



HỆ CƠ SỞ DỮ LIỆU

GV: ThS.Trịnh Thị Ngọc Linh
Email: ngoclinhnl@yahoo.com

Chương 2

MÔI TRƯỜNG CƠ SỞ DỮ LIỆU

- 2.1. Kiến trúc ANSI-PARC 3-mức
- 2.2. Các ngôn ngữ cơ sở dữ liệu
- 2.3. Mô hình dữ liệu
- 2.4. Hệ quản trị CSDL đa người dùng
- 2.5. Từ điển dữ liệu

2.1. Kiến trúc ANSI-PARC 3-mức

- Một CSDL có 3 mức biểu diễn:
 - Mức vật lý (mức trong)
 - Mức logic (mức khái niệm)
 - Mức khung nhìn (mức ngoài)
- Mục đích: giải phóng đa số người dùng khỏi sự quan tâm về lưu trữ và bảo trì dữ liệu

2.1.1. Mức vật lý

- Mức vật lý mô tả dữ liệu được lưu trữ như thế nào trong CSDL
- Đây là mức thể hiện các cài đặt có tính chất vật lý của CSDL để đạt được sự tối ưu khi thực hiện các thao tác tìm kiếm và lưu trữ



2.1.1. Mức vật lý (tt)

- Ví dụ: Biểu diễn mức vật lý đơn giản dữ liệu về nhân viên của công ty
 - mã nhân viên - kiểu số nguyên
 - mã chi nhánh mà nhân viên đó làm việc - kiểu số nguyên
 - họ đệm - kiểu chuỗi
 - tên - kiểu chuỗi
 - ngày sinh - kiểu ngày
 - lương - kiểu số thực
 - và một con trỏ đến bản ghi tiếp theo

2.1.2. Mức logic

- Mức logic mô tả những dữ liệu nào được lưu trữ trong CSDL và có những mối quan hệ nào giữa chúng
- Không quan tâm đến cách thức để lưu trữ
- Mức logic biểu diễn:
 - Các thực thể, các thuộc tính, và các mối quan hệ giữa các thực thể đó
 - Các ràng buộc trên dữ liệu, các thông tin về ngữ nghĩa của dữ liệu
 - Các thông tin về an ninh và toàn vẹn của dữ liệu

2.1.2. Mức logic (tt)

- Ví dụ: Biểu diễn mức logic như sau:
 - Công ty gồm các phòng ban (Department), mỗi phòng ban có một số hiệu, tên gọi khác nhau, một địa chỉ (Location), các số điện thoại (Telephone)
 - Có một người làm trưởng phòng ban, hàng năm được cấp một khoản kinh phí để hoạt động (Expense_Budget), và phải đạt một doanh thu (Revenue_Budget)
 - Mỗi phòng ban có thể có từ một đến nhiều nhân viên (Employee). Mỗi nhân viên có một mã số nhân viên (EmpNo), tên gọi, một công việc làm (Job), một khoản tiền lương hàng tháng (Salary), số hiệu phòng ban. Có thể theo dõi thêm ngày sinh (BirthDay), giới tính (Sex)...

2.1.3. Mức khung nhìn

- Mức khung nhìn mô tả phần CSDL liên quan đến NSD hay các chương trình ứng dụng
- NSD hay chương trình ứng dụng:
 - Có thể không được biết về cấu trúc tổ chức lưu trữ thông tin trong CSDL, tên gọi của các loại dữ liệu hay tên gọi của các thuộc tính
 - Chỉ làm việc trên phần CSDL theo cách "nhìn" do người quản trị hay chương trình ứng dụng quy định, gọi là khung nhìn (View)

2.1.3. Mức khung nhìn (tt)

■ Ví dụ:

- Phòng Kế toán
 - Thấy danh sách nhân viên đang làm các công việc cụ thể trong từng phòng ban với các mức lương thỏa thuận
 - không được thấy lý lịch của các nhân viên
- Lãnh đạo công ty
 - Thấy số lượng nhân viên, tổng số lương phải trả và ai là người lãnh đạo của từng phòng ban
- Phòng Tổ chức nhân sự
 - Có người được xem lý lịch của tất cả cán bộ, công nhân viên của công ty
 - Nhưng có người chỉ được xem lý lịch của những cán bộ, công nhân viên với mức lương từ n đồng trở xuống

Mức khung nhìn

Khung nhìn 1

Khung nhìn 2

...

Khung nhìn n

Khung logic

Khung vật lý

Thuận lợi của kiến trúc ANSI-PARC 3-mức

- Mỗi người dùng có thể thay đổi khung nhìn của họ, không làm ảnh hưởng đến những khung nhìn dữ liệu của những người dùng khác đang dùng chung CSDL
- Những tương tác của người dùng với CSDL không phụ thuộc vào những vấn đề chi tiết trong lưu trữ dữ liệu
- Người quản trị CSDL có thể thay đổi cấu trúc lưu trữ của CSDL mà không làm ảnh hưởng đến những khung nhìn của NSD

Thuận lợi của kiến trúc ANSI-PARC 3-mức (tt)

- Những thay đổi về khía cạnh vật lý trong lưu trữ, chẳng hạn như thay một thiết bị nhỏ thứ cấp mới, có thể không ảnh hưởng đến cấu trúc bên trong của CSDL
- Người quản trị CSDL có thể thay đổi cấu trúc tổng quát hay cấu trúc khái niệm của CSDL mà không làm ảnh hưởng đến tất cả người dùng

2.1.4. Lược đồ và thể hiện của CSDL

- Toàn bộ mô tả CSDL được gọi là lược đồ CSDL (database schema)
- Tương ứng với ba mức truy xuất dữ liệu nói trên có ba loại lược đồ:
 - Ở mức khung nhìn có nhiều lược đồ ngoài (hay còn gọi là lược đồ con)
 - Ở mức logic có lược đồ khái niệm (hay còn gọi là lược đồ logic)
 - Ở mức vật lý có lược đồ trong (hay còn gọi là lược đồ vật lý)

