thiết lưu trữ dữ liệu

- Đặt vấn đề

- CSDL là gì?
- Tại sao chúng ta cần CSDL
- Tại sao chúng ta cần sử dụng các hệ quản trị CSDL

# 1. Vấn đề cần thiết lưu trữ dữ liệu

- Dữ liệu (data)?
  - Là các thông tin của đối tượng (Ví dụ: người, vật, sự kiện,)
  - Có thể truy cập vào dữ liệu để truy xuất các thông tin
- Dữ liệu được mô tả dưới nhiều dạng khác nhau: ví dụ, ký tự, số, hình ảnh, văn bản.
- Dữ liệu về các đối tượng khác nhau thì khác nhau, tùy thuộc vào ngữ cảnh và mục đích sử dụng
- Ví dụ: Dữ liệu về đối tượng sinh viên có thể khác nhau tùy vào mục đích quản lý:
  - Mục đích quản lý điểm: Tên, mã sv, điểm môn 1, điểm môn 2
  - Mục đích là quản lý nhân thân: Tên, địa chỉ, ngày sinh, quê quán, giới tính, lớp.

# Vấn đề quản lý dữ liệu về nhân sự của một công ty?

Phòng Tổ chức: Lưu trữ thông tin về lý lịch nhân viên

Phòng Tài vụ: Lưu trữ về họ tên, lương

Phòng Văn thư: Tạo các công văn về các nhân viên

→ Để phát triển một hệ thống quản lý:

→ Những chương trình ứng dụng mới phải được viết khi một nhu cầu nảy sinh.

→ Các tập tin thường trực mới được tạo ra theo yêu cầu.

→ Trong một thời gian dài các tập tin có thể ở các dạng thức khác nhau.

→ Những chương trình ứng dụng có thể viết bởi các ngôn ngữ khác nhau.

# Vấn đề quản lý dữ liệu về nhân sự của một công ty?

- Ưu điểm:
  - Việc xây dựng những hệ thống tập tin riêng lẻ tại từng đơn vị quản lý ít tốn thời gian bởi lượng thông tin cần quản lý và khai thác là nhỏ, không đòi hỏi đầu tư nhiều nên triển khai nhanh.
  - Thông tin được khai thác chỉ phục vụ cho mục đích hẹp nên khả năng đáp ứng nhanh chóng, kịp thời.

# Vấn đề quản lý dữ liệu về nhân sự của một công ty?

- Nhược điểm
  - Dữ liệu dư thừa và không nhất quán
  - Khó khăn trong việc truy xuất dữ liệu
  - Khó khăn khi viết chương trình ứng dụng đòi hỏi thông tin liên quan đến nhiều đơn vị
  - Nhiều người sử dụng
  - Các vấn đề bảo mật
  - Khó khăn khi nâng cấp hệ thống.

# Vấn đề quản lý dữ liệu về nhân sự của một công ty?

- Yêu cầu đặt ra: Cần một hệ quản trị cơ sở dữ liệu
- Trước tiên: Cần phải định nghĩa CƠ SỞ DỮ LIỆU Là gì?



# Cơ sở dữ liệu là gì?

- **Ví dụ về quản lý chương trình đào tạo?**

- Khoá học: lớp học, sinh viên, giáo viên, môn học,...
- Sinh viên: Thông tin các nhân, thông tin nhập học, môn học, khoá học.
- Môn học: Số tín chỉ, giáo viên, lịch học.

Thông tin rất đa dạng: Số, văn bản, ảnh....: -> Cần lưu trữ thông tin một cách khoa học -> Cơ sở dữ liệu:

- **Cơ sở dữ liệu:**

Cơ sở dữ liệu (database) là một tập hợp các thông tin có cấu trúc có liên quan với nhau; mô tả các vấn đề của thế giới thực; được thiết kế logic nhằm mục đích thoả mãn yêu cầu khai thác và truy xuất thông tin dựa trên CSDL được tập hợp đó.

