BỘ GIÁO DỤC VÀ ĐẠO TẠO

**TRƯỜNG ĐẠI HỌC NÔNG LÂM TP HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN MÔN NHẬP MÔN TIN HỌC**

ĐỀ TÀI:

**CƠ SỞ DỮ LIỆU**

**GVHĐ: ĐẶNG KIÊN CƯỜNG**

**Sinh viên: NGUYỄN PHI HÙNG 18130090**

**MAI TRANG MỘNG LINH 21130424**

**NGUYỄN ĐÌNH VIỆT ANH 21130272**

**TRẦN MINH PHÁT 21130470**

**TRƯƠNG VĂN THÀNH LỢI 21130431**

**VÕ MINH PHI** **21130471**

**Ngành/ chuyên ngành: Công nghệ thông tin**

**TP Hồ Chí Minh, tháng 11 năm 2021**

# **LỜI MỞ DẦU**

Công nghệ thông tin trong giai đoạn hiện nay đang có những bước phát triển như vũ bão trên mọi lĩnh vực hoạt động ở khắp nơi trên toàn thế giới. Dữ liệu đã và đang là một trong những vấn đề không thể thiếu đối với bất kỳ một tổ chức nào, công ty nào. Đặc biệt cơ sở dữ liệu ngày càng có vai trò quan trọng trong vấn đề quản lí dữ liệu tại các cơ quan, tổ chức, cá nhân trong lĩnh vực thu thập thông tin nói riêng và công nghệ thông tin nói chung.

Qua thời gian tìm hiểu về cơ sở dữ liệu là đề tài tụi em thấy cơ sở dữ liệu trong thời buổi công nghệ số hiện nay, nhiều quy trình, công đoạn hay các hệ thống quản trị đều được mã hóa và vận hành bởi các thiết bị, phần mềm nhằm giúp doanh nghiệp đạt được hiệu suất làm việc tốt nhất. Trên cơ sở đó, các hệ thống quản trị cơ sở dữ liệu ra đời và đóng vai trò quan trọng trong xử lý và kiểm soát nguồn thông tin

Lợi ích của cơ sở dữ liệu trong thực tế là không hề nhỏ, nó được ứng dụng ở hầu hết các mặt trong thực tế. Nhận biết được tầm quan trọng đó sau thời gian tìm hiểu được của nhóm em đã có những thông tin hiểu biết về cơ sở dữ liệu gồm những phần sau:

* **Chương 1:** GIỚI THIỆU SƠ LƯỢC CƠ SỞ DỮ LIỆU
* **Chương 2:** NỘI DUNG CỦA CƠ SỞ DỮ LIỆU
* **Chương 3:** KẾT LUẬN

Trong bản báo cáo này được tạo điều kiện sự hướng dẫn đề tài của Thầy và tìm kiếm thông tin, sự hiểu biết của nhóm em chỉ trình bày tóm tắt, sơ lược những kiến thức của cơ sở dữ liệu. Mặc dù nhóm em đã rất cố gắng nhưng chắc chắc có những thiếu sót. Mong quý thầy cô và các bạn sửa chữa bổ sung nhận xết đánh giá để bài báo cáo được hoàn thiện tốt hơn.

Nhóm em xin chân thành cảm ơn!

# **LỜI CẢM ƠN**

Đầu tiên, chúng em xin gửi lời cảm ơn chân thành đến Trường Đại học Nông Lâm TP.HCM đã đưa môn Nhập môn tin học vào giảng dạy. Đặc biệt chúng em xin gửi lời cảm ơn sâu sắc đến giáo viên bộ môn – PGS.TS. Đặng Kiên Cường đã dạy dỗ, truyền đạt những kiến thức quý báu, là hành trang để chúng em có thể vững bước cho con đường sau này.

Môn học “Nhập môn tin học” là môn học rất thú vị, vô cùng bổ ích và có tính thực tế cao, cung cấp đầy đủ kiến thức, gắn liền với cuộc sống của con người trong thời đại mới. Tuy nhiên do vốn kiến thức còn hạn chế và khả năng tiếp thu thực tế còn nhiều bỡ ngỡ.

Để đồ này đạt kết quả tốt đẹp, em và các thành viên đã hỗ trợ, giúp đỡ và kết nối tới nhau để hoàn thành. Với tình cảm sâu sắc, chân thành, cho phép em được bày tỏ lòng biết ơn sâu sắc đến tất cả các cá nhân và Thầy đã tạo điều kiện giúp đỡ trong quá trình học tập và nghiên cứu tìm hiểu đề tài.

Mặc dù chúng em đã cố gắng hết sức nhưng chắc chắn bài báo cáo khó có thể tránh khỏi nhưng thiếu sót và nhiều chỗ còn chưa chính xác, kính mong Thầy xem xét và góp ý để bài đồ án của chúng em được hoàn thiện hơn.

Nhóm em xin chân thành cảm ơn!

# **NHẬN XÉT CỦA GIẢNG VIÊN**

**….........................................................................................**

**...................................................................................................**

**...................................................................................................**

**...................................................................................................**

**...................................................................................................**

**...................................................................................................**

**...................................................................................................**

**...................................................................................................**

**...........................................................................**

**Ngày .... tháng .... năm .....**

**(Ký tên)**

**MỤC LỤC**

[**LỜI MỞ DẦU** 1](#_Toc86850885)

[**LỜI CẢM ƠN** 2](#_Toc86850886)

[**NHẬN XÉT CỦA GIẢNG VIÊN** 3](#_Toc86850887)

[**MỤC LỤC** 4](#_Toc86850888)

[**CHƯƠNG 1: GIỚI THIỆU SƠ LƯỢC CƠ SỞ DỮ LIỆU** 5](#_Toc86850889)

[**1.1** **Giới thiệu** 5](#_Toc86850890)

[**1.2** **Sơ lược lịch sử phát triển** 5](#_Toc86850891)

[**CHƯƠNG 2: NỘI DUNG CỦA CƠ SỞ DỮ LIỆU** 6](#_Toc86850892)