Khung nhìn 1

MaNV Hodem Ten Tuoi Luong

Khung nhìn 2

MaNV Ten Ma_chi_nhanh



Mức logic

MaNV Hodem Ten Tuoi Luong Ma_chi_nhanh



Mức vật lý

```
struct NHANVIEN {
    int MaNV;
    int Ma_chi_nhanh;
    char Hodem[15];
    char Ten[15];
    struct date Ngay_sinh;
    float Luong;
    struct NHANVIEN next; /*Con trỏ đến bản ghi tiếp của tệp
NHANVIEN*/
}
index MaNV; /*Xác định các chỉ mục cho tệp NHANVIEN*/
index Ma_chi_nhanh;
```

2.1.4. Lược đồ và thể hiện của CSDL (tt)

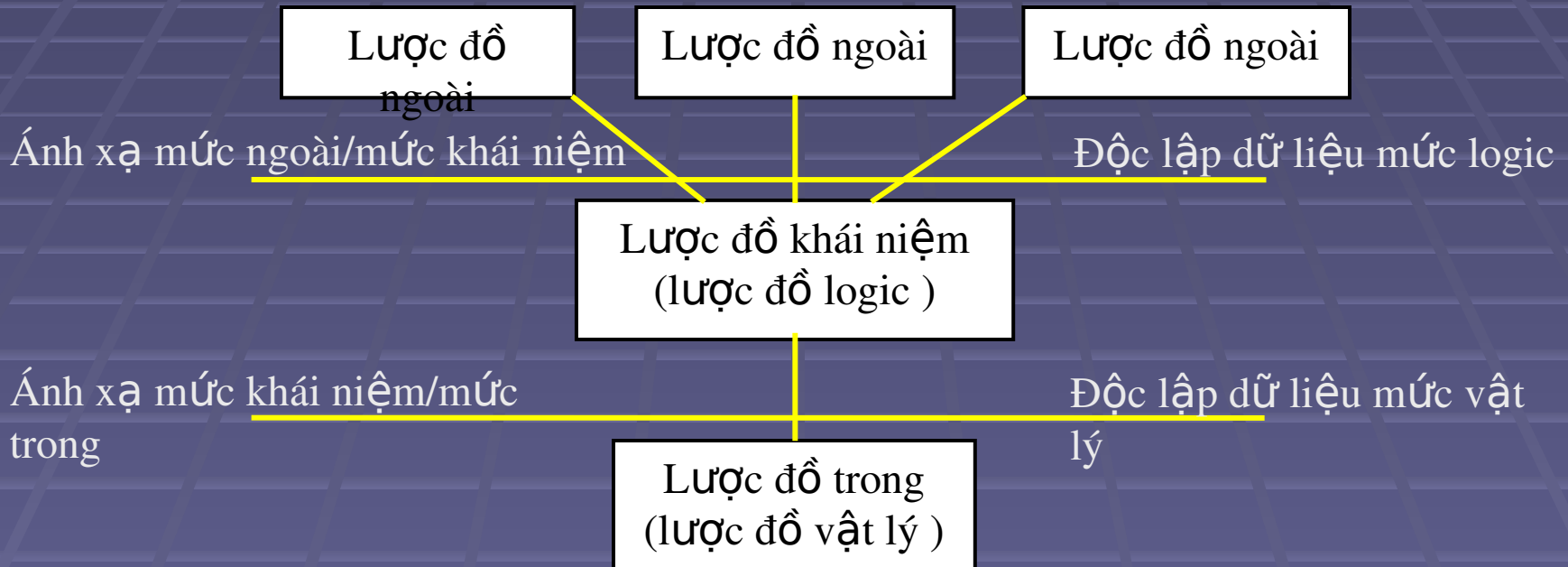
- Toàn bộ dữ liệu trong CSDL tại một thời điểm nhất định được gọi là một thể hiện của CSDL (database instance)

TenKH	SoCMT	NoiO	TaiKhoan
Trần Văn Ban	031803491	Hà Nội	A-101
Nguyễn Thị Giao	044803581	Hải Phòng	A-215
Hoàng Thị Kim Dung	037120582	Hà Nội	A-102
Trần Thị Lan Anh	035671241	Bắc Ninh	A-305

2.1.5. Tính độc lập dữ liệu

- **Độc lập dữ liệu** được hiểu theo nghĩa các lược đồ ở mức trên không bị ảnh hưởng khi có sự thay đổi các lược đồ ở các mức dưới
- Có 2 loại:
 - **Độc lập dữ liệu mức vật lý**: Là khả năng sửa đổi lược đồ vật lý mà không thay đổi lược đồ logic, như vậy không đòi hỏi viết lại các trình ứng dụng
 - **Độc lập dữ liệu mức logic**: Là khả năng sửa đổi lược đồ logic mà không làm thay đổi các lược đồ ngoài (các khung nhìn), như vậy không đòi hỏi viết lại các trình ứng dụng

2.1.5. Tính độc lập dữ liệu (tt)



2.2. Các ngôn ngữ cơ sở dữ liệu

- Một hệ CSDL cung cấp hai kiểu ngôn ngữ khác nhau:
 - Một ngôn ngữ đặc tả sơ đồ dữ liệu, gọi là **ngôn ngữ định nghĩa dữ liệu (DDL - Data Definition Language)**
 - Một ngôn ngữ biểu diễn các truy vấn và cập nhật CSDL, gọi là **ngôn ngữ thao tác dữ liệu (DML - Data Manipulation Language)**

2.2.1. Ngôn ngữ định nghĩa dữ liệu (DDL)

- Ngôn ngữ định nghĩa dữ liệu cho phép khai báo, hiệu chỉnh cấu trúc CSDL, mô tả các mối quan hệ của dữ liệu, các quy tắc áp đặt lên dữ liệu
- Kết quả biên dịch các lệnh của DDL là tập hợp các bảng được lưu trữ trong một tập tin đặc biệt được gọi từ điển dữ liệu hay thư mục dữ liệu
- CREATE (tạo), ALTER(sửa) và DROP

