

Chương 1

TỔNG QUAN VỀ CÁC HỆ CSDL

Bộ môn Hệ thống thông tin
Khoa CNTT – Đại học Khoa học tự nhiên, TpHCM

Ref: TS. Hồ Bảo Quốc



Nội dung

- Giới thiệu
- Các hệ thống tập tin
- Các hệ thống CSDL và định nghĩa
- Một số đặc trưng theo hướng CSDL
- Người sử dụng CSDL
- Kiến trúc của HQT CSDL
- Các tính năng của HQT CSDL
- Mô hình dữ liệu
- Ngôn ngữ CSDL



Cơ sở dữ liệu

Programs = Algorithms + Data Structure

*Phát triển
Các thuật toán*

*Phát triển
các cấu trúc lưu trữ*

**Vậy trước đó,
tổ chức dữ liệu bằng gì ?**



Phù hợp cho các ứng dụng

- Có khối lượng dữ liệu lớn
- Không có tính toán phức tạp

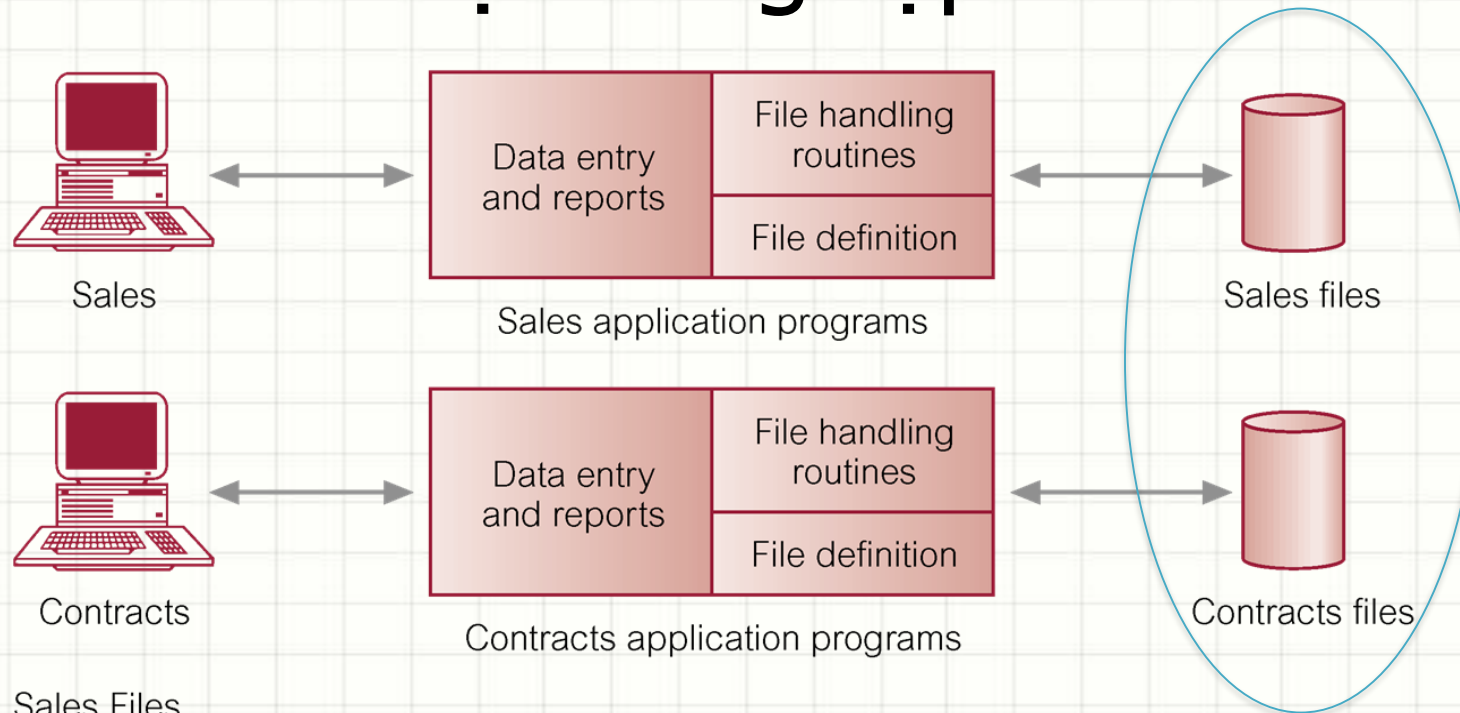
Ví dụ: ứng dụng

- Quản lý nhân sự
- Quản lý hàng không



CÁC HỆ THỐNG TẬP TIN

Xử lý dữ trên hệ thống tập tin



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)

Hạn chế của Hệ thống tập tin

- **Dữ liệu tách biệt và không chia sẻ**
 - Mỗi ứng dụng sử dụng một hệ thống tập tin riêng
 - Dữ liệu không thể chia sẻ giữa các ứng dụng
- **Sự trùng lặp**
 - Dữ liệu bị trùng lặp giữa các ứng dụng
 - Tốn không gian lưu trữ
 - Có thể không nhất quán

