

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG
KHOA CÔNG NGHỆ THÔNG TIN**



Ths. Nguyễn Thu Hương
TS. Nguyễn Thị Thanh Nhân
Ths. Lê Thu Trang
Ths. Trịnh Văn Hà
Ths. Ngô Thị Lan Phương
Ths. Phạm Thị Liên
Ths. Đào Trần Chung
Ths. Hồ Thị Tuyên
Ths. Vũ Thị Nguyệt
Ths. Dương Thị Mai Thương

**BÀI GIẢNG
CƠ SỞ DỮ LIỆU**

Tài liệu lưu hành nội bộ

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG
KHOA CÔNG NGHỆ THÔNG TIN

TS. Nguyễn Thị Thanh Nhân
Ths. Nguyễn Thu Hương
Ths. Lê Thu Trang
Ths. Trịnh Văn Hà
Ths. Ngô Thị Lan Phương
Ths. Phạm Thị Liên
Ths. Đào Trần Chung
Ths. Hồ Thị Tuyền
Ths. Vũ Thị Nguyệt
Ths. Dương Thị Mai Thương

BÀI GIẢNG
CƠ SỞ DỮ LIỆU

Thái Nguyên, tháng 4 năm 2022

MỤC LỤC

CÁC TỪ VIẾT TẮT	7
MỞ ĐẦU	8
Chương 1: TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU	9
Nội dung chính của chương.....	9
Mục tiêu cần đạt được của chương.....	9
Bài 1: TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU (Số tiết: 3 tiết).....	9
1.1 Cơ sở dữ liệu	9
1.1.1 Định nghĩa cơ sở dữ liệu.....	9
1.1.2 Phân loại cơ sở dữ liệu.....	9
1.1.3 Ưu điểm của cơ sở dữ liệu	12
1.1.4 Những vấn đề mà cơ sở dữ liệu cần giải quyết.....	12
1.1.5 Các đối tượng sử dụng cơ sở dữ liệu	13
1.1.6 Các ứng dụng của cơ sở dữ liệu.....	13
1.2 Hệ cơ sở dữ liệu	14
1.2.1 Khái niệm hệ cơ sở dữ liệu	14
1.2.2 Các thành phần của hệ cơ sở dữ liệu	14
1.2.3 Ưu điểm của hệ cơ sở dữ liệu	16
1.2.4 Các loại kiến trúc hệ cơ sở dữ liệu	17
Câu hỏi	18
Bài 2: TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU (TIẾP) (Số tiết: 3 tiết).....	18
1.3 Các mô hình dữ liệu	18
1.3.1 Mô hình dữ liệu mạng.....	18
1.3.2 Mô hình dữ liệu phân cấp	20
1.3.3 Mô hình dữ liệu thực thể kết hợp	20
1.3.4 Mô hình dữ liệu quan hệ.....	21
1.3.5 Mô hình dữ liệu hướng đối tượng.....	22
1.3.6 Mô hình dữ liệu phân tán.....	23
1.3.7 Mô hình cơ sở dữ liệu suy diễn	24
1.4 Hệ quản trị cơ sở dữ liệu	24
1.4.1 Khái niệm hệ quản trị cơ sở dữ liệu.....	24

1.4.2 Kiến trúc hệ quản trị cơ sở dữ liệu	25
1.4.3 Chức năng của hệ quản trị cơ sở dữ liệu	26
Câu hỏi cuối chương 1	27
CHƯƠNG 2: MÔ HÌNH DỮ LIỆU QUAN HỆ	28
Nội dung chính của chương:.....	28
Mục tiêu cần đạt được của chương:.....	28
Bài 3: MÔ HÌNH DỮ LIỆU QUAN HỆ (Số tiết: 3 tiết)	28
2.1 Các khái niệm cơ bản.....	28
2.1.1 Thuộc tính	28
2.1.2 Quan hệ	28
2.1.3 Khóa của một quan hệ	29
2.2 Các phép toán đại số quan hệ	30
2.2.1 Phép hợp	31
2.2.2 Phép giao.....	32
2.2.3 Phép trừ.....	33
2.2.4 Phép tích Đề-các	34
2.2.5 Phép chiếu.....	36
2.2.6 Phép chọn.....	37
2.2.7 Phép kết nối	39
2.2.8 Phép chia.....	40
Bài tập cuối chương 2	43
Chương 3: LÝ THUYẾT THIẾT KẾ CƠ SỞ DỮ LIỆU	46
Nội dung chính của chương.....	46
Mục tiêu cần đạt được của chương.....	46
Bài 4: PHỤ THUỘC HÀM (Số tiết: 1.5 tiết)	46
3.1 Phụ thuộc hàm và các dạng chuẩn	46
3.1.1. Phụ thuộc hàm (Functional Dependencies).....	46
Câu hỏi	50
Bài 5: CÁC DẠNG CHUẨN (Số tiết: 1.5 tiết)	51
3.1 Phụ thuộc hàm và các dạng chuẩn (Tiếp).....	51
3.1.2 Các dạng chuẩn.....	51
Câu hỏi	53
Bài tập	53

