

CHƯƠNG 1

Tổng quan về Cơ sở dữ liệu

- Giới thiệu
- Tiếp cận CSDL so với tập tin
- Một số đặc tính của CSDL
- Các vai trò trong CSDL
- Các tính năng của HQT CSDL
- Khái niệm mô tả CSDL
- Kiến trúc ba lược đồ
- Ngôn ngữ CSDL

Thế giới thực – lĩnh vực
áp dụng

Quản trị, kinh doanh, sản xuất, ngân hàng,
giáo dục, giải trí, xã hội, ... → **phức tạp,
giàu ngữ nghĩa**

Phát triển hệ thống

Tự động hoá, hỗ trợ

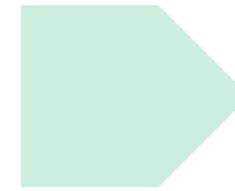
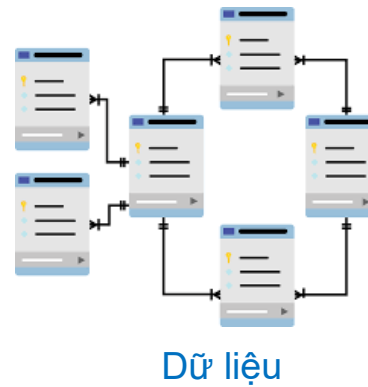


**Làm sao máy tính có
thể hiểu lĩnh vực thế
giới thực để số hóa &
hỗ trợ tự động hóa ?**

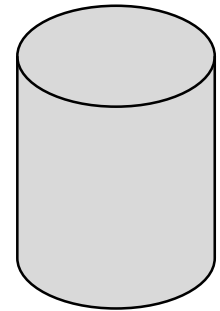
■ Làm sao?



Tìm hiểu, phân tích
Thiết kế



Cài đặt



Quy trình, qui định, giao
diện, v.v...



Ứng dụng/
phần mềm

■ Dữ liệu (Data)

- Dữ liệu là những số liệu rời rạc mô tả về sự kiện, sự vật, hiện tượng được chọn lọc để lưu trữ với một mục đích xác định.

- Ví dụ:
 - “Nguyễn Văn A” là tên của sinh viên
 - “11.12.008 ” là mã số một sinh viên
 - “19-02-2015” là ngày Tết âm lịch năm 2015

■ Cơ sở dữ liệu (Database) -CSDL

- Định nghĩa: “**tập dữ liệu có liên quan với nhau gắn với một ngữ nghĩa**”
- Ví dụ:
 - Danh sách sinh viên → tập dữ liệu về sinh viên
 - Danh mục lớp học → tập dữ liệu về lớp học
 - Danh mục các đề án, danh mục nhân viên, danh mục phòng ban, v.v...
- Các dữ liệu được lưu trữ dùng chung một cấu trúc → CSDL cấu trúc
- Ngoài ra còn có các dạng CSDL khác: CSDL phi cấu trúc, CSDL tài liệu, v.v...

■ Ví dụ về CSDL

MÔN HỌC	Tên MH	Mã MH	Số TC	Khoa
	Khoa học máy tính	CS1310	4	CNTT
	Cấu trúc dữ liệu	CS3320	4	CNTT
	Toán rời rạc	MATH2410	3	TOÁN
	Cơ sở dữ liệu	CS3380	3	CNTT

SINH VIÊN	Tên	MSSV	Lớp	Khoa
	Trang	17	1	CNTT
	Ngọc	8	2	CNTT

HỌC PHẦN	Mã HP	Mã MH	Học Kỳ	Năm	Giáo Viên
	85	MATH2410	1	2008	Anh
	92	CS1310	1	2007	Tiền
	112	MATH2410	2	2008	Anh
	119	CS1310	2	2007	Tiền

KẾT QUẢ	MSSV	Mã HP	Điểm
	17	112	10
	17	119	7
	8	85	6
	8	92	9

ĐIỀU KIỆN	Mã MH	Mã MH Trước
	CS3380	CS3320
	CS3380	MATH2410
	CS3320	CS1310

❑ Ví dụ - CSDL “Quản lý đề án” của một công ty

NHANVIEN	HONV	TENLOT	TENNV	MANV	NGSINH	MA_NQL	PHG
	Tran	Hong	Quang	987987987	03/09/1969	987654321	4
	Nguyen	Thanh	Tung	333445555	12/08/1955	888665555	5
	Nguyen	Manh	Hung	666884444	09/15/1962	333445555	5
	Tran	Thanh	Tam	453453453	07/31/1972	333445555	5

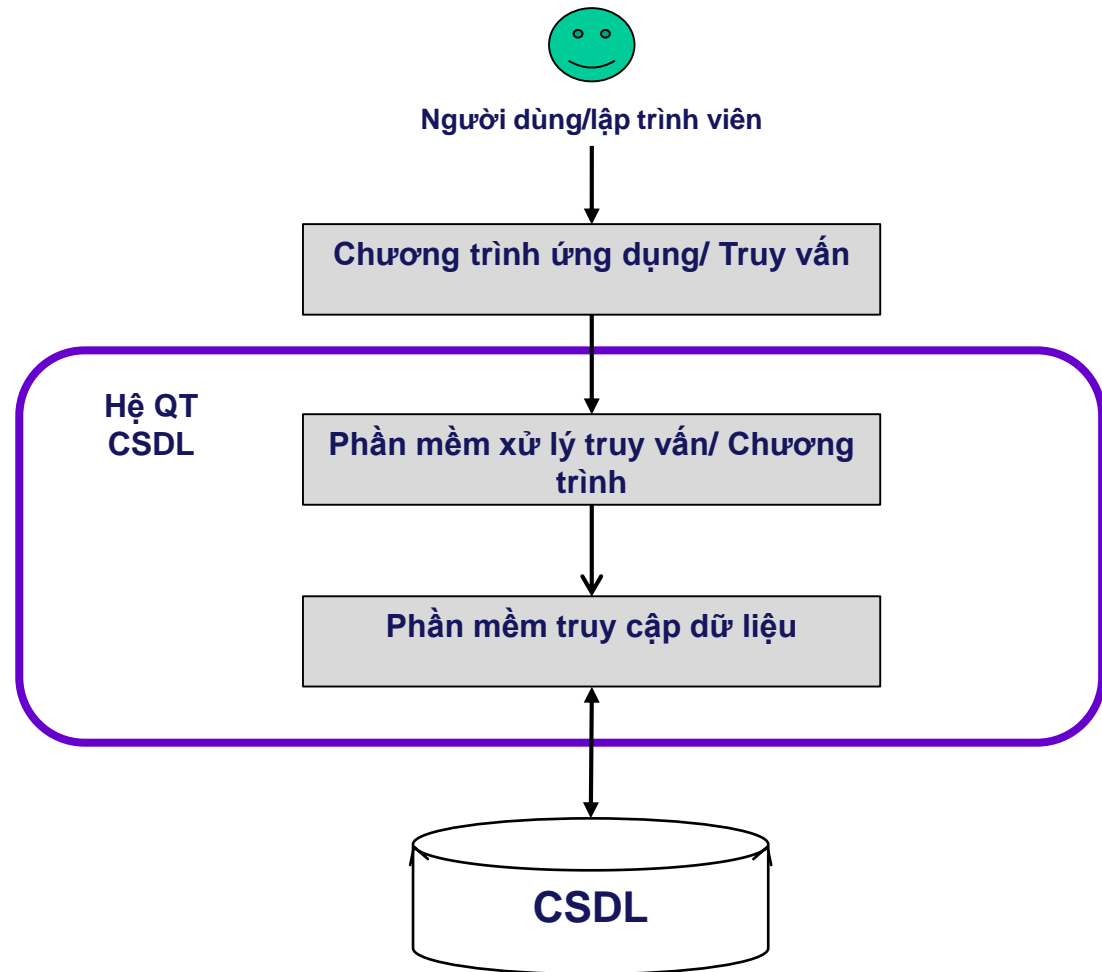
DEAN	TENDA	MADA	DDIEM_DA	PHONG
	San pham X	1	VUNG TAU	5
	San pham Y	2	NHA TRANG	5
	San pham Z	3	TP HCM	5
	Tin hoc hoa	10	HA NOI	4

