



Trường Đại học Nha Trang
Khoa Công nghệ Thông tin

CƠ SỞ DỮ LIỆU



Giới thiệu môn học

TS. Phạm Thị Thu Thúy
thuthuy@ntu.edu.vn



Đề cương chi tiết học phần



Thông tin về học phần và lớp học

- ❖ Tên học phần:
 - Tiếng Việt: Cơ sở dữ liệu
 - Tiếng Anh: Database
- ❖ Mã học phần: INS330 Số tín chỉ: 3TC (2LT + 1BT)
- ❖ Đào tạo trình độ: Đại học
- ❖ Học phần tiên quyết:
 - Tin học đại cương A
 - Nhập môn lập trình

(*) SV có thể xem lại tại trang khoá học ở trang: elearning.ntu.edu.vn



Thông tin về giảng viên:

- ❖ Họ và tên: Phạm Thị Thu Thúy
- ❖ Chức danh, học hàm, học vị: TS, GVC
- ❖ Điện thoại: 0901905679
- ❖ Email: thuthuy@ntu.edu.vn
- ❖ Địa chỉ trang web/nguồn dữ liệu internet của giảng viên: elearning.ntu.edu.vn (Khoá học Cơ sở dữ liệu, GV. Phạm Thị Thu Thúy)
- ❖ Địa điểm, lịch tiếp SV:
 - Email, điện thoại
 - <https://meet.google.com/bvy-iibi-ave>



Mô tả tóm tắt học phần

- ❖ Học phần trang bị các kiến thức cơ bản về cơ sở dữ liệu quan hệ (Relational Databases).
- ❖ Hiểu được ý nghĩa và vai trò của cơ sở dữ liệu quan hệ, các phép toán đại số quan hệ, chuẩn hóa cơ sở dữ liệu, làm việc với dữ liệu thông qua ngôn ngữ định nghĩa và xử lý cơ sở dữ liệu. Kiến thức học phần làm tiền đề cho các môn học phân tích, thiết kế và cài đặt cơ sở dữ liệu.



Mục tiêu

- ❖ Giới thiệu các khái niệm tính chất của cơ sở dữ liệu và hệ quản trị CSDL.
- ❖ Hiểu được các mô hình dữ liệu, đặc biệt là mô hình dữ liệu quan hệ.
- ❖ Phát hiện các Phụ thuộc hàm, Khóa và chuẩn hóa cơ sở dữ liệu quan hệ.
- ❖ Trình bày truy vấn dữ liệu bằng các phép toán Đại số quan hệ.
- ❖ Khai thác CSDL với SQL.



Kết quả học tập mong đợi (KQHT):

Sau khi học xong học phần, sinh viên có thể:

- a) Hiểu các khái niệm cơ sở mô hình dữ liệu quan hệ
- b) Xác định được tập Phụ thuộc hàm, Khóa, Phủ tối thiểu, Phân rã thông tin
- c) Biểu diễn các truy vấn dữ liệu dựa trên các phép toán đại số quan hệ và SQL
- d) Chuẩn hóa cơ sở dữ liệu
- e) Có khả năng tiếp cận với các hệ cơ sở dữ liệu tiên tiến.

Thang điểm học phần

TT.	Hoạt động đánh giá	Hình thức/công cụ đánh giá	Nhằm đạt CLOs	Trọng số (%)
1	Đánh giá quá trình	<ul style="list-style-type: none"> - Chuyên cần - Hỏi/đáp, thảo luận trên lớp - Mức độ hoàn thành các hoạt động trên NTU elearning (xem bài giảng, làm bài tập, tham gia diễn đàn) 	a,b,c	30
2	Thi giữa kỳ	Tự luận/Tiểu luận	a,b,c	30
3	Thi cuối kỳ	Tự luận/Tiểu luận	c,d	40

Tài liệu học tập

TT.	Tên tác giả	Tên tài liệu	Năm xuất bản	Nhà xuất bản	Địa chỉ khai thác tài liệu	Mục đích sử dụng	
						Tài liệu chính	Tham khảo
1	Nguyễn Đức Thuần	CSDLQH	2007	ĐH Nha Trang	elearning	x	
2	Maier,D	The Theory of Relational Databases	1983	Computer Science Press, Rockville	elearning	x	
3	Nguyễn Đức Thuần, Trương Ngọc Châu	Phương pháp giải bài tập CSDLQH	2012	NXB Khoa học & Kỹ Thuật	Thư viện	x	
4	Ramez Elmasri, Shamkant B. Navathe	Fundamentals of Database Systems (7 th)	2016	Pearson	https://vn1lib.org/s/-Fundamentals%20of-%20Database%20-Systems%20		x