# Khai thác thông tin từ CSDL

- Xác định các vấn đề cần khai thác, truy xuất thông tin
  - Từ “Sinh Viên”: Tình Trạng, học qua bao nhiêu môn, nợ bao nhiêu, điểm tổng kết
  - Từ “Khoá học”: Số lượng tốt nghiệp, tỷ lệ nam nữ,..
  - Từ “Môn học”: Tỷ lệ qua môn, số lần thi
- 
- Chúng ta cần có một cách thức để quản lý, khai thác thông tin hiệu quả -> Phần mềm ứng dụng hay chuyên biệt

Người sử dụng CSDL

- **Quản trị viên :**

- Có trách nhiệm quản lý hệ CSDL
  - Cấp quyền truy cập CSDL
  - Điều phối và giám sát việc sử dụng CSDL

# Người sử dụng CSDL

- **Người thiết kế**
- Chịu trách nhiệm về
  - Lựa chọn cấu trúc phù hợp để lưu trữ dữ liệu
  - Quyết định những dữ liệu nào cần được lưu trữ
- Liên hệ với người dùng để nắm bắt được những yêu cầu và đưa ra một thiết kế CSDL thỏa yêu cầu này
- Có thể là 1 nhóm các DBA quản lý các CSDL sau khi việc thiết kế hoàn tất

# Người sử dụng CSDL

- **Người dùng cuối**

- Người ít sử dụng

- Ít khi truy cập CSDL, nhưng cần những thông tin khác nhau trong mỗi lần truy cập và dùng những câu truy vấn phức tạp
- Người quản lý

- Người sử dụng thường xuyên

- Thường xuyên truy vấn và cập nhật CSDL nhờ vào một số các chức năng đã được xây dựng sẵn
- Nhân viên

- Người sử dụng đặc biệt

- Thông thạo về HQT CSDL, tự xây dựng những truy vấn phức tạp cho công việc
- Kỹ sư, nhà khoa học, người phân tích,...

## 1.2 Hệ quản trị CSDL

- Yêu cầu:
  - Lưu trữ thông tin cần thiết và xác thực
  - Truy xuất thông tin hiệu quả (chính xác và nhanh chóng)
  - Đáp ứng nhu cầu của người sử dụng CSDL
- Thực hiện:
  - Làm sạch dữ liệu
  - Xác định thông tin cần lưu trữ
  - Xác định mô hình (cách thức) lưu trữ.
  - Xác định vấn đề cần truy xuất.