[**2.1** **Khái niệm cơ sở dữ liệu, thuật ngữ liên quan** 6](#_Toc86850893)

[**2.2** **Phân loại cơ sở dữ liệu** 7](#_Toc86850894)

[**2.3** **Phương pháp tổ chức dữ liệu** 10](#_Toc86850895)

[**2.4** **Quản lí cơ sở dữ liệu** 13](#_Toc86850896)

[**2.5** **Tại sao phải sử dụng cơ sở dữ liệu** 15](#_Toc86850897)

[**2.6** **Vai trò và tầm quan trọng của cơ sở dữ liệu** 16](#_Toc86850898)

[**CHƯƠNG 3: KẾT LUẬN** 18](#_Toc86850899)

[**TÀI LIỆU THAM KHẢO** 19](#_Toc86850900)

# **CHƯƠNG 1:** **GIỚI THIỆU SƠ LƯỢC CƠ SỞ DỮ LIỆU**

1. **Giới thiệu**

* Dữ liệu có vai trò quan trọng trong nền kinh tế và bối cảnh cách mạng công nghiệp 4.0. Mỗi ngày, chúng ta gặp gỡ rất nhiều con người, sự vật, sự việc và tiếp cận rất nhiều thông tin thuộc nhiều lĩnh vực trong cuộc sống.
* Tuy nhiên, chúng ta không cần phải nhớ hết tất cả các thông tin. Từ vô vàn thông tin đó, chúng ta lọc ra những thông tin cần thiết và hữu ích để lưu lại và từ đó khái niệm cơ sở dữ liệu ra đời.

1. **Sơ lược lịch sử phát triển**

* Lịch sử của cơ sở dữ liệu được xem như bắt đầu vào những năm 60 của thế kỷ trước, khi máy tính bắt đầu được sử dụng tương đối rộng rãi.
* Năm 1970, Edgar Codd, một nhà khoa học làm việc cho IBM, đề xuất một số khái niệm về hệ cơ sở dữ liệu quan hệ (relational) đánh dấu một bước chuyển biến quan trọng trong lĩnh vực này.
* Năm 1976, Peter Chen đã hoàn thiện thêm hệ CSDL này bằng mô hình thực thể - liên kết (Entity-Relationship hay E-R).
* Năm 1979, Relational Software Inc (RSI), sau đổi tên là Oracle, là công ty đầu tiên sử dụng mô hình quan hệ để thương mại hóa và dần dần chiếm lĩnh thị trường quản lý dữ liệu. Cũng trong thời gian này IBM cũng bắt đầu thương mại hóa SQL. Chẳng bao lâu SQL trở thành tiêu chuẩn để sử dụng với CSDL quan hệ.
* Năm 1979, Relational Software Inc (RSI), sau đổi tên là Oracle, là công ty đầu tiên sử dụng mô hình quan hệ để thương mại hóa và dần dần chiếm lĩnh thị trường quản lý dữ liệu. Cũng trong thời gian này IBM cũng bắt đầu thương mại hóa SQL. Chẳng bao lâu SQL trở thành tiêu chuẩn để sử dụng với CSDL quan hệ. Vì thế từ thập niên 1980, HQTCSDL hướng đối tượng (object-oriented DBMS) bắt đầu được nghiên cứu và đến những năm 1990, chúng bắt đầu được đưa ra thị trường.
* Cũng bắt đầu từ thập niên 1990, cùng với sự bùng nổ của internet và các dịch vụ trực tuyến, các HQTCSDL phục vụ cho thị trường này phát triển mạnh mẽ.
* Bên cạnh đó, một xu hướng khác cũng xuất hiện từ đầu thế kỷ 21 là các HQTCSDL theo hướng NoSQL (được hiểu là “not only SQL”), như MongoDB, có khả năng làm việc với các dữ liệu có cấu trúc ít bị ràng buộc chặt chẽ hơn.