PHANCONG	MA_NVIENT	SODA	THOIGIAN
	123456789	1	32.5
	123456789	2	7.5
	666884444	3	40.0
	453453453	1	20.0

■ Đặc trưng CSDL

- Một CSDL biểu diễn một phần của thế giới thực (thế giới thu nhỏ)
- CSDL được thiết kế, xây dựng, và lưu trữ với một mục đích xác định, phục vụ cho một số ứng dụng và người dùng
- Tập ngẫu nhiên của các dữ liệu không thể xem là một CSDL

- Hệ quản trị CSDL
(Database Management System)
 - HQT CSDL
 - “Tập hợp các chương trình cho phép người dùng tạo, vận hành và duy trì CSDL”



Môi trường 1 HQT CSDL cơ bản

Định nghĩa

- khai báo bộ khung dữ liệu cùng với các mô tả chi tiết về dữ liệu

Xây dựng

- lưu trữ dữ liệu lên các phương tiện lưu trữ

Xử lý

- Truy vấn, cập nhật và phát sinh báo cáo

Chia sẻ

- Cho phép nhiều người dùng và chương trình truy cập đồng thời CSDL

Bảo vệ

- Đảm bảo dữ liệu được lưu trữ an toàn từ các sự cố, ngăn cản truy cập không được phép

...

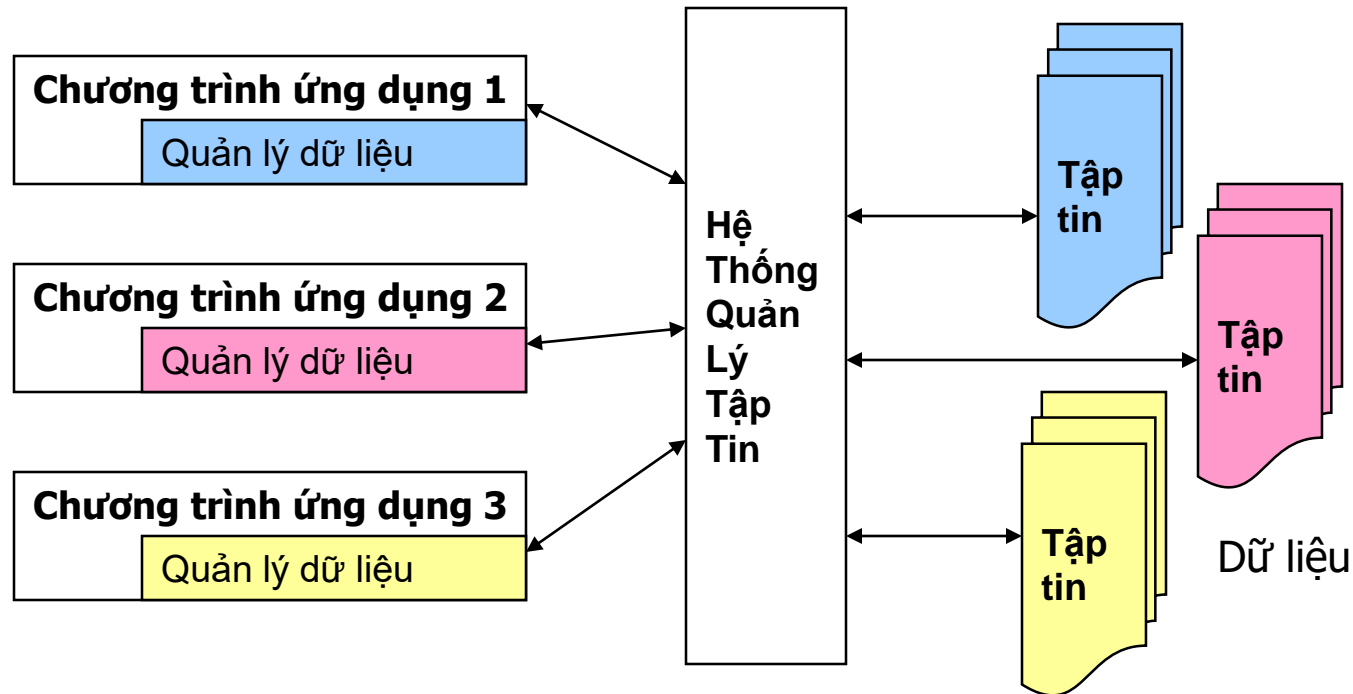
- CSDL “Quản lý đề án” của một công ty
 - Định nghĩa CSDL
 - Định nghĩa cấu trúc cho: NHANVIEN, DEAN và PHANCONG
 - Xây dựng CSDL
 - Đưa dữ liệu vào các bảng
 - Xử lý CSDL
 - Thực hiện các truy vấn: “Cho biết những nhân viên thuộc phòng 5”
 - Thực hiện các phép cập nhật: “Chuyển nhân viên Nguyễn Thanh Tùng sang phòng số 1”

- Mục tiêu của CSDL là gì?

- HQT CSDL là:
 - ☐ Một tập dữ liệu có chung cấu trúc
 - ☐ Các dữ liệu rời rạc
 - ☐ Công cụ cho hỗ trợ lập trình dữ liệu và ứng dụng
 - ☐ Tập chương trình giúp tạo, vận hành và duy trì CSDL

- Giới thiệu
- **Tiếp cận CSDL so với tập tin**
- Một số đặc tính của CSDL
- Các vai trò trong CSDL
- Các tính năng của HQT CSDL
- Khái niệm mô tả CSDL
- Kiến trúc ba lược đồ
- Ngôn ngữ CSDL

- Tiếp cận quản lý dữ liệu theo tập tin (File)

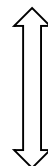


→ Chương trình khai báo cấu trúc CSDL

- Tiếp cận quản lý dữ liệu theo tập tin – Ví dụ:

Chương trình

```
...  
struct SINHVIEN  
{  
    char[10] masv;  
    char[100] hoten;  
};  
...
```