2.2.1. Ngôn ngữ định nghĩa dữ liệu (DDL)

- Định nghĩa miền:

```
CREATE DOMAIN < tên miền > < Type >
```

```
Ví dụ: CREATE DOMAIN hoten char( 30 );
```

- Tạo bảng

```
CREATE TABLE < tên bảng > (
```

```
< Thuộc tính 1 > < miền giá trị thuộc tính 1 > ,
```

```
...
```

```
< Thuộc tính n > < miền giá trị thuộc tính n > ,
```

```
< ràng buộc toàn vẹn 1 > ,
```

```
...
```

```
< ràng buộc toàn vẹn k > )
```

Ví dụ:

```
CREATE TABLE customer (  
customer_name CHAR( 20 ) not null,  
customer_street CHAR( 30 ),  
customer_city CHAR( 30 ),  
PRIMARY KEY( customer_name) );
```

2.2.1. Ngôn ngữ định nghĩa dữ liệu (DDL)

- Xoá bảng

DROP TABLE < tên bảng >

- Thêm thuộc tính vào bảng

ALTER TABLE < tên bảng > ADD < thuộc tính > < miền giá trị >

- Xoá bỏ một thuộc tính khỏi bảng

ALTER TABLE < Tên bảng > DROP < tên thuộc tính >

2.2.2. Ngôn ngữ thao tác dữ liệu DML

- Ngôn ngữ thao tác dữ liệu cho phép người dùng thực hiện các thao tác trên dữ liệu như tìm kiếm, chèn, sửa đổi, xoá bỏ thông tin
- INSERT, UPDATE và DELETE
- Có hai kiểu ngôn ngữ thao tác dữ liệu:
 - **DML thủ tục (procedural DML):** Yêu cầu NSD phải xác định dữ liệu nào họ đang cần và xác định cách thức để có được dữ liệu đó
 - **DML phi thủ tục (Nonprocedural DML):** Yêu cầu NSD xác định dữ liệu nào họ đang cần, chứ không yêu cầu NSD xác định cách thức để có dữ liệu đó.
 - SQL (Structured Query Language)

2.3. Mô hình dữ liệu

- Mô hình dữ liệu là một tập các khái niệm và ký pháp dùng để mô tả dữ liệu, các mối quan hệ của dữ liệu, và ràng buộc trên dữ liệu của một tổ chức
- Như vậy, có thể xem một mô hình dữ liệu có ba thành phần:
 - Phần mô tả cấu trúc của CSDL
 - Phần mô tả các thao tác, định nghĩa các phép toán được phép trên dữ liệu
 - Phần mô tả các ràng buộc toàn vẹn để đảm bảo sự chính xác của dữ liệu

