



HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



BÀI GIẢNG MÔN  
CƠ SỞ DỮ LIỆU

Giảng viên: TS. Nguyễn Thị Thu Hiền

Email: [csdl.nth@gmail.com](mailto:csdl.nth@gmail.com)

# Nội dung & Mục tiêu học phần

Mục tiêu Nội dung	Bậc 1	Bậc 2	Bậc 3
Chương 1: <b>Đại cương về các hệ cơ sở dữ liệu</b>	Nắm được các nhược điểm của hệ thống lưu theo file truyền thống; Nắm được các hệ CSDL: khái niệm, khả năng, kiến trúc, người dùng của một hệ quản trị CSDL ; Nắm được khái niệm ngôn ngữ cơ sở dữ liệu;	Phân tích được các bước thiết kế một cơ sở dữ liệu. Một số hệ quản trị cơ sở dữ liệu phổ biến	
Chương 2: <b>Mô hình dữ liệu</b>	Nắm được các khái niệm cơ bản của mô hình thực thể liên kết và mô hình quan hệ; Mô hình phân cấp; Mô hình hướng đối tượng	So sánh và đánh giá các mô hình dữ liệu; Hiểu được các bước xây dựng một hệ CSDL; Nắm được cách ánh xạ từ mô hình thực thể liên kết sang mô hình quan hệ	Biết thiết kế mô hình thực thể liên kết và mô hình quan hệ cho bài toán trong thực tế

## Nội dung & Mục tiêu học phần

Chương 3: <b>Ngôn ngữ định nghĩa và thao tác dữ liệu đối với mô hình quan hệ</b>	Nắm được các phép toán của đại số quan hệ; Nắm được các thành phần của SQL; Nắm được Ngôn ngữ định nghĩa dữ liệu (DDL); Ngôn ngữ thao tác dữ liệu (DML); cú pháp các câu lệnh SQL	Biết cách biểu diễn các truy vấn cho trước bằng đại số quan hệ; Biết cách chuyển đổi câu truy vấn từ đại số quan hệ sang câu lệnh SQL	Biết cách sử dụng SQL để định nghĩa CSDL và thao tác dữ liệu
Chương 4: <b>Lý thuyết thiết kế cơ sở dữ liệu quan hệ</b>	Hiểu được các vấn đề đối với CSDL; Nắm được các khái niệm chung liên quan tới phụ thuộc hàm: Hệ tiên đề Armstrong đối với phụ thuộc hàm, bao đóng, phủ của một tập phụ thuộc hàm; Nắm được khái niệm các dạng chuẩn 1,2,3, Boye Codd; Nắm được lý do tại sao cần chuẩn hóa dữ liệu và sự cần thiết bảo toàn thông tin và phụ thuộc hàm	Nắm được cách sử dụng thuật toán tìm bao đóng của một tập thuộc tính, thuật toán tìm phủ tối thiểu của một tập phụ thuộc hàm ; Nắm được cách kiểm tra bảo toàn phụ thuộc hàm; Nắm được cách kiểm tra CSDL quan hệ đã ở dạng 1NF, 2NF, 3NF, BCNF chưa ; Nắm được cách phân tách một quan hệ để đưa về dạng chuẩn	Hiểu được tại sao lại cần các khái niệm bao đóng, phủ tối thiểu. Biết cách xác định phụ thuộc hàm trong hệ CSDL thực thể. Biết thiết kế một CSDL quan hệ ở dạng chuẩn 3NF hoặc BCNF

- **Học liệu bắt buộc:**

- ❖ [1] Bài giảng CSDL, Học viện công nghệ bưu chính viễn thông
- ❖ [2] Ramakrishna, R. (2004). Database Management Systems 2nd Ed.

- *Học liệu tham khảo:*

- ❖ [3] Database Systems: A Practical Approach to Design, Implementation, and Management By T. Connolly, and C. Begg, 6th edition, Addison-Wesley, 2014, ISBN: 0-132-94326-3
- ❖ [4] Silberschatz, Korth and Sudarshan, Database System Concepts (6th Edition), McGrawHill, 2010
- ❖ [5] Poulton, N. (2014). Data Storage Networking: Real World Skills for the CompTIA Storage+ Certification and Beyond. John Wiley & Sons.

- ❖ Tham dự lớp đầy đủ, tích cực xây dựng bài
- ❖ Kiểm tra giữa kỳ
- ❖ Kiểm tra cuối kỳ

Hình thức kiểm tra	Tỷ lệ đánh giá	Đặc điểm đánh giá
- Tham gia học tập trên lớp (đi học đầy đủ, tích cực thảo luận)	10%	Cá nhân
- Bài tập lớn theo nhóm	20%	Nhóm
- Kiểm tra giữa kỳ	10%	Cá nhân
- Thi cuối kỳ	60%	Cá nhân

- ❖ Bài tập lớn phải làm theo nhóm từ 2-3 người và cần nộp đúng hạn. Nếu không đúng hạn sẽ bị 0 điểm;
- ❖ Thiếu 1 điểm thành phần (bài tập lớn, bài kiểm tra giữa kỳ), hoặc nghỉ quá 20% tổng số giờ của môn học, không được thi hết môn.



- www.ptit.edu.vn**



1. Tổng quan về lý thuyết CSDL: CSDL, hệ CSDL và hệ quản trị CSDL
2. Kiến trúc và người dùng của hệ quản trị CSDL
3. Phân loại các hệ CSDL: lịch sử phát triển, những ưu thế của CSDL quan hệ, mô hình QH đang phổ biến hiện nay.
4. Ngôn ngữ CSDL

Chapter 1	Introduction to Databases
1.1	Introduction
1.2	Traditional File-Based Systems
1.2.1	File-Based Approach
1.2.2	Limitations of the File-Based Approach
1.3	Database Approach
1.3.1	The Database
1.3.2	The Database Management System (DBMS)
1.3.3	(Database) Application Programs
1.3.4	Components of the DBMS Environment
1.3.5	Database Design: The Paradigm Shift
1.4	Roles in the Database Environment
1.4.1	Data and Database Administrators
1.4.2	Database Designers
1.4.3	Application Developers
1.4.4	End-Users
1.5	History of Database Management Systems
1.6	Advantages and Disadvantages of DBMSs

# 1. Tổng quan về lý thuyết CSDL

## 1.1 Giới thiệu

- ❖ CSDL và hệ thống CSDL là một **thành phần thiết yếu** của cuộc sống trong xã hội hiện đại: hầu hết chúng ta đều gặp phải một số hoạt động hàng ngày liên quan đến một số tương tác với CSDL.
- ❖ Ví dụ:
  - Hệ thống ngân hàng: Gửi tiền hoặc rút tiền, vay tiền
  - Hệ thống quản lý khách sạn: Đặt/trả phòng
  - Hệ thống thư viện: truy cập danh mục thư viện trên máy tính để tìm kiếm tài liệu
  - Hệ thống thương mại điện tử: mua hàng trực tuyến, tự động cập nhật CSDL chứa hàng tồn kho của các mặt hàng tạp hóa
  - Hệ thống hàng không,...

là các hoạt động sẽ liên quan đến ai đó hoặc chương trình máy tính nào đó truy cập CSDL.



# 1. Tổng quan về lý thuyết CSDL

## 1.1 Giới thiệu

### ❖ Ví dụ: Mua hàng từ siêu thị

Khi bạn mua hàng từ siêu thị địa phương, CSDL có thể được truy cập.

Trợ lý thanh toán sử dụng đầu đọc mã vạch để quét từng giao dịch mua hàng của bạn. Điều này được liên kết với một chương trình ứng dụng sử dụng mã vạch để tìm ra giá của mặt hàng từ CSDL sản phẩm.

Sau đó, chương trình sẽ giảm số lượng mặt hàng đó trong kho và hiển thị giá trên máy tính tiền.

Nếu mức đặt hàng lại giảm xuống dưới ngưỡng quy định, hệ thống CSDL có thể tự động đặt hàng để có thêm lượng hàng tồn kho của mặt hàng đó.

Nếu khách hàng gọi điện đến siêu thị, người trợ lý có thể kiểm tra xem mặt hàng đó còn hàng hay không bằng cách chạy một chương trình ứng dụng xác định tình trạng còn hàng từ CSDL.

### ❖ Ví dụ khác: Mua hàng bằng thẻ tín dụng; Đặt kỳ nghỉ tại các đại lý du lịch,...

# 1. Tổng quan về lý thuyết CSDL

## 1.1 Giới thiệu

- ❖ Những tương tác này là ví dụ về những gì chúng ta có thể gọi là các **ứng dụng CSDL truyền thống**, trong đó hầu hết thông tin được lưu trữ và truy cập là văn bản hoặc số.
- ❖ **Ứng dụng CSDL chỉ đơn giản là một chương trình tương tác với CSDL tại một thời điểm nào đó trong quá trình thực thi.**
- ❖ Nhiều năm qua, những tiến bộ trong công nghệ đã dẫn đến những **ứng dụng mới của hệ thống CSDL**. Sự phát triển nhanh chóng của các trang web truyền thông xã hội, chẳng hạn như Facebook, Twitter và Flickr, cùng nhiều trang khác, đã yêu cầu tạo ra các CSDL khổng lồ lưu trữ dữ liệu phi truyền thống, chẳng hạn như các bài đăng, tweet (bài viết ngắn trên mạng xã hội Twitter, có giới hạn 280 ký tự), hình ảnh và video clip.

# 1. Tổng quan về lý thuyết CSDL

## 1.1 Giới thiệu

- ❖ Các loại hệ thống CSDL mới, thường được gọi là **hệ thống lưu trữ dữ liệu lớn** hoặc hệ thống NOSQL, đã được tạo để quản lý dữ liệu cho các ứng dụng truyền thông xã hội.
- ❖ Những loại hệ thống này cũng được các công ty như Google, Amazon và Yahoo sử dụng để quản lý dữ liệu cần thiết trong các công cụ tìm kiếm Web của họ cũng như cung cấp lưu trữ đám mây, nhờ đó người dùng được cung cấp khả năng lưu trữ trên Web để quản lý tất cả loại dữ liệu bao gồm tài liệu, chương trình, hình ảnh, video và email.
- ❖ Rõ ràng: CSDL và công nghệ CSDL đã có tác động lớn đến việc sử dụng máy tính ngày càng tăng. Công bằng mà nói thì **CSDL đóng vai trò quan trọng trong hầu hết các lĩnh vực sử dụng máy tính, bao gồm kinh doanh, thương mại điện tử, truyền thông xã hội, kỹ thuật, y học, di truyền học, luật, giáo dục và khoa học thư viện.**
- ❖ Vì CSDL từ được sử dụng phổ biến như vậy nên để hiểu sâu CSDL là gì, đi theo **cách tiếp cận: Hệ thống dựa trên tệp truyền thống → Tiếp cận CSDL**

# 1. Tổng quan về lý thuyết CSDL

## 1.2 Các hệ thống dựa trên tệp truyền thống

- ❖ <https://www.geeksforgeeks.org/traditional-file-system/>
- ❖ Trước đây, các ứng dụng CSDL được xây dựng trên hệ thống tệp. Cách tiếp cận này hầu hết đã lỗi thời nhưng -Việc hiểu các vấn đề cố hữu trong các hệ thống dựa trên tệp có thể **giúp tránh lặp lại những vấn đề này trong hệ thống CSDL của mình. Hiểu cách hoạt động của hệ thống tệp** sẽ giúp ích khi chuyển đổi hệ thống dựa trên tệp sang hệ thống CSDL.
- ❖ Hệ thống dựa trên tệp là nỗ lực ban đầu nhằm **tin học hóa hệ thống lưu trữ thủ công** mà tất cả chúng ta đều quen thuộc.
- ❖ **Ví dụ:** trong một tổ chức, một tệp thủ công được thiết lập để chứa tất cả thư từ bên ngoài và nội bộ liên quan đến một dự án, sản phẩm, nhiệm vụ, khách hàng hoặc nhân viên. Thông thường, có nhiều tệp tin như vậy và để đảm bảo an toàn, chúng được dán nhãn và lưu trữ trong một hoặc nhiều tủ. Để đảm bảo an ninh, tủ có thể có khóa hoặc có thể được đặt ở những khu vực an toàn của tòa nhà.

# 1. Tổng quan về lý thuyết CSDL

## 1.2 Các hệ thống dựa trên tệp truyền thống

- ❖ <https://www.geeksforgeeks.org/traditional-file-system/>
- ❖ **Hệ thống tệp** là tập hợp dữ liệu. Trong hệ thống này, người dùng phải viết các thủ tục quản lý CSDL. Nó cung cấp chi tiết về biểu diễn dữ liệu và lưu trữ dữ liệu. Trong đó:
  - Dữ liệu được lưu trữ trong các tệp tin.
  - Mỗi tệp tin có định dạng cụ thể.
  - Các chương trình sử dụng các tệp này phụ thuộc vào kiến thức về định dạng đó.
- ❖ Trước đây, các ứng dụng CSDL được xây dựng trên hệ thống tệp. Cách tiếp cận này hầu hết đã lỗi thời nhưng -Việc hiểu các vấn đề cố hữu trong các hệ thống dựa trên tệp có thể **giúp tránh lặp lại những vấn đề này trong hệ thống CSDL của mình**. Hiểu cách hoạt động của hệ thống tệp sẽ giúp ích khi chuyển đổi hệ thống dựa trên tệp sang hệ thống CSDL.

# 1. Tổng quan về lý thuyết CSDL

## 1.2 Các hệ thống dựa trên tệp truyền thống

Trong nhà riêng của chúng ta:

- ❖ Có thể có **một số loại hệ thống lưu trữ hồ sơ** bao gồm các biên lai, bảo lãnh, hóa đơn, báo cáo ngân hàng, v.v. Khi cần tra cứu thứ gì đó, chúng ta vào hệ thống lưu trữ và tìm kiếm trong hệ thống bắt đầu từ mục nhập đầu tiên cho đến khi tìm thấy thứ mình muốn.
- ❖ Ngoài ra, có thể có **một hệ thống lập chỉ mục** giúp xác định những gì chúng ta muốn một cách nhanh chóng hơn. Ví dụ: có thể **phân chia trong hệ thống lưu trữ hoặc các thư mục riêng biệt cho các loại mục khác nhau có liên quan về mặt logic theo cách nào đó.**
- ❖ Hệ thống lưu trữ thủ công **hoạt động tốt trong khi số lượng hồ sơ cần lưu trữ ít, thậm chí còn khá hiệu quả khi có số lượng lớn vật phẩm và chỉ phải lưu trữ và lấy chúng.**



# 1. Tổng quan về lý thuyết CSDL

## 1.2 Các hệ thống dựa trên tệp truyền thống

- ❖ Tuy nhiên, hệ thống lưu trữ thủ công **bị lỗi khi chúng ta phải đối chiếu chéo hoặc xử lý thông tin trong hồ sơ.**
- ❖ Ví dụ: **văn phòng đại lý bất động sản điển hình có thể có một tệp riêng cho từng tài sản để bán hoặc cho thuê, từng người mua và người thuê tiềm năng cũng như từng nhân viên.** Hãy xem xét để trả lời các câu hỏi sau:
  - Bạn có căn hộ ba phòng ngủ nào có sân vườn và gara để bán?
  - Những căn hộ nào bạn có cho thuê trong vòng ba km từ trung tâm thành phố?
  - Giá thuê trung bình cho một căn hộ hai phòng ngủ là bao nhiêu?
  - Tổng hóa đơn tiền lương hàng năm cho nhân viên là bao nhiêu?
  - Doanh thu tháng trước so với dự kiến tháng này như thế nào?
  - Doanh thu hàng tháng dự kiến trong năm tài chính tiếp theo là bao nhiêu?

# 1. Tổng quan về lý thuyết CSDL

## 1.2 Các hệ thống dựa trên tệp truyền thống

Ví dụ: **văn phòng đại lý bất động sản**

- ❖ Ngày nay, khách hàng, quản lý cấp cao và nhân viên ngày càng muốn có nhiều thông tin hơn. Ở một số khu vực, pháp luật yêu cầu phải lập báo cáo chi tiết hàng tháng, hàng quý và hàng năm.
- ❖ Rõ ràng, hệ thống thủ công không phù hợp với loại công việc này. Hệ thống dựa trên tệp tin được phát triển để đáp ứng nhu cầu truy cập dữ liệu hiệu quả hơn của ngành. Tuy nhiên, thay vì thiết lập một kho lưu trữ tập trung cho dữ liệu hoạt động của tổ chức, một cách tiếp cận phi tập trung đã được thực hiện, trong đó mỗi bộ phận, với sự hỗ trợ của nhân viên Xử lý Dữ liệu (DP), lưu trữ và kiểm soát dữ liệu của chính mình.
- ❖ Để hiểu điều này có nghĩa là gì, hãy xem xét ví dụ DreamHome.

# 1. Tổng quan về lý thuyết CSDL

## 1.2 Các hệ thống dựa trên tệp truyền thống: DreamHome

- ❖ **Phòng Kinh doanh:** chịu trách nhiệm về việc bán và cho thuê tài sản.
- ❖ Ví dụ: bất cứ khi nào khách hàng tiếp cận Phòng Kinh doanh nhằm tiếp thị tài sản cho thuê của mình, một biểu mẫu sẽ được hoàn thành, tương tự như biểu mẫu trong Hình 1.1(a). Điều này cung cấp thông tin chi tiết về tài sản như địa chỉ và số phòng cùng với thông tin chi tiết về chủ sở hữu.

DreamHome Property for Rent Details Property Number: <u>FG21</u>	
Address <u>18 Dale Rd</u> City <u>Glasgow</u> Postcode <u>G12</u> Type <u>House</u> Rent <u>600</u> No of Rooms <u>5</u>	Allocated to Branch: <u>163 Main St, Glasgow</u> Branch No <u>B003</u>  Staff Responsible <u>Ann Beech</u>
Owner's Details	
Name <u>Carol Farrel</u> Address <u>6 Achray St,</u> <u>Glasgow G32 9DX</u> Tel No. <u>0141-357-7419</u> Owner No. <u>G087</u>	Business Name _____ Address _____ Tel No. _____ Owner No. _____ Contact Name _____ Business Type _____

# 1. Tổng quan về lý thuyết CSDL

## 1.2 Các hệ thống dựa trên tập truyền thống: DreamHome

❖ Phòng Kinh doanh cũng xử lý các yêu cầu từ khách hàng và một biểu mẫu tương tự như trong Hình 1.1(b) được hoàn thành cho mỗi khách hàng.

<b>DreamHome</b> <b>Client Details</b> <b>Client Number:</b> <u>CR74</u>	
<b>First Name</b> <u>Mike</u>	<b>Last Name</b> <u>Ritchie</u>
<b>Address</b> <u>18 Tain St.</u> <u>PAIG TNG</u>	<b>Tel No.</b> <u>01475-392178</u>
<b>Property Requirement Details</b>	
<b>Preferred Property Type</b> <u>House</u>	<b>Maximum Monthly Rent</b> <u>750</u>
<b>General Comments</b> <u>Currently living at home with parents</u> <u>Getting married in August</u>	
<b>Seen By</b> <u>Ann Bosch</u>	<b>Date</b> <u>24-Mar-04</u>
<b>Branch No</b> <u>B003</u>	<b>Branch City</b> <u>Glasgow</u>

# 1. Tổng quan về lý thuyết CSDL

## 1.2 Các hệ thống dựa trên tệp truyền thống: DreamHome

- ❖ Với sự hỗ trợ của Phòng DP, Phòng Kinh doanh **xây dựng hệ thống thông tin xử lý việc cho thuê bất động sản.**
- ❖ Hệ thống **bao gồm ba tệp chứa thông tin chi tiết về: tài sản, chủ sở hữu và khách hàng**, như minh họa trong Hình 1.2.

(Để đơn giản, bỏ qua các chi tiết liên quan đến nhân viên, văn phòng chi nhánh và chủ doanh nghiệp.)

PropertyForRent

propertyNo	street	city	postcode	type	rooms	rent	ownerNo
PA14	16 Holhead	Aberdeen	AB7 5SU	House	6	650	CO46
PL94	6 Argyll St	London	NW2	Flat	4	400	CO87
PG4	6 Lawrence St	Glasgow	G11 9QX	Flat	3	350	CO40
PG36	2 Manor Rd	Glasgow	G32 4QX	Flat	3	375	CO93
PG21	18 Dale Rd	Glasgow	G12	House	5	600	CO87
PG16	5 Novar Dr	Glasgow	G12 9AX	Flat	4	450	CO93

# 1. Tổng quan về lý thuyết CSDL

## 1.2 Các hệ thống dựa trên tệp truyền thống: DreamHome

- ❖ Với sự hỗ trợ của Phòng DP, Phòng Kinh doanh **xây dựng hệ thống thông tin xử lý việc cho thuê bất động sản.**
- ❖ Hệ thống **bao gồm ba tệp chứa thông tin chi tiết về: tài sản, chủ sở hữu và khách hàng**, như minh họa trong Hình 1.2.

(Để đơn giản, bỏ qua các chi tiết liên quan đến nhân viên, văn phòng chi nhánh và chủ doanh nghiệp.)

PrivateOwner

ownerNo	fName	lName	address	telNo
CO46	Joe	Keogh	2 Fergus Dr, Aberdeen AB2 7SX	01224-861212
CO87	Carol	Farrel	6 Achray St, Glasgow G32 9DX	0141-357-7419
CO40	Tina	Murphy	63 Well St, Glasgow G42	0141-943-1728
CO93	Tony	Shaw	12 Park Pl, Glasgow G4 0QR	0141-225-7025



# 1. Tổng quan về lý thuyết CSDL

## 1.2 Các hệ thống dựa trên tập truyền thống: DreamHome

❖ **Phòng Hợp đồng:** có trách nhiệm xử lý các hợp đồng cho thuê liên quan đến tài sản cho thuê. Bất cứ khi nào khách hàng đồng ý thuê một bất động sản, một trong các nhân viên Bán hàng sẽ điền vào một biểu mẫu để cung cấp cho khách hàng thông tin chi tiết về bất động sản, như trong Hình 1.3.