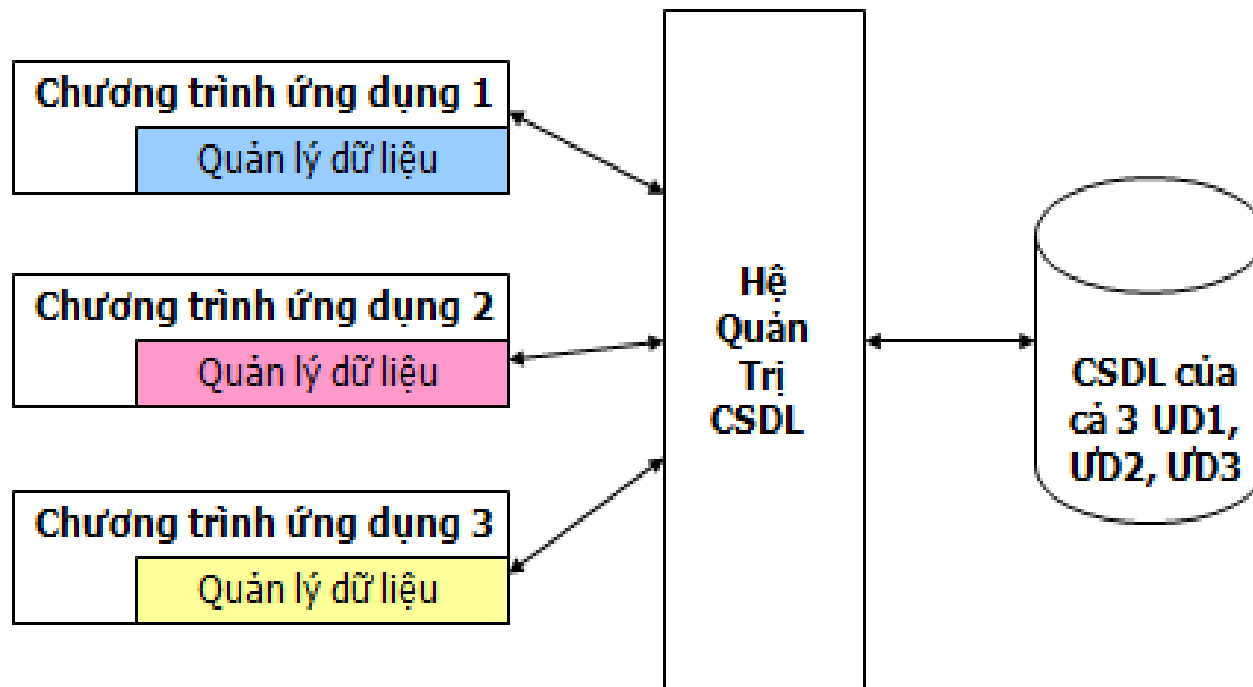
Tập tin dữ liệu

```
<12345, Nguyen Van X>  
<54321, Tran Van Y>  
<21345, Vo Van Z>  
...
```


■ Hạn chế

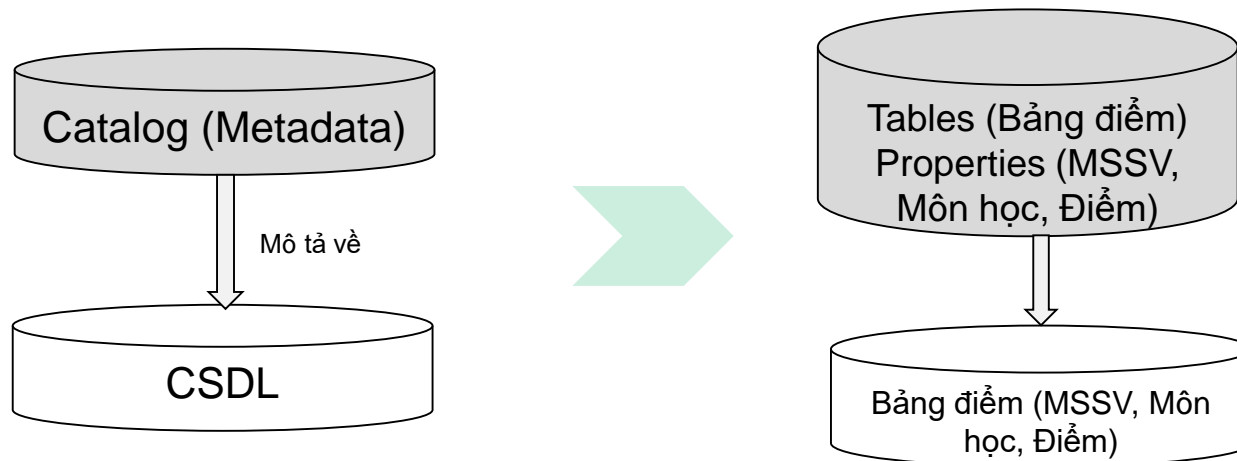
- Khó kiểm soát trùng lặp và dư thừa dữ liệu
- Khó đảm bảo tính nhất quán giữa các dữ liệu
- Khó khăn trong việc truy xuất
- Việc chia sẻ dữ liệu bị hạn chế
- Khó khôi phục

- Tiếp cận Cơ sở dữ liệu (Database)

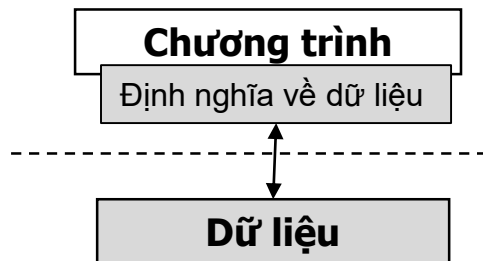


- Giới thiệu
- Tiếp cận CSDL so với tập tin
- **Một số đặc tính của CSDL**
 - Tính tự mô tả
 - Tính cô lập giữa chương trình và dữ liệu
 - Tính trừu tượng hóa dữ liệu
 - Hỗ trợ nhiều khung nhìn dữ liệu
- Các vai trò trong CSDL
- Các tính năng của HQT CSDL
- Khái niệm mô tả CSDL
- Kiến trúc ba lược đồ
- Ngôn ngữ CSDL

- Hệ CSDL không chỉ chứa bản thân CSDL mà còn chứa thông tin định nghĩa đầy đủ (mô tả catalog) của CSDL
- Các định nghĩa được lưu trữ trong catalog gọi là metadata
 - 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
- Các CTƯD có thể truy xuất đến nhiều CSDL nhờ thông tin cấu trúc được lưu trữ trong catalog về những CSDL đó.

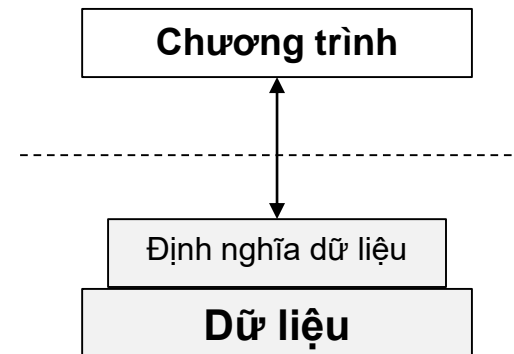


- CSDL và mô tả của nó (catalog) được quản lý chung → quản lý CSDL độc lập tương đối với chương trình ứng dụng



Hệ thống xử lý tập tin:

→ Chương trình chứa mô tả về dữ liệu → tạo ra sự phụ thuộc giữa chương trình và dữ liệu mà nó sử dụng