# **CHƯƠNG 2: NỘI DUNG CỦA CƠ SỞ DỮ LIỆU**

1. **Khái niệm cơ sở dữ liệu, thuật ngữ liên quan**

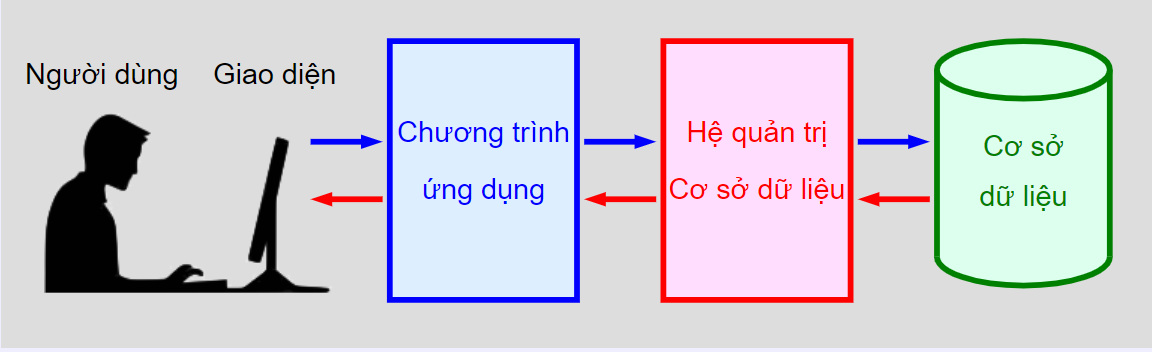
*Hình 1: Hình minh họa cơ sở dữ liệu*

* Dữ liệu
* **Dữ liệu**là một tập hợp các dữ kiện, chẳng hạn như số, từ, hình ảnh, nhằm đo lường, quan sát hoặc chỉ là mô tả về sự vật. Sự phát triển trong lĩnh vực công nghệ, đặc biệt là trong điện thoại thông minh đã dẫn đến việc văn bản, video và âm thanh được đưa vào dữ liệu cùng với nhật ký web. Hầu hết dữ liệu này là không có cấu trúc.
* Cơ sở dữ liệu
* **Cơ sở dữ liệu** là hệ thống bao gồm rất nhiều thông tin, dữ liệu được xây dựng theo một cấu trúc nhất định nhằm đáp ứng nhu cầu khai thác, sử dụng của nhiều người hay chạy nhiều chương trình ứng dụng cùng một lúc.
* Khi áp dụng hình thức lưu trữ này, nó sẽ giúp khắc phục được những điểm yếu của việc lưu file thông thường trên máy tính. Các thông tin lưu trữ sẽ đảm bảo được nhất quán, hạn chế tình trạng trùng lặp thông tin.
* Một số tính chất của cơ sở dữ liệu
* Qua phần trình bày trên, ta có thể rút ra một số tính chất của CSDL:
* CSDL tương ứng với một phần nào đó của thế giới thực (miniworld hay universe of discourse UoD). Khi phần thế giới thực ấy có sự thay đổi thì nội dung của CSDL phải thay đổi theo.
* Các nội dung trong CSDL phải có một ý nghĩa nhất định để thể hiện phần thế giới thực tương ứng với nó. Nói cách khác, giữa CSDL và phần thế giới thực ấy có những quan hệ chặt chẽ với nhau.
* CSDL được thiết kế, xây dựng để đáp ứng những mục đích xác định, phục vụ cho những đối tượng (người dùng – user) xác định.
* Thông tin
* Thông thường, các dữ liệu rời rạc không có nhiều giá trị. Để tăng giá trị của chúng, ta phải chuyển chúng thành thông tin. Trong tiếng Việt, thông tin có một số nghĩa ít nhiều khác nhau. Trong lĩnh vực cơ sở dữ liệu, ta định nghĩa thông tin như sau:
* "Thông tin là một tập hợp các dữ liệu được sắp xếp, xử lý theo phương thức nào đó."
* Sau quá trình chuyển đổi, giá trị của thông tin tăng lên đáng kể so với những dữ liệu ban đầu.