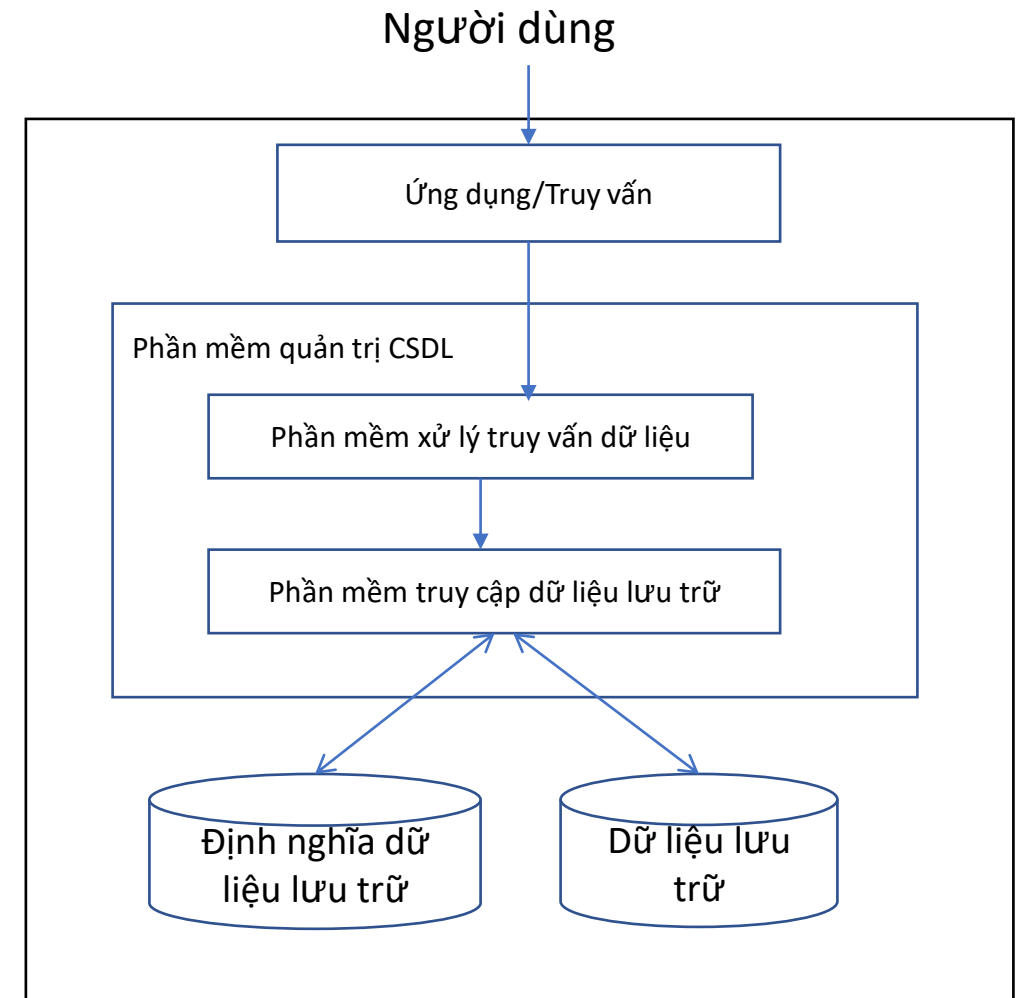
## 1.2 Hệ quản trị CSDL

- Hệ thống này được gọi là hệ quản trị CSDL (Database Management System - DBMS), là công cụ hỗ trợ tích cực cho các nhà phân tích, thiết kế và khai thác CSDL.
- Hiện nay có nhiều hệ quản trị CSDL trên thị trường như: SQL server, DB2, Microsoft Access, Oracle, **MySQL**.
- Các hệ quản trị này cho phép:
  - Định nghĩa xác định kiểu, cấu trúc, ràng buộc dữ liệu.
  - Tạo các tệp lưu trữ dữ liệu trên các thiết bị vật lý
  - Cho phép thao tác truy vấn, cập nhật, truy xuất

## Hệ cơ sở dữ liệu

### Hệ cơ sở dữ liệu (Database management system (DBMS):

- Cho phép quản lý dữ liệu
- Là một tập hợp các chương trình cho phép người dùng tạo ra và bảo trì một CSDL
- Cho phép nhiều truy cập đồng thời vào một CSDL
- Cho phép định nghĩa, cấu trúc, nhân bản và chia sẻ các CSDL giữa các người dùng với nhau và các ứng dụng.
- Định nghĩa về CSDL hay là thông tin mô tả cũng được lưu trữ trong CSDL dưới dạng catalog hay là dictionary (Hay gọi là meta-data).

