

# CƠ SỞ DỮ LIỆU

# GIỚI THIỆU MÔN HỌC

## ■ Nội dung môn học

- |  |         |
|--|---------|
| 1. Tổng quan về CSDL   | 8 tiết  |
| 2. Ngôn ngữ và mô hình quan hệ <ul style="list-style-type: none"><li>• Mô hình quan hệ</li><li>• Đại số quan hệ và tính toán trên quan hệ</li><li>• Ngôn ngữ SQL</li><li>• Hệ quản trị cơ sở dữ liệu quan hệ hướng đối tượng</li></ul>             | 20 tiết |
| 3. Phân tích và thiết kế CSDL <ul style="list-style-type: none"><li>■ Vòng đời phát triển của hệ thống cơ sở dữ liệu</li><li>■ Phân tích cơ sở dữ liệu trên bài toán thực tế</li><li>■ Mô hình thực thể liên hệ (ER)</li><li>■ Chuẩn hóa</li></ul> | 16 tiết |
| 4. Các phương pháp thiết kế cho mô hình quan hệ  | 10 tiết |

# GIỚI THIỆU MÔN HỌC

## ■ Cách đánh giá

### ■ 30% Điểm QT

- 50% đánh giá trên lớp: Lên bảng, chấm kiểm tra nhanh...
- 50% kiểm tra (Bài KT1-ĐSQH, Bài KT2-Lý thuyết thiết kế)

### ■ 70% Thi

- 1. Tự thiết kế ER cho 1 bài toán, dùng SQL Server cài đặt
- 2. Thực hành SQL trên SQL Server

# GIỚI THIỆU MÔN HỌC (TIẾP)

- Tài liệu tham khảo
  - Jan L. Harrington - Relational Database Design and Implementation-Morgan Kaufmann (2016)

# TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU

- Một số hệ thống cơ sở dữ liệu phổ biến
- Hệ thống dựa trên tệp tin
  - Đặc điểm
  - Các vấn đề với hệ thống dựa trên tệp tin
- Ý nghĩa của cơ sở dữ liệu (CSDL)
- Hệ quản trị CSDL(DBMS)
  - Ý nghĩa và chức năng điển hình của DBMS
  - Các thành phần chính của DBMS
  - Nhân sự tham gia vào DBMS
  - Lịch sử phát triển của DBMS
  - Ưu điểm và nhược điểm của DBMS

# MỘT SỐ HỆ THỐNG CSDL PHỔ BIẾN

- Mua hàng trong siêu thị
- Mua bằng thẻ tín dụng
- Đặt tour du lịch tại đại lý du lịch
- Hệ thống thư viện
- Mua bảo hiểm
- Thuê phim
- Quản lý học tập tại trường đại học

# HỆ THỐNG DỰA TRÊN TẬP TIN (1)

- Ví dụ chương trình sử dụng CSDL DreamHome, thông tin từ phòng Bán hàng

**DreamHome**  
Property for Rent Details  
Property Number: PG21

Address <u>18 Dale Rd</u>	Allocated to Branch: <u>163 Main St, Glasgow</u>
City <u>Glasgow</u>	Branch No. <u>8003</u>
Postcode <u>G12</u>	Staff Responsible <u>Ann Beech</u>
Type <u>House</u> Rent <u>600</u>	
No. of Rooms <u>5</u>	

**Owner's Details**

Name <u>Carol Farrel</u>	Business Name _____
Address <u>6 Achray St,</u> <u>Glasgow G32 9DX</u>	Address _____
Tel. No. <u>0141-357-7419</u>	Tel. No. _____
Owner No. <u>CO87</u>	Owner No. _____
	Contact Name _____
	Business Type _____

**DreamHome**  
Client Details  
Client Number: CR74

First Name <u>Mike</u>	Last Name <u>Ritchie</u>
Address <u>18 Tain St,</u> <u>PA1G 1YQ</u>	Tel. No. <u>01475-392178</u>

**Property Requirement Details**

Preferred Property Type <u>House</u>	Maximum Monthly Rent <u>750</u>
General Comments <u>Currently living at home with parents</u> <u>Getting married in August</u>	

Seen By <u>Ann Beech</u>	Date <u>24-Mar-13</u>
Branch No. <u>8003</u>	Branch City <u>Glasgow</u>

(b)

PropertyForRent

propertyNo	street	city	postcode	type	rooms	rent	ownerNo
PA14	16 Holthead Rd	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

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

Figure 1.2 The PropertyForRent, PrivateOwner, and Client files used by Sales.

Figure 1.1 Sales Department forms (a) Property for Rent Details form; (b) Client Details form.

# HỆ THỐNG DỰA TRÊN TẬP TIN (2)

- Ví dụ chương trình sử dụng CSDL DreamHome, thông tin từ phòng Hợp đồng

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-13</u> <b>Rent Finish Date</b> <u>30-Jun-14</u> <b>Duration</b> <u>1 Year</u>

Lease

leaseNo	propertyNo	clientNo	rent	payment Method	deposit	paid	rentStart	rentFinish	duration
10024	PA14	CR62	650	Visa	1300	Y	1-Jun-13	31-May-14	12
10075	PL94	CR76	400	Cash	800	N	1-Aug-13	31-Jan-14	6
10012	PG21	CR74	600	Cheque	1200	Y	1-Jul-13	30-Jun-14	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