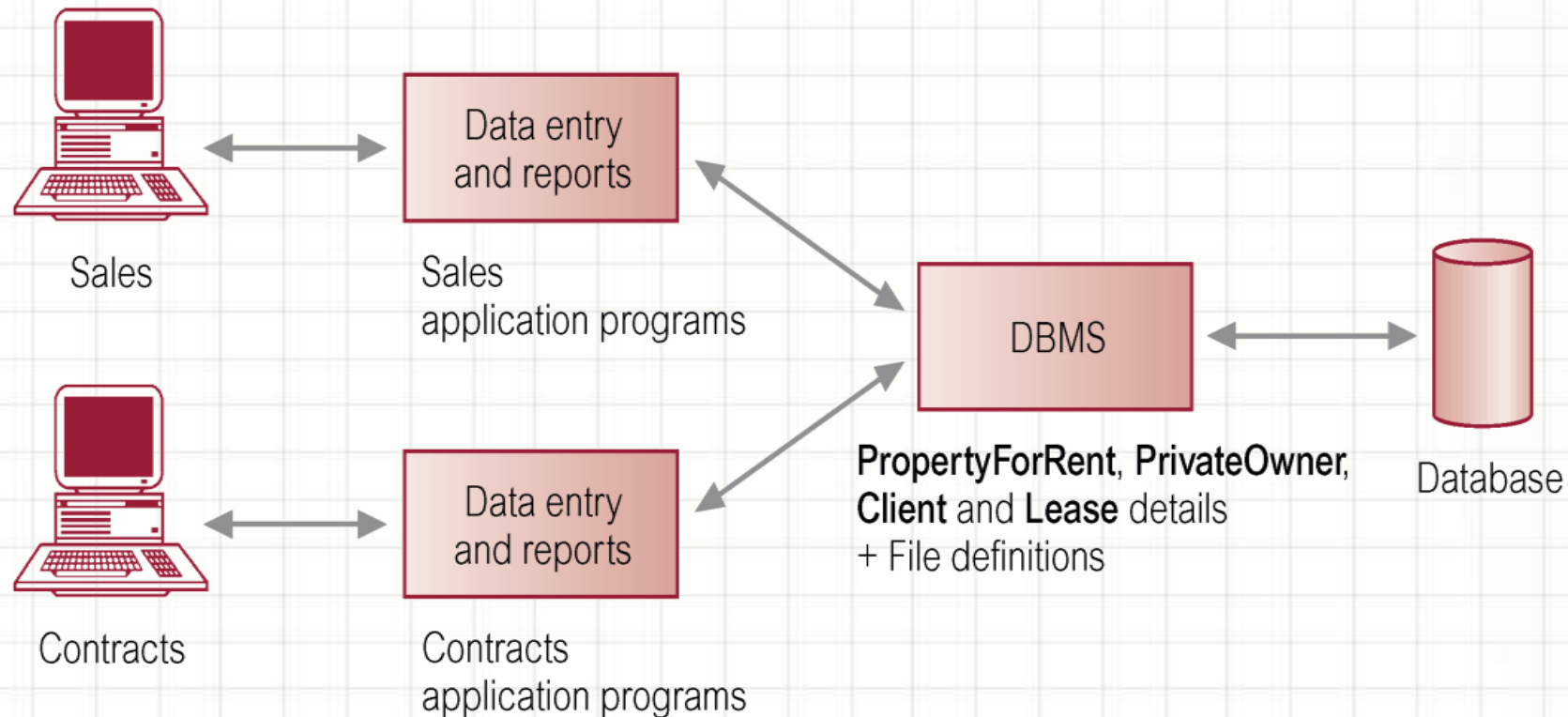
Hạn chế của Hệ thống tập tin

- Phụ thuộc dữ liệu
 - Các cấu trúc tập tin được định nghĩa cố định trong chương trình
- Định dạng tập tin không tương thích
 - Các tập tin được định dạng theo một ngôn ngữ lập trình nào đó → khó chia sẻ
- Câu truy vấn bị cố định trước
 - Mỗi chương trình truy xuất được phát triển cố định cho một câu hỏi



CÁC HỆ THỐNG CSDL

Cách tiếp cận CSDL



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)

Đặc điểm của Dữ liệu lưu trên CSDL

- Dữ liệu **lưu trữ độc lập** với chương trình
- Dữ liệu được **quản lý tập trung**
- Dữ liệu được **chia sẻ** cho nhiều ứng dụng
- Dữ liệu được bảo đảm **an toàn**
- Dữ liệu **ít dư thừa**



MỘT SỐ ĐỊNH NGHĨA

Định nghĩa Cơ sở dữ liệu

- **Cơ sở dữ liệu (Database)**
- Là một *tập hợp dữ liệu* được tổ chức theo một *cấu trúc chặt chẽ* nhằm *phục vụ (chia sẻ) cho nhiều mục tiêu* khác nhau một cách *có chọn lọc*
- Ví dụ :
 - Cơ sở dữ liệu nhân viên,
 - cơ sở dữ liệu hàng hóa ...

Định nghĩa Cơ sở dữ liệu

Tập ngẫu nhiên của các dữ liệu
không thể xem là một CSDL

NHANVIEN	HONV	TENLOT	TENNV	MANV
	Tran	Hong	Quang	NHQuang
	Tran	Thanh	Tam	TTTam

DEAN	TENDA	MADA	DDIEM_DA
	IBM Leader	1	VUNG TAU
	Photoshop	2	NHA TRANG
	Management	3	TP HCM

PHANCONG	MA_NVIENT	SODA	THOIGIAN
	NHQuang	1	32.5
	NHQuang	2	7.5
	TTTam	3	40.0

Ví dụ về một CSDL

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

CSDL
Quản lý đề án công ty

Ví dụ về một CSDL

NHANVIEN	HONV	TENLOT	TENNV	MANV
	Tran	Hong	Quang	NHQuang
	Tran	Thanh	Tam	TTTam

DEAN	TENDA	MADA	DDIEM_DA
	IBM Leader	1	VUNG TAU
	Photoshop	2	NHA TRANG
	Management	3	TP HCM

PHANCONG	MA_NVEN	SODA	THOIGIAN
	NHQuang	1	32.5
	NHQuang	2	7.5
	TTTam	3	40.0

- **Quản lý đề án của một công ty**

- Định nghĩa CSDL

- Cấu trúc bảng, bao gồm các thành phần dữ liệu và kiểu dữ liệu tương ứng

- Xây dựng CSDL

- Đưa dữ liệu vào các bảng

- Xử lý CSDL

- Thực hiện truy vấn: “Cho biết các nhân viên thuộc phòng 5”
- Thực hiện phép cập nhật: “Chuyển nhân viên Nguyễn Thanh Tùng sang phòng số 1”

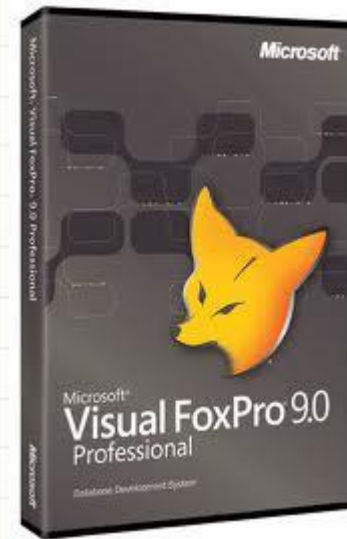
Định nghĩa **Hệ quản trị cơ sở dữ liệu**

- **HQT CSDL (Database Management System)**

- Hệ thống phần mềm, cung cấp công cụ:
 - Xây dựng CSDL
 - Thao tác trên dữ liệu của CSDL
 - Kiểm soát việc truy xuất dữ liệu
- Cho phép **định nghĩa**, **xây dựng** và **xử lý** dữ liệu
 - Định nghĩa – khai báo bộ khung, các mô tả chi tiết về dữ liệu
 - Xây dựng – lưu trữ dữ liệu lên bộ nhớ phụ
 - Xử lý – truy vấn, cập nhật và phát sinh báo cáo