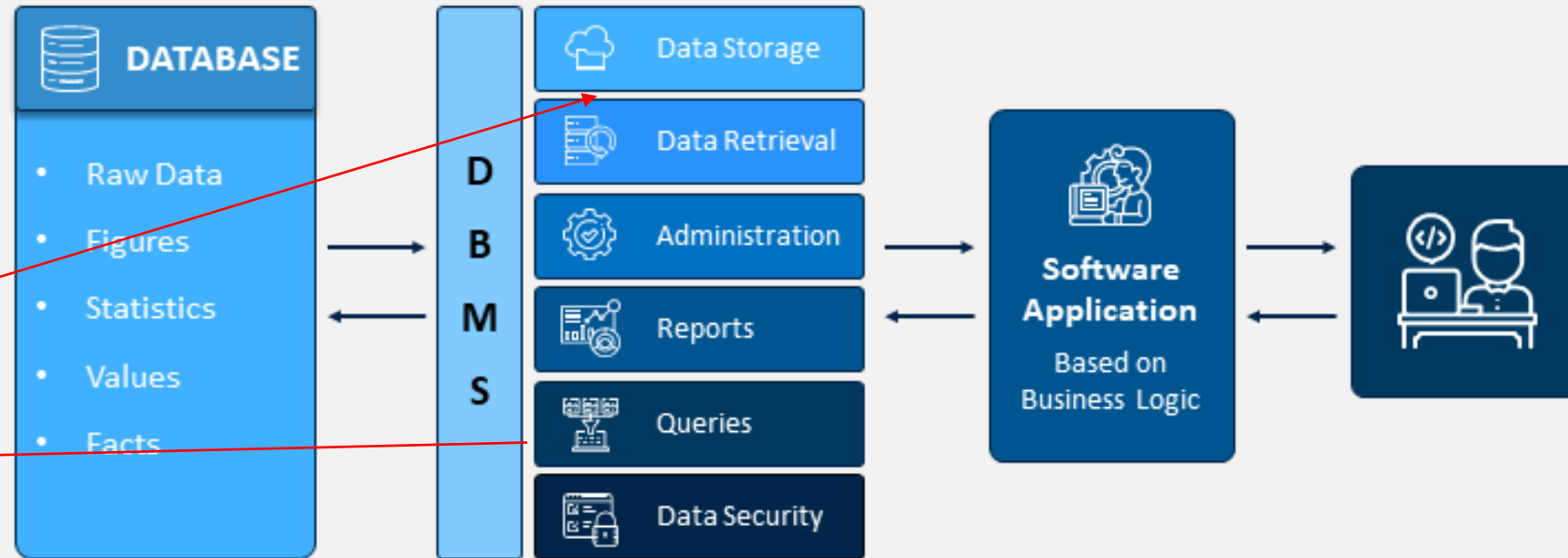




# DATABASE MANAGEMENT SYSTEM (DBMS)

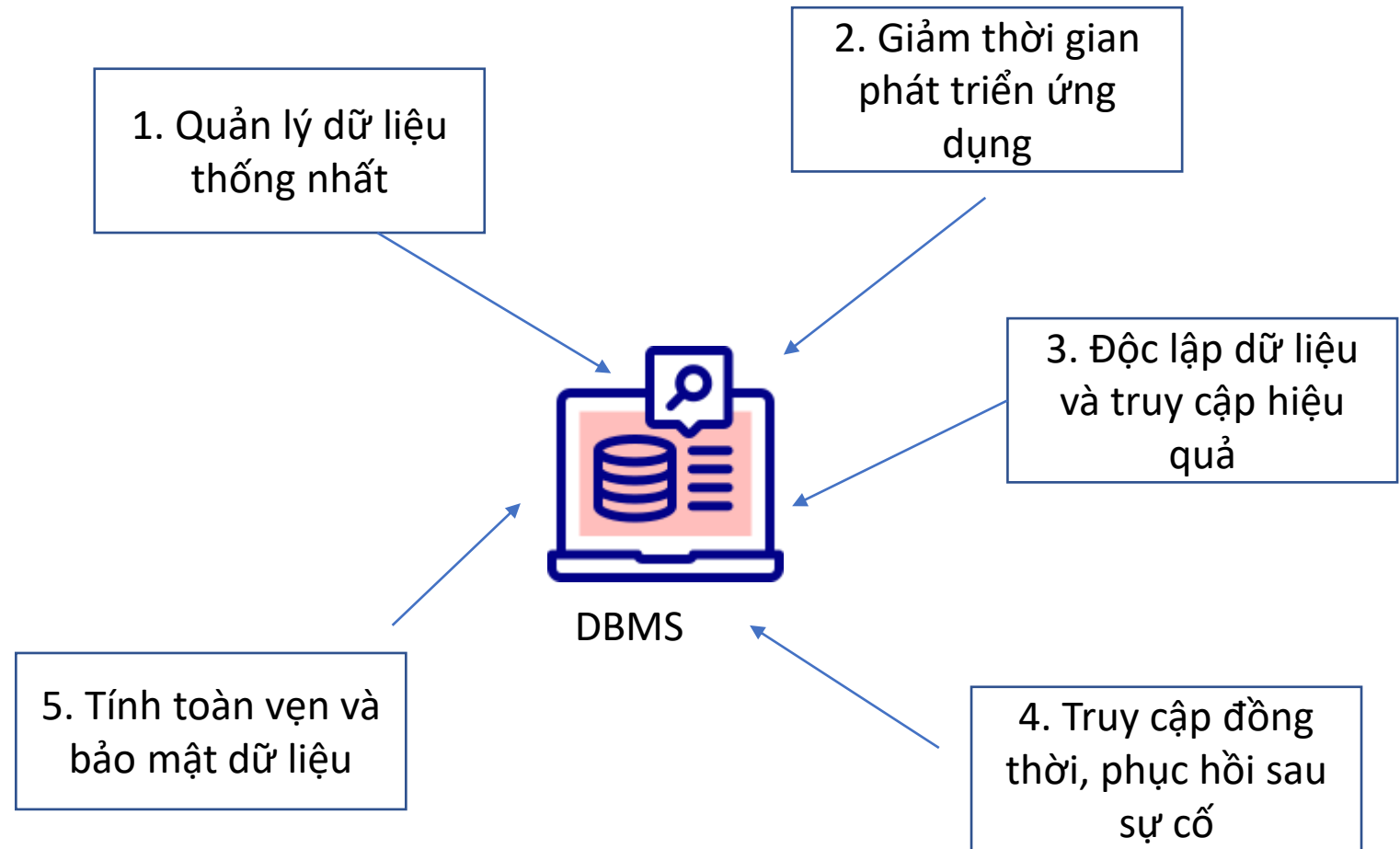
Enter your sub headline here

Chúng ta sẽ học  
trong môn học này



## 1.2 Hệ quản trị CSDL

Tại sao phải sử dụng DBMS??



# Các tính năng của HQT CSDL

- Kiểm soát được tính dư thừa của dữ liệu
  - Tích hợp các nhu cầu dữ liệu của người dùng để xây dựng một CSDL thống nhất
- Chia sẻ dữ liệu
  - Trong môi trường đa người dùng, các HQT phải cho phép truy xuất dữ liệu đồng thời
- Hạn chế những truy cập không cho phép
  - Từng người dùng và nhóm người dùng có một tài khoản và mật mã để truy xuất dữ liệu
- Cung cấp nhiều giao diện
  - HQT cung cấp ngôn ngữ giữa CSDL và người dùng

# Các tính năng của HQT CSDL

- Đảm bảo các ràng buộc toàn vẹn (RBTV)
  - RBTV (Integrity Constraints) là những qui định cần được thỏa mãn để đảm bảo dữ liệu luôn phản ánh đúng ngữ nghĩa của thế giới thực
  - Một số RB có thể được khai báo với HQT và HQT sẽ tự động kiểm tra. Một số RB khác được kiểm tra nhờ CTƯD
- Khả năng sao lưu dự phòng khi gặp sự cố
  - Có khả năng khôi phục dữ liệu khi có sự hư hỏng về phần cứng hoặc phần mềm

# Các tính năng của HQT CSDL

- Các tính năng khác
  - Chuẩn hóa