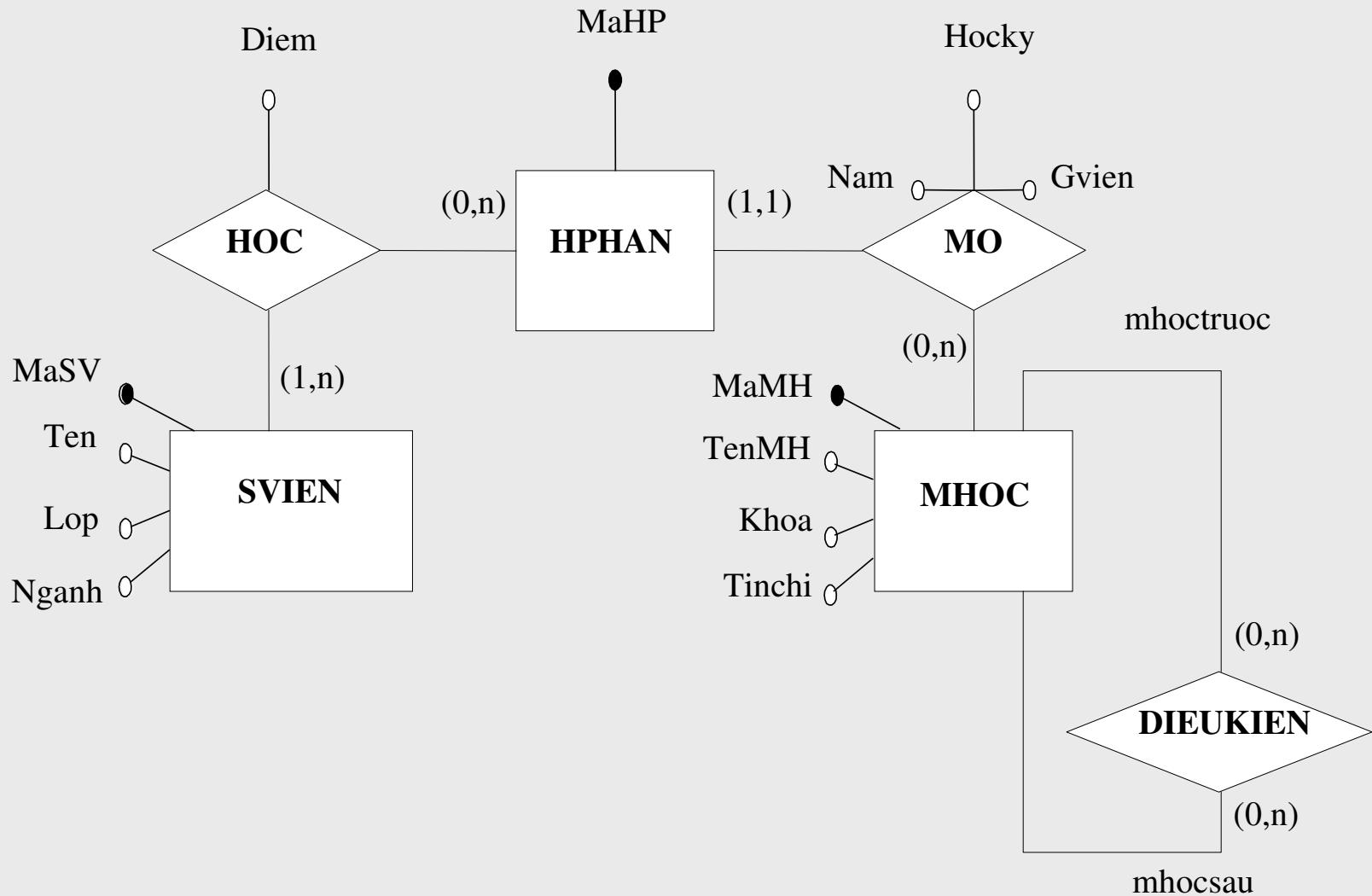
2.3. Mô hình dữ liệu

- Mô hình dữ liệu được chia thành các nhóm sau:
 - Mô hình logic trên cơ sở đối tượng:
 - Mô hình thực thể kết hợp
 - Mô hình hướng đối tượng
 - Mô hình ngữ nghĩa
 - Mô hình dữ liệu chức năng
 - Mô hình logic trên cơ sở bản ghi:
 - Mô hình quan hệ
 - Mô hình mạng
 - Mô hình phân cấp

2.3.1. Mô hình thực thể kết hợp

- Mô hình thực thể kết hợp được xây dựng dựa trên nhận thức rằng thế giới thực mà chúng ta muốn phản ánh là một tập hợp các đối tượng cơ sở và các mối quan hệ giữa chúng
- Dùng các khái niệm “thực thể” (entity) và “mối quan hệ” (relationship)

2.3.1. Mô hình thực thể kết hợp (tt)



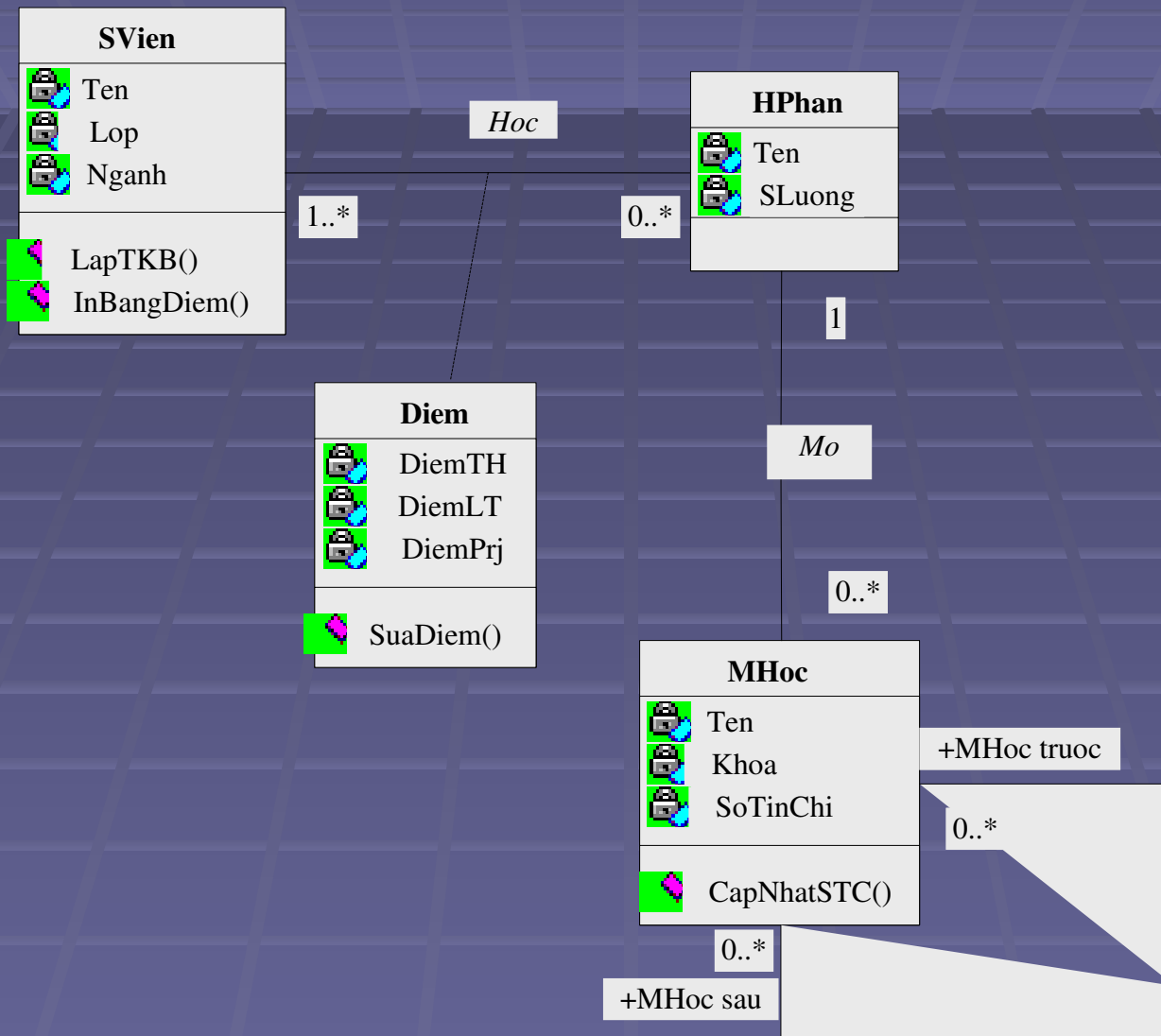
2.3.1. Mô hình thực thể kết hợp (tt)

