**Tổng quan về cơ sở dữ liệu**

1. **Các khái niệm cơ bản**

* Dữ liệu: tập hợp những thông tin mà ta muốn lưu trữ.

+ Dữ liệu thường là các thông tin có liên quan nhau.

+ Dữ liệu được mô tả dưới nhiều dạng khác nhau (các ký tự, ký số, hình ảnh, ký hiệu, âm thanh…). Mỗi cách mô tả gắn với một ngữ nghĩa nào đó.

+ Dữ liệu về đối tượng có thể khác nhau, tùy thuộc vào ngữ cảnh.

VD: dữ liệu về sinh viên trong trường học thì bao gồm dữ liệu về sinh viên, giảng viên, khóa học, môn học, điểm thi, kết quả học tập …

VD: Dữ liệu sinh viên trong hệ thống quản lý sinh viên có thể khác sinh viên trong hệ thống quản lý thư viện.

+ Dữ liệu cũng có thể là thông tin về thực thể thật ( vật lý ) trong thực tế mà nó cũng có thể là các thông tin mô tả trừu tượng đại diện cho mô hình thực thể

* Cơ sở dữ liệu (DATABASE) (CSDL)

+ CSDL là tập hợp dữ liệu được tổ chức có cấu trúc liên quan với nhau và được lưu trữ trong máy tính.

+ CSDL được thiết kế, xây dựng cho phép người dùng lưu trữ dữ liệu, truy xuất thông tin hoặc cập nhật dữ liệu.

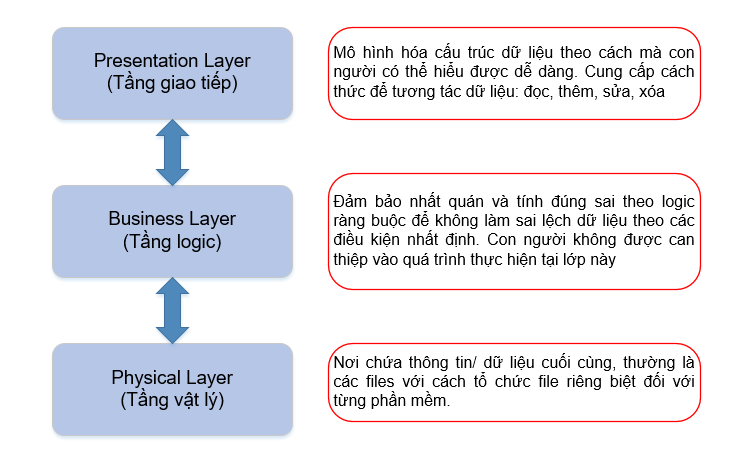
+ CSDL được tổ chức có cấu trúc: Các dữ liệu được lưu trữ có cấu trúc thành các bản ghi (record), các trường dữ liệu (field). Các dữ liệu lưu trữ có mối quan hệ (relation) với nhau

+ CSDL được cấu trúc để dễ dàng truy cập, quản lý và cập nhật.

Vd: 1 website có 1 cơ sở dữ liệu, nhưng được nhiều người dùng, nhiều vai trò cùng thực thi tới.



* Kiến trúc cơ bản của CSDL thông thường: thường gồm 3 lớp



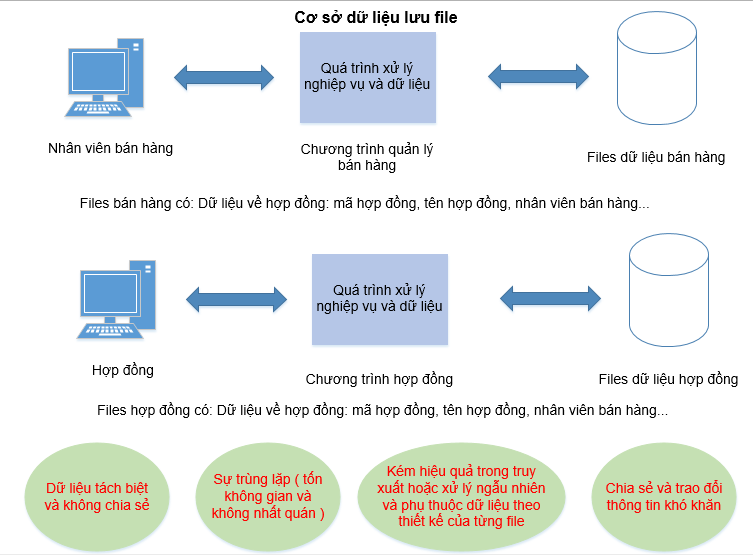
1. **Quản lý dữ liệu**

Quản lý dữ liệu là quản lý một số lượng lớn dữ liệu, bao gồm cả việc lưu trữ và cung cấp cơ chế cho phép đọc (truy vấn), thêm, sửa, xóa dữ liệu.