    - Cho phép DBA định nghĩa và bắt buộc áp dụng một chuẩn thống nhất cho mọi người dùng
  - Uyển chuyển
    - Khi nhu cầu công việc thay đổi, cấu trúc CSDL rất có thể thay đổi, HQT cho phép thêm hoặc mở rộng cấu trúc mà không làm ảnh hưởng đến CTƯD
  - Giảm thời gian phát triển ứng dụng
  - Tính khả dụng
    - Khi có một sự thay đổi lên CSDL, tất cả người dùng đều thấy được

## 1.2 Hệ quản trị CSDL

- **Một hệ quản trị CSDL phải có:**

- Ngôn ngữ giao tiếp giữa users và CSDL:

- Ngôn ngữ mô tả dữ liệu (Data Definition Language - DDL) để cho phép khai báo cấu trúc của CSDL, khai báo các mối liên hệ của dữ liệu (Data Relationship) và các quy tắc (Rules, Constraint) quản lý áp đặt trên dữ liệu.
- Ngôn ngữ thao tác dữ liệu (Data Manipulation Language - DML) cho phép users có thể thêm, xoá, dữ liệu trong CSDL.
- Ngôn ngữ truy vấn dữ liệu (Structured Query Language - SQL) cho phép users truy vấn các thông tin cần thiết.
- Ngôn ngữ quản lý dữ liệu (Data Control Language - DCL) cho phép những người quản trị hệ thống thay đổi cấu trúc của các bảng dữ liệu, khai báo bảo mật thông tin và cấp quyền khai thác CSDL cho users.
- Từ điển dữ liệu (Data dictionary) dùng để mô tả các ánh xạ liên kết, ghi nhận các thành phần cấu trúc của CSDL, các chương trình ứng dụng, mật mã, quyền sử dụng...