Phương pháp học tập

❖ Chuẩn bị:

- Đọc slide, tài liệu, xem video tại khoá học trên elearning.ntu.edu.vn.
- Làm ví dụ trong tài liệu số [3]
- bài tập trắc nghiệm/ở mức dễ.

❖ Đến lớp:

- Hỏi/đáp.
- củng cố kiến thức cốt lõi.
- Làm bài tập tự luận.
- Sửa bài tập tự luận.



Nội dung

1. Các khái niệm cơ bản
2. Các phép toán đại số quan hệ
3. Phụ thuộc hàm – Hệ tiên đề Armstrong
4. Phủ tối thiểu của tập phụ thuộc hàm
5. Khóa của quan hệ - lược đồ quan hệ
6. Phép phân rã bảo toàn thông tin
7. Ngôn ngữ SQL
8. Chuẩn hóa CSDL



Tại sao phải cần CSDL?

1



Hỏi/Đáp?

- ❖ Hãy nêu cách tiếp cận giải quyết bài toán quản lý bán hàng, quản lý lương nhân viên với các môn mà SV đã học như Nhập môn Lập trình, Kỹ thuật lập trình hay Excel?
- ❖ Đề xuất cách giải quyết trên Excel?

Những vấn đề xảy ra khi lưu trữ dữ liệu trong một danh sách

- ❖ Giả sử quản lý việc **đặt hàng** của khách hàng với danh sách (dạng bảng) như sau:

MSKH	TÊNKH	TP	PVC	MSMH	TÊNMH	ĐG	SL
S1	An	HCM	01	P1	Táo	650	300
S1	An	HCM	01	P2	Cam	500	200
S1	An	HCM	01	P3	Chanh	450	400
S2	Hòa	HN	02	P1	Táo	650	100
S2	Hoà	HN	02	P3	Chanh	450	300
S3	Thanh	NT	03	P2	Cam	500	200
S4	Trang	NT	03	P2	Cam	500	210

❖ Trong đó:

- Mã số khách hàng (MSKH),
- Tên khách hàng (TENKH),
- Thành phố (TP),
- Phí vận chuyển đến thành phố của khách (PVC),
- Mã số mặt hàng (MSMH),
- Tên mặt hàng (TENMH),
- Đơn giá (ĐG): đơn giá hàng khách đặt,
- Số lượng (SL): số lượng hàng khách đặt.



Hỏi/Đáp?

- ❖ Hãy nhận xét: Những hạn chế, khó khăn sẽ phát sinh khi thêm, xoá, sửa dữ liệu. Cho ví dụ cụ thể.



A: Những vấn đề phát sinh khi cập dữ liệu

- ➡ **Thêm:** Khi thêm thông tin cho một khách hàng chưa đặt hàng thì các thông tin về mặt hàng để trống → phí không gian lưu trữ.
- ➡ **Xoá:** Chẳng hạn, ông S3 chỉ mới đặt một mặt hàng duy nhất là P2, nếu đơn hàng này bị hủy bỏ → xoá dòng → mất hoàn toàn thông tin của khách hàng S3.
- ➡ **Sửa:** Giả sử đơn giá mặt hàng P2 thay đổi, ta phải đổi đơn giá của mặt hàng này trên toàn bộ quan hệ → sửa lâu.



Q&A: Nguyên nhân phát sinh các vấn đề đó???

❖ Gợi ý:

- Có bao nhiêu loại dữ liệu được lưu trữ trong bảng?



A: Nguyên nhân phát sinh các vấn đề đó???

❖ Các loại dữ liệu được lưu trữ trong bảng?

▪ Ít nhất là 3 loại dữ liệu về:

- Khách hàng
- Mặt hàng
- Dữ liệu về việc đặt hàng của khách
- Và dữ liệu về phí vận chuyển đến các thành phố

❖ Tất cả dữ liệu đưa vào 1 bảng → Một bảng phải chứa nhiều hơn một khái niệm (concept) hoặc nhiều hơn một chủ đề (theme) như: khách hàng, mặt hàng, và việc đặt hàng của khách...