1. **Phân loại cơ sở dữ liệu**

* Khi nhắc đến phân loại cơ sở dữ liệu, có vô vàn kiểu phân loại cơ sở dữ liệu. Một số loại cơ sở dữ liệu thường được nhắc đến như: Cơ sở dữ liệu lớn, cơ sở dữ liệu hướng tài liệu, cơ sở dữ liệu hướng đối tượng, Cơ sở dữ liệu đồ thị, cơ sở dữ liệu thời gian thực, cơ sở dữ liệu tri thức, cơ sở dữ liệu không gian, cơ sở dữ liệu thời gian, cơ sở dữ liệu tập trung, cơ sở dữ liệu phân tán, cơ sở dữ liệu đám mây, cơ sở dữ liệu quan hệ, cơ sở dữ liệu ngữ nghĩa…. Với mỗi loại cơ sở dữ liệu này lại kéo theo rất nhiều các công nghệ liên quan, mục đích sử dụng và kỹ thuật thực hiện.
* Tuy nhiên, theo một cách thông dụng nhất, phân loại về cơ sở dữ liệu trong máy tính thường được thực hiện theo một số hình thức của dữ liệu cơ bản sau:
* Phân loại theo loại dữ liệu
* Cơ sở dữ liệu có cấu trúc (structured database): có nghĩa là cơ sở dữ liệu dữ liệu được định hình theo một cấu trúc xác định từ trước. Chúng ta có thể hình dung như một văn bản đã được xác định tiêu đề, có các dòng và cột với tiêu đề xác định trước, các thông tin chi tiết được lấp đầy các bảng này và không thay đổi khi cập nhật. Một hình dung khác về cơ sở dữ liệu dữ liệu có cấu trúc là một thư viện với các tủ hồ sơ được đánh nhãn, trong mỗi tủ được phân ngăn rõ ràng. Cơ sở dữ liệu có cấu trúc được xây dựng sẽ dễ dàng quản lý và truy cập thông tin.
* Cơ sở dữ liệu phi cấu trúc (unstructured database): là cơ sở dữ liệu không được xác định cấu trúc thông tin từ trước. Thường là tập hợp các dữ liệu thô, hỗn tạp và không đồng nhất. Các thành phần của cơ sở dữ liệu không có đặc điểm chung. Chúng ta có thể hình dung cơ sở dữ liệu này là tập hợp các thông tin, dữ liệu bao gồm: thư điện tử, dữ liệu ảnh, video, âm thanh, các bài viết,…Dữ liệu phi cấu trúc có mặt ở khắp mọi nơi và được sản sinh ra từ các nguồn khác nhau. Để quản lý, dữ liệu phi cấu trúc cần được chuyển đổi thành dữ liệu có cấu trúc qua quá trình chuẩn hóa.
* Cơ sở dữ liệu bán cấu trúc (semi-structured database): thường là dữ liệu có cấu trúc nhưng không đồng nhất. Cấu trúc của dữ liệu phụ thuộc vào chính nội dung của dữ liệu ấy. Chúng ta có thể thấy được rằng trong thực tế dữ liệu được lưu dưới dạng XML tự do (không kèm theo lược đồ), với định dạng này thông tin mô tả về đối tượng thể hiện trong các thẻ. Đây là cơ sở dữ liệu có nhiều ưu điểm do lưu trữ được hầu hết các loại dữ liệu khác nhau nên cơ sở dữ liệu bán cấu trúc là hướng mới trong nghiên cứu và ứng dụng và được sử dụng thông dụng trên mạng Internet. Tuy nhiên cũng cần lưu ý rằng XML cũng có thể được mô tả dữ liệu có cấu trúc bằng cách kèm xây dựng và lưu trữ dữ liệu tuân thủ lược đồ.
* Phân loại theo hình thức lưu trữ, mô hình tổ chức
* Cơ sở dữ liệu quan hệ (relational database): dữ liệu được lưu trữ trong các bảng dữ liệu gọi là các thực thể, giữa các thực thể này có mối liên hệ với nhau gọi là các quan hệ, mỗi quan hệ có các thuộc tính, trong đó có một thuộc tính là khóa chính. Các hệ quản trị hỗ trợ cơ sở dữ liệu quan hệ như: MS SQL server, Oracle, MySQL… là đặc trưng thể hiện của các cơ sở dữ liệu này.
* Cơ sở dữ liệu dạng tệp (file database) dữ liệu được lưu trữ dưới dạng các file có thể là văn bản, ảnh, thông tin nhị phân, hoặc phát triển hơn là tệp cơ sở dữ liệu nhỏ gọn của các phần mềm quản lý dữ liệu. Tiêu biểu cho cơ sở dữ liệu dạng tệp là\*.mdb Foxpro, Microsoft Access…dạng cơ sở dữ liệu này thường phù hợp với phạm vi nhỏ hoặc theo cách thức tổ chức quản lý dạng cũ.
* Cơ sở dữ liệu phân cấp (herachical database): Một mô hình cơ sở dữ liệu phân cấp là một mô hình dữ liệu trong đó các dữ liệu được tổ chức thành một cây cấu trúc. Các dữ liệu được lưu trữ như các hồ sơ đó được kết nối với nhau thông qua các liên kết. Một thực thể là một tập hợp của các thực thể con, mỗi thực thể con cuối cùng (gọi là lá) chỉ chưa giá trị. Thể hiện thực tế của loại cơ sở dữ liệu này là cơ sở dữ liệu được lưu như hệ thống thư mục trên ổ đĩa. Mỗi thư mục cha chứa các thư mục con và tệp nằm trong các thư mục. Một thể hiện khác là các cơ sở dữ liệu quản lý tài khoản, người dùng như LDAP, AD trong đó có chứa các tài khoản tổ chức, cá nhân có cấu trúc lồng nhau.
* Phân loại theo đặc tính sử dụng
* Cơ sở dữ liệu hoạt động (operational databases): Trong hoạt động của mỗi cơ quan, đơn vị hay doanh nghiệp luôn sản sinh ra một lượng lớn thông tin. Các thông tin này lại là đầu vào cho quá trình thực hiện một nghiệp vụ khác. Các thông tin này được đưa vào các cơ sở dữ liệu để quản lý và truy xuất bởi các đối tượng khác nhau. Đây chính là hình thức của các cơ sở dữ liệu hoạt động. Một cơ sở dữ liệu hoạt động thường là cực kỳ quan trọng đối với các tổ chức vì chúng bao gồm các cơ sở dữ liệu đối tác, khách hàng, cơ sở dữ liệu cá nhân và cơ sở dữ liệu sản phẩm hoạt động… Các dữ liệu được lưu trữ trong cơ sở dữ liệu hoạt động có thể được thay đổi và thao tác tùy thuộc vào những gì các tổ chức yêu cầu.
* Cơ sở dữ liệu kho (data warehouse): Các tổ chức được yêu cầu phải giữ tất cả các dữ liệu có liên quan trong nhiều năm. Những thông tin này cũng là một nguồn quan trọng của thông tin để phân tích và so sánh các dữ liệu năm nay với các năm trước đó cũng làm cho nó dễ dàng hơn để xác định xu hướng chính phát triển của các tổ chức đang diễn ra. Tất cả các dữ liệu này từ năm trước đó được lưu trữ trong một kho dữ liệu. Kể từ khi dữ liệu được lưu trữ đã trải qua tất cả các loại sàng lọc, chỉnh sửa và tích hợp. Cơ sở dữ liệu này đóng vai trò quan trọng trong việc hoạch định chính sách, định hướng phát triển và hỗ trợ ra quyết định. Đặc điểm của cơ sở dữ liệu này là không được sửa đổi nội dung mà chỉ có làm đầy thêm theo thời gian.
* Cơ sở dữ liệu ngữ nghĩa (semantic database): Đây là cơ sở dữ liệu mới thường được nhắc đến trong thời đại Internet. Là một cơ sở dữ liệu mềm dẻo và linh hoạt lưu trữ ngữ nghĩa của thông tin như như các sự kiện của các đối tượng. Cơ sở dữ liệu dữ liệu ngữ nghĩa được thiết kế để đại diện cho thế giới thực một cách chính xác nhất có thể trong tập hợp dữ liệu. Ký hiệu dữ liệu được tổ chức tuyến tính và phân cấp để cung cấp cho những ý nghĩa nhất định như một trong những mô tả ở trên. Bằng đại diện cho thế giới thực trong bộ dữ liệu, dữ liệu ngữ nghĩa cho phép các máy để tương tác với thông tin của thế gian mà không cần giải thích bởi con người.
* Phân loại theo mô hình triển khai
* Cơ sở dữ liệu tập trung (centralized database): Một cơ sở dữ liệu tập trung là một cơ sở dữ liệu được đặt, lưu trữ, và duy trì trong một địa điểm duy nhất. Đây là vị trí thường xuyên nhất thiết đặt một hệ thống máy tính hoặc cơ sở dữ liệu hệ thống trung tâm, ví dụ một máy tính chủ, hoặc một hệ thống máy tính máy tính lớn (Wikipedia). Thông thường, một cơ sở dữ liệu tập trung sẽ được duy trì và quản lý bởi một đầu mối, một tổ chức hoặc một cơ quan. Người dùng truy cập vào một cơ sở dữ liệu tập trung thông qua hệ thống mạng nội bộ hoặc mạng diện rộng, internet để truy cập vào các cơ sở dữ liệu để cập nhật hoặc khai thác trung tâm CPU, do đó duy trì cơ sở dữ liệu riêng của mình.
* Cơ sở dữ liệu phân tán (distributed database) là cơ sở dữ liệu không được lưu trữ và xử lý bởi nhiều máy tính, nhiều hệ thống thông tin và thường được đặt ở nhiều vị trí khác nhau. Các vị trí được kết nối với nhau bằng hệ thống mạng (có thể không thường xuyên trực tuyến). cơ sở dữ liệu phân tán có thể được quản lý bởi nhiều cơ quan, đơn vị và tổ chức khác nhau nhưng cần có một chính sách thống nhất để các cơ sở dữ liệu tương hợp và trao đổi thông tin với nhau.
* Cơ sở dữ liệu tập trung có bản sao: Các cơ sở dữ liệu tập trung và phân tán thường có những ưu điểm và nhược điểm của mình. Quyết định sử dụng loại nào còn phụ thuộc vào yếu tố quản lý và kỹ thuật như sự tham gia của các cơ quan, chính sách quản lý và khai thác dữ liệu, sự ổn định và tốc độ kết nối mạng, mức độ đáp ứng người sử dụng… cơ sở dữ liệu tập trung bản sao là cơ sở dữ liệu được lai giữa hai loại tập trung và phân tán trong đó một cơ sở dữ liệu tập trung lưu toàn bộ dữ liệu theo phạm vi quản lý, triển khai thêm các cơ sở dữ liệu thành phần bản sao theo từng phạm vi dữ liệu và kết nối đồng bộ đến cơ sở dữ liệu trung tâm. Các cơ sở dữ liệu bản sao thành phần này có thể phục vụ các mục đích khác nhau hoặc đặt ở các vị trí khác nhau để tối ưu và khắc phục các nhược điểm đã kể trên nhưng vẫn đảm bảo tính thống nhất như cơ sở dữ liệu tập trung.