## 1.2 Hệ quản trị CSDL

- **Một hệ quản trị CSDL phải có:**
- Có biện pháp bảo mật tốt.
- Có cơ chế giải quyết tranh chấp dữ liệu.
- Có cơ chế sao lưu và phục hồi dữ liệu khi có sự cố xảy ra.
- Có giao diện tốt, dễ sử dụng.
- Bảo đảm tính độc lập giữa dữ liệu và chương trình: khi có sự thay đổi dữ liệu (như sửa đổi cấu trúc lưu trữ các bảng dữ liệu, thêm các chỉ mục,...) thì các chương trình ứng dụng đang chạy trên CSDL vẫn không cần phải viết lại hay cũng không ảnh hưởng đến users khác.

## 1.3 Các đặc tính CSDL

- CSDL cung cấp khả năng trừu tượng hoá dữ liệu thông qua ba mức:
  - Mức trong (mức vật lý - Physical level):
  - Mức khái niệm hay mức logic (conception level, logical level)
  - Mức ngoài hay mức nhìn (view level)
- Sự phân biệt giữa các mức tạo nên 2 tầng độc lập
  - Độc lập dữ liệu logic
  - Độc lập dữ liệu vật lý



**Mức ngoài:** mô tả bởi một sơ đồ. Bao gồm định nghĩa các bản ghi và mối quan hệ. Người dùng nhìn thấy và hiểu về dữ liệu



End Users



Độc lập dữ liệu logic

**Mức khái**

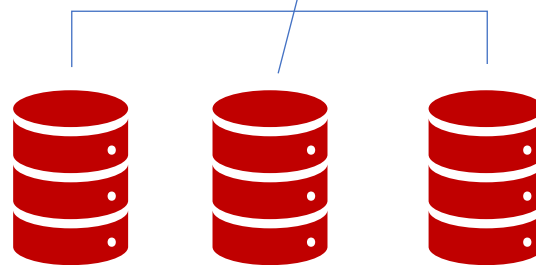
**niệm:** diễn tả toàn bộ dữ liệu, tập ghi và mối quan hệ giữa chúng

Sơ đồ khái niệm

Độc lập dữ liệu vật lý

Sơ đồ bên trong

**Mức trong:** Cho biết dữ liệu được lưu trữ thế nào và mô tả cấu trúc của dữ liệu, phương pháp truy cập được sử dụng bởi CSDL