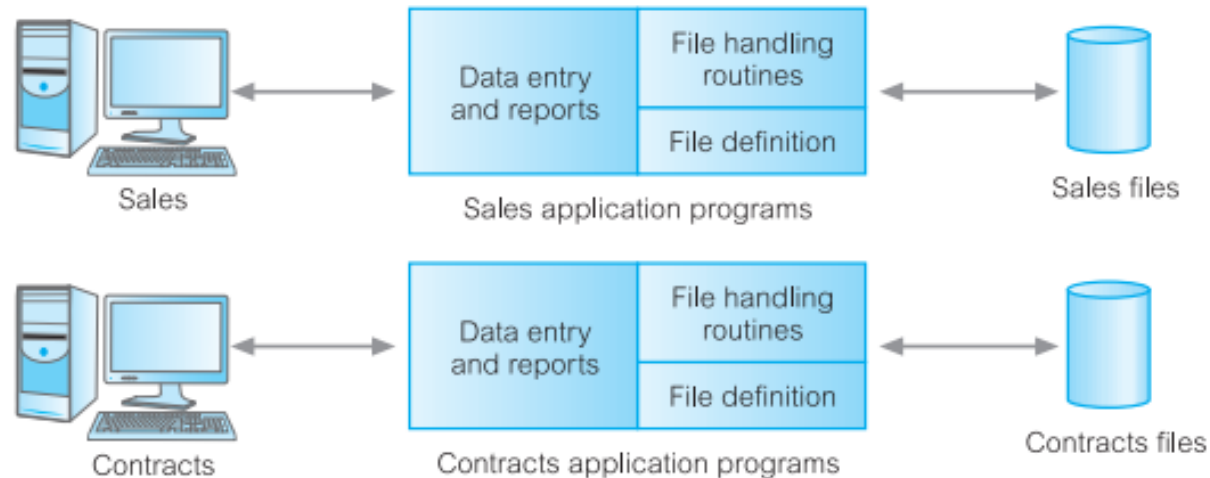
Client

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



# HỆ THỐNG DỰA TRÊN TẬP TIN (3)

- Cơ chế truy cập của hệ thống dựa trên tập tin của DreamHome



**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)

- Tập hợp các ứng dụng thực hiện các dịch vụ cho người dùng cuối
- Mỗi chương trình xác định và quản lý dữ liệu riêng của mình.

# CÁC VẤN ĐỀ VỚI HỆ THỐNG DỰA TRÊN TẬP TIN

- DL được định nghĩa và tổ chức bằng các cấu trúc của ngôn ngữ lập trình.

Ví dụ trong **C**:

```
struct Sinhvien { char masv[6]; char tensv[30];  
float diem };  
Sinhvien x;
```

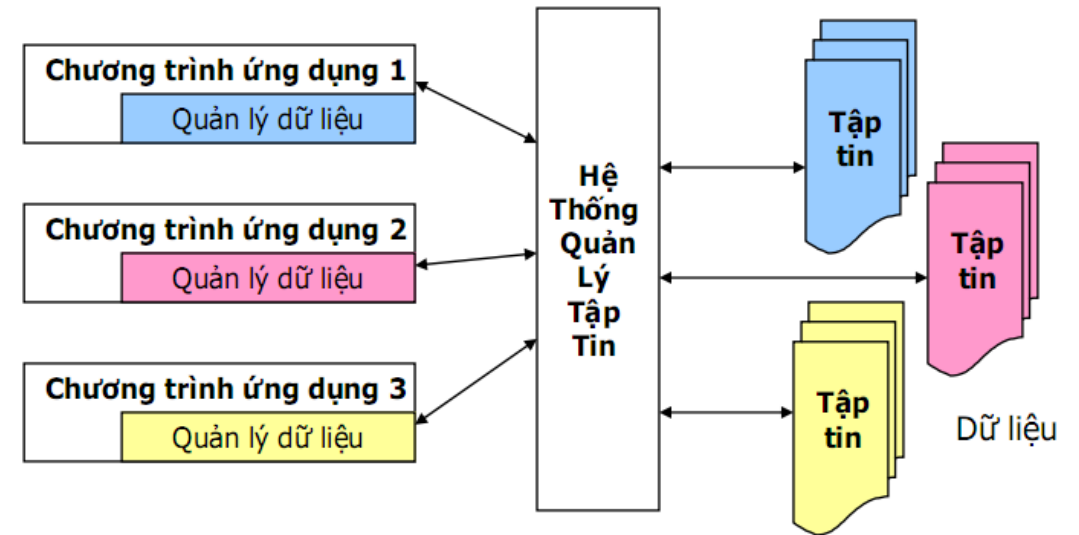
```
/* Đọc DL từ file */  
ifstream fin("dulieu.dat");  
fin.read ((char*) &x, sizeof(Sinhvien));  
cout << x.masv << x.tensv << x.diem;
```

Ví dụ trong **Pascal**:

```
Type Sinhvien = record  
    masv string[6], tensv string[30],  
    diem real  
end;  
Var x: Sinhvien;  
    f: file of Sinhvien;  
/* Đọc DL từ file */  
reset (f, "dulieu.dat");  
while not eof(f) do  
    read (f, x)  
end;  
writeln(x.masv, ",", x.tensv, ",", x.diem);
```

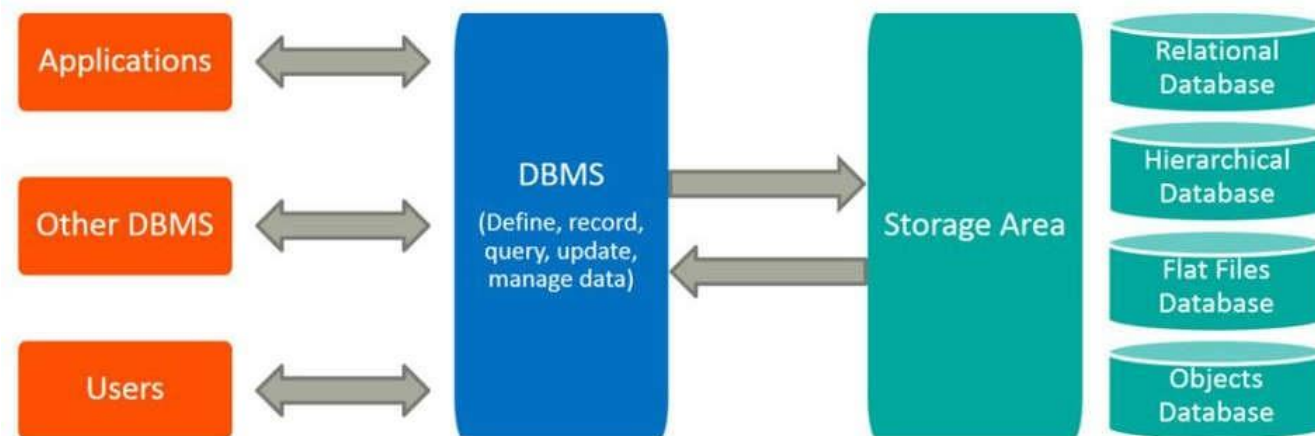
# CÁC VẤN ĐỀ VỚI HỆ THỐNG DỰA TRÊN TẬP TIN (2)