Hai phương pháp quản lý dữ liệu: Hệ thống quản lý bằng file và Hệ thống quản lý bằng CSDL

1. Hệ thống quản lý bằng file

* Dữ liệu được lưu trữ trong các file riêng biệt. Ví dụ: các chương trình lưu trữ thông tin bằng hệ thống các file dạng text.



1. Hệ thống quản lý bằng CSDL

Quản lý dữ liệu bằng CSDL giúp dữ liệu được lưu trữ một cách hiệu quả và có tổ chức, cho phép quản lý dữ liệu nhanh chóng và hiệu quả.

Đặc điểm của dữ liệu lưu trữ tại 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

* Nên cần phải có Hệ quản trị CSDL ( Database Management System – DBMS )

1. Hệ quản trị cơ sở dữ liệu DBMS:

* Là 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

…

Chú ý: Nói tới CSDL -> nói tới dữ liệu lưu trữ, còn hệ quản trị cơ sở dữ liệu -> phần mềm công cụ.

VD: Microsoft Access, SQL Server (Microsoft), Oracle (oracle), DB2( IBM), MySQL, MariaDB…

* DBMS thường có:

+ Ngôn ngữ mô tả, định nghĩa dữ liệu – DDL.

+ Ngôn ngữ thao tác dữ liệu – DML.

+ Ngôn ngữ truy vấn dữ liệu có cấu trúc – SQL Có cơ chế an toàn, bảo mật cao

* Các tính năng của hệ quản trị CSDL:

+ Khả năng quản lý dữ liệu tồn tại lâu dài: đặc điểm này chỉ ra rằng có một cơ sở dữ liệu tồn tại trong một thời gian dài, nội dung của cơ sở dữ liệu này là các dữ liệu mà hệ quản trị CSDL truy nhập và quản lý.

+ Khả năng truy nhập các khối lượng dữ liệu lớn một cách hiệu quả.

+ Hỗ trợ ít nhất một mô hình dữ liệu hay một sự trừu tượng toán học mà qua đó người sử dụng có thể quan sát dữ liệu.

+ Ðảm bảo tính độc lập dữ liệu hay sự bất biến của chương trình ứng dụng đối với các thay đổi về cấu trúc trong mô hình dữ liệu.

+ Hỗ trợ các ngôn ngữ cao cấp nhất định cho phép người sử dụng định nghĩa cấu trúc dữ liệu, truy nhập dữ liệu và thao tác dữ liệu.

+ Quản lý giao dịch, có nghĩa là khả năng cung cấp các truy nhập đồng thời, đúng đắn đối với CSDL từ nhiều người sử dụng tại cùng một thời điểm.

+ Ðiều khiển truy nhập, có nghĩa là khả năng hạn chế truy nhập đến các dữ liệu bởi những người sử dụng không được cấp phép và khảnăng kiểm tra tính đúng đắn của CSDL.

+ Phục hồi dữ liệu, có nghĩa là có khả năng phục hồi dữ liệu, không làm mất mát dữ liệu với các lỗi hệ thống.