→ Mỗi dòng chứa nhiều loại dữ liệu khác nhau.

→ Nên khi lưu chung vào 1 bảng sẽ bị các hạn chế như trên.



Q&A: Đề xuất phương án giải quyết ???

- ❖ Hãy đưa ra phương án giải quyết các vấn đề phát sinh khi thêm, xoá, sửa.
- ❖ Gợi ý: Dựa vào nguyên nhân để đưa ra đề xuất.



A: Đề xuất phương án giải quyết???

- ❖ Do lưu chung vào 1 bảng sẽ bị các hạn chế như trên.
- ➔ Tách 1 bảng đó ra thành nhiều bảng. Mỗi bảng chứa mỗi khái niệm, mỗi chủ đề khác nhau.

A: Phương án đề xuất 1:

KHACHHANG

MSKH	TÊNKH	TP	PVC
S1	An	HCM	01
S2	Hoà	HN	02
S3	Bình	NT	03
S4	Trang	NT	03

MATHANG

MSMH	TÊNMH	ĐG
P1	Táo	650
P2	Cam	500
P3	Chanh	450

DATTHANG

MSKH	MSMH	SL
S1	P1	300
S1	P2	200
S1	P3	400
S2	P1	100
S2	P3	300
S3	P2	200
S4	P2	210

→ gọi là CSDL quan hệ.

A: Phương án đề xuất 2:

KHACHHANG

MSKH	TÊNKH	TP
S1	An	HCM
S2	Hoà	HN
S3	Bình	NT
S4	Trang	NT

VANCHUYEN

TP	PVC
HCM	01
HN	02
NT	03

MATHANG

MSMH	TÊN MH	ĐG
P1	Táo	650
P2	Cam	500
P3	Chanh	450

DATTHANG

MSKH	MSMH	SL
S1	P1	300
S1	P2	200
S1	P3	400
S2	P1	100
S2	P3	300
S3	P2	200
S4	P2	210

→ gọi là CSDL quan hệ.



CSDL quan hệ

- ➔ **CSDL quan hệ** giải quyết được vấn đề về cập nhật (thêm, xoá, sửa) dữ liệu.
- ❖ Hơn nữa, trong thế giới nghiệp vụ, có nhiều **mối quan hệ phức tạp giữa các khái niệm** khác nhau. Ví dụ: một khách hàng có thể đặt nhiều mặt hàng, một mặt hàng có thể được đặt bởi nhiều khách hàng, hoặc 1 khách hàng thuộc một thành phố, một thành phố có nhiều khách hàng...
- ❖ **CSDL quan hệ** không chỉ giải quyết các **vấn đề phát sinh những vấn đề với danh sách** mà còn cho phép chúng **mô hình hoá** các quan hệ vốn có giữa những dữ liệu nghiệp vụ.



Q&A?

- ❖ Vậy làm sao để có lại được toàn bộ dữ liệu như danh sách ban đầu?
- ❖ Hãy nêu cách làm thủ công?



A

- ❖ Nếu cần có lại toàn bộ thông tin như danh sách ban đầu
 - ➔ Ngôn ngữ truy vấn mang tính cấu trúc (Structured Query Language – SQL).



Môn học CSDL

- ❖ CSDL là gì?
 - ❖ MS SQL Server?
 - ❖ Cơ sở để có được các phương án giải quyết cho vấn đề cập nhật đã nêu? Tại sao là 3 hay 4 bảng? Còn phương án nào khác không? Có chuẩn gì để đánh giá không? Trong ví dụ này 3 bảng hay 4 bảng là tốt hơn?
- ➔ Môn học CSDL.

CSDL quan hệ là gì?