- Tổ chức lưu trữ DL phụ thuộc hoàn toàn vào ngôn ngữ sử dụng để xây dựng CT ứng dụng
  - Chương trình được viết bởi các ngôn ngữ khác nhau có cấu trúc tệp khác nhau
  - Chương trình được viết để đáp ứng 1 chức năng cụ thể
  - Bất kỳ yêu cầu mới nào cũng đều phải viết chương trình mới
- Dữ liệu bị trùng lặp và dư thừa
  - Cùng một dữ liệu ở nhiều chương trình khác nhau.
  - Lãng phí bộ nhớ và tiềm ẩn khả năng cùng 1 mục nội dung nhưng ở những chương trình khác nhau có định dạng khác nhau.
- Việc chia sẻ dữ liệu với các hệ thống khác hạn chế, mỗi ỨD cần một hệ thống tệp riêng
- Nếu cải tiến ỨD → có thể phải thay đổi cả cấu trúc hệ thống file DL
- Tìm kiếm dữ liệu chậm do duyệt tuần tự trên file DL
- Chi phí bảo trì dữ liệu lớn



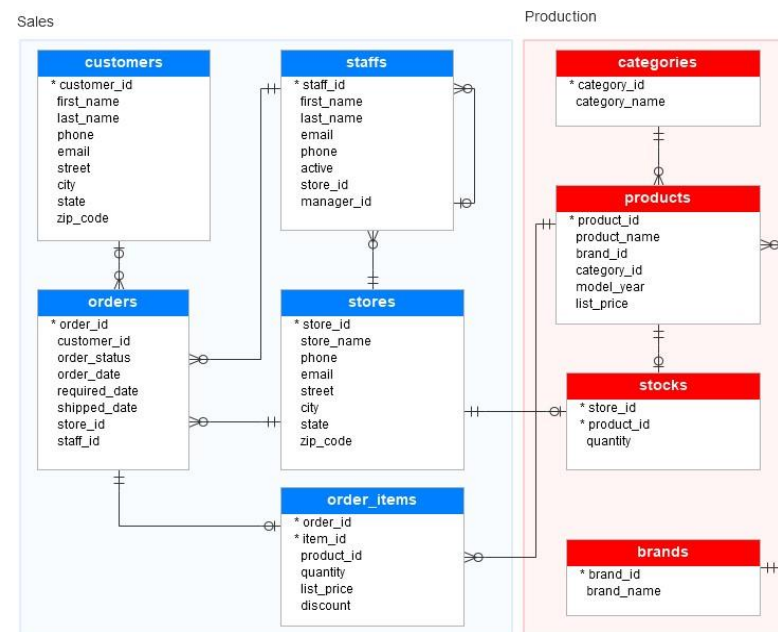
# CÁCH TIẾP CẬN THEO CƠ SỞ DỮ LIỆU

- Xuất hiện vấn đề:
  - Định nghĩa dữ liệu được nhúng trong một chương trình, thay vì lưu trữ riêng biệt và độc lập.
  - Chương trình sẽ không kiểm soát quyền truy cập và thao tác dữ liệu.
- Kết quả:
  - Cần một CSDL (Database) và hệ thống quản lý CSDL (Database Management System - DBMS)