DreamHome Lease Details Lease Number: <u>10012</u>	
<b>Client No.</b> <u>CR74</u> <b>Full Name</b> <u>Mike Ritchie</u> <b>Address (previous)</b> <u>18 Tain St,</u> <u>PA1G 1YQ</u> <b>Tel No.</b> <u>01475-392178</u>	<b>Property No.</b> <u>PG21</u> <b>Address</b> <u>18 Dale Rd,</u> <u>Glasgow G12</u>
<b>Payment Details</b>	
<b>Monthly Rent</b> <u>600</u> <b>Payment Method</b> <u>Cheque</u> <b>Deposit</b> <u>1200</u> <b>Paid (Y or N)</b> <u>Y</u>	<b>Rent Start Date</b> <u>1-Jul-04</u> <b>Rent Finish Date</b> <u>30-Jun-05</u> <b>Duration</b> <u>1Year</u>

# 1. Tổng quan về lý thuyết CSDL

## 1.2 Các hệ thống dựa trên tệp truyền thống: DreamHome

### Phòng Hợp đồng:

- ❖ Biểu mẫu này được chuyển đến Phòng Hợp đồng để phân bổ số hợp đồng thuê và hoàn thành các chi tiết thanh toán và thời gian thuê.
- ❖ Với sự hỗ trợ của Phòng DP, Phòng Hợp đồng tạo ra một hệ thống thông tin để xử lý các hợp đồng cho thuê.
- ❖ Hệ thống bao gồm ba tệp lưu trữ thông tin chi tiết về hợp đồng thuê, tài sản và khách hàng, chứa dữ liệu tương tự như dữ liệu do Phòng Kinh doanh nắm giữ, như minh họa trong Hình 1.4.

DreamHome Lease Details Lease Number: <u>10012</u>	
Client No. <u>CR74</u> Full Name <u>Mike Ritchie</u> Address (previous) <u>18 Tain St,</u> <u>PA1G 1YQ</u> Tel No. <u>01475-392178</u>	Property No. <u>PG21</u> Address <u>18 Dale Rd,</u> <u>Glasgow G12</u>
Payment Details	
Monthly Rent <u>600</u> Payment Method <u>Cheque</u> Deposit <u>1200</u> Paid (Y or N) <u>Y</u>	Rent Start Date <u>1-Jul-04</u> Rent Finish Date <u>30-Jun-05</u> Duration <u>1 Year</u>

# 1. Tổng quan về lý thuyết CSDL

## 1.2 Các hệ thống dựa trên tệp truyền thống - Phòng Hợp đồng:

- ❖ Hệ thống bao gồm ba tệp lưu trữ thông tin chi tiết về hợp đồng thuê, tài sản và khách hàng, chứa dữ liệu tương tự như dữ liệu do Phòng Kinh doanh nắm giữ, như minh họa trong Hình 1.4.

Lease

leaseNo	propertyNo	clientNo	rent	payment Method	deposit	paid	rentStart	rentFinish	duration
10024	PA14	CR62	650	Visa	1300	Y	1-Jun-05	31-May-05	12
10075	PL94	CR76	400	Cash	800	N	1-Aug-05	31-Jan-05	6
10012	PG21	CR74	600	Cheque	1200	Y	1-Jul-05	30-Jun-05	12

PropertyForRent

propertyNo	street	city	postcode	rent
PA14	16 Holhead	Aberdeen	AB7 5SU	650
PL94	6 Argyll St	London	NW2	400
PG21	18 Dale Rd	Glasgow	G12	600

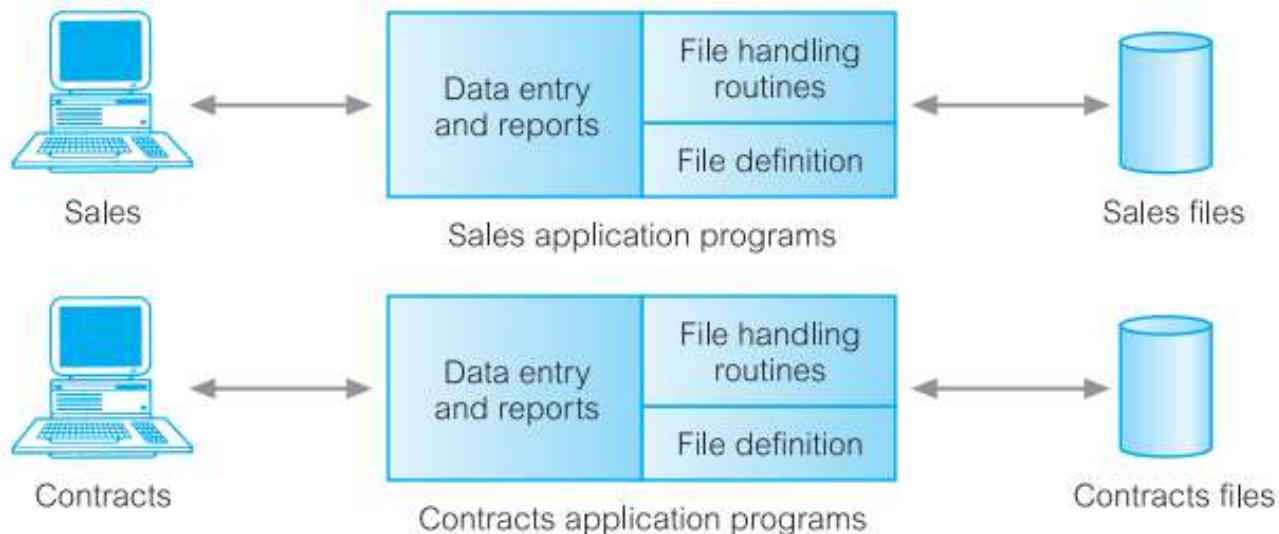
Client

clientNo	fName	lName	address	telNo
CR76	John	Kay	56 High St, London SW1 4EH	0171-774-5632
CR74	Mike	Ritchie	18 Tain St, PA1G 1YQ	01475-392178
CR62	Mary	Tregear	5 Tarbot Rd, Aberdeen AB9 3ST	01224-196720

# 1. Tổng quan về lý thuyết CSDL

## Các hệ thống dựa trên tệp truyền thống: DreamHome

❖ Tình huống này được minh họa trong Hình 1.5.



**Figure 1.5**  
File-based  
processing.

### Sales Files

**PropertyForRent** (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)

**PrivateOwner** (ownerNo, fName, lName, address, telNo)

**Client** (clientNo, fName, lName, address, telNo, prefType, maxRent)

### Contracts Files

**Lease** (leaseNo, propertyNo, clientNo, rent, paymentMethod, deposit, paid, rentStart, rentFinish, duration)

**PropertyForRent** (propertyNo, street, city, postcode, rent)

**Client** (clientNo, fName, lName, address, telNo)

# 1. Tổng quan về lý thuyết CSDL

## Các hệ thống dựa trên tệp truyền thống: DreamHome

Tình huống này được minh họa trong Hình 1.5. Cho thấy:

- ❖ Mỗi bộ phận đang truy cập các tệp tin riêng của họ thông qua các chương trình ứng dụng được viết riêng cho họ.
- ❖ Mỗi bộ chương trình ứng dụng của phòng ban xử lý việc nhập dữ liệu, bảo trì tệp và tạo một bộ báo cáo cụ thể cố định.
- ❖ Điều quan trọng hơn là **cấu trúc vật lý và việc lưu trữ các tệp dữ liệu và bản ghi được xác định trong mã ứng dụng.**
- ❖ Có thể tìm thấy những ví dụ tương tự ở các bộ phận khác. Ví dụ:
  - Phòng Tiền lương lưu trữ các thông tin chi tiết liên quan đến tiền lương của từng nhân viên, cụ thể là:
    - StaffSalary(staffNo, fName, lName, sex, salary, branchNo)
  - Phòng Nhân sự còn lưu trữ thông tin chi tiết của nhân viên, cụ thể:
    - Staff(staffNo, fName, lName, position, sex, dateOfBirth, salary, branchNo)

# 1. Tổng quan về lý thuyết CSDL

## Các hệ thống dựa trên tệp truyền thống:

- ❖ Có thể thấy khá rõ ràng rằng: **có một lượng dữ liệu trùng lặp đáng kể trong các bộ phận này** và điều này nói chung đúng với các hệ thống dựa trên tệp.
- ❖ Trước khi chúng ta thảo luận về những hạn chế của phương pháp này, có thể **hữu ích nếu hiểu được thuật ngữ được sử dụng trong các hệ thống dựa trên tệp tin**:
- ❖ Một **tệp** chỉ đơn giản là một tập hợp các bản ghi chứa dữ liệu liên quan đến logic. Ví dụ: tệp PropertyForRent trong Hình 1.2 chứa sáu bản ghi, một bản ghi cho mỗi thuộc tính.

PropertyForRent

propertyNo	street	city	postcode	type	rooms	rent	ownerNo
PA14	16 Holhead	Aberdeen	AB7 5SU	House	6	650	CO46
PL94	6 Argyll St	London	NW2	Flat	4	400	CO87
PG4	6 Lawrence St	Glasgow	G11 9QX	Flat	3	350	CO40
PG36	2 Manor Rd	Glasgow	G32 4QX	Flat	3	375	CO93
PG21	18 Dale Rd	Glasgow	G12	House	5	600	CO87
PG16	5 Novar Dr	Glasgow	G12 9AX	Flat	4	450	CO93



# 1. Tổng quan về lý thuyết CSDL

## Các hệ thống dựa trên tệp truyền thống:

Ví dụ: tệp PropertyForRent trong Hình 1.2 chứa sáu bản ghi, một bản ghi cho mỗi thuộc tính.

Mỗi bản ghi chứa một tập hợp được kết nối logic gồm một hoặc nhiều trường, trong đó mỗi trường đại diện cho một số đặc điểm của đối tượng trong thế giới thực đang được mô hình hóa. Trong Hình 1.2, các trường của tệp PropertyForRent thể hiện các đặc điểm của thuộc tính, chẳng hạn như địa chỉ, loại thuộc tính và số phòng.

PropertyForRent

propertyNo	street	city	postcode	type	rooms	rent	ownerNo
PA14	16 Holhead	Aberdeen	AB7 5SU	House	6	650	CO46
PL94	6 Argyll St	London	NW2	Flat	4	400	CO87
PG4	6 Lawrence St	Glasgow	G11 9QX	Flat	3	350	CO40
PG36	2 Manor Rd	Glasgow	G32 4QX	Flat	3	375	CO93
PG21	18 Dale Rd	Glasgow	G12	House	5	600	CO87
PG16	5 Novar Dr	Glasgow	G12 9AX	Flat	4	450	CO93

# 1. Tổng quan về lý thuyết CSDL

## Các hệ thống dựa trên tệp truyền thống: DreamHome

- ❖ Với sự hỗ trợ của Phòng DP, Phòng Kinh doanh **xây dựng hệ thống thông tin xử lý việc cho thuê bất động sản.**
- ❖ Hệ thống **bao gồm ba tệp chứa thông tin chi tiết về: tài sản, chủ sở hữu và khách hàng**, như minh họa trong Hình 1.2.

(Để đơn giản, bỏ qua các chi tiết liên quan đến nhân viên, văn phòng chi nhánh và chủ doanh nghiệp.)

Client

clientNo	fName	lName	address	telNo	prefType	maxRent
CR76	John	Kay	56 High St, London SW1 4EH	0207-774-5632	Flat	425
CR56	Aline	Stewart	64 Fern Dr, Glasgow G42 0BL	0141-848-1825	Flat	350
CR74	Mike	Ritchie	18 Tain St, PA1G 1YQ	01475-392178	House	750
CR62	Mary	Tregear	5 Tarbot Rd, Aberdeen AB9 3ST	01224-196720	Flat	600

# 1. Tổng quan về lý thuyết CSDL

## Hạn chế của phương pháp tiếp cận dựa trên tệp

### ❖ Tách biệt và cô lập dữ liệu

- Khi dữ liệu bị tách biệt trong các tệp riêng biệt, việc truy cập dữ liệu có sẵn sẽ khó khăn hơn.
- Ví dụ: nếu muốn tạo danh sách tất cả các ngôi nhà phù hợp với yêu cầu của khách hàng, trước tiên cần tạo một tệp tạm thời gồm những khách hàng có prefType là 'house'. Sau đó, tìm kiếm tệp PropertyForRent để tìm những tài sản có loại tài sản là 'house' và giá thuê thấp hơn giá thuê tối đa của khách hàng.
- Với hệ thống tập tin, việc xử lý như vậy rất khó khăn. Nhà phát triển ứng dụng phải đồng bộ hóa việc xử lý hai tệp để đảm bảo dữ liệu được trích xuất chính xác.
- Khó khăn này càng tăng thêm nếu chúng ta yêu cầu dữ liệu từ nhiều hơn hai tệp.

Client

clientNo	fName	lName	address	telNo	prefType	maxRent
CR76	John	Kay	56 High St, London SW1 4EH	0207-774-5632	Flat	425
CR56	Aline	Stewart	64 Fern Dr, Glasgow G42 0BL	0141-848-1825	Flat	350
CR74	Mike	Ritchie	18 Tain St, PA1G 1YQ	01475-392178	House	750
CR62	Mary	Tregear	5 Tarbot Rd, Aberdeen AB9 3ST	01224-196720	Flat	600

# 1. Tổng quan về lý thuyết CSDL

## Hạn chế của phương pháp tiếp cận dựa trên tệp

### ❖ Trùng lặp dữ liệu

Do cách tiếp cận phi tập trung được thực hiện bởi mỗi bộ phận, cách tiếp cận dựa trên tệp đã làm cho việc sao chép dữ liệu không được kiểm soát.

Ví dụ, trong Hình 1.5, chúng ta có thể thấy rõ rằng có sự trùng lặp về cả chi tiết tài sản và khách hàng trong Phòng Bán hàng và Hợp đồng.

#### Sales Files

**PropertyForRent** (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)

**PrivateOwner** (ownerNo, fName, lName, address, telNo)

**Client** (clientNo, fName, lName, address, telNo, prefType, maxRent)

#### Contracts Files

**Lease** (leaseNo, propertyNo, clientNo, rent, paymentMethod, deposit, paid, rentStart, rentFinish, duration)

**PropertyForRent** (propertyNo, street, city, postcode, rent)

**Client** (clientNo, fName, lName, address, telNo)

# 1. Tổng quan về lý thuyết CSDL

## Hạn chế của phương pháp tiếp cận dựa trên tệp

### ❖ Trùng lặp dữ liệu

Việc sao chép dữ liệu không được kiểm soát là điều không mong muốn vì một số lý do, bao gồm:

Trùng lặp là lãng phí. Việc nhập dữ liệu nhiều lần sẽ tốn thời gian và tiền bạc.

Nó chiếm thêm không gian lưu trữ và kéo theo các chi phí liên quan. Thông thường, có thể tránh được việc trùng lặp dữ liệu bằng cách chia sẻ tệp dữ liệu.

Trùng lặp có thể dẫn đến **mất tính toàn vẹn dữ liệu; dữ liệu không còn nhất quán** nữa.

# 1. Tổng quan về lý thuyết CSDL

## Hạn chế của phương pháp tiếp cận dựa trên tệp

### ❖ Trùng lặp dữ liệu

Ví dụ: Xét sự trùng lặp dữ liệu giữa Phòng Nhân sự và Tiền lương được mô tả ở trên.

- Phòng Tiền lương lưu trữ các thông tin chi tiết liên quan đến tiền lương của từng nhân viên, cụ thể là:
  - StaffSalary(staffNo, fName, lName, sex, salary, branchNo)
- Phòng Nhân sự còn lưu trữ thông tin chi tiết của nhân viên, cụ thể:
  - Staff(staffNo, fName, lName, position, sex, dateOfBirth, salary, branchNo)

Nếu một thành viên của nhân viên chuyển nhà, việc thay đổi địa chỉ chỉ được thông báo cho Nhân sự chứ không phải cho phòng Tiền lương, phiếu lương của người đó sẽ được gửi đến địa chỉ sai.



# 1. Tổng quan về lý thuyết CSDL

## Hạn chế của phương pháp tiếp cận dựa trên tệp

### ❖ Trùng lặp dữ liệu

Nghiêm trọng hơn: nếu một nhân viên được thăng chức kèm theo tăng lương. Thay đổi này được thông báo cho Nhân sự nhưng lại không được lọc qua Tiền lương → nhân viên bị nhận lương sai.

Khi lỗi này được phát hiện thì sẽ mất rất nhiều thời gian và công sức để giải quyết.

Cả hai ví dụ này đều minh họa sự **mâu thuẫn có thể xảy ra do trùng lặp dữ liệu**. Vì không có cách nào tự động để Nhân sự cập nhật dữ liệu trong các tệp Tiền lương nên không khó để thấy trước những mâu thuẫn như vậy sẽ phát sinh.

Ngay cả khi Tiền lương được thông báo về những thay đổi, vẫn có thể dữ liệu được nhập không chính xác.

# 1. Tổng quan về lý thuyết CSDL

## Hạn chế của phương pháp tiếp cận dựa trên tệp

### ❖ **Sự phụ thuộc dữ liệu**

- Do cấu trúc vật lý và việc lưu trữ các tệp dữ liệu và bản ghi được xác định trong mã ứng dụng → rất khó thực hiện những thay đổi đối với cấu trúc hiện tại.

Ví dụ: việc tăng kích thước của trường địa chỉ PropertyForRent từ **40 lên 41 ký tự** tưởng đơn giản nhưng nó yêu cầu tạo chương trình một lần (nghĩa là một chương trình chỉ chạy một lần và sau đó có thể bị loại bỏ) chuyển đổi tệp PropertyForRent sang định dạng mới. Chương trình này phải:

- Mở tệp PropertyForRent gốc để đọc;
- Mở một tệp tin tạm thời với cấu trúc mới;
- Đọc bản ghi từ tệp gốc, chuyển đổi dữ liệu để phù hợp với cấu trúc mới và ghi nó vào tệp tạm thời. Lặp lại bước này cho tất cả các bản ghi trong tệp gốc;

# 1. Tổng quan về lý thuyết CSDL

## Hạn chế của phương pháp tiếp cận dựa trên tệp

### ❖ Sự phụ thuộc dữ liệu

- Chương trình này phải:
  - Xóa tệp PropertyForRent gốc;
  - Đổi tên tệp tạm thời thành PropertyForRent.
  - Tất cả các chương trình truy cập tệp PropertyForRent phải được sửa đổi để phù hợp với cấu trúc tệp mới. Có thể có nhiều chương trình như vậy truy cập vào tệp PropertyForRent. Vì vậy, lập trình viên cần xác định tất cả các chương trình bị ảnh hưởng, sửa đổi chúng và sau đó kiểm tra lại chúng.
  - Lưu ý rằng một chương trình thậm chí không cần phải sử dụng trường địa chỉ để bị ảnh hưởng: nó chỉ phải sử dụng tệp PropertyForRent. Rõ ràng, điều này có thể rất tốn thời gian và có thể xảy ra lỗi. Đặc điểm này của các hệ thống dựa trên tệp được gọi là sự phụ thuộc dữ liệu-chương trình.
- Định dạng tệp không tương thích
- Truy vấn cố định/sự phổ biến của các chương trình ứng dụng

# 1. Tổng quan về lý thuyết CSDL

## Hạn chế của phương pháp tiếp cận dựa trên tệp

### ❖ Định dạng tệp không tương thích

- Vì cấu trúc của tệp được nhúng trong các chương trình ứng dụng nên cấu trúc phụ thuộc vào ngôn ngữ lập trình ứng dụng.
- Ví dụ: cấu trúc của tệp do chương trình COBOL tạo ra có thể khác với cấu trúc của tệp do chương trình 'C' tạo ra. Sự không tương thích trực tiếp của các tệp như vậy khiến chúng khó xử lý chung.
- Ví dụ: giả sử Phòng Hợp đồng muốn tìm tên và địa chỉ của tất cả các chủ sở hữu có tài sản hiện đang được cho thuê. Nhưng Hợp đồng không lưu giữ thông tin chi tiết về chủ sở hữu tài sản; chỉ có Phòng Kinh doanh mới có thông tin này.

#### Sales Files

**PropertyForRent** (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)

**PrivateOwner** (ownerNo, fName, lName, address, telNo)

**Client** (clientNo, fName, lName, address, telNo, prefType, maxRent)

#### Contracts Files

**Lease** (leaseNo, propertyNo, clientNo, rent, paymentMethod, deposit, paid, rentStart, rentFinish, duration)

**PropertyForRent** (propertyNo, street, city, postcode, rent)

**Client** (clientNo, fName, lName, address, telNo)

# 1. Tổng quan về lý thuyết CSDL

## Hạn chế của phương pháp tiếp cận dựa trên tệp

### ❖ Định dạng tệp không tương thích

- Tuy nhiên, Hợp đồng có số tài sản (propertyNo), sử dụng số này để tìm số thuộc tính tương ứng trong tệp PropertyForRent của Phòng Kinh doanh. Tệp này chứa số chủ sở hữu (ownerNo), số này được sử dụng để tìm thông tin chi tiết về chủ sở hữu trong Tập tin PrivateOwner.
- Các chương trình của Phòng Hợp đồng trong COBOL và các chương trình của Phòng Bán hàng trong 'C'. Do đó, để khớp các trường thuộc tính không có trong hai tệp PropertyForRent yêu cầu nhà phát triển ứng dụng **phải viết phần mềm để chuyển đổi các tệp sang một số định dạng phổ biến để tạo điều kiện xử lý** → tốn kém thời gian và chi phí.

#### Sales Files

**PropertyForRent** (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)

**PrivateOwner** (ownerNo, fName, lName, address, telNo)

**Client** (clientNo, fName, lName, address, telNo, prefType, maxRent)

#### Contracts Files

**Lease** (leaseNo, propertyNo, clientNo, rent, paymentMethod, deposit, paid, rentStart, rentFinish, duration)

**PropertyForRent** (propertyNo, street, city, postcode, rent)

**Client** (clientNo, fName, lName, address, telNo)

# 1. Tổng quan về lý thuyết CSDL

## Hạn chế của phương pháp tiếp cận dựa trên tệp

### ❖ Truy vấn cố định/sự gia tăng của các chương trình ứng dụng

- Từ quan điểm của người dùng cuối, các **hệ thống dựa trên tệp đã chứng tỏ là một cải tiến lớn so với các hệ thống thủ công**. Do đó, yêu cầu về các truy vấn mới hoặc sửa đổi ngày càng tăng.
- Tuy nhiên, hệ thống dựa trên tệp phụ thuộc rất nhiều vào nhà phát triển ứng dụng - người phải viết truy vấn hoặc báo cáo được yêu cầu.
- Có hai điều đã xảy ra:
  - Ở một số tổ chức, loại truy vấn hoặc báo cáo được tạo ra là cố định. Không có phương tiện để hỏi các câu hỏi ngoài kế hoạch về bản thân dữ liệu hoặc về loại dữ liệu nào có sẵn.
  - Ở các tổ chức khác, có sự gia tăng nhanh chóng các tập tin và chương trình ứng dụng làm cho nguồn lực Phòng DP hiện tại không thể xử lý hết công việc, gây áp lực rất lớn cho nhân viên DP, dẫn đến các chương trình không đầy đủ hoặc kém hiệu quả trong việc đáp ứng nhu cầu của người dùng, tài liệu bị hạn chế và việc bảo trì gặp khó khăn.



