

SAU BÀI HỌC NÀY EM SẼ:

- Hiểu được mô hình CSDL quan hệ.
- Hiểu được các thuật ngữ và khái niệm liên quan: bản ghi, trường (thuộc tính), khoá, khoá chính, khoá ngoài, liên kết dữ liệu.



Trong bài trước các em đã biết khái niệm CSDL. Đã có khá nhiều mô hình CSDL khác nhau. Từ những năm 1970, Edgar Frank Codd (1923 – 2003) đã đề xuất mô hình CSDL quan hệ. Mô hình này nhanh chóng trở thành mô hình được dùng phổ biến nhất, nó xuất hiện trong hầu khắp các ứng dụng quản lí, kể cả trong các ứng dụng thư tin điện tử, mạng xã hội,... Vậy mô hình CSDL quan hệ là gì?

1. KHÁI NIỆM CƠ SỞ DỮ LIỆU QUAN HỆ

Hoạt động 1 Tìm hiểu về một CSDL thông tin âm nhạc

Một CSDL các bản nhạc, trên một website âm nhạc, được tổ chức như mô tả trong Hình 13.1, gồm có danh sách các tên nhạc sĩ với mã (định danh) là **Aid** (Hình 13.1a), danh sách các tên ca sĩ với mã (định danh) là **Sid** (Hình 13.1b), danh sách các bản nhạc với tên bản nhạc, mã nhạc sĩ (tác giả bản nhạc) và mã **Mid** – định danh bản nhạc (Hình 13.1c), danh sách các bản thu âm gồm có mã bản nhạc và mã ca sĩ (Hình 13.1d). Hãy quan sát và trả lời các câu hỏi sau:

1. Nhạc sĩ sáng tác bản nhạc "Trường ca sông Lô" là nhạc sĩ nào? Nhạc sĩ sáng tác bản nhạc "Xa khơi" là nhạc sĩ nào?
2. Bản thu âm trong Hình 13.1d tương ứng với dòng 0005 TN là bản thu âm của bản nhạc nào, do ca sĩ nào thể hiện?

Nhạc sĩ		Ca sĩ		Bản nhạc			Bản thu âm	
Aid	TenNS	Sid	TenCS	Mid	Aid	TenBN	Mid	Sid
1	Đỗ Nhuận	TK	Trần Khánh	0001	1	Du kích sông Thao	0001	TK
2	Văn Cao	LD	Lê Dung	0002	2	Trường ca sông Lô	0001	LD
3	Hoàng Việt	TN	Tân Nhân	0003	3	Tình ca	0005	TK
4	Nguyễn Tài Tuệ	QH	Quốc Hương	0004	4	Xa khơi	0005	TN
				0005	1	Việt Nam quê hương tôi	0004	QH
				0006	2	Tiến về Hà Nội	0005	QH

a)

b)

c)

d)

Hình 13.1. CSDL âm nhạc



CSDL này được xây dựng nhằm đáp ứng nhu cầu tìm kiếm các bản nhạc, bản thu âm lưu trữ trong hệ thống theo tên bản nhạc, tên nhạc sĩ hay tên ca sĩ. Có thể thấy tất cả các dữ liệu của CSDL này được tổ chức ở dạng các bảng gồm các hàng và các cột. Mỗi bảng lưu trữ dữ liệu của các đối tượng có các thuộc tính (được hiểu là các đặc tính xác định đối tượng) giống nhau, mỗi đối tượng một hàng. Dữ liệu của mỗi thuộc tính được ghi trong một cột. Ví dụ bảng Nhạc sĩ lưu trữ những thông tin liên quan đến nhạc sĩ với các cột thuộc tính là mã định danh (**Aid**) và tên nhạc sĩ (**TenNS**). Còn bảng Ca sĩ mang thông tin của các ca sĩ với các cột thuộc tính là mã định danh (**Sid**) và tên ca sĩ (**TenCS**).

Mỗi bảng có quan hệ với một số bảng còn lại thông qua một thuộc tính nào đó: bảng Bản nhạc và bảng Nhạc sĩ có chung thuộc tính **Aid**, bảng Bản thu âm và bảng Bản nhạc có chung thuộc tính **Mid**.

Từ các phân tích trên, có thể thấy Hình 13.2 chính là tóm tắt cách tìm câu trả lời cho các câu hỏi nêu trong Hoạt động 1.