Bài 6: CÁC DẠNG CHUẨN (Số tiết: 3 tiết)	54
3.1 Phụ thuộc hàm và các dạng chuẩn (Tiếp)	54
3.1.2 Các dạng chuẩn (Tiếp).....	54
Bài tập	60
Bài 7: CÁC DẠNG CHUẨN (tiếp) (Số tiết: 03 tiết)	62
3.1 Phụ thuộc hàm và các dạng chuẩn (Tiếp)	62
3.1.2 Các dạng chuẩn (Tiếp).....	62
Câu hỏi, Bài tập	66
Bài 8: THIẾT KẾ CƠ SỞ DỮ LIỆU (Số tiết: 3 tiết)	68
3.2 Thiết kế cơ sở dữ liệu	68
3.2.1 Quy trình thiết kế cơ sở dữ liệu	68
3.2.2 Mô hình thực thể kết hợp (ERD)	70
3.2.3 Các khái niệm cơ bản trong ERD	71
Câu hỏi	84
Bài tập	85
Bài 9: THIẾT KẾ CƠ SỞ DỮ LIỆU (tiếp) (Số tiết: 3 tiết)	85
3.2 Thiết kế cơ sở dữ liệu (Tiếp)	86
3.2.4 Mô hình thực thể kết hợp mở rộng (EER)	86
3.2.5 Chuyển mô hình ERD sang mô hình quan hệ	92
Bài tập	96
Bài tập cuối chương III	97
Chương 4: NGÔN NGỮ ĐỊNH NGHĨA VÀ THAO TÁC DỮ LIỆU	101
Nội dung chính của chương.....	101
Mục tiêu cần đạt được của chương.....	101
Bài 10: GIỚI THIỆU VỀ HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU (Số tiết: 3 tiết) ..	101
4.1 Giới thiệu về hệ quản trị CSDL SQL Server	101
4.1.1 Các editions của SQL Server	102
4.1.2 Các Version của SQL Server	102
4.1.3 Các thành phần quan trọng trong SQL Server.....	103
4.2 Ngôn ngữ định nghĩa dữ liệu	104
4.2.1 Cú pháp của T-SQL	104
4.2.2 Lệnh tạo cơ sở dữ liệu.....	111
Câu hỏi	112

Bài 11: NGÔN NGỮ ĐỊNH NGHĨA DỮ LIỆU (Tiếp) (Số tiết: 3 tiết).....	112
4.2.3 Lệnh tạo và sửa đổi cấu trúc bảng dữ liệu	112
Câu hỏi	116
Bài 12: NGÔN NGỮ THAO TÁC DỮ LIỆU (Số tiết: 3 tiết)	116
4.3 Ngôn ngữ thao tác dữ liệu	116
4.3.1 Lệnh truy vấn dữ liệu – Select.....	117
Câu hỏi	126
Bài 13: LỆNH TRUY VẤN DỮ LIỆU LỒNG NHAU (Số tiết: 3 tiết)	126
4.3 Ngôn ngữ thao tác dữ liệu (Tiếp).....	126
4.3.2 Lệnh truy vấn dữ liệu lồng nhau.....	126
Câu hỏi	128
Bài 14: CÁC THAO TÁC DỮ LIỆU KHÁC TRONG SQL SERVER (Số tiết: 2 tiết).....	128
4.3 Ngôn ngữ thao tác dữ liệu (Tiếp).....	128
4.3.3 Lệnh nhập dữ liệu	128
4.3.4 Lệnh cập nhật dữ liệu.....	130
4.3.5 Lệnh xóa dữ liệu	130
Câu hỏi	131
Bài tập cuối chương 4	131
Tài liệu tham khảo	139
Các câu hỏi thường gặp	140
Bài tập thực hành.....	149
Bài thực hành số 1 (Số tiết: 5 tiết).....	149
Bài thực hành số 2 (số tiết: 5 tiết)	153
Bài thực hành số 3 (số tiết: 5 tiết)	157