- Thực thể SVIEN (sinh viên): gồm thuộc tính khóa là MaSV (mã sinh viên), các thuộc tính Ten (tên sinh viên), Lop (lớp), Ngành (ngành)
- Thực thể HPHAN (học phần): gồm thuộc tính khóa là MaHP (mã học phần)
- Thực thể MHOC (môn học): gồm thuộc tính khóa là MaMH (mã môn học), các thuộc tính TenMH (tên môn học), Khoa (Khoa), Tinchi (tín chỉ)
- Giữa thực thể SVIEN và HPHAN có quan hệ HOC (học), quan hệ này có thuộc tính Diem (điểm của sinh viên khi học học phần đó)
- Giữa thực thể HPHAN và MHOC có quan hệ MO (mở), quan hệ này có các thuộc tính Nam (năm), HKy (học kỳ), Gvien (giáo viên)
- Ngoài ra thực thể MHOC còn có quan hệ đệ quy DIEUKIEN (điều kiện) cho biết môn nào học trước, môn nào học sau

2.3.2. Mô hình hướng đối tượng

- Mô hình hướng đối tượng dựa trên cơ sở các đối tượng
- Một đối tượng chứa các thuộc tính được lưu trữ trong các biến thể hiện ở bên trong đối tượng
- Một đối tượng còn chứa các phần mã thao tác trên đối tượng, phần mã đó gọi là **phương thức**
- Các đối tượng chứa cùng các kiểu thuộc tính và cũng các **phương thức** như nhau được nhóm thành các lớp
- Chỉ có một cách theo đó một đối tượng có thể truy cập dữ liệu của một đối tượng khác là gọi tới phương thức của đối tượng khác đó (gửi một thông báo tới đối tượng)

2.3.2. Mô hình hướng đối tượng (tt)



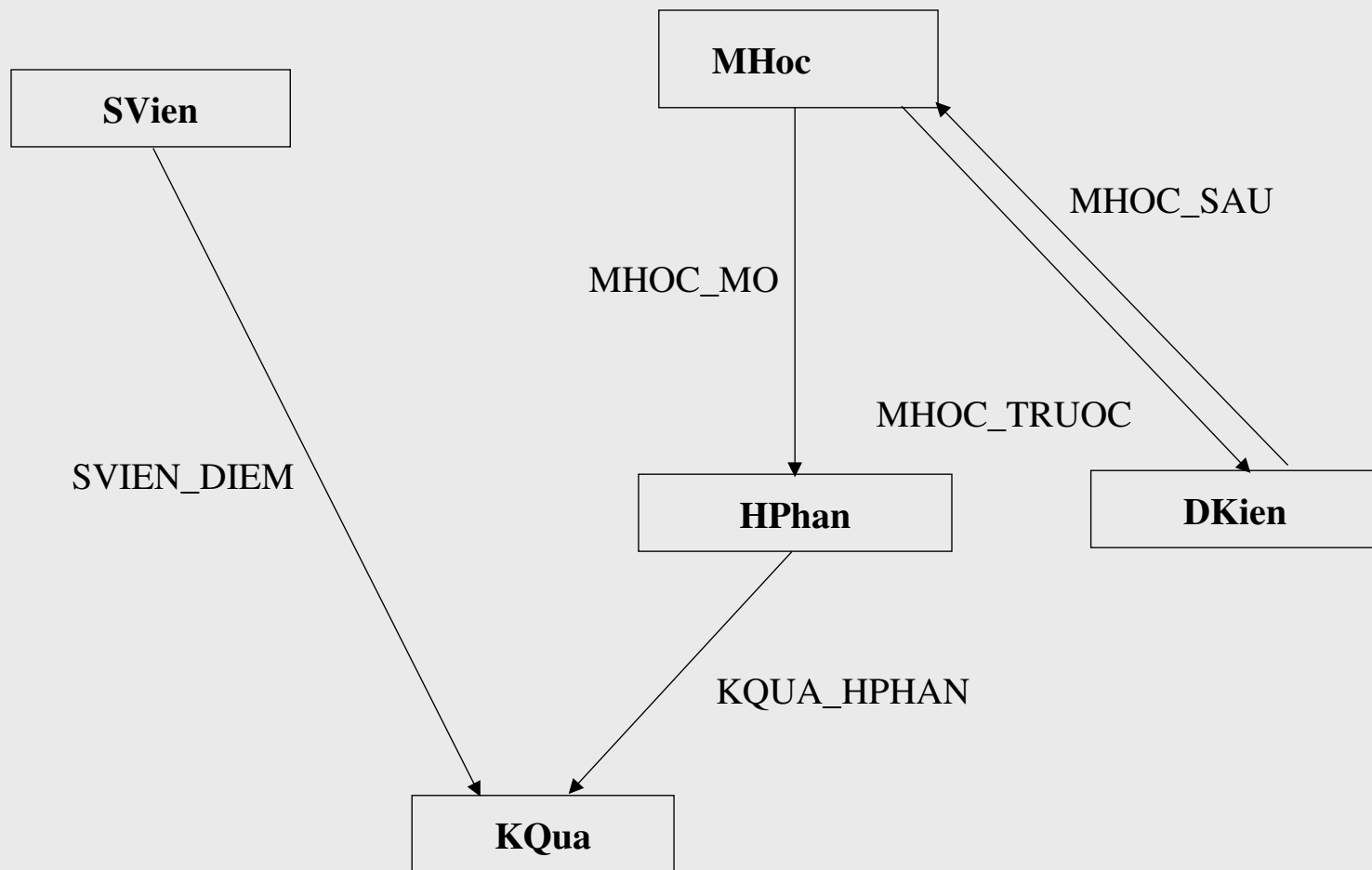
2.3.2. Mô hình hướng đối tượng (tt)

- Đối tượng SVien (sinh viên): gồm các biến Ten (tên sinh viên), Lop (lớp), Nganh (ngành) và các phương thức LapTKB (lập thời khóa biểu), InBangDiem (in bảng điểm)
- Đối tượng HPhan (học phần) gồm các biến Ten (tên học phần), SLuong (số lượng)
- Đối tượng Diem (điểm) gồm các biến DiemTH (điểm thực hành), DiemLT (điểm lý thuyết), DiemPrj (điểm đồ án), và phương thức SuaDiem (sửa điểm)
- Đối tượng MHọc gồm các biến Ten (tên), Khoa, SoTinChi (số tín chỉ) và phương thức CapNhatSTC (cập nhật số tín chỉ)