1. **Phương pháp tổ chức dữ liệu**

* Việc tổ chức và sắp xếp dữ liệu thông minh của mỗi doanh nghiệp nói chung hay cá nhân nói riêng đều vô cùng cần thiết. Có nhiều phương pháp để bạn có thể tổ chức và sắp xếp dữ liệu khác nhau. Song tất cả đều hướng đến việc tổ chức hợp lý, thuận lợi cho người sử dụng dữ liệu mỗi khi cần tìm kiếm.
* Có nhiều các phương pháp tổ chức và sắp xếp dữ liệu khác nhau. Song dưới đây là 3 phương pháp chính:
* Sử dụng tập tin
* Trong thời kỳ đầu của việc ứng dụng tin học trong quản lý và sản xuất, khoảng những năm 60 của thế kỷ trước, người ta sử dụng những tập tin để lưu trữ các dữ liệu (về nhân sự, tài chính, vật tư, giao dịch, . . .). Để thuận tiện trong việc thao tác với các tập tin ấy, người ta xây dựng các chương trình ứng dụng khác nhau như tính toán nguyên liệu, hàng hóa xuất nhập, phân tích kết quả kinh doanh, tính điểm trung bình, đánh giá kết quả học tập của sinh viên.
* Việc sử dụng các tập tin ấy có thể dẫn đến một số điểm bất cập chính như sau:
* Có sự phụ thuộc giữa chương trình ứng dụng và dữ liệu. Một sự thay đổi nhỏ trong cấu trúc của dữ liệu có thể làm chương trình ứng dụng không hoạt động được.
* Không có sự nhất quán (consistency). Các tập tin thường được ghi nhận ở một số bộ phận khác nhau của cơ quan nên nếu không có những hướng dẫn thật cụ thể và chi tiết thì cách ghi chép trong những tập tin khác nhau có thể không giống nhau. Thí dụ cùng một địa chỉ nhưng ta thấy những cách ghi khác nhau : Nha Trang, Khánh Hòa, Phú Khánh. Điều này có thể dẫn đến việc xử lý bị sai lạc.
* Thừa dữ liệu (redundancy). Do được thực hiện bởi một số bộ phận trong tổ chức nên thường xuyên xẩy ra hiện tượng thừa dữ liệu do có sự trùng chập, cùng một dữ liệu lại được trình bày trong một số tập tin khác nhau. Các thông số của một sản phẩm có thể nằm ở các tập tin của phân xưởng sản xuất, phòng kế hoạch, bộ phận bán hàng, phòng điều độ. Điều này làm tăng dung lượng lưu trữ và chi phí. Đôi khi cách trình bày các dữ liệu trùng chập này lại khác nhau như trên đã đề cập, dẫn đến kết quả xử lý không còn chính xác.
* Việc tiếp cận dữ liệu khó khăn. Đối với một số thông tin, việc nắm bắt khá khó khăn do các dữ liệu có liên quan nằm rải rác ở một số tập tin. Thí dụ để so sánh thị hiếu của khách hàng ở các khu vực Đồng bằng sông Cửu Long, Tây Nguyên, và Thành phố Hồ Chí Minh cần sử dụng dữ liệu ở một số tập tin.
* Sự phân tán của dữ liệu: Do dữ liệu được phân tán trong nhiều tập tin, định dạng của các tập tin có thể khác nhau, xử lý bằng những chương trình khác nhau, có thể viết bằng những ngôn ngữ lập trình khác nhau, với những phong thái khác nhau. Vì thế việc liên kết chúng gặp nhiều khó khăn.
* Khi trong cùng một lúc, có nhiều người dùng khác nhau cùng sử dụng, nhiều chương trình cùng được thực thi, sẽ xẩy ra các sự tranh chấp (concurrency). Việc quản lý các tranh chấp này rất khó khăn, gần như là không thể thực hiện được.
* Các phiền toái xẩy ra trong các quá trình hiệu chỉnh, cập nhập và xóa bỏ dữ liệu vì ta phải thực hiện đồng thời trên một số tập tin một cách thích hợp. Người ta thấy rằng các quá trình này dễ xẩy ra những sai sót.
* An toàn dữ liệu: Vì dữ liệu có thể được thao tác từ một số chương trình khác nhau nên việc giữ cho dữ liệu an toàn tương đối khó khăn.
* Sử dụng bảng dữ liệu
* Ở thời kỳ đầu của việc ứng dụng tin học trong quản lý, người ta thường sử dụng nhiều tập tin để có thể lưu trữ dữ liệu. Theo đó, nhằm thuận tiện hơn trong việc thao tác với các tập tin này, người ta xây dựng nên những ứng dụng khác nhau. Có thể kể đến một số ứng dụng như tính hàng hóa xuất nhập, tính điểm trung bình, kết quả kinh doanh,…
* Tuy nhiên, phương pháp tổ chức và sắp xếp dữ liệu này còn nhiều bất cập. Do dữ liệu có thể được thao tác từ các chương trình khác nhau nên vấn đề an toàn dữ liệu, sự phân tán của dữ liệu và các phiền toái xảy ra trong quá trình hiệu chỉnh là những bất cập mà người dùng phương pháp này gặp phải.
* Sử dụng cơ sở dữ liệu
* Sử dụng cơ sở dữ liệu (CSDL) là phương pháp thông dụng nhất hiện nay. Toàn bộ dữ liệu được tập trung và tổ chức theo những khuôn mẫu thống nhất. Khi CSDL được tổ chức theo mô hình quan hệ, dạng phổ biến nhất, dữ liệu được lưu trữ trong các bảng, còn gọi là các [quan hệ](http://xdulieu.com/co-so-du-lieu/cs3-mo-hinh-quan-he/mq5-mo-hinh-co-so-du-lieu-quan-he.html) theo thuật ngữ của CSDL. Nhưng không giống như các bảng dữ liệu thông thường, trong các quan hệ có những cơ chế để các bảng có thể móc nối, liên kết với nhau.
* Ngoài ra, cấu trúc của hệ thống CSDL cũng có một số điểm khác biệt. Các chương trình ứng dụng không giao tiếp trực tiếp với CSDL mà qua một trung gian gọi là hệ quản trị cơ sở dữ liệu (HQTCSDL) như trên Hình 2.