1. Học CSDL như thế nào?

* Có nhiều góc nhìn để làm việc với dữ liệu:

+ Người sử dụng phần mềm/ nghiệp vụ: quan tâm tới dữ liệu hiển thị và thực thi kết quả thay đổi

+ Người lập trình: quan tâm tới thao tác và thực thi trên dữ liệu ( tạo, cập nhật, xóa, truy vấn ) và cấu trúc lưu trữ

+ Người vận hành, bảo trì DBA: quan tâm tới quản trị CSDL

+ Người phát triển phần mềm DBMS: quan tâm tới việc phát triển chương trình DBMS

* Nên bắt đầu học từ trên xuống dưới ( càng xuống tới các level dưới thì cần phải có khả năng và hiểu biết rộng )

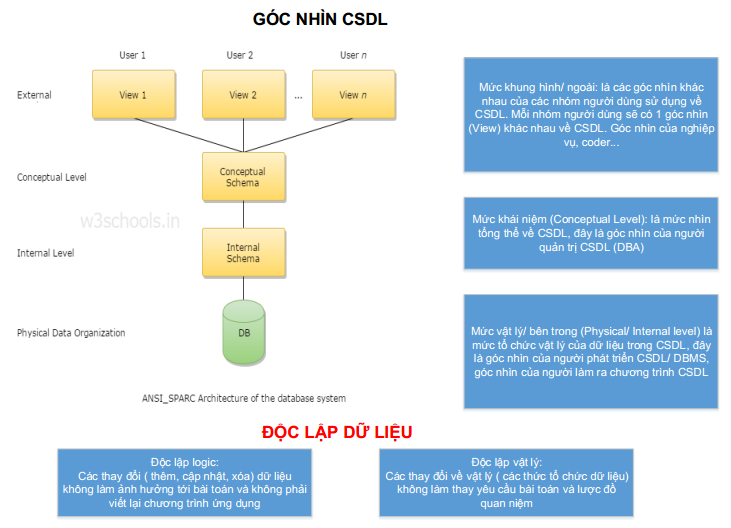
+ Tìm hiểu dữ liệu gồm những gì ( phân tích )

+ Thiết kế dữ liệu

+ Lập trình

+ DBA

+ ….



1. **Hệ quản trị cơ sở dữ liệu quan hệ (Relational Database Management System RDBMS)**

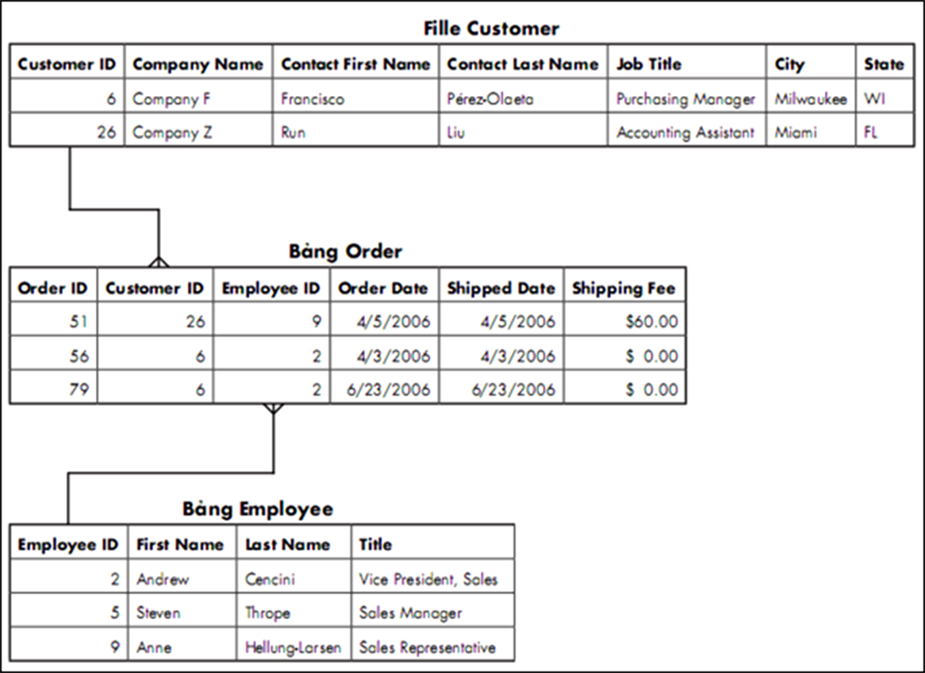
Giữa dữ liệu có mối quan hệ với nhau về mặt logic thể hiện các khía cạnh của bài toán thực tế. Dữ liệu được biểu diễn dưới dạng bảng với các hàng và các cột:

+ CSDL là tập hợp các bảng (còn gọi là quan hệ).