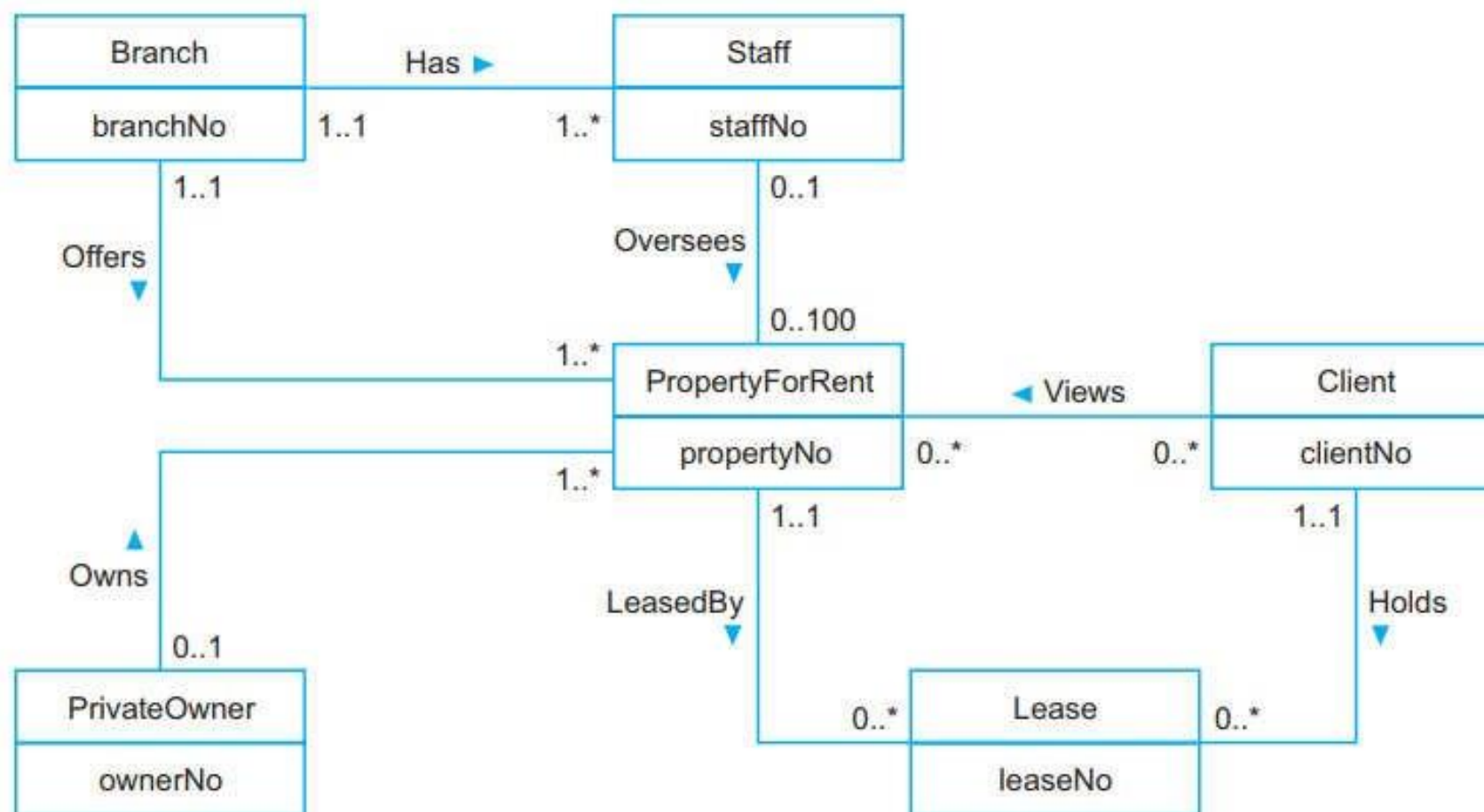


# CƠ SỞ DỮ LIỆU LÀ GÌ ?

- Tập hợp dữ liệu có quan hệ về mặt logic, được thiết để đáp ứng nhu cầu thông tin của một tổ chức
- Danh mục hệ thống cung cấp các mô tả dữ liệu cho phép độc lập giữa dữ liệu và chương trình
- Dữ liệu có quan hệ về mặt logic gồm: thực thể, thuộc tính, mối quan hệ về thông tin của một tổ chức



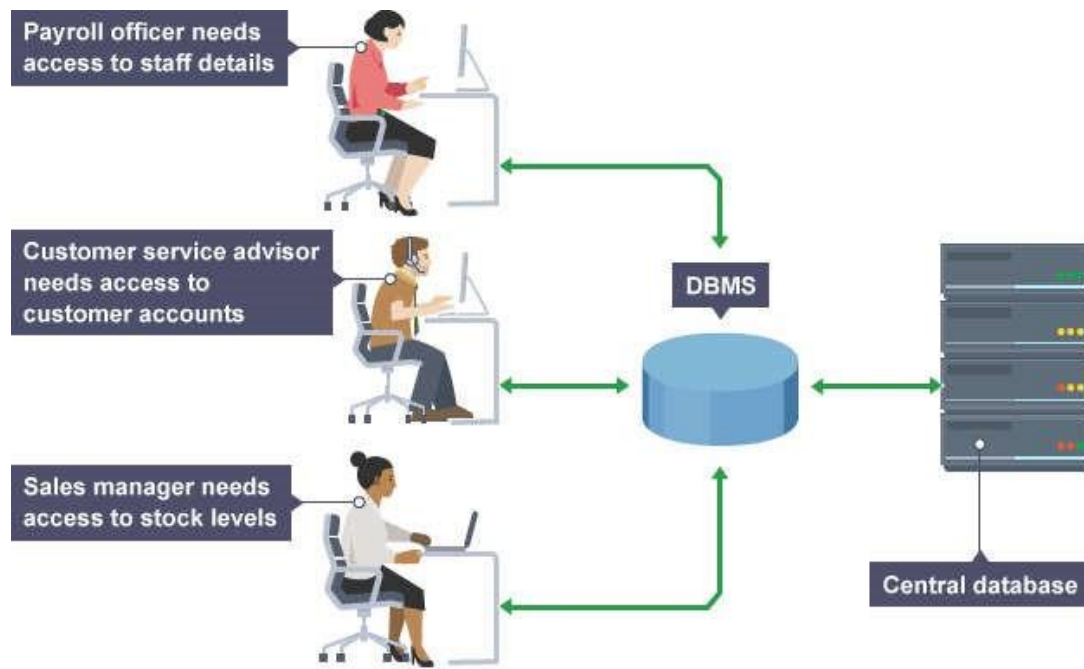
## ■ Ví dụ: Lược đồ ER của CSDL DreamHome



**Figure 1.6** Example Entity–Relationship diagram.

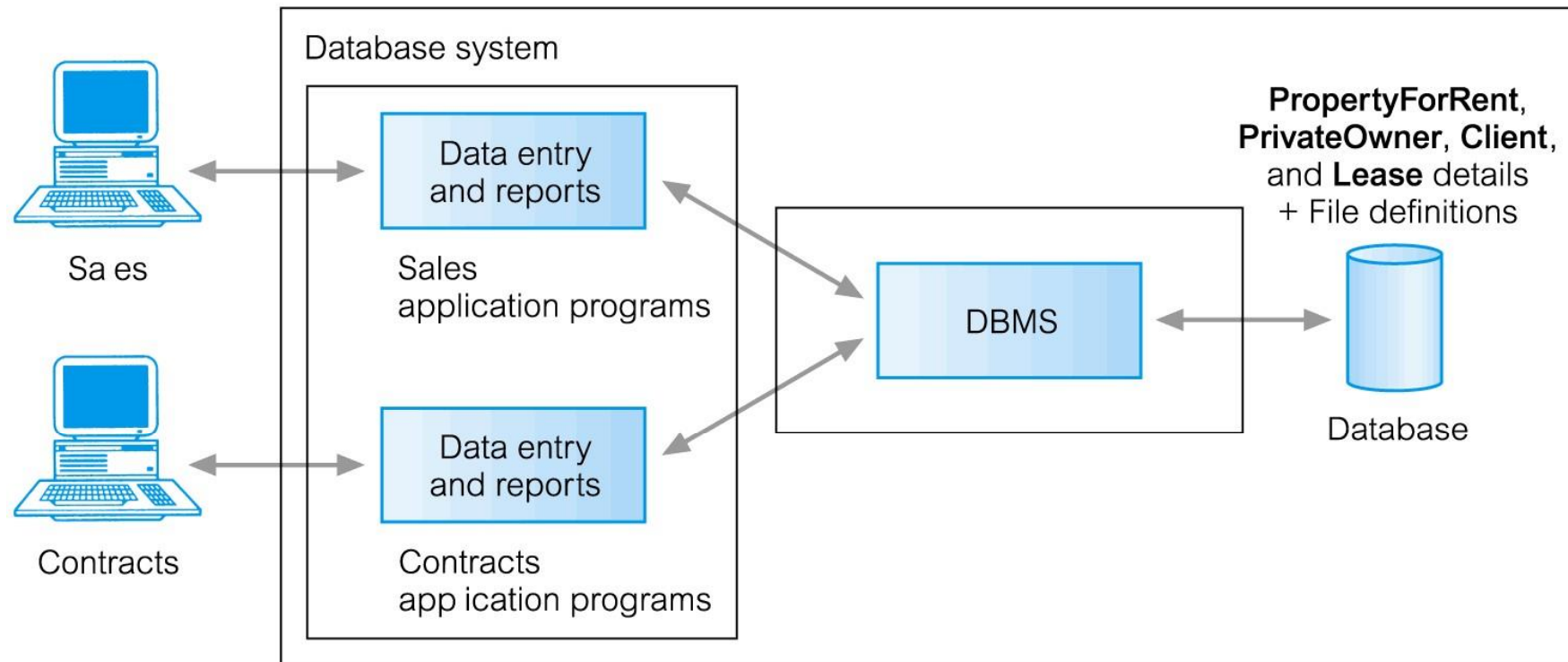
# HỆ QUẢN TRỊ CSDL – DBMS (1)