CÁC TỪ VIẾT TẮT

TT	Từ viết tắt	Ý nghĩa của từ
1	CSDL	Cơ sở dữ liệu
2	ERD	Entity Relationship Diagram – Sơ đồ thực thể liên kết
3	SQL	Structured Query Language – Ngôn ngữ truy vấn có cấu trúc
4	IBM	International Business Machines
5	NSD	Người sử dụng
6	DBA	Database Administrator – Người quản trị csdl
7	DOM	Domain – Miền trị của thuộc tính
8	PTH	Phụ thuộc hàm
9	1NF	1 st Normal Form. Dạng chuẩn thứ nhất
10	2NF	2 st Normal Form. Dạng chuẩn thứ hai
11	3NF	3 rd Normal Form. Dạng chuẩn thứ ba
12	BCNF	Dạng chuẩn Boye - Code
13	EER	Enhanced Entity-Relationship – Quan hệ thực thể nâng cao
14	CSDL	Cơ sở dữ liệu
15	SQL	Structured Query Language -Ngôn ngữ truy vấn có cấu trúc
16	RDBMS	Relational Database Management System -Hệ quản trị CSDL quan hệ
17	RAM	Random Access Memory – Bộ nhớ truy cập ngẫu nhiên
18	CPU	Central Processing Unit – Đơn vị xử lý trung tâm
19	GB	Gigabyte
20	T-SQL	Transact-SQL

MỞ ĐẦU

Mô hình dữ liệu quan hệ được E. Codd đề xuất năm 1970 đã đánh dấu khởi đầu cho sự phát triển quan trọng về cơ sở lý thuyết cũng như khả năng ứng dụng của các hệ cơ sở dữ liệu. Bên cạnh đó là sự bùng nổ về nhu cầu xây dựng các hệ thống thông tin, trước hết là các hệ thống thông tin quản lý cũng đã kéo theo nhiều hãng máy tính lớn vào cuộc với sự xuất hiện của các hệ quản trị cơ sở dữ liệu danh tiếng như DB2, Sybase, Oracle... Đặc biệt, sự phát triển ngoạn mục của máy tính cá nhân đã tạo ra một bước chuyển biến mạnh mẽ về cách tiếp cận cũng như về triết lý triển khai các hệ cơ sở dữ liệu, khi đó việc tiếp cận và học tập là tất yếu. Để đáp ứng nhu cầu học tập của sinh viên Khoa Công nghệ Thông tin, bài giảng “Cơ sở dữ liệu” được nhóm tác giả bộ môn Hệ thống thông tin biên soạn theo đề cương môn học. Bài giảng này với mục tiêu cung cấp các kiến thức cơ bản về lý thuyết cơ sở dữ liệu.

Nội dung bài giảng gồm 4 chương:

Chương 1. Tổng quan về cơ sở dữ liệu. Trình bày các khái niệm cơ bản về cơ sở dữ liệu; hệ cơ sở dữ liệu; các mô hình dữ liệu và hệ quản trị cơ sở dữ liệu cũng như các cấu trúc của hệ cơ sở dữ liệu và hệ quản trị cơ sở dữ liệu.

Chương 2. Mô hình dữ liệu quan hệ. Trình bày các lý thuyết cơ bản về cơ sở dữ liệu quan hệ, cùng với đối tượng và các phép toán được xây dựng trên cơ sở toán học chặt chẽ và logic, có tính độc lập dữ liệu cao và có cấu trúc biểu diễn đơn giản.

Chương 3. Lý thuyết thiết kế cơ sở dữ liệu. Giới thiệu thiết kế cấu trúc cơ sở dữ liệu. Các dạng chuẩn trong cơ sở dữ liệu, mối quan hệ giữa các thực thể trong cơ sở dữ liệu. Kết quả chúng ta sẽ thu được các thành phần dữ liệu, xác định được các bảng và các cột cho cơ sở dữ liệu, cuối cùng là chúng ta sẽ thu được mô hình thực thể ERD.