Stored  
database

## 1.3 Các đặc tính CSDL

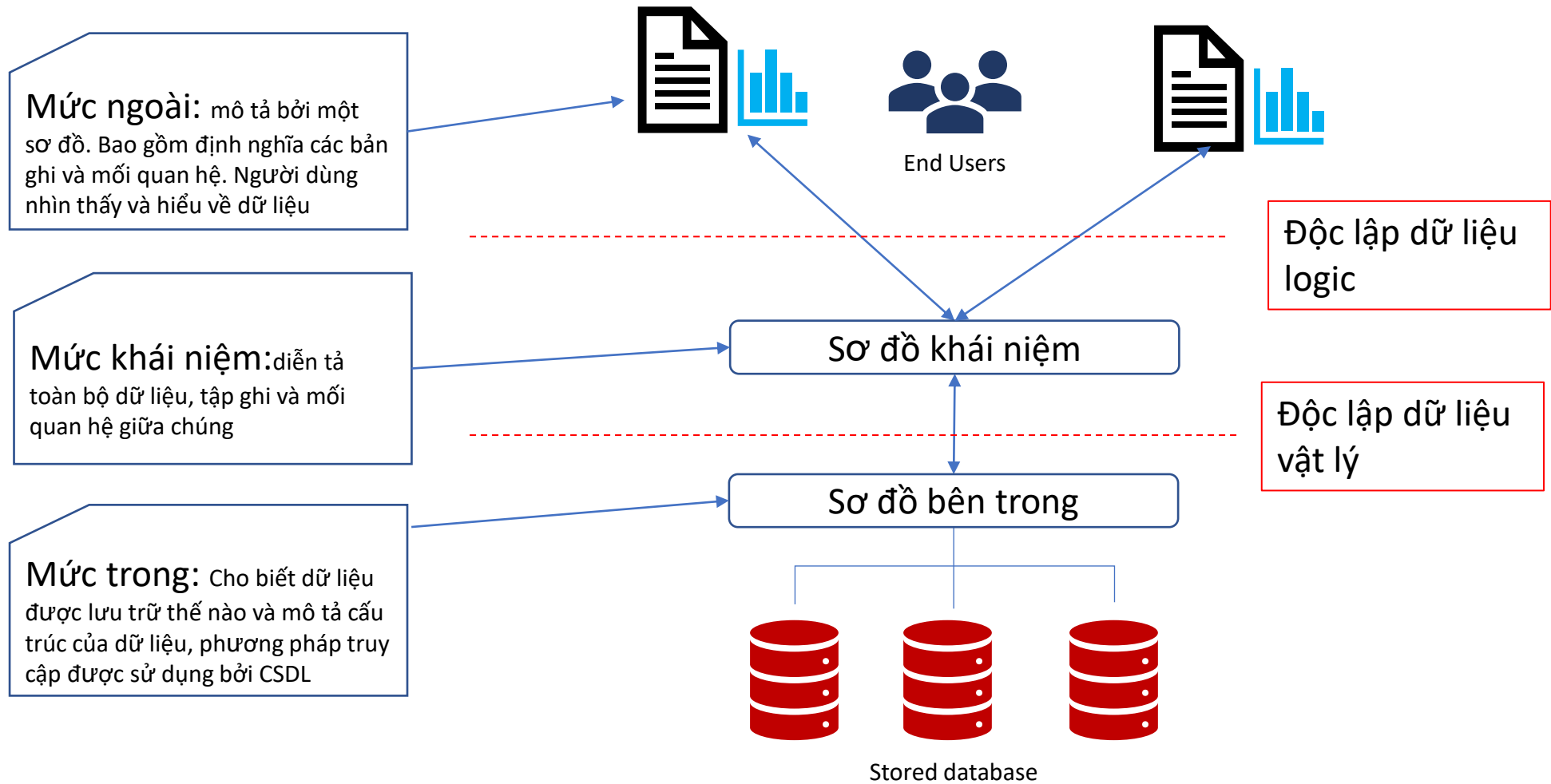
- Mức trong (Mức vật lý):
  - Lưu trữ toàn bộ dữ liệu.
  - Cách thức dữ liệu được lưu trữ thực sự. Trả lời câu hỏi dữ liệu gì và lưu trữ chúng như thế nào? Cần các chỉ mục gì?
  - Ví dụ như chỉ mục, B-tree.
  - Những người thao tác tại mức này là người quản trị CSDL và những người sử dụng chuyên môn.
  - Người dùng CSDL không nhất thiết phải nắm được toàn bộ cấu trúc của các file dữ liệu khi sử dụng một cơ sở dữ liệu.

## 1.3 Các đặc tính CSDL

- Mức khái niệm hay mức logic (conceptual level, logical level)
  - Trả lời câu hỏi cần phải lưu trữ bao nhiêu loại dữ liệu? Đó là những dữ liệu gì? Mối quan hệ giữa chúng như thế nào?
  - Lớp logic có thể chứa một tập hợp các bảng hai chiều, một cấu trúc phân cấp tương tự như sơ đồ tổ chức của một công ty hay vài cấu trúc khác.
  - Cấu trúc dữ liệu trừu tượng được tạo thành từ lớp vật lý
  - Như vậy CSDL mức vật lý là sự cài đặt cụ thể của CSDL mức quan niệm.