# 1. Tổng quan về lý thuyết CSDL

## Hạn chế của phương pháp tiếp cận dựa trên tệp

### ❖ Truy vấn cố định/sự gia tăng của các chương trình ứng dụng

- Thông thường, một số chức năng nhất định đã bị bỏ qua, bao gồm:
  - Không có điều khoản về an ninh hoặc tính toàn vẹn;
  - Khả năng phục hồi, trong trường hợp xảy ra lỗi phần cứng hoặc phần mềm, bị hạn chế hoặc không tồn tại;
  - Quyền truy cập vào các tập tin bị hạn chế ở một người dùng tại một thời điểm – không có quy định nào về quyền truy cập chung của nhân viên trong cùng một bộ phận.
- Trong cả hai trường hợp, kết quả đều không được chấp nhận.  
Một giải pháp khác được yêu cầu → Tiếp cận **CSDL**

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

- ❖ Tất cả những hạn chế nêu trên của cách tiếp cận dựa trên tệp có thể là do hai yếu tố:
  - (1) định nghĩa của dữ liệu được nhúng vào các chương trình ứng dụng, thay vì được lưu trữ riêng biệt và độc lập;
  - (2) không có sự kiểm soát nào đối với việc truy cập và thao tác dữ liệu ngoài những gì do các chương trình ứng dụng áp đặt.
- ❖ Để hiệu quả hơn, cần phải có một cách tiếp cận mới. Những gì nổi lên là **CSDL và Hệ thống quản trị CSDL (DBMS)**.
- ❖ Cung cấp một định nghĩa chính thức hơn về các thuật ngữ này và kiểm tra các thành phần mà mong đợi trong môi trường DBMS.

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### CSDL

❖ Là một tập chung các dữ liệu có liên quan về mặt logic và mô tả về dữ liệu này, được thiết kế để đáp ứng nhu cầu thông tin của một tổ chức.

- Dữ liệu

- Dữ liệu là các giá trị phản ánh về sự vật, hiện tượng trong thế giới khách quan.
- Dữ liệu là các giá trị thô, chưa có ý nghĩa với người sử dụng. Có thể là một tập hợp các giá trị mà không biết được sự liên hệ giữa chúng
  - Ví dụ: Nguyễn Văn Nam, 845102, 14 / 10 / 02, 18
- Dữ liệu có thể biểu diễn dưới nhiều dạng khác nhau (âm thanh, văn bản, hình ảnh, v.v...)
- 2 loại dữ liệu:
  - Tĩnh: Tên, ngày sinh, địa chỉ...
  - Động: Doanh thu, số lượng kho...

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### CSDL

❖ Là một tập chung các dữ liệu có liên quan về mặt logic và mô tả về dữ liệu này, được thiết kế để đáp ứng nhu cầu thông tin của một tổ chức.

- Dữ liệu

- Dữ liệu về đối tượng có thể khác nhau, tùy thuộc vào ngữ cảnh.
- Ví dụ: dữ liệu về sinh viên có thể khác nhau tùy thuộc vào mục đích quản lý
  - Quản lý điểm: tên, mã sinh viên, điểm môn A, điểm môn B,...
  - Quản lý nhân thân: tên, địa chỉ, ngày sinh, quê quán,...

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### CSDL

❖ Là một tập chung các dữ liệu có liên quan về mặt logic và mô tả về dữ liệu này, được thiết kế để đáp ứng nhu cầu thông tin của một tổ chức.

- Phân biệt dữ liệu & thông tin

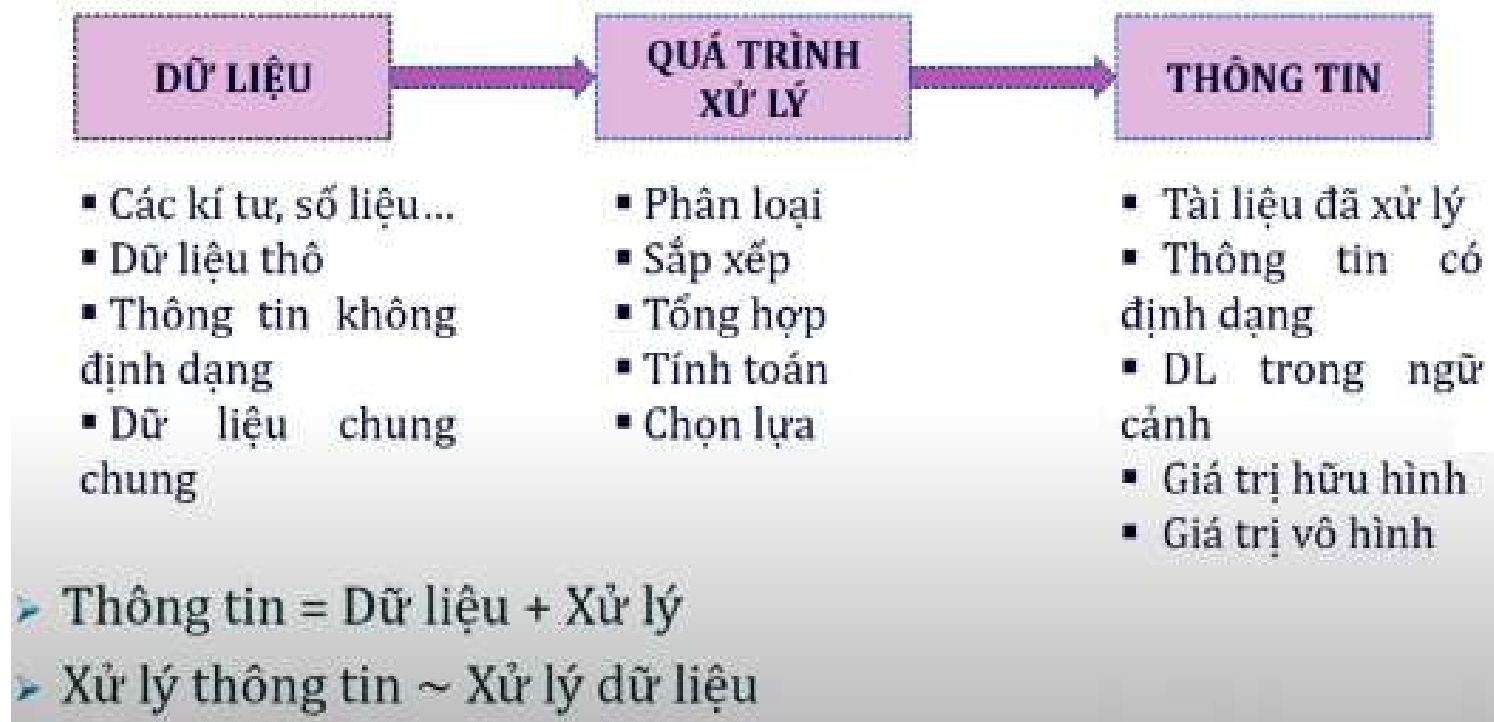
- Thông tin là ý nghĩa được rút ra từ dữ liệu thông qua quá trình xử lý (phân tích, tổng hợp, v.v.), phù hợp với mục đích cụ thể của người sử dụng.
- Thông tin có thể gồm nhiều giá trị dữ liệu được tổ chức sao cho nó mang lại một ý nghĩa cho một đối tượng cụ thể, trong một ngữ cảnh cụ thể.
  - Ví dụ với dữ liệu trên có thông tin như sau: Thủ kho Nguyễn Văn Nam xuất mặt hàng có danh mục là 845102 vào ngày 14/10/ 02 với số lượng 18.

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### CSDL

- ❖ Là **một tập chung** các dữ liệu có liên quan về mặt logic và mô tả về **dữ liệu** này, được thiết kế để đáp ứng nhu cầu thông tin của một tổ chức.
  - Phân biệt dữ liệu & thông tin





# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

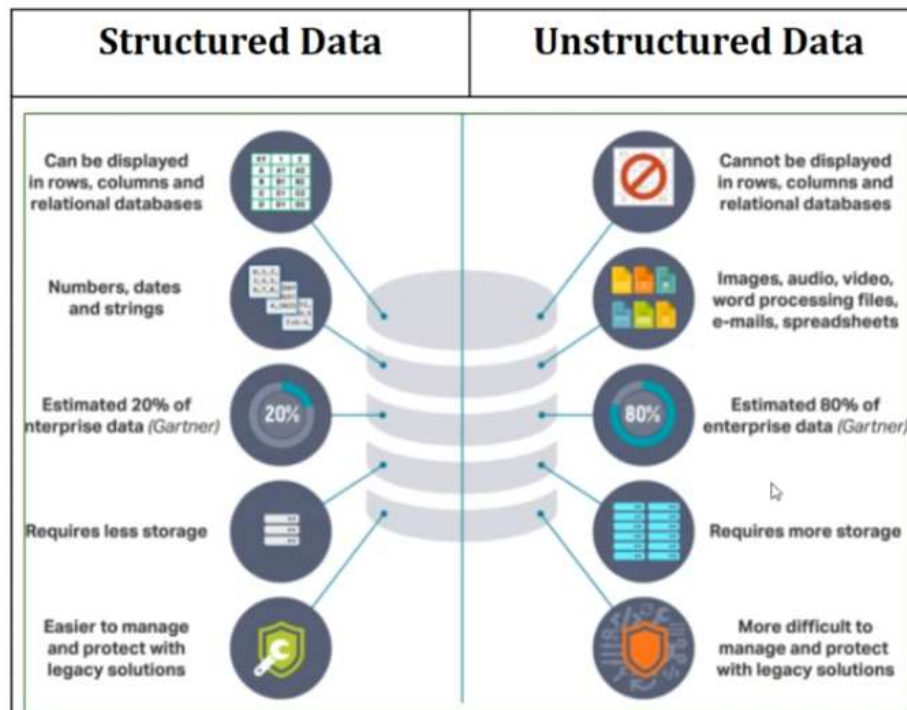
### CSDL

- ❖ Là một tập chung các dữ liệu có liên quan về mặt logic và mô tả về dữ liệu này, được thiết kế để đáp ứng nhu cầu thông tin của một tổ chức.
  - Các dạng của dữ liệu

#### ➤ Có cấu trúc



#### ➤ Phi cấu trúc



# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### CSDL

- ❖ CSDL là một kho dữ liệu lớn, duy nhất có thể được sử dụng đồng thời bởi nhiều bộ phận và người dùng. Thay vì các tệp bị ngắt kết nối với dữ liệu dư thừa, tất cả các mục dữ liệu đều được tích hợp với mức độ trùng lặp tối thiểu.
- ❖ CSDL không còn thuộc sở hữu của một bộ phận nữa mà là tài nguyên chung của công ty.
- ❖ CSDL không chỉ chứa dữ liệu hoạt động của tổ chức mà còn chứa mô tả về dữ liệu này. CSDL cũng được định nghĩa là một tập hợp các bản ghi tích hợp tự mô tả.
- ❖ Mô tả dữ liệu được gọi là danh mục hệ thống (hoặc từ điển dữ liệu hoặc siêu dữ liệu (metadata)- 'dữ liệu về dữ liệu'). Bản chất tự mô tả của CSDL cung cấp tính độc lập giữa chương trình và dữ liệu.

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### CSDL

- ❖ Cách tiếp cận được thực hiện với các hệ thống CSDL, trong đó việc xác định dữ liệu được tách biệt khỏi các chương trình ứng dụng, tương tự như cách tiếp cận được thực hiện trong phát triển phần mềm hiện đại, trong đó cung cấp định nghĩa bên trong của một đối tượng và định nghĩa bên ngoài riêng biệt.
- ❖ Người sử dụng một đối tượng chỉ nhìn thấy định nghĩa bên ngoài và không biết đối tượng đó được định nghĩa như thế nào và nó hoạt động như thế nào.
- ❖ Một ưu điểm của phương pháp này, được gọi là trừu tượng hóa dữ liệu, là chúng ta có thể thay đổi định nghĩa bên trong của một đối tượng mà không ảnh hưởng đến người dùng đối tượng đó, miễn là định nghĩa bên ngoài vẫn giữ nguyên.

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### CSDL

- ❖ Theo cách tương tự, cách tiếp cận CSDL tách cấu trúc dữ liệu khỏi các chương trình ứng dụng và lưu trữ nó trong CSDL. Nếu cấu trúc dữ liệu mới được thêm vào hoặc cấu trúc hiện có được sửa đổi thì các chương trình ứng dụng sẽ không bị ảnh hưởng, miễn là chúng không phụ thuộc trực tiếp vào những gì đã được sửa đổi.
- ❖ Ví dụ: nếu chúng ta thêm trường mới vào bản ghi hoặc tạo tệp mới thì các ứng dụng hiện có sẽ không bị ảnh hưởng. Tuy nhiên, nếu xóa một trường khỏi tệp mà chương trình ứng dụng sử dụng thì chương trình ứng dụng đó sẽ bị ảnh hưởng bởi thay đổi này và phải được sửa đổi cho phù hợp.

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### CSDL

- ❖ Thuật ngữ '**liên quan về mặt logic**': Khi phân tích nhu cầu thông tin của một tổ chức, cố gắng **xác định các thực thể, thuộc tính và mối quan hệ**.
  - Thực thể là một đối tượng riêng biệt (người, địa điểm, sự vật, khái niệm hoặc sự kiện) trong tổ chức và được thể hiện trong CSDL.
  - Thuộc tính là một thuộc tính mô tả một số khía cạnh của đối tượng mà chúng ta muốn ghi lại
  - Mối quan hệ là sự liên kết giữa các thực thể.

Ví dụ: Hình 1.6 hiển thị sơ đồ Thực thể–Mối quan hệ (ER) cho một phần của nghiên cứu điển hình DreamHome. Nó bao gồm:

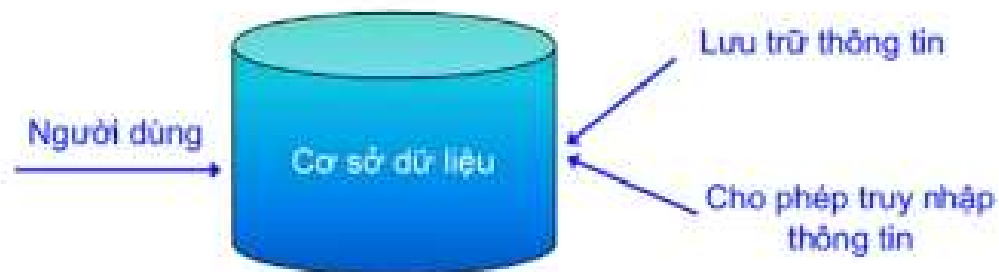
- 6 thực thể (hình chữ nhật): Branch, Staff, PropertyForRent, Client, PrivateOwner, and Lease;
- 7 mối quan hệ : *Has, Offers, Oversees, Views, Owns, LeasedBy*, và *Holds*;
- 6 thuộc tính, mỗi thuộc tính cho mỗi thực thể: branchNo, staffNo, propertyNo, clientNo, ownerNo, and leaseNo.

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### CSDL

- ❖ **Định nghĩa:** CSDL (database) = Tập hợp dữ liệu được tổ chức có cấu trúc liên quan với nhau và được lưu trữ trong máy tính.
- ❖ Được thiết kế, xây dựng cho phép người dùng lưu trữ dữ liệu, truy xuất thông tin hoặc cập nhật dữ liệu.
- ❖ **Mục đích:** Nhằm thoả mãn nhu cầu khai thác thông tin của một hay nhiều người sử dụng phục vụ cho nhiều mục đích khác nhau.





# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### CSDL

❖ CSDL được tổ chức có cấu trúc:

- Dữ liệu được lưu trữ thành các bản ghi (record), các trường (field)
- Dữ liệu có mối quan hệ với nhau

❖ Khả năng truy xuất thông tin từ CSDL:

- CSDL được cấu trúc để dễ dàng truy cập, quản lý và cập nhật dữ liệu

→ Cần phải quản trị CSDL

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### CSDL

#### ❖ Ví dụ CSDL:

SINHVIEN	TENSV	MASV	NAMSINH
	Hoang	1002013105	1995
	Van	1002013303	2004

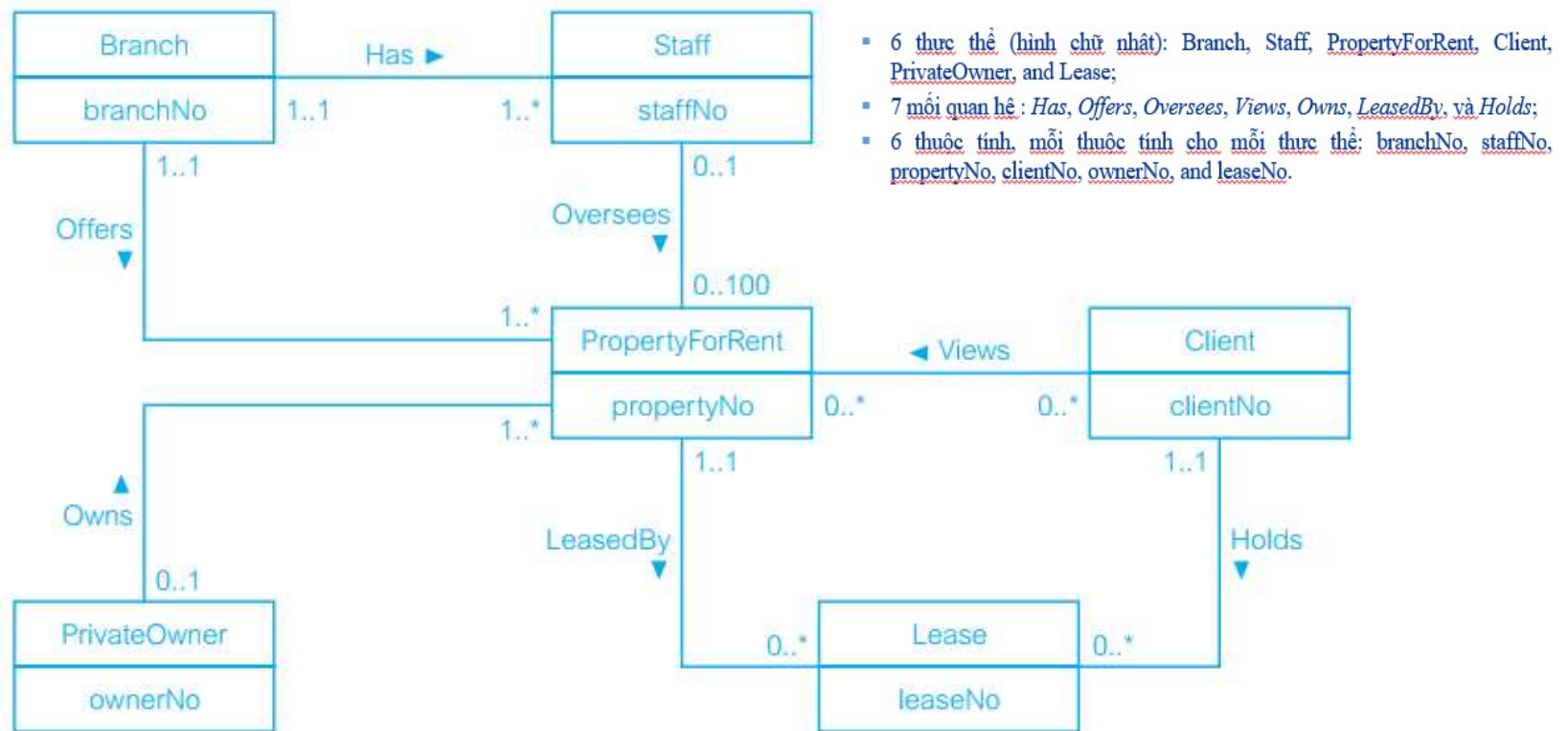
MONHOC	TENMH	MAHP	TC
	Cơ sở dữ liệu	841107	4
	HTTT DN	841048	3

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### CSDL

Ví dụ: Hình 1.6 hiển thị sơ đồ Thực thể–Mối quan hệ (ER)



Cơ sở dữ liệu đại diện cho các thực thể, thuộc tính và mối quan hệ logic giữa các thực thể. Nói cách khác, cơ sở dữ liệu chứa dữ liệu có liên quan về mặt logic.

# 1. Tổng quan về lý thuyết CSDL

## Các loại cơ sở dữ liệu

- ❖ Thực tế, cơ sở dữ liệu bao gồm rất nhiều dạng khác nhau và rất khó để phân loại chúng. Tuy nhiên, để có thể quản lý các dữ liệu một cách dễ dàng, cơ sở dữ liệu thường được phân loại như sau:
- ❖ Phân loại theo loại dữ liệu: Cơ sở dữ liệu có cấu trúc; Cơ sở dữ liệu phi cấu trúc; Cơ sở dữ liệu bán cấu trúc.
- ❖ Phân loại theo hình thức lưu trữ, mô hình tổ chức: Cơ sở dữ liệu dạng tệp; Cơ sở dữ liệu quan hệ; Cơ sở dữ liệu phân cấp.
- ❖ Phân loại theo đặc tính sử dụng: Cơ sở dữ liệu kho; Cơ sở dữ liệu ngữ nghĩa; Cơ sở dữ liệu hoạt động.
- ❖ Phân loại theo mô hình triển khai: Cơ sở dữ liệu tập trung; Cơ sở dữ liệu phân tán; Cơ sở dữ liệu tập trung có bản sao.