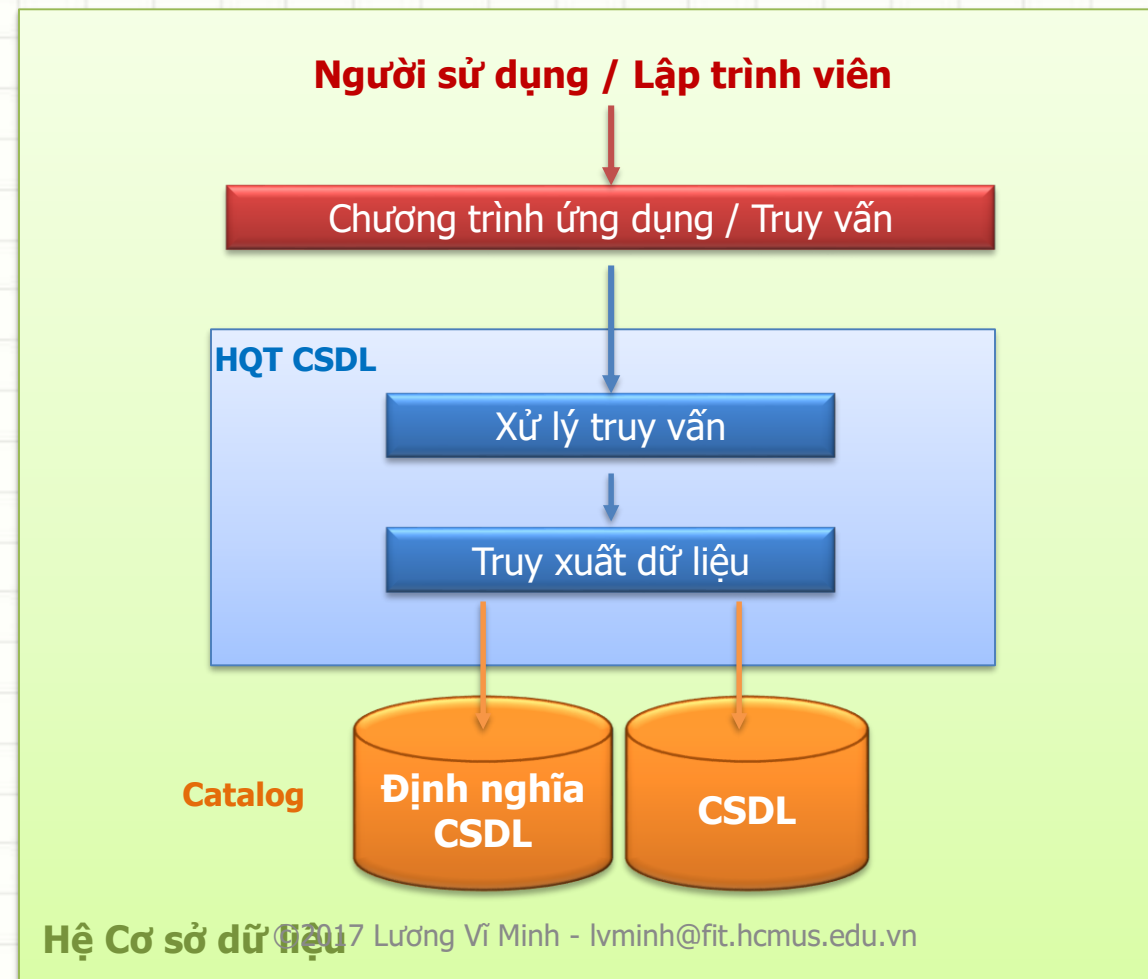
Định nghĩa Hệ quản trị cơ sở dữ liệu

- Một số Hệ quản trị CSDL phổ biến



Định nghĩa – Hệ Cơ sở dữ liệu

Hệ Cơ sở dữ liệu (Database System)





ĐẶC TRƯNG THEO HƯỚNG CSDL

Đặc trưng theo hướng CSDL

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

1. Tính tự mô tả

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

2. Tính độc lập

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



3. Tính trừu tượng

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

4. Tính nhất quán

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

5. Đa khung nhìn

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



NGƯỜI SỬ DỤNG CSDL

Người sử dụng CSDL



Quản trị viên

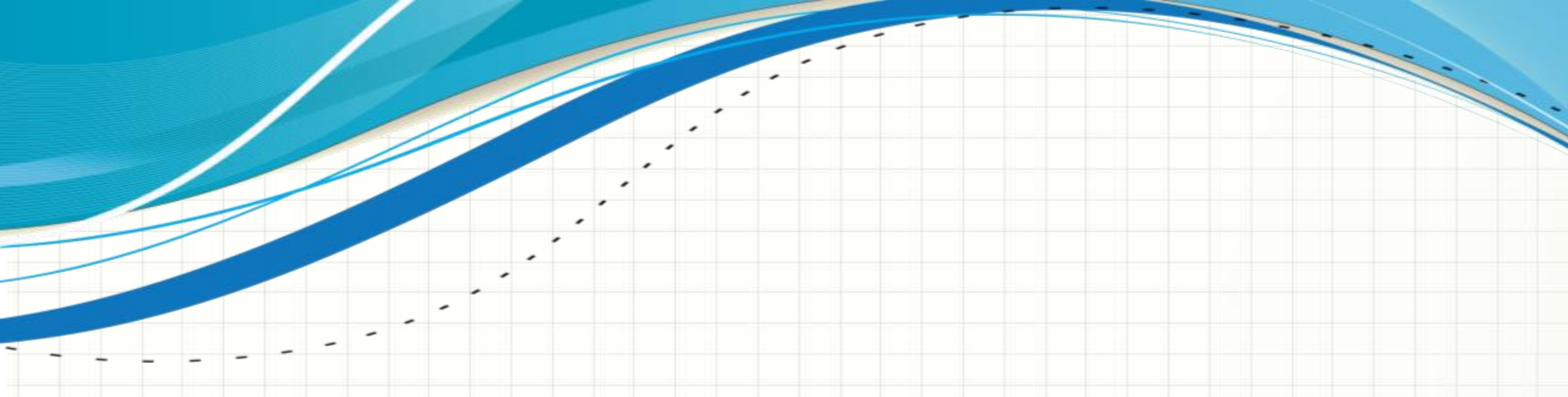
- **DBA – Database Administrator**
- **Có trách nhiệm** quản lý hệ CSDL
 - Cấp quyền truy cập
 - Điều phối việc sử dụng
 - Giám sát việc sử dụng
 - Tìm kiếm & đảm bảo phần cứng, phần mềm liên quan
- **Chịu trách nhiệm** cho các vấn đề liên quan đến
 - Tính an toàn của hệ CSDL
 - Khả năng đáp ứng của hệ CSDL

Thiết kế viên

- DBD – Database Designer
- Chịu trách nhiệm về vấn đề :
 - Quyết định những dữ liệu nào cần được lưu trữ
 - Lựa chọn cấu trúc phù hợp để lưu trữ dữ liệu
- Trao đổi với người dùng để nắm bắt được những yêu cầu và đưa ra một thiết kế CSDL thỏa yêu cầu này
- Có thể là thành viên của
 - Nhóm DBA quản lý các CSDL
 - Các nhóm khác sau khi việc thiết kế hoàn tất (nhóm lập trình,...)