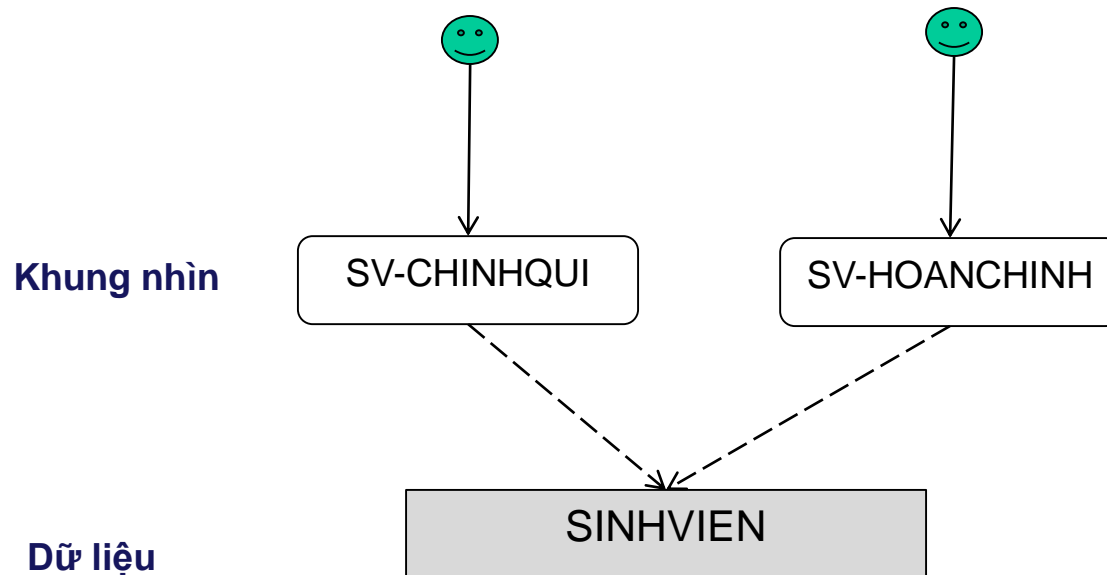


Hệ CSDL:

→ Hệ QT CSDL chứa dữ liệu và mô tả về dữ liệu → tạo ra sự độc lập giữa chương trình và dữ liệu mà nó sử dụng

- Hệ CSDL cho phép trình bày dữ liệu ở một mức trừu tượng (gần với lãnh vực ứng dụng) nhằm che bớt những chi tiết lưu trữ và cài đặt của dữ liệu
 - Người dùng và chương trình truy xuất đến các “thành phần” trừu tượng thay vì các chi tiết vật lý.
- Trừu tượng hóa dữ liệu thông qua mô hình dữ liệu:
 - Cho phép mô tả dữ liệu dùng các khái niệm luận lý như:
 - Đối tượng
 - Thuộc tính
 - Liên kết

- Hỗ trợ định nghĩa nhiều khung nhìn (view) khác nhau về cùng một dữ liệu



- Giới thiệu
- Tiếp cận CSDL so với tập tin
- Một số đặc tính của CSDL
- **Các vai trò trong CSDL**
 - Quản trị viên (Database Administrator - DBA)
 - Thiết kế viên (Database Designer)
 - Lập trình viên CSDL (database programmer)
 - Người dùng cuối (End User)
- Các tính năng của HQT CSDL
- Khái niệm mô tả CSDL
- Kiến trúc ba lược đồ
- Ngôn ngữ CSDL

- **Quản trị viên CSDL (DBA – Database Administrator)**
 - Có trách nhiệm quản lý hệ CSDL
 - Cấp quyền truy cập CSDL
 - Điều phối và giám sát việc sử dụng CSDL
- **Thiết kế viên CSDL**
 - 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
- **Lập trình viên CSDL**
 - Lập trình các chức năng nhằm quản lý và khai thác CSDL

■ Người dùng cuối

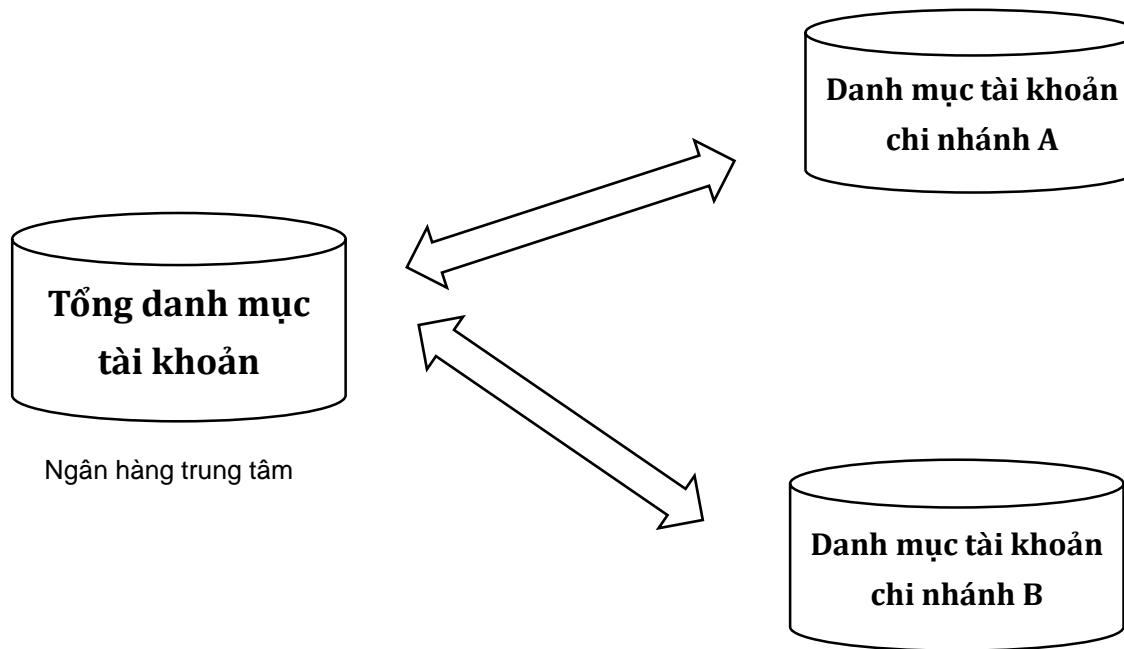
- Người ít sử dụng
 - Ít khi truy cập CSDL, nhưng cần những thông tin khác nhau trong mỗi lần truy cập và dùng những câu truy vấn phức tạp
 - Người quản lý
- Người sử dụng thường xuyên
 - Thường xuyên truy vấn và cập nhật CSDL nhờ vào một số các chức năng đã được xây dựng sẵn
 - Nhân viên
- Người sử 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 kinh doanh,...