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### CSDL

- **Quản lý dữ liệu:** là quản lý số lượng lớn dữ liệu, bao gồm các việc lưu trữ và cung cấp cơ chế cho phép **thao tác** (thêm, sửa, xóa) và **truy vấn** dữ liệu.
- **Các phương pháp quản lý dữ liệu:**
  - Hệ thống quản lý bằng file
  - Hệ thống quản lý bằng cơ sở dữ liệu

#### Quản lý dữ liệu bằng file (tập tin)

##### ❖ Nhược điểm

- Mỗi người dùng khác nhau sẽ lưu trữ tập tin với thông tin theo nhu cầu riêng → Dữ liệu lưu trữ rời rạc
- Dữ liệu dễ bị trùng lặp.
- Dễ bị thiếu nhất quán, mâu thuẫn
- Chia sẻ dữ liệu rất khó khăn.
- Kém hiệu quả trong truy xuất ngẫu nhiên hoặc truy cập đồng thời. Truy xuất khó.
- Khó khôi phục dữ liệu khi có sự cố
- Có vấn đề về an toàn và bảo mật dữ liệu

#### Quản lý dữ liệu bằng CSDL

##### ❖ Ưu điểm:

- Tối thiểu hoá dư thừa thông tin
- Nâng cao tính nhất quán dữ liệu
- Nâng cao tính chia sẻ dữ liệu
- Nâng cao chất lượng dữ liệu
- Nâng cao khả năng truy cập và trả lại dữ liệu
- Giảm chi phí bảo trì chương trình.
- Bảo mật dữ liệu

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### **CSDL: Tiêu chuẩn**

- Tính bảo mật
- Tính độc lập
- Tránh sự dư thừa thông tin
- Có khả năng lưu trữ, phục hồi khi gặp sự cố
- Hiệu suất sử dụng tốt, đảm bảo chế độ ưu tiên khi truy nhập dữ liệu.

### **Đặc tính của CSDL**

- Tính tự mô tả 1
- Tính độc lập 2
- Tính trừu tượng 3
- Tính nhất quán 4
- Đa khung nhìn 5



# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### Đặc tính của CSDL

Tính tự mô tả 1

Tính độc lập 2

Tính trừu tượng 3

Tính nhất quán 4

Đa khung nhìn 5

#### ➤ Tính tự mô tả

- ✓ Hệ CSDL không chỉ chứa bản thân CSDL mà còn chứa định nghĩa đầy đủ mô tả (cấu trúc+ràng buộc) của CSDL
- ✓ Các định nghĩa được lưu trữ trong catalog: Chứa các thông tin về cấu trúc tập tin, kiểu và dạng thức lưu trữ của mỗi thành phần dữ liệu và những ràng buộc dữ liệu
- ✓ Dữ liệu trong catalog gọi là meta-data
- ✓ Thông tin lưu trong catalog được sử dụng bởi:
  - ✓ Hệ QTCSDL
  - ✓ Người lập trình CSDL

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### Đặc tính của CSDL

Tính tự mô tả 1

Tính độc lập 2

Tính trừu tượng 3

Tính nhất quán 4

Đa khung nhìn 5

#### ➤ Tính độc lập giữa chương trình và dữ liệu

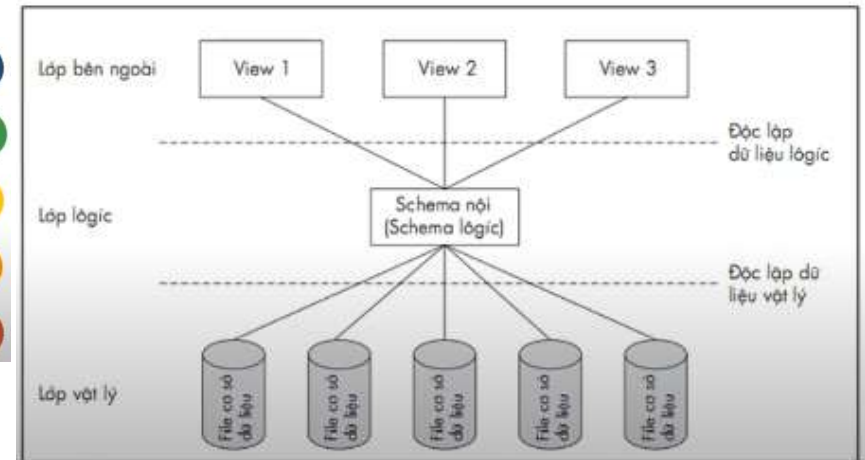
- Cung cấp khả năng trừu tượng hóa thông qua các lớp: vật lý, logic, bên ngoài.
- Sự phân biệt giữa các lớp tạo nên 2 tầng độc lập: độc lập dữ liệu vật lý và độc lập dữ liệu logic
- Một CSDL là độc lập dữ liệu nếu nó có hoặc có thể có khả năng phát triển mà không ảnh hưởng tới các hệ ứng dụng.

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### Đặc tính của CSDL

- Tính tự mô tả 1
- Tính độc lập 2
- Tính trừu tượng 3
- Tính nhất quán 4
- Đa khung nhìn 5



### ➤ Tính độc lập giữa chương trình và dữ liệu

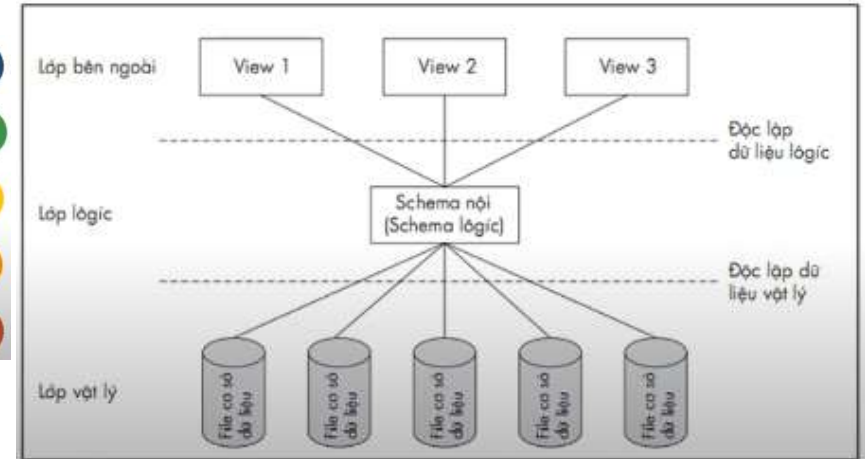
- Cung cấp khả năng trừu tượng hóa thông qua các lớp: vật lý, logic, bên ngoài.
- Sự phân biệt giữa các lớp tạo nên 2 tầng độc lập: độc lập dữ liệu vật lý và độc lập dữ liệu logic
- Một CSDL là độc lập dữ liệu nếu nó có hoặc có thể có khả năng phát triển mà không ảnh hưởng tới các hệ ứng dụng.

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### Đặc tính của CSDL

- 1 Tính tự mô tả
- 2 Tính độc lập
- 3 Tính trừu tượng
- 4 Tính nhất quán
- 5 Đa khung nhìn



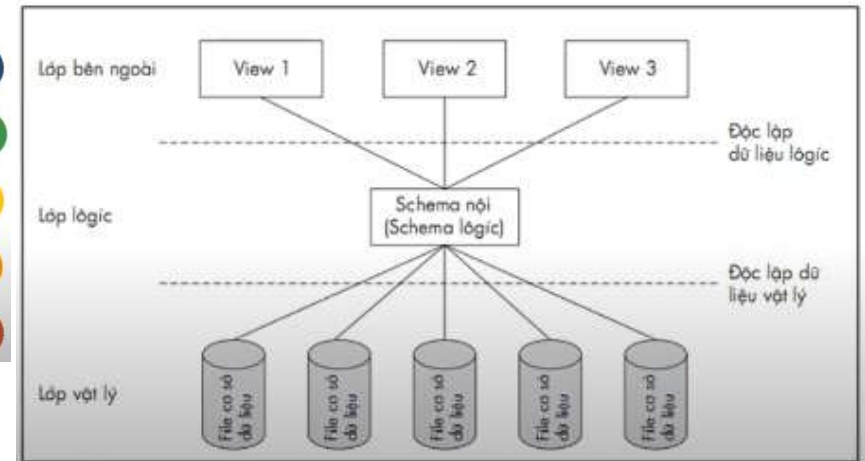
- **Lớp bên ngoài** : là tập tất cả các dữ liệu mà người sử dụng cụ thể có thể nhìn thấy và được phép truy cập (khung nhìn), là mức của người sử dụng và các chương trình ứng dụng.
- **Lớp logic (quan niệm)**: giải quyết vấn đề lưu trữ dữ liệu loại gì, lưu trữ bao nhiêu, mối quan hệ trong CSDL và mối quan hệ giữa các loại dữ liệu này như thế nào?
- **Lớp vật lý**: Đây là mức lưu trữ dữ liệu. Mục đích của mức này giải quyết vấn đề dữ liệu là gì, dữ liệu được lưu trữ như thế nào và nhằm mục đích gì. Người dùng CSDL không nhất thiết phải nắm được cấu trúc tổ chức của các file vật lý khi sử dụng một CSDL.

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### Đặc tính của CSDL

- 1 Tính tự mô tả
- 2 Tính độc lập
- 3 Tính trừu tượng
- 4 Tính nhất quán
- 5 Đa khung nhìn



#### ➤ Tính độc lập giữa chương trình và dữ liệu

##### ➤ Độc lập dữ liệu mức vật lý :

- ✦ Khả năng làm thay đổi cấu trúc file vật lý của một CSDL mà không làm gián đoạn người dùng đang truy cập vào các quá trình đang diễn ra
- ✦ Việc phân biệt lớp vật lý và lớp logic tạo ra sự độc lập vật lý

##### ➤ Độc lập dữ liệu mức 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
- ✦ Việc phân biệt giữa lớp logic và lớp bên ngoài gọi là độc lập dữ liệu logic



# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### Đặc tính của CSDL

Tính tự mô tả	1
Tính độc lập	2
Tính trừu tượng	3
Tính nhất quán	4
Đa khung nhìn	5

#### ➤ Tính trừu tượng

- ✓ HQT CSDL chỉ cung cấp biểu diễn về dữ liệu ở mức khái niệm và che dấu nhiều chi tiết về cách thức lưu trữ
- ✓ Sự trừu tượng hóa giúp cho tính độc lập chương trình và dữ liệu
- ✓ Sự trừu tượng hóa được thể hiện qua mô hình dữ liệu

#### ➤ 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 tại mọi thời điểm

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL - **Đặc tính của CSDL**

Tính tự mô tả

1

Tính độc lập

2

Tính trừu tượng

3

Tính nhất quán

4

Đa khung nhìn

5

### ➤ **Tính trừu tượng**

- ✓ HQT CSDL chỉ cung cấp biểu diễn về dữ liệu ở mức khái niệm và che dấu nhiều chi tiết về cách thức lưu trữ
- ✓ Sự trừu tượng hóa giúp cho tính độc lập chương trình và dữ liệu
- ✓ Sự trừu tượng hóa được thể hiện qua mô hình dữ liệu

### ➤ **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 tại mọi thời điểm

### ➤ **Đa khung nhìn:** Hỗ trợ nhiều cách nhìn dữ liệu

- ✓ Một CSDL có nhiều người sử dụng
- ✓ Mỗi người đòi hỏi cách nhìn (view) khác nhau về CSDL
- ✓ Một khung nhìn (view) là một phần của CSDL hoặc các dữ liệu được tổng hợp từ CSDL



# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL - **Khung nhìn – xem (View)**

❖ Một số lợi ích khác của chế độ xem:

- Chế độ xem **cung cấp mức độ bảo mật**. Chế độ xem có thể được thiết lập để loại trừ dữ liệu mà một số người dùng không nên xem.

Ví dụ: có thể tạo chế độ xem cho phép Giám đốc chi nhánh và Phòng tính lương xem tất cả dữ liệu nhân viên, bao gồm chi tiết tiền lương và có thể tạo chế độ xem thứ hai cho các nhân viên khác sử dụng (loại trừ chi tiết tiền lương).

- Chế độ xem **cung cấp cơ chế để tùy chỉnh giao diện của CSDL**.

Ví dụ: Phòng Hợp đồng có thể muốn gọi trường tiền thuê hàng tháng (tiền thuê) bằng tên rõ ràng hơn: Tiền thuê hàng tháng.

- Chế độ xem trình bày một bức tranh nhất quán, không thay đổi về cấu trúc của CSDL, ngay cả khi CSDL cơ bản bị thay đổi.

Ví dụ: các trường được thêm hoặc xóa, các mối quan hệ đã thay đổi, chia tách, cấu trúc lại hoặc đổi tên tệp. Nếu các trường được thêm hoặc xóa khỏi tệp và chế độ xem không yêu cầu các trường này thì chế độ xem sẽ không bị ảnh hưởng bởi thay đổi này. Do đó, **một khung nhìn giúp cung cấp tính độc lập giữa chương trình và dữ liệu đã đề cập ở trước.**

# 1. Tổng quan về lý thuyết CSDL

## Khung nhìn – xem (View)

- ❖ Với chức năng này, DBMS là một công cụ mạnh và hữu ích. Tuy nhiên, do người dùng cuối không quá quan tâm đến mức độ phức tạp hay dễ dàng của một nhiệm vụ đối với hệ thống, nên có thể DBMS sẽ khiến mọi thứ trở nên phức tạp hơn vì giờ đây họ thấy nhiều dữ liệu hơn mức họ thực sự cần hoặc muốn.

Ví dụ: các chi tiết mà Phòng Hợp đồng muốn xem về tài sản cho thuê, như trong Hình 1.5, đã thay đổi trong cách tiếp cận CSDL, như trong Hình 1.7. Bây giờ CSDL cũng chứa loại tài sản, số phòng và thông tin chi tiết về chủ sở hữu.

- ❖ DBMS cung cấp một tiện ích khác được gọi là cơ chế xem, cho phép mỗi người dùng có chế độ xem CSDL của riêng mình (1 chế độ xem là một tập hợp con của CSDL).

Ví dụ: chúng ta có thể thiết lập chế độ xem cho phép Phòng Hợp đồng chỉ xem dữ liệu mà họ muốn xem về tài sản cho thuê.

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### **Hệ quản trị CSDL (DBMS)**

- DBMS là hệ thống phần mềm cho phép người dùng xác định, tạo, duy trì và **kiểm soát** quyền truy cập vào CSDL.
- DBMS là phần mềm tương tác với các chương trình ứng dụng và CSDL của người dùng.
- Thông thường, **DBMS cung cấp các tiện ích** sau:
  - Nó cho phép người dùng xác định CSDL, thường thông qua Ngôn ngữ định nghĩa dữ liệu (DDL). DDL cho phép người dùng **chỉ định** các kiểu và cấu trúc dữ liệu cũng như các ràng buộc đối với dữ liệu sẽ được lưu trữ trong CSDL.
  - Nó cho phép người dùng chèn, cập nhật, xóa và truy xuất dữ liệu từ CSDL, thường thông qua Ngôn ngữ thao tác dữ liệu (DML).

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL - **Hệ quản trị CSDL (DBMS)**

### ■ Thông thường, DBMS cung cấp các tiện ích sau:

- Nó cho phép người dùng xác định CSDL, thường thông qua Ngôn ngữ định nghĩa dữ liệu (DDL).
- Nó cho phép người dùng chèn, cập nhật, xóa và truy xuất dữ liệu từ CSDL, thường thông qua Ngôn ngữ thao tác dữ liệu (DML).
  - Việc có một kho lưu trữ trung tâm cho tất cả dữ liệu và mô tả dữ liệu cho phép DML cung cấp phương tiện truy vấn chung cho dữ liệu này, được gọi là ngôn ngữ truy vấn.
  - Việc cung cấp ngôn ngữ truy vấn giúp giảm bớt các vấn đề với hệ thống dựa trên tệp trong đó người dùng phải làm việc với một bộ truy vấn cố định hoặc có sự phát triển nhanh chóng của các chương trình, gây ra các vấn đề lớn về quản lý phần mềm.
  - Ngôn ngữ truy vấn phổ biến nhất là Ngôn ngữ truy vấn có cấu trúc (SQL), hiện là ngôn ngữ tiêu chuẩn chính thức và thực tế cho các DBMS quan hệ.

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL - **Hệ quản trị CSDL (DBMS)**

### ■ Thông thường, DBMS cung cấp các tiện ích sau:

- Nó cho phép người dùng xác định CSDL, thường thông qua Ngôn ngữ định nghĩa dữ liệu (DDL).
- Nó cho phép người dùng chèn, cập nhật, xóa và truy xuất dữ liệu từ CSDL, thường thông qua Ngôn ngữ thao tác dữ liệu (DML).
- Nó cung cấp quyền truy cập có kiểm soát vào CSDL.

Ví dụ: nó có thể cung cấp:

- một hệ thống bảo mật ngăn chặn người dùng trái phép truy cập CSDL;
- một hệ thống toàn vẹn, duy trì tính nhất quán của dữ liệu được lưu trữ;
- một hệ thống kiểm soát đồng thời, cho phép truy cập chung vào CSDL;
- hệ thống kiểm soát phục hồi có chức năng khôi phục CSDL về trạng thái nhất quán trước đó sau lỗi phần cứng hoặc phần mềm;
- một danh mục mà người dùng có thể truy cập, chứa các mô tả về dữ liệu trong CSDL.

# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL - **Hệ quản trị CSDL (DBMS)**

❖ Một số ví dụ về hệ quản trị CSDL



# 1. Tổng quan về lý thuyết CSDL

## 1.3 Tiếp cận CSDL

### **Chương trình ứng dụng**

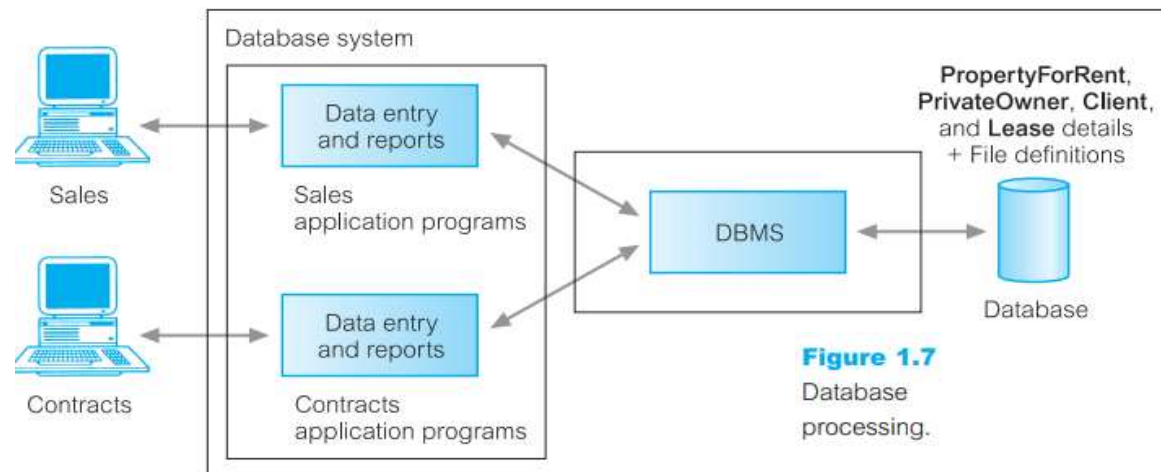
- ❖ Là chương trình máy tính tương tác với CSDL bằng cách đưa ra một yêu cầu thích hợp (thường là câu lệnh SQL) tới DBMS.
- ❖ Người dùng tương tác với CSDL thông qua một số chương trình ứng dụng được sử dụng để tạo và duy trì CSDL cũng như tạo ra thông tin.
- ❖ Các chương trình này có thể là các ứng dụng hàng loạt thông thường hoặc là các ứng dụng trực tuyến.
- ❖ Các chương trình ứng dụng có thể được viết bằng một số ngôn ngữ lập trình hoặc bằng một số ngôn ngữ thế hệ thứ tư mức cao hơn.



## 1.3 Tiếp cận CSDL

### Chương trình ứng dụng

Cách tiếp cận CSDL được minh họa trong Hình 1.7, dựa trên cách tiếp cận tệp của Hình 1.5.