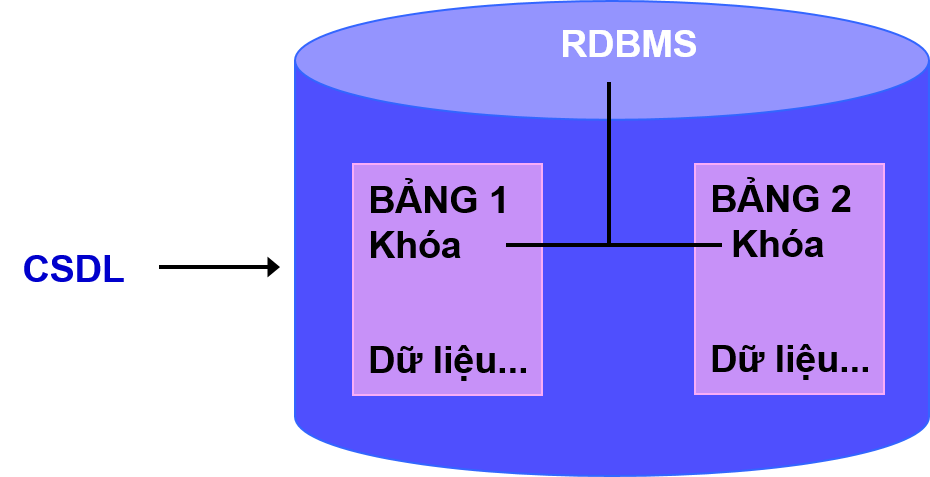
+ Mỗi hàng là một bản ghi (record), còn được gọi là bộ (tuple).

+ Mỗi cột là một thuộc tính, còn được gọi là trường (field)

VD:



Hệ quản trị cơ sở dữ liệu quan hệ (Relation Database Management System - RDBMS) là một hệ quản trị cơ sở dữ liệu theo mô hình quan hệ. Nên nó cũng cung cấp phần mềm, công cụ lưu trữ và truy cập và quản lý dữ liệu.



1. **Sơ đồ đồ thực thể/ dữ liệu quan hệ (Entity Relationship Diagram – ERD)**
2. Khái niệm

**Mô hình thực thể quan hệ** được sử dụng để biểu diễn cơ sở dữ liệu ở mức khái niệm.

* *Entity Relationship* Diagram.

*+ Entity* nghĩa là các **thực thể**

*+ Relationship* là các **mối quan hệ**, (giữa các thực thể đó).

=> *ERD là một* ***sơ đồ****, thể hiện* ***các thực thể*** *có trong database, và* ***mối quan hệ giữa chúng*** *với nhau.*

- Tại sao phải tạo ERD:

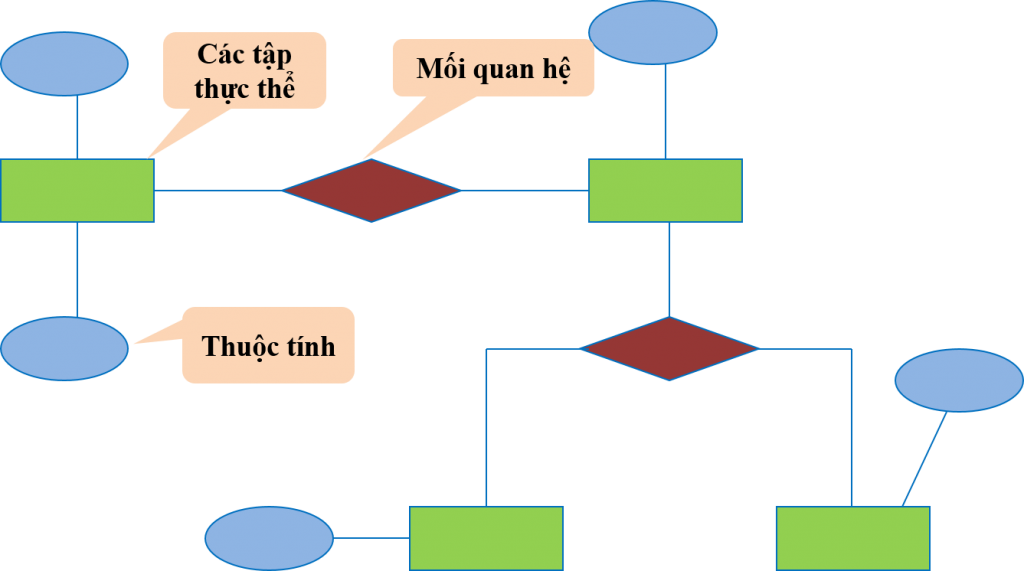
+ Giúp mường tượng tổng quan hệ thống có gì