- Giới thiệu
- Tiếp cận CSDL so với tập tin
- Một số đặc tính của CSDL
- Các vai trò trong CSDL
- **Các tính năng của HQT CSDL**
- Khái niệm mô tả CSDL
- Kiến trúc ba lược đồ
- Ngôn ngữ CSDL

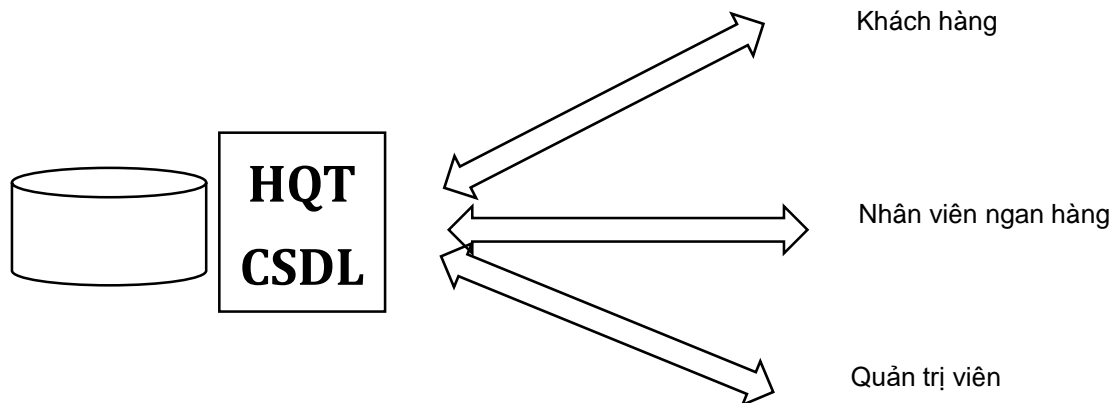
- HQT CSDL (DBMS) là phần mềm hệ thống được thiết kế để định nghĩa, thao tác, truy vấn và quản lý dữ liệu trong CSDL.
- HQT CSDL giúp người dùng cuối có thể tạo, đọc, cập nhật và xóa dữ liệu trong CSDL, đóng vai trò là giao diện giữa CSDL và người dùng cuối hoặc chương trình ứng dụng, đảm bảo dữ liệu được tổ chức nhất quán và vẫn có thể truy cập dễ dàng.



- 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



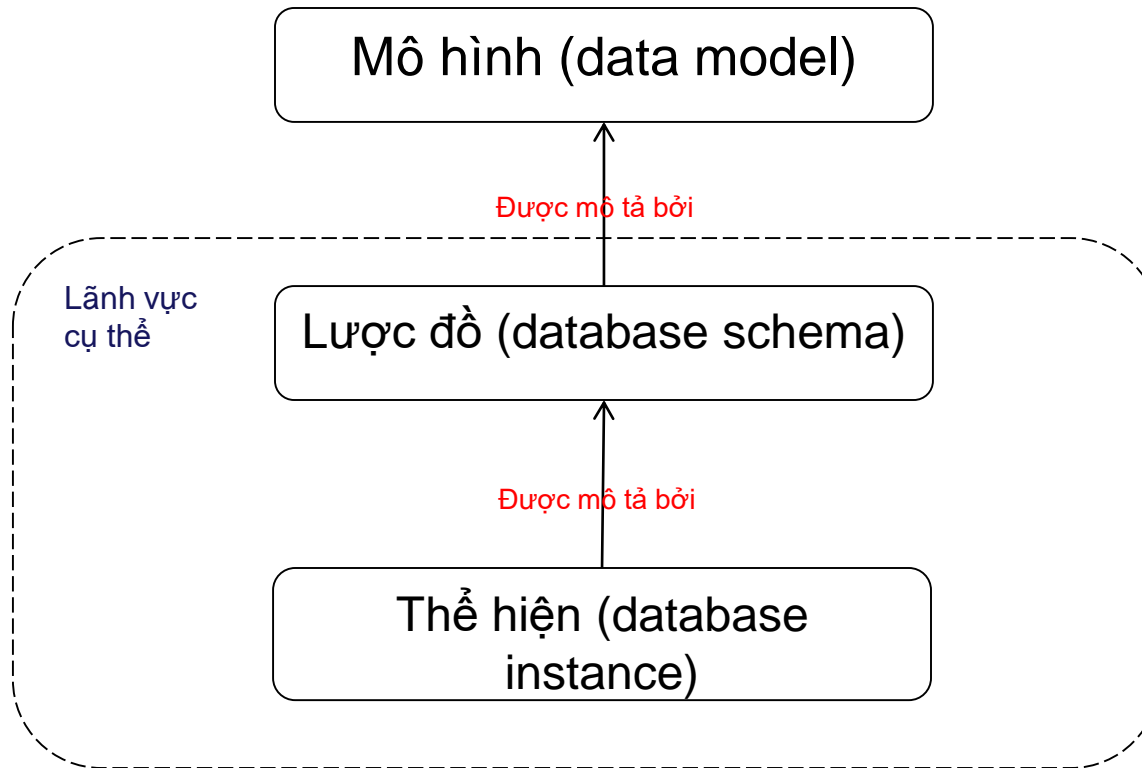
- **Đảm bảo các ràng buộc toàn vẹn**
 - RBTV (Integrity Constraints) là những qui định cần được thỏa mãn để đảm bảo dữ liệu luôn phản ánh đúng ngữ nghĩa của thế giới thực
 - Một số RB có thể được khai báo với HQT và HQT sẽ tự động kiểm tra. Một số RB khác được kiểm tra nhờ CTƯD
- **Khả năng sao lưu dự 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 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

- Giới thiệu
- Tiếp cận CSDL so với tập tin
- Một số đặc tính của CSDL
- Người sử dụng CSDL
- Các tính năng của HQT CSDL
- **Khái niệm mô tả CSDL**
 - Mô hình dữ liệu
 - Lược đồ
 - Thể hiện
- Kiến trúc ba lược đồ
- Ngôn ngữ CSDL

- Liên hệ giữa các khái niệm biểu diễn CSDL



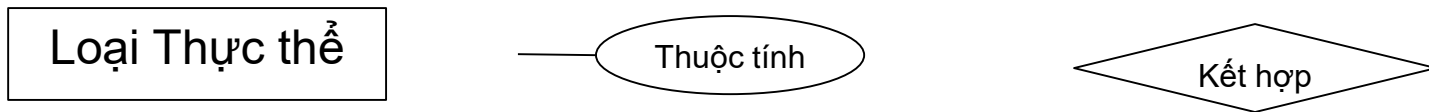
- Mô hình dữ liệu (Data model) bao gồm
 - Cung cấp tập khái niệm dùng mô tả CSDL
 - Ký hiệu, kiểu dữ liệu, quan hệ, ràng buộc...
 - Cung cấp các phép toán xử cơ bản lý dữ liệu:
 - Truy vấn và cập nhật CSDL

■ Các loại mô hình dữ liệu

- Mô hình mức cao hoặc mô hình dữ liệu quan niệm
 - Cung cấp tập các khái niệm mô tả CSDL gần gũi với người dùng
 - Tự nhiên và giàu ngữ nghĩa
 - VD: mô hình thực thể kết hợp (ER), mô hình đối tượng...
- Mô hình cài đặt
 - Cung cấp tập các khái niệm mô tả CSDL mà người dùng có thể hiểu được nhưng không quá xa với cách dữ liệu được tổ chức thật sự trên máy tính
 - VD: mô hình quan hệ, mô hình mạng, mô hình phân cấp
- Mô hình mức thấp (mô hình vật lý)
 - Cung cấp tập các khái niệm mô tả chi tiết về cách thức CSDL được lưu trữ trong máy tính

■ Ví dụ 1 – Mô hình thực thể kết hợp (ER)

- Một số khái niệm:



■ Mô hình mạng

- Một số khái niệm:



■ Lược đồ CSDL (Database Schema)

- Là các mô tả về cấu trúc và ràng buộc trên CSDL về một ứng dụng xác định (ngân hàng, y tế, giáo dục, bán hàng, ...)

