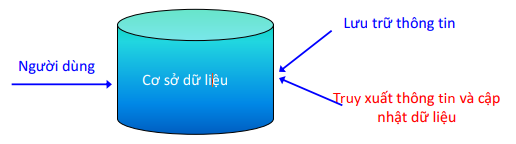
Cơ sở dữ liệu - COM2012

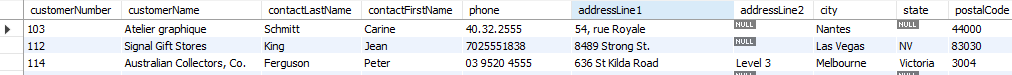
1. Giới thiệu về cơ sở dữ liệu
2. **Tại sao phải học môn cơ sở dữ liệu?** Bởi vì nó là một kỹ năng quan trọng và không thể thiếu trong lĩnh vực công nghệ thông tin. Giúp bạn phát triển ra các sản phẩm động có khả năng lưu trữ các thông tin và cuối cùng là khi đi ứng tuyển tại các vị trí công việc sau này.
3. **Lưu trữ dữ liệu dưới dạng hệ thống tệp tin?** Lưu trữ thông tin dưới dạng file và mỗi file sẽ đặt nó trong những folder khác nhau nhằm mục đích dễ dàng tìm đến.
   1. Ưu điểm:
      * Với việc lưu trữ này thì rõ ràng khi ta cần là lưu ngay nên tốc độ triển khai sẽ nhanh
      * Rõ ràng và trực quan với người không có chuyên môn công nghệ thông tin
   2. Khuyết điểm:
      * Dữ liệu không nhất quán, đôi khi có trường hợp nhiều người sử dụng một file và sẽ xảy ra hiện tượng lưu đè.
      * Trùng lặp dữ liệu nhiều
   3. Tính chia sẽ dữ liệu không cao (kém)
4. **Cơ sở dữ liệu là gì?** Cơ sở dữ liệu là một hệ thống các thông tin có cấu trúc, được lưu trữ trên các thiết bị lưu trữ nhằm thõa mãn yêu cầu khai thác thông tin đồng thời của nhiều người sử dụng hay nhiều chương trình ứng dụng chạy cùng một lúc với những mục đích khác nhau.
   1. Việc sử dụng hệ thống CSDL này sẽ khắc phục được những khuyết điểm của cách lưu trữ dươi dạng hệ thống tập tin:
      * Giảm trùng lặp thông tin ở mức thấp nhất, đảm bảo tính nhất quán và toàn vẹn dữ liệu
      * Đảm bảo dữ liệu được truy xuất theo nhiều cách khác nhau, từ nhiều người khác nhau và nhiều ứng dụng khác nhau.
      * Tăng khả năng chia sẻ thông tin. Ví dụ nếu ta đặt hệ thống dữ liệu online tại Server thì ở bất cứ đâu nếu có password logi vào thì ta hoàn toàn có thể vào hệ thống để truy vấn thông tin
   2. Tuy nhiên việc sử dụng hệ quản trị CSDL lại có những phiền hà không hề nhỏ sau đây:
      * Phải đảm bảo tính chủ quyền của dữ liệu, vì khi sử dụng có tính chất chia sẻ cao
      * Bảo mật quyền khai thác thông tin
      * Bảo đảm vấn đề tranh chấp dữ liệu khi xảy ra
      * Khi gặp các trục trặc sự cố thì phải bảo đảm vấn đề an toàn dữ liệu, không bị mất dữ liệu
5. **Các hệ cơ sở dữ liệu phổ thông hiện nay?** có rất nhiều hệ quản trị **CSDL** như **MYSQL, SQL SERVER, Oracle, MS Access.**
6. **Tại sao lại chọn hệ CSDL MySQL vào môn học này?** Mysql là một hệ quản trị **CSDL** được sử dụng rất rộng rãi trên thế giới, đặc biệt là các ứng dụng website và nó thường đi kết hợp với ngôn ngữ lập trình **PHP** để xây dựng các ứng dụng website. Các hệ thống web ưa chuộng **MYSQL** là tại vì tốc độ xử lý của nó cao, tính dễ sử dụng và thương thích với các hệ điều hành thông dụng hiện nay như **Linix**, **Window**, ... **Mysql** có sử dụng ngôn ngữ truy vấn T-**SQL** để thao tác dữ liệu.
7. **Hiện nay đang đình đám các hệ quản trị CSDL với tên gọi NO-SQL, nghĩa là hệ quản trị không sử dụng ngôn ngữ có cấu trúc truy vấn (T-SQL), vì thế trong tương lai không biết MYSQL liệu có bị tiêu diệt hay không?**