2.3.3. Mô hình mạng

- Trong mô hình mạng, dữ liệu được biểu diễn bởi một tập các bản ghi, còn các mối quan hệ được biểu diễn bởi các mối nối có thể xem như những con trỏ (giống như đồ thị)
- Xuất phát từ một đối tượng (biểu diễn bằng một bản ghi) có thể có nhiều mối quan hệ đến những đối tượng khác nhau. Trong những quan hệ đó luôn phân biệt đối tượng là chủ của quan hệ và những đối tượng thành phần của quan hệ

2.3.3. Mô hình mạng (tt)



2.3.4. Mô hình phân cấp

- Dữ liệu biểu diễn bằng tập các bản ghi, và mối quan hệ giữa dữ liệu biểu diễn bằng các mối nối như các con trỏ
- Mối quan hệ giữa hai đối tượng trong mô hình phân cấp thể hiện theo kiểu cha-con
- Sơ đồ các bản ghi cùng các quan hệ giữa chúng có cấu trúc như các cây chứ không phải các đồ thị

2.3.4. Mô hình phân cấp (tt)

Mức 1:

KQua

DiemTH

DiemLT

Mức 2:

HPhan

SVien

TenHP

SLuong

TenSV

Lop

Nganh

Mức 3:

MHoc

TenMH

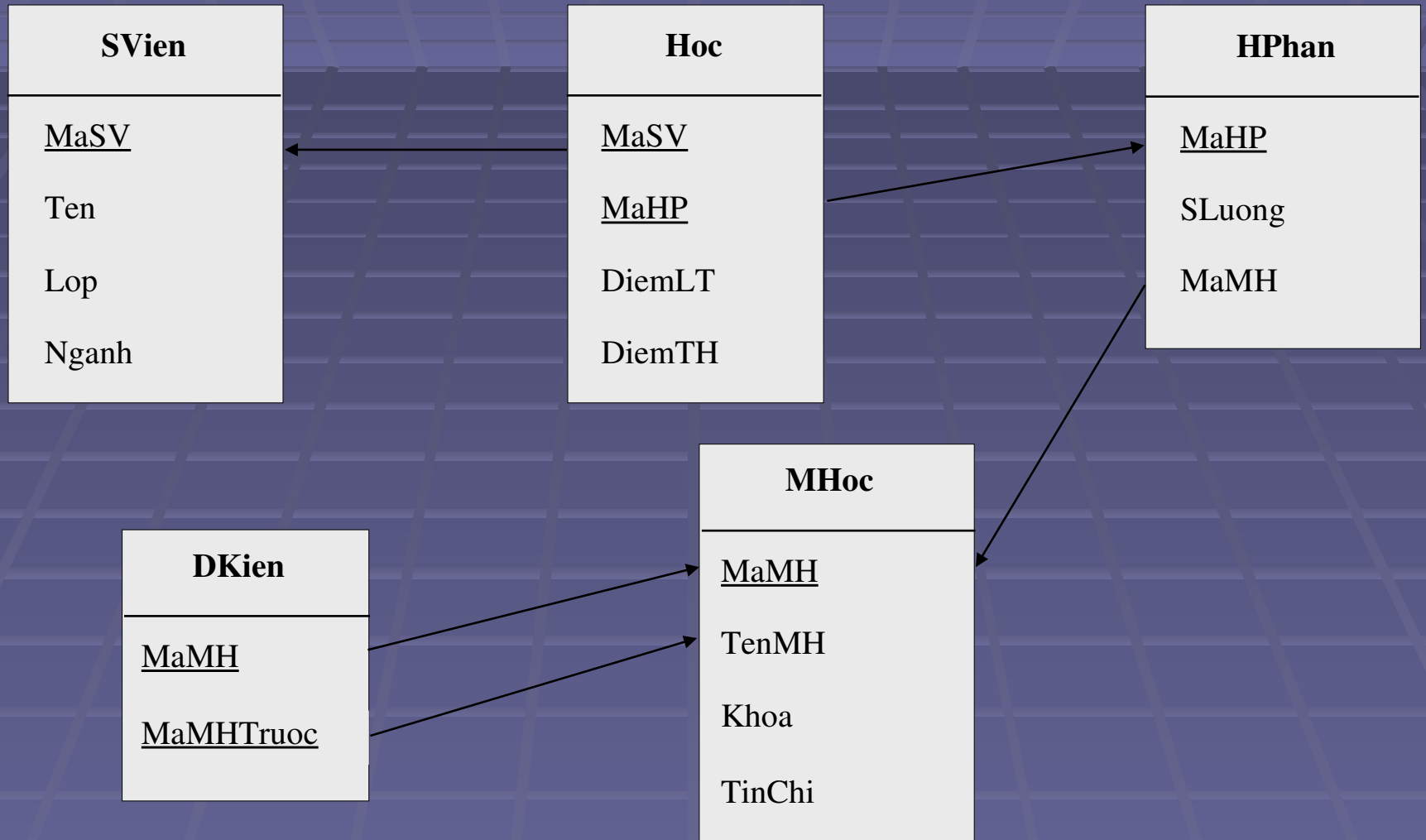
Khoa

TinChi

2.3.5. Mô hình quan hệ

- Trong mô hình quan hệ, dữ liệu được thể hiện trong các bảng
- Mỗi bảng gồm các dòng (thường gọi là bản ghi hay bộ) và cột (thường gọi là trường)

2.3.5. Mô hình quan hệ (tt)



2.3.5. Mô hình quan hệ (tt)

Cột

MaSV	Ten	Lop	Nganh
T001	Trần Thị Lan	03T1	CNTT
T002	Nguyễn Văn Dũng	03T1	CNTT
T003	Lý Hải	03T1	CNTT