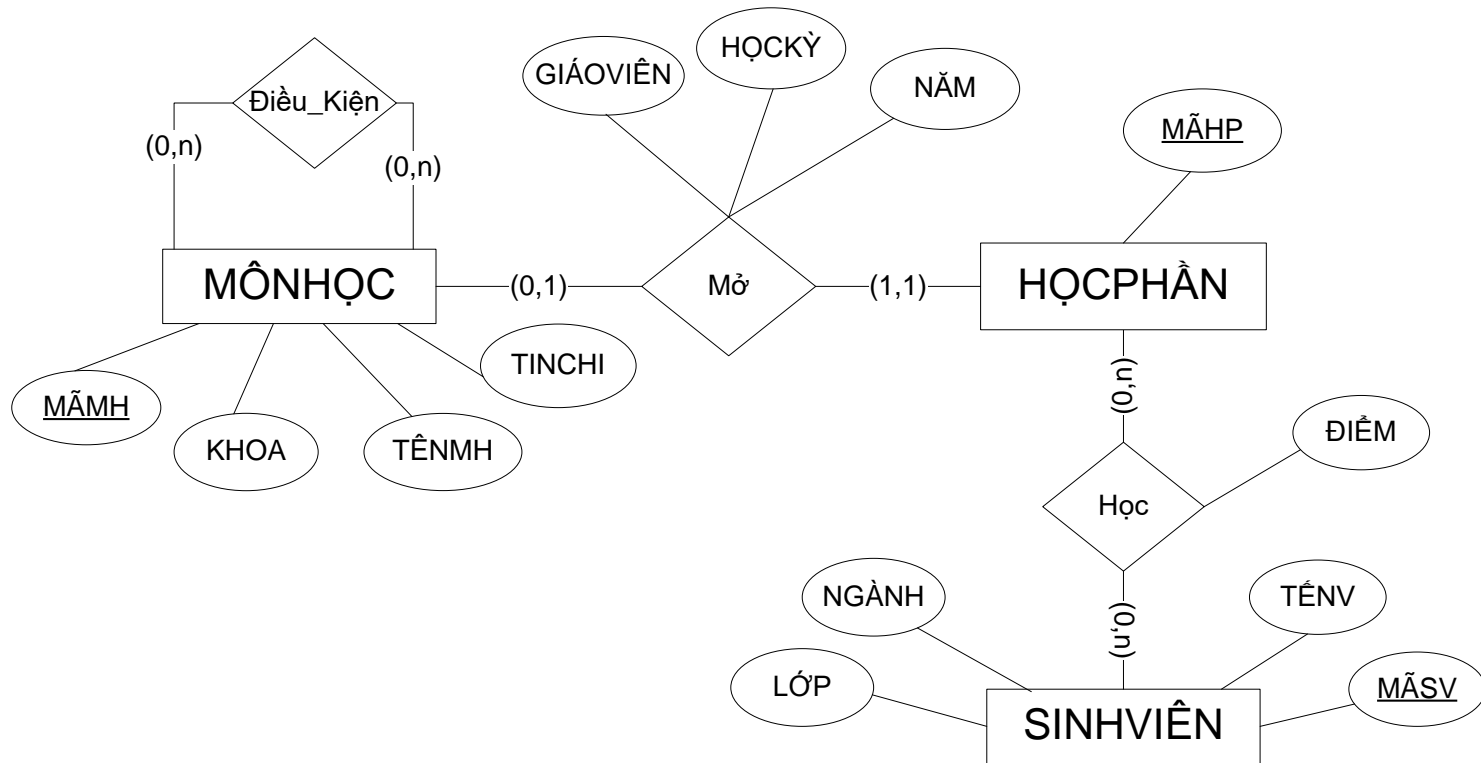
SINH VIÊN	TÊN SV	<u>MÃ SV</u>	LỚP	NGÀNH
-----------	--------	--------------	-----	-------

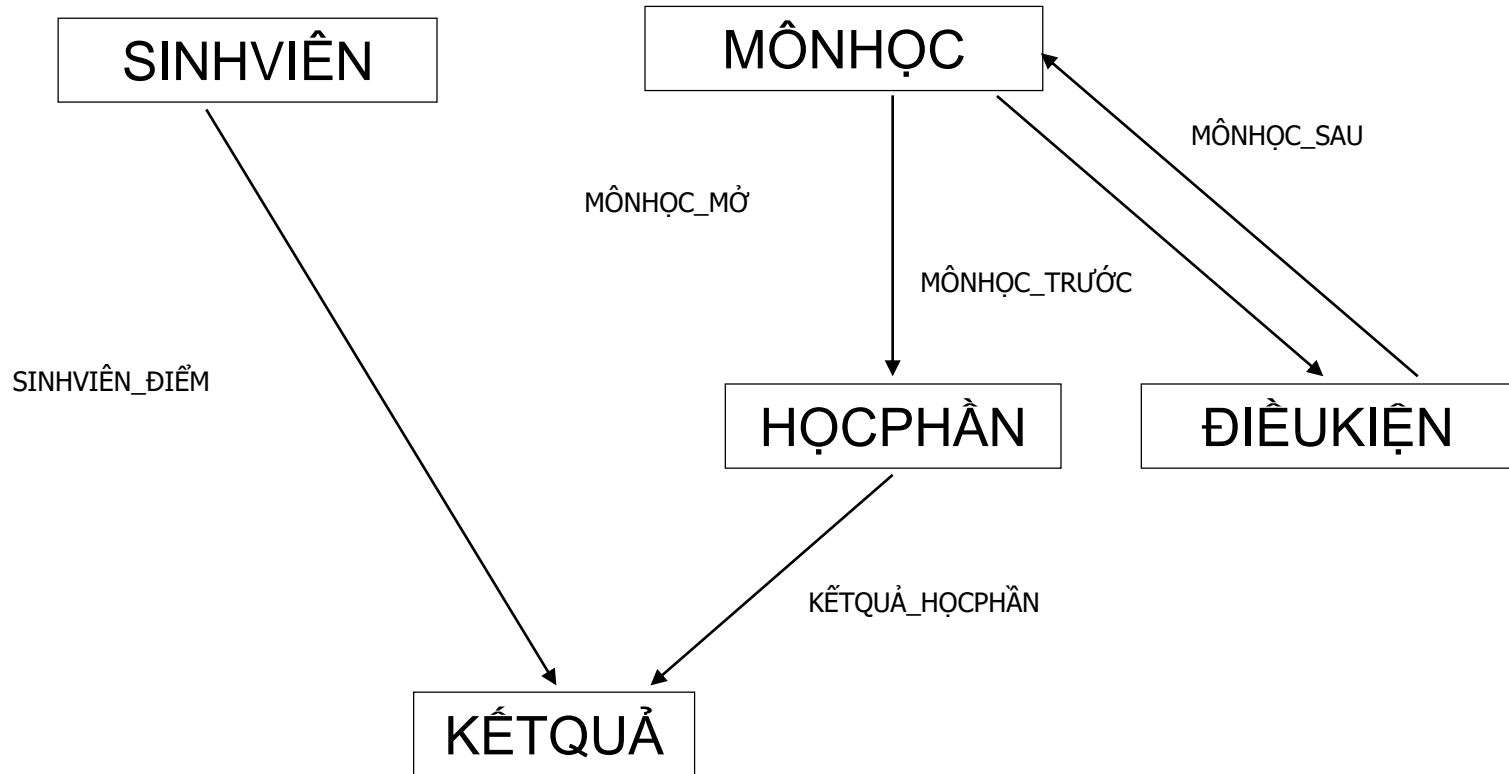
MÔN HỌC	TÊN MH	<u>MÃ MH</u>	KHOA	TỈNH
---------	--------	--------------	------	------