+ Giúp phân tích hệ thống

+ Giúp xây dựng DB tốt

+ Tạo báo cáo tốt hơn

1. Các thành phần của ERD



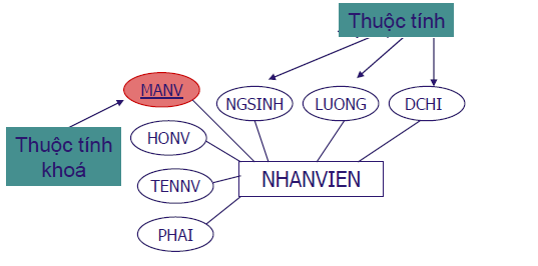
ERD có 3 thành phần chính:

* **Entity**: *thực thể (hoặc đối tượng hoặc dữ liệu) mà hệ thống quản lý.*

Thực thể là một đối tượng cụ thể hay trừu tượng trong thếgiới thực mà nó tồn tại và có thể phân biệt được với các đối tượng khác, được biểu diễn thực thể bởi hình chữ nhật và có danh sách các thuộc tính.

Chú ý: Entity luôn là danh từ.

VD: Entity Nhân viên:



* **Attribute**: *thuộc tính của các đối tượng.*

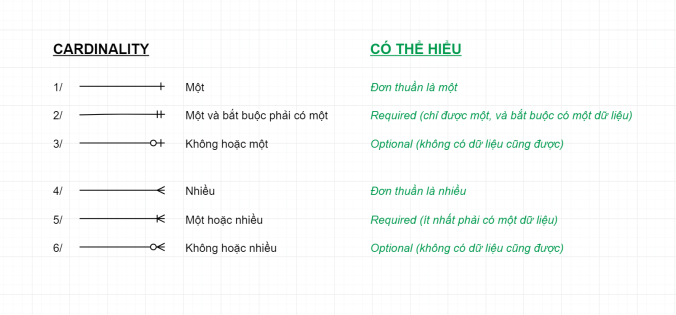
Thuộc tính là các **đặc tính** của một đối tượng, và là những thông tin **riêng biệt** của đối tượng đó.

VD: Thực thể là NHANVIEN có các thuộc tính: MANV ( PK – khóa chính ), HONV, TENNV, PHAI, NGSINH, LUONG, DCHI

* **Relationship**: *mối quan hệ giữa các đối tượng.*
* Thể hiện mối liên quan giữa hai hay nhiều thực thể. Mỗi liên kết có một tên gọi và thường dùng động từ.

VD: một Nhân viên Làm việc tại một Phòng ban nào đó và một Nhân viên có thể là Trưởng phòng của một Phòng ban.

* Các mối quan hệ và cách biểu diễn:

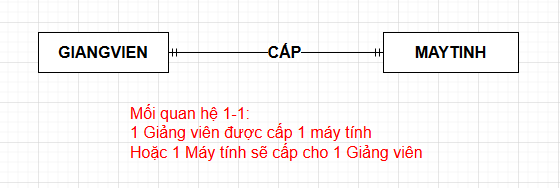


* Có 3 quan hệ chính:

+ One-to-One: quan hệ 1-1

Mỗi đối tượng của thực thể thứ nhất tương ứng với 1 và chỉ 1 đối tượng của thực thể thứ 2 và ngược lại.

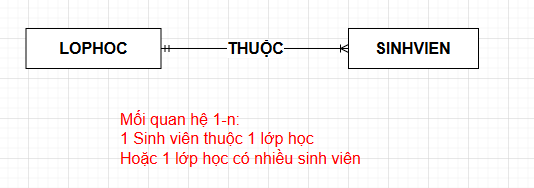
VD: mỗi giảng viên được cấp 1 và chỉ một máy tính. Một máy tính chỉ được cấp cho 1 và chỉ một giảng viên.