*Hình 2: Sơ đồ cấu trúc của hệ thống cơ sở dữ liệu.*

* Việc sử dụng cơ sở dữ liệu để quản lý và khai thác dữ liệu có các ưu điểm sau:
* Có sự độc lập giữa chương trình ứng dụng và dữ liệu. Điều này cho phép cấu trúc dữ liệu có thể thay đổi rất nhiều mà vẫn không ảnh hưởng đến chương trình ứng dụng.
* Do dữ liệu được tập trung nên giải quyết được hay kiểm soát được vấn đề trùng chập dữ liệu và không nhất quán trong trình bày. Do đó giảm thiểu các sai sót gây ra do hai vấn đề trên.
* Chất lượng dữ liệu được nâng cao. HQTCSDL cung cấp những cơ chế sàng lọc để chỉ đưa các dữ liệu đạt các điều kiện nhất định mới được đưa vào cơ sở dữ liệu.
* Dữ liệu được chia sẻ tốt hơn, khả năng truy cập đến nhiều loại dữ liệu hơn.
* HQTCSDL đã cung cấp một số công cụ chính để làm việc với cơ sở dữ liệu nên việc thực hiện các chương trình ứng dụng được đơn giản hơn, rút ngắn hơn, quá trình triển khai dễ dàng hơn.
* Bảo trì hệ thống dữ liệu dễ dàng hơn.