ĐIỀU KIỆN	<u>MÃ MH TRƯỚC</u>	<u>MÃ MH</u>
-----------	--------------------	--------------

HỌC PHẦN	<u>MÃ HP</u>	GIÁO VIÊN	HỌC KỲ	NĂM
----------	--------------	-----------	--------	-----

KQ_HỌC	<u>MÃ SV</u>	<u>MÃ HP</u>	ĐIỂM
--------	--------------	--------------	------





Ví dụ lược đồ phân cấp



Mức 1:

PHÒNGBAN			
TÊNPHG	MÃPHG	TRPHG	NG_NHẬNCHỨC

Mức 2:

NHÂNVIÊN			
TÊN	MÃNV	NGSINH	ĐCHỈ

ĐỀÁN		
TÊNĐA	MÃĐA	ĐĐIỂM_ĐA

Mức 3:

THÂN NHÂN		
TÊN TN	PHÁI	NGSINH

NGƯỜI QUẢN LÝ	
TÊN_NQL	MÃNV

NHÂN CÔNG		
TÊN	MÃNV	THỜI GIAN

- Thể hiện CSDL (Database Instance hoặc Database Status)
 - Là dữ liệu hiện thời được lưu trữ trong CSDL ở một thời điểm nào đó
 - Tình trạng của CSDL

MÔN HỌC	Tên MH	Mã MH	Số TC	Khoa
	Khoa học máy tính	CS1310	4	CNTT
	Cấu trúc dữ liệu	CS3320	4	CNTT
	Toán rời rạc	MATH2410	3	TOÁN
	Cơ sở dữ liệu	CS3380	3	CNTT

HỌC PHẦN	Mã HP	Mã MH	Học Kỳ	Năm	Giáo Viên
	85	MATH2410	1	2008	Anh
	92	CS1310	1	2007	Tiền
	112	MATH2410	2	2008	Anh
	119	CS1310	2	2007	Tiền

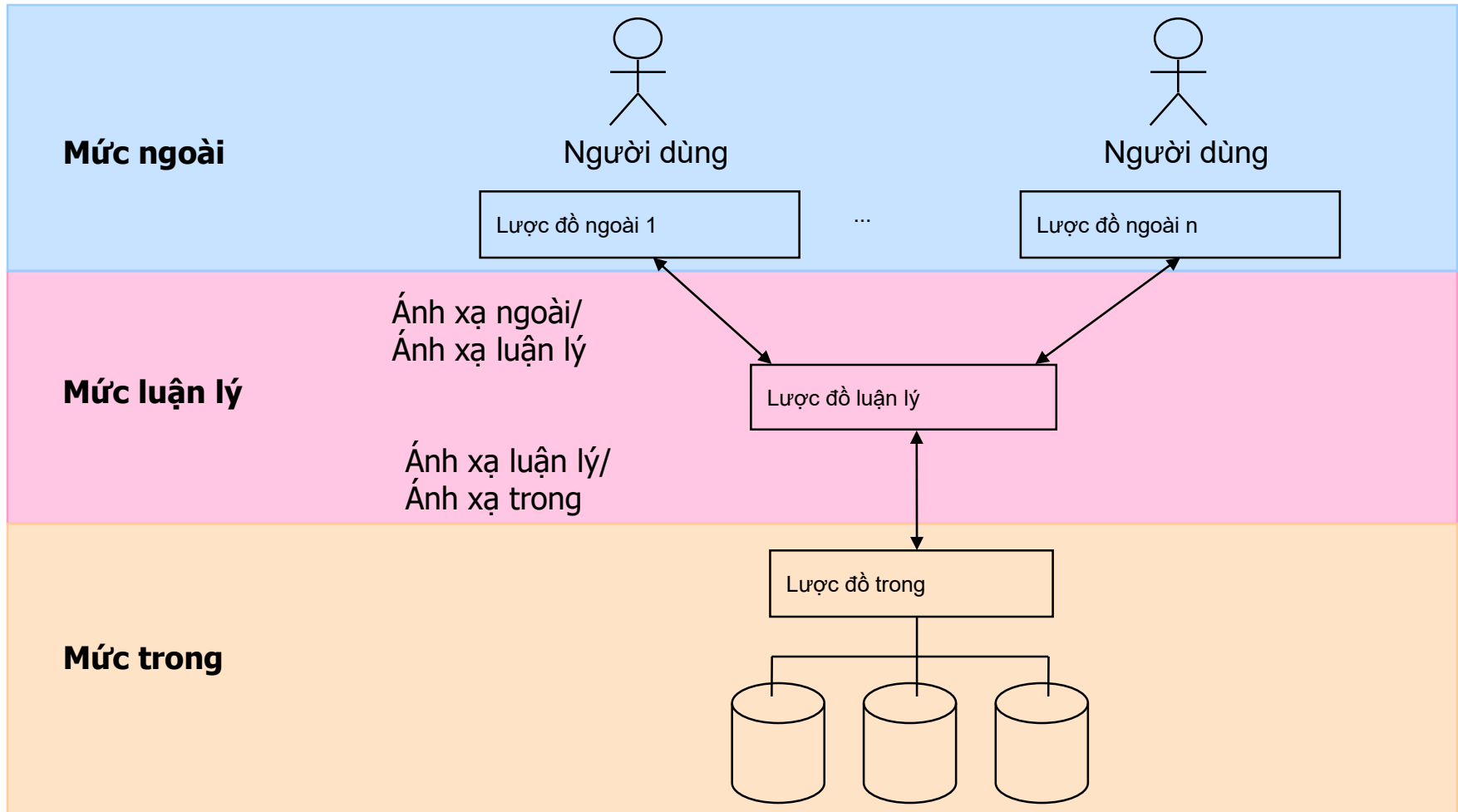
SINH VIÊN	Tên	MSSV	Lớp	Khoa
	Trang	17	1	CNTT
	Ngọc	8	2	CNTT