Người dùng cuối

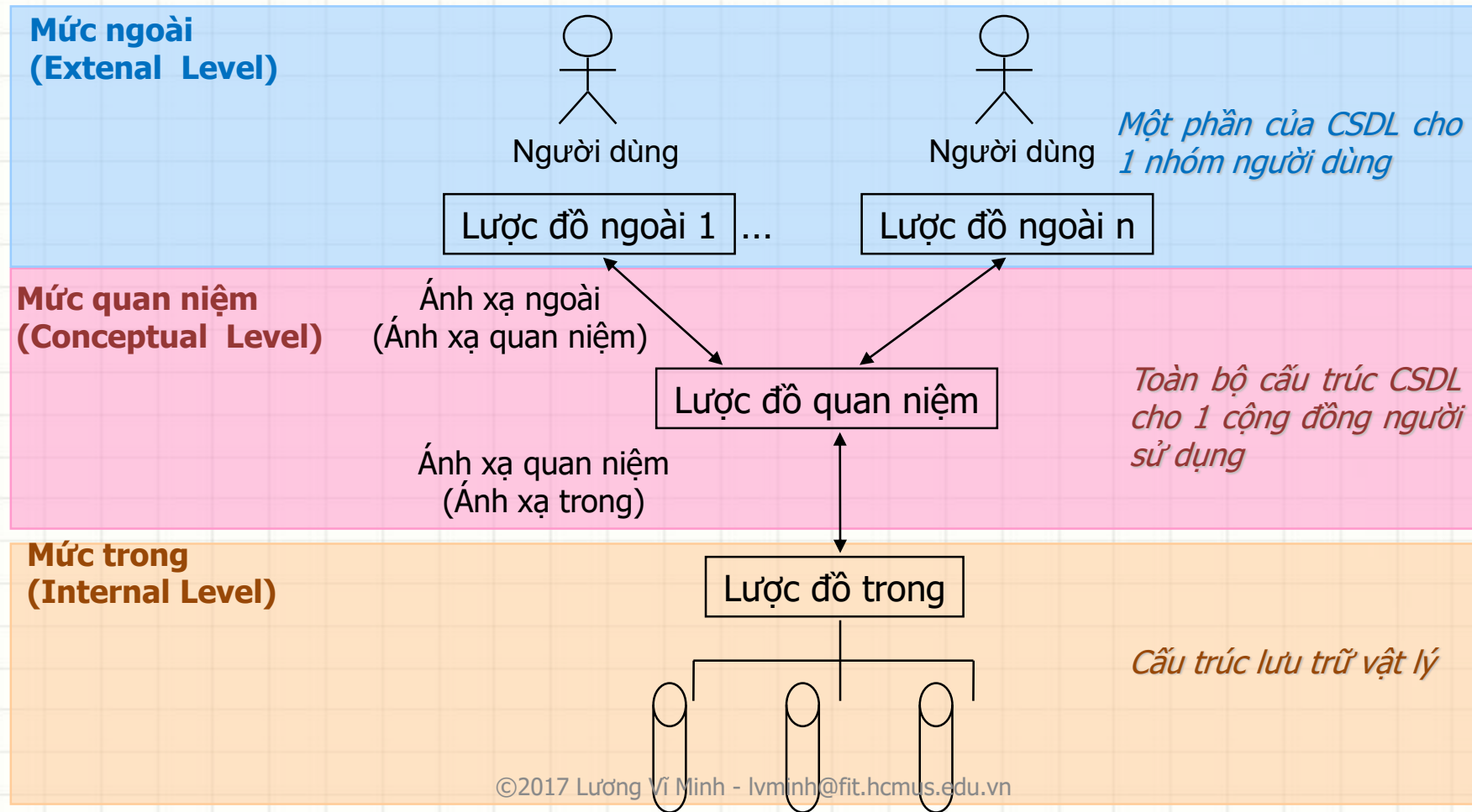
- End Users
- Người ít sử dụng (Casual users)
 - Ít khi truy cập CSDL, nhưng cần những thông tin khác nhau trong mỗi lần truy cập và dùng những câu truy vấn phức tạp
 - Người quản lý
- Người sử dụng thường xuyên (Naïve users)
 - Thường xuyên truy vấn và cập nhật CSDL nhờ vào một số các chức năng đã được xây dựng sẵn và kiểm thử
 - Nhân viên
- Người sử dụng đặc biệt (Sophisticated users)
 - Thông thạo về HQT CSDL, tự xây dựng những truy vấn phức tạp cho công việc
 - Kỹ sư, nhà khoa học, người phân tích,...
- Người sử dụng cá nhân (stand-alone users)
 - Sử dụng các ứng dụng truy cập HQT CSDL để phục vụ công việc cá nhân
 - Người sử dụng chương trình quản lý thuế cá nhân, ...



KIẾN TRÚC CỦA HQTCSĐL

Kiến trúc 3 lược đồ của HQT CSDL

- Kiến trúc 3 lược đồ



Kiến trúc 3 lược đồ

- **Mức khung nhìn / ngoài (External Level):** là các góc nhìn khác nhau của các nhóm người sử dụng về CSDL. Mỗi nhóm người dùng có một góc nhìn (View) khác nhau về CSDL
- **Mức quan niệm (Conceptual Level):** là mức nhìn tổng thể về CSDL, đây là góc nhìn của người có trách nhiệm quản trị CSDL
- **Mức vật lý / trong (Physic/ Internal level):** là mức tổ chức vật lý của dữ liệu trong CSDL, đây là góc nhìn của người phát triển CSDL

Ví dụ về 3 mức nhìn về CSDL

External view 1

Sno	FName	LName	Age	Salary
-----	-------	-------	-----	--------

External view 2

Staff_No	LName	Bno
----------	-------	-----

Conceptual level

Staff_No	FName	LName	DOB	Salary	Branch_No
----------	-------	-------	-----	--------	-----------

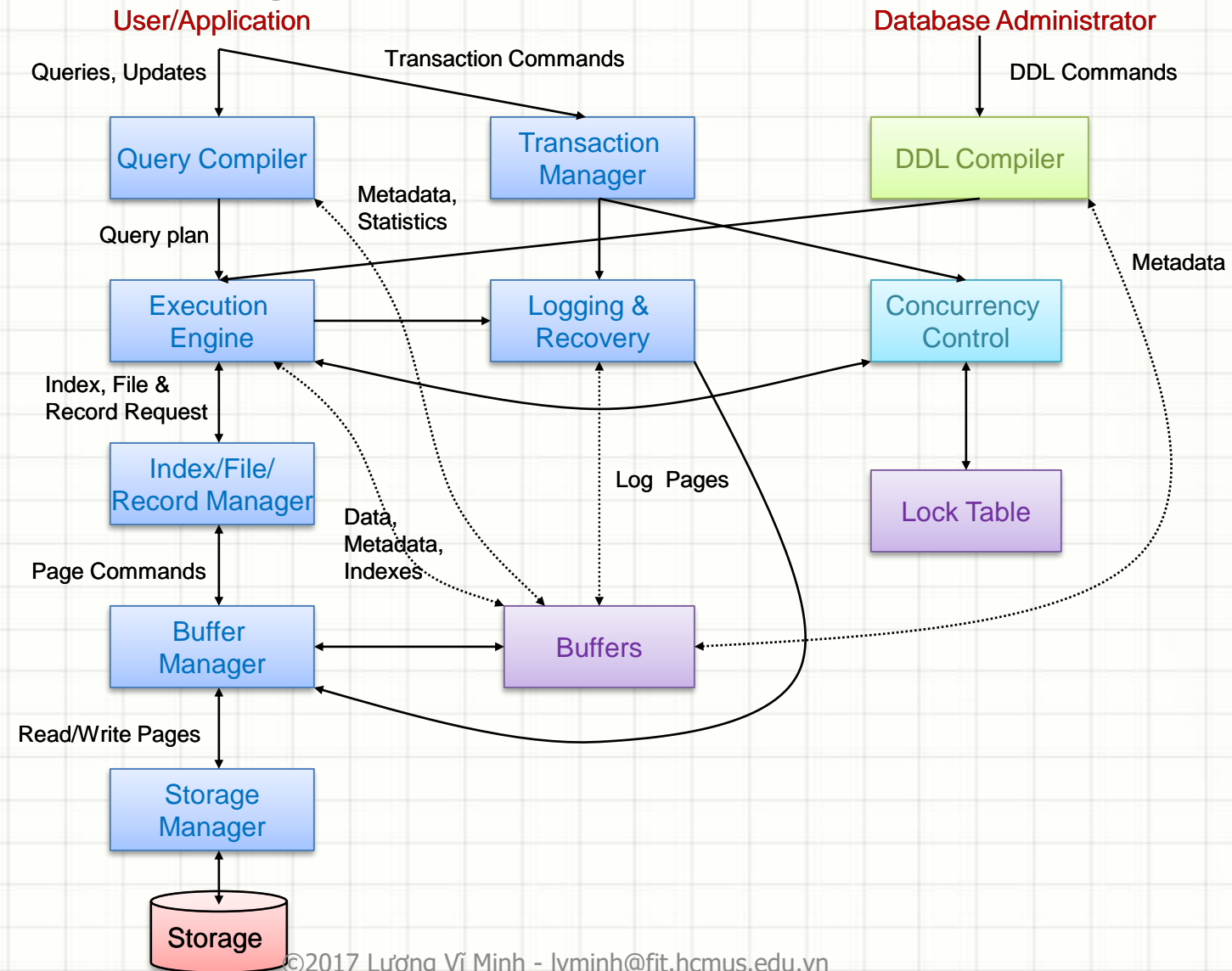
Internal level

```
struct STAFF {  
    int Staff_No;  
    int Branch_No;  
    char FName [15];  
    char LName [15];  
    struct date Date_of_Birth;  
    float Salary;  
    struct STAFF *next;           /* pointer to next Staff record */  
};  
index Staff_No; index Branch_No; /* define indexes for staff */
```