- ❖ Cơ sở dữ liệu quan hệ lưu trữ dữ liệu dưới dạng bảng (table).
- ❖ Một bảng đơn giản là một mảng 2 chiều gồm các cột (column) và các dòng (row).
 - Column thể hiện các thuộc tính khác nhau của một đối tượng.
 - Row thể hiện 1 thực thể của đối tượng.
- ❖ VD:

KHACHHANG

MSKH	TÊNKH	TP
S1	An	HCM
S2	Hoà	HN
S3	Bình	NT
S4	Trang	NT



CSDL quan hệ

- ❖ Việc liên kết các dòng ở các bảng khác nhau lại dựa trên sự tương ứng giữa các cặp giá trị (matched pair value).
- ❖ Các cặp giá trị tương ứng này cho phép chúng ta liên kết một dòng ở một bảng với một hoặc nhiều dòng ở một bảng khác.
- ❖ Ví dụ:
 - 2 bảng: KHACHHANG và ĐATHANG:
 - sử dụng thuộc tính MSKH để xác định một khách hàng là duy nhất
 - dùng chính MSKH để đưa vào bảng ĐATHANG.
 - Và dùng MSKH của mỗi khách hàng trong mỗi bảng để liên kết các dòng cùng chung MSKH lại với nhau.
- ➔ CSDL tối ưu được dữ liệu dư thừa.
- ➔ Cho phép ta mô hình hoá các quan hệ tự nhiên tồn tại



CSDL quan hệ

- CSDL hạn chế được dữ liệu dư thừa.
- Cho phép ta mô hình hoá các quan hệ tự nhiên tồn tại giữa các đối tượng nghiệp vụ trong thế giới thực.
- Có cơ sở vững chắc để tạo ra các báo cáo hoặc tạo giao diện cho người dùng.



Ngôn ngữ truy vấn mang tính cấu trúc (The Structured Query Language - SQL)

- ❖ Trong database quan hệ, ta sử dụng ngôn ngữ truy vấn (Structured Query Language – SQL).
- ❖ Và đây chính là phương tiện cơ bản để làm việc với CSDL.
- ❖ SQL được thiết kế dễ hiểu và dễ dùng.

VD SQL: The Art Course Database[8]

CUSTOMER

CustomerNumber	CustomerLastName	CustomerFirstName	Phone
1	Johnson	Ariel	206-567-1234
2	Green	Robin	425-678-2345
3	Jackson	Charles	500-789-3456
4	Pearson	Jeffery	206-567-3345
5	Sears	Miguel	500-789-1567
6	Kyle	Leah	425-678-7054
7	Myers	Lynnda	500-789-5678
(New)			

COURSE

CourseNumber	CourseTitle	CourseDate	Fee
1	Acryl Pastels	10/1/2011	\$500.00
2	Reg Oils	9/15/2011	\$350.00
3	Int Pastels	8/15/2011	\$350.00
4	Reg Oils	10/15/2011	\$350.00
5	Acryl Pastels	11/15/2011	\$400.00
(New)			\$0.00

ENROLLMENT

CustomerNumber	CourseNumber	EnrollmentFee
1	1	\$500.00
1	3	\$350.00
2	3	\$350.00
3	1	\$500.00
4	1	\$500.00
5	2	\$350.00
6	2	\$350.00
7	4	\$0.00
8	5	\$0.00

Can change COURSE
CourseDate without
problems

Can insert new
COURSE data as
needed

Can delete
ENROLLMENT
rows as needed—
no adverse
consequences



VD SQL: The Art Course Database[8]

- ❖ Để gộp toàn bộ dữ liệu từ 3 bảng, ta sử dụng câu lệnh SELECT sau đây:

```
SELECT Customer.customerLastName,  
       Customer.customerFirstName,  
       Customer.phone,  
       Course.courseDate, Enrollment.amountPaid,  
       Course.course, Course.fee  
FROM   Customer, Enrollment, Course  
WHERE  Customer.customerNumber = Enrollment.customerNumber  
       AND Course.courseNumber = Enrollment.courseNumber;
```



VD SQL: The Art Course Database[8]

❖ Kết quả như sau:

Art Course List						
CustomerLastName	CustomerFirstName	Phone	CourseDate	AmountPaid	Course	Fee
Johnson	Ariel	206-567-1234	10/1/2011	\$250.00	Adv Pastels	\$500.00
Johnson	Ariel	206-567-1234	3/15/2011	\$350.00	Int Pastels	\$350.00
Green	Robin	425-678-8765	9/15/2011	\$350.00	Beg Oils	\$350.00
Jackson	Charles	360-789-3456	10/1/2011	\$500.00	Adv Pastels	\$500.00
Pearson	Jeffery	206-567-2345	10/1/2011	\$500.00	Adv Pastels	\$500.00
Sears	Miguel	360-789-4567	9/15/2011	\$350.00	Beg Oils	\$350.00
Kyle	Leah	425-678-7654	11/15/2011	\$250.00	Adv Pastels	\$500.00
Myers	Lynda	360-789-5678	10/15/2011	\$0.00	Beg Oils	\$350.00