1. **Quản lí cơ sở dữ liệu**

* Sau khi được thực hiện, để cơ sở dữ liệu hoạt động được hiệu quả, ổn định trong thời gian dài, đáp ứng được các mục đính đã đề ra và có thể ứng phó được với một số biến động của môi trường hoạt động, cơ sở dữ liệu phải được quản lý một cách chặt chẽ.
* Khái quát về quản lí cơ sở dữ liệu
* Đối với rất nhiều cơ quan, dữ liệu là một tài sản có giá trị như những dạng tài sản khác như con người, trang thiết bị, công nghệ, tài chính. Vì thế nó cần được quản lý một cách chặt chẽ, hợp lý, khoa học.
* Hai khía cạnh quan trọng của quản lý dữ liệu là an toàn và chất lượng phục vụ. Một hệ thống dữ liệu hoạt động ì ạch, thường xuyên hỏng hóc, dữ liệu dễ bị mất mát, sai lạc sẽ ảnh hưởng đáng kể đến nhiều hoạt động của cơ quan.
* Trong những cơ quan lớn, quản lý cơ sở dữ liệu là một bộ phận của quản lý hệ thống thông tin của cơ quan. Vai trò chủ yếu của bộ phận quản lý cơ sở dữ liệu là làm cho hệ thống này hoạt động trôi chảy, đáp ứng yêu cầu của người dùng. Ngoài ra trong chừng mực nhất định, bộ phận này có thể tham gia vào việc thiết kế, xây dựng cơ sở dữ liệu cũng như trang bị hạ tầng, phần mềm và/hay phần cứng. Ngoài ra bộ phận này cũng có thể tham gia vào việc xử lý dữ liệu.
* Các nội dung chính của quản lý cơ sở dữ liệu là:
* quản lý nội dung của cơ sở dữ liệu,
* quản lý kỹ thuật cơ sở dữ liệu,
* đảm bảo an toàn cho dữ liệu.
* Quản lí nội dung cơ sở dữ liệu
* Trong đa số trường hợp, cơ sở dữ liệu có tính động, thường xuyên thay đổi. Cấu trúc ban đầu có thể không đáp ứng được những yêu cầu mới. Trong quá trình hoạt động, nội dung cơ sở dữ liệu có thể thường xuyên thay đổi do thêm, bớt dữ liệu. Ngoài ra còn có sự xuống cấp trong quá trình hoạt động. Tất cả những điều này khiến chất lượng phục vụ của cơ sở dữ liệu giảm sút, hỏng hóc có thể xẩy ra. Vì vậy bộ phận quản lý cơ sở dữ liệu phải có những biện pháp phù hợp.
  + Giám sát hoạt động của cơ sở dữ liệu: Cần thường xuyên theo dõi tình hình hoạt động của cơ sở dữ liệu, khả năng đáp ứng các yêu cầu của người dùng. Các thông số kỹ thuật như thời gian đáp ứng yêu cầu của người dùng, dung lượng truyền, lượng bộ nhớ sử dụng, số báo lỗi trong một đơn vị thời gian. Đặc biệt là các chỉ tiêu liên quan đến sự truy vấn cơ sở dữ liệu.
* Xử lí sự cố: Trong quá trình hoạt động, có thể xẩy ra các sự cố lớn nhỏ khác nhau như không cập nhật được dữ liệu, thời gian trả lời các tham vấn quá dài, . . . Nguyên nhân của các sự cố này có nhiều như số người dùng tăng lên đột biến, vượt quá tiên liệu, bộ nhớ cấp phát không đủ, có sự xung đột phần mềm, có sự phá hoại từ bên ngoài, . . . Bộ phận quản lý phải xác định chính xác nguyên nhân các sự cố và có biện pháp khắc phục kịp thời.
* Sao lưu phục hồi dữ liệu: Một số sự cố như mất điện, hỏng ổ cứng, bị virus tấn công có thể làm dữ liệu bị hư hỏng, mất mát, trong một số trường hợp là rất nghiêm trọng. Vì thế dữ liệu cần được thường xuyên sao lưu để cơ sở dữ liệu có thể phục hồi lại trong thời gian ngắn. Hiện nay các phần mềm quản lý dữ liệu thường có công cụ để thực hiện này một cách dễ dàng, hiệu quả. Vì vậy ta chỉ cần có kế hoạch hợp lý để thực hiện việc sao lưu.
* Quản lí cơ sở dữ liệu về mặt kỹ thuật
* Hệ cơ sở dữ liệu cần phải được hoạt động trong một môi trường đáng tin cậy, có khả năng phục vụ lâu dài. Môi trường này gồm cả phần cứng và phần mềm.
* Về phần cứng: Cấu hình của máy chủ (server), phải đủ mạnh, dung lượng ổ cứng, RAM, bộ nhớ đệm phải đủ lớn, có thể hoạt động liên tục trong thời gian dài, đường truyền phải có băng thông thích hợp.
* Về phần mềm: Phần mềm của hệ quản trị cơ sở dữ liệu, các chương trình ứng dụng, cũng như các tiện ích phải được cài đặt một cách đúng đắn, các chế độ, các tùy chọn phải được lựa chọn một cách hợp lý.
* Phần cứng và phần mềm cần được thường xuyên cập nhật, và nâng cấp để đáp ứng nhu cầu của người dùng, tiến bộ của công nghệ, sự tăng trưởng của đơn vị.
* Với kinh nghiệm và trình độ của mình, bộ phận quản lý cơ sở dữ liệu có thể đưa ra những đóng góp tư vấn quan trọng trước khi trang bị và triển khai HQTCSDL.
* An toàn cơ sở dữ liệu
* Các truy cập không hợp lệ hay sửa đổi dữ liệu với ý đồ xấu, các tấn công phá hoại có thể làm mất mát, sai lạc dữ liệu, gây hư hỏng nghiêm trọng. Vì thế cần phải có những biện pháp an ninh hiệu quả như:
* Chống các thâm nhập hay tấn công từ bên ngoài bằng tường lửa, các phần mềm chống các chương trình phá hoại, thường xuyên cập nhật các bản vá lỗi của các hệ điều hành.
* Cấp phát quyền sử dụng cơ sở dữ liệu một cách chặt chẽ, chi tiết cho các đối tượng người dùng.
* Xây dựng các cơ chế giám sát các hoạt động của mọi người dùng, cả bên ngoài và trong nội bộ cơ quan.

1. **Tại sao phải sử dụng cơ sở dữ liệu**

* Nếu CSDL thực chất là một tập hợp các danh sách được lưu trữ trong các bảng và có thể tạo bảng trong Excel thì tại sao lại cần CSDL và phải dùng đến Access? Excel thiên về lưu trữ và xử lý dữ liệu dạng số, trong khi đó Access mạnh hơn ở việc xử lý dữ liệu không phải dạng số, như tên, mô tả. Dữ liệu không phải là số đóng vai trò quan trọng trong hầu hết các CSDL và việc sắp xếp, phân tích nó mới là điều cần thực hiện thường xuyên.
* Hơn nữa, sự khác biệt chính trong việc tổ chức dữ liệu trong CSDL với Access và theo các cách khác là sự kết nối. Những CSDL như vậy được gọi là CSDL quan hệ, nhìn vào đó, chúng ta có thể hiểu được các danh sách và đối tượng trong CSDL liên hệ với nhau như thế nào. Quay trở lại với CSDL chỉ có hai hàng là tên bạn bè và loại bánh đã làm. Giờ bạn muốn tạo thêm danh sách thứ ba là những loại bánh đã làm và tặng chúng cho ai. Vì bạn chỉ làm bánh đã biết công thức và chỉ mang cho bạn bè nên danh sách mới này sẽ bao gồm thông tin từ hai danh sách đã có trước đó.

 *Hình 3. Tại sao phải sử dụng cơ sở dữ liệu?*

* Lý giải thêm sao nên sử dụng CSDL
* Nhìn cách danh sách thứ 3 được tạo ra bằng cách sử dụng các từ có trong hai danh sách trước đó bạn có thể hiểu được rằng Dad và Oatmeal trong Batches chính là Dad và Oatmeal trong 2 danh sách ban đầu. Mối quan hệ này có vẻ hiển nhiên, tuy nhiên, nếu là một bảng tính Excel thì sẽ khó hiểu hơn.
* Excel sẽ coi tất cả những dữ liệu đã nhập vào là những mẩu thông tin riêng biệt, không liên quan đến nhau. Trong Excel, bạn phải nhập mọi thông tin về một người hoặc một loại bánh mỗi khi đề cập đến nó, vì CSDL trong Excel không phải là CSDL quan hệ như trong Access. Để đơn giản có thể hiểu CSDL quan hệ có thể nhận ra những gì mà con người có thể nhận: Nếu một từ cùng xuất hiện trong nhiều danh sách, chúng đang đề cập đến cùng một thứ.
* Với cách xử lý thông tin như vậy, CSDL quan hệ giúp nhập, tìm kiếm, phân tích dữ liệu trong nhiều bảng cùng lúc. Tất cả những điều này sẽ rất khó thực hiện trong Excel, nhưng trong Access thậm chí là các tác vụ phức tạp hơn cũng có thể được đơn giản hóa và thân thiện với người dùng.