Khái niệm cơ bản về dữ liệu và cơ sở dữ liệu

1. **Dữ liệu là gì?** **Dữ liệu(Data)** là Là các thông tin của đối tượng (người, vật, một khái niệm, sự việc…) được lưu trữ trên máy tính. Có thể truy nhập vào dữ liệu để trích xuất thông tin. Dữ liệu được **mô tả dưới nhiều dạng khác nhau** (các ký tự, 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.
   1. **Ví dụ:** dữ liệu về đối tượng sinh viên có thể khác nhau tùy vào mục đích quản lý
      * Quản lý điểm: Tên, mã sinh viên, điểm môn 1, điểm môn 2, điểm môn 3
      * Quản lý nhân thân: Tên, địa chỉ, ngày sinh, quê quán, lớp
2. **CSDL**: 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.



1. **Trong cơ sở dữ liệu gồm có những gì?**
   1. **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.
   2. Khả năng truy xuất thông tin từ CSDL:
      * CSDL được cấu trúc để dễ dàng truy cập, quản lý và cập nhật.

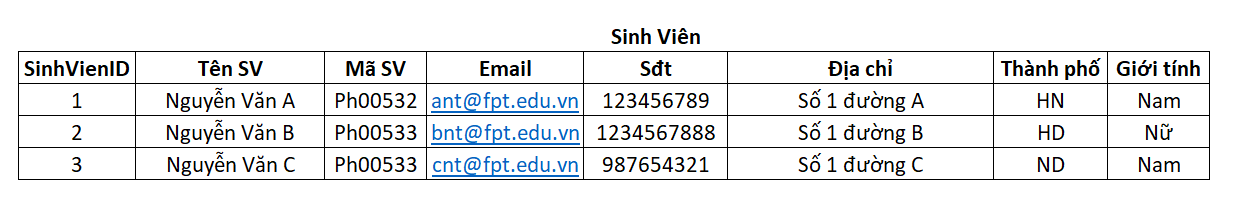


1. **Phương pháp trong quản lý dữ liệu DBMS và RDBMS**
   1. **RDBMS** là viết tắt của **Relational Database Management System** có nghĩa là hệ quản trị cơ sỡ **dữ liệu quan hệ**. RDBMS là cơ sở cho SQL, và cho tất cả các hệ thống cơ sở dữ liệu hiện đại như MS SQL Server, IBM DB2, Oracle, MySQL và Microsoft Access.
   2. **Database Management System (DBMS)** là hệ thống quản lý cơ sở dữ liệu, hỗ trợ phát triển, quản trị và sử dụng cơ sở dữ liệu. DBMS giống RDBMS là cũng được sử dụng để lưu trữ, quản lý cơ sở dữ liệu, nhưng về bản chất thì hai hệ thống này có rất nhiều điểm khác biệt. Ngoài chức năng giúp quản lý dữ liệu thì các hệ thống quản trị cơ sở dữ liệu có cung cấp cho người dùng chức năng kiểm soát quyền truy cập đọc / ghi, chỉ định tạo báo cáo và phân tích việc sử dụng.