Records: 8 1 of 8 No Filter Search



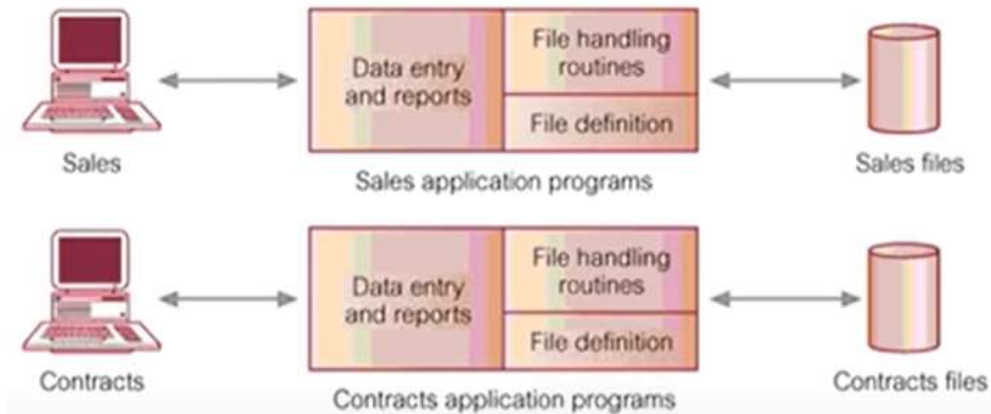
Tại sao phải cần Cơ sở dữ liệu???

2

Xử lý dựa trên hệ thống tập tin

- ❖ Mỗi loại người dùng định nghĩa và xử lý các tập tin bằng một phần mềm ứng dụng cụ thể.
- ❖ Mỗi ứng dụng sẽ có một tập hợp các tập tin riêng chứa dữ liệu riêng.

❖ Ví dụ:



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)



Hỏi/Đáp?

- ❖ Như vậy, tiếp cận tập tin thì có hạn chế gì?
 - Gợi ý để SV trả lời:
 - Thông tin có thể bị trùng lặp không?
 - Gây ra hậu quả gì khi cập nhật trên hệ thống tập tin này mà không cập nhật bên hệ thống tập tin kia?
- ❖ Hãy cho thêm VD khác minh họa sự hạn chế này.



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.
- ❖ Phụ thuộc dữ liệu
 - Các cấu trúc tập tin được định nghĩa cố định trong chương trình.
- ❖ Định dạng tập tin không tương thích → khó chia sẻ
- ➔ Giải pháp: Giảm bớt các hạn chế của cách tiếp cận theo hệ thống tập tin.
- ➔ **Cách tiếp cận CSDL.**



Cách tiếp cận cơ sở dữ liệu

- ❖ Thay vì thiết kế dữ liệu ở nhiều nơi, thành 2 hệ thống tập tin riêng lẻ thì ta thiết kế dữ liệu tập trung vào 1 chỗ và mỗi dữ liệu được lưu 1 lần, sau đó các ứng dụng lấy dữ liệu từ 1 nơi nhất quán, không sai lệch, hy vọng chuyện trùng lặp là không tồn tại, chuyện không nhất quán là không xảy ra, các chuyện không phụ thuộc vào dữ liệu, câu truy vấn linh hoạt hơn... họ đề ra những mô hình để bảo đảm chuyện đó.
- ❖ Nếu có mô hình dữ liệu tốt sẽ bảo đảm:
 - độc lập với cấu trúc dữ liệu bên dưới;
 - Câu truy vấn không còn phụ thuộc cấu trúc dữ liệu ;
 - Có thể chia sẻ dữ liệu.