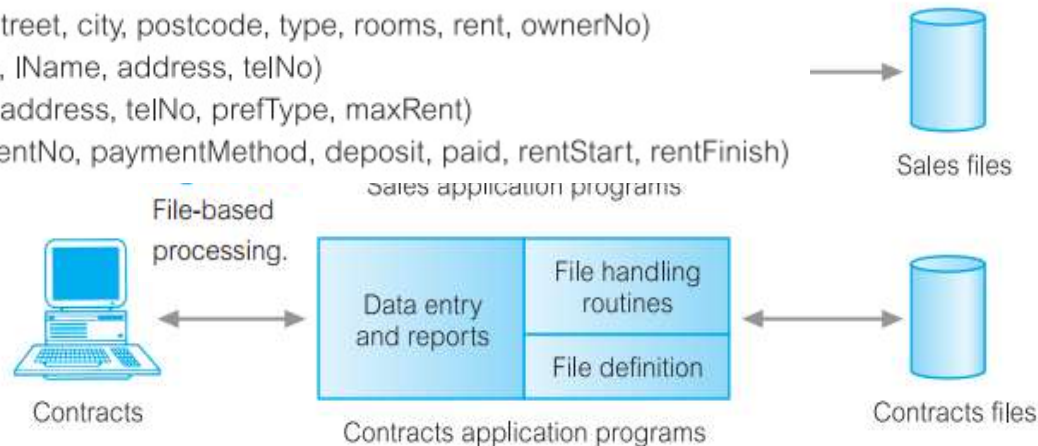


**PropertyForRent** (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)

**PrivateOwner** (ownerNo, fName, lName, address, telNo)

**Client** (clientNo, fName, lName, address, telNo, prefType, maxRent)

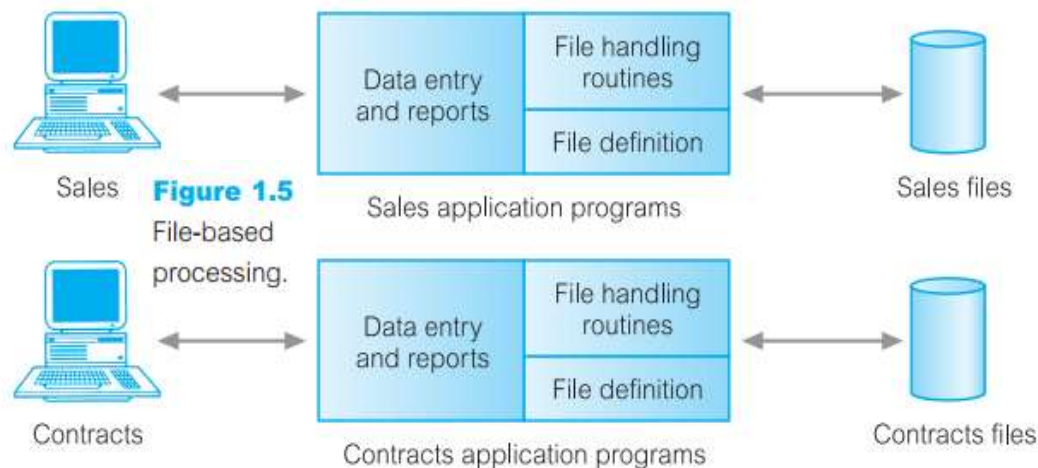
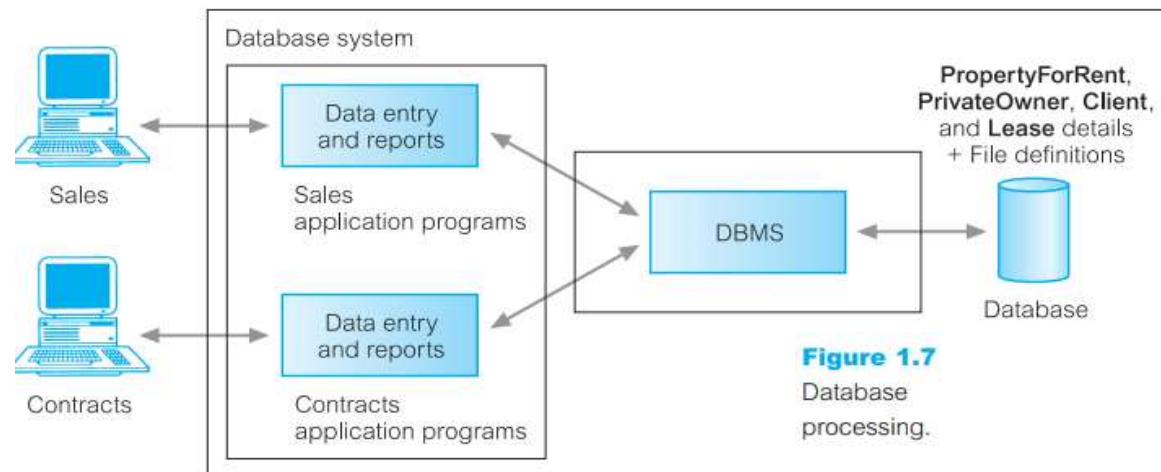
**Lease** (leaseNo, propertyNo, clientNo, paymentMethod, deposit, paid, rentStart, rentFinish)



## 1.3 Tiếp cận CSDL

### Chương trình ứng dụng

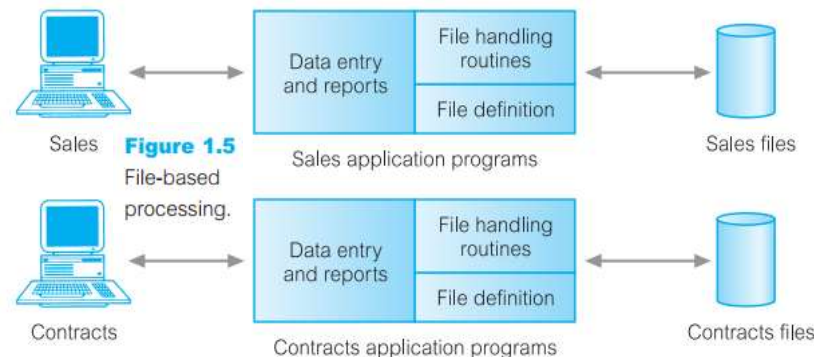
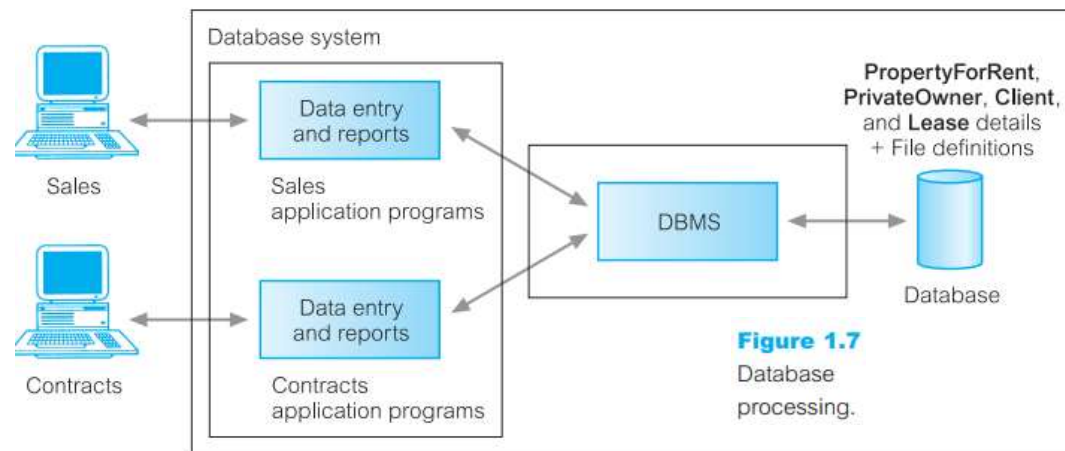
Nó cho thấy Phòng Bán hàng và Hợp đồng sử dụng các chương trình ứng dụng của họ để truy cập CSDL thông qua DBMS.



## 1.3 Tiếp cận CSDL

### Chương trình ứng dụng

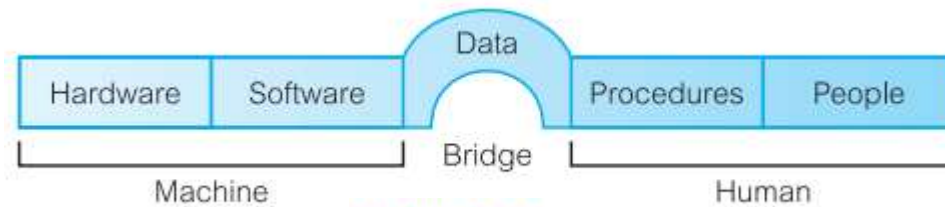
Mỗi chương trình ứng dụng của phòng ban xử lý việc nhập dữ liệu, bảo trì dữ liệu và tạo báo cáo. Tuy nhiên, so với cách tiếp cận dựa trên tệp, cấu trúc vật lý và lưu trữ dữ liệu hiện được quản lý bởi DBMS.



## 1.3 Tiếp cận CSDL

### Các thành phần của môi trường DBMS: 5 thành phần chính

- Phần cứng
- Phần mềm
- Dữ liệu
- Thủ tục
- Con người



**Figure 1.8**  
DBMS environment.

### Phần cứng

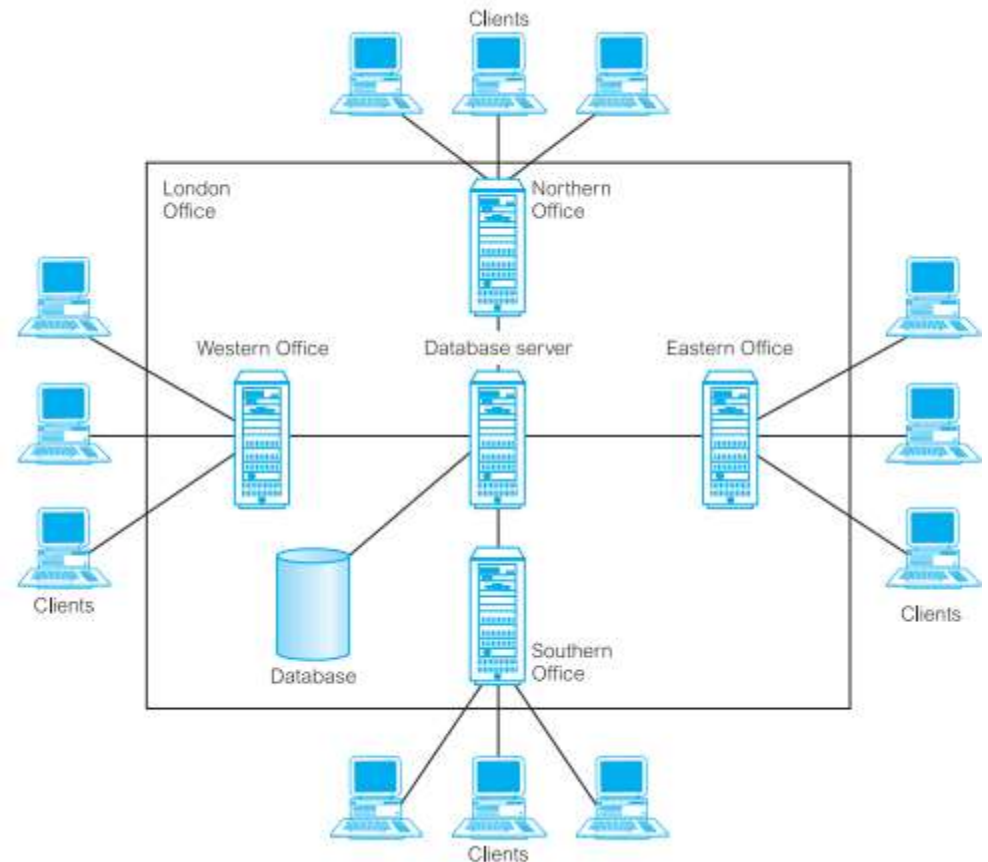
- ❖ DBMS và các ứng dụng yêu cầu phần cứng để chạy.
- ❖ Phần cứng bao gồm từ một máy tính cá nhân, đến một máy tính lớn, hoặc đến một mạng máy tính.
- ❖ Phần cứng cụ thể phụ thuộc vào yêu cầu của tổ chức và DBMS được sử dụng.
- ❖ Một số DBMS chỉ chạy trên phần cứng hoặc hệ điều hành cụ thể, trong khi một số khác chạy trên nhiều loại phần cứng và hệ điều hành.

## 1.3 Tiếp cận CSDL

### Các thành phần của môi trường DBMS: 5 thành phần chính

**Phần cứng:** Cấu hình phần cứng đơn giản hóa cho DreamHome được minh họa trong Hình 1.9.

- ❖ Gồm 1 mạng lưới các máy tính mini, với một máy tính trung tâm đặt tại London chạy **backend** của DBMS, tức là một phần của DBMS quản lý và kiểm soát quyền truy cập vào CSDL.

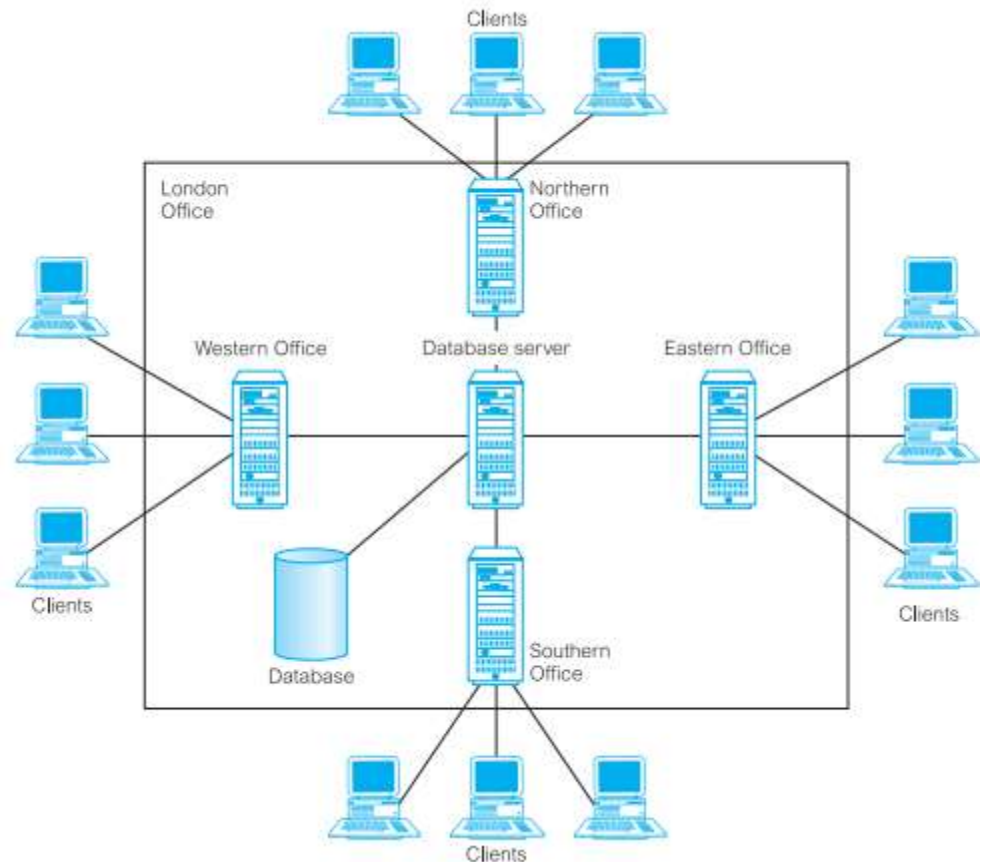


## 1.3 Tiếp cận CSDL

### Các thành phần của môi trường DBMS: 5 thành phần chính

**Phần cứng:** Cấu hình phần cứng đơn giản hóa cho DreamHome được minh họa trong Hình 1.9.

- ❖ Nó cũng cho thấy một số máy tính ở nhiều vị trí khác nhau đang chạy frontend của DBMS, nghĩa là một phần của DBMS có giao diện với người dùng. Đây được gọi là kiến trúc máy khách-máy chủ: phần backend là máy chủ và các frontend là máy khách.





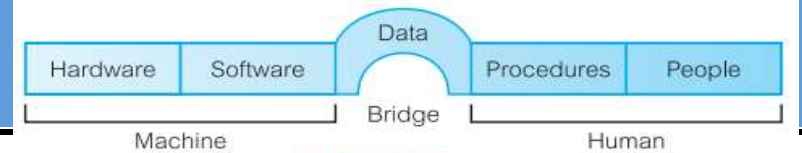
### **Các thành phần của môi trường DBMS: 5 thành phần chính**

#### **Phần mềm**

- ❖ Gồm chính phần mềm DBMS và các chương trình ứng dụng, cùng với hệ điều hành, bao gồm cả phần mềm mạng nếu DBMS đang được sử dụng qua mạng.
- ❖ Thông thường, các chương trình ứng dụng được viết bằng ngôn ngữ lập trình thế hệ thứ ba (3GL), chẳng hạn như 'C', C++, Java, Visual Basic, COBOL, Fortran, Ada hoặc Pascal hoặc sử dụng ngôn ngữ thế hệ thứ tư (4GL) (như SQL), được nhúng trong ngôn ngữ thế hệ thứ ba.
- ❖ DBMS mục tiêu có thể có các công cụ thế hệ thứ tư của riêng nó cho phép phát triển nhanh các ứng dụng thông qua việc cung cấp các ngôn ngữ truy vấn phi thủ tục, trình tạo báo cáo, trình tạo biểu mẫu, trình tạo đồ họa và trình tạo ứng dụng.
- ❖ Việc sử dụng các công cụ thế hệ thứ tư có thể cải thiện đáng kể năng suất và tạo ra các chương trình dễ bảo trì hơn.



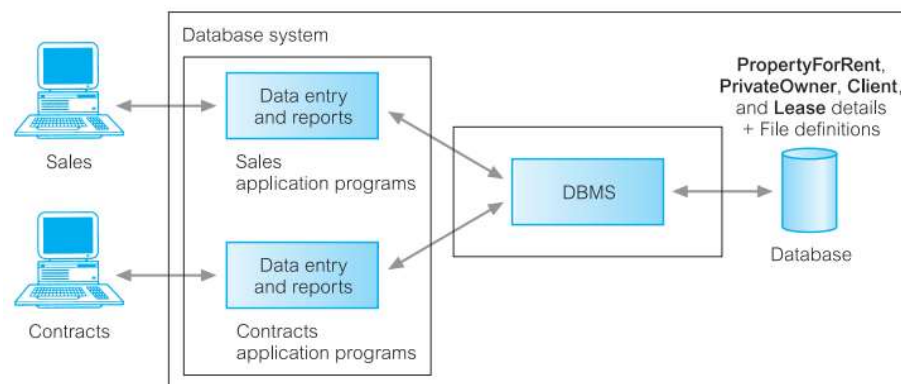
## 1.3 Tiếp cận CSDL



**Figure 1.8**  
DBMS environment.

### Các thành phần của môi trường DBMS: 5 thành phần chính Dữ liệu

- ❖ Là thành phần quan trọng nhất của môi trường DBMS,
- ❖ Từ Hình 1.8: dữ liệu đóng vai trò là cầu nối giữa máy và con người. CSDL chứa cả dữ liệu vận hành và siêu dữ liệu, 'dữ liệu về dữ liệu'. Cấu trúc của CSDL được gọi là lược đồ.
- ❖ Trong Hình 1.7, lược đồ bao gồm bốn tệp hoặc bảng, cụ thể là: PropertyForRent, PrivateOwner, Client và Lease.



**Figure 1.7**  
Database processing.

**PropertyForRent** (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)  
**PrivateOwner** (ownerNo, fName, lName, address, telNo)  
**Client** (clientNo, fName, lName, address, telNo, prefType, maxRent)  
**Lease** (leaseNo, propertyNo, clientNo, paymentMethod, deposit, paid, rentStart, rentFinish)

## 1.3 Tiếp cận CSDL

### Các thành phần của môi trường DBMS: 5 thành phần chính

**PropertyForRent** (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)

**PrivateOwner** (ownerNo, fName, lName, address, telNo)

**Client** (clientNo, fName, lName, address, telNo, prefType, maxRent)

**Lease** (leaseNo, propertyNo, clientNo, paymentMethod, deposit, paid, rentStart, rentFinish)

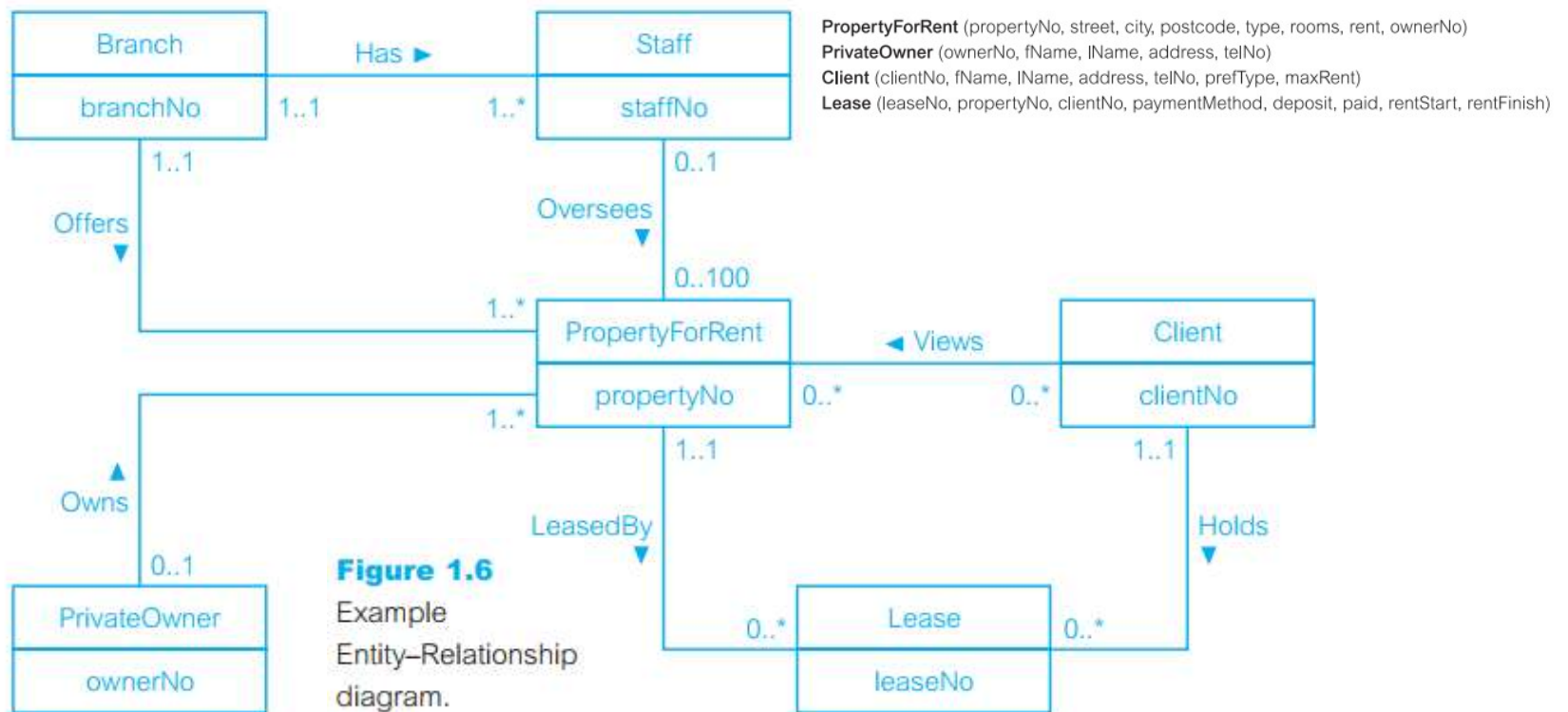
### Dữ liệu

- ❖ Bảng PropertyForRent có 8 trường hoặc thuộc tính, cụ thể là: Số tài sản, đường phố, thành phố, mã bưu điện, loại (loại tài sản), số phòng, tiền thuê (tiền thuê hàng tháng) và số chủ sở hữu.
- ❖ Thuộc tính số chủ sở hữu (ownerNo) mô hình hóa mối quan hệ giữa PropertyForRent và PrivateOwner: nghĩa là chủ sở hữu sở hữu một tài sản cho thuê, như được mô tả trong sơ đồ Thực thể—Mối quan hệ của Hình 1.6.

## 1.3 Tiếp cận CSDL

### Các thành phần của môi trường DBMS: 5 thành phần chính Dữ liệu

❖ Sơ đồ Thực thể–Mối quan hệ của Hình 1.6.



## 1.3 Tiếp cận CSDL

### Các thành phần của n Dữ liệu

**PropertyForRent** (propertyNo, street, city, postcode, type, rooms, rent, ownerNo)

**PrivateOwner** (ownerNo, fName, lName, address, telNo)

**Client** (clientNo, fName, lName, address, telNo, prefType, maxRent)

**Lease** (leaseNo, propertyNo, clientNo, paymentMethod, deposit, paid, rentStart, rentFinish)

Ví dụ, trong Hình 1.2, thấy chủ sở hữu CO46, Joe Keogh, sở hữu tài sản PA14.