- Là một phần mềm cho phép người dùng định nghĩa, tạo, duy trì và kiểm soát dữ liệu vào CSDL.
- Chương trình ứng dụng: là một chương trình máy tính tương tác với CSDL bằng cách đưa ra một yêu cầu (câu lệnh SQL) đến Hệ quản trị CSDL.





# HỆ QUẢN TRỊ CSDL – DBMS (2)



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

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

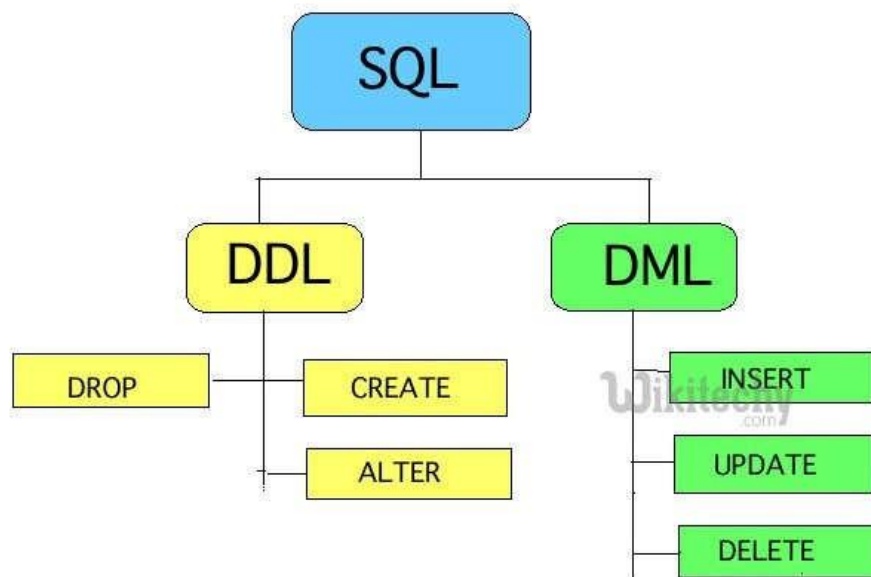
**Client** (clientNo, fName, lName, address, telNo, prefType, maxRent)

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



# HỆ QUẢN TRỊ CSDL – DBMS (3)

- Hệ quản trị CSDL cung cấp:
  - Ngôn ngữ định nghĩa dữ liệu (Data definition language – DDL):
    - Đặc tả các loại dữ liệu, cấu trúc và ràng buộc dữ liệu.
    - Các đặc tả đều được lưu trữ trong CSDL
  - Ngôn ngữ thao tác dữ liệu (Data manipulation language - DML):
    - Dễ dàng đưa ra yêu cầu về dữ liệu bằng ngôn ngữ truy vấn (query language)



# HỆ QUẢN TRỊ CSDL – DBMS (4)

- DBMS kiểm soát quyền truy cập vào CSDL bao gồm:
  - **Hệ thống bảo mật:** Ngăn người dùng truy cập trái phép CSDL
  - **Hệ thống toàn vẹn:** Đảm bảo tính nhất quán của dữ liệu
  - **Hệ thống kiểm soát đồng thời:** chia sẻ quyền truy cập dữ liệu
  - **Hệ thống kiểm soát khôi phục:** khôi phục dữ liệu khi có sự cố xảy ra
  - **Danh mục người dùng có quyền truy cập:** Những người dùng được phép sử dụng dữ liệu

# HỆ QUẢN TRỊ CSDL – DBMS (5)

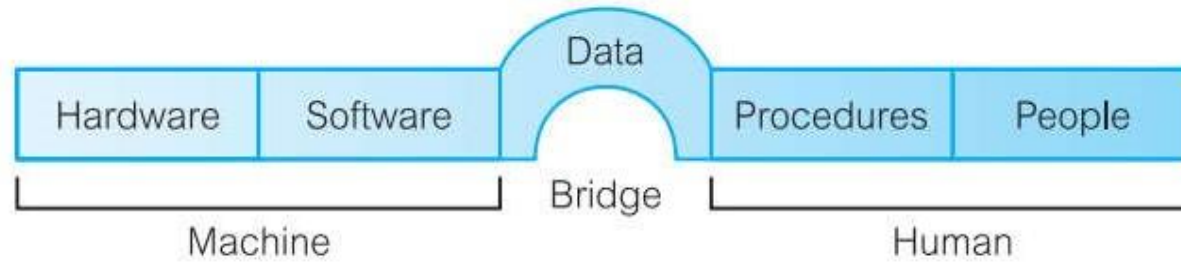
## ■ **View** – Khung nhìn

- Cho phép mỗi người dùng có cái nhìn riêng về dữ liệu
- Về cơ bản, nó giống như một tập hợp con của CSDL

## ■ **Lợi ích của View:**

- Đảm bảo an toàn dữ liệu, hạn chế một số dữ liệu không cho người dùng thấy.
- Cung cấp cơ chế tùy chỉnh hình thức của CSDL. Ví dụ, thay vì làm việc với thuộc tính StaffNo thì người dùng có thể làm việc với tên Mã nhân sự.
- Cung cấp cái nhìn nhất quán về CSDL, nhưng không làm thay đổi về cấu trúc của dữ liệu.