Cách tiếp cận cơ sở dữ liệu

❖ Thay vì thiết kế 2 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

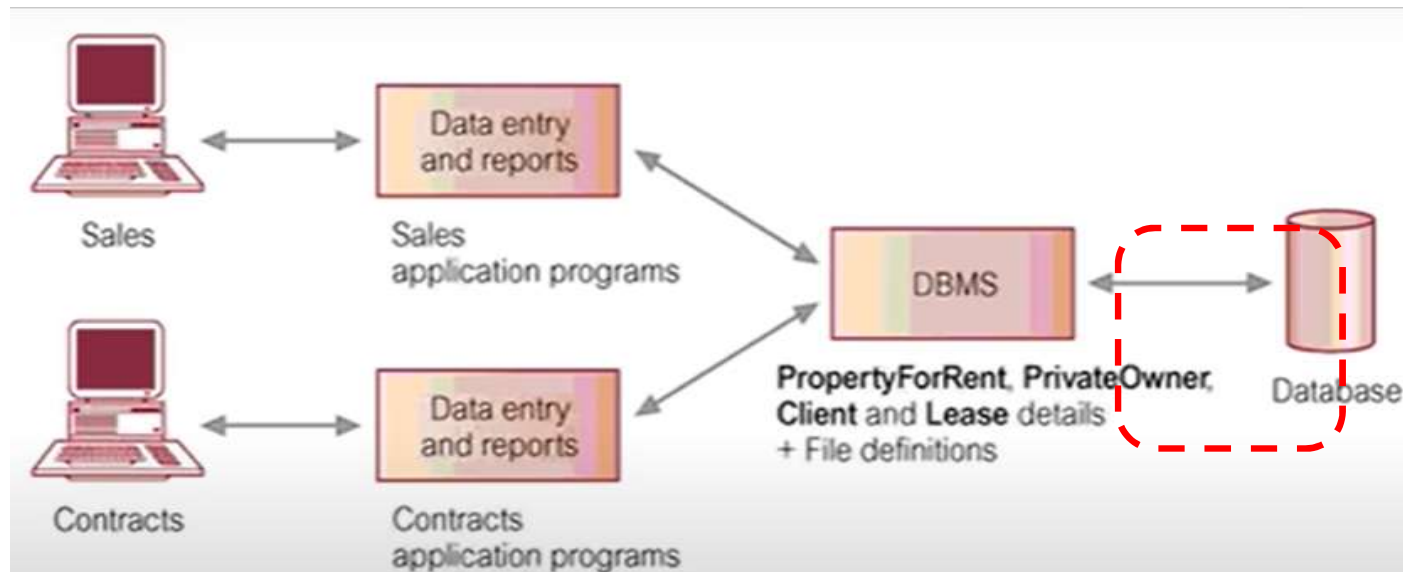
Lease (leaseNo, propertyNo, clientNo, rent, paymentMethod, deposit, paid, rentStart, rentFinish, duration)

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

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

Cách tiếp cận cơ sở dữ liệu

- ❖ Dữ liệu được tập trung vào 1 nơi duy nhất:



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)



Cách tiếp cận cơ sở dữ liệu

- ❖ Tập trung DL vào 1 chỗ và mỗi dữ liệu lưu 1 lần ở 1 hình trụ duy nhất, dữ liệu được chứa trong hình trụ đó gọi là CSDL.
- ❖ Ta không cần 2 hệ thống quản lý tập tin riêng lẻ mà chỉ cần 1 hệ thống duy nhất gọi là Hệ quản trị CSDL (DataBase Management System – DBMS)
- ❖ Sau đó các chương trình họ phát triển đều thông qua HQT CSDL này để truy cập dl trong CSDL.
- ❖ Với cách tổ chức này giúp giảm dư thừa, bảo đảm nhất quán và nếu DBMS này dựa trên mô hình tốt thì bảo đảm được tính độc lập với dl được lưu trữ bên dưới, cho phép phát triển ứng dụng mà ko quan tâm đến ctdl được lưu trữ bên dưới, rất linh hoạt và lúc đó dl cũng có thể được chia sẻ hiệu quả nữa.
- ❖ Như vậy khi làm việc với CSDL phải chú ý 2 điểm:
 - Phải có cách tổ chức CSDL tốt.
 - Phải có hệ phần mềm QT CSDL giúp thiết kế, triển khai và khai thác csdl.



Đặc điểm của dữ liệu trên CSDL

- ❖ Dữ liệu lưu trữ trên CSDL 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