**PropertyForRent**

propertyNo	street	city	postcode	type	rooms	rent	ownerNo
PA14	16 Holhead	Aberdeen	AB7 5SU	House	6	650	CO46
PL94	6 Argyll St	London	NW2	Flat	4	400	CO87
PG4	6 Lawrence St	Glasgow	G11 9QX	Flat	3	350	CO40
PG36	2 Manor Rd	Glasgow	G32 4QX	Flat	3	375	CO93
PG21	18 Dale Rd	Glasgow	G12	House	5	600	CO87
PG16	5 Novar Dr	Glasgow	G12 9AX	Flat	4	450	CO93

**PrivateOwner**

ownerNo	fName	lName	address	telNo
CO46	Joe	Keogh	2 Fergus Dr, Aberdeen AB2 7SX	01224-861212
CO87	Carol	Farrel	6 Achray St, Glasgow G32 9DX	0141-357-7419
CO40	Tina	Murphy	63 Well St, Glasgow G42	0141-943-1728
CO93	Tony	Shaw	12 Park Pl, Glasgow G4 0QR	0141-225-7025

**Figure 1.2**

The PropertyForRent, PrivateOwner, and Client files used by Sales.

## 1.3 Tiếp cận CSDL

### Thủ tục

- ❖ Đề cập đến các hướng dẫn và quy tắc chi phối việc thiết kế và sử dụng CSDL.
- ❖ Người dùng hệ thống và nhân viên quản lý CSDL yêu cầu các thủ tục dạng văn bản về cách sử dụng hoặc vận hành hệ thống. Bao gồm các hướng dẫn về cách:
  - đăng nhập vào DBMS;
  - sử dụng một cơ sở DBMS hoặc chương trình ứng dụng cụ thể;
  - khởi động và dừng DBMS;
  - tạo bản sao lưu của CSDL;
  - xử lý các lỗi phần cứng hoặc phần mềm. Bao gồm các quy trình về cách xác định thành phần bị lỗi, cách khắc phục thành phần bị lỗi (ví dụ: gọi điện cho kỹ sư phần cứng thích hợp) và sau khi sửa lỗi, cách khôi phục CSDL;
  - thay đổi cấu trúc của bảng, sắp xếp lại CSDL trên nhiều đĩa, cải thiện hiệu suất hoặc lưu trữ dữ liệu vào bộ lưu trữ thứ cấp.

## 1.3 Tiếp cận CSDL

Vai trò trong môi trường CSDL: Con người (thành phần thứ 5)

### Quản trị viên dữ liệu và CSDL

- ❖ Quản trị viên **CSDL** (DBA) chịu trách nhiệm hiện thực hóa vật lý **CSDL**, bao gồm thiết kế và triển khai **CSDL** vật lý, kiểm soát tính bảo mật và tính toàn vẹn, bảo trì hệ thống vận hành và đảm bảo hiệu suất thỏa đáng của ứng dụng cho người dùng.
- ❖ Vai trò của DBA thiên về mặt kỹ thuật hơn vai trò của DA, đòi hỏi kiến thức chi tiết về DBMS mục tiêu và môi trường hệ thống.

## 1.3 Tiếp cận CSDL

### Vai trò trong môi trường CSDL: Con người (thành phần thứ 5)

- ❖ 4 loại người riêng biệt tham gia vào môi trường DBMS: quản trị viên CSDL và dữ liệu, nhà thiết kế CSDL, nhà phát triển ứng dụng và người dùng cuối.

### Quản trị viên dữ liệu và CSDL

- ❖ CSDL và DBMS là các tài nguyên của công ty phải được quản lý giống như bất kỳ tài nguyên nào khác.
- ❖ Quản trị dữ liệu và CSDL là các vai trò thường liên quan đến việc quản lý và kiểm soát DBMS và dữ liệu của nó.
- ❖ Quản trị viên dữ liệu (DA) chịu trách nhiệm quản lý tài nguyên dữ liệu bao gồm lập kế hoạch CSDL, phát triển và duy trì các tiêu chuẩn, chính sách và thủ tục cũng như thiết kế CSDL khái niệm/logic. DA tư vấn và tư vấn cho các nhà quản lý cấp cao, đảm bảo rằng hướng phát triển CSDL cuối cùng sẽ hỗ trợ các mục tiêu của công ty.



## 1.3 Tiếp cận CSDL

Vai trò trong môi trường CSDL: Con người (thành phần thứ 5)

### Nhà thiết kế CSDL

- ❖ Trong các dự án thiết kế CSDL lớn, có thể phân biệt hai loại người thiết kế: người thiết kế CSDL logic và người thiết kế CSDL vật lý.
- ❖ Người thiết kế CSDL logic quan tâm đến việc xác định dữ liệu (nghĩa là các thực thể và thuộc tính), mối quan hệ giữa dữ liệu và các ràng buộc đối với dữ liệu sẽ được lưu trữ trong CSDL.
- ❖ Người thiết kế CSDL logic phải có sự hiểu biết thấu đáo và đầy đủ về dữ liệu của tổ chức cũng như mọi ràng buộc đối với dữ liệu này (các ràng buộc đôi khi được gọi là quy tắc kinh doanh). Những ràng buộc này mô tả các đặc điểm chính của dữ liệu được xem bởi tổ chức.

Ví dụ về các ràng buộc đối với DreamHome là:

- ❖ một nhân viên không thể quản lý hơn 100 bất động sản cho thuê hoặc bán cùng một lúc;
- ❖ nhân viên không thể xử lý việc bán hoặc cho thuê tài sản của chính mình;
- ❖ luật sư không thể đại diện cho cả người mua và người bán tài sản.

## 1.3 Tiếp cận CSDL

**Vai trò trong môi trường CSDL: Con người (thành phần thứ 5)**

### **Nhà thiết kế CSDL**

- ❖ Để có hiệu quả, người thiết kế CSDL logic phải thu hút tất cả người dùng CSDL tiềm năng tham gia vào quá trình phát triển mô hình dữ liệu và sự tham gia này phải bắt đầu càng sớm càng tốt trong quá trình.
- ❖ Công việc của người thiết kế CSDL logic chia thành hai giai đoạn:
  - Thiết kế CSDL khái niệm, độc lập với các chi tiết triển khai như DBMS mục tiêu, chương trình ứng dụng, ngôn ngữ lập trình hoặc bất kỳ cân nhắc vật lý nào khác;
- ❖ Thiết kế CSDL logic, nhắm đến một mô hình dữ liệu cụ thể, chẳng hạn như quan hệ, mạng, phân cấp hoặc hướng đối tượng.

## 1.3 Tiếp cận CSDL

Vai trò trong môi trường CSDL: Con người (thành phần thứ 5)

**Nhà thiết kế CSDL:** Người thiết kế **CSDL** vật lý

- ❖ Người thiết kế **CSDL** vật lý quyết định cách thực hiện thiết kế **CSDL** logic về mặt vật lý. Điều này liên quan đến:
  - ánh xạ thiết kế CSDL logic vào một tập hợp các bảng và các ràng buộc toàn vẹn;
  - lựa chọn cấu trúc lưu trữ cụ thể và phương pháp truy cập dữ liệu để đạt được hiệu suất tốt;
  - thiết kế mọi biện pháp bảo mật cần thiết đối với dữ liệu.
- ❖ Nhiều phần của thiết kế **CSDL** vật lý phụ thuộc nhiều vào DBMS mục tiêu và có thể có nhiều cách triển khai một cơ chế. Do đó, người thiết kế CSDL vật lý phải **nhận thức đầy đủ về chức năng của DBMS mục tiêu và phải hiểu những ưu điểm và nhược điểm của từng phương án đối với việc triển khai cụ thể.**
- ❖ Người thiết kế **CSDL** vật lý phải có khả năng lựa chọn chiến lược lưu trữ phù hợp có tính đến việc sử dụng. Trong khi thiết kế **CSDL** logic và khái niệm quan tâm đến cái gì thì thiết kế **CSDL** vật lý quan tâm đến cách thức.

## 1.3 Tiếp cận CSDL

### Thiết kế CSDL: Sự thay đổi mô hình

- ❖ Dữ liệu trong CSDL có cấu trúc. Ví dụ: 4 bảng trong Hình 1.7: PropertyForRent, PrivateOwner, Client và Lease. Nhưng làm thế nào có được cấu trúc này? Câu trả lời khá đơn giản: **cấu trúc của CSDL được xác định trong quá trình thiết kế CSDL.**
- ❖ Tuy nhiên, việc thực hiện thiết kế CSDL có thể cực kỳ phức tạp. Để tạo ra một hệ thống đáp ứng nhu cầu thông tin của tổ chức đòi hỏi một cách tiếp cận khác với cách tiếp cận của các hệ thống dựa trên tệp, trong đó công việc được thúc đẩy bởi nhu cầu ứng dụng của từng bộ phận.
- ❖ Để cách tiếp cận CSDL thành công, tổ chức hiện phải nghĩ đến dữ liệu trước tiên và ứng dụng thứ hai. Sự thay đổi trong cách tiếp cận này đôi khi được gọi là **sự thay đổi mô hình.**
- ❖ Để hệ thống được người dùng cuối chấp nhận, hoạt động thiết kế CSDL là rất quan trọng. CSDL được thiết kế kém sẽ tạo ra các lỗi có thể dẫn đến việc đưa ra các quyết định tồi, điều này có thể gây hậu quả nghiêm trọng cho tổ chức.
- ❖ Mặt khác, một CSDL được thiết kế tốt sẽ tạo ra một hệ thống cung cấp thông tin chính xác để quá trình ra quyết định thành công một cách hiệu quả.

## 1.3 Tiếp cận CSDL

Vai trò trong môi trường CSDL: Con người (thành phần thứ 5)

### Nhà phát triển ứng dụng

- ❖ Khi CSDL đã được triển khai, các chương trình ứng dụng cung cấp chức năng cần thiết cho người dùng cuối phải được triển khai. Đây là trách nhiệm của các nhà phát triển ứng dụng.
- ❖ Thông thường, các nhà phát triển ứng dụng làm việc dựa trên đặc tả do các nhà phân tích hệ thống tạo ra.
- ❖ Mỗi chương trình chứa các câu lệnh yêu cầu DBMS thực hiện một số thao tác trên CSDL. Điều này bao gồm truy xuất dữ liệu, chèn, cập nhật và xóa dữ liệu.
- ❖ Các chương trình có thể được viết bằng ngôn ngữ lập trình thế hệ thứ ba hoặc ngôn ngữ thế hệ thứ tư.

### Người dùng cuối

- ❖ Người dùng cuối là 'khách hàng' của CSDL đã được thiết kế, triển khai và đang được duy trì để phục vụ nhu cầu thông tin của họ.
- ❖ Người dùng cuối có thể được phân loại theo cách họ sử dụng hệ thống:

## 1.3 Tiếp cận CSDL

Vai trò trong môi trường CSDL: Con người (thành phần thứ 5)

**Người dùng ngây thơ (không biết về DBMS).**

- ❖ Họ truy cập CSDL thông qua các chương trình ứng dụng được viết đặc biệt nhằm cố gắng thực hiện các thao tác đơn giản nhất có thể.
- ❖ Họ gọi các hoạt động CSDL bằng cách nhập các lệnh đơn giản hoặc chọn các tùy chọn từ menu. Điều này có nghĩa là họ không cần biết gì về CSDL hoặc DBMS.
- ❖ Ví dụ: nhân viên thanh toán tại siêu thị sử dụng đầu đọc mã vạch để tìm hiểu giá của mặt hàng. Tuy nhiên, có một chương trình ứng dụng có chức năng đọc mã vạch, tra cứu giá của mặt hàng trong CSDL, giảm trường CSDL chứa số lượng mặt hàng đó trong kho và hiển thị giá trên quầy tính tiền.

**Người dùng cuối tinh tế.**

- ❖ Là người đã quen thuộc với cấu trúc của CSDL và các tiện ích do DBMS cung cấp.
- ❖ Họ có thể sử dụng ngôn ngữ truy vấn cấp cao như SQL để thực hiện các thao tác được yêu cầu, thậm chí có thể viết các chương trình ứng dụng để sử dụng cho riêng họ.

## 1.3 Tiếp cận CSDL

### ❖ Hệ quản trị CSDL – Lợi ích

- ❖ Quản trị CSDL: hạn chế trùng lặp, chia sẻ dữ liệu, Phân quyền, quản lý tài khoản đăng nhập, đảm bảo các ràng buộc giữa CSDL.
- ❖ Cung cấp giao diện làm việc để che dấu các đặc tính phức tạp về mặt cấu trúc tổ chức dữ liệu vật lý.
- ❖ Hỗ trợ ngôn ngữ giao tiếp CSDL.
- ❖ Có cơ chế bảo mật an toàn, khả năng sao lưu, backup.



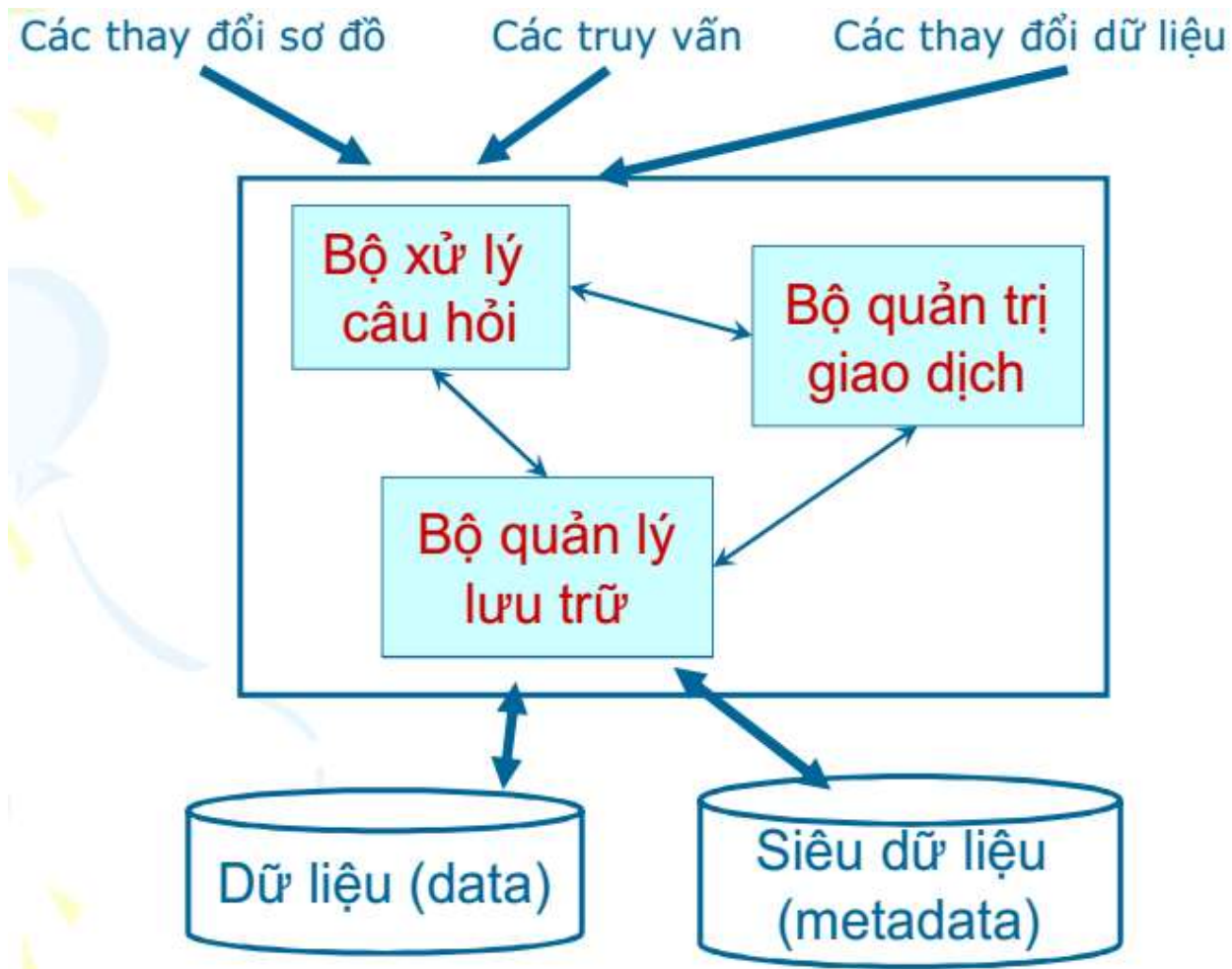
## 1.3 Tiếp cận CSDL

### Các tính năng của hệ quản trị CSDL

- ❖ Quản lý dữ liệu tồn tại lâu dài
  - Định nghĩa dữ liệu
  - Quản lý lưu trữ
- ❖ Truy xuất dữ liệu một cách hiệu quả
  - Biểu diễn các thao tác dữ liệu
  - Xử lý câu hỏi
  - Quản trị giao dịch
- ❖ Hỗ trợ ít nhất một mô hình dữ liệu
- ❖ Đảm bảo tính độc lập dữ liệu
- ❖ Hỗ trợ các ngôn ngữ cấp cao nhất định cho phép người sử dụng định nghĩa cấu trúc của dữ liệu, truy nhập và thao tác dữ liệu
- ❖ Điều khiển truy nhập
- ❖ Phục hồi dữ liệu

## 1.3 Tiếp cận CSDL

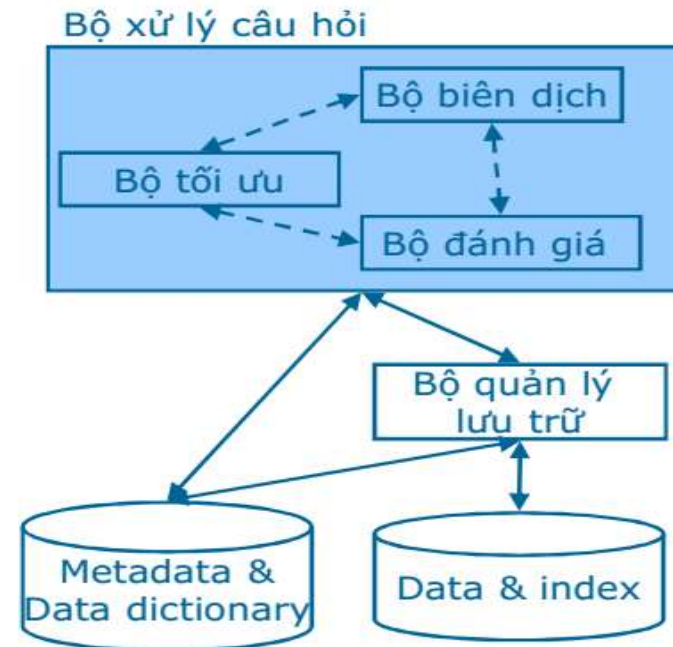
### ❖ Kiến trúc của hệ quản trị CSDL



## 1.3 Tiếp cận CSDL

### Xử lý câu hỏi

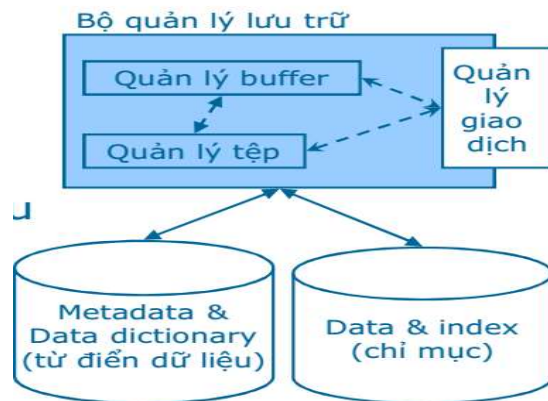
- Yêu cầu
  - Tìm kiếm dữ liệu trả lời cho một yêu cầu truy vấn.
- Thực hiện
  - Biến đổi truy vấn ở mức cao thành các yêu cầu có thể hiểu được bởi hệ CSDL.
  - Lựa chọn một kế hoạch tốt nhất để trả lời truy vấn này.



## 1.3 Tiếp cận CSDL

### Quản lý lưu trữ

- Yêu cầu
  - Lưu trữ và truy xuất dữ liệu trên các thiết bị nhớ
- Thực hiện
  - Tổ chức tối ưu dữ liệu trên thiết bị nhớ
  - Tương tác hiệu quả với bộ quản lý tệp



### Quản trị giao dịch

- Yêu cầu
  - Định nghĩa giao dịch: một tập các thao tác được xử lý như một đơn vị không chia cắt được.
  - Đảm bảo tính đúng đắn và tính nhất quán của dữ liệu.
- Thực hiện
  - Quản lý điều khiển tương tranh.
  - Phát hiện lỗi và phục hồi CSDL

### Ưu điểm của DBMS:

- ❖ Cung cấp tính toàn vẹn và bảo mật dữ liệu
- ❖ Quy trình quản lý dữ liệu thống nhất
- ❖ Các lập trình viên ứng dụng không được tiếp xúc với các chi tiết về biểu diễn và lưu trữ dữ liệu.
- ❖ DBMS có các kỹ thuật khác nhau để lưu trữ và truy xuất dữ liệu.
- ❖ DBMS sử dụng nhiều tính năng mạnh mẽ khác nhau để lưu trữ và truy xuất dữ liệu một cách hiệu quả.
- ❖ DBMS ngụ ý các ràng buộc toàn vẹn để duy trì mức độ bảo vệ cao chống lại việc truy cập trái phép vào dữ liệu.
- ❖ Giảm thời gian phát triển ứng dụng
- ❖ Giảm sự dư thừa.
- ❖ Cung cấp sự độc lập về dữ liệu

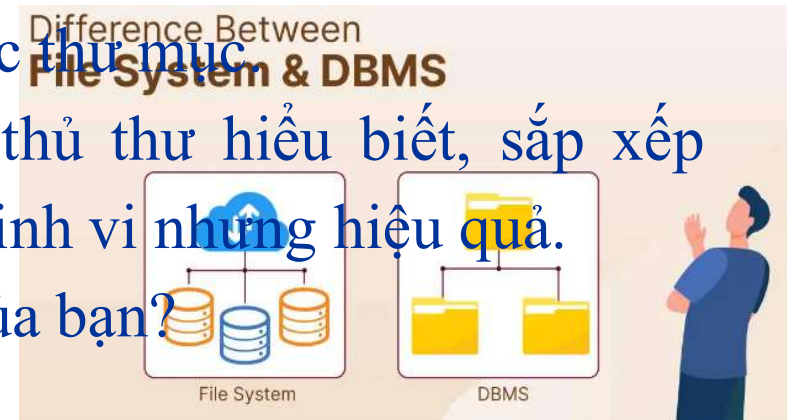
## 1.3 Tiếp cận CSDL

### Nhược điểm của DBMS:

- ❖ Hầu hết các hệ thống quản lý CSDL đều phức tạp và yêu cầu đào tạo người dùng để sử dụng DBMS.
- ❖ Một số dữ liệu có thể bị mất khi nhiều người dùng sử dụng cùng một chương trình cùng một lúc.
- ❖ Chi phí phần cứng và phần mềm DBMS rất cao và làm tăng thêm ngân sách của tổ chức.
- ❖ DBMS không thể thực hiện các phép tính nâng cao
- ❖ Bộ dữ liệu lớn hơn vì chúng cung cấp thời gian phản hồi truy vấn dễ dự đoán hơn.
- ❖ Cần có bộ xử lý nhanh để xử lý dữ liệu.
- ❖ Mất điện có thể khiến CSDL bị lỗi và thậm chí làm hỏng toàn bộ hệ thống.