+ One-to-Many: quan hệ 1-nhiều hoặc 1 –N

Mỗi đối tượng của thực thể thứ nhất tương ứng với nhiều đối tượng của thực thể thứ 2 và ngược lại.

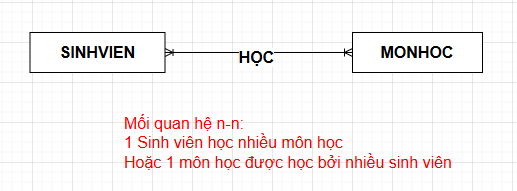
VD: mỗi sinh viên thuộc một và chỉ một lớp. Một lớp có nhiều sinh viên.



+ Many-to-Many: quan hệ nhiều-nhiều HOẶC N – N

Mỗi đối tượng của thực thể thứ nhất tương ứng với nhiều đối tượng của thực thể thứ 2 và ngược lại.

VD: **Mỗi sinh viên** học **một hoặc nhiều môn học. Một môn học** có **một hoặc nhiều sinh viên** học.



VD: Tìm các mối quan hệ sau:

|  |  |  |
| --- | --- | --- |
| Thực thể | Mối quan hệ | Diễn giải |
| Khách hàng – Đơn hàng | 1-n | 1 khách hàng có nhiều đơn hàng  1 đơn hàng chỉ thuộc 1 khách hàng |
| Người dùng – Đặt chỗ (Booking) | 1-n | 1 người dùng có thể đặt được nhiều chỗ  1 chỗ chỉ được đặt được nhiều người |
| Dịch vụ – Hợp đồng | n -n | 1 Dịch vụ nằm trong nhiều Hợp đồng 1 Hợp đồng có nhiều dịch vụ |
| Nhân viên CSKH – Khách hàng | n-n | 1 NV có thể chăm sóc nhiều KH  1 KH có thể được chăm sóc nhiều NV |
| Sinh Viên – Sách | 1-n | 1 SV có thể mượn được nhiều sách  1 sách(seri) có thể được mượn bởi 1 SV |
| Khách hàng – Hoạt động tương tác | n-n | 1KH có thể tham gia nhiều HĐTT  1 HĐTT được tham gia bởi nhiều KH |
| Trường học - Lớp học | 1-n | 1 TH có nhiều LH  1 LH chỉ thuộc 1 TH |
| Lớp học - Sinh viên | 1-n | 1 LH có nhiều SV  1 SV thuộc 1 LH |
| Khoa - Sinh viên |  |  |
| Trường học - Khoa | 1 - n |  |
| Sách - Thể loại | 1-n | 1 Sách có 1 Thể loại  1 Thể loại thì có nhiều sách |

Draw.io

* **Các bước tạo sơ đồ ERP**

+ **Bước 1.** Xác định thực thể/ đối tượng và thuộc tính của nó bao gồm thuộc tính khoá và thuộc tính đa trị

**+ Bước 2.** Xác định mối quan hệ có thể có giữa các thực thể

* Chú ý:

+ Thực thể là **DANH TỪ**

+ Thuộc tính ám chỉ **TÍNH TỪ**, chỉ những tính chất – đặc điểm của thực thể *(VD: khách hàng thì cao 175cm, nặng 65kg, giới tính Nam, địa chỉ nhà, email…)*

+ Còn mối quan hệ thì ám chỉ **ĐỘNG TỪ** (có, thuộc, đặt, chăm sóc, mượn, thực hiện…), nếu đọc ngược lại thì chuyển về thể bị động

1. **BÀI TẬP**

VẼ CÁC BIỂU ĐỒ ERD sử dụng online trên web draw.io

**Bài 1: Thiết kế cơ sở dữ liệu quản lý đào tạo**

Mỗi sinh viên gồm họ tên, ngày sinh, giới  tính, nơi sinh, số điện thoại và e-mail. Được cấp một mã số sinh viên duy nhất và mỗi sinh viên chỉ thuộc về một lớp.

SINHVIEN(HOTEN, NGAYSINH, GIOITINH, NOISINH, SODT, EMAIL, MASV, MALOP)

LOP(MALOP, TENLOP)

Mỗi lớp học có một mã số lớp duy nhất, tên lớp và mỗi lớp chỉ thuộc về một khoa.