## 1.3 Các đặc tính CSDL

- Mức bên ngoài
- Bao gồm các khung nhìn (view)
- Khung nhìn: là khái niệm cho phép nhiều người dùng quan sát dữ liệu theo nhiều cách khác nhau trong khi dữ liệu lưu bên dưới tầng vật lý là duy nhất.
- Là mức của users và các chương trình ứng dụng.
- Làm việc tại mức này là các nhà chuyên môn, kỹ sư tin học, users không chuyên.
- Mỗi users hay mỗi chương trình ứng dụng có thể được nhìn CSDL theo một góc độ khác nhau. Tùy thuộc vào vai trò của users mà có thể thấy tất cả hoặc một phần của CSDL. users hay chương trình ứng dụng có thể không được biết về cấu trúc tổ chức lưu trữ thông tin trong CSDL. 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.



## 1.3 Các đặc tính CSDL

- Độc lập dữ liệu vật lý:
  - Khả năng thay đổi cấu trúc file vật lý của một cơ sở dữ liệu mà không làm gián đoạn người dùng đang truy cập vào các quá trình đang diễn ra được gọi độc lập dữ liệu vật lý.
  - Việc phân biệt lớp vật lý với lớp logic tạo ra sự độc lập dữ liệu vật lý
  - Khả năng cập nhật lược đồ vật lý không làm cho các chương trình ứng dụng phải bị viết lại.
  - Việc cập nhật ở mức này thường cần thiết để cải tiến hiệu suất.

## 1.3 Các đặc tính CSDL

- Độc lập dữ liệu logic
  - Khả năng tạo ra các thay đổi tới lớp logic mà không làm gián đoạn người dùng hiện tại và các quá trình đang diễn ra được gọi là độc lập dữ liệu logic
  - Việc phân biệt giữa lớp logic và lớp bên ngoài tạo ra tính độc lập của dữ liệu logic.
  - Khả năng cập nhật lược đồ logic mà không làm cho các chương trình ứng dụng phải bị viết lại.
  - Những cập nhật thường được thực hiện khi cấu trúc logic của cơ sở dữ liệu thay đổi.

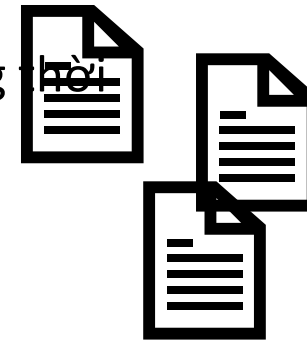
## 1.4 Quản lý dữ liệu

- Quản lý dữ liệu: là quản lý một số lượng lớn dữ liệu, bao gồm cả việc lưu trữ và cung cấp cơ chế cho phép Thao tác (thêm, sửa, xoá dữ liệu) và Truy vấn dữ liệu
- Hai phương pháp quản lý dữ liệu:
  - Hệ thống quản lý bằng file
  - Hệ thống quản lý bằng CSDL



# 1.4 Quản lý dữ liệu

- Quản lý dữ liệu bằng file
  - Dữ liệu được lưu trữ trong các file riêng biệt
  - Ví dụ: Các chương trình lưu trữ thông tin bằng hệ thống các dạng file text
- Nhược điểm của việc quản lý bằng file:
  - Dư thừa và mâu thuẫn dữ liệu
  - Kém hiệu quả trong truy xuất ngẫu nhiên, hoặc xử lý đồng thời
  - Dữ liệu lưu trữ rời rạc
  - Gặp vấn đề về an toàn và bảo mật



# Quản lý dữ liệu bằng CSDL

- Quản lý bằng CSDL giúp dữ liệu được lưu trữ một cách hiệu quả và có tổ chức, cho phép quản lý dữ liệu nhanh chóng và hiệu quả
- Lợi ích của hệ thống quản lý bằng CSDL:  
Tránh dư thừa, trùng lặp dữ liệu
- Đảm bảo sự nhất quán trong CSDL
- Các dữ liệu lưu trữ có thể được chia sẻ
- Có thể thiết lập các chuẩn dữ liệu
- Duy trì tính toàn vẹn dữ liệu
- Đảm bảo bảo mật dữ liệu



# Tóm Tắt

- CSDL là gì
- Hệ cơ sở dữ liệu
- Hệ quản trị cơ sở dữ liệu
- Các đặc tính của cơ sở dữ liệu