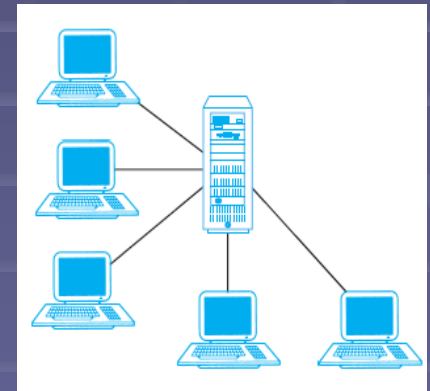
Dòng

2.4. Hệ quản trị CSDL đa người dùng

- Một hệ xử lý từ xa bao gồm một máy tính và một số trạm đầu cuối
- Tất cả các xử lý đều được thực hiện trên cùng một máy tính
- Các trạm đầu cuối đều được nối với máy tính trung tâm bằng dây cáp
- Trạm đầu cuối gửi các thông điệp yêu cầu đến chương trình ứng dụng của người dùng (khi sử dụng các dịch vụ của hệ quản trị CSDL), nhờ vào hệ thống điều khiển truyền thông của hệ điều hành
- Các thông điệp gửi trả về cho trạm đầu cuối của người dùng cũng theo con đường đó

2.4. Hệ quản trị CSDL đa người dùng

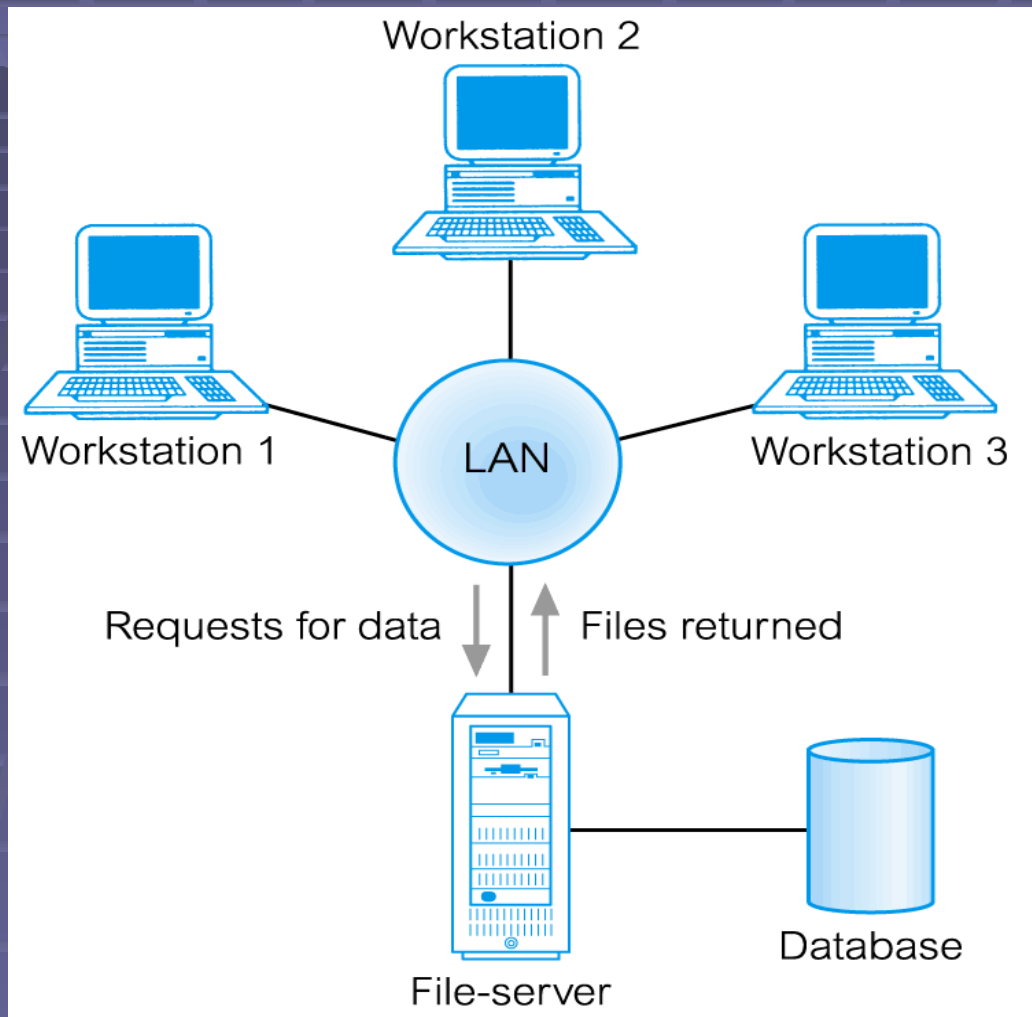
- Trong những năm qua con người đã đạt được những thành công lớn trong việc nâng cao khả năng của máy tính cá nhân và phát triển công nghệ mạng máy tính
- Nảy sinh hai kiến trúc mới cho hệ thống đa người dùng
 - Kiến trúc máy chủ - tập
 - Kiến trúc máy khách - chủ



2.4.1. Kiến trúc máy chủ - tệp (File - Server)

- Trong kiến trúc máy chủ - tệp, các xử lý không tập trung vào một máy tính trung tâm mà được phân tán trên mạng, thường là mạng cục bộ
- Máy chủ (File-Server) lưu giữ các tập tin dữ liệu (database). Máy chủ hoạt động đơn giản như một đĩa cứng chứa dữ liệu có thể chia sẻ
- Các ứng dụng và các hệ quản trị CSDL chạy trên mỗi trạm làm việc (workstation) yêu cầu các tập tin dữ liệu ở máy chủ khi cần.