# THÀNH PHẦN TRONG MÔI TRƯỜNG DBMS



- **Phần cứng:**
  - Các thiết bị cần thiết để triển khai DBMS
  - ví dụ như máy tính, mạng máy tính ...
- **Phần mềm:** DBMS, Hệ điều hành, phần mềm mạng, chương trình ứng dụng
- **Dữ liệu:** Được sử dụng bởi tổ chức và các mô tả dữ liệu
- **Thủ tục:** Là các quy tắc và hướng dẫn về thiết kế và sử dụng dữ liệu
- **Con người:** người tham gia vào hệ thống

# THÀNH PHẦN TRONG MÔI TRƯỜNG DBMS

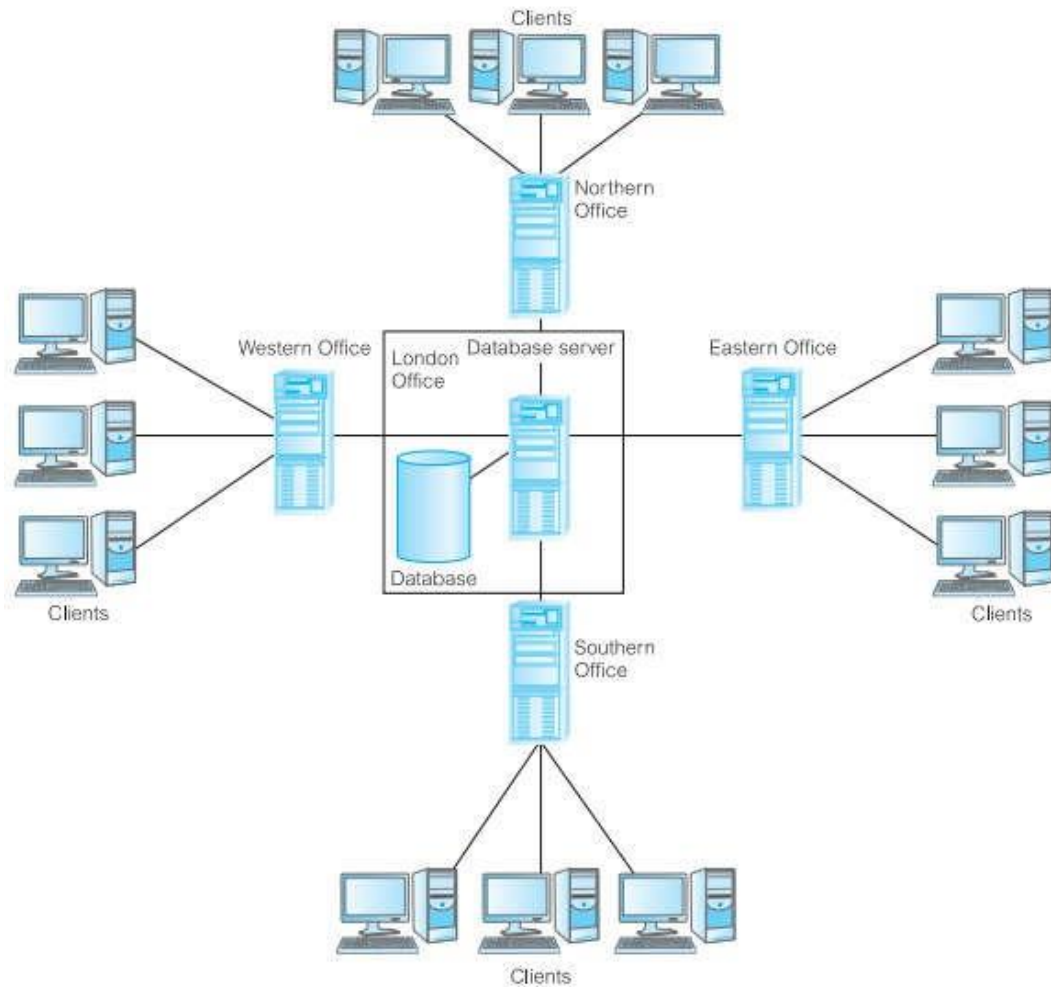
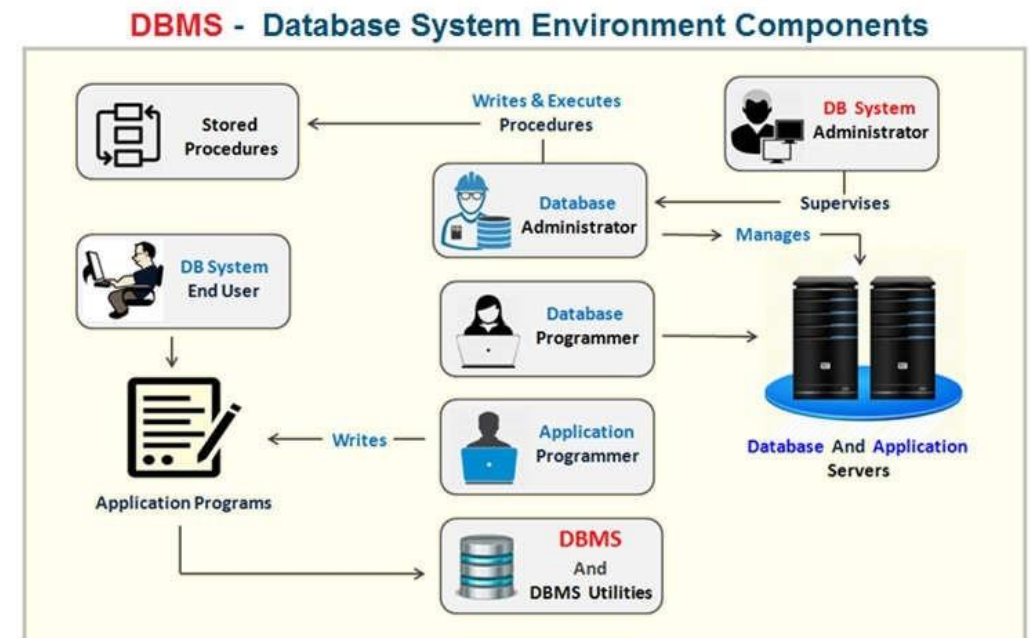


Figure 1.9 DreamHome hardware configuration.