KẾT QUẢ	MSSV	Mã HP	Điểm
	17	112	10
	17	119	7
	8	85	6
	8	92	9

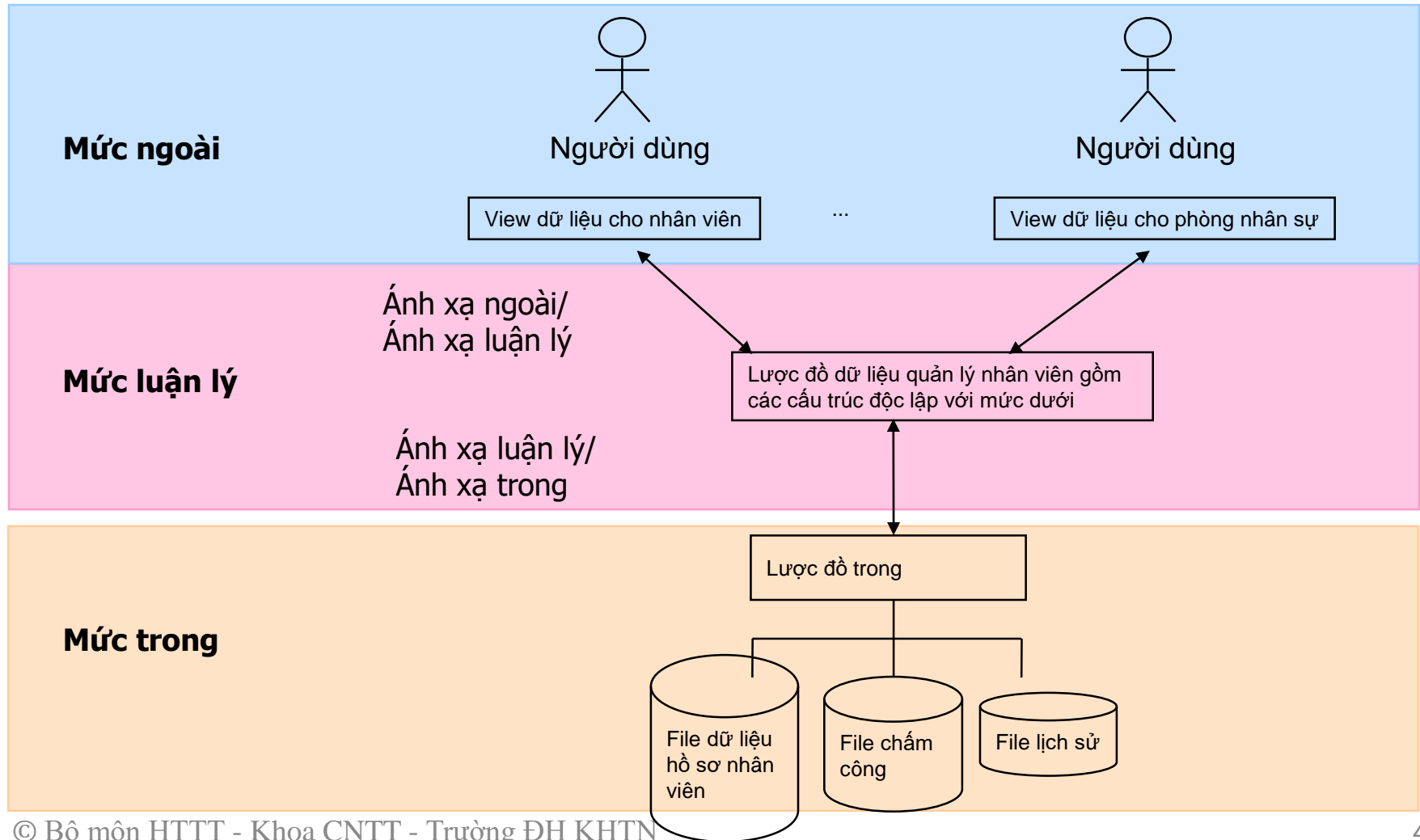
ĐIỀU KIỆN	Mã MH	Mã MH Trước
	CS3380	CS3320
	CS3380	MATH2410
	CS3320	CS1310

- Giới thiệu
- Tiếp cận CSDL so với tập tin
- Một số đặc tính của CSDL
- Người sử dụng CSDL
- Các tính năng của HQT CSDL
- Khái niệm mô tả CSDL
- **Kiến trúc ba lược đồ**
- Ngôn ngữ CSDL

Kiến trúc ba lược đồ của hệ CSDL



■ Ví dụ



- **Mức trong (lược đồ trong)**
 - Mô tả cấu trúc lưu trữ vật lý CSDL
- **Mức luận lý (lược đồ quan niệm)**
 - 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 ngoài (lược đồ ngoài)**
 - Còn gọi là mức khung nhìn (view)
 - 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

■ Độc lập dữ liệu

- Độc lập logic
 - Khả năng thay đổi lược đồ luận lý mà không thay đổi lược đồ ngoài hoặc các CTƯD
- Độc lập vật lý
 - Khả năng thay đổi lược đồ trong mà không làm thay đổi lược đồ luận lý cũng như lược đồ ngoài

- Giới thiệu
- Tiếp cận CSDL so với tập tin
- Một số đặc tính của CSDL
- Người sử dụng CSDL
- Các tính năng của HQT CSDL
- Khái niệm mô tả CSDL
- Kiến trúc ba lược đồ
- **Ngôn ngữ CSDL**

- **Ngôn ngữ định nghĩa dữ liệu (DDL – Data Definition Language)**
 - Ngôn ngữ lưu trữ dữ liệu (SDL): Ngôn ngữ định nghĩa lược đồ trong (cấu trúc lưu trữ và kiểu dữ liệu)
 - Ngôn ngữ định nghĩa khung nhìn (VDL): Ngôn ngữ định nghĩa lược đồ ngoài
- **Ngôn ngữ thao tác dữ liệu (DML – Data Manipulation Language)**
 - Cho phép truy xuất, thêm, xóa, sửa dữ liệu
 - Mức cao (phi thủ tục)
 - Mức thấp (thủ tục)

<p>DDL: Data Definition Language SDL: Storage Definition Language VDL: View Definition Language DML: Data Manipulation Language</p>

■ Yêu cầu: Xác định các thông tin cần lưu trữ của hệ thống sau:

QUẢN LÝ ĐỀ ÁN:

CSDL đề án của một công ty theo dõi các thông tin liên quan đến nhân viên, phòng ban và đề án

- Cty có nhiều phòng ban, mỗi phòng ban có tên duy nhất, mã phòng duy nhất, một trưởng phòng và ngày nhận chức. Mỗi phòng ban có thể ở nhiều **địa điểm** khác nhau.
- Đề án có tên duy nhất, mã duy nhất, do 1 một phòng ban chủ trì và được triển khai ở 1 **địa điểm**.
- Nhân viên có mã số, tên, địa chỉ, ngày sinh, phái và lương. Mỗi nhân viên làm việc ở 1 phòng ban, tham gia vào các đề án với số giờ làm việc khác nhau. Mỗi nhân viên đều có một người quản lý trực tiếp.
- Một nhân viên có thể có nhiều thân nhân. Mỗi thân nhân có tên, phái, ngày sinh và mối quan hệ với nhân viên đó.

Các đối tượng và thông tin cần lưu trữ:

Đối tượng	Thông tin chi tiết
...	

Các mối quan hệ cần lưu trữ:

- ☐ ...
- ☐ ...
- ☐ ...

Các đối tượng và mối quan hệ cần lưu trữ:

Đối tượng	Thông tin chi tiết
Phòng ban	Tên phòng
Nhân viên	Mã số, họ tên, địa chỉ, ngày sinh, phái, lương
Địa điểm	Tên địa điểm
...	

Các mối quan hệ cần lưu trữ:

- ☐ Mỗi **nhân viên** làm việc cho 1 **phòng ban**
- ☐ Một **phòng ban** sẽ có 1 **nhân viên** làm trưởng phòng, khi nhân viên làm trưởng phòng thì sẽ có ngày nhận chức tương ứng
- ☐ Mỗi **phòng ban** có thể ở nhiều **địa điểm**
- ☐ Một **nhân viên** có 1 **nhân viên** là người quản lý trực tiếp
- ☐ ...