LOP(MALOP, TENLOP, MAKHOA)

KHOA(TENKHOA, MAKHOA)

=>

SINHVIEN(HOTEN, NGAYSINH, GIOITINH, NOISINH, SODT, EMAIL, MASV, MALOP)

LOP(MALOP, TENLOP, MAKHOA)

KHOA(TENKHOA, MAKHOA)

~~SINHVIEN\_MONHOC( MASV, MAMON)~~

DIEMTHI (MASV, MAMON, LANTHI, DIEMTHI)

Mỗi khoa có một tên khoa và một mã số khoa duy nhất.

Mỗi  môn  học  có  tên môn học,  số giờ và  mã  số  môn học duy nhất.

Mỗi giảng viên cần quản lý các thông tin như họ và tên giảng viên, một chuyên ngành và được cấp một mã số  duy nhất. Mỗi giảng viên thuộc một khoa.

Mỗi  sinh  viên  với  một môn  học được phép  thi  tối  đa  2  lần,  mỗi  lần  thi, điểm thi.Mỗi môn học ở  mỗi lớp học chỉ  phân công cho một giảng viên dạy và một giảng viên thì có thể dạy nhiều môn ở một lớp.

=>MONHOC (MAMON, TENMON, SOGIO)

LOP(MALOP, TENLOP, MAKHOA)

GIANGVIEN(MAGV, TENGV, CHUYENNGANH, MAKHOA)

MONHOC\_LOP\_GIANGVIEN(MAMON, MALOP, MAGV, THOIGIAN, DIADIEM)

Sinhvien(masv, hoten, ngaysinh, gioitinh, sodt, email, malop)

Lop ( malop, tenlop, makhoa )

Khoa (makhoa, tenkhoa)

Monhoc (mamonhoc, tenmonhoc, sogio)

Giangvien (magv, tengv, chuyennganh, makhoa)

Sinhvien\_monhoc (masv, mamonhoc, lanthi, diemthi)

Dayhoc (Lop\_giangvien\_monhoc) (malop, mamonhoc, magv, thoigian)

**Bài 2: Thiết kế cơ sở dữ liệu quản lý siêu thị**

Mỗi khu vực có một mã số phân biệt, tên, chuyên bán 1 loại hàng.

Mỗi loại hàng có một mã số, có tên và bao gồm nhiều mặt hàng.

Mỗi mặt hàng có một mã số phân biệt, có tên, đơn vị tính, đơn giá bán hiện hành. Mỗi mặt hàng có thể cung cấp bởi nhiều nhà cung ứng.

Mỗi nhân viên chỉ làm việc tại một khu vực và có một mã số phân biệt, và các thông tin như: họ tên, ngày sinh ,địa chỉ, ngày vào làm việc.

Mỗi nhà cung cấp có một mã số phân biệt, tên công ty, địa chỉ, phone, fax. Lưu ý, mỗi nhà cung cấp có thể có nhiều số điện thoại.

Mỗi khách hàng khi mua hàng, sẽ có một hóa đơn gồm số hóa đơn, ngày lập hóa đơn, tên khách mua, địa chỉ, và danh sách các mặt hàng mua kèm theo đơn giá bán.

Bài 3: Thiết kế cơ sở dữ liệu hội thảo

Người ta cần phải tổ chức một cơ sở dữ liệu để chuẩn bị cho một hội thảo khoa học.

Hội thảo sẽ được chia thành các tiểu ban khác nhau ( thông tin về mỗi tiêu ban bao gồm: tên tiểu ban, tên người điều khiển và tên thư ký).

Các tác giả muốn tham gia báo cáo trong hội thảo phải gửi trước bản tóm tắt của báo cáo ( bao gồm tên và tóm tắt báo cáo ) đến một ban tổ chức ( BTC ).

Một tác giả có thể gửi nhiều báo cáo đến BTC và mỗi một báo cáo có thể là của nhiều tác giả.

Sau khi nhận được bản tóm tắt báo cáo, BTC phải tiến hành bố trí các báo cáo vào trong các tiểu ban ( mỗi một báo cáo chỉ được trình bày ở 1 tiểu ban nhất định )