# THIẾT KẾ CƠ SỞ DỮ LIỆU

- Mỗi CSDL cần có cấu trúc. Ví dụ CSDL DreamHome xác định được cấu trúc gồm 4 bảng

**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)

- Nhưng làm thế nào để có cấu trúc này?
- Thiết kế này đã tốt chưa?
- Có đáp ứng yêu cầu người dùng không?



Học Phần **Thiết kế dữ liệu** sẽ giúp chúng ta trả lời các câu hỏi trên

# VAI TRÒ (ROLES) TRONG MÔI TRƯỜNG DBMS

## ■ Quản trị CSDL và dữ liệu:

- Quản trị dữ liệu (DA): lập kế hoạch; phát triển và duy trì; thiết kế logic;...
- Quản trị CSDL (DBA): thiết CSDL vật lý và triển khai; kiểm soát tính toàn vẹn và bảo mật; duy trì vận hành;...

## ■ Thiết kế CSDL:

- Thiết kế logic: nắm rõ nghiệp vụ, nhu cầu của tổ chức
- Thiết kế vật lý: chuyển từ thể kế logic sang triển khai cụ thể

## ■ Phát triển ứng dụng:

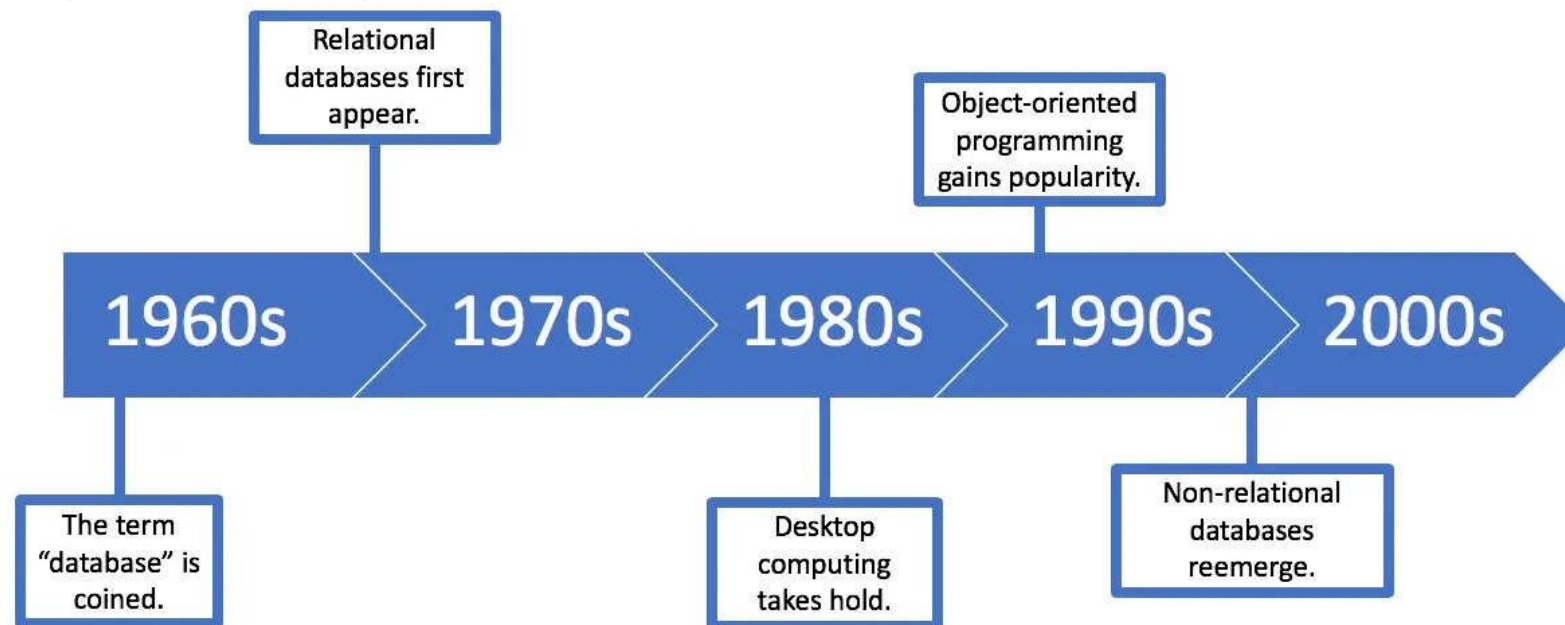
- Xây dựng chương trình cung cấp các chức năng cho người dùng.
- Mỗi chức năng có thể gồm các thao tác trên CSDL