## Thảo luận: So sánh Hệ thống tệp & DBMS

- Cách thức máy tính của bạn **sắp xếp các tệp** hoặc **cách thức các trang web xử lý hàng núi dữ liệu?**
- Dữ liệu và cách thức tổ chức dữ liệu từ trước đến nay (cách khai thác, lưu trữ dữ liệu) và những hạn chế (nhu cầu cần xây dựng CSDL)
- Câu trả lời nằm ở hai công cụ mạnh mẽ: **Hệ thống tệp và DBMS.**
- Hệ thống Tệp giống như tủ hồ sơ kỹ thuật số, lưu trữ tài liệu của bạn một cách gọn gàng trong các **thư mục**.
- DBMS hoạt động giống như một thủ thư hiểu biết, sắp xếp lượng thông tin khổng lồ một cách tinh vi nhưng hiệu quả.
- Cái nào phù hợp hơn cho nhu cầu của bạn?

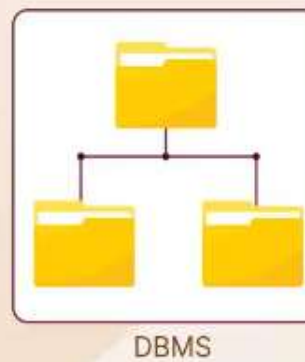
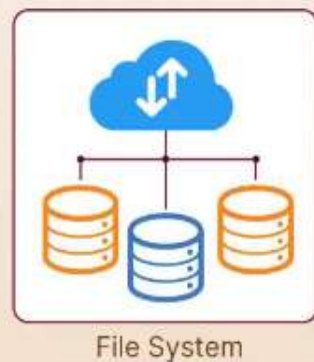




## Thảo luận: So sánh Hệ thống tệp & DBMS

- Hệ quản trị CSDL (DBMS) là một ứng dụng quản lý CSDL.
- Hệ thống tệp là một ứng dụng quản lý hệ thống tệp tin.
- Cả hai đều là giải pháp hợp lý để lưu trữ thông tin - vì vậy điều quan trọng là phải hiểu sự khác biệt của chúng.

### Difference Between **File System & DBMS**



## Sự khác biệt chính giữa Hệ thống tệp và DBMS

- ❖ Hệ thống tệp là tập hợp dữ liệu và bạn phải viết các thủ tục để quản lý nó, trong khi DBMS là tập hợp dữ liệu và bạn không phải viết các thủ tục để làm việc với CSDL.
- ❖ Hệ thống tệp là phần mềm quản lý và sắp xếp các tệp trên phương tiện lưu trữ và DBMS là một ứng dụng phần mềm được sử dụng để truy cập, tạo và quản lý CSDL.
- ❖ Hệ thống tệp không có cơ chế khôi phục sự cố nhưng DBMS thì có.
- ❖ Các hệ thống tệp không hỗ trợ các giao dịch phức tạp, nhưng hệ thống DBMS có thể nhanh chóng triển khai các giao dịch phức tạp bằng SQL.
- ❖ DBMS cung cấp khả năng đồng thời, nhưng hệ thống tệp thì không.
- ❖ Sự không nhất quán dữ liệu cao trong các hệ thống tệp tin. Ngược lại, sự không nhất quán về dữ liệu trong hệ thống quản lý CSDL là rất hiếm.

# Sự khác biệt chính giữa Hệ thống tệp và DBMS

	File System	DBMS
<b>Data Organization</b>	Phân cấp, được lưu trữ trong các tập tin	Có cấu trúc, được lưu trữ trong bảng
<b>Data Integrity (Tính toán vẹn)</b>	Hạn chế, dễ bị dư thừa dữ liệu	Cao, có cơ chế tránh dư thừa
<b>Data Security</b>	Cơ bản, dựa vào bảo mật hệ thống	Nâng cao, với các tính năng bảo mật tích hợp
<b>Data Accessibility</b>	Truy cập trực tiếp, kém linh hoạt	Linh hoạt, với khả năng truy vấn phức tạp
<b>Data Recovery</b>	Hạn chế, Yêu cầu sao lưu thủ công	Toàn diện, với các tùy chọn sao lưu tự động
<b>Scalability</b>	Hạn chế, không lý tưởng cho các tập dữ liệu lớn	Cao, được thiết kế để xử lý các tập dữ liệu lớn
<b>Complexity</b>	Đơn giản, dễ cài đặt	Phức tạp, yêu cầu thiết lập và quản lý ban đầu
<b>Concurrency (Tính đồng thời)</b>	Hạn chế, nguy cơ hỏng dữ liệu	Cao, hỗ trợ sử dụng nhiều lần

## 2. Hệ cơ sở dữ liệu

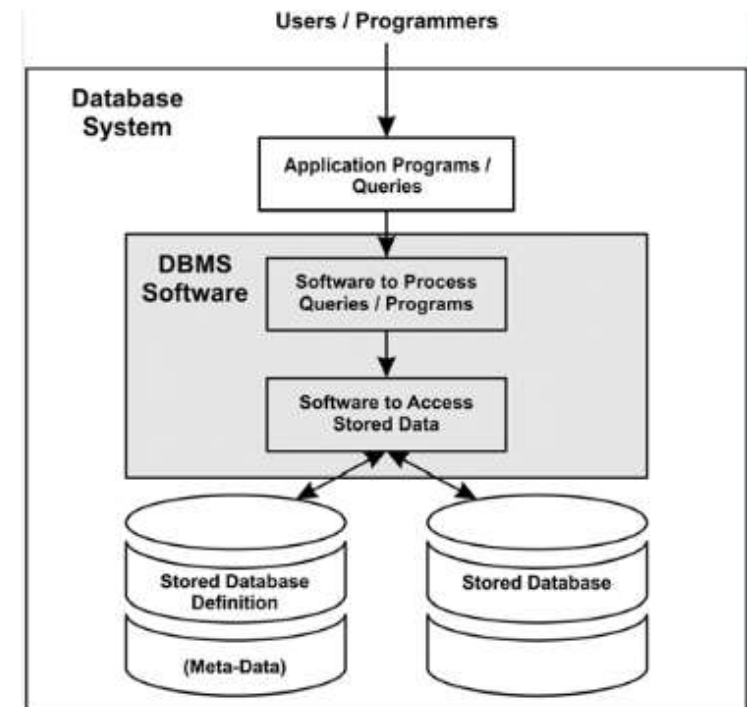
- ❖ Nếu cơ sở dữ liệu là phần thông tin bên trong thì hệ cơ sở dữ liệu được hiểu đơn giản giống như những vỏ bọc cho thông tin.
- ❖ Hệ cơ sở dữ liệu là một tập hợp các phần mềm cho phép người dùng **xác định cấu trúc, nhập dữ liệu và thao tác với dữ liệu** để đảm bảo thông tin được bảo mật và an toàn. Trong đó:
  - Cấu trúc bao gồm các nội dung như: xác định kiểu dữ liệu, xác định cấu trúc dữ liệu, xác định giới hạn dữ liệu.
  - Nhập dữ liệu là một quá trình lưu trữ dữ liệu trên thiết bị lưu trữ. Dữ liệu sau đó được kiểm tra bởi hệ thống quản lý.
  - Thao tác dữ liệu bao gồm những việc như lấy dữ liệu, cập nhật dữ liệu và tổng hợp dữ liệu để báo cáo

## 2. Hệ cơ sở dữ liệu

- **Hệ quản trị cơ sở dữ liệu (DBMS):**
  - Một gói/hệ thống phần mềm hỗ trợ việc tạo và duy trì cơ sở dữ liệu trên máy tính.
- **Hệ cơ sở dữ liệu:**
  - Phần mềm DBMS cùng với chính dữ liệu đó. Đôi khi, các chương trình ứng dụng và giao diện cũng được đưa vào.

### Hệ CSDL:

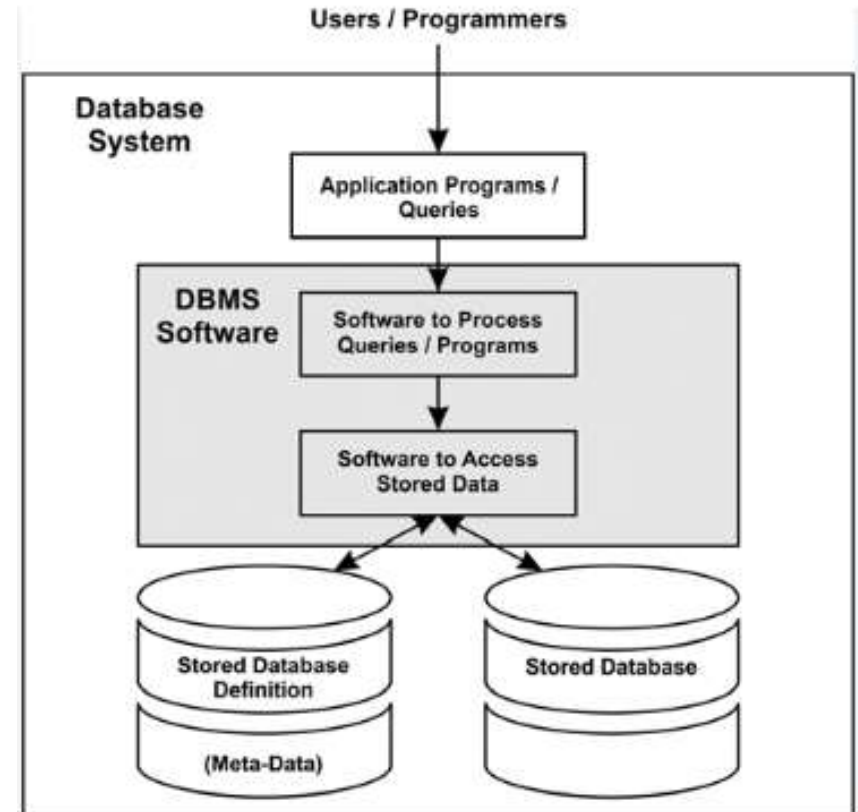
- ❖ Là một hệ thống gồm 4 thành phần:
  - Hệ quản trị CSDL
  - Phần cứng
  - CSDL và phần mềm ứng dụng
  - Những người dùng
- ❖ Ví dụ: Hệ quản lý đào tạo, hệ quản lý nhân sự, hệ quản lý kinh doanh,...



## 2. Hệ cơ sở dữ liệu

**Hệ CSDL:** Ví dụ về Hệ quản lý đào tạo

- Định nghĩa CSDL
  - Định nghĩa cấu trúc cho: MÔN HỌC, SINH VIÊN, HỌC PHẦN, KẾT QUẢ
- Xây dựng CSDL
  - Đưa dữ liệu vào các bảng
- Xử lý CSDL
  - Thực hiện các truy vấn: Tìm kiếm sinh viên được điểm 10 môn Toán rồi rạc
  - Thực hiện cập nhật: Chuyển lớp



## 2. Hệ cơ sở dữ liệu

### Kiến trúc hệ CSDL

❖ Kiến trúc 3 mức: Trục quan hóa 3 trong 4 đặc điểm của tiếp cận CSDL:

- Bản chất tự mô tả của hệ thống cơ sở dữ liệu: sử dụng danh mục để lưu trữ mô tả cơ sở dữ liệu (lược đồ) để làm cho nó tự mô tả
- Cách ly giữa các chương trình và dữ liệu (tính độc lập) và 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

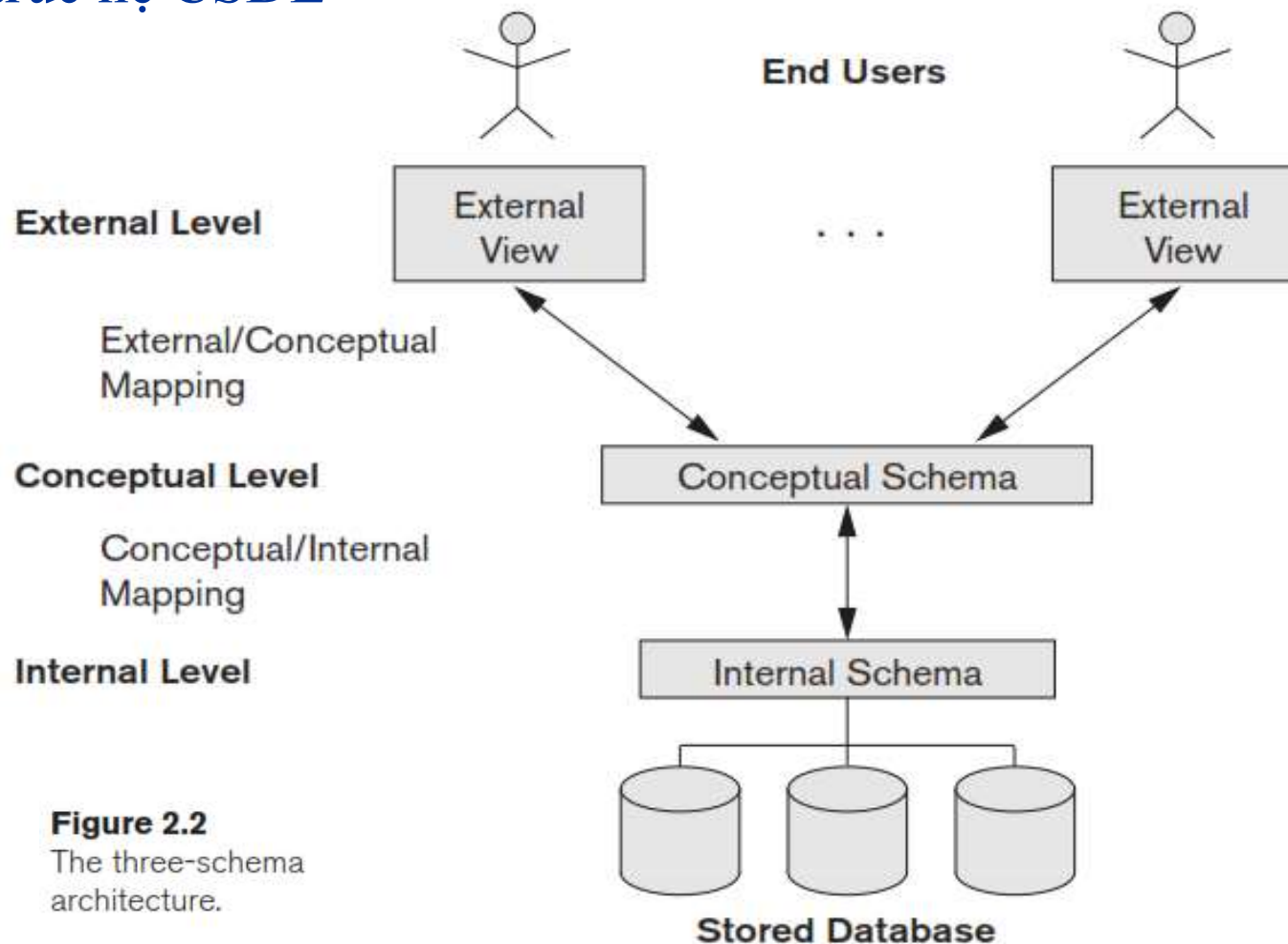
❖ Mục tiêu của kiến trúc 3 mức: là tách các ứng dụng người dùng khỏi cơ sở dữ liệu vật lý. Gồm có:

- Mức lưu trữ (trong)
- Mức quan niệm (logic)
- Mức khung nhìn (ngoài)



## 2. Hệ cơ sở dữ liệu

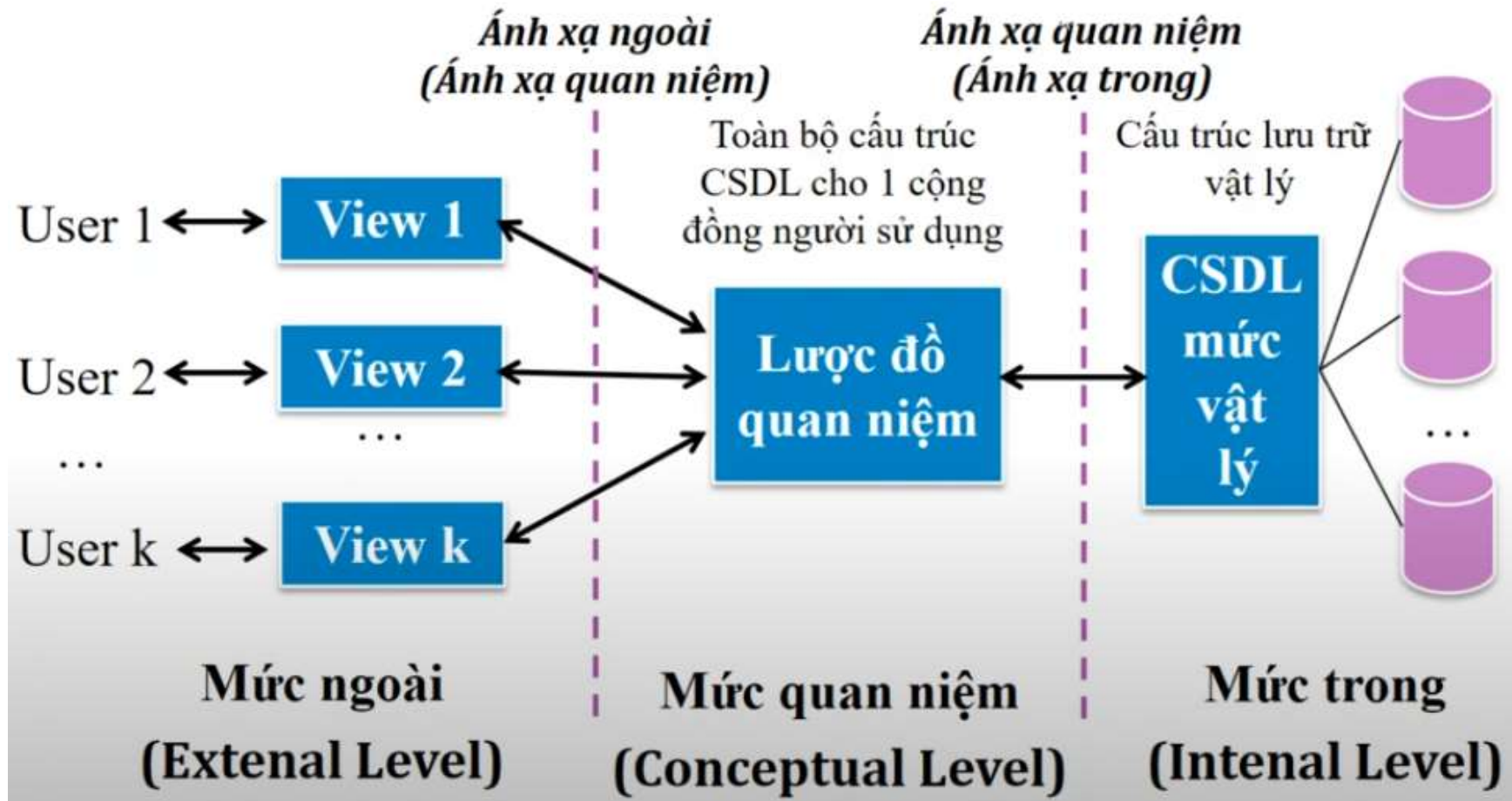
### Kiến trúc hệ CSDL



**Figure 2.2**  
The three-schema architecture.

## 2. Hệ cơ sở dữ liệu

### Kiến trúc hệ CSDL



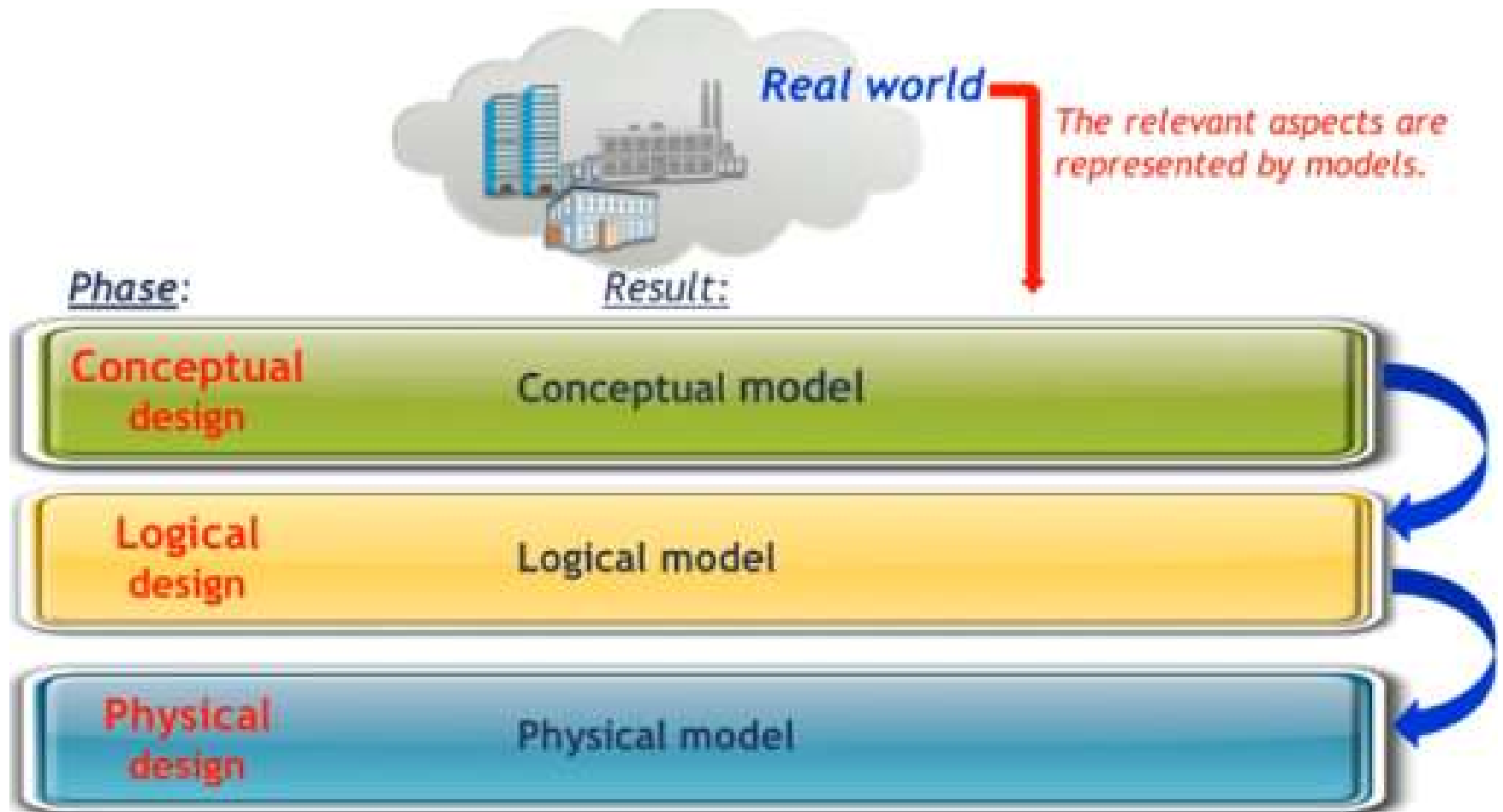
## 2. Kiến trúc hệ CSDL

### Kiến trúc 3 mức:

- Mức lưu trữ (mức trong)
  - Mô tả cấu trúc lưu trữ vật lý CSDL
- Mức quan niệm (mức logic)
  - Mô tả cấu trúc của toàn thể CSDL cho 1 cộng đồng người sử dụng, gồm thực thể, kiểu dữ liệu, mối liên hệ và ràng buộc
  - Che bớt các chi tiết của cấu trúc lưu trữ vật lý
- Mức khung nhìn (mức ngoài)
  - Mô tả một phần của CSDL mà 1 nhóm người dùng quan tâm đến và che dấu phần còn lại của CSDL đối với nhóm người dùng đó

## 2. Kiến trúc hệ CSDL

### ❖ Thiết kế CSDL



## 2. Kiến trúc hệ CSDL

### ❖ Thiết kế CSDL



## 2. Kiến trúc hệ CSDL

### Vòng đời phát triển CSDL



## 2. Kiến trúc hệ CSDL

### ❖ Vòng đời phát triển CSDL





### 3. Phân loại các hệ CSDL

- ❖ Mô hình dữ liệu
  - Mạng vs. phân cấp vs. quan hệ vs. hướng đối tượng vs.
  - ...
- ❖ Số người sử dụng
  - Một người dùng vs. nhiều người dùng
- ❖ Tính phân tán của CSDL
  - Tập trung vs. Phân tán
- ❖ Tính thống nhất của dữ liệu
  - Đồng nhất vs. Không đồng nhất
- ❖ ...