Hình 13.2. Quan hệ giữa các bảng trong CSDL âm nhạc

Mô hình tổ chức dữ liệu thành các bảng dữ liệu của các đối tượng có các thuộc tính giống nhau, có thể có quan hệ với nhau theo cách tương tự như trên được gọi là *mô hình dữ liệu quan hệ* và các CSDL tương ứng được gọi là *CSDL quan hệ*.

Sau đây, chúng ta sẽ chỉ xem xét các CSDL quan hệ. Vì vậy để cho ngắn gọn, nếu không gây nhầm lẫn, ta sẽ dùng thuật ngữ CSDL thay cho CSDL quan hệ.

CSDL quan hệ là CSDL lưu trữ dữ liệu dưới dạng các bảng có quan hệ với nhau.



1. Hãy chỉ ra các cột của bảng Bản nhạc.
2. Bảng Bản thu âm và bảng Ca sĩ có chung thuộc tính nào?

2. MỘT SỐ THUẬT NGỮ, KHÁI NIỆM LIÊN QUAN



a) Bản ghi, trường

Về mặt cấu trúc, CSDL quan hệ tổ chức lưu trữ dữ liệu dưới dạng các *bảng* gồm các hàng và cột. Mỗi hàng của bảng được gọi là một *bản ghi* (record), là tập hợp các thông tin về một đối tượng cụ thể được quản lý trong bảng. Mỗi cột trong bảng được gọi là *trường* (field) thể hiện *thuộc tính* của đối tượng được quản lý trong bảng.

Nói cách khác mỗi hàng là một bản ghi chép dữ liệu cho một đối tượng, gồm một bộ giá trị ghi trong các trường tương ứng với các thuộc tính của đối tượng.

Ví dụ, bảng **Nhạc sĩ** có bốn bản ghi là (1, Đỗ Nhuận), (2, Văn Cao), (3, Hoàng Việt), (4, Nguyễn Tài Tuệ). Bảng này có hai trường dữ liệu là **Aid** và **TenNS**. Cũng nói mỗi bản ghi của bảng này có hai thuộc tính là **Aid** và **TenNS**.

Nhờ sự nhất quán về cấu trúc như thế, việc thực hiện các *thao tác dữ liệu* (cập nhật, truy xuất dữ liệu) sẽ có nhiều thuận lợi.

b) Khoá chính

Mỗi bảng có thể có một hay một nhóm trường mà giá trị của chúng tại các bản ghi không trùng nhau, xác định duy nhất một bản ghi, nói cách khác là bộ giá trị của chúng cho phép xác định phân biệt các bản ghi của bảng. Trường hay nhóm trường ấy được gọi là *khoá* của bảng.

Ví dụ, ở bảng **Nhạc sĩ**, trường **Aid** có giá trị phân biệt tại mỗi bản ghi của bảng, nó là khoá của bảng **Nhạc sĩ**.

Một bảng có thể có nhiều khoá. Người ta có thể chọn (chỉ định) một khoá trong các khoá đó làm *khoá chính* của bảng và thường chọn khoá có số trường ít nhất.

Ví dụ, ở bảng **Bản nhạc**, nhóm ba trường **Mid**, **Aid**, **TenBN** có giá trị không trùng nhau tại các bản ghi, chúng làm thành một khoá của bảng **Bản nhạc**. Mỗi bản nhạc cũng được phân biệt bởi **Aid** và **TenBN**, vì vậy nhóm hai trường **Aid** và **TenBN** cũng làm thành một khoá của bảng **Bản nhạc**. Mỗi giá trị của trường **Mid** cũng xác định phân biệt một bản nhạc vì vậy **Mid** cũng là một khoá của bảng **Bản nhạc**. Để chọn khoá chính của bảng này, rõ ràng chọn **Mid** là tốt hơn cả. Khoá chính xác định duy nhất bản ghi nên có vai trò quan trọng trong sắp xếp, truy xuất dữ liệu.

c) Khoá ngoài

Mỗi bảng (A) có thể có trường hay nhóm các trường (*k*) làm thành khoá chính ở một bảng khác (B). Khi đó *k* được gọi là *khoá ngoài* của bảng A. Hai bảng A và B được gọi là có *quan hệ* với nhau qua khoá ngoài *k* của bảng A.