## ■ Người dùng cuối:

- Là người được thiết kế, cài đặt và duy trì nhằm đáp ứng nhu cầu thông tin

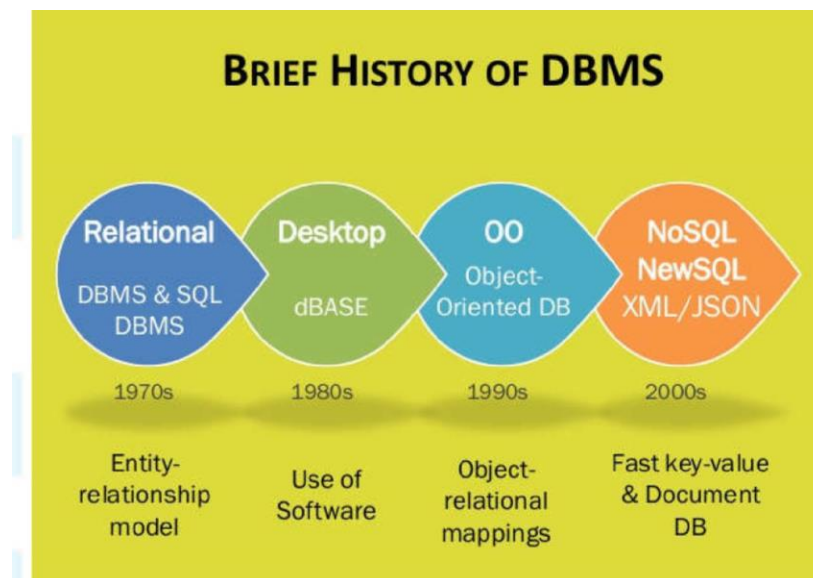
# LỊCH SỬ HỆ THỐNG CSDL

- Thế hệ thứ nhất: Cơ sở dữ liệu phân cấp và mạng
- Thế hệ thứ 2: Cơ sở dữ liệu quan hệ
- Thế hệ thứ 3:
  - CSDL lai giữa quan hệ và hướng đối tượng
  - CSDL hướng đối tượng





TIMEFRAME	DEVELOPMENT	COMMENTS
1960s (onwards)	File-based systems	Precursor to the database system. Decentralized approach: each department stored and controlled its own data.
Mid-1960s	Hierarchical and network data models	Represents first-generation DBMSs. Main hierarchical system is IMS from IBM and the main network system is IDMS/R from Computer Associates. Lacked data independence and required complex programs to be developed to process the data.
1970	Relational model proposed	Publication of E. F. Codd's seminal paper "A relational model of data for large shared data banks," which addresses the weaknesses of first-generation systems.
1970s	Prototype RDBMSs developed	During this period, two main prototypes emerged: the Ingres project at the University of California at Berkeley (started in 1970) and the System R project at IBM's San José Research Laboratory in California (started in 1974), which led to the development of SQL.
1976	ER model proposed	Publication of Chen's paper "The Entity-Relationship model—Toward a unified view of data." ER modeling becomes a significant component in methodologies for database design.
1979	Commercial RDBMSs appear	Commercial RDBMSs like Oracle, Ingres, and DB2 appear. These represent the second generation of DBMSs.
1987	ISO SQL standard	SQL is standardized by the ISO (International Standards Organization). There are subsequent releases of the standard in 1989, 1992 (SQL2), 1999 (SQL:1999), 2003 (SQL:2003), 2008 (SQL:2008), and 2011 (SQL:2011).
1990s	OODBMS and ORDBMSs appear	This period initially sees the emergence of OODBMSs and later ORDBMSs (Oracle 8, with object features released in 1997).
1990s	Data warehousing systems appear	This period also see releases from the major DBMS vendors of data warehousing systems and thereafter data mining products.
Mid-1990s	Web-database integration	The first Internet database applications appear. DBMS vendors and third-party vendors recognize the significance of the Internet and support web-database integration.
1998	XML	XML 1.0 ratified by the W3C. XML becomes integrated with DBMS products and native XML databases are developed.



# ƯU ĐIỂM CỦA DBMS

- Kiểm soát dư thừa dữ liệu
- Đảm bảo tính nhất quán dữ liệu
- Cung cấp thông tin từ dữ liệu
- Chia sẻ dữ liệu
- Cải thiện tính toàn vẹn dữ liệu
- Cải thiện tính bảo mật dữ liệu
- Đòi hỏi tính tiêu chuẩn dữ liệu
- Tiết kiệm chi phí duy trì, bảo dưỡng

## Advantages of DBMS over a File system

Data  
concurrency

Data  
searching

Data  
redundancy

Data  
inconsistency

Data  
integrity

Data  
sharing

# ƯU ĐIỂM CỦA DBMS

- Cân bằng yêu cầu xung đột tài nguyên
- Cải thiện khả năng truy cập và phản hồi
- Tăng năng suất
- Cải thiện việc duy trì dữ liệu
- Tăng tính đồng thời
- Cải thiện dự phòng và khôi phục dữ liệu

# NHƯỢC ĐIỂM CỦA DBMS

- Sự phức tạp: đòi hỏi người thiết kế, quản trị, người dùng cuối hiểu rõ chức năng và sử dụng nó
- Kích thước: Không gian lưu trữ
- Chi phí cho DBMS: chi phí phải trả cho thuê và duy trì máy chủ dữ liệu
- Chi phí phần cứng bổ sung
- Chi phí chuyển đổi
- Hiệu suất: vì có thể phải chia sẻ cho nhiều người dùng
- Tác động xấu khi xảy ra sự cố.

## ■ Sinh viên trả lời các câu hỏi trang 80

### Review Questions

- 1.1 List four government sectors in your country that use database systems.
- 1.2 Discuss each of the following terms:
  - (a) data
  - (b) database
  - (c) database management system
  - (d) database application program
  - (e) data independence
  - (f) security
  - (g) integrity
  - (h) views
- 1.3 Describe the role of database management systems (DBMS) in the database approach. Discuss why knowledge of DBMS is important for database administrators.
- 1.4 Describe the main characteristics of the database approach and contrast it with the file-based approach.
- 1.5 Describe the five components of the DBMS environment and discuss how they relate to each other.
- 1.6 Discuss the roles of the following personnel in the database environment:
  - (a) data administrator
  - (b) database administrator
  - (c) logical database designer
  - (d) physical database designer
  - (e) application developer
  - (f) end-users
- 1.7 Discuss the three generations of DBMSs.
- 1.8 Why are views an important aspect of database management systems?

- Tổ chức dữ liệu dưới dạng tệp: text, Excel, Word, CSV,... cho các loại giấy tờ sau:
  - Chứng minh thư
  - Hóa đơn bán hàng
  - Hóa đơn tiền điện
- Thực hiện các bài tập 1.9 -> 1.14 trang 80 và 81.