Kiến trúc của HQT CSDL

- Độc lập dữ liệu
 - Độc lập logic
 - Khả năng thay đổi lược đồ quan niệm (VD: Thêm/xóa thực thể) mà không thay đổi lược đồ ngoài hoặc phải viết lại các chương trình ứng dụng
 - Độc lập vật lý
 - Khả năng thay đổi lược đồ trong (VD: thay đổi cách thức lưu trữ dữ liệu) mà không làm thay đổi lược đồ quan niệm

Kiến trúc của HQT CSDL





TÍNH NĂNG CỦA HQT CSDL

Tính năng của HQT CSDL

- Hạn chế sự trùng lặp 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 nhiều người dùng, các HQT phải cho phép truy xuất dữ liệu đồng thời
- Hạn chế những truy cập không được phép
 - Từng người dùng và nhóm người dùng có một tài khoản và mật mã để truy xuất dữ liệu

Tính năng của HQT CSDL

- **Cung cấp nhiều giao diện giao tiếp với dữ liệu**
 - Các ngôn ngữ truy vấn
 - Giao diện ngôn ngữ lập trình
 - Form và giao diện dòng lệnh
 - Giao diện hướng menu và Giao diện ngôn ngữ tự nhiên
- **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 RBTV Có thể được khai báo với HQT và HQT sẽ tự động kiểm tra.
 - Một số RBTV khác được kiểm tra nhờ ứng dụng
- **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

Tính năng của HQT CSDL

- Các tính năng khác
 - Chuẩn hóa
 - Cho phép DBA định nghĩa và bắt buộc áp dụng một chuẩn thống nhất cho mọi người dùng
 - Uyển chuyển
 - Khi nhu cầu công việc thay đổi, cấu trúc CSDL rất có thể thay đổi, HQT cho phép thêm hoặc mở rộng cấu trúc mà không làm ảnh hưởng đến CTƯD
 - Giảm thời gian phát triển ứng dụng
 - Khả năng cung cấp thông tin mới nhất
 - Cân đối được sự tiết kiệm đầu tư về chia sẻ dữ liệu



Mô HÌNH DỮ LIỆU (DATA MODEL)

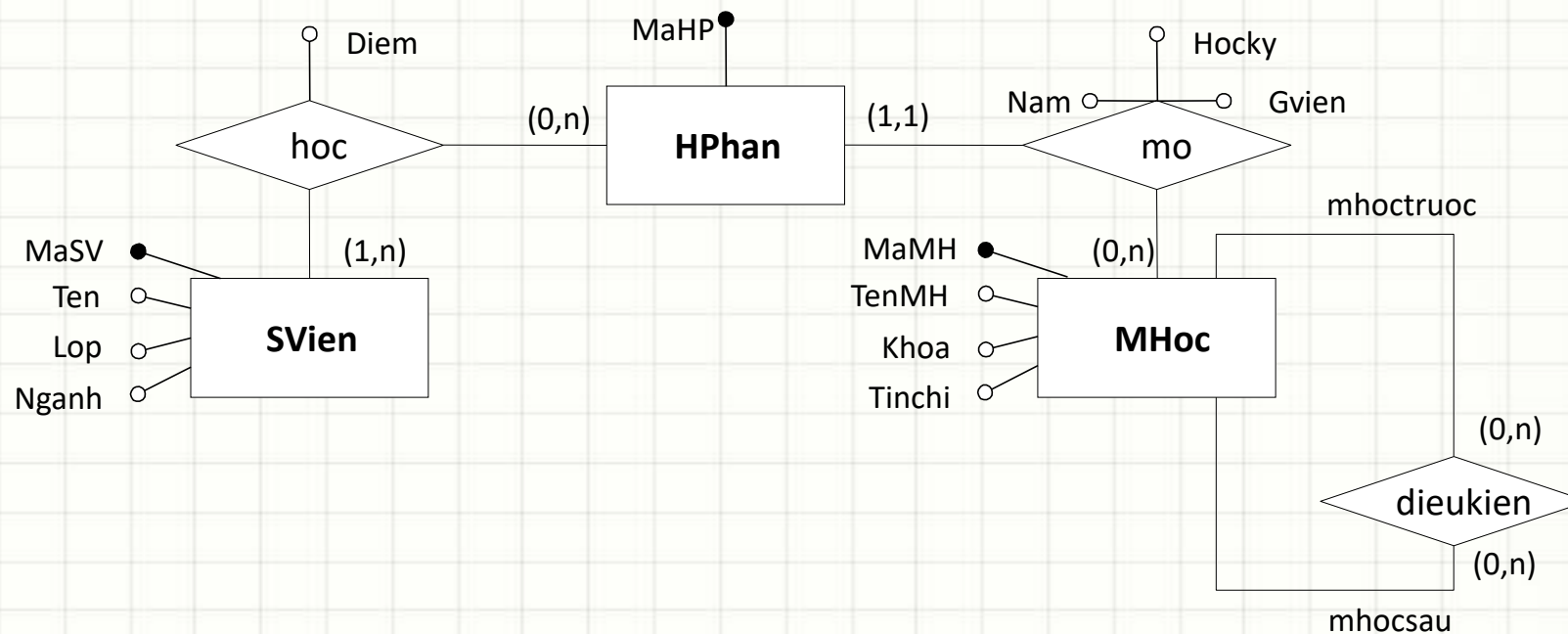
Mô hình dữ liệu

- Bao gồm
 - Các **khái niệm** biểu diễn dữ liệu
 - Các **phép toán xử lý** dữ liệu
- Phân loại
 - Mô hình mức cao
 - Mô hình mức cài đặt
 - Mô hình mức thấp

Mô hình dữ liệu

- Mô hình mức cao
 - Cung cấp các khái niệm gần gũi với người dùng
 - Mô hình phải tự nhiên và giàu ngữ nghĩa
 - VD: mô hình thực thể kết hợp (ER), mô hình đối tượng...
- Mô hình mức cài đặt
 - Đưa ra các khái niệm 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)
 - Đưa ra các khái niệm mô tả chi tiết về cách thức dữ liệu được lưu trữ trong máy tính