2.4.1. Kiến trúc máy chủ - tệp (tt)



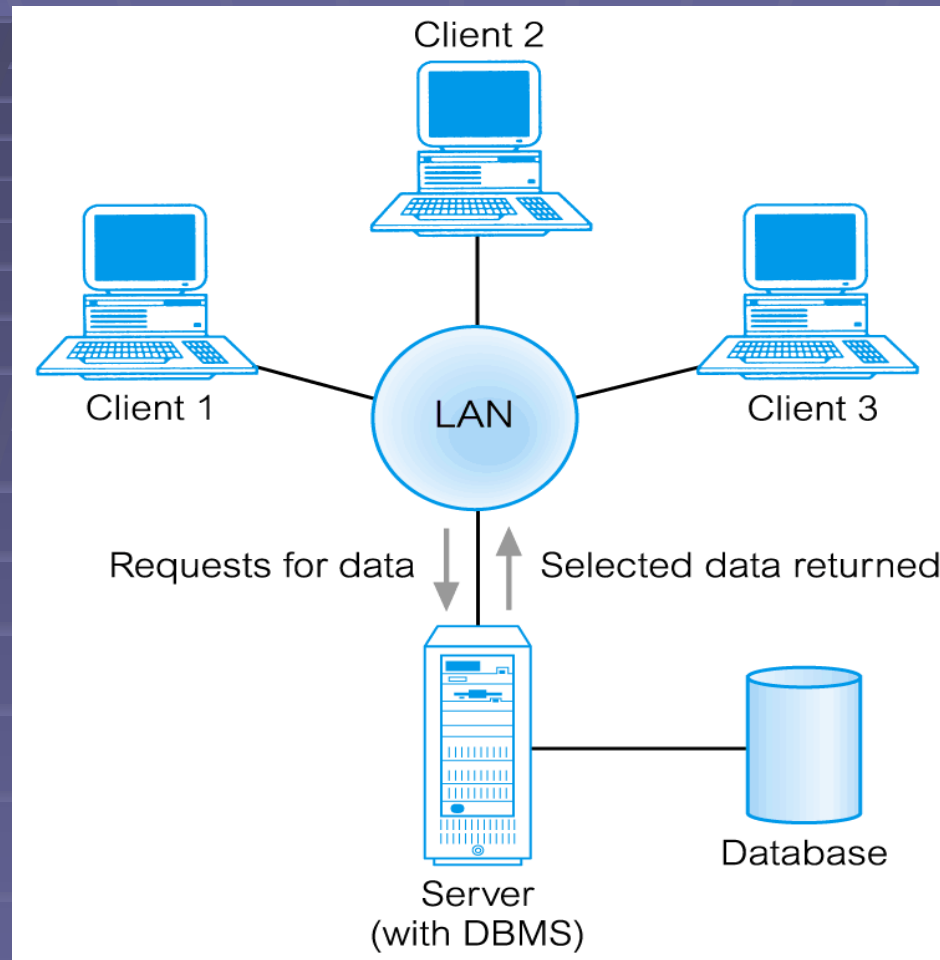
2.4.1. Kiến trúc máy chủ - tập (tt)

- Kiến trúc này có những nhược điểm chính sau đây:
 - Lượng dữ liệu truyền qua lại trên mạng rất nhiều
 - Mỗi trạm làm việc phải cài đặt một bản sao đầy đủ của hệ quản trị CSDL
 - Việc giải quyết các vấn đề tương tranh, khôi phục dữ liệu và bảo đảm tính nhất quán của dữ liệu sẽ phức tạp hơn do có nhiều hệ quản trị CSDL truy cập vào cùng các tập tin dữ liệu

2.4.2. Kiến trúc máy khách - chủ (Client - Server)

- Trong kiến trúc máy khách - chủ, các bộ phận phần mềm tương tác với nhau tạo nên hệ thống:
 - Tiến trình máy khách (client) yêu cầu cung cấp tài nguyên nào đó và tiến trình máy chủ (server) cung cấp tài nguyên đó
 - Thông thường tiến trình máy chủ và tiến trình máy khách đặt tại các điểm khác của mạng

2.4.2. Kiến trúc máy khách - chủ (tt)



2.4.2. Kiến trúc máy khách - chủ (tt)

- Một số ưu điểm
 - Khả năng truy cập rộng rãi đến các CSDL
 - Nâng cao khả năng thực hiện: Nếu tiến trình máy chủ và các máy khách làm việc ở trên các máy tính khác nhau thì các CPU khác nhau có thể cùng chạy song song, mỗi CPU thực hiện tiến trình của nó
 - Chi phí cho phần cứng có thể được giảm do chỉ cần máy chủ có cấu hình đủ mạnh để lưu trữ và quản trị CSDL

2.4.2. Kiến trúc máy khách - chủ (tt)

- Một số ưu điểm (tt)
 - Chi phí cho truyền thông được giảm do một phần trong các thao tác của ứng dụng được giải quyết trên máy khách, truyền thông trên mạng chỉ bao gồm: yêu cầu về truy cập CSDL của máy khách gửi đến máy chủ và dữ liệu kết quả từ máy chủ gửi cho máy khách
 - Nâng cao được khả năng đảm bảo tính nhất quán của dữ liệu. Máy chủ có thể kiểm soát được tính toàn vẹn bởi các ràng buộc này được định nghĩa và kiểm tra chỉ tại đó

2.5. Từ điển dữ liệu

- Trong một hệ quản trị CSDL, một từ điển dữ liệu là một tập hợp chỉ đọc (read only) của các bảng (table) và khung nhìn (view)
- Thông thường một hệ thống từ điển dữ liệu chứa các thông tin:
 - Tên, kiểu, kích thước các bản ghi
 - Tên của các mối quan hệ, các ràng buộc toàn vẹn trên dữ liệu
 - Tên những người có quyền truy cập vào CSDL
 - Các lược đồ trong, lược đồ khái niệm, lược đồ ngoài và các ánh xạ giữa chúng

2.5. Từ điển dữ liệu

- Microsoft Access: các bảng có tên bắt đầu bằng chữ MSys, chẳng hạn:
 - MSysACEs
 - MSysColumn
 - MSysIMEXColumn
 - MSysIMEXSpecs
 - MSysIndexes
 - MSysMacros
 - MSysObjects
 - MSysQueries
 - MSysRelationships

Tools / Options / View tabs / System Objects