Ví dụ: Trường **Aid** trong bảng **Bản nhạc** là một khoá ngoài của bảng này vì **Aid** là khoá chính ở bảng **Nhạc sĩ**.

d) Liên kết dữ liệu

Có thể dùng khoá ngoài của bảng để thực hiện ghép nối dữ liệu hai bảng với nhau. Người ta gọi việc ghép nối như thế là liên kết (join) dữ liệu theo khoá. Ví dụ: Có thể liên kết bảng **Bản nhạc** với **Nhạc sĩ** theo trường **Aid** để biết được tên nhạc sĩ sáng tác bản nhạc.

Nhạc sĩ		Bản nhạc			Bảng NB			
Aid	TenNS	Mid	Aid	TenBN	Mid	Aid	TenNS	TenBN
1	Đỗ Nhuận	0001	1	Du kích sông Thao	0001	1	Đỗ Nhuận	Du kích sông Thao
2	Văn Cao	0002	2	Trường ca sông Lô	0002	2	Văn Cao	Trường ca sông Lô
3	Hoàng Việt	0003	3	Tình ca	0003	3	Hoàng Việt	Tình ca
4	Nguyễn Tài Tuệ	0004	4	Xa khơi	0004	4	Nguyễn Tài Tuệ	Xa khơi
		0005	1	Việt Nam quê hương tôi	0005	1	Đỗ Nhuận	Việt Nam quê hương tôi
		0006	2	Tiến về Hà Nội	0006	2	Văn Cao	Tiến về Hà Nội

Hình 13.3. Mô tả liên kết dữ liệu hai bảng **Nhạc sĩ** và **Bản nhạc**

Nếu liên kết bảng **Bản thu âm** với bảng **Ca sĩ** theo khoá **Sid** và liên kết với **Bảng NB** theo khoá **Mid**, sẽ có được kết quả là một bảng với dữ liệu đầy đủ nhất.

Bản thu âm		Ca sĩ		Bảng NB			
Mid	Sid	Sid	TenCS	Mid	Aid	TenNS	TenBN
0001	TK	TK	Trần Khánh	0001	1	Đỗ Nhuận	Du kích sông Thao
0001	LD	LD	Lê Dung	0002	2	Văn Cao	Trường ca sông Lô
0005	TK	TN	Tân Nhân	0003	3	Hoàng Việt	Tình ca
0005	TN	QH	Quốc Hương	0004	4	Nguyễn Tài Tuệ	Xa khơi
0002	QH			0005	1	Đỗ Nhuận	Việt Nam quê hương tôi
0002	QH			0006	2	Văn Cao	Tiến về Hà Nội

↓

Bảng NBC					
Mid	Sid	TenCS	Aid	TenNS	TenBN
0001	TK	Trần Khánh	1	Đỗ Nhuận	Du kích sông Thao
0001	LD	Lê Dung	2	Văn Cao	Trường ca sông Lô
0005	TK	Trần Khánh	3	Hoàng Việt	Tình ca
0005	TN	Tân Nhân	4	Nguyễn Tài Tuệ	Xa khơi
0002	QH	Quốc Hương	1	Đỗ Nhuận	Việt Nam quê hương tôi
0002	QH	Quốc Hương	2	Văn Cao	Tiến về Hà Nội

Hình 13.4. Mô tả liên kết dữ liệu ba bảng

Quan sát **Bảng NBC** (Bản thu âm – Bản nhạc – Ca sĩ), em lại thấy những điều quen thuộc: Bản thu âm thứ hai là bản thu âm do ca sĩ Lê Dung thể hiện bản nhạc Trường ca sông Lô của nhạc sĩ Văn Cao, bản thu âm thứ năm là bản thu âm ca sĩ Quốc Hương thể hiện bản nhạc Việt Nam quê hương tôi của nhạc sĩ Đỗ Nhuận.

Lí do tạo CSDL âm nhạc với bốn bảng Nhạc sĩ, Ca sĩ, Bản nhạc, Bản thu âm như ở Hình 13.1 thay vì chỉ lập một bảng tương tự **Bảng NBC** có thể được nêu ngắn gọn như sau:

- Nếu chỉ lập một bảng như **Bảng NBC**, dữ liệu tên các ca sĩ, tên các nhạc sĩ và tên các bản nhạc có thể lặp lại nhiều lần với độ dài xâu kí tự khá lớn, gây ra dư thừa dữ liệu, tiêu tốn không gian lưu trữ và dễ mắc lỗi làm mất tính nhất quán của dữ liệu: chẳng hạn ở dòng trên ghi là Văn Cao, dòng dưới là Van Cao,... Mỗi lần sửa đổi tên một ca sĩ chẳng hạn cần phải tìm soát sửa đổi ở rất nhiều dòng khác.