Ví dụ - Mô hình ER



Ví dụ - Mô hình phân cấp

Cấp 1:

KQua	
DiemTH	DiemLT

Cấp 2:

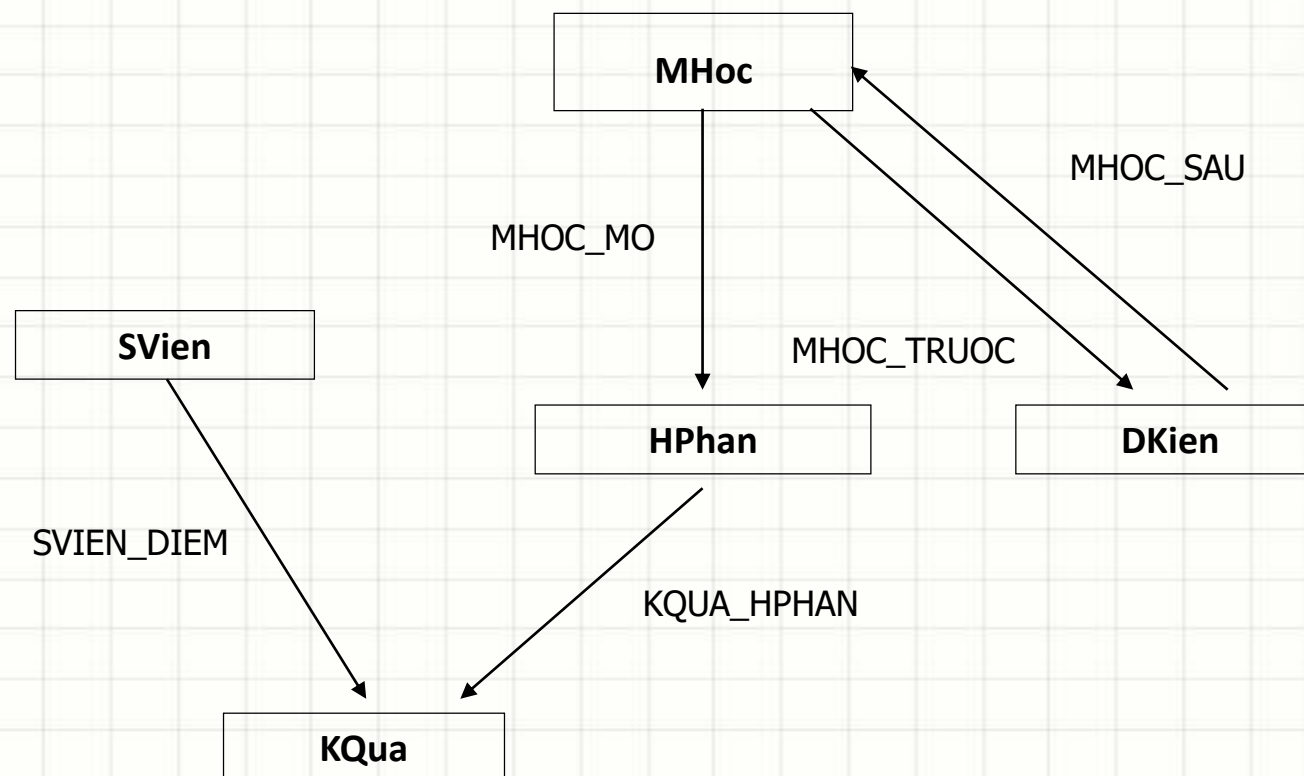
HPhan	
TenHP	SLuong

SVien		
TenSV	Lop	Nganh

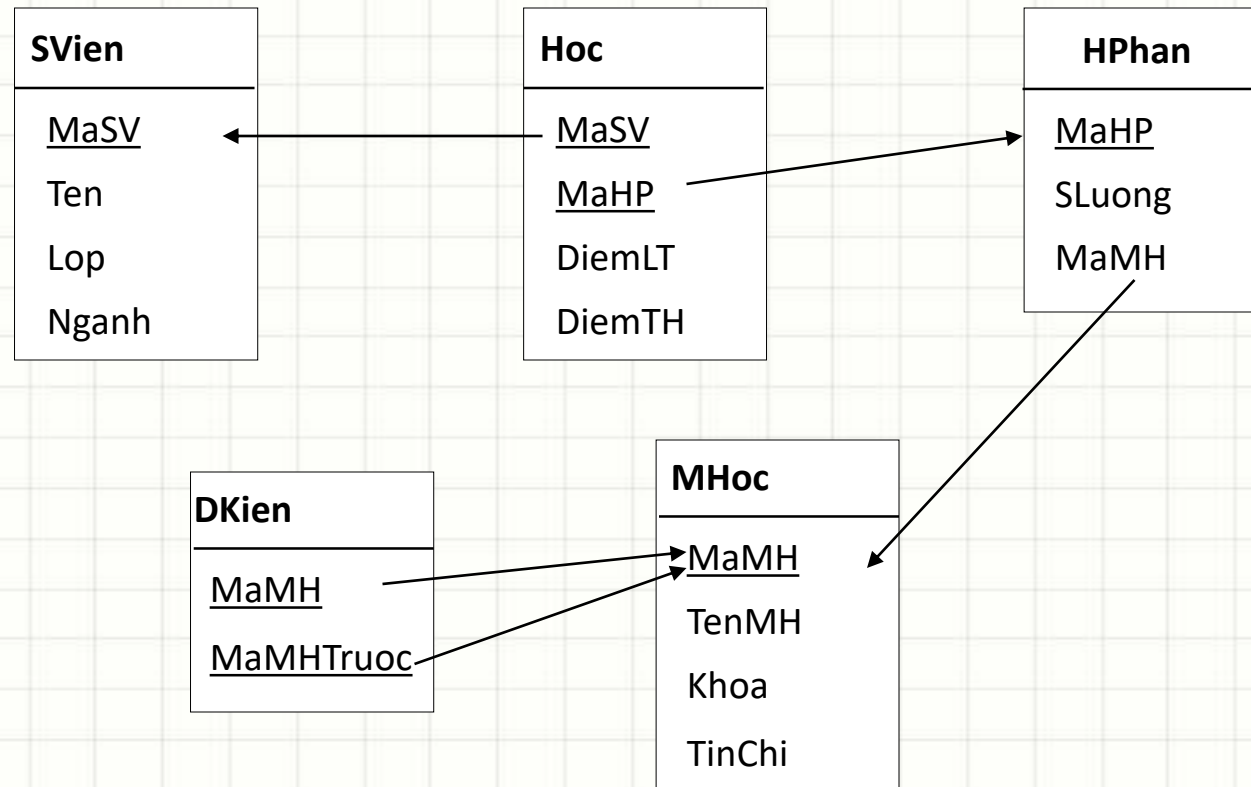
Cấp 3:

MHoc		
TenMH	Khoa	TinChi

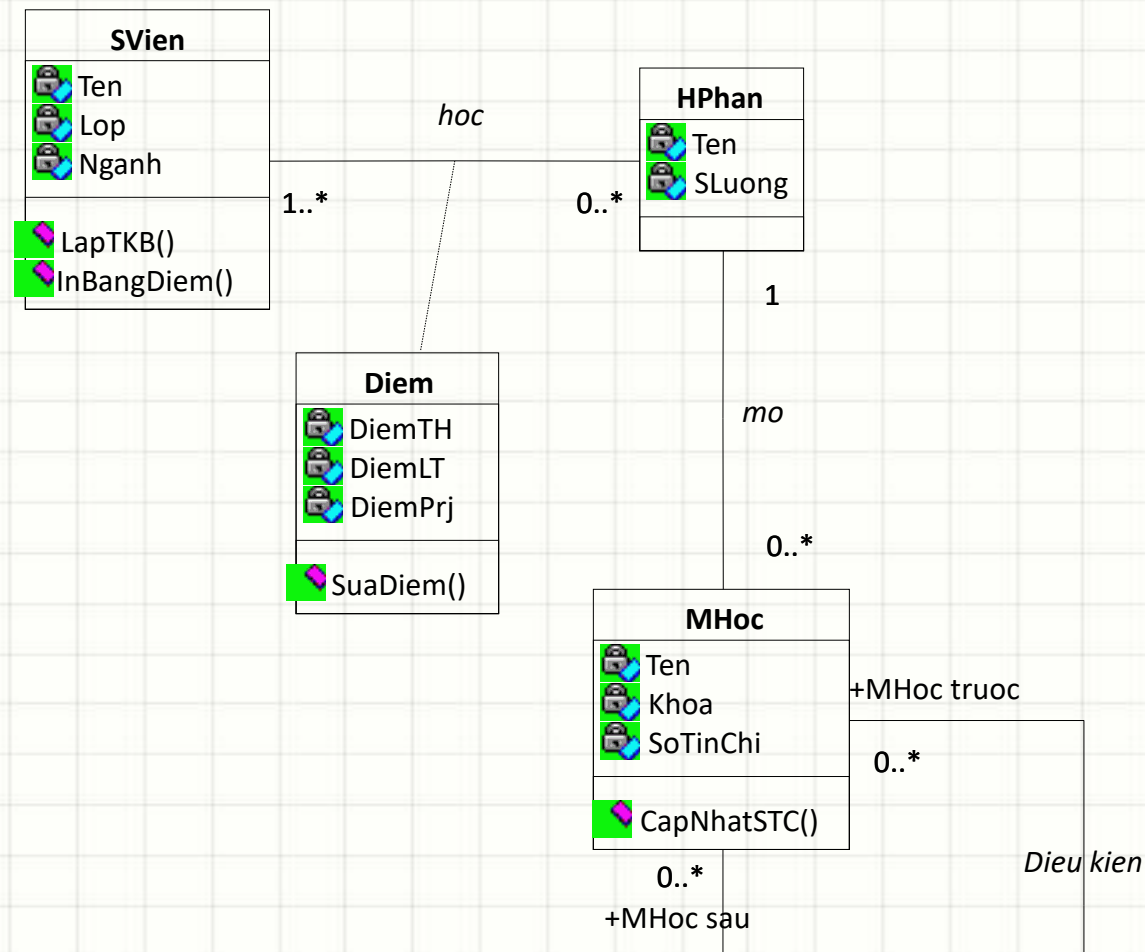
Ví dụ - Mô hình mạng



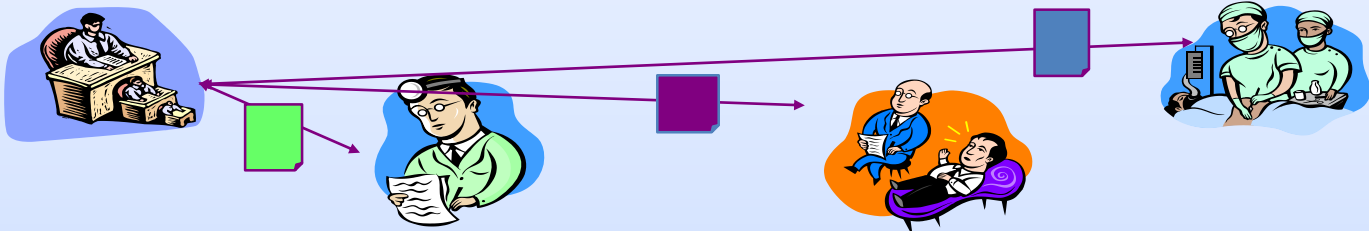
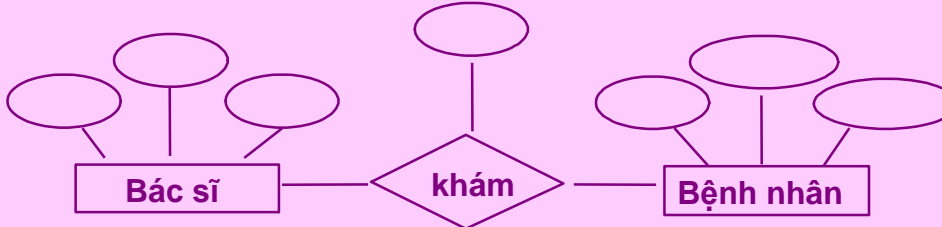
Ví dụ - Mô hình quan hệ