## 3. Phân loại các hệ CSDL

### 3.1. Các hệ cơ sở dữ liệu tập trung

- ❖ Đặc trưng cơ bản để phân biệt kiến trúc một hệ CSDL là **cách tổ chức lưu trữ CSDL**.
- ❖ Với hệ CSDL tập trung, **toàn bộ dữ liệu được lưu tại một máy hoặc một dàn máy**. Những người dùng từ xa có thể truy cập vào CSDL thông qua các phương tiện truyền thông dữ liệu.
- ❖ Việc phân loại phụ thuộc vào **cách tổ chức khai thác**, cụ thể là:
  - Những ai sẽ truy cập vào CSDL?
  - Việc truy cập được thực hiện từ đâu
  - Các môđun của hệ QTCSDL được lưu trữ ở đâu?

## 3. Phân loại các hệ CSDL

### 3.1. Các hệ cơ sở dữ liệu tập trung

❖ Dựa vào các tiêu chí trên người ta chia các **hệ CSDL tập trung thành 3 loại:**

- Hệ CSDL cá nhân;
- Hệ CSDL trung tâm;
- Hệ CSDL khách - chủ.



Hình 1. Hệ cơ sở dữ liệu cá nhân

#### a. Hệ CSDL cá nhân

❖ **Khái niệm:** Hệ CSDL cá nhân là hệ CSDL có một người dùng. Thông thường, người này vừa thiết kế, tạo lập và bảo trì CSDL, đồng thời cũng là người khai thác thông tin, tự lập và hiển thị các báo cáo.

❖ **Đặc điểm:**

- Dữ liệu được tập trung ở một máy;

## 3. Phân loại các hệ CSDL

### 3.1. Các hệ cơ sở dữ liệu tập trung

#### a. Hệ CSDL cá nhân

##### ❖ Đặc điểm:

- Dữ liệu được tập trung ở một máy;
- Chỉ một người hoặc một nhóm người truy cập theo nguyên tắc lần lượt (không có tương tranh, xung đột);
- Toàn bộ hệ QTCSDL được lưu ngay tại hệ thống chứa CSDL;
- Việc truy cập vào CSDL được thực hiện ngay hệ thống chứa CSDL.



Hình 1. Hệ cơ sở dữ liệu cá nhân

##### ❖ Ưu điểm và nhược điểm:

- **Ưu điểm:** Việc phát triển và sử dụng các hệ CSDL cá nhân khá đơn giản và dễ dàng.
- **Nhược điểm:** Tình an toàn thường không cao.

## 3. Phân loại các hệ CSDL

### 3.1. Các hệ cơ sở dữ liệu tập trung

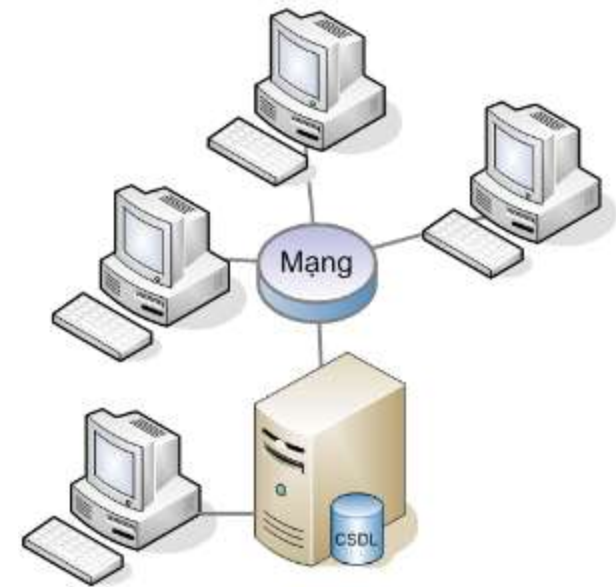
#### b. Hệ CSDL trung tâm

❖ **Khái niệm:** Hệ cơ sở dữ liệu trung tâm là hệ CSDL được cài đặt trên máy tính trung tâm.

❖ **Đặc điểm:**

- Dữ liệu được lưu trữ trên máy tính trung tâm.
- Nhiều người dùng từ xa có thể truy cập CSDL.

❖ Tùy thuộc vào việc tổ chức: máy tính trung tâm có thể là một máy hay một dàn máy và có rất nhiều người dùng. Việc truy cập thông qua thiết bị đầu cuối và phương tiện truyền thông.



Hình 2. Hệ cơ sở dữ liệu trung tâm

## 3. Phân loại các hệ CSDL

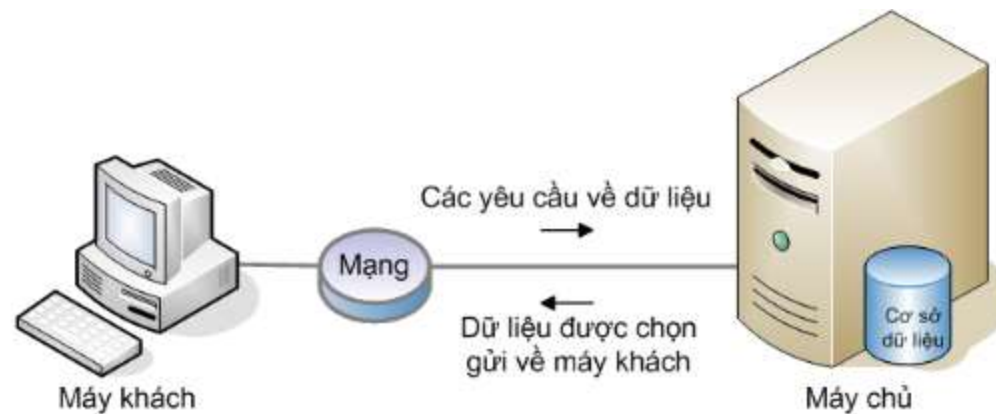
### 3.1. Các hệ cơ sở dữ liệu tập trung

#### c. Hệ CSDL khách - chủ

❖ Trong kiến trúc khách-chủ, các thành phần của hệ QTCSDL:

- Bộ phận cung cấp tài nguyên được đặt ở máy chủ.
- Bộ phận yêu cầu cấp phát tài nguyên đặt ở máy khách.

Hai thành phần này không nhất thiết phải cài đặt trên cùng một máy tính.



Hình 3. Hệ CSDL khách-chủ

## 3. Phân loại các hệ CSDL

### 3.1. Các hệ cơ sở dữ liệu tập trung

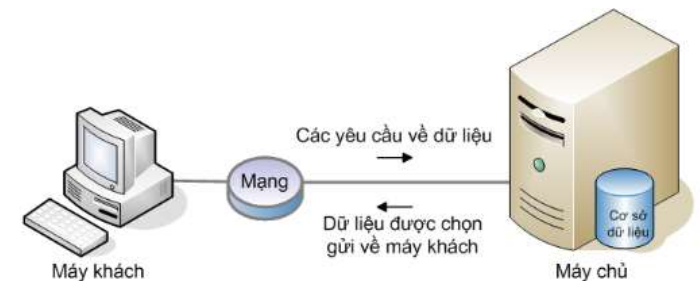
#### c. Hệ CSDL khách - chủ

##### ❖ Phần mềm máy khách:

- Tiếp nhận kết quả và tổ chức đưa ra khuôn dạng phù hợp.
- Chuyển tới yêu cầu của máy chủ và chờ đợi trả lời;
- Tổ chức giao diện, tiếp nhận yêu cầu truy cập thông tin;

##### ❖ Phần mềm máy chủ:

- Tiếp nhận yêu cầu truy vấn thông tin;
- Xử lý
- Gửi kết quả tới máy khách.



Hình 3. Hệ CSDL khách-chủ

- ❖ Việc xử lý yêu cầu truy cập thông tin được thực hiện theo nguyên lý truy cập từ xa (**RPC – Remote Procedure Call**).



## 3. Phân loại các hệ CSDL

### 3.1. Các hệ cơ sở dữ liệu tập trung

#### c. Hệ CSDL khách - chủ

##### ❖ Kiến trúc loại này có một số ưu điểm sau:

- Khả năng truy cập rộng rãi đến các CSDL;
- Nâng cao khả năng thực hiện của CPU: các CPU ở máy chủ và máy khách khác nhau có thể cùng chạy song song, mỗi CPU thực hiện một nhiệm vụ;
- Chi phí cho phần cứng giảm do chỉ cần máy chủ có cấu hình mạnh để lưu trữ và quản trị dữ liệu;
- Chi phí truyền thông được giảm do một phần các thao tác được giải quyết trên máy khách, chỉ cần: yêu cầu về truy cập CSDL gửi đến máy chủ và dữ liệu kết quả gửi về máy khách;
- Nâng cao khả năng đảm bảo tính nhất quán của dữ liệu vì các ràng buộc được định nghĩa và kiểm tra trên máy chủ.

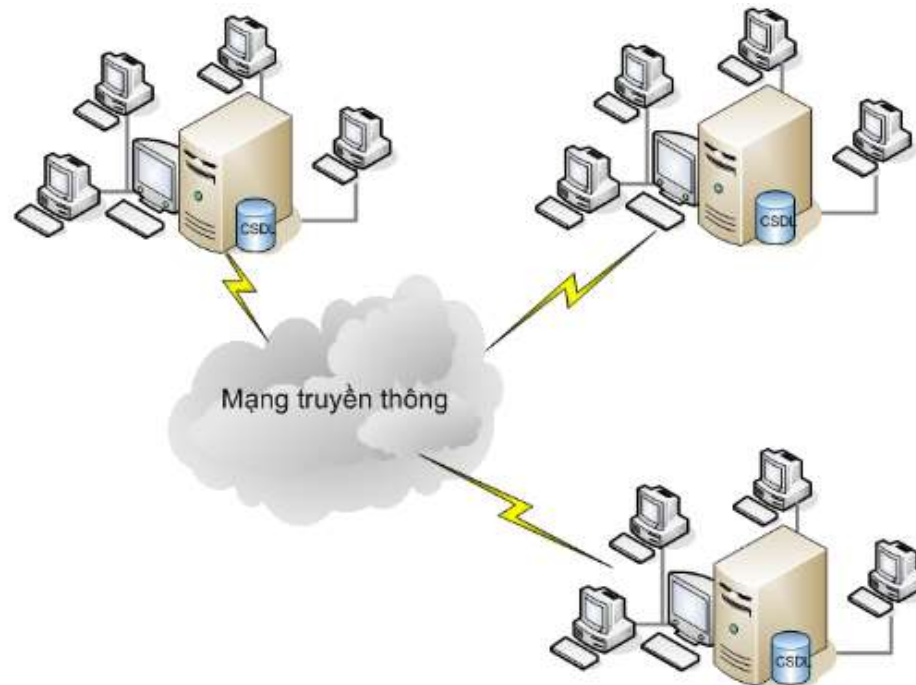
Kiến trúc này phù hợp với việc xây dựng các hệ thống mở.

## 3. Phân loại các hệ CSDL

### 3. 2. Các hệ CSDL phân tán

#### a. Khái niệm CSDL phân tán

- ❖ **CSDL phân tán** là một tập hợp dữ liệu có liên quan (về logic) được dùng chung và phân tán về mặt vật lí trên một mạng máy tính.



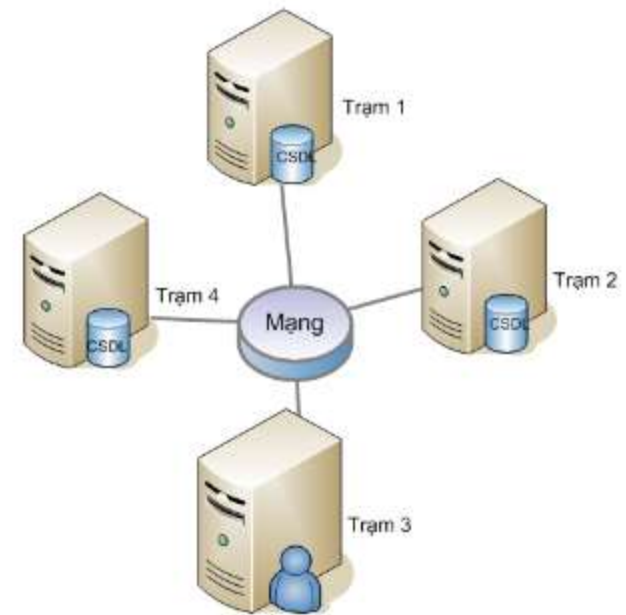
Hình 4. Cơ sở dữ liệu phân tán

## 3. Phân loại các hệ CSDL

### 3.2. Các hệ CSDL phân tán

#### a. Khái niệm CSDL phân tán

- ❖ Một **hệ QTCSDL phân tán** là một hệ thống phần mềm cho phép quản trị CSDL phân tán và làm cho người sử dụng không nhận thấy sự phân tán về lưu trữ dữ liệu.
- ❖ **Người dùng truy cập vào CSDL phân tán thông qua chương trình ứng dụng.** Các chương trình ứng dụng được chia làm hai loại:
  - Chương trình không yêu cầu dữ liệu từ nơi khác;
  - Chương trình có yêu cầu dữ liệu từ nơi khác.



Hình 5. Hệ CSDL phân tán

## 3. Phân loại các hệ CSDL

### 3.2. Các hệ CSDL phân tán

#### a. Khái niệm CSDL phân tán

- ❖ Có thể chia các hệ CSDL phân tán thành 2 loại chính
  - **Hệ CSDL phân tán thuần nhất:** các nút trên mạng đều dùng cùng một hệ QTCSDL.
  - **Hệ CSDL phân tán hỗn hợp:** các nút trên mạng có thể dùng các hệ QTCSDL khác nhau.

#### b. Một số ưu điểm và hạn chế của các hệ CSDL phân tán

- ❖ Sự phân tán dữ liệu và các ứng dụng có **một số ưu điểm** so với các hệ CSDL tập trung:
  - Cấu trúc phân tán dữ liệu thích hợp cho bản chất phân tán của nhiều người dùng.
  - Dữ liệu được chia sẻ trên mạng nhưng vẫn cho phép quản trị dữ liệu địa phương (dữ liệu đặt tại mỗi trạm)

## 3. Phân loại các hệ CSDL

### 3.2. Các hệ CSDL phân tán

#### b. Một số ưu điểm và hạn chế của các hệ CSDL phân tán

- ❖ Sự phân tán dữ liệu và các ứng dụng có **một số ưu điểm** so với các hệ CSDL tập trung:
  - Dữ liệu có tính sẵn sàng cao.
  - Dữ liệu có tính tin cậy cao vì khi một nút gặp sự cố, có thể khôi phục được dữ liệu tại đây do bản sao của nó có thể được lưu trữ tại một nút khác nữa.
  - Hiệu năng của hệ thống được nâng cao hơn.
  - Cho phép mở rộng các tổ chức một cách linh hoạt. Có thể thêm nút mới vào mạng máy tính mà không ảnh hưởng đến hoạt động của các nút sẵn có.

## 3. Phân loại các hệ CSDL

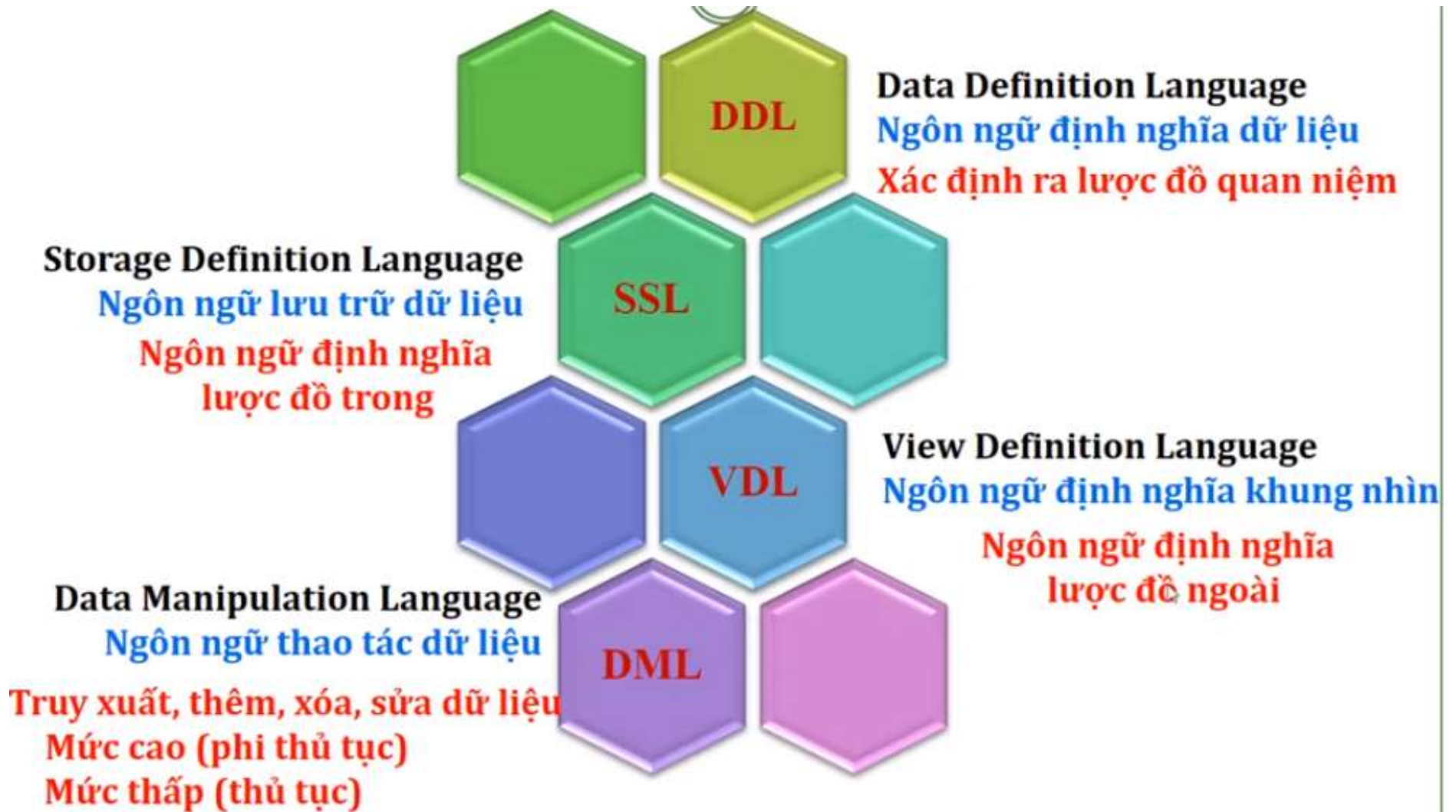
### 3.2. Các hệ CSDL phân tán

#### b. Một số ưu điểm và hạn chế của các hệ CSDL phân tán

❖ So với các hệ CSDL tập trung, hệ CSDL phân tán có **một số hạn chế** như sau:

- Hệ thống phức tạp hơn vì phải làm ẩn đi sự phân tán dữ liệu đối với người dùng.
- Chi phí cao hơn.
- Đảm bảo an ninh khó khăn hơn.
- Đảm bảo tính nhất quán dữ liệu khó hơn.
- Việc thiết kế CSDL phân tán phức tạp hơn.

## 4. Ngôn ngữ CSDL





## 4. Ngôn ngữ CSDL

- Ngôn ngữ định nghĩa dữ liệu (Data Definition Language - DDL)
  - Cấu trúc dữ liệu
  - Mối liên hệ giữa các dữ liệu và quy tắc, ràng buộc áp đặt lên dữ liệu
- Ngôn ngữ thao tác dữ liệu (Data Manipulation Language - DML)
  - Tìm kiếm, thêm, xóa, sửa dữ liệu trong CSDL
- Ngôn ngữ điều khiển dữ liệu (Data Control Language - DCL)
  - Thay đổi cấu trúc của các bảng dữ liệu
  - Khai báo bảo mật thông tin
  - Quyền hạn của người dùng trong khai thác CSDL