Chương 4. Ngôn ngữ định nghĩa và thao tác dữ liệu. Giới thiệu ngôn ngữ định nghĩa và thao tác dữ liệu trên SQL Server 2014.

Bài giảng được biên soạn trên cơ sở kế thừa những nội dung của các giáo trình về cơ sở dữ liệu đã được xuất bản trong và ngoài nước nhưng được cấu trúc lại, bổ sung và mở rộng cho phù hợp với sinh viên Khoa Công nghệ thông tin. Ngoài ra, tài liệu đã trình bày một số kỹ năng cần thiết để thiết kế và cài đặt các hệ cơ sở dữ liệu cụ thể. Chúng tôi hy vọng sẽ là tài liệu có ích trong học tập và ứng dụng trong việc xây dựng các hệ thống tin học ứng dụng phục vụ cho sản xuất, quản lý trong các doanh nghiệp.

Mặc dù bài giảng đã được biên soạn nghiêm túc, cẩn thận, nhưng chắc chắn không thể tránh khỏi những thiếu sót trong biên soạn. Chúng tôi rất mong nhận được sự góp ý của các nhà khoa học, các đồng nghiệp và các em sinh viên đóng góp ý kiến để cuốn bài giảng được hoàn thiện hơn. Xin trân trọng cảm ơn.

Nhóm tác giả

Chương 1: TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU

Nội dung chính của chương

Chương 1 trình bày những kiến thức cơ bản về cơ sở dữ liệu (CSDL); các đặc điểm của CSDL; người sử dụng CSDL; các ứng dụng của CSDL; hệ CSDL; các mô hình dữ liệu và Hệ quản trị CSDL. Nội dung chính của chương bao gồm:

- Kiến thức cơ bản về cơ sở dữ liệu
- Hệ cơ sở dữ liệu
- Các mô hình dữ liệu
- Giới thiệu tổng quan về hệ quản trị CSDL

Mục tiêu cần đạt được của chương

Sinh viên hiểu được các kiến thức cơ bản về: cơ sở dữ liệu; ý nghĩa, vai trò và ứng dụng của cơ sở dữ liệu; các mô hình dữ liệu; hệ quản trị CSDL.

Bài 1: TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU (Số tiết: 3 tiết)

1.1 Cơ sở dữ liệu

1.1.1 Định nghĩa cơ sở dữ liệu

Chúng ta cùng xem xét hệ thống bán vé máy bay bằng máy tính. Dữ liệu lưu trữ trong máy tính bao gồm thông tin về hành khách, chuyến bay, đường bay, v.v... Mọi thông tin về mối quan hệ này được biểu diễn trong máy thông qua việc đặt chỗ của khách hàng. Vậy làm thế nào để biểu diễn được dữ liệu đó và để đảm bảo cho khách hàng đi đúng chuyến.

Một tập hợp có cấu trúc của dữ liệu được lưu trong máy tính, theo một cách đặc biệt nào đó có thể được tiếp cận theo những cách khác nhau được gọi là cơ sở dữ liệu.

Ví dụ 1.1: Để quản lý việc học tập trong một trường đại học, các dữ liệu là các thông tin về sinh viên, về các môn học, điểm thi... Chúng ta tổ chức các dữ liệu đó thành các bảng và lưu giữ chúng vào sổ sách hoặc sử dụng một phần mềm máy tính để lưu giữ chúng trên máy tính. Ta có một tập các dữ liệu có liên quan đến nhau và mang nhiều ý nghĩa, đó là một CSDL.

1.1.2 Phân loại cơ sở dữ liệu

Có nhiều kiểu phân loại CSDL. Một số loại CSDL thường được nhắc đến như: CSDL lớn, CSDL hướng tài liệu, CSDL hướng đối tượng, CSDL đồ thị, CSDL thời gian thực, CSDL tri thức, CSDL không gian, CSDL thời gian, CSDL tập trung, CSDL phân tán, CSDL đám mây, CSDL quan hệ, CSDL ngữ nghĩa... Với mỗi loại CSDL 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: là cơ sở 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 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: 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: 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 dạng tệp 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 quan hệ: 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 phân cấp: 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 chứa 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.