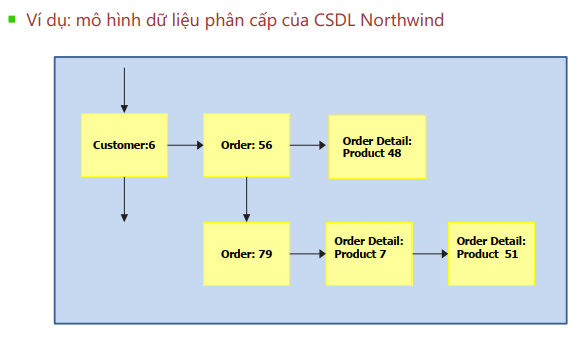
|  |  |  |
| --- | --- | --- |
| **STT** | **DBMS** | **RDBMS** |
| 1 | **DBMS lưu dữ liệu dưới dạng file** | **RDBMS lưu dữ liệu dưới dạng bảng** |
| 2 | Dữ liệu được lưu trữ ở dạng phân cấp hoặc điều hướng | Bảng có một định danh gọi là khóa chính và giá trị dữ liệu được lưu trữ dưới dạng các bảng |
| 3 | **Không có chuẩn hóa dữ liệu** | **Có chuẩn hóa dữ liệu** |
| 4 | Không có sự bảo vệ dữ liệu khi thực hiện các thao tác | Các ràng buộc toàn vẹn đảm bảo toàn vẹn dữ liệu khi thực hiện các thao tác dữ liệu như thêm, sửa, xóa,... |
| 5 | **Không có mỗi quan hệ giữa các bảng (do lưu dưới dạng file)** | **Có mối quan hệ giữa các bảng** |
| 6 | Cung cấp những phương thức thống nhất để truy cập vào thông tin được lưu trữ | Hỗ trợ một cấu trúc dạng bảng và quan hệ giữa chúng để truy cập vào thông tin lưu trữ |
| 7 | **Không hỗ trợ cơ sở dữ liệu phân tán** | **Hỗ trợ cơ sở dữ liệu phân tán** |
| 8 | Thường dùng trong tổ chức nhỏ, xử lý khối lượng dữ liệu nhỏ | Được thiết kế để xử lý lượng dữ liệu lớn, hỗ trợ nhiều người dùng |
| 9 | Ví dụ của DBMS là file system, xml,... | Ví dụ về RDBMS là MySQL, Oracle hay SQL Server,... |

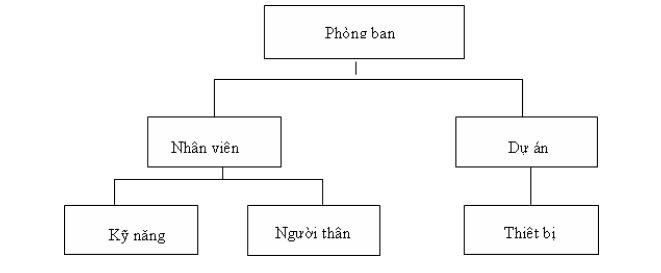
Trong RDBMS, dữ liệu được biểu diễn bởi các hàng. Relational Database là cơ sở dữ liệu được sử dụng phổ biến nhất. Nó chứa các bảng và mỗi bảng có **Primary Key** riêng. Bởi vì các bảng này được tổ chức chặt chẽ nên việc truy cập dữ liệu trở nên dễ dàng hơn trong RDBMS.

1. **Mô hình CSDL:**
   1. Các CSDL có thể khác nhau về chức năng và mô hình dữ liệu **(data model).**
   2. Mô hình dữ liệu sẽ quyết định cách thứ lưu trữ và truy cập dữ liệu.
   3. Tùy từng ngữ cảnh quan hệ giữa các thành phần dữ liệu trong CSDL, mô hình phức hợp được áp dụng để việc lưu trữ và truy xuất dữ liệu đạt hiệu quả cao nhất.
   4. **Các mô hình:** Mô hình dữ liệu file phẳng **(Flat file)** - Mô hình dữ liệu mạng **(Network model)**
      * Mô hình dữ liệu file phẳng **(Flat file): Chỉ dùng cho các CSDL đơn giản – CSDL dạng file phẳng thường là file kiểu văn bản chứa dữ liệu dạng bảng.**
      * Ví dụ về một bảng sinh viên:

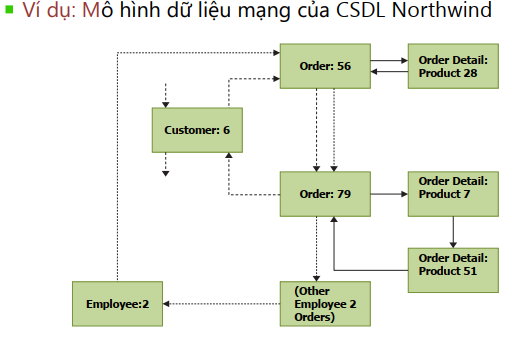


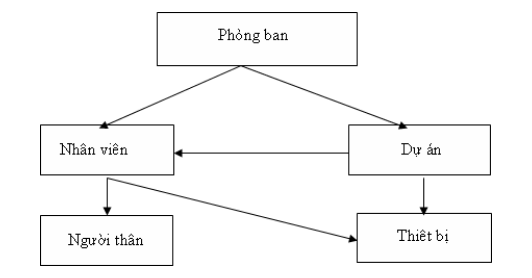
* 1. **Mô hình phân cấp (Hierarchical model)** trong mô hình này dữ liệu được tổ chức thành cấu trúc cây, các **nút (node)** là tập các thực thể, các cành là các mối quan hệ giữa hai nút theo mối quan hệ nhẩt định, cứng nhắc. Hay nói cách khác:
     + Là mô hình dữ liệu trong đó các bản ghi được sắp xếp theo cấu trúc **top-down(tree).**
     + Một con chỉ có một cha -> chỉ có một đường truy nhập tới dữ liệu đó trước.



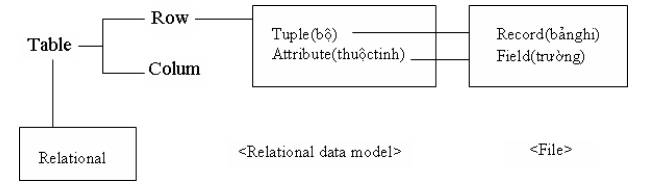


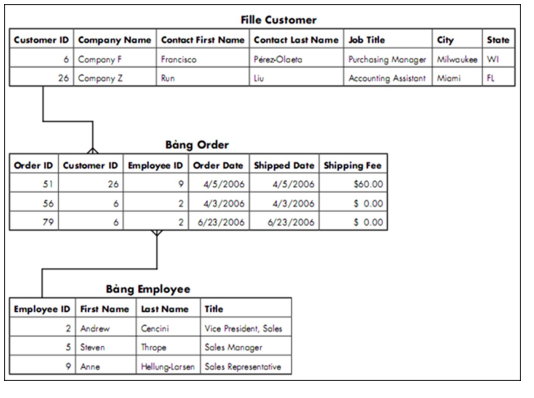
* 1. **Mô hình mạng(Network model)**
     + Các file riêng biệt trong hệ thống file phẳng được gọi là các bản ghi . Tập hợp bản ghi cùng kiểu tạo thành một kiểu thực thể dữ liệu.
     + Các kiểu thực thể kết nối với nhau thông qua mối quan hệ cha-con.
     + Mô hình dữ liệu mạng biểu diễn bởi một đồ thị có hướng, và các mũi tên chỉ từ kiểu thực thể cha sang kiểu thực thể con.





* + - **Tính chất:**
      * **Ưu điểm:**
        + Dễ biểu diễn mô hình
        + Diễn đạt được các liên hệ dữ liệu phức tạp
      * **Nhược điểm:**
        + Truy xuất chậm
        + Không thích hợp với các CSDL có quy mô lớn
  1. **Mô hình dữ liệu quan hệ(Relational model)** : mô hình này dựa trên lý thuyết tập hợp và đại số quan hệ. Vì tính chất chặt chẽ của toán học về lí thuyết tập hợp nên mô hình này đã mô tả dữ liệu một cách rõ ràng, mềm dẻo và là mô hình thông dụng nhất hiện nay. Hầu hết các hệ QTCSDL đều tổ chức dữ liệu theo mô hình dữ liệu quan hệ. Trong đó dữ liệu được tổ chức dưới dạng bảng các phép toán thao tác trên dữ liệu dựa trên lý thuyết tập hợp của toán học. Tập dữ liệu được tổ chức theo cấu trúc của mô hình dữ liệu quan hệ gọi là CSDL quan hệ.
     + **Nói ngắn gọn lại:**
       - 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)**
       - Dữ liệu trong hai bảng **liên hệ với nhau** thông qua các **cột chung.**





* 1. **Các khái niệm căn bản trong CSDL:**
     + Relational database
     + Table – Bảng
     + Column – Cột
     + Row - Hàng (bản ghi)
     + Primary Key – Khoá chính
     + Unique Key – Khoá duy nhất
     + Index – Chỉ mục
     + Quan hệ một- một (1-1)
     + Quan hệ một – nhiều (1-N)
     + Quan hệ nhiều nhiều (N-N)