1. **Vai trò và tầm quan trọng của cơ sở dữ liệu**

* Trong thời đại công nghệ 4.0, database chiếm vị trí quan trọng và trở thành một phần không thể thiếu trong hầu hết các hoạt động đời sống hàng ngày.
* Lưu trữ thông tin có hệ thống
* Cấu trúc sắp xếp có tính hệ thống – đây là điều làm nên sự khác biệt lớn nhất giữa dữ liệu thông thường và CSDL database. Dữ liệu sẽ được lưu trữ theo một cấu trúc nhất định, có tính nhất quán cao.
* Với đặc điểm này, database giúp người dùng thuận tiện trong việc tạo lập, lưu trữ, tìm kiếm và sử dụng dữ liệu một cách chính xác và nhanh chóng.
* Đảm bảo an toàn dữ liệu
* Đảm bảo toàn vẹn và an toàn dữ liệu là điểm quan trọng hàng đầu trong công tác lưu trữ dữ liệu. Và database xứng đáng nhận “điểm 10” về độ an toàn dữ liệu.
* Nếu muốn lưu dữ liệu ra file text, bạn nên sử dụng cách lưu ra File .xml và file.[csv](https://www.semtek.com.vn/csv-file/) là tốt nhất. Khi lưu ra hai loại file bạn hoàn toàn có thể chuyển đổi vào database dễ dàng và an toàn. Thông thường cách này thường dùng trong việc sao lưu database ra ngoài để đọc được dễ dàng.
* Đảm bảo khả năng truy xuất đồng thời của nhiều người dùng trên dữ liệu
* Nhiều người có thể sử dụng database cùng lúc mà không phải qua các khâu rườm rà phức tạp nhờ vào việc truy xuất từ các cách khác nhau. Do đó, bạn sẽ gặp nhiều thuận lợi trong việc việc sử dụng, quản lý, truy cập dữ liệu,…
* Linh hoạt thay đổi theo nhu cầu của người dùng
* Bạn có thể linh hoạt thay đổi kích cỡ và độ phức tạp của một database. Có những database chỉ gồm vài trăm bản ghi (danh sách học sinh của một lớp) và có những database có dung lượng rất lớn (như database quản lí hàng hoá của một hệ thống siêu thị).
* Song song đó, hình thức lưu trữ database cũng khá đa dạng. Database có thể được lưu trữ dưới nhiều dạng khác nhau như ổ cứng, USB hay đĩa CD.
* Công tác quản lí dễ dàng hơn
* Một database được thiết kế, hình thành, lưu trữ để dễ dàng trong việc tạo lập, cập nhập và khai thác thông tin. Dữ liệu sẽ được cập nhật thường xuyên và hoàn toàn không trùng lặp. Sử dụng database giúp tạo ra các sản phẩm chuyên nghiệp hơn, lưu trữ có hệ thống, dễ dàng trong công tác quản lí.
* Với ưu điểm đó, database ngày càng phổ biến trong lĩnh vực lập trình ứng dụng nói riêng và công nghệ thông tin nói chung.
* Ví dụ: Hệ thống đặt vé máy bay của Vietnam Airlines: ứng dụng database được phát huy tính năng khá tốt. Trong trường hợp nhiều hành khách đặt vé cùng lúc, database sẽ giúp doanh nghiệp tránh khỏi những sai sót không mong muốn. Điển hình như: [khách hàng](https://www.semtek.com.vn/khai-niem-khach-hang/) mua phải vé đã bán cho người khác, một vé nhưng nhiều [khách hàng](https://www.semtek.com.vn/khai-niem-khach-hang/) đặt, ….
* Tích hợp database với hệ thống website là xu hướng hiện nay của các tổ chức, công ty doanh nghiêp. Việc tích hợp này cho phép tổ chức, công ty đó gửi và thu thập thông tin với người dùng nhanh gọn và hiệu quả. Bạn có thể tạo ra database và duy trì một cách thủ công hoặc tin học hoá database. Khi được tin học hoá, một database được tạo ra và duy trì bằng bằng một nhóm chương trình ứng dụng hoặc bằng một hệ quản trị cơ sở dữ liệu.

# **CHƯƠNG 3: KẾT LUẬN**

* Công nghệ chắc chắn sẽ không biến mất. Như vậy cơ sở dữ liệu là điều tất yếu không thể thiếu trong trong nền công nghệ hóa hiện đại đất nước và trên toàn thế giới.
*  Qua thời gian tìm hiểu, nghiên cứu và đang học tập về nghành công nghệ thông tin thì tụi em đã trang bị cho mình được kiến thức thấy được tầm quan trọng của Cơ sở dữ liệu trong các lĩnh vực thiết yếu như thế nào.

*Hình 4: Minh họa của CSDL thúc đẩy quá trình của doanh nghiệp*

* Tóm lại, cơ sở dữ liệu là hết sức cần thiết nhằm bước tiến tới công đoạn quản lí dữ liệu đối với tổ chức và doanh nghiệp để thúc đẩy quá trình ra quyết định vận hành và hoạch định chiến lược trong doanh nghiệp hay tổ chức nào đó.

# **TÀI LIỆU THAM KHẢO**

**Nguồn internet:**

1. <https://vi.wikipedia.org/wiki/C%C6%A1_s%E1%BB%9F_d%E1%BB%AF_li%E1%BB%87u>
2. <http://xdulieu.com/co-so-du-lieu/cs1-khai-quat/index.html>
3. <https://edata.it-care.vn/cac-phuong-phap-to-chuc-va-sap-xep-du-lieu-can-phai-biet-58>
4. <https://www.semtek.com.vn/co-so-du-lieu/#Vai_tro_va_tam_quan_trong_cua_co_so_du_lieu>
5. <https://hocjavascript.net/co-so-du-lieu/tim-hieu-ve-co-so-du-lieu/>