Ví dụ - Mô hình đối tượng



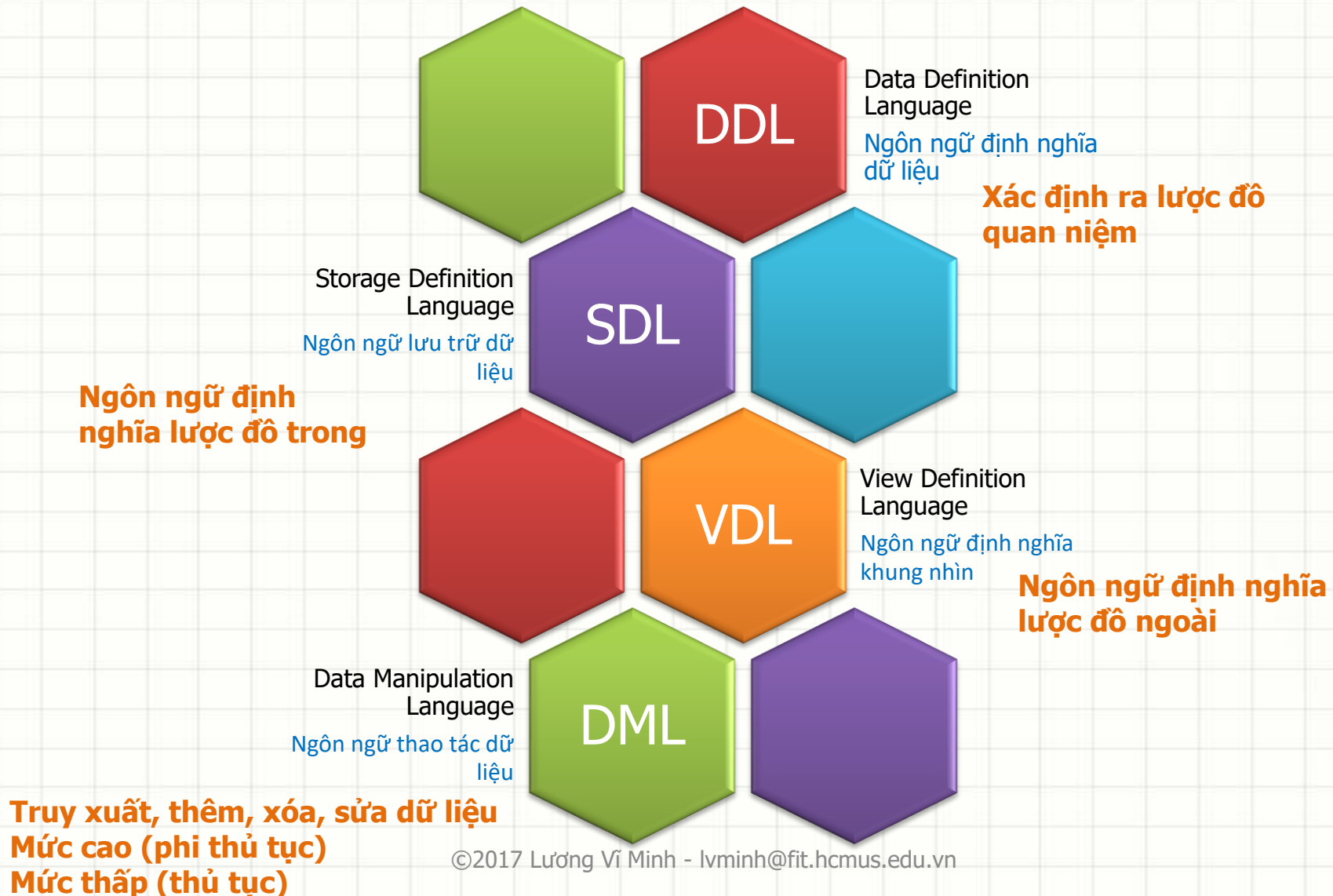
Mô hình hóa nhiều cấp

Thực tế					
Mô hình quan niệm	<ul style="list-style-type: none">♦ Độc lập với mô hình dữ liệu♦ Độc lập với DBMS				
Mô hình logic	<ul style="list-style-type: none">♦ Phụ thuộc mô hình dữ liệu♦ Phụ thuộc DBMS	Codasyl	Relational	Object	XML
Mô hình vật ký	<ul style="list-style-type: none">♦ Phụ thuộc mô hình dữ liệu♦ Phụ thuộc DBMS	<ul style="list-style-type: none">♦ Tổ chức vật lý của dữ liệu♦ Cấu trúc lưu trữ dữ liệu♦ Các cấu trúc chỉ mục (index)			



NGÔN NGỮ CSDL

Ngôn ngữ Cơ sở dữ liệu



Câu hỏi

- Khi nào nên dùng và không nên dùng cách tiếp cận CSDL?
- Các vấn đề ảnh hưởng
 - Đầu tư cao cho phần cứng, phần mềm, đào tạo
 - Chức năng chính chỉ gồm định nghĩa và xử lý dữ liệu
 - Tốn chi phí cho vấn đề bảo mật, điều khiển trùng lặp thông tin, chức năng phục hồi và toàn vẹn dữ liệu
- Không phù hợp với ứng dụng :
 - Nhỏ, có tổ chức dữ liệu ít và tốt, không có nhu cầu thay đổi
 - Có yêu cầu về xử lý thời gian thực nghiêm ngặt
 - Không có nhiều người truy cập dữ liệu đồng thời

Câu hỏi

- Dữ liệu ?
- Thông tin ?
- Tri thức ?



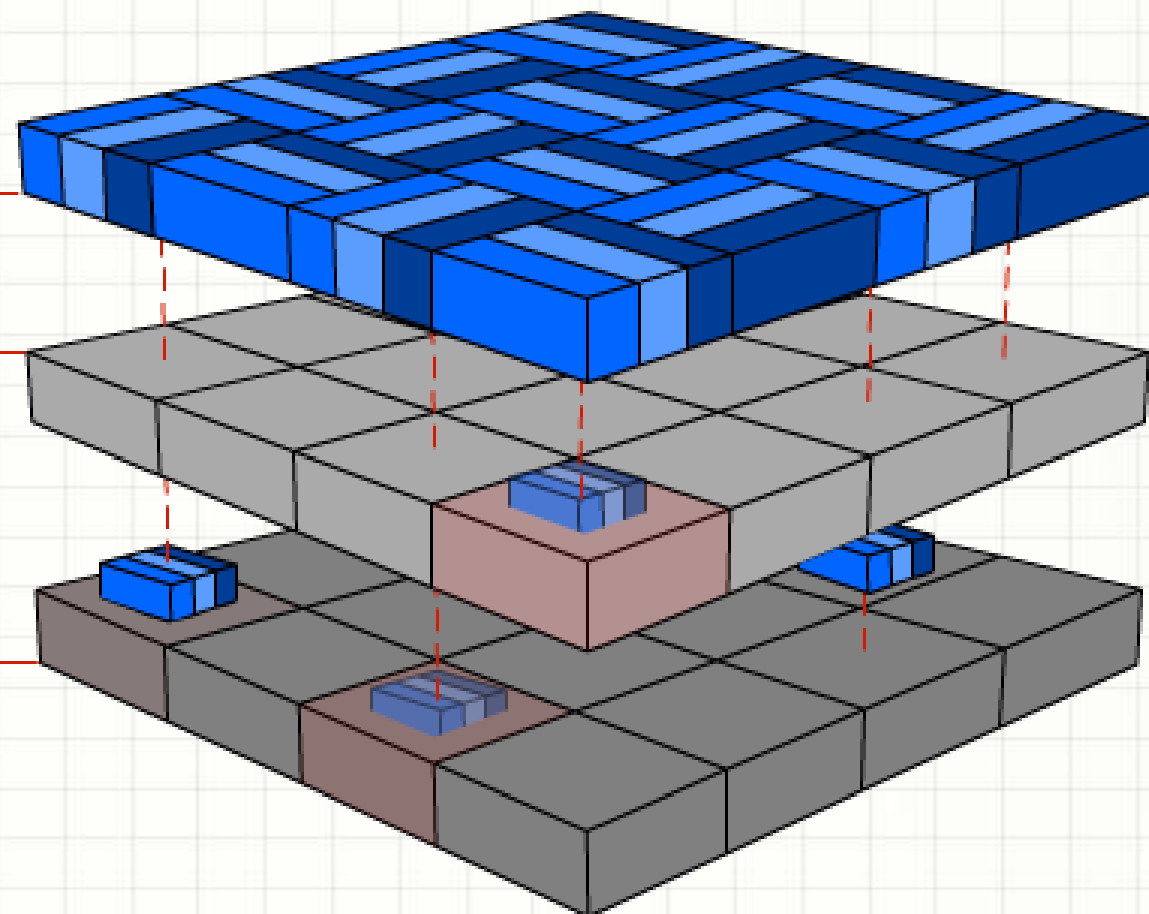
ENTERPRISE KNOWLEDGE INFRASTRUCTURE –

Metadata Repository

Content Repositories

Intranets, Portals, Extranets, Document Management, Supply Chain Management, Customer Relationship Management, E-mail Servers, Personal Computers, Internet, File Shares, And More...

Employees





CÂU HỎI ?