- Bằng cách lập bốn bảng Nhạc sĩ, Ca sĩ, Bản nhạc, Bản thu âm như ở Hình 13.1, mỗi lần thay đổi một tên nhạc sĩ hay một tên ca sĩ ta chỉ cần sửa đổi một lần ở bảng **Nhạc sĩ** hay bảng **Ca sĩ** và khi thực hiện liên kết sẽ có được **Bảng NBC** với đầy đủ dữ liệu đã sửa đổi.

e) Các trường và dữ liệu

Hãy cùng xem xét một bảng phức tạp hơn - bảng điểm thi tiếng Anh ở một trung tâm ngoại ngữ.

Bảng 13.1. Bảng điểm thi tiếng Anh

SBD	Họ và tên	Giới tính	Ngày sinh	Điểm	Mã kì thi
1	Phan Việt An	Nam	17/10/2005	6,25	A1001
2	Trần Duy Anh	Nam	11/09/2004	8,00	A1001

3	Bùi Thị Hương	Nữ	20/03/2006	9,50	A1001
4	Đỗ Hồng Hoa	Nữ	15/09/2005	7,75	A1001
5	Nguyễn Văn Khoa	Nam	22/11/2003	7,25	A1001
	...				

Giả sử người ta chỉ thiết lập một bảng trong CSDL để lưu trữ dữ liệu từ bảng điểm trên, với tên bảng là **bangdiem** và các trường tương ứng là **sbd**, **hoten**, **gt**, **ngaysinh**, **diem**, **makithi**. Chúng ta cùng phân tích các đặc điểm dữ liệu trong các trường này.

- Trường **sbd** có các giá trị là các số nguyên.
- Trường **hoten** có các giá trị là chuỗi ký tự, độ dài giới hạn, ví dụ không quá 64 ký tự.
- Trường **gt** chỉ có hai giá trị là Nữ hoặc Nam, có thể xem nó là trường có giá trị logic là đúng (1) hoặc sai (0) tương ứng thay thế cho các giá trị là Nữ hoặc Nam.
- Trường **ngaysinh** có các giá trị là ngày/tháng/năm.
- Trường **diem** có các giá trị kiểu số thập phân có tối đa 5 chữ số với 2 chữ số sau dấu phẩy thập phân.
- Trường **makithi** có các giá trị kiểu chuỗi với độ dài 5 ký tự.

Những phân tích trên cho thấy mỗi trường có các dữ liệu cùng một kiểu. Chúng được gọi là kiểu dữ liệu của trường. Trường **sbd** có kiểu số nguyên, trường **hoten** có kiểu chuỗi ký tự độ dài không quá 64 ký tự, trường **gt** có kiểu logic, trường **ngaysinh** có kiểu ngày tháng, trường **diem** có kiểu số thập phân, trường **makithi** có kiểu chuỗi ký tự độ dài cố định 5 ký tự.

Việc xác định kiểu dữ liệu của các trường có mục đích:

- Hạn chế việc lãng phí dung lượng lưu trữ dữ liệu.
- Kiểm soát tính đúng đắn về logic của dữ liệu được nhập vào bảng. Chẳng hạn sẽ không thể nhập các ký tự dạng chữ vào trường **sbd**, không thể nhập giá trị số vào trường **ngaysinh**,...



1. Hãy chỉ ra khoá chính của bảng Ca sĩ và bảng Bản nhạc.
2. Hãy chỉ ra các khoá ngoài của bảng Bản nhạc và bảng Bản thu âm.



LUYỆN TẬP

Cho CSDL học tập có các bảng sau: **Hocsinh** (họ tên, số CCCD, số thẻ học sinh, ngày sinh, địa chỉ), **monhoc** (tên, mã môn), **Diem** (số thẻ học sinh, mã môn, năm, học kì, loại điểm, điểm), trong đó loại điểm chỉ các loại ĐDG thường xuyên, ĐDG giữa kì, ĐDG cuối kì.

Hãy xác định các khoá chính và các khoá ngoài của từng bảng, có thể lấy số CCCD làm khoá chính được không.



VẬN DỤNG

Trong kì thi tốt nghiệp trung học phổ thông, học sinh được đánh số báo danh, có thể thi một số môn, được chia vào các phòng thi được đánh số, sau khi chấm sẽ có điểm thi với các môn đăng kí dự thi. Em hãy đề xuất một số bảng dữ liệu và các trường làm khoá chính và khoá ngoài